

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

---

**На правах рукописи**

**ШЕВЧЕНКО АЛЕКСАНДР ПЕТРОВИЧ**

**ТЕОРИЯ И МЕТОДЫ КОМПЬЮТЕРНОГО  
ГЕОМЕТРИКО-ТОПОЛОГИЧЕСКОГО АНАЛИЗА И  
ПРОГНОЗИРОВАНИЯ СТРОЕНИЯ И ФИЗИЧЕСКИХ СВОЙСТВ  
КООРДИНАЦИОННЫХ СОЕДИНЕНИЙ**

**1.4.4. Физическая химия**

**ДИССЕРТАЦИЯ  
на соискание ученой степени  
доктора химических наук**

**Научный консультант:  
д.х.н., профессор Блатов В.А.**

**Самара – 2025**

## Оглавление

ВВЕДЕНИЕ .....	5
ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ И СОКРАЩЕНИЙ .....	13
Глава 1. Обзор литературы.....	14
1.1. Хранилища кристаллоструктурных данных .....	14
1.2. Модели строения кристаллических соединений.....	17
1.2.1. Разнообразие моделей.....	17
1.2.2. Модель жестких сфер .....	18
1.2.3. Полиэдрическое представление атомов.....	19
1.2.4. Методы расчета и хранения связанности атомов .....	24
1.3. Кристаллохимические дескрипторы .....	28
1.3.1. Химические дескрипторы .....	30
1.3.2. Геометрические дескрипторы.....	33
1.3.3. Топологические дескрипторы.....	36
1.3.4. Дескрипторы свободного пространства .....	48
1.4. Хранилища структурных данных и их дескрипторов .....	52
1.4.1. Интеллектуальные и экспертные системы .....	54
1.4.2. Топологические библиотеки .....	57
1.4.3. Библиотеки металл-органических координационных полимеров .....	62
1.5. Анализ структурных данных машинными методами.....	63
1.5.1. Поверхности потенциальной энергии.....	64
1.5.2. Машинный прогноз степени окисления .....	65
1.5.3. Дизайн координационных соединений.....	67
1.5.4. Общая схема обработки и использования кристаллографических данных.....	70
Глава 2. Программный пакет ToposPro.....	74
2.1. Общее описание.....	74
2.1.1. Система управления структурными базами данных .....	74
2.1.2. Прикладные программы ToposPro.....	75
2.2. Методы и дескрипторы для распознавания формы координационных фигур .....	80

2.2.1. Обозначение координационной фигуры.....	80
2.2.2. Топология полиэдров Вороного .....	81
2.2.3. Максимизация объема пересечения координационных полиэдров	84
2.2.4. Сравнение угловых фингерпринтов .....	86
2.3. Определение свободного пространства в структуре кристаллов .....	90
2.3.1. Оценка доли свободного пространства.....	90
2.3.2. Генерация подпространства пустот.....	90
2.3.3. Определение геометрических характеристик пустот.....	94
2.3.4. Топология системы каналов и расположения пустот.....	97
2.4. Топологические библиотеки .....	97
2.4.1. Библиотеки топологических типов .....	97
2.4.2. Библиотеки структурных единиц .....	99
2.5. Модули для обработки и анализа данных.....	100
2.6. Общая схема анализа данных.....	101
Глава 3. Корреляции «состав-структура-свойство» в координационных полимерах.....	105
3.1. Координационные фигуры атомов меди и цинка .....	105
3.1.1. Выборка.....	105
3.1.2. Определение формы координационной фигуры.....	106
3.1.3. Правила и закономерности.....	107
3.2. Прогноз размерности и топологии кристаллических структур машинными методами .....	115
3.2.1. Выборка.....	115
3.2.2. Выбор оптимального машинного метода анализа данных .....	123
3.2.3. Применение полученных результатов .....	127
3.3. Прогноз свободного пространства в структуре металлоорганических каркасных соединениях.....	135
3.3.1. Выборка.....	135
3.3.2. Топологическое устройство кристаллов.....	138
3.4. Анализ и прогноз степени окисления металлов.....	148

3.4.1. Выборка.....	149
3.4.1. Методы машинного анализа .....	154
3.4.2. Результаты и обсуждение .....	156
3.5. Схема топологической сборки кристаллов.....	163
3.5.1. Выборка.....	163
3.5.2. Топологическое устройство кристаллов.....	165
3.5.3. Сборка каркасных структур .....	168
3.5.4. Сборка молекулярных структур .....	171
Глава 4. Геометрический и топологический анализ кристаллических структур с иной природой связывания атомов.....	173
4.1. Изменение размера атомов инертных газов при фазовых переходах.....	173
4.2. Динамика изменения атомных дескрипторов и Периодический закон .....	182
4.3. Использование полиэдров Вороного для построения карты миграции подвижных катионов .....	189
4.4. Использование полиэдров Вороного для анализа кристаллических структур интерметаллидов.....	191
4.5. Использование полиэдров Вороного для анализа строения кристаллов насыщенных углеводородов .....	193
Заключение .....	207
Библиографический список .....	209
Приложение .....	239



## ВВЕДЕНИЕ

**Актуальность темы исследования и степень ее разработанности.** Возможности современной экспериментальной техники и программного обеспечения позволяют быстро и с высокой точностью определять кристаллическую структуру практически любого химического соединения. Экспериментальная структурная информация аккумулируется в электронных хранилищах, наиболее известные из которых – Кембриджская структурная база данных (CSD), База данных неорганических структур (ICSD), Файл порошковых данных (PDF4), Банк белковых данных (PDB), База кристаллических данных Пирсона (PCD). На начало 2024 года они содержат сведения о строении более 2 млн. кристаллических веществ.

Несмотря на уже длительную историю своего существования (первая из указанных баз данных, CSD, была основана в 1965 г.) электронные хранилища в течение длительного времени использовались только как справочники. Ситуация качественно изменилась, когда в 2011 г. в США была принята программа «Геном материалов», основной задачей которой является ускорение разработки новых материалов за счет внедрения компьютерных методов и инструментов прогнозирования. В результате ее выполнения планируется сократить срок разработки новых материалов с 10-15 лет до 2-3 лет и существенно уменьшить стоимость этого процесса. Для осуществления указанных задач необходимо создание новых методов моделирования строения и свойств химических веществ, аккумулирование накопленной химической и структурной информации в единых электронных базах данных, использование современных методов машинной обработки и машинного обучения для анализа собранных данных в целях извлечения из них новых полезных знаний («вторичной» информации). Эта информация, собранная в базы знаний, станет основанием будущих систем искусственного интеллекта (ИИ), способных с высокой точностью прогнозировать структуру и свойства веществ и материалов.

Именно разработка баз знаний, соответствующих методов и программного инструментария становится общемировым трендом, в русле которого лежит данная работа. В настоящее время существует лишь несколько примеров баз знаний по кристаллохимии и материаловедению, а системы ИИ все еще находятся в стадии разработки. Одной из таких баз знаний является Mogul, которая интегрирована в систему CSD. Однако разработка баз знаний существенно ограничивается необходимостью подготовки надежных исходных кристаллографических данных, создания новых дескрипторов для характеристики особенностей кристаллической структуры и ее свойств, а также разработки специальных методов и инструментов для создания баз знаний из исходных наборов кристаллографических данных. CSD является хорошим примером тщательной проработки этих проблем в течение длительного времени, однако даже информация из CSD требует дополнительной подготовки для машинного анализа. База знаний Mogul содержит параметры только локального окружения молекул в кристаллах, но не рассматривает свойства структуры как целого, в частности, способ связанности структурных строительных блоков. Кроме того, Mogul ориентирована только на молекулярные кристаллы и не содержит дескрипторов для полимерных структур, как неорганических, так и координационных. Разработка новых дескрипторов, в свою очередь, требует развития теоретического базиса описания и моделирования кристаллических веществ.

**Целью работы** являлась разработка теоретических моделей строения кристаллических веществ, алгоритмов и методов автоматизированной обработки объемных выборок кристаллографических данных для поиска закономерностей «состав-структура-свойство» кристаллов, а также прогнозирования их строения и физических свойств на примере координационных соединений (КС) и их важнейшего подкласса, металл-органических координационных полимеров (МОКП).

Для достижения указанной цели были поставлены следующие **задачи**:

- разработка теоретических основ геометрического и топологического моделирования атомных ансамблей и свободного пространства в кристаллических структурах с помощью их представления в виде периодических графов и разбиения Вороного;
- создание системы геометрико-топологических дескрипторов кристаллической структуры и ее структурных строительных блоков (ССБ) для прогнозирования свойств и поиска закономерностей строения простых веществ, твердых электролитов, КС и МОКП;
- разработка универсальных алгоритмов и программного обеспечения для автоматизированного расчета геометрико-топологических дескрипторов в КС и МОКП любого химического состава и строения;
- разработка методов машинного поиска корреляций между геометрико-топологическими дескрипторами, химическим составом КС и его физическими свойствами, а также создание системы баз знаний, содержащих найденные корреляции;
- разработка на основе корреляций, содержащихся в созданных базах знаний, схем прогнозирования для дизайна КС, обладающих заданной структурой и физическими свойствами.

**Положения, выносимые на защиту:**

- универсальный геометрико-топологический подход к моделированию и анализу кристаллической структуры, основанный на представлении ее атомов в виде полиэдров Вороного или Ван-дер-Ваальсовых полиэдров, связанных в трехмерную периодическую сетку, и позволяющий проводить классификацию систем химических связей, поиск взаимосвязей «химический состав – структура – свойство» и

прогнозирование новых структур в соответствии со строгими алгоритмами;

- совокупность геометрико-топологических дескрипторов и методов их расчета для описания структурной роли атомов и ССБ, а также строения и физических свойств кристаллических структур;
- правила, принципы и закономерности, сформулированные на основании геометрико-топологического подхода, в том числе: классификация лигандов по форме их координационных фигур, схема машинного анализа структурных данных на основе геометрико-топологических дескрипторов, численный критерий оценки жесткости катионов в структуре КС и МОКП, условия миграции катиона в структуре твердого электролита, взаимосвязи радиуса сферического домена атомов в структуре простых веществ с длинами волн де Бройля валентных электронов, рассчитанными из потенциалов ионизации изолированных атомов;
- топологические алгоритмы, программное обеспечение и методы анализа кристаллических веществ, основанные на понятии о множественности топологических представлений кристаллической структуры и способов ее топологического упрощения;
- базы знаний химических, геометрических и топологических дескрипторов простых веществ, а также известных на настоящий момент структур КС и МОКП, библиотеки и методы сборки их ССБ, рассчитанные на основе геометрико-топологического подхода, реализованного в программном комплексе ToposPro и интернет-сервисах TopCryst;
- прогностическая модель для оценки степени окисления атомов-комплексобразователей в кристаллической структуре;

- схема прогнозирования и дизайна структуры МОКП, опирающаяся на корреляции между его локальными и глобальными топологическими дескрипторами;
- метод анализа свободного пространства в кристаллической структуре, основанный на представлении атомов в виде полиэдров Вороного и Ван-дер-Ваальсовых полиэдров и позволяющий прогнозировать сорбционные свойства и ионную проводимость кристалла.

**Научная новизна.** Создано новое междисциплинарное направление, объединяющее методы и модели кристаллохимии и материаловедения с информационными технологиями и ориентированное на создание универсальных систем ИИ, пригодных для прогнозирования структуры и свойств кристаллических веществ любой природы. Впервые объединены топологический и геометрический подходы при анализе и моделировании структур КС и МОКП. Развита теория топологических представлений кристаллических структур, в которой формализованы методы выделения ССБ в зависимости от их химической природы и топологических свойств для изучения и контроля механизмов их сборки. Предложена схема прогнозирования ряда топологических свойств структур КС и МОКП, в том числе размерности и топологии их базовой сетки на основании разработанных количественных методов определения формы координатных фигур ССБ. С использованием современных алгоритмов машинного обучения разработана прогностическая модель для оценки степени окисления атомов-комплексобразователей в кристаллической структуре, исходя из геометрических дескрипторов их полиэдров Вороного, а также положения в Периодической системе Д. И. Менделеева. Впервые на примере кристаллов простых веществ исследованы и проанализированы взаимосвязи радиуса сферического домена атомов с длинами волн де Бройля валентных электронов, рассчитанными из потенциалов ионизации изолированных атомов. Найденные корреляции позволяют объяснить

геометрическую модель атома в структуре кристаллов, как упаковку деформируемых, но несжимаемых сфер, с позиции электронного строения атома.

**Теоретическая и практическая значимость работы.** Разработанный геометрико-топологический подход к анализу и прогнозированию кристаллических структур и их свойств активно используется исследователями во всем мире. Созданные базы данных топологических типов и ССБ, а также разработанные на их основе сервисы, востребованы при топологическом и кристаллохимическом описании новых синтезируемых КС и МОКП, при поиске закономерностей строения кристаллических веществ, анализе свободного пространства и скрининге структурной информации для отбора потенциальных твердых электролитов. На конец 2023 г. программный комплекс ToposPro, в котором реализован указанный подход, используют более 7800 исследователей из 103 стран мира; ежегодно публикуются около 500 научных статей, в которых ToposPro применяется для анализа кристаллических веществ. Созданные топологические базы данных и использующие их сервисы, доступны на сайте <https://topocryst.com>; на конец 2023 г. указанные сервисы были применены для анализа более 33 тыс. кристаллических структур пользователями из 74 стран мира.

**Методология и методы исследования.** В качестве экспериментальных методов исследования в диссертационной работе использовались методы геометрико-топологического анализа кристаллоструктурных данных большого объема. Математическая обработка рассчитанных данных осуществлялась методами машинного обучения, такими как случайный лес, метод опорных векторов, дерево решений, К-ближайших соседей. Для обработки экспериментальных данных и построения моделей интеллектуального анализа использовалась среда разработки Python's Integrated Development Environment, а также программные пакеты ToposPro, Excel, PostgreSQL.

**Личный вклад автора.** Выбор темы, постановка задач и целей исследования, обсуждение и обобщение всех полученных результатов, формулировка научных положений и выводов, которые выносятся на защиту, принадлежат лично автору. Все вычислительные эксперименты, описанные в диссертации, выполнены автором в сотрудничестве с коллегами Международного научно-исследовательского центра по теоретическому материаловедению (МНИЦТМ, г. Самара).

**Степень достоверности и апробация результатов.** Достоверность полученных теоретических результатов подтверждается их воспроизводимостью в численных экспериментах, проведенных как автором, так и сторонними научными группами, а также большими объемами выборок экспериментальных структурных данных, используемых для машинного обучения прогностических моделей.

Основные положения работы и результаты докладывались на Национальных кристаллохимических конференциях (Черноголовка 2003, Суздаль 2016, Суздаль 2018, Чеget 2021, Нальчик 2024), II Международной научной конференции «Наука будущего» (Казань 2016), IUCr 2017 (Hyderabad, India), Intermetallics 2017 (Bad Staffelstein, Germany), IV Международной конференции по информационным технологиям и нанотехнологиям (Самара 2018), 9-й Международной научной конференции «Numerical Geometry, Grid Generation and Scientific Computing» (Москва 2018), XXI Менделеевском съезде по общей и прикладной химии (Санкт-Петербург 2019), 1-м Международном симпозиуме по компьютерному прогнозированию структуры и перспективным материалам (Xi'an 2019, China), XII Международной научной конференции «Кинетика и механизм кристаллизации. Кристаллизация и материалы нового поколения» (Иваново 2023), 7-й Международной конференции «Глубокое обучение в вычислительной физике» (Санкт-Петербург 2023), XXV Международной

конференции по анализу и управлению данными в областях с интенсивным использованием данных (Москва 2023).

**Публикации.** По материалам диссертации опубликованы 2 главы в монографиях и 35 статей в рецензируемых журналах. Получены 18 свидетельств о регистрации программ для ЭВМ и баз данных.

**Объем и структура работы.** Диссертационная работа состоит из введения, обзора литературы, экспериментальной части, обсуждения результатов, выводов, заключения, списка литературы (284 источника) и приложения (изложенного на 15 страницах и содержащего 12 таблиц, 2 рисунка). Текст диссертационной работы изложен на 204 страницах, включает 36 таблиц и 91 рисунок.



## ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ И СОКРАЩЕНИЙ

ВСЕ – вторичная строительная единица

КП – координационный полиэдр

КС – координационное соединение

КФ – координационная фигура

МНК – метод наименьших квадратов

МОКП - металлоорганический координационный полимер

ПВ – полиэдр Вороного

СГ – свернутый граф

CIF – Crystallographic Information File

CN – Coordination number

CSD – Cambridge Structural Database

EPINET – Euclidean Patterns in Non-Euclidean Tilings

ICSD – Inorganic Crystal Structure Database

PDB – Protein Databank

RCSR – Reticular Chemistry Structure Resource

$R_{SD}$  – радиус сферического домена, объем которого равен  $V_{VP}$

$S_{VP}$  – площадь поверхности граней полиэдра Вороного

SD – spherical domain

$V_{VP}$  – объем полиэдра Вороного

VP – Voronoi polyhedron

w – встречаемость в процентах от общего числа записей

$\Omega$  – телесный угол грани полиэдра Вороного

## Глава 1. Обзор литературы

### 1.1. Хранилища кристаллоструктурных данных

Кристаллографическая информация о кристаллических структурах является одним из наиболее организованных и полных хранилищ данных в естественных науках. Со второй половины 1960-х годов эти данные начали преобразовываться из печатных справочников в электронные базы данных; к настоящему времени известно семь основных источников кристаллографических данных [1-7], которые охватывают все классы химических соединений, от металлов до белков. Рассмотрим подробнее среди них два источника, содержащих информацию о координационных соединениях.

Cambridge Structural Database (CSD) [1] это всемирное хранилище низкомолекулярных органических и металлоорганических кристаллических структур. CSD постоянно дополняется новыми структурами (более 50 тыс. новых структур ежегодно, Рис. 1). На ноябрь 2023 года CSD содержит 1 240 тыс. записей. В систему CSD включен набор программных инструментов, необходимых для поиска, анализа кристаллических структур, а также для экспорта данных в другие приложения. CSD поддерживается Cambridge Crystallographic Data Centre (CCDC).

Inorganic Crystal Structure Database (ICSD) [2] является крупнейшей в мире базой данных полностью расшифрованных неорганических кристаллических структур (Рис. 2). На ноябрь 2023 года ICSD содержит более 220 тыс. рецензируемых записей структурных данных. Эта база данных поддерживается центром Технической Информации в г. Карлсруе (Германия), а также Министерством торговли США.

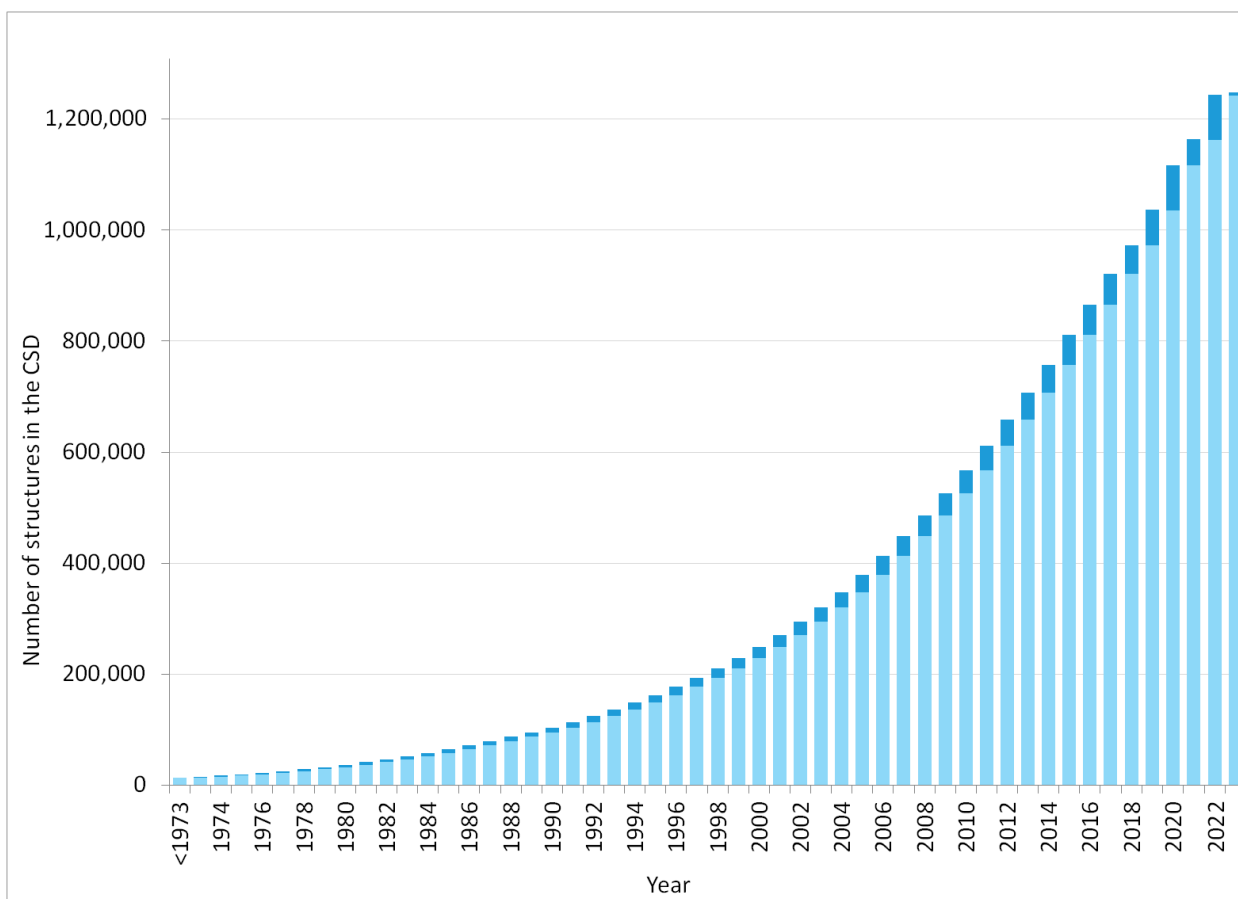


Рис. 1. Динамика общего количества записей в CSD по годам.

Каждая база данных имеет свой уникальный формат. Для обмена структурной информацией между ними, а также при передаче ее сторонним программным продуктам используется файлы CIF-формата. Crystallographic Information File (CIF) - это стандартный текстовый формат файла для обмена кристаллографической информацией, утвержденный Международным союзом кристаллографов. Формат файла был разработан Холлом, Алленом и Брауном [8], а затем пересмотрен и усовершенствован [9]. Подробное описание CIF-формата доступно на веб-сайте IUCr<sup>1</sup>.

<sup>1</sup> <https://www.iucr.org/resources/cif/spec>

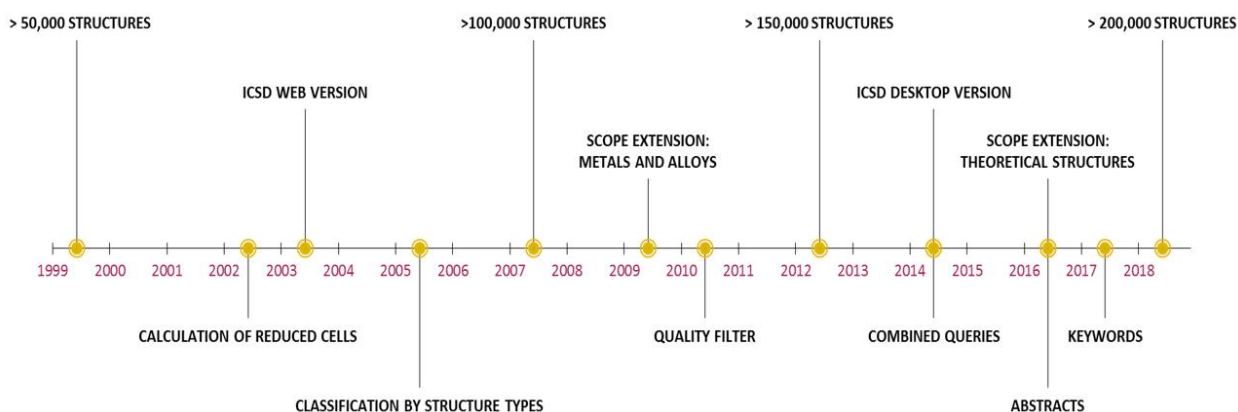


Рис. 2. Временная линия роста ICSD [2].

Как известно, чем больше объем накопленной информации, тем больше возможностей появляется для добычи знаний и поиска скрытых закономерностей [10, 11, 12]. Несмотря на новые возможности, эти базы данных часто используются пользователями только как электронные справочники. CSD наиболее активно развивает свои программные продукты, предназначенные для изучения геометрии ближайшего окружения атомов и выявления их предпочтительных взаимодействий (Mercury), анализа и инженерии материалов (Mogul), проверки соответствий белок-лиганд (GOLD, Hermes) и прогнозирования лекарственных препаратов на основе базы знаний (SuperStar). Кроме того, разработчики CSD предлагают пакетную выгрузку информации о кристаллических структурах в табличном виде (так называемые tab-файлы), а также API-интерфейс для прямого доступа к базе данных CSD из сторонних программных продуктов.

## 1.2. Модели строения кристаллических соединений

### 1.2.1. Разнообразие моделей

Кристаллическая структура - это весьма сложная система с большим числом степеней свободы, поэтому она трудно исследуема и плохо предсказуема. Чтобы найти закономерности ее строения и свойств, требуется упрощение и представление кристаллической структуры в виде некоторой геометрической, топологической или квантово-химической модели, а ее структурных единиц - в виде определенных геометрических, топологических или квантово-механических объектов, соответственно. Например, если структурные единицы кристалла представлены в виде сфер или полиэдров, то им отвечают геометрические модели жестких [13, 14] и мягких сфер [15, 16, 17], соответственно. Если моделью кристаллической структуры является сетка, тогда структурными единицами будут координационные фигуры [18]. Квантово-химической модели отвечает представление кристаллической структуры в виде ансамбля электронов и атомных ядер. Хотя эта модель является наиболее точным представлением кристалла, но она самая сложная в трактовании, расчет ее весьма затратный и требует специального программного обеспечения. Эта модель может быть использована только после введения определенных приближений и допущений, что вносит свои неоднозначности в рассчитанные ею характеристики. По перечисленным выше причинам квантово-химическая модель не подходит для обработки больших массивов информации и не рассматривается в настоящем исследовании. Нами будет только выполнено сравнение различных систем атомных радиусов с длинами волн де Бройля [19], рассчитанным из потенциалов ионизации изолированных атомов. Напротив, расчет геометрических и топологических представлений кристаллической структуры проще, быстрее и однозначнее в трактовании. Рассмотрим эти два варианта подробнее.

### 1.2.2. Модель жестких сфер

В модели жестких сфер [20] атомы представляют в виде шаров определенного радиуса. Хорошо известны системы атомных, слейтеровских, ван-дер-Ваальсовых, ионных и других типов радиусов [21]. Каждая из таких систем весьма успешно используется для кристаллических структур с определенным типом межатомных связей, хотя большая часть известных структур содержит связи разных типов. Кроме радиуса атома, основными дескрипторами модели жестких сфер являются межатомные расстояния, валентные и торсионные углы.

Так как квантово-химическая и геометрическая модели описывают один и тот же объект, кристаллическую структуру, они должны быть взаимосвязаны друг с другом. Попытки найти взаимосвязи этих моделей предпринимались, практически, с момента их создания. Для выяснения общей составляющей моделей обычно исходят из положения основного максимума электронной плотности атомной орбитали [22] или длины волны де Бройля  $\lambda_{\text{дБ}}^i$ , рассчитанной из  $i$ -го потенциала ионизации изолированного атома  $I_i$  по уравнению (1) [19].

$$\lambda_{\text{дБ}}^i = \frac{12.26}{\sqrt{I_i}} \quad (1)$$

Эту формулу легко получить из выражения для кинетической энергии электрона ( $E$ )

$$E = \frac{p^2}{2m_e} \quad (2)$$

и уравнения де Бройля

$$\lambda = \frac{h}{p} \quad (3)$$

где  $m_e$  – масса электрона,  $p$  – его импульс,  $\lambda$  – длина волны де Бройля, а  $h$  – постоянная Планка. Учитывая потенциалы ионизации атома разных порядков  $I_i$  [23], можно охарактеризовать каждый атом набором длин волн де Бройля

$\lambda_{дв}^i$  (**Таблица 1**). В статье [19] длины волн де Бройля были использованы для оценки ван-дер-Ваальсовых радиусов атомов.

**Таблица 1.** Рассчитанные из потенциалов ионизации различного порядка [23] длины электронных волн де Бройля для атомов инертных газов и щелочных металлов

Атом А	$\lambda_B^I, \text{\AA}$	$\lambda_B^{II}, \text{\AA}$	$\lambda_B^{III}, \text{\AA}$	Атом	$\lambda_B'^I, \text{\AA}$	$\lambda_B'^{II}, \text{\AA}$
He	2.47	1.66	—	Li	5.28	1.41
Ne	2.64	1.92	1.54	Na	5.41	1.78
Ar	3.09	2.33	1.92	K	5.89	2.18
Kr	3.28	2.48	2.02	Rb	6.00	2.35
Xe	3.52	2.68	2.20	Cs	6.21	2.55
Rn	3.74	—	—	Fr	—	2.59

Жесткие шары заполняют пространство кристалла лишь частично, что дает возможность оценить долю свободного пространства кристалла. Например, для плотнейшей упаковки равных шаров, которой описываются структуры многих металлов, доля заполнения пространства составляет 0.7405. *Упаковкой* [24] называют заполнение пространства некоторыми геометрическими телами без их перекрывания. Если принять во внимание, что атомы в структуре кристаллов могут быть представлены деформируемыми сферами, то результатом их деформации в кристаллическом поле соседних атомов могут быть полиэдры.

### 1.2.3. Полиэдрическое представление атомов

Учитывая известный принцип, гласящий, что природа не терпит пустоты, допустимо предположение о полном заполнении атомами кристаллического пространства, то есть его представление в виде *разбиения* [25]. При таком подходе чаще всего используется разбиение Вороного, важным преимуществом которого является однозначность построения для любой кристаллической решетки. При этом все пространство делится на полиэдры Вороного (ПВ) так, чтобы каждая их грань расположена на

плоскости, равноудаленной от двух соседних атомов, ее образующих (Рис. 3 слева). Ребра ПВ лежат на линиях, равноудаленных не менее, чем от трех соседних атомов, тогда как вершины равноудалены не менее, чем от четырех соседних атомов. Минимальное ближайшее окружение соответствует общему асимметричному случаю. Для ребер и вершин число ближайших атомов может быть больше в особых симметричных случаях. Сетка Вороного, построенная из таких вершин и ребер, отображает пространство пустот, окружающих атомы, поэтому ее анализ позволяет рассчитать геометрические и топологические дескрипторы свободного пространства. Разбиение Вороного активно используется при анализе кристаллических материалов [26] и их пустот [27], а также мембран [28], или в качестве инструмента для исследования путей миграции ионов в кристаллах [29].

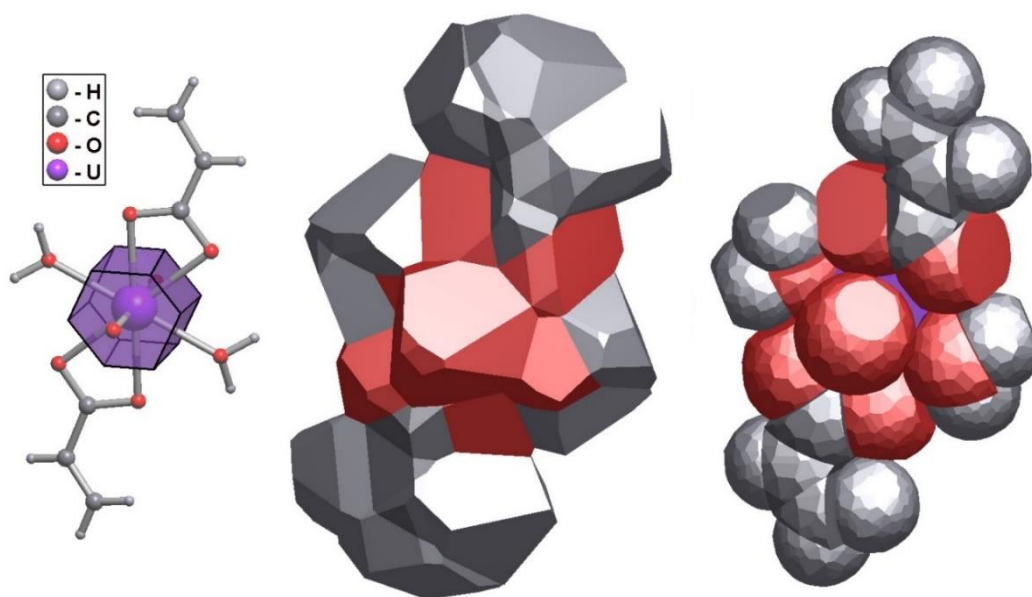


Рис. 3. Молекулярный фрагмент структуры  $\text{UO}_2(\text{C}_3\text{H}_3\text{O}_2)_2(\text{H}_2\text{O})_2$  [30] (NAFRIY<sup>2</sup>, слева), где  $\text{C}_3\text{H}_3\text{O}_2^-$  - акрилат анион, с ПВ атома урана и представления этого фрагмента в виде ПВ (в центре) или ван-дер-Ваальсовых полиэдров (справа). Доля незанятого атомами пространства в этой структуре составляет 33.5%.

<sup>2</sup> Здесь и далее приводятся справочные коды определений структуры кристаллов в CSD или ICSD.



Важно, что грани в ПВ некоторого атома А образуют лишь те соседние атомы В, взаимодействие которых с атомом А наиболее существенно. При этом число граней ПВ атома А в общем случае больше, чем его координационное число (CN) вследствие того, что некоторые грани отвечают атомам В, которые не образуют валентных связей с центральным атомом полиэдра (**Таблица 2**, Рис. 4).

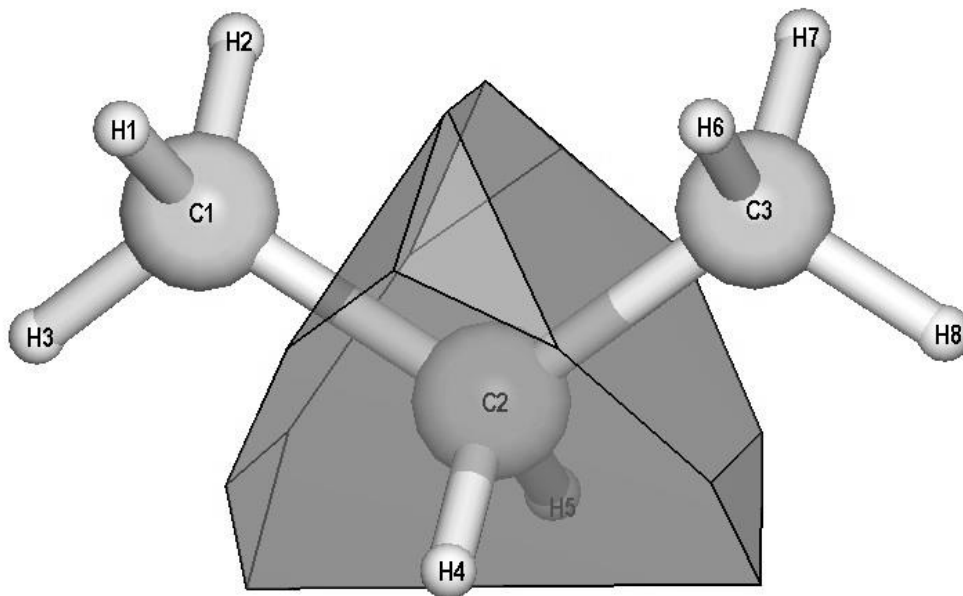


Рис. 4. ПВ атома углерода C(2) в структуре пропана [31] (JAYDUI).

Валентным связям А-В обычно отвечают короткие расстояния  $r(A-B)$  и большие  $\Omega(A-B)$ , причем отрезок А-В всегда пересекает соответствующую ему грань ПВ (прямой контакт). Невалентным же контактам  $A \cdots B$  соответствуют сравнительно длинные  $r(A \cdots B)$  и маленькие  $\Omega(A \cdots B)$ , при этом отрезок  $A \cdots B$  может и не пересекать соответствующую ему грань ПВ (непрямой контакт). Межатомные контакты А-В или  $A \cdots B$  в ПВ можно классифицировать также по числу химических связей в кратчайшей цепи, соединяющей атомы А и В в структуре кристалла. Примем, что число таких связей определяет ранг ( $\tau$ ) межатомного контакта. Тогда для любых валентных связей  $\tau=1$  (например, C(2)-C(3), C(2)-H(5), H(2)-C(1) и др., **Таблица 2**), а для атомов А и В, входящих в состав цепи А-D-B,  $\tau=2$

(контакты C(2)···H(2), H(2)···H(1) и др., **Таблица 2**) и т.д. Если грань ПВ атома А образуется атомом В из другой молекулы (цепи или слоя), то ранг контакта А···В считается равным нулю, поскольку отсутствует цепь связей, соединяющих атомы А и В (контакты H(2)···H(2), H(2)···H(8) и др., **Таблица 2**) в структуре соединения.

**Таблица 2.** Характеристики ПВ атомов C(2) и H(2) в структуре пропана (JAYDUI)\* [31]

Центральный атом: C(2)									
Характеристики ПВ: $R_{SD}=0.910 \text{ \AA}$ ; $V_{ПВ}=3.2 \text{ \AA}^3$ ; $G_3=0.1015$									
№	Атом	r, Å	Ω, %	τ	№	Атом	r, Å	Ω, %	τ
1	H(5)	0.969	28.20	1	6	H(6) #	2.080	1.78	2
2	H(4)	0.985	27.80	1	7	H(7) #	2.088	1.88	2
3	C(3)	1.530	17.24	1	8	H(3) #	2.094	1.14	2
4	C(1)	1.531	17.29	1	9	H(1) #	2.099	1.61	2
5	H(2) #	2.079	1.97	2	10	H(8) #	2.101	1.08	2
Центральный атом: H(2)									
Характеристики ПВ: $R_{SD}=1.366 \text{ \AA}$ ; $V_{ПВ}=10.7 \text{ \AA}^3$ ; $G_3=0.0983$									
№	Атом	r, Å	Ω, %	τ	№	Атом	r, Å	Ω, %	τ
1	C(1)	0.963	26.45	1	10	H(8)	2.861	6.68	0
2	H(1) #	1.552	8.58	2	11	H(1)	2.891	7.44	0
3	H(3) #	1.561	8.78	2	12	H(8)	2.996	5.13	0
4	C(2) #	2.079	1.97	2	13	H(6) #	3.034	0.04	4
5	H(5) #	2.403	3.33	3	14	H(3)	3.048	4.49	0
6	H(7)	2.586	7.69	4	15	H(6)	3.082	5.08	0
7	C(3) #	2.735	0.02	3	16	H(3) #	3.549	0.12	0
8	H(2)	2.769	7.23	0	17	H(7) #	3.847	0.06	0
9	H(1)	2.804	6.89	0	18	H(4) #	4.260	0.03	0

\* r – длина контакта, Ω - соответствующая ему величина телесного угла, а τ - ранг контакта.  
# Непрямой контакт.

Как отмечалось ранее, при построении разбиения Вороного каждый контакт, независимо от природы атомов его образующих, делится на равные

части, что является причиной критического отношения к данному представлению структуры [26] и поиска лучших вариантов неполновинного деления [32]. Однако во многих случаях, когда анализируются контакты между однотипными атомами (например, при поиске путей миграции катионов металлов в каналах, образованных только ионами кислорода [33]), деление пополам межатомных контактов является физически обоснованным. Использование разбиения Вороного позволяет однозначно определять ближайшее окружение атомов, а также множество дополнительных геометрических дескрипторов самого атома и его соседних атомов [34]. Наиболее важными из них являются объем, площадь поверхности и сферичность атомного домена, площадь, телесный угол грани, ее ранг, равный числу валентных связей в цепи, соединяющей атомы контакта [35]. В полиэдрическом представлении атома мы также получаем характеристики тройных и четверных межатомных взаимодействий, которым в тривиальном случае отвечают ребра и вершины ПВ, соответственно.

Применяя полиэдрическую модель, можно рассчитать и дескрипторы многоатомной структурной единицы кристалла (Рис. 3), проведя суммирование характеристик ПВ содержащихся в ней атомов. Так, суммирование объемов ПВ атомов структурной единицы дает приходящийся на нее объем, а сумма площадей всех граней ПВ, отвечающих межмолекулярным контактам, позволяет определить общую площадь ее поверхности. В частности, координационные соединения являются результатом химического взаимодействия лигандов и атомов или ионов металлов. В их кристаллах могут также содержаться внешнесферные частицы - противоионы и (или) молекулы растворителя. Таким образом, лиганды и внешнесферные частицы являются структурными единицами координационных соединений, так как именно они, а не составляющие их атомы, формируют кристаллическую структуру. Прогнозирование

параметров кристаллической структуры, исходя из ее структурных единиц, является важной задачей современного дизайна кристаллов.

#### **1.2.4. Методы расчета и хранения связанности атомов**

Большинство методов кристаллохимического анализа традиционно рассматривают кристалл как совокупность атомов, представляемых точками, сферами или эллипсоидами, связанными между собой системой связей (в случае ковалентных или координационных соединений) или образующими упаковку (если межатомные взаимодействия носят преимущественно ненаправленный характер, как в случае ионных или молекулярных веществ). Традиционный геометрический анализ ближайшего окружения атомов включает процедуру определения химической связи по соответствующим межатомным расстояниям и сравнение последних с суммами радиусов взаимодействующих атомов. Следовательно, все атомы в веществе рассматриваются как жесткие сферы определенных радиусов. Хотя такой подход ранее имел жизненно важное значение для развития кристаллохимии, жесткое (недеформируемое) сферическое приближение вступало в противоречие с современной концепцией природы химических связей.

Даже если предположить, что распределение электронной плотности в изолированных атомах  $A$  и  $X$  является сферическим, образование сильной химической (например, ковалентной) связи между ними является результатом перекрытия внешних (валентных) атомных орбиталей, что может рассматриваться как «взаимопроникновение». Поскольку трудно оценить степень такого взаимопроникновения (или перекрытия) только из структурных данных, даже самая простая и стандартная система атомных радиусов (ковалентных или металлических) теряет свой физический смысл. Если предположить, что распределение электронной плотности в атоме несферично в кристаллическом поле, само понятие атомного радиуса в кристаллической структуре становится довольно условным.

Чтобы устранить эти неопределенности, был предложен [36] и усовершенствован [37] альтернативный метод, основанный на использовании ПВ, для анализа межатомных связей в кристаллической структуре (метод *Domain*). На настоящий момент использованы следующих требований [37] к существованию валентного взаимодействия А-Х:

- (1) Для каждого неэквивалентного атома А строится его ПВ, а при последующем шаге учитываются только атомы Х, являющиеся непосредственными соседями атома А [38] (линии, соединяющие их с А, пересекают соответствующие грани ПВ).
- (2) Для всех контактов А – Х<sub>і</sub> вычисляются телесные углы Х<sub>і</sub> пирамид с базовой гранью ПВ, соответствующей контакту А – Х<sub>і</sub>. Если Х<sub>і</sub> < 1.5% от полного телесного угла 4π стерadiana, то контакт называется невалентным, а его природа определяется с помощью дополнительных критериев [39].
- (3) Для каждого типа атома определены два вида радиусов: радиус Слейтера r<sub>s</sub>, [40] (это постоянная величина, заданная в параметрах ToposPro) и усредненный радиус сферического домена (<R<sub>sd</sub>>, радиус сферы с объемом ПВ, усредненный по всем атомам данного сорта в структуре). Первый радиус характеризует размер атома, принятый в кристаллохимии; последний оценивает реальный размер данного атома в зависимости от положения атомов в конкретной структуре.
- (4) Для каждой пары А–Х определяется эталонное расстояние D(А–Х), которое равно минимуму двух сумм: R<sub>sd</sub>(А) + R<sub>sd</sub>(Х) или [r<sub>s</sub>(А) + r<sub>s</sub>(Х)] (1+δ), где по умолчанию δ=0.25. D(А–Х) означает максимальное расстояние, на котором контакт А–Х следует рассматривать как прочную (валентную) связь; если расстояние d(А–Х) между атомами в структуре превышает эталонное расстояние D(А–Х), контакт считается невалентным, и его природа определяется с помощью дополнительных критериев [39].

- (5) Значения  $D(A-X)$  изменены с учетом особенностей структуры. Для этой цели, если  $A$  - атом металла или оба  $A$  и  $X$  - неметаллы, все расстояния  $d(A-X)$  сгруппированы, если они не отличаются больше, чем на значение по умолчанию  $0.3 \text{ \AA}$ . Если после этого значение  $D(A-X)$  попадает в одну из групп, то оно увеличивается до максимального значения группы. Эта операция позволяет избежать ситуации, когда контакты, расстояния которых очень близки, но больше  $D(A-X)$ , не рассматриваются как связи.
- (6) Если  $A$  является атомом металла, и он не образует валентной связи или образует только одну связь в соответствии со значениями  $D(A-X)$ , полученными на предыдущих этапах, соответствующее значение  $D(A-X)$  увеличивается до максимального значения расстояния ближайшей группы контактов. В результате самые близкие контакты  $A-X$  будут считаться валентными (как правило, атом металла не может иметь нулевое координационное число или быть однокоординированным).
- (7) Если  $A$  и  $X$  являются металлами и хотя бы один из них является  $s$ - или  $p$ -металлом, то валентный контакт между ними может существовать, только если между ними и атомами неметаллов нет валентных связей; в противном случае  $D(A-X)$  указывается равным нулю. Это предотвращает образование валентных связей атомов  $s$ - или  $p$ -металлов с другими металлами, за исключением структур интерметаллических соединений. Также предполагается, что атомы  $s$ - или  $p$ -металлов не образуют кластеров; прочная связь между ними и другими металлами может существовать только в интерметаллических соединениях.
- (8) Если  $A$  - атом металла, а  $X$  - атом бора, связанный с атомами  $O$  или  $F$ , то связь  $A-X$  (если есть) разрывается ( $D(A-X) = 0$ ).
- (9) Если имеется несколько химически эквивалентных, но кристаллографически неэквивалентных пар  $(A, X)$  и, следовательно, на предыдущих шагах было выбрано несколько значений  $D_i(A-X)$ , то

максимальное значение  $D_i(A-X)$  принимается равным финальному эталонному расстоянию для всех контактов определенного типа  $A-X$ .

В отличие от классического метода, определение координационных чисел атома по методу Domain не требует ни априорного знания типа межатомных взаимодействий, ни использования различных систем кристаллохимических атомных радиусов (не считая использования радиусов Слейтера в качестве параметров). Несмотря на большой успех в автоматическом определении методом Domain матрицы смежности в структуре координационных соединений, проблемы ее расчета все же остались. Среди них отнесение к валентным или невалентным контактам связей между атомами металла и с искажениями Яна-Теллера. Есть вопросы по нахождению связей между атомами металла и атомами водорода или дейтерия в структуре гидридов.

Несмотря на то, что представление кристаллической структуры как трехмерного графа широко принято, методы его топологического анализа отсутствовали в течение длительного времени. Метод, предложенный в [41] и разработанный в [42, 43], был впервые реализован на компьютере. Этот метод тесно связан со стереоатомной моделью [20], поскольку межатомные связи могут быть корректно определены с использованием ПВ. Метод состоит из преобразования бесконечного структурного графа в конечный «свернутый» граф (СГ). Процедуру свертки можно представить наглядно следующим образом (Рис. 5). Те ребра бесконечного графа, которые выходят за пределы элементарной ячейки или заканчиваются на ее границах, зацикливают на трансляционно идентичных вершинах, расположенных внутри ячейки или на ее границах. СГ может содержать циклы и кратные ребра.

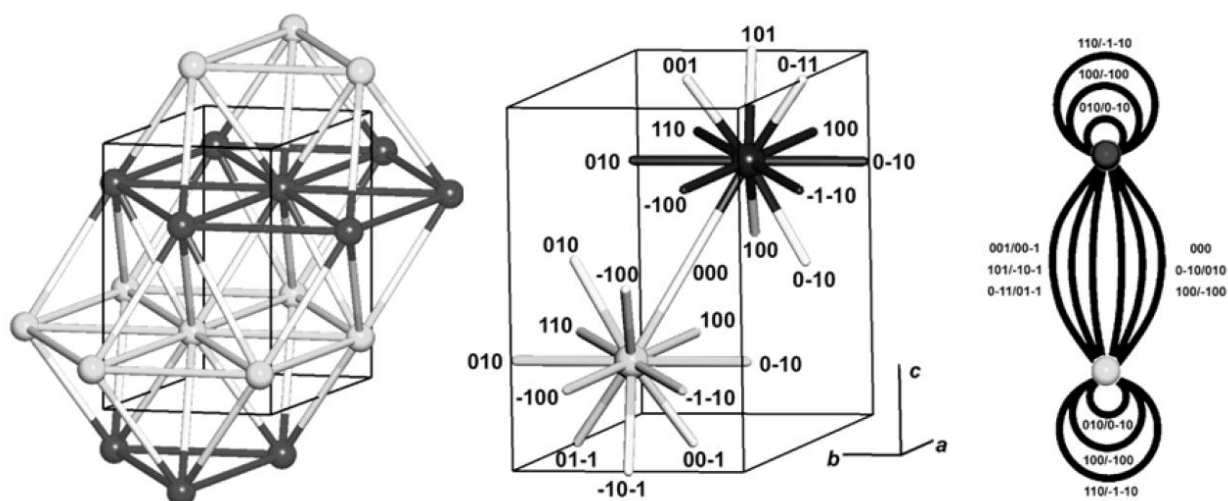


Рис. 5. Слева - фрагмент вложения сетки **hcp**, которое соответствует гексагональной плотнейшей упаковке. Трансляционно-эквивалентные узлы раскрашены одним цветом. В центре - элементарная ячейка вложения. Справа - соответствующий сетке свёрнутый граф. Ребра графа помечены в соответствии с трансляциями, связывающими эквивалентные узлы.

Петли появляются, если трансляционно эквивалентные атомы связаны друг с другом, а кратные ребра возникают, если атом связан с несколькими взаимодополняющими эквивалентными атомами. В общем случае каждая вершина СГ соответствует примитивной (без базиса) периодической подрешетке, а каждое ребро соответствует системе связей между наложенными примитивными подрешетками.

### 1.3. Кристаллохимические дескрипторы

Кристаллическая структура и ее структурные единицы могут быть охарактеризованы химическими, геометрическими и топологическими дескрипторами (Таблица 3).



**Таблица 3.** Список использованных дескрипторов межатомного контакта, атома, многоатомной структурной единицы и кристаллической структуры

Обозначение <sup>#</sup>	Дескриптор
<i>Межатомные контакты</i>	
<b>Rank</b>	наименьшее число валентных связей, соединяющих атомы, которыми образована грань ПВ
<i>Атом</i>	
AtNum	порядковый номер атома в Периодической системе
Group	номер группы атома в Периодической системе
Mult	кратность кристаллографической позиции
CN	координационное число
<b>WCN</b>	взвешенное координационное число
V <sub>CP</sub>	объем координационного полиэдра (Å <sup>3</sup> )
G <sub>3</sub>	второй момент инерции ПВ
Δ <sub>CM</sub>	смещение атома металла из геометрического центра тяжести его ПВ (Å)
R <sub>sd</sub>	радиус сферического домена, объем которого равен объему ПВ (Å)
V <sub>VP</sub>	объем ПВ (Å <sup>3</sup> )
S <sub>VP</sub>	площадь поверхности ПВ (Å <sup>2</sup> )
ΣΩ(X)	доля телесных углов граней, которые образованы соседними атомами X определенного химического сорта в ПВ этого атома
OxSt	степень окисления
T <sub>VP</sub>	топология реберной сетки ПВ
<i>Многоатомная структурная единица</i>	
TF	химический состав в виде валовой формулы (Total Formula)
CT	тип координации лиганда (Coordination Type)
ξ	ионный заряд
V <sub>VP</sub> <sup>M</sup>	молекулярный объем
S <sub>VP</sub> <sup>M</sup>	площадь внешней поверхности
CN <sub>UN</sub>	координационное число в базовой сетке
CF <sub>UN</sub>	координационная фигура в базовой сетке

<i>Кристаллическая структура</i>	
<b>CCF</b>	кристаллохимическая формула соединения
$\chi$	соотношение структурных единиц в соединении
$T_{UN}^{stand}$	топология базовой сетки в стандартном представлении
$T_{UN}^{cluster}$	топология базовой сетки в кластерном представлении при заданной величине максимального размера колец ( $\eta$ ) в структурной единице
<b>Z</b>	количество взаимопроникающих структурных сеток
<b>P</b>	доля свободного пространства, рассчитанная по объему, занятому ван-дер-Ваальсовыми полиэдрами атомов
$D_{chan}$	размерность системы каналов при заданном радиусе пробной частицы
$\Sigma \xi_i$	дисбаланс положительных и отрицательных зарядов
<b>TAM</b>	топологический механизм сборки из первичных и вторичных структурных единиц (Topological Assembly Mechanism)

#Жирным шрифтом выделены дескрипторы, которые впервые введены в этой работе.

### 1.3.1. Химические дескрипторы

*Химический состав* определяется химической природой атомов структурной единицы и представляется в виде валовой формулы. Атомы-комплексообразователи могут быть охарактеризованы их порядковыми номерами, а также номером группы в Периодической системе Д.И. Менделеева. Если структурная единица содержит несколько атомов, то дополнительно к ее химическому составу может быть учтена топология межатомных связей, то есть ее *внутренняя структура*, определяемая *молекулярным графом*. Это связный помеченный неориентированный граф, находящийся во взаимно-однозначном соответствии со структурной формулой структурной единицы таким образом, что вершинам графа соответствуют ее атомы, а ребрам графа — химические связи между этими атомами. Обычно, для определения внутренней структуры формируется

библиотека известных типов, на основании которой происходит отнесение новой структурной единицы к конкретному типу из этой библиотеки. В результате, каждой многоатомной структурной единице ставится в соответствие валовая формула и идентификатор топологического типа.

*Степень окисления* атомов металла и *заряд* лигандов определяют, как правило, исходя из баланса отрицательных и положительных зарядов в кристаллической структуре.

*Тип координации лиганда*, согласно номенклатуре [44], задается в виде символа  $L^{mbtkpghond}$ , в котором лиганд L обозначается одной из латинских букв M, B, T, K, P, G, H, O, N или D в зависимости от количества донорных атомов, связанных с металлом (1, 2, 3, 4, 5, 6, 7, 8, 9 или 10, соответственно). Общее число атомов металла, координируемых лигандом ( $v$ ), задается строкой mbtkpghond в виде верхнего индекса, где каждое целое число m, b, t, k, p, g, h, o, n, d равно количеству атомов металла, координируемых одним, двумя, тремя и так далее донорными атомами лиганда;  $v = m + b + t + k + p + g + h + o + n + d$ . (Рис. 6). Локальная топология комплексной группы записывается в виде *координационной* (или *кристаллохимической*) *формулы*  $A_n L_{i1} L_{i2} L_{i3} \dots$ , где A обозначает атом металла, n - число атомов металла в полиядерном звене,  $L_1, L_2, L_3 \dots$  обозначают типы координации лигандов, а  $i_1, i_2, i_3 \dots$  количество таких лигандов. Например, цепочечная структура KEPFEQ (Рис. 7) с координационной формулой  $AB^{11}_2$  содержит атомы металла (A) и бидентатные мостиковые лиганды ( $B^{11}$ ) в соотношении 1:2, и все дескрипторы этих структурных групп будут иметь веса  $1/3$  и  $2/3$ , соответственно.

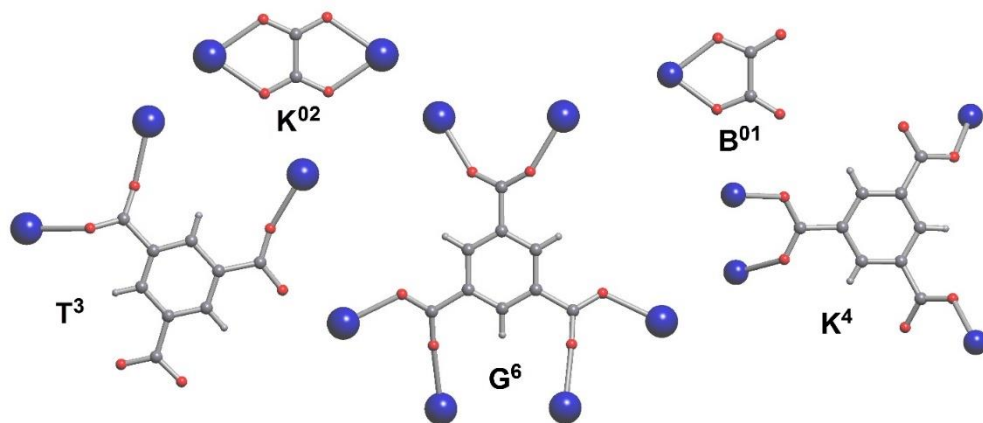


Рис. 6. Часто встречающиеся типы координации оксалатных и тримезатных лигандов. Атомы металла показаны синими шарами.

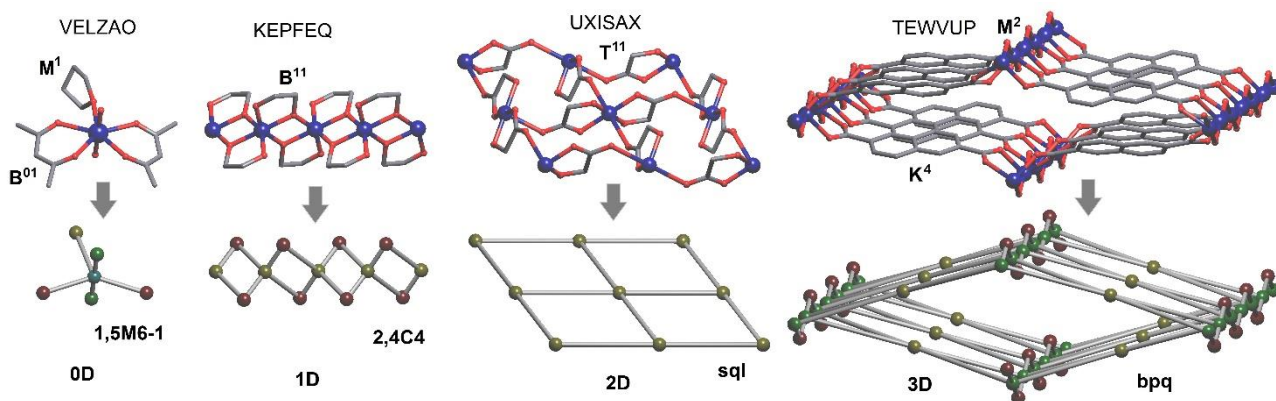


Рис. 7. Фрагменты координационных 0D, 1D, 2D и 3D сеток (вверху) и соответствующих базовые сетки с названиями их топологических типов (внизу) в кристаллических структурах  $\text{UO}_2(\text{acac})_2\text{thf}$  [45] (*acac* = ацетилацетонат, *thf* = тетрагидрофуран),  $(\text{этилендиолато})_2\text{Ti}$  [46],  $(\text{гликолато})_2\text{Cu}$  [47],  $\text{гидроксо-(нафталене-2,6-дикарбоксилато)V}$  [48]. Координация в этих соединениях описывается кристаллохимическими формулами  $\text{AB}^{01}_2\text{M}^1_3$ ,  $\text{AB}^{11}_2$ ,  $\text{AT}^{11}_2$  и  $\text{AK}^4\text{M}^2$ , соответственно. Референтные коды структур в базе данных CSD указаны в верхней части рисунка.

Для классификации лигандов по структурным фрагментам используются *бинарные векторные фингерпринты* [49], которые применяются для методов машинного обучения в фармацевтике [50, 51] и биохимии [52, 53]. Обычно, фингерпринт - это число или вектор чисел,

однозначно рассчитываемые на основании некоторых параметров объекта. Сравнение фингерпринтов двух объектов при определении их сходства происходит на порядки быстрее, чем сопоставление самих объектов.

Например, для лигандов  $C_nH_mO_k$  можно выделить пять трехатомных ( $O-C-O$ ,  $C-O-C$ ,  $C-C-O$ ,  $H-O-C$ ,  $H-C-O$ ) и три четырехатомных ( $C-C(O)-H$ ,  $C-C(O)-C$ ,  $C-C(O)-O$ ) структурных фрагментов и присвоить им номера 1–8. Затем каждый лиганд кодируется восьмизначным двоичным вектором, где «0» или «1» соответствуют отсутствию или присутствию определенного фрагмента. Например, фингерпринт 00100010 соответствует кетонам и поликетонам, а векторы 00111100 и 00111110 кодируют спирты. Структурные фингерпринты могут быть непосредственно получены из графов лигандов.

### 1.3.2. Геометрические дескрипторы

Объем структурной единицы в кристалле определяется по суммарному объему ван-дер-Ваальсовых сфер всех его атомов [54] без учета их перекрывания. Если допустить, что объем кристалла заполнен полностью, без промежутков, то для оценки объема может быть использован суммарный объем ПВ атомов, составляющих структурную единицу.

*Координационным полиэдром* атома называется выпуклый многогранник, вершинами которого являются соседние атомы, определяющие значение координационного числа атома (Рис. 8). Координационный полиэдр характеризует положение атомов окружения относительно центрального атома [55]. В качестве центрального атома обычно выступает атом металла, а атомами окружения являются атомы неметаллов.

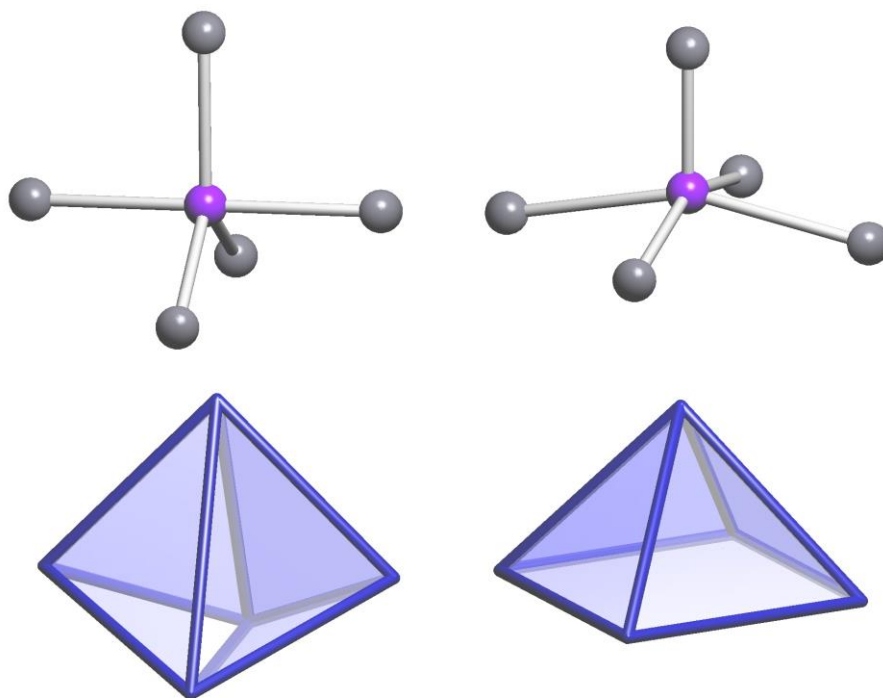


Рис. 8. Пятикоординационные фигуры (вверху) и соответствующие им координационные полиэдры (внизу): тригональная бипирамида (слева) и квадратная пирамида (справа).

Определение формы координационного полиэдра атома является важной задачей кристаллографии, так как эта характеристика связана с топологией кристаллической решетки в целом [56]. Чтобы решить эту задачу необходимо сопоставить исследуемому координационному полиэдру (образцу) многогранник-эталон из библиотеки. Известно несколько машинных методов определения формы координационного полиэдра. Наиболее часто [57] в них используется алгоритм итерационного совмещения исследуемого образца с эталоном. Достижение наибольшего их совпадения определяется по минимальному значению суммы расстояний между соответствующими парами вершин эталона и образца. В некоторых работах также рассматривается точечная симметрия координационного полиэдра [57] или топология связанности его ребер [58], которую затем сравнивают с соответствующей характеристикой эталона. Распределение плоских углов, опирающихся на координированные атомы, с вершиной в центральном атоме, может быть представлено инвариантами сферических гармоник [59].

В ряде случаев при описании окружения, вместо координационного полиэдра используют понятие *координационная фигура* (КФ) [60] (Рис. 8). КФ - это геометрическое тело, образованное центральным атомом и атомами, связанными с ним. Топология кристаллической структуры существенно зависит от формы КФ, а иногда полностью определяется ею [56, 60]. Для оценки формы КФ и координационных полиэдров могут быть использованы одни и те же алгоритмы [57, 58, 61]. Так как требуется обработка больших объемов информации, то весьма перспективным в этом случае являются дескрипторы формы КФ, которые в англоязычной терминологии называют *фингерпринтами* [62, 63, 64].

Чтобы охарактеризовать химическую природу *окружения* атома металла необходимо оценить энергию его взаимодействий с соседними атомами одного и того же химического сорта. В качестве геометрического параметра для качественной оценки энергии могут выбраны площадь или телесный угол грани ПВ. Известны и другие характеристики [65] окружения атома, рассчитываемые по характеристикам граней ПВ. К их числу относится *взвешенное координационное число* ( $WCN$ ), которое определяется по уравнению (4) как сумма всех телесных углов граней, нормированная на максимальный телесный угол.

$$WCN = \frac{\sum_{i=1}^{CN} \Omega_i}{\max(\Omega)}, \quad (4)$$

Если все телесные углы имеют одинаковое значение,  $WCN$  совпадает с координационным числом. Для оценки степени отклонения формы структурной единицы от геометрически правильного варианта удобно использовать значение нормализованного безразмерного *второго момента инерции* ПВ ( $G_3$ ), в общем случае рассчитываемого по формуле:

$$G_3 = \frac{1}{3} \int_{VP} R^2 dV_{VP} / V_{VP}^{5/3}, \quad (5)$$

где  $V_{VP}$  – объем ПВ,  $R$  – расстояние от центра структурной единицы до точки внутри ее ПВ. Как известно [24], для произвольного тела, разбитого на симплексы, величина  $G_3$  может быть рассчитана более простым способом:

$$G_3 = \frac{1}{3} \sum_i V_i I_i / V_{VP}^{5/3}, \quad (6)$$

где суммирование проводится по всем симплексам,  $V_i$  – объем  $i$ -го симплекса,  $I_i$  – нормализованный второй момент инерции симплекса относительно центра ПВ:

$$I = \frac{4}{5} \|\bar{v}\|^2 + \frac{1}{20} \sum_{j=0}^3 \|v_j\|^2. \quad (7)$$

В формуле (7) суммирование проводится по всем вершинам симплекса,  $\|v_j\|$  – норма радиус-вектора  $j$ -ой вершины симплекса,  $\|\bar{v}\|$  – норма радиус-вектора центра тяжести симплекса относительно центра ПВ. В ряде работ [20, 66, 67]) было показано, что характеристика  $G_3$  позволяет наиболее детально, по сравнению с другими известными критериями, оценить степень искажения полиэдрической области действия атома по сравнению со сферой, имеющей минимальное значение  $G_3$ , или по сравнению с правильными полиэдрическими формами (**Таблица 4**).

**Таблица 4.** Значения  $G_3$  для некоторых полиэдров и шара

Полиэдр	$G_3$	Полиэдр	$G_3$
Тетраэдр	0.10400	Ромбододекаэдр	0.07875
Куб	0.08333	Кубооктаэдр	0.07854
Октаэдр	0.08255	Додекаэдр	0.07813
Гексагональная призма	0.08122	Икосаэдр	0.07782
Шар	0.07697		

### 1.3.3. Топологические дескрипторы

*Координационное число (CN)* определяется как для узлов *структурной* сетки, так и для соответствующей ей *базовой* сетки. В то время как структурная сетка включает все атомы в качестве узлов и все межатомные



связи в качестве ребер, узлы и ребра базовой сетки соответствуют структурным единицам (как правило, их центрам тяжести) и связям между ними. Фактически, любое структурное представление, которое выделяет структурные единицы (например, молекулы в молекулярных упаковках или лиганды в координационных соединениях), неявно применяет концепцию базовой сетки. Такое более простое представление позволяет найти структурное сходство между химически совершенно разными соединениями. Если лиганд связан с атомом металла хелатным типом, то координационные числа для структурной и базовой сеток будут различны. Например, как для атомов металла, так и для лигандов в 0D, 1D и 2D структурах, изображенных на Рис. 7, координационное число в базовой сетке меньше, чем в структурной сетке.

*Топология реберной сетки ПВ ( $T_{VP}$ )* однозначно определяется по ее графу, который может быть представлен в виде диаграммы Шлегеля [68] (Рис. 9), являющейся проекцией на наибольшую по размеру грань всех оставшихся граней ПВ. Для обозначения  $T_{VP}$  ПВ могут быть использованы символы в виде последовательности чисел вершин в гранях, расположенных по возрастанию, с верхними индексами, равными их количеству ( $4^6 6^2$  для представленного на рисунке полиэдра в форме гексагональной призмы) или в виде набора чисел граней/ребер/вершин (8/18/12 для того же случая). Если символ отвечает ПВ с разной топологией, то его дополняют номером для каждого уникального случая ( $4^6 6^2-1$  или 8/18/12-1), собирая их в некоторой топологической библиотеке ТТР (см. Раздел 2.4.1) для однозначной идентификации аналогичных случаев.

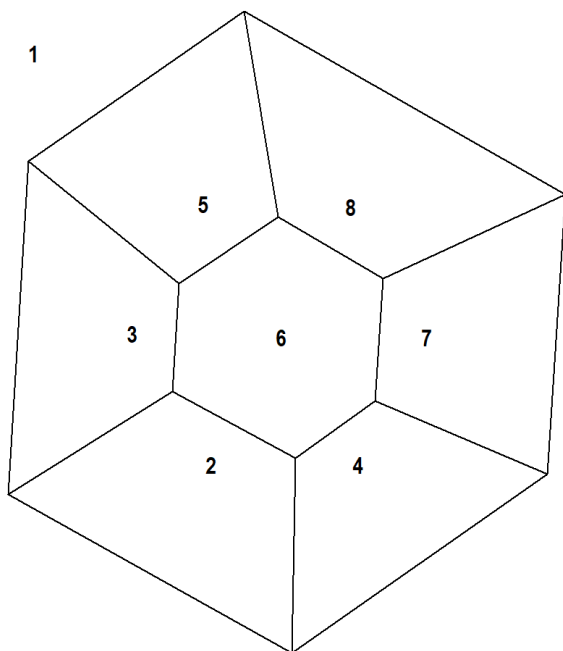


Рис. 9. Диаграмма Шлегеля [68] ПВ атома U (см. Рис. 3 слева) в кристаллической структуре  $\text{UO}_2(\text{C}_3\text{H}_3\text{O}_2)_2(\text{H}_2\text{O})_2$  [30] (NAFRIY). Все грани пронумерованы. Номер грани проекции указывается за пределами многоугольника.

Часто ПВ содержит небольшие по площади грани (Рис. 10 слева, грани с номерами 7, 11, 13 и 16), которые, практически не меняют форму полиэдра, но сказываются на его топологии. В подобных случаях, для устранения искажения и корректной идентификации, ПВ упрощают (Рис. 10 справа) стягиванием малых по размеру ребер или граней.

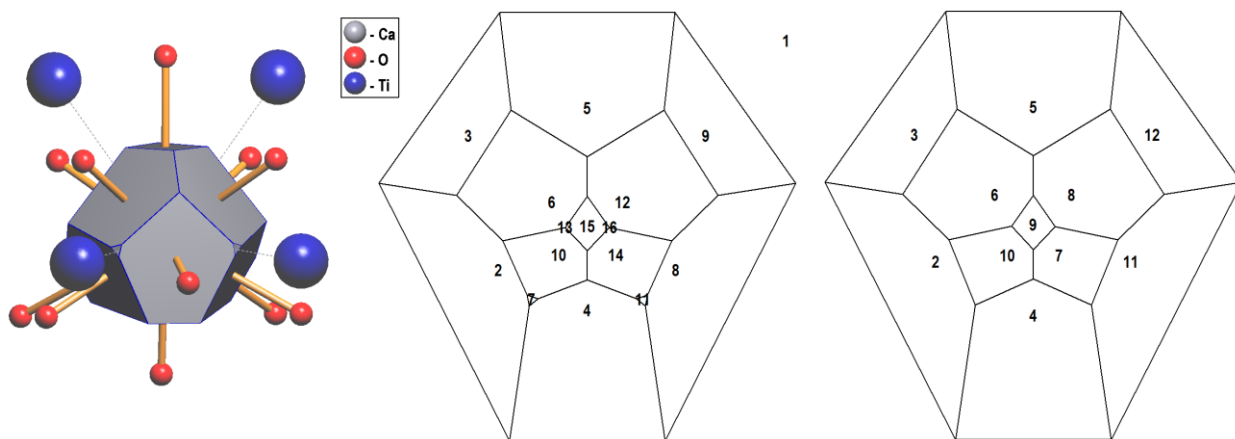


Рис. 10. Полный ПВ (справа) атома Ca в структуре  $\text{CaTiO}_3$  (162910) со слабыми контактами Ca-Ti, его (в центре) и результата упрощения (слева) диаграммы Шлегеля. В упрощенном ПВ стянуты все ребра, которые видны из центрального атома под углом не более  $10^\circ$ .

Передвигаясь по валентным связям структуры, можно оставаться в пределах конечной или бесконечной атомной группировки, у которой *периодичность* будет, соответственно, отсутствовать или наблюдаться в одном, двух или трех кристаллографических направлениях (Рис. 7). Таким образом, можно выделить нульмерные (островные или молекулярные), одномерные (цепи, ленты, трубки), двумерные (слои) и трехмерные (каркасы) группировки атомов, в которых атомы связаны ковалентными или координационными связями.

*Общая топология базовой сетки.* Для определения топологии сетки используется ряд топологических индексов [69]. Генерация базовой сетки и идентификация ее топологии формализованы в рамках строгих алгоритмов [60]. В зависимости от выбора структурных групп, для одной и той же кристаллической структуры могут быть рассмотрены разные представления, и каждое представление характеризуется собственной базовой сеткой. Очевидно, что топологии этих представлений взаимосвязаны [60, 70].

Для *полного представления* определяется топология структурной сетки, состоящей из всех атомов и всех валентных связей структуры (Рис. 11). Это представление обычно используется для неорганических соединений.

*Стандартным упрощением КС* (Рис. 7 или 12 справа вверху) является представление их лигандов в виде геометрических центров тяжести при сохранении связанности с атомами металла-комплексобразователя, не меняющими при таком упрощении своих позиции в структуре. Затем из полученной базовой сетки удаляют узлы, которые не влияют на ее топологию при данной периодичности.

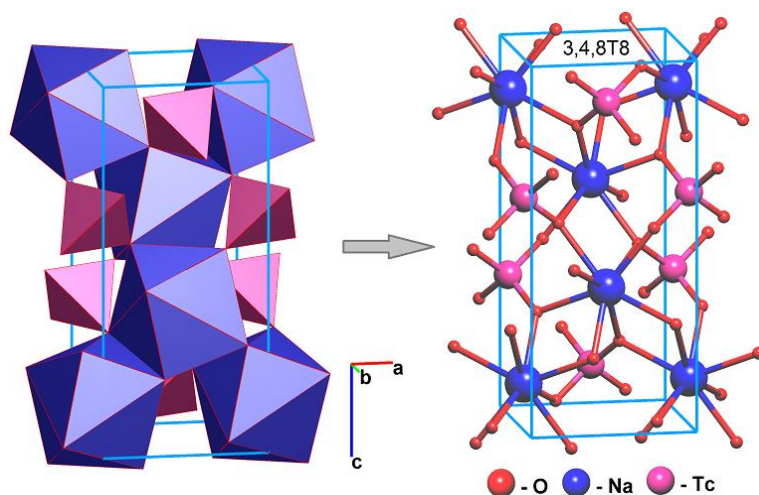


Рис. 11. Кристаллическая структура  $\text{NaTcO}_4$  в полиэдрическом представлении (слева) и в виде сетки топологического типа 3,4,8T8 (справа). Стандартное упрощение со всеми атомами металлов-комплексобразователей (Na и Tc) не меняет исходную сетку.

Для молекулярных структур удаляются только изолированные узлы, для цепочечных – изолированные и концевые узлы, для слоистых и каркасных структур – изолированные и концевые узлы, а мостиковый узлы преобразуются в ребра базовой сетки. Например, в структуре UXISAX, изображенной на Рис. 7, на финальной стадии стандартного упрощения будут удалены все узлы, отвечающие лигандам. Представление обычно используется для координационных соединений с органическими и неорганическими лигандами. Если координационная фигура атома металла окружена только одноатомными лигандами, то в стандартном представлении она идентична его координационному полиэдру.

*Кластерное упрощение КС* основано на определении для каждой связи кольца минимального размера, которому эта связь принадлежит. Далее эти кольца группируют по размеру, и все связи, относящиеся к кольцам, не превосходящим определенный размер, относят к внутрикластерным, а связи, принадлежащие бóльшим кольцам, рассматривают как межкластерные, что позволяет однозначно определить наиболее энергетически устойчивые фрагменты или вторичные структурные единицы (тектоны) [71, 72]. В

зависимости от представления лигандов данное упрощение допускает два варианта [73]: одноузловой, когда каждый полидентатный лиганд представляется в виде одного узла базовой сетки и полиузловой, когда такой лиганд может быть дополнительно разбит на несколько частей, представляемых в базовой сетке связанными центрами тяжести. Как и для стандартного представления, на финальной стадии кластерного упрощения удаляются все узлы, не влияющие на топологию сетки. Так, например, комплекс бис(5-хлороникотинато)-меди(II) (Рис. 12 слева) может быть охарактеризован двумя различными представлениями: *стандартным*, в котором структурными единицами являются атомы меди и лиганды 5-хлороникотината (Рис. 12 вверху справа), и *кластерным*, где в качестве вторичной структурной единицы рассматриваются биметаллические кластеры, размер колец в которых равен 8 (Рис. 12 внизу справа). Это минимальный размер колец, формирующихся в процессе роста кристалла в результате образования координационных связей, в то время как шестиатомные пиридиновые кольца не являются структурообразующими, так как уже присутствуют в лиганде – первичной структурной единице. Следующие по величине кольца содержат 24 атома, что в три раза больше 8, соответственно принадлежащие им связи, для которых эти кольца являются кратчайшими, должны быть отнесены к межкластерным. Как правило, возможность кластерного упрощения кристаллической структуры свидетельствует о многоэтапном механизме ее сборки. В данном случае сначала четыре лиганда координируются карбооксильными группами к двум ионам  $\text{Cu}^{2+}$ , образуя «китайские фонарики» - вторичные структурные единицы состава  $\text{Cu}_2(\text{C}_6\text{H}_3\text{NO}_2\text{Cl})_4$ , из которых затем формируется вся каркасная структура за счет образования связей между атомами меди и атомами азота пиридинового кольца (Рис. 12 слева).

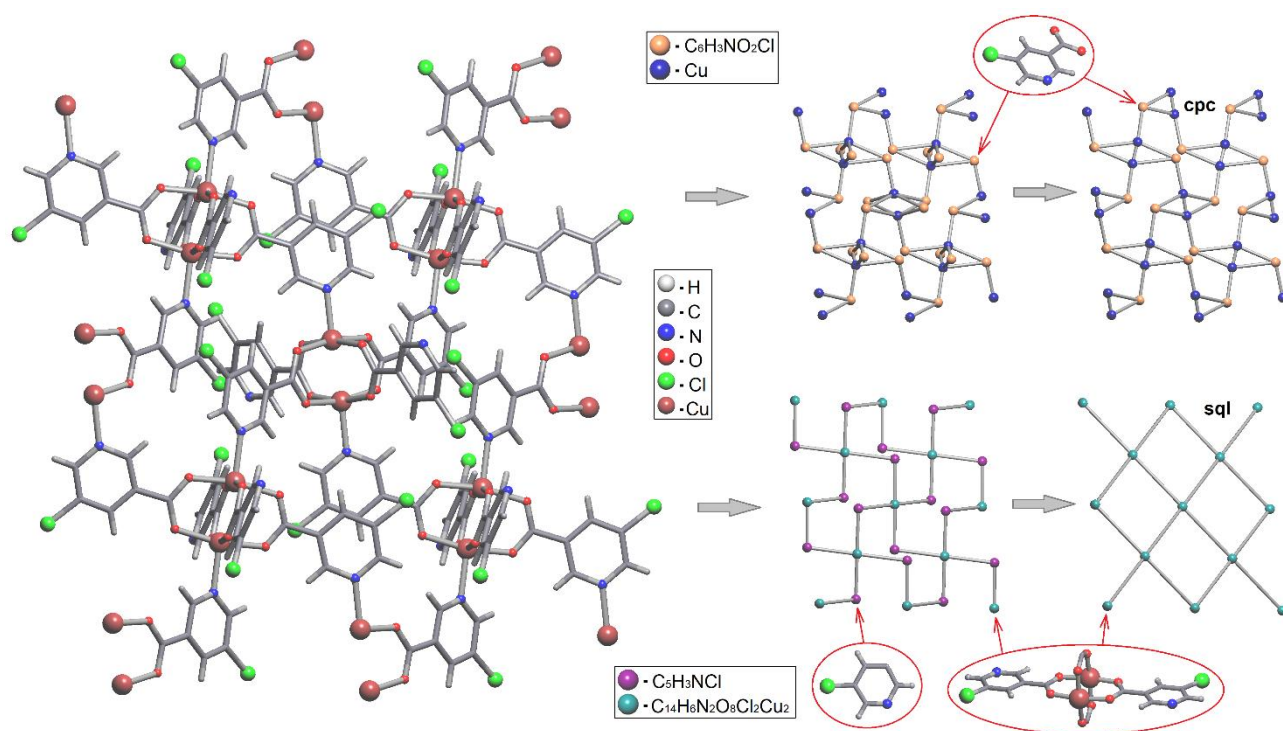


Рис. 12. Представления кристаллической структуры бис(5-хлороникотinato)-медь(II)  $[\text{Cu}(\text{C}_6\text{H}_3\text{NO}_2\text{Cl})_2]$  (слева) (АСАКАТ) [74] и соответствующих базовых сеток: стандартное представление (вверху справа), кластерное представление (внизу справа) с топологическими типами **cpc** и **sql**, соответственно. Базовые сетки до удаления 2-координированных узлов показаны в центре рисунка. В легенде окраски узлов указаны формулы упрощенных структурных единиц.

Если это *стандартное представление для молекулярных структур с водородными или (и) специфическими связями*, то топология определяется для сетки молекулярных центров масс, которые связаны на основе межмолекулярных водородных или (и) специфических связей (Рис. 13). Это представление обычно используется для молекулярных структур и конечных координационных группировок [75].

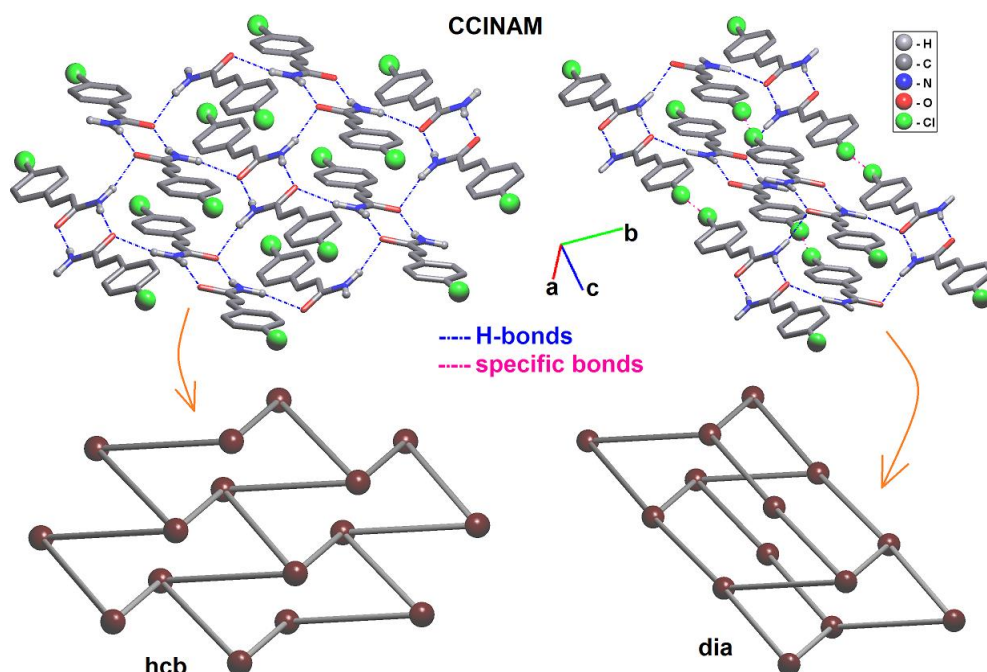


Рис. 13. Фрагменты структуры (вверху) и соответствующие базовые сетки (внизу) кристаллической структуры *p*-хлорциннамида (CCINAM). Учитываются только водородные связи (слева) или одновременно водородные и специфические связи (справа).

При стандартном представлении кулоновских или ван-дер-Ваальсовых молекулярных структур топология определяется для сетки центров масс молекул, которые связаны между собой межмолекулярными кулоновскими или ван-дер-Ваальсовыми связями [76, 77]. Это представление обычно используется для структур, состоящих из молекул или молекулярных ионов без водородных и специфических связей. Пример сетки, упрощенной таким образом, с топологией **bcsu**-х показан на Рис. 14. Данное представление часто характеризуется уникальной топологией, поэтому для него выполняют *генерацию представлений* [77] по мере роста прочности контактов, оцениваемой по величине их телесного угла (Рис. 14).



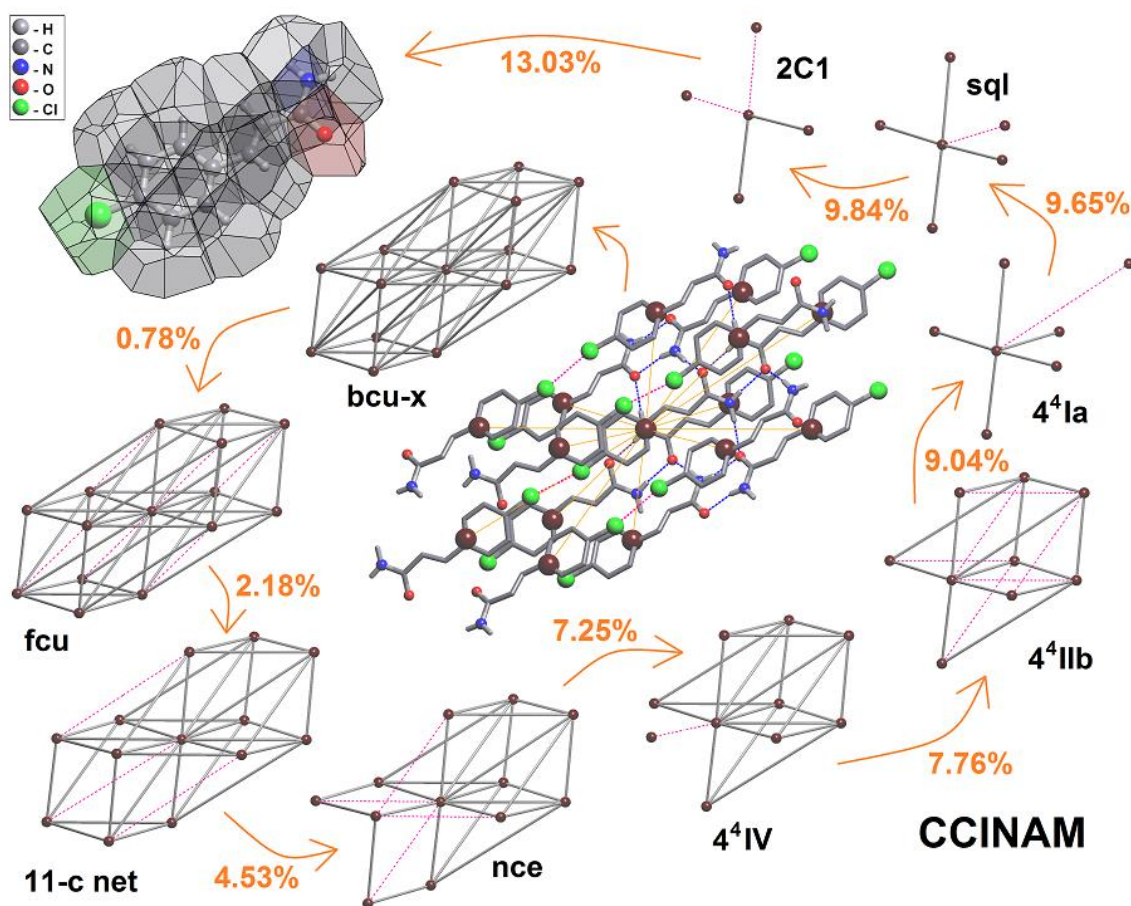


Рис. 14. Иерархический анализ упаковки молекул *p*-хлорциннамида (CCINAM) [78]. Межмолекулярные контакты расположены по телесным углам (в процентах к общему телесному углу) соответствующих общих граней молекулярных ПВ [79, 80] (вверху слева): от 13.03% до 0.78% для наиболее прочных и наиболее слабых контактов, соответственно. Полное представление со всеми межмолекулярными контактами имеет объемноцентрированную кубическую (**bcu-x**) топологию, а игнорирование самых слабых контактов приводит к гранецентрированной кубической (**fcu**) топологии. Последующий разрыв наиболее слабых сохранившихся контактов приводит к образованию серии *трехпериодических* 11-координатных и 8-координатных (**nce**), *двухпериодических* (слоистых) 7-координатных ( $4^4\text{IV}$ ), 6-координационных ( $4^4\text{Ib}$ ), 5-координационных ( $4^4\text{Ia}$ ) и 4-координированные (**sql**), и *однопериодических* (цепочечных) 2C1 мотивов. Контакты, которые необходимо разрывать на каждом этапе, показаны тонкими красными линиями. Обратную последовательность топологических типов можно трактовать, как последовательную сборку структуры из отдельных молекул.



Базовую сетку можно рассматривать как код сборки структуры из соответствующих структурных единиц. Эта особенность привела к появлению ретикулярной химии и специальных методов дизайна кристаллов с использованием базовых сеток в качестве шаблонов, которые должны быть декорированы соответствующими структурными единицами для получения целевой архитектуры. Этот подход наиболее интенсивно разрабатывался для металлоорганических каркасов [81, 82], но также обсуждалась его возможная применимость к цеолитам [83]. Ключевая роль базовых сеток в моделировании роста кристаллов проявилась в коде *CrystalGrower* [84], который расширяет топологический подход термодинамическими и кинетическими условиями, чтобы получить полное совпадение с экспериментом по выращиванию кристаллов совершенно разных химических классов.

*Параметры переплетения.* В структуре может присутствовать не одна полимерная группировка, а несколько, и в этом случае структура представляется набором сеток, которые могут переплетаться, а в 3D варианте обязательно переплетаются. Известны различные варианты переплетения (Рис. 15) [87, 88, 89]. В общем случае две (или более) сетки *переплетаются*, если они не имеют общих узлов и ребер, но не могут быть отделены друг от друга без разрыва ребер в данном пространственном вложении. Если система переплетающихся сеток имеет ту же периодичность, что и сами сетки, то она называется *взаимопроникающей* (Рис. 15, слева), если периодичность системы переплетающихся сеток выше периодичности отдельной сетки в системе, то такая система называется *поликатенирующей*. Поликатенация может быть *наклонной* (Рис. 15, справа) или *параллельной* (Рис. 16), когда переплетающиеся группы имеют соответственно одинаковую или разную ориентацию в пространстве [87]. Количество поликатенирующих сеток всегда бесконечно, что и определяет повышение периодичности всего массива переплетающихся сеток в этом случае.

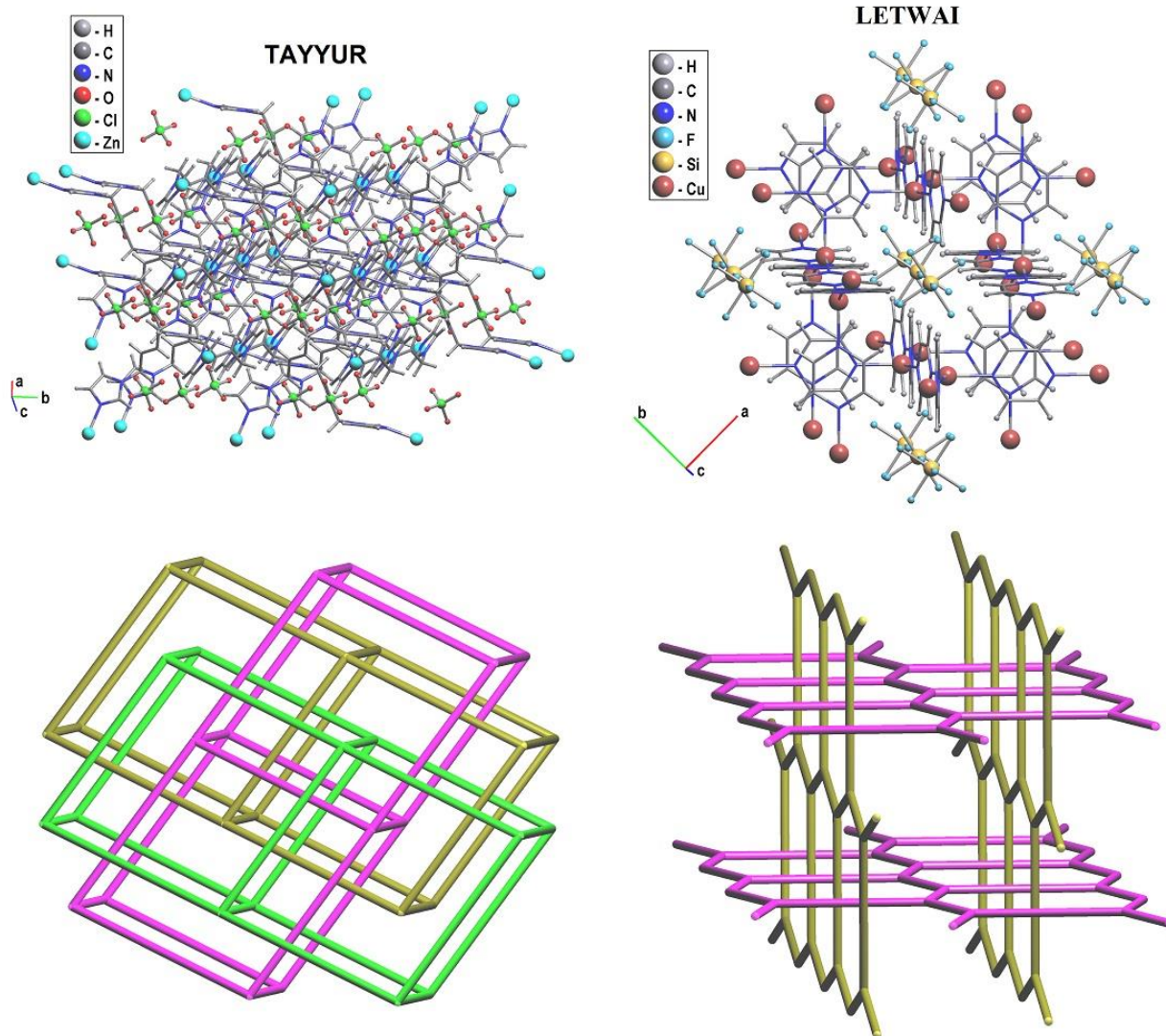


Рис. 15. Взаимопроникновение (слева,  $Z=3$ ) и наклонная (справа) поликатенация в структурах трис( $\mu$ -1,1'-(бут-2-ен-1,4-диил)бис(1H-имидазол))-диперхлорат цинка  $[\text{Zn}(\text{C}_{10}\text{H}_{12}\text{N}_4)_3] \cdot 2\text{ClO}_4$  (TAYYUR) [85] и гексакис( $\mu_2$ -пирозин)-тетра-медь(I)-бис(гексафторосиликат)  $[\text{C}_{24}\text{H}_{24}\text{Cu}_4\text{N}_{12}] \cdot 2(\text{SiF}_6)$  (LETWAI) [86], соответственно. Показаны структурные (вверху) и базовые (внизу) сетки.

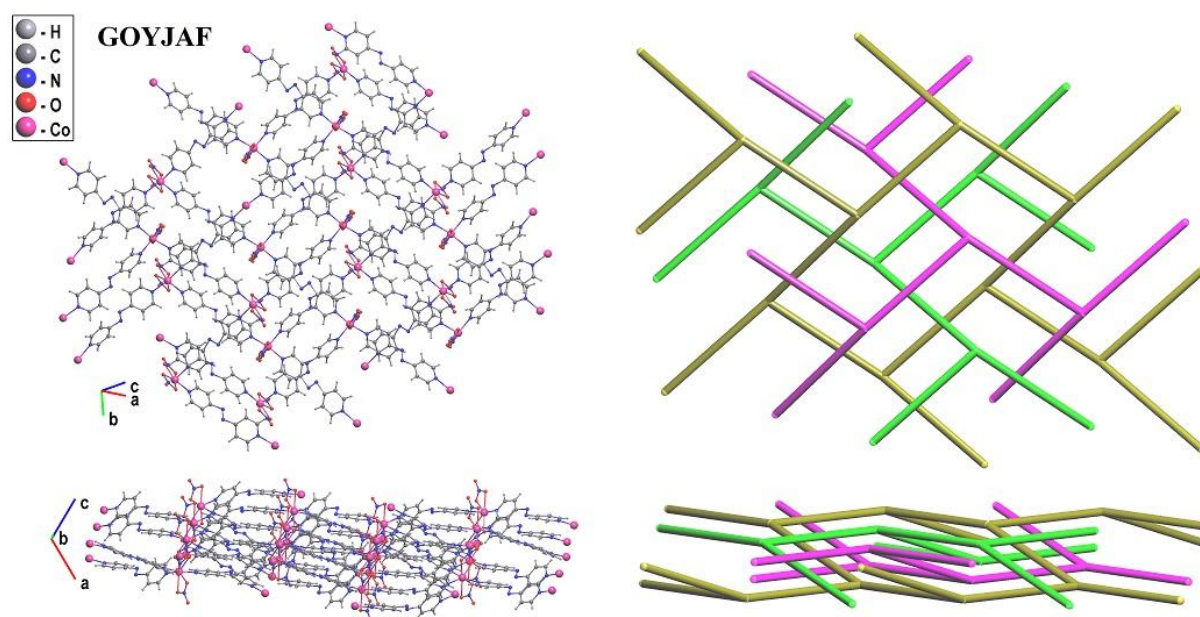


Рис. 16. Параллельная 3-кратно взаимопроникающая поликатенация в структуре трис( $\mu_2$ -транс-4,4'-азо-бис(пиридин))-тетракис (нитрато-О,О')-ди-кобальта(II)  $[\text{Co}_2(\text{C}_{10}\text{H}_8\text{N}_4)_3(\text{NO}_3)_2] \cdot \text{CH}_2\text{Cl}_2$  (GOYJAF) [90]. Показаны структурные (слева) и базовые (справа) сетки в двух различных проекциях.

Напротив, число взаимопроникающих сеток всегда конечно и называется *степенью взаимопроникновения* (**Z**). Поликатенация возможна только для 1- и 2-периодических сеток, в то время как взаимопроникновение реализуется для сеток любой периодичности. Известны случаи самопереплетения сетки, когда фрагмент сетки проходит сквозь ее кольцо (Рис. 17).

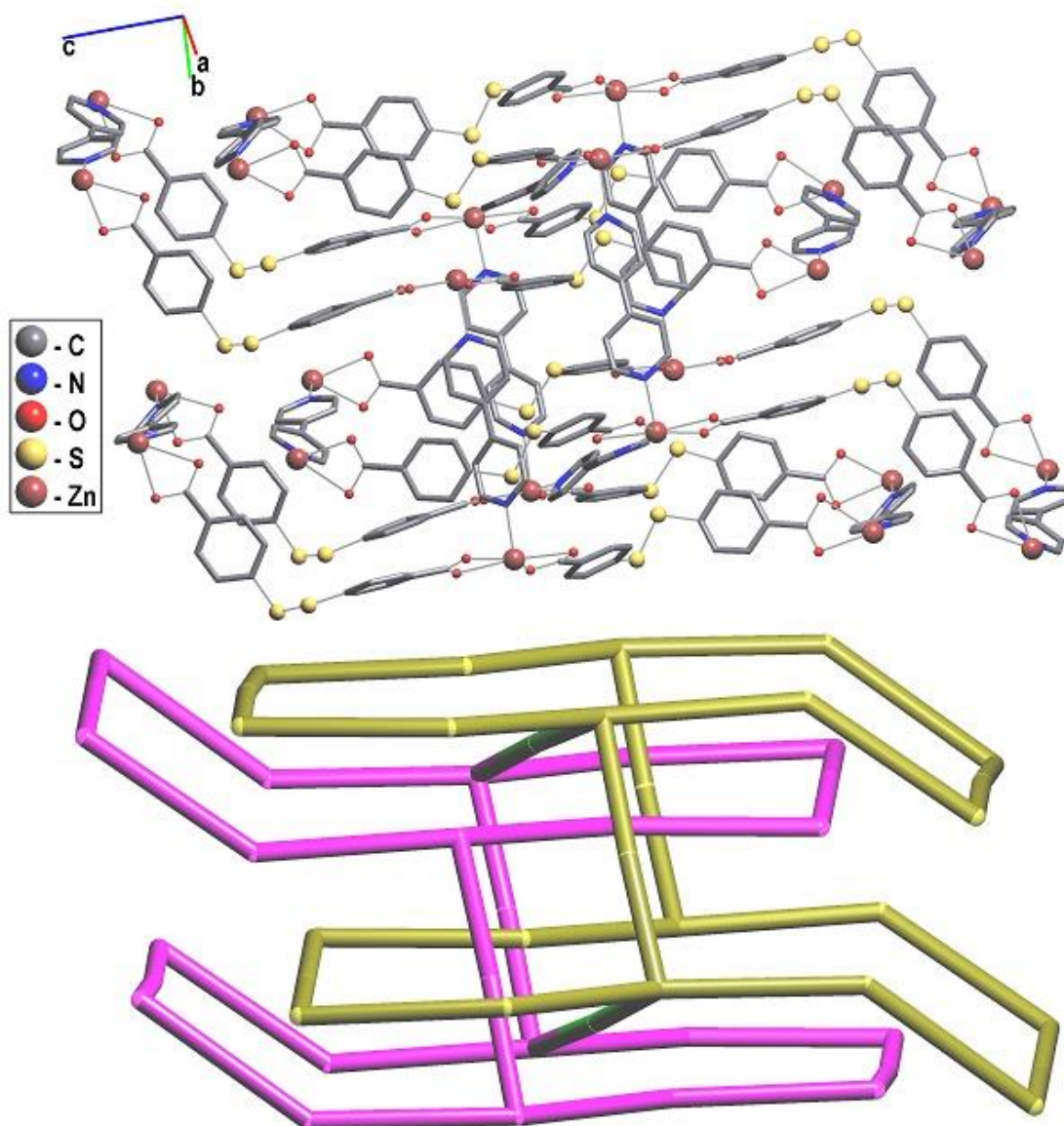


Рис. 17. Самовзаимопроникновение на примере структурной (вверху) и базовой (внизу) сеток структуры трис( $\mu$ -4,4'-бипиридин)-бис( $\mu$ -4,4'-дисульфандиилдибензоато)-ди-цинк  $\text{Zn}_2(\text{C}_{10}\text{H}_8\text{N}_2)_3(\text{C}_{14}\text{H}_8\text{O}_4\text{S}_2)_2$  (VOHVAT) [91]. Мостиковые лиганды 4,4'-бипиридина пронизывают кольца. На верхнем рисунке атомы водорода не указаны для наглядности.

#### 1.3.4. Дескрипторы свободного пространства

В геометрико-топологических моделях кристаллическое пространство естественным образом разделяется на две непересекающихся части — подпространство атомов, топология на котором задается системой

межатомных контактов той или иной природы, и подпространство межатомных пустот, связанных между собой системой каналов. Как правило, кристаллохимическое исследование конкретной кристаллической структуры сводится к анализу подпространства первого типа, тогда как подпространство пустот рассматривается лишь в особых случаях, когда с ним связывают специальные свойства вещества (такие как ионная проводимость, абсорбция, способность образовывать соединения внедрения, выступать в роли молекулярного сита и т. д.). Надежные результаты в оценке положений, числа и размеров пустот, топологии системы каналов между ними обычно возможны лишь в случае высокосимметричных структур несложного химического состава, в которых пустоты имеют правильное полиэдрическое атомное окружение (как, например, в плотнейших упаковках или в цеолитах [55]). В остальных случаях приходится довольствоваться полуколичественными или качественными результатами, и судить о свойствах пустот по косвенным данным (например, по способности вещества абсорбировать ионы или молекулы определенного размера или по анизотропии его ионной проводимости).

Известно несколько алгоритмов и программ оценки размеров пустот в структуре кристаллов [92, 93, 94]; все они требуют использования систем атомных и ван-дер-Ваальсовых радиусов. Следует отметить, что в современной кристаллохимии роль анализа свободного пространства возрастает в связи с изучением новых классов веществ (ионных и молекулярных сит, ионных проводников, металлоорганических каркасов и электридов).

Указанное выше деление кристаллического пространства на две части является условным, так как в нем предполагается, что атом имеет сферическую форму и размер, определяемый одной из систем кристаллохимических радиусов. До исследований автора разбиение Вороного практически не применялось для исследования пространства



пустот в структуре кристаллов, хотя на принципиальную возможность такого применения неоднократно указывалось [55]. Известен также ряд алгоритмов анализа пор в непериодических системах, использующих элементы разбиения Вороного [95]. В настоящее время для оценки степени *пористости* разработано несколько компьютерных программ. Так программа Mercury [96] оценивает размер незанятого атомами пространства и визуализирует его. Программа ZEO++ [97] выполняет подробный геометрический анализ пористых материалов и их пустот с использованием ПВ.

К пористым структурам обычно относят цеолиты [98], ковалентные органические полимеры [99], пористые полимерные сетки [100] и молекулярные материалы [101, 102] а также металлоорганические координационные полимеры (МОКП) или Metal-Organic Frameworks (MOF) в англоязычной терминологии [103]. МОКП построены из ионов металла или кластеров, связанных между собой органическими лигандами (*линкерами*). Варьируя состав и строение лигандов можно получить поры нужного размера [104], зависящего также от топологии базисной сетки и наличия в структуре взаимопроникающих каркасов [105, 106]. Некоторые координационные соединения с небольшими порами при внедрении в структуру гостевых молекул могут значительно увеличивать размер пор посредством деформации лиганда и углов координационного полиэдра атома металла (эффект «дыхания») [107].

Помимо пористости, свободное пространство может быть охарактеризовано набором параметров, определяющих возможность миграции в структуре сферического зонда с радиусом  $R_{\text{probe}}$  [108], размер которого определяет *периодичность* ( $n$ ) доступной сетки Вороного. Предельные значения  $R(n)$  могут быть найдены для максимальных значений  $R_{\text{probe}}$ , которые все еще сохраняют периодичность  $n$  и соответствуют наиболее широким системам каналов заданной периодичности. Сетка теряет

периодичность при некотором граничном значении  $R_{\text{probe}}$ , соответствующему наибольшему зонду радиуса  $R_f$ , который может свободно перемещаться по кристаллу. Максимальное значение  $R_f$  соответствует радиусу самого большого сферического зонда ( $R_{if}$ ), который может быть размещен в соответствующей системе каналов и полостей (но, возможно, не может свободно перемещаться по ним). Если сетка Вороного расщепляется на несколько симметрично-неэквивалентных несвязанных сеток, это означает, что существуют независимые пути миграции, и для каждого из них могут быть определены параметры  $R_f$  и  $R_{if}$ . Максимальное значение всех  $R_{if}$  соответствует радиусу самой большой полости в структуре и обозначается  $R_i$ . Пример канала и его характеристики показаны на Рис. 18.

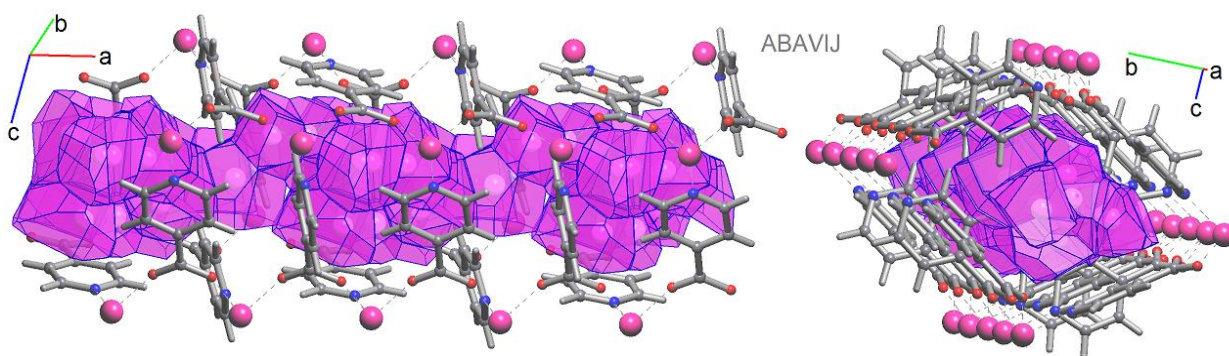


Рис. 18. Фрагмент структуры  $[\text{Co}_2(\text{INA})_4] \cdot \text{EtOH}$  [109] (INA = изоникотинат,  $\text{NC}_5\text{H}_4\text{-4-CO}_2^-$ ) и одномерный канал в направлении  $[1,0,0]$  с  $R_i = 1.86 \text{ \AA}$  и  $R_f = 0.98 \text{ \AA}$ . Сольватные молекулы этилового спирта удалены из структуры перед расчетами.

Для характеристики свободного пространства, не занятого атомами, используются и топологические методы. Наиболее известной является модель натурального тайлинга [108, 110], элементы которого, натуральные тайлы, представляют собой обобщенные многогранники, образованные узлами и ребрами атомной сетки и заполняющие кристаллическое пространство без пропусков и пересечений (Рис. 19). Натуральные тайлы характеризуют размер и форму минимальных клеток (полостей) каркаса, а

тайлинг в целом кодирует способ соединения клеток через общие грани тайлов, соответствующие каналам между полостями [110].

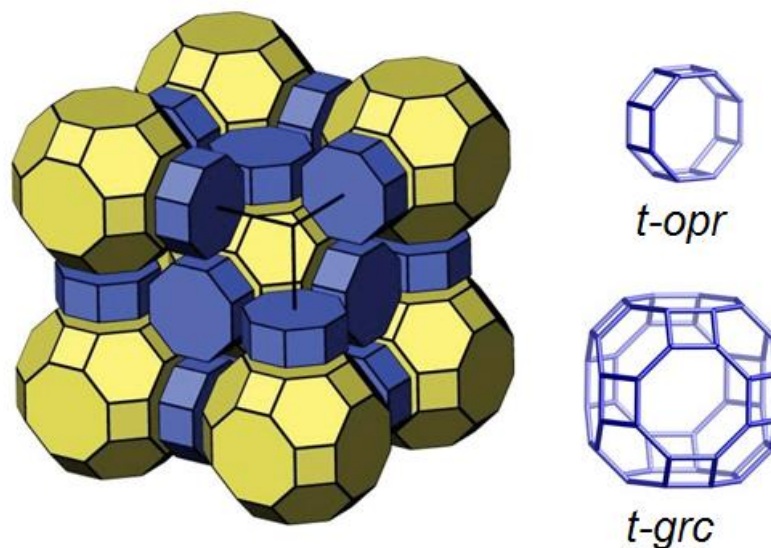


Рис. 19. Натуральный тайлинг в цеолите RHO (слева), который состоит из двух типов натуральных тайлов с именами *t-opr* и *t-grc* (справа).

#### 1.4. Хранилища структурных данных и их дескрипторов

Очевидно, что огромное количество информации, накопленной в базах структурных данных, скрывает много общих законов и особых закономерностей, которые управляют кристаллической архитектурой. В 2008 году У.Д. Мазервелл опубликовал статью [111] с интригующим названием: «CSD - 400 000 ответов, но какие вопросы?» Действительно, экспериментальные данные можно рассматривать как доказательство некоторых гипотез, которые на самом деле являются «вопросами», которые мы задаем природе. Чтобы сформулировать правильные гипотезы, нам нужны специальные инструменты, которые позволяют нам обрабатывать, понимать и упорядочивать информацию из баз данных, а также находить корреляции в данных [112]. Таким образом, следующим шагом в разработке кристаллографических информационных систем является создание «вторичных» баз данных, в которые извлечены полезные «ответы» из



наборов данных. Такие базы данных являются неотъемлемой частью систем искусственного интеллекта (ИИ), которые используют «ответы», чтобы делать надежные прогнозы. На настоящий момент известно лишь несколько их примеров в области кристаллохимии и материаловедения, в то время как системы искусственного интеллекта все еще находятся на стадии разработки. Одной из таких является база данных Mogul [113], которая интегрирована в систему CSD. Разработка подобных баз данных существенно ограничена следующими проблемами [114]: (i) подготовка надежных исходных кристаллографических данных, (ii) поиск новых структурных дескрипторов для характеристики особенностей структуры и (iii) разработка специальных методов и инструментов для создания «вторичных» баз данных из исходной кристаллографической информации. CSD является хорошим примером, где все эти проблемы решались в течение длительного времени, однако даже эти данные нуждаются в специальной подготовке для машинного анализа. База данных Mogul имеет дело с параметрами локального окружения молекул в кристаллах, но не учитывает свойства всей структуры, в частности тип молекулярной упаковки. Возможный путь дальнейшего развития этих инструментов состоит в расширении списка дескрипторов теми, которые характеризуют особенности структуры и имеют ясный физический смысл.

Для извлечения знаний из кристаллографических данных используются кристаллохимические [115], квантово-химические [116, 117], геометрические, топологические и электронные [118] дескрипторы, а также алгоритмы машинного обучения [119], которые стали популярными в материаловедении в последние несколько лет: метод опорных векторов [120], k-ближайших соседей, случайный лес [121], нейронные сети [12] и др. Машинное обучение применялось для прогнозирования механических [122, 123], адсорбционных [123, 124], каталитических [125, 126, 127] и других свойств [128] кристаллических веществ и материалов.

### 1.4.1. Интеллектуальные и экспертные системы

Интеллектуальные и экспертные системы [129] стали неотъемлемой частью современных технологий и науки, имеющих дело с огромными объемами данных [10, 130, 131, 132]. Основным компонентом таких систем является база знаний, которая содержит набор правил или обученные модели, используемые для решения задач данной предметной области. Именно разработка баз знаний, соответствующих методов и программного инструментария становится общемировым трендом.

В настоящее время известно несколько баз знаний, созданных на основании кристаллоструктурных данных, в их числе электронная база геометрических параметров Mogul [113], базы данных физических свойств (matweb.com [133]), результатов квантово-механических расчетов (materialsproject.org, aflow.org [134], topologicalquantumchemistry.com [135]), библиотеки топологических типов сеток (RCSR [136], Epinet [137]).

На основе данных баз знаний создаются веб-сервисы, которые предоставляют пользователю интересующую его информацию о кристаллических структурах и (или) прогнозируют ее свойства. Обычно, химический состав соединения или CIF-файл структуры используются в качестве первичных данных таких сервисов. Например, сервис materialsatlas.org [138] определяет для неорганических соединений по химическому составу ширину запрещенной зоны, твердость, кристаллическое строение и др. Сервис oxiMACHINE [121] использует машинную модель для прогноза степени окисления атомов металла в МОКП. Прогноз доступности МОКП для гостевых молекул с вероятностью 80.5% по характеристикам металла и линкера выполнен в работе [139], где данные для обучение машинной модели были получены путем разложения на металл и линкер экспериментальных трехпериодичных структур МОКП в Кембриджской структурной базе данных. Авторы исключили из рассмотрения гостевые молекулы и ионы, которые во многих случаях играют

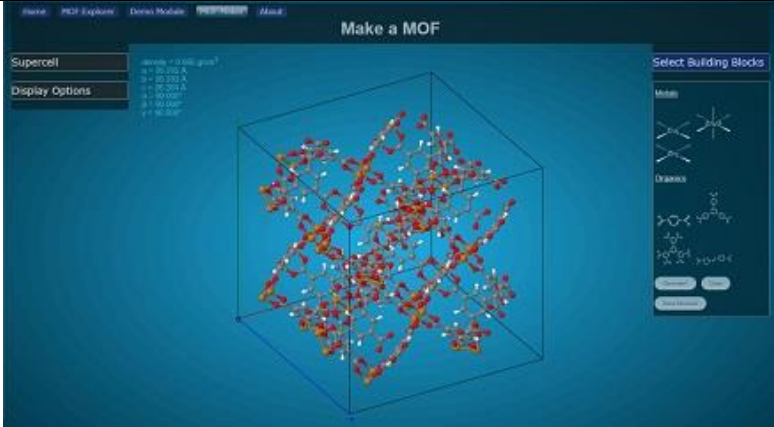
важную роль в процессе сборки кристаллической структуры, а иногда, устойчивость самой структуры и размеры ее пор зависят от наличия таких частиц. Гарсиа и Гаранчик [140] разработали предикативную модель для оценки пористости в молекулярных кристаллических структурах, изучили корреляции между дескрипторами пористости молекулярных клеток и образуемых ими кристаллов и установили, что пористость кристаллов сильно коррелирует с пористостью молекулярных структурных единиц.

Перечень программ для изучения особенностей строения кристаллических структур и их назначение приведены в **Таблице 5**.

**Таблица 5.** Современное программное обеспечение (ПО) и веб-сервисы для решения кристаллохимических задач

Имя приложения	Назначение	Веб-адрес
<i>Визуализация и анализ кристаллических структур</i>		
Mercury	ПО для 3D визуализации структуры, исследования кристаллической упаковки и статистического анализа данных поиска CSD.	<a href="https://www.ccdc.cam.ac.uk/solutions/software/mercury/">https://www.ccdc.cam.ac.uk/solutions/software/mercury/</a>
WebCSD	Сервис для поиска кристаллических структур через совместную службу доступа CCDC и FIZ Karlsruhe. Позволяют просматривать и загружать отдельные структуры и связанные с ними данные.	<a href="https://www.ccdc.cam.ac.uk/structures/?">https://www.ccdc.cam.ac.uk/structures/?</a>
VESTA	ПО для 3D-визуализации структурных моделей и объемных данных, таких как электронная плотность и морфология кристаллов	<a href="https://jp-minerals.org/vesta/en/">https://jp-minerals.org/vesta/en/</a>
<i>Определение свободного пространства кристалла и молекул его наполнения</i>		
Zeo++	ПО для расчета геометрических параметров пор в кристаллических структурах.	<a href="http://zeoplusplus.org/">http://zeoplusplus.org/</a>
MOF Data Explorer	Сервис для анализа свойств МОКП: диаметра пор, теплоты адсорбции, плотности и др.	<a href="https://aam.ceb.cam.ac.uk/mof-">https://aam.ceb.cam.ac.uk/mof-</a>

		explorer/
<i>Предсказание структуры кристаллов</i>		
USPEX	Метод компьютерного предсказания кристаллических структур, исходя из знания только химического состава материала	<a href="https://uspex-team.org/">https://uspex-team.org/</a>
CrySPY	ПО для предсказания кристаллической структуры, написанное на Питоне.	<a href="https://github.com/Tomoki-YAMASHITA/CrySPY">https://github.com/Tomoki-YAMASHITA/CrySPY</a>
CrystalGrower	ПО для выращивания кристаллов методом Монте-Карло для моделирования габитуса кристаллов и наноразмерной топографии поверхности.	<a href="https://crystalgrower.org/">https://crystalgrower.org/</a>
<i>Предсказание физических свойств кристаллического вещества по его структуре</i>		
Materials Atlas	Сервис для прогноза по химическому составу теплопроводности, твердости, сверхпроводимости и др.	<a href="http://www.materialsatlas.org/">http://www.materialsatlas.org/</a>
Topological Material Database	Сервис для поиска материалов с заданными топологическими свойствами	<a href="https://www.topologicalquantumchemistry.com/">https://www.topologicalquantumchemistry.com/</a>
<i>Предсказание атомных свойств по кристаллической структуре</i>		
OxiMachine	Сервис для прогноза степени окисления атомов металла CIF-файлу кристаллической структуры	<a href="https://oximachine.materialscloud.io/">https://oximachine.materialscloud.io/</a>
<i>Конструкторы кристаллических структур</i>		
MOF Maker	Сервис для генерации структуры МОКП из структурных единиц	<a href="https://mausdin.github.io/MOFsite/maker.html">https://mausdin.github.io/MOFsite/maker.html</a>

		
ToBaCCo	ПО для генерации МОКП из структурных единиц для последующей оценки с помощью молекулярного моделирования	<a href="https://github.com/tobacco-mofs/tobacco_3.0">https://github.com/tobacco-mofs/tobacco_3.0</a>
<i>Генераторы отчетов по кристаллическим структурам</i>		
Materials Explorer	Поиск информации о материалах по химическому составу или свойствам	<a href="https://materialsproject.org/">https://materialsproject.org/</a>
Robocrystallographer	ПО для создания текстовых описаний кристаллических структур	MRS Commun. 2019, 9, 874–881.

#### 1.4.2. Топологические библиотеки

Топологический тип сетки является важным дескриптором кристаллической структуры. Для его определения необходимы библиотеки известных топологии кристаллических структур, а также методы и программные средства, которые позволяют однозначно определять топологию атомной сетки. До разработки программного комплекса ToposPro единственной известной программой, позволяющей определять топологический тип сетки и создавать топологические библиотеки, являлась программа *Systre* (<http://gavrog.org/>). Она предназначена для анализа только периодических сеток при исследовании (полимерных или немолекулярных) кристаллических структур. Акроним расшифровывается как Симметрия, Структура и Уточнение (или Распознавание) (**S**ymmetry, **S**tructure **R**efinement (или **R**ecognition)). *Systre* использует метод, называемый барицентрическим размещением, для определения идеальной (максимальной) пространственной

симметрии кристаллической сетки и анализа ее топологии. Затем программа *Systre* генерирует уникальный ключ, кодирующий матрицу смежности данной сетки, и использует этот ключ для поиска топологии во встроенной базе данных. Рассчитанные программой *Systre* сетки и тайлы в них позволяет визуализировать программа *3dt* (Рис. 20).

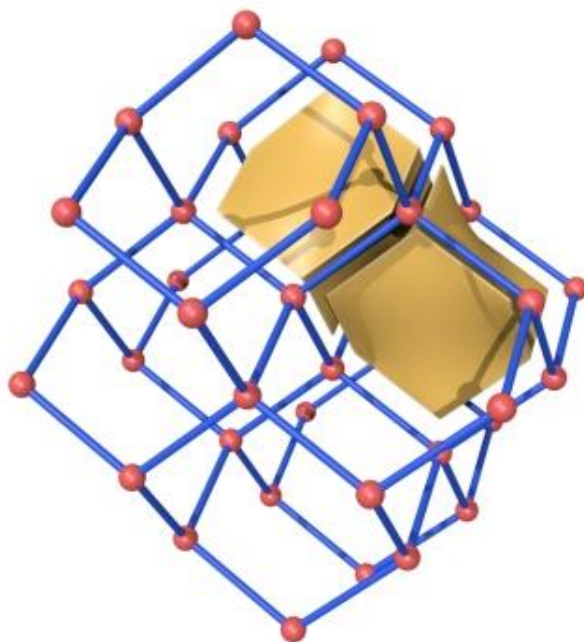


Рис. 20. Алмазная сетка, изображенная с помощью программы *3dt*, с двумя выделенными адамантановыми клетками.

Встроенная база данных *Systre* содержит 176 топологических типов цеолитных каркасов и большую часть (более тысячи) сеток, содержащихся в RCSR. Пользователи также могут создавать свои собственные базы данных и использовать ключ *Systre* для поиска структур в них. Примерами проектов, которые уже используют *Systre*, являются проект EPINET в Австралийском национальном университете и ресурс RCSR, размещенный в Австралийском национальном университете в сотрудничестве с государственным университетом Аризоны и Калифорнийским университетом в Лос-Анджелесе.

Общая топология сетки обозначается одним из следующих вариантов:

- строчные жирные трехбуквенные символы RCSR [136], например, **dia** обозначает алмазоподобную сетку;

- символы ToposPro  $ND_n$ , где  $N$  – последовательность координационных чисел всех независимых узлов сетки;  $D$  - одна из букв  $C$ ,  $L$  или  $T$ , обозначающая периодичность сети ( $C$  – цепочечная,  $L$  – слоистая,  $T$  – каркасная);  $n$  нумерует неизоморфные сетки с заданной последовательностью  $ND$ . Например, символом  $3,3,4T3$  обозначена третья по порядку трехпериодическая трехузловая сетка с двумя 3-координированными и одним 4-координированным независимыми узлами. Для конечных (молекулярных) графов используются символы  $NM_k-n$ , где  $M$  обозначает разновидность молекул, а  $k$  — количество вершин (атомов) в графе;
- символы EPINET  $sqcN$ , где  $N$  - целое число [137], например, алмазоподобная сетка обозначается  $sqc6$ ;
- символы Фишера  $k/m/fn$  для трехпериодических сферических упаковок [141], в этих обозначениях алмазоподобная сетка записывается как  $4/6/c1$ .
- Трехбуквенные обозначения сеток цеолитов пишутся заглавными буквами (см. пример далее).

База данных цеолитных структур (*Database of Zeolite Structures*, <http://www.iza-structure.org/databases/>) содержит информацию о структурных типах цеолитов, которые были одобрены Структурной комиссией Международной цеолитной ассоциации. На конец 2023 года общее количество составило 264 структурных типа. Пример дескрипторов структурного типа цеолита FAU приведен на Рис. 21.

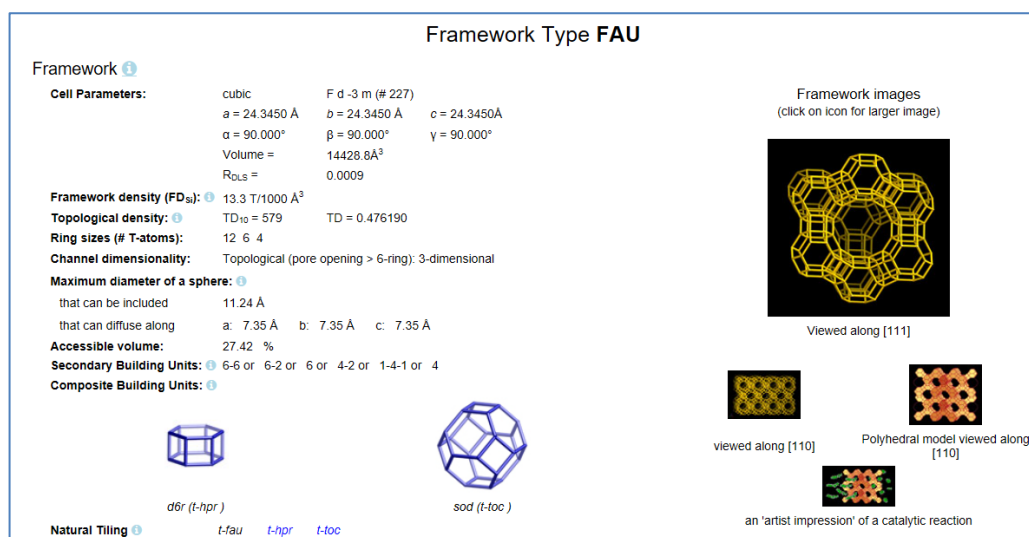


Рис. 21. Сведения о типе каркаса FAU из базы данных цеолитных структур.

Ресурс сеток химических структур (*Reticular Chemistry Structure Resource*, RCSR, <http://rcsr.anu.edu.au/nets>) создан для разработки новых и анализа известных кристаллических структур. Содержит 3D-, 2D-, 1D- и 0D-сетки и их дескрипторы (соответственно, 3579, 203, 11 и 143 на окончание 2023 года). Набор дескрипторов одной из сеток приведен на Рис. 22.

**fau**

RCSR reference: <http://rcsr.net/nets/fau>

**names:** faujasite, 4/4/c17, sqc13519, 4\*3.6 D  
**key words:** zeolite net, simple tiling, good  
**references:** J. Solid State Chem. 178, 2533-2554 (2005)

embed type	space group	volume	density	genus	td10	deg freedom
1a	Fd-3m	505.6330	0.3797	49	579	4

a	b	c	alpha	beta	gamma
7.9667	7.9667	7.9667	90.0	90.0	90.0

**vertices: 1**

vertex	cn	x	y	z	symbolic	Wyckoff	symmetry	order
V1	4	0.0362	0.1250	0.3025	x, y, z	192 I	1	1

vertex	CS <sub>1</sub>	CS <sub>2</sub>	CS <sub>3</sub>	CS <sub>4</sub>	CS <sub>5</sub>	CS <sub>6</sub>	CS <sub>7</sub>	CS <sub>8</sub>	CS <sub>9</sub>	CS <sub>10</sub>	cum <sub>10</sub>	vertex symbol
V1	4	9	16	25	37	53	73	96	120	145	579	4.4.4.6.6.12

**edges: 4**

edge	x	y	z	symbolic	Wyckoff	symmetry
E1	0.0806	0.0806	0.3025	x, x, z	96 g	m
E2	0.0806	0.1694	0.3025	x, 1/4-x, z	96 g	m
E3	0.0000	0.1612	0.3388	0, y, 1/2-y	96 h	2
E4	0.9919	0.1250	0.2581	x, y, 1/4-x	96 g	m

**tiling:**

tiling	dual	vertices	edges	faces	tiles	D-symbol
$2[4^6.6^6.2] + [4^6.6^6.8] + [4^6.18.6^4.12^4.4]$	-	1	4	5	3	24

Рис. 22. Сведения о типе каркаса FAU из базы данных RCSR.

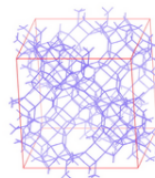


Проект EPINET (*Euclidean Patterns in Non-Euclidean Tilings*, <http://epinet.anu.edu.au>) исследует двумерные гиперболические ( $H^2$ ) тайлинги, как источник кристаллических каркасов (или сеток) в трехмерном евклидовом пространстве ( $E^3$ ). Трехпериодичные сетки получены при помощи абстрактной математической процедуры проектированием на трехмерное евклидово пространство тайлингов, которые построены в двумерном гиперболическом пространстве. EPINET содержит 3D- и 2D-сетки и их дескрипторы (соответственно, 14 646 и 2451 на окончание 2023 года). Набор дескрипторов одной из сеток приведен на Рис. 23. База EPINET уникальна тем, что сетки в ней получены с помощью абстрактной процедуры, не имеющей никакого отношения к реальным кристаллическим веществам. Поэтому EPINET-сетки могут рассматриваться как своего рода «случайный» набор, и анализ встречаемости их в реальных структурах позволяет оценить вероятность реализации произвольной, случайным образом выбранной сетки.

### sqc13519

Index
« sqc13518 sqc13520 »

Vertex Degree:	{4}
Transitivity (vertex,edge):	(1,4)
Nodes/Primitive Unit Cell:	48
Nodes/Asymmetric Unit:	1
Other Names:	fau, Faujasite, dia-i, 4/4/c17
Spacegroup:	Fd-3m
Symmetry Class:	cubic
Long Vectors:	N



CrystalMaker Unit Cell CM File

### Unit Cell:

a	b	c	alpha	beta	gamma
7.96626	7.96626	7.96626	90.0	90.0	90.0

### Atoms:

### Coordination Sequences:

Number	Coord	X-pos	Y-pos	Z-pos	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
1	4	0.03624	0.62500	0.30253	4	9	16	25	37	53	73	96	120	145

Рис. 23. Сведения о типе каркаса FAU из базы данных EPINET.

### 1.4.3. Библиотеки металл-органических координационных полимеров

Современное развитие химии координационных соединений неразрывно связано со стремлением заменить классический метод проб и ошибок компьютерным дизайном новых материалов с требуемыми свойствами. В свою очередь, существует необходимость в создании баз данных известных и гипотетических МОКП и их пополнении информацией о структурных характеристиках и свойствах. Создание таких баз данных делает возможным быстрый поиск требуемых материалов. Далее перечислены известные базы данных МОКП:

1) Первая база данных 137 953 гипотетических МОКП Уилмера [142], построенных на основе геометрических принципов связывания 120 структурных единиц. Этот метод не ограничивается известными топологическими типами и гораздо быстрее метода автоматизированной сборки вторичных строительных единиц (BCE) (см. Раздел 1.5.3), поскольку не требует минимизации энергии на каждом шагу. Тем не менее перебор ограничен одним типом узла и двумя типами линкеров.

2) База данных цеолитоподобных цинк-имидазолатных МОКП [143] полученных из 300 000 структур гипотетических цеолитов заменой атомов кремния на атомы цинка и атомы кислорода на имидазолатные связки.

3) База данных реальных МОКП Голдсмита [144]. Получена в результате анализа реальных структур, опубликованных в Кембриджском банке структурных данных [1]. Содержит структурную информацию и данные о пористости и площади поверхности пор для около 22 700 неупорядоченных МОКП, которая может быть использована для квантовохимических расчетов и других видов моделирования (удалены сольватные и клатратные молекулы). В частности, авторы использовали ее для поиска материалов, оптимальных для хранения водорода.

4) База данных CoRE MOF готовых для расчетов реальных МОКП Чунга [145] содержит структуры более 14 тыс. МОКП и была построена похожим на предыдущий вариант способом анализа CSD, но с использованием более строгих критериев отбора. Имеется информация о максимальном размере пор и минимальном размере каналов.

5) Фирст с соавторами [146, 147] создали базу данных нанопористых материалов MOFOMICS. Авторы разделили пористую структуру на каналы и поры, которые они аппроксимировали простейшими геометрическими фигурами, а именно сферами и цилиндрами. В базе данных размещена информация о размерах этих сфер и цилиндров для 251 структуры реальных МОКП и 1424 гипотетических.

6) Сотрудники Кембриджского центра структурных данных создали подмножество CSD MOF [148], которое на начало 2022 г. содержало 114 тыс. структур МОКП, включающих 33, 24 и 27 тысяч 1D-, 2D- и 3D-периодических, соответственно.

## **1.5. Анализ структурных данных машинными методами**

Добыча знаний из экспериментальных данных является необходимым шагом в развитии науки. Когда объем данных достигает критического уровня, их классификация, а затем разработка моделей, концепций и теорий, базирующихся на созданной классификации и объясняющих корреляции в данных, становятся необходимыми шагами для достижения уровня, на котором возможно достоверное прогнозирование. Кристаллохимия не является исключением: первые кристаллохимические закономерности, такие как правила Полинга или правила Юма-Розери, были сформулированы в 1920-х годах, когда число определений кристаллической структуры исчислялось десятками. В настоящее время в кристаллографических базах данных хранится более миллиона структурных экспериментов, но мы по-прежнему пользуемся старыми правилами, хотя и знаем множество

исключений. Как мы можем извлечь новые надежные закономерности из этого огромного количества данных? Очевидно, что человек вряд ли способен среди множества особенностей строения выделить важнейшие параметры, отвечающие за архитектуру и свойства тысяч кристаллических веществ. Накопление кристаллографических данных в электронных базах данных приводит к привлечению машинных методов для решения этой задачи [10, 11, 12]. Будучи несомненно перспективными и основанными на хорошо разработанных математических алгоритмах, эти методы, однако, сталкиваются с серьезными трудностями при применении к исходной информации о кристаллических структурах. Неоднородность информации, собранной в различных базах данных, ошибки, неточности и неопределенности в данных, отсутствие надежных структурных дескрипторов и другие проблемы до сих пор сдерживают широкое распространение машинных методов в кристаллохимии и делают их текущие результаты весьма ограниченными. Таким образом, для упорядочения и понимания остро необходима разработка универсальных схем (моделей) машинной обработки экспоненциально возрастающей кристаллографической информации [149]. Геометрические и топологические модели базируются на богатом и достоверном экспериментальном материале, который может быть использован для открытия новых знаний с помощью современных машинных методов анализа данных [120, 150].

### **1.5.1. Поверхности потенциальной энергии**

При изучении межатомных взаимодействий в структурных единицах кристалла и между ними, а также для поиска их оптимальной конформации используют поверхности потенциальной энергии, которые характеризуют зависимость общей энергии рассматриваемой системы от некоторых ее параметров. Для простых систем форму таких поверхностей рассчитывают квантово-химическими методами [151, 152], для сложных - методами

молекулярной механики [153]. При анализе сложных систем, как правило, ограничивают число параметров, изучая сечения поверхности потенциальной энергии с помощью корреляционных диаграмм.

Хотя молекулярная механика рассматривает атомные системы с позиций классической физики, а не квантовой механики, ее методы широко применяются в настоящее время. Немалую роль в корректности расчетов играют среднестатистические геометрические характеристики молекул, которые используют для параметризации методов молекулярной механики и, в некоторых случаях, получают при анализе строения кристаллических веществ на основе информации, содержащейся в компьютерных банках данных, таких как Кембриджская база структурных данных [1] и база данных о строении неорганических соединений [2]. Многие публикации посвящены оценке величин ван-дер-Ваальсовых радиусов [154], а также поиску взаимосвязей геометрических параметров структуры друг с другом [155]. Авторы таких работ при кристаллохимическом анализе ограничиваются только классическими характеристиками структуры: межатомными расстояниями, а также валентными и торсионными углами, что, на наш взгляд, существенно снижает информативность получаемых результатов. Кроме того, при анализе трехмерных объектов чаще рассматриваются одномерные распределения величин [156] и лишь в редких случаях обсуждаются двумерные распределения [157].

### **1.5.2. Машинный прогноз степени окисления**

Степень окисления атома является важным химическим дескриптором, однако его прямое определение требует сложного эксперимента, например, с использованием спектроскопии Мессбауэра [158] на основании слегка различающихся значения электронной плотности в области, занимаемой ядром. Формальное определение на основе электростатического баланса структуры обоснованно только для ионных соединений, но и в этом случае

не всегда однозначно. Поэтому зачастую степень окисления не содержится в базах структурных данных. Устойчивость кристаллических веществ, характер их взаимодействия с другими веществами, их каталитическая активность, оптические, адсорбционные и многие другие свойства зависят от степени окисления атомов, составляющих кристаллическую структуру [159]. Для конкретного атома металла набор значений степени окисления обусловлен его электронным строением [160, 161], а редкие значения этого дескриптора реализуются из-за особого химического состава и/или особенностей кристаллической структуры. Хотя определение степени окисления в большинстве случаев тривиально, для некоторых соединений эта задача является сложной.

Так, формальная степень окисления атомов металла может быть нецелой, как в  $\text{Co}^{+2.5}_{12}\text{O}_3(\text{cpb}^{3-})_8$ , [162] где  $\text{cpb}$  = 1,3,5-трис(4-карбоксилатофенил) бензол (NIGPAW), нулевой, как в  $\text{Ni}(\text{CO})_4$  [163] (FOJBUB03) или даже отрицательный, как в  $\text{K}^+\text{Na}^-(12\text{-crown-4})_2$  (REGCAN) [164]. В других случаях степень окисления кажется необычной, исходя из химической формулы, но ее можно определить после тщательного анализ кристаллической структуры, например, для  $(\text{LiH})\text{H}_n$ ,  $n=2, 6, 8$  [165]. Еще один сложный пример — электриды, где электростатический баланс не может быть достигнут без учета внеатомных электронов, как в  $\text{Sr}^{2+}_5\text{P}^{3-}_3\text{:e}^-$  [166] и  $\text{Na}^+\text{Ba}^{2+}_2\text{O}^{2-}_2\text{:3e}^-$  [167, 168].

Последние несколько лет проблема автоматизированного определения степени окисления неоднократно исследовалась для отдельных классов химических соединений. Так, модуль *pyratgen* [169] содержит код Python для определения степени окисления неорганических веществ. Состояния окисления переходных металлов в координационных соединениях, депонированные в Кембриджской структурной базе данных (CSD) [1], были определены Ривзом и др. [170] с использованием метода суммы валентностей связей и специальной методики определения зарядов лигандов.

Та же задача была решена методами машинного обучения [121], где в качестве признаков модели авторы рассматривали положение атома в Периодической таблице Д.И. Менделеева, электроотрицательность атома, форму атомного координационного полиэдра и распределение электронов по s, p, d и f-орбиталям. Площадь грани ПВ использовалась в качестве нормировочного коэффициента [65] при расчете различных характеристик (в том числе степени окисления) атомов в металлоорганических каркасах; для атомов меди доказана высокая точность модели – 98%.

Другие геометрические параметры ПВ использовались в качестве атомных дескрипторов и для определения степени окисления некоторых атомов (см. [26] и ссылки в ней). Однако до сих пор не была разработана универсальная схема автоматизированного определения степеней окисления всех видов атомов металлов как в неорганических, так и в металлоорганических кристаллических структурах.

### **1.5.3. Дизайн координационных соединений**

За последние десять лет совершен большой скачок в возможностях моделирования и предсказания структур КС [152] и МОКП, особенно, и практикуются следующие подходы:

1) Последовательная автоматизированная сборка ВСЕ [171]. В рамках метода рассматривается сборка многоатомных структурных единиц в структуру посредством потенциалов взаимодействия между определёнными атомами. Процесс моделируется множеством серий самосборки и минимизации энергии. С его помощью удалось предсказать ряд известных (HKUST-1, MOF-5) и гипотетических каркасов [172]. Однако эти расчеты требуют больших затрат компьютерных ресурсов и времени.

2) Прямой перебор возможных топологических типов для заданных структурных единиц или метод декорирования [56, 103]. Этим методом

возможно воспроизвести практически неограниченное количество полиморфов (изомеров) заданного состава с помощью замены узлов и ребер известных [60, 136] и математически сгенерированных [137] сеток на выбранные ВСЕ [173], которые затем могут быть проверены на стабильность молекулярно-механическим моделированием или квантовохимическими расчетами. Этот метод был использован для генерации гипотетических и известных структур карбоксилатных МОКП [174, 175], цеолитоподобных имидазолатных каркасов [176], пористых цианидов [177] и ковалентных органических каркасов [178].

3) Использование принципа изоретикулярности для варьирования лигандов [179, 180]. Подход состоит в искусственной замене однотипных по координационным возможностям лигандов в структуре известных каркасов. При этом можно получить однотипные структуры одного топологического мотива, но с разными свойствами (например, с разной плотностью и пористостью). Эта стратегия была использована для моделирования каркасных структур MIL-88B, MIL-88C и MIL-88D [181], аналогов MOF-5 [182], и предсказания высокопористого металлоорганического каркаса NU-100 [183].

4) Обратный топологический подход [184]. Этот метод основан на комбинировании принципов изоретикулярности и декорирования. Он был применен для генерации 204 гипотетических МОКП с использованием структурных единиц  $\{Zr_6O_4\}$  с 12 дикарбоксилатными и 36 тетракарбоксилатными лигандами для сеток топологии **fcu**, **ftw**, **scu** и **csq** [178]. Этот метод также был использован для генерации 48 новых каркасов на основе топологических мотивов **acs**, **cds**, **dia**, **hxx**, **lvt**, **nbo**, **pcu**, **rhr**, **sod** [185].

Несмотря на многообразие методов, существующие инструменты конструирования имеют существенные недостатки, а именно они работают



только с молекулярными высоко симметричными структурными единицами, и, следовательно, не способны перебирать все возможные варианты локального связывания несимметричных блоков или конструировать каркасы из полимерных одно- и двух-периодичных структурных единиц. Поэтому результаты конструирования микропористых каркасов с их использованием сильно ограничены, это подтверждается тем фактом, что многие синтезированные МОКП (из около 25000) даже весьма простого строения не были предсказаны этим методами. К ним относятся, например, МОКП, построенные на основе стержневых структурных единиц [186].

Известно несколько программных инструментов для создания структур кристаллических веществ – конструкторов. Авторы [187, 188] предлагают алгоритм сборки непосредственно из структурных единиц. Их графы с обозначенными точками связывания и сетка их центров тяжести (базовый топологический мотив) указываются в качестве исходной информации конструктора. Затем выполняется однозначное связывание структурных единиц по заданному периодическому мотиву, а полученный гипотетический каркас оптимизируется методами молекулярной механики и расчетами теории функционала плотности. Авторы [189] модифицируют известные кристаллические структуры, находя и заменяя в них одни структурные фрагменты на другие с помощью алгоритмов Ульмана и Прокруста. Первый алгоритм используется для поиска подграфов графа кристаллической структуры, которые изоморфны графу искомого фрагмента. Второй алгоритм выполняет замещение в кристаллической структуре найденного фрагмента на заменяемый фрагмент.

Описанные выше конструкторы используют файлы конкретных структурных единиц, а не их библиотеки, и работают с ограниченным набором сеток из библиотеки RCSR [136]. При таком подходе нельзя перебирать возможные варианты сборки структурных единиц в каркас и оценивать вероятность их реализации, так как рассматривается только один

строго детерминированный способ координации структурных единиц. Известны программные инструменты для сборки молекулярных и надмолекулярных фрагментов структур [190], которые могут быть полезны для генерации ВСЕ.

Упаковка структурных единиц редко анализируется и, как правило, только для простейших систем (высокая кристаллическая симметрия и относительно симметричные структурные единицы довольно простого химического состава). Отсутствуют унифицированные методы, подходящие для анализа способов упаковки в веществах различной химической природы. Традиционно упаковка одного из наиболее распространенных структурных типов берется как образец, а расположение структурных единиц в конкретном соединении описывается как один из его искаженных вариантов. Наиболее интересные результаты были получены при анализе молекулярных упаковок (теория Китайгородского [191, 192] и понятие структурных классов [193]), структур металлов (теория плотной упаковки) и ионных соединений (правило 12 соседей [194], модель катионных матриц [195] и кристаллохимическая модель атомных взаимодействий [196]).

#### **1.5.4. Общая схема обработки и использования кристаллографических данных**

В заключении первой главы рассмотрим общую схему обработки и использования кристаллографических данных (Рис. 24). Если требуется найти новые материалы с нужным свойством, то, как правило, исходят (Этап I) из известных кристаллоструктурных данных и результатов квантово-химических расчетов кристаллических веществ. Такие данные получают либо из CIF-файлов, хранящих отдельные структуры, либо из кристаллоструктурных баз данных (ICSD и CSD), либо веб-сервисов ([materialsproject.org](http://materialsproject.org), [aflow.org](http://aflow.org) [134]). Кроме того, максимально полно собирают из первоисточников и баз данных информацию о прогнозируемом

свойстве и его измерениях для конкретных веществ и материалов. В качестве свойств могут быть выбраны окраска, плотность и форма кристаллов, их температура плавления, показатель преломления, проводимость, люминисцентные характеристики, каталитическая активность, размерность и топологический тип структурной сетки и многое другое.



Рис. 24. Стадии интеллектуального поиска материалов с заданным свойством.

После выбора прогнозируемого свойства, на Этапе II разрабатывают геометрико-топологическую модель, формируют список дескрипторов, оказывающих влияние на это свойство, и рассчитывают их. Набор подходящих дескрипторов существенно (иногда полностью) зависит от

выбранного свойства. Рассчитанные данные собирают в таблицы, обрабатывают (исключают ошибки, противоречия, заполняют пропуски) и преобразуют, при необходимости, в нужный формат, формируя, в итоге на Этапе III, пространство признаков для поиска корреляций и машинного обучения, которые осуществляют на Этапе IV.

На этом этапе, по найденным корреляциям, создается база знаний об изучаемом свойстве и прогностическая схема. Строго говоря, при построении такой схемы мы не знаем *a priori*, какой метод машинного обучения лучше всего подходит для прогнозирования конкретных целевых параметров с конкретным набором дескрипторов, указанным для данного набора данных. Например, *Python*-библиотека *Scikit-learn* [197] содержит классификаторы случайного леса (RF), метода опорных векторов (SVC), дерева решений (CART), К-ближайших соседей (KNN), логистической регрессии (LR) и гауссовского наивного байесовского (GNB) классификаторов. Выбор наиболее подходящего классификатора является одной из задач машинного обучения. Для ее решения модель обучают на некоторой выборке с размеченными данными, то есть разделенными на классы по изучаемому свойству. Затем проверяют ее качество на отложенной выборке, сравнивая прогнозируемый моделью класс с действительным классом.

При оценке надежности выбранного алгоритма используют численные критерии *Достоверность* (Accuracy), *Точность* (Precision), *Полнота* (Recall) и *F1-мера* [119, 197, 198]. *Достоверность* отвечает доли правильных предсказаний. Критерий *Полнота* демонстрирует способность алгоритма обнаруживать конкретный класс и является мерой правильно идентифицированных положительных идентификаций; *Точность* показывает способность отличать этот класс от других классов, в то время как *F1-мера* есть среднее гармоническое *Точности* и *Полноты* или мера точности схемы с учетом распределения данных по классам. Все критерии варьируются от 0 до 1 и достигают максимального значения при наилучшем качестве. В итоге

выбирается модель с оптимальными значениями перечисленных выше критериев качества.

Для создания надежной модели применяют перекрестную проверку, при которой вся выборка делится на несколько частей. Одна часть используется для тестирования модели, а остальные части – для ее обучения. Затем меняют часть для тестирования и проводится повторное обучение на оставшихся частях. Так повторяется до тех пор, пока все части не будут использованы в качестве тестового набора. Наконец, критерии качества усредняются по всем комбинациям обучающего и тестового подмножества. В результате получают оценки качества выбранной модели с наиболее равномерным использованием имеющихся данных. Если выборку делят на пять частей, то такой метод называют пятикратной перекрестной проверкой.

На финальном Этапе V осуществляется прогноз новых материалов, из которых выбираются наиболее подходящие, в том числе с учетом других параметров (цены, токсичности, устойчивости и др.). При необходимости осуществляется проверка результатов прогноза с помощью более точной модели или прецензионных квантово-химических расчетов. В заключении материал синтезируют и проводят его экспериментальные испытания.

## Глава 2. Программный пакет ToposPro

### 2.1. Общее описание

#### 2.1.1. Система управления структурными базами данных

Разработанный нами пакет программ ToposPro представляет собой интегрированную интерактивную многопользовательскую среду, позволяющую редактировать и анализировать кристаллографическую информацию (Схема 1). При разработке пакета преследовались две основные цели:

- (1) компьютерная реализация и комбинация вышеупомянутых методов кристаллохимического анализа в рамках единой информационно-аналитической системы, чтобы пользователь мог провести полное геометрико-топологическое исследование структуры конкретного кристаллического соединения по исходным кристаллографическим данным;
- (2) автоматизированный анализ большого числа кристаллических структур из базы данных с помощью унифицированных алгоритмов в целях поиска кристаллохимических корреляций и закономерностей.

Базовой программой пакета является система управления базами данных (СУБД), в которой реализованы все стандартные операции редактирования и поиска кристаллографической информации в базах данных уникального формата. Хотя некоторые из возможностей СУБД ToposPro аналогичны возможностям СУБД ICSD [2] и CSD [1], ToposPro предоставляет средства для организации своих собственных баз данных, которые отсутствуют во всех крупных кристаллографических информационных системах. Все прикладные программы, за исключением StatPack, интегрированы в СУБД и позволяют анализировать активное соединение или группу соединений в базе данных.



**Схема 1.** СУБД программного комплекса ToposPro и его модули.

### 2.1.2. Прикладные программы ToposPro

Комплекс ToposPro содержит 8 прикладных программ, которые имеют следующие возможности (с учетом новых разработок):

ADS (Automatic Description of Structure). Выявление структурных групп, определение их состава, ориентации, размерности и связывания в различных структурных представлениях. Вычисление топологических инвариантов (координационных последовательностей, размеров циклов разного типа) и выполнение топологической классификации. Построение молекулярных ПВ и вычисление их геометрических характеристик. Построение разбиений для 3D-сеток. Поиск и классификация переплетений

1D, 2D или 3D расширенных структур. Автор реализовал в этой программе расчет молекулярных ПВ и их характеристик.

AutoCN. Выполняет автоматический поиск химических связей в кристалле для определения матрицы смежности структуры и представляет ее, как СГ, с использованием алгоритмов Domains [37], а также Ranges и Solid Angles. Метод Ranges использует классический подход, основанный на межатомных расстояниях и атомных радиусах. Метод Solid Angles рассчитывает матрицу смежности по величине телесных углов граней ПВ и предназначен для систем с ненаправленными связями (интерметаллиды, ионные подрешетки). Программа AutoCN также определяет специфические и ван-дер-Ваальсовы межатомные контакты. Это единственная в мире программа, которая позволяет сохранять полную информацию обо всей системе связей в кристаллической структуре, используя модель помеченного фактор-графа (см. Раздел 1.2.4) и уникальный формат.

DiAn. Расчет межатомных расстояний и углов связи, определение формы координационных фигур с помощью угловых фингерпринтов или методом их совмещения с эталоном. Расчет угловых фингерпринтов реализован в рамках данной работы.

Dirichlet. Построение ПВ и ван-дер-Ваальсовых полиэдров для атомов и пустот. Вычисление геометрических характеристик атомных доменов, молекулярных полиэдров и доменов пустот. Поиск позиций пустот и каналов. В рамках данной работы методы «заворачивания подарка» [199] и расчета положений элементарных пустот, каналов и их конгломератов впервые реализованы и адаптированы для кристаллических и низкосимметричных систем. Предложены новые дескрипторы межатомных контактов, атомов и структурных единиц (**Таблица 3**), созданы программные инструменты для подготовки объемных выборок дескрипторов, которые могут быть использованы для машинного обучения.



HSite. Генерация позиций атомов водорода [200] на основании геометрических принципов с использованием полиэдрической модели.

IsoCryst. Визуализация кристаллической структуры и расчет ее геометрических параметров, построение ПВ и координационных полиэдров. В рамках данной работы, в программе впервые в мире реализована возможность осуществлять рост кристаллической структуры по контактам заданного типа, используя данные матрицы смежности, представленные в виде СГ.

IsoTest. Классификация кристаллических структур по топологическим и структурным типам. Сравнительный анализ атомных сеток и упаковок.

StatPack. Статистическая обработка файлов данных, создаваемых программами Dirichlet, DiAn и ADS, визуализация диаграмм с данными о ПВ атомов, а также их граней, ребер и вершин.

Автором были созданы в комплексе ToposPro программы Dirichlet, IsoCryst и StatPack. Разработка остальных программ осуществлялась в соавторстве с проф. В.А. Блатовым.

Возможности ToposPro активно используются в настоящее время мировым научным сообществом для решения различных задач физической химии, в том числе, прикладного характера. Таковыми являются создание минибаз структурных данных для поиска закономерностей в их строении, исследование и анализ свободного пространства пористых структур [201] (Рис. 25), скрининг структурной информации для отбора потенциальных твердых электролитов [128, 202] (Рис. 26), расчет натуральных тайлов [84] (Рис. 27), а также топологический анализ МОКП [203] и систем водородных связей [75, 204]. Программный комплекс ToposPro задействован в университетских курсах «Кристаллохимия» и «Строение вещества», а также при подготовке многих квалификационных работ.

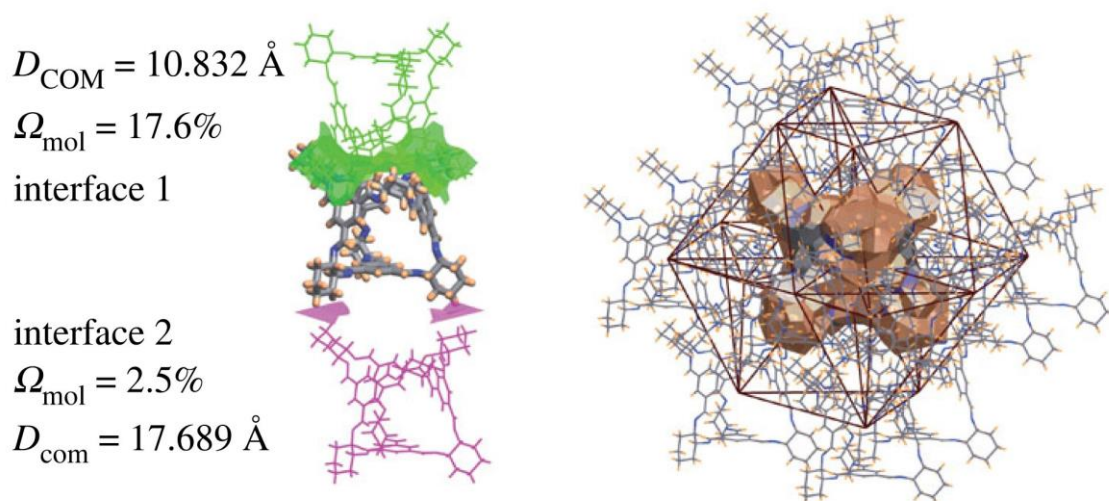


Рис. 25. Анализ свободного пространства в структуре  $\text{C}_{72}\text{H}_{84}\text{N}_{12},3(\text{CHCl}_3)$  (FOXLAG) - молекулярный ПВ и базовая сетка молекулярной упаковки, фрагмент Рис. 2 из [201].

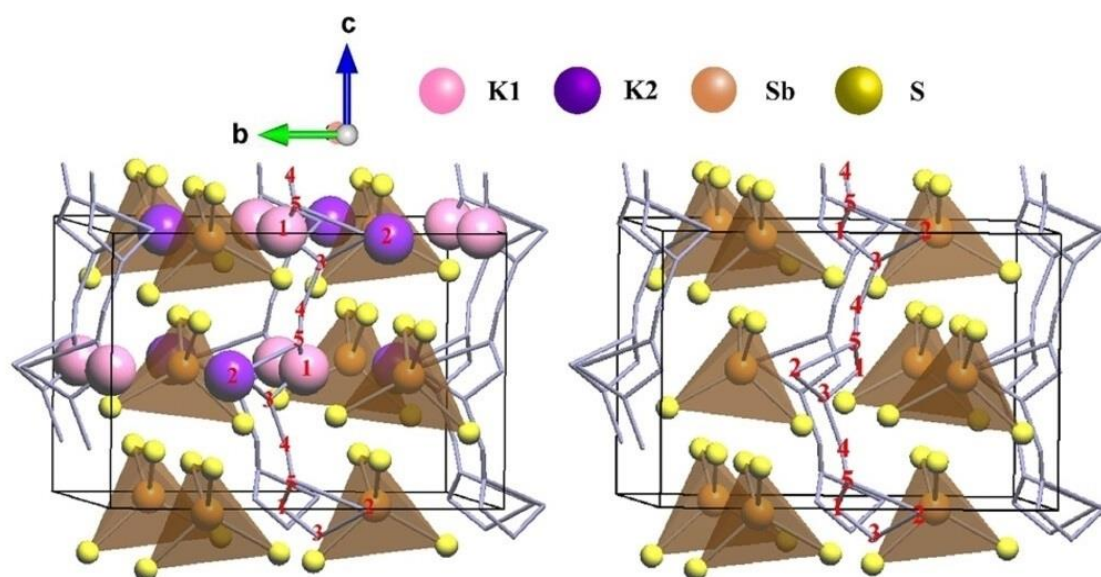


Рис. 26. Расчет карты миграции ионов  $\text{K}^+$  в виде графов Вороного для структуры  $\beta\text{-K}_3\text{SbS}_4$ , Рис. 1 из [202].

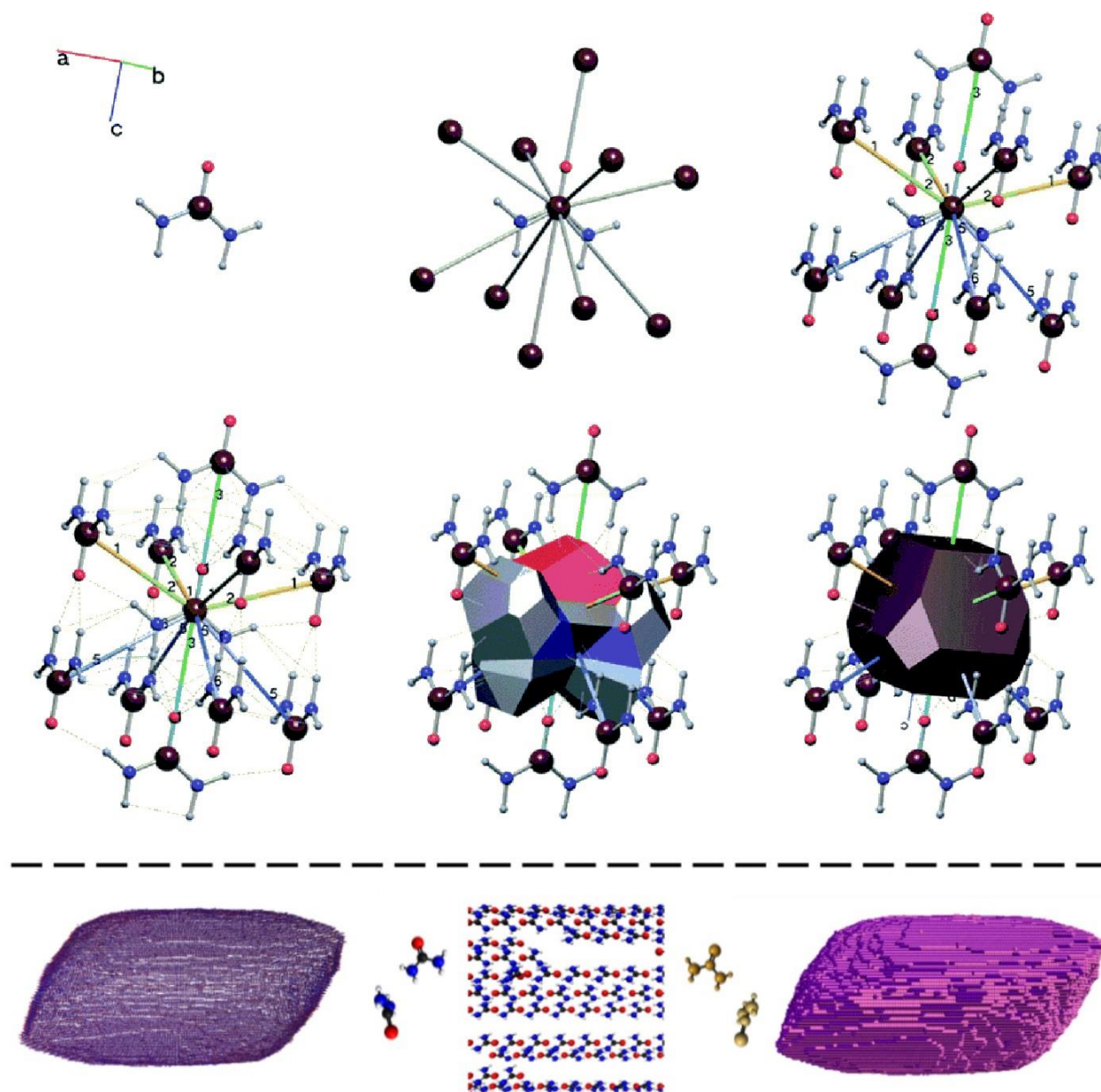


Рис. 27. Представления сборки кристаллической структуры мочевины  $(\text{NH}_2)_2\text{CO}$  (вверху, Рис. 13 из [84]) и изображения кристалла мочевины, рассчитанные с помощью программы CrystalGrower на основании данных программного комплекса ToposPro (внизу, фрагмент Рис. S13 из [84]): слева и в центре - для построения кристалла используются атомы и связи, справа - для построения кристалла используются сферы среднего размера, помещенные в центр каждой молекулы.

## 2.2. Методы и дескрипторы для распознавания формы координационных фигур

### 2.2.1. Обозначение координационной фигуры

Как уже указывалось выше, *координационная фигура атома металла или лиганда* определяется по их ближайшему окружению в базовой сетке (Рис. 7) [60, 103, 110, 114] и может быть охарактеризована угловым фингерпринтом. Классификация координационных фигур не является тривиальной задачей, особенно для лигандов, однако она играет важную роль в прогнозировании структуры, поскольку общая топология координационной группы существенно зависит от формы координационной фигуры [103]. Мы использовали названия координационных фигур по Хартшорну [205] (если имеются, иначе были предложены дополнительные наименования) вида NAME- $n\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ , где NAME - имя класса фигур,  $n$  - число вершин у фигуры (координационное число атома или лиганда), а  $\{\alpha_i\}$  - набор углов фигуры (в градусах), расположенных в порядке возрастания и делимых на  $\delta$ ;  $m = n(n-1)/2$ . Поскольку углы фигуры могут немного отличаться от эталона  $\alpha_i$ , мы относим фигуру к определенному типу, если ее углы попадают в группу вокруг  $\alpha_i$ ; кластерный анализ пространства углов координационных фигур был выполнен с использованием стандартных библиотек Python [206]. Например, координационные фигуры атома лиганда и металла в кристаллической структуре (этиленедиолolato)<sub>2</sub>Ti [46] относятся к классам уголковым A-2{75} и тетраэдраэдрическим T-4{90,105<sup>4</sup>,135}, соответственно (Рис. 28), хотя их углы ({69.13°} и {88.84°, 104.66°, 104.66°, 110.87°, 110.87°, 129.50°}) отличаются от значений эталона.

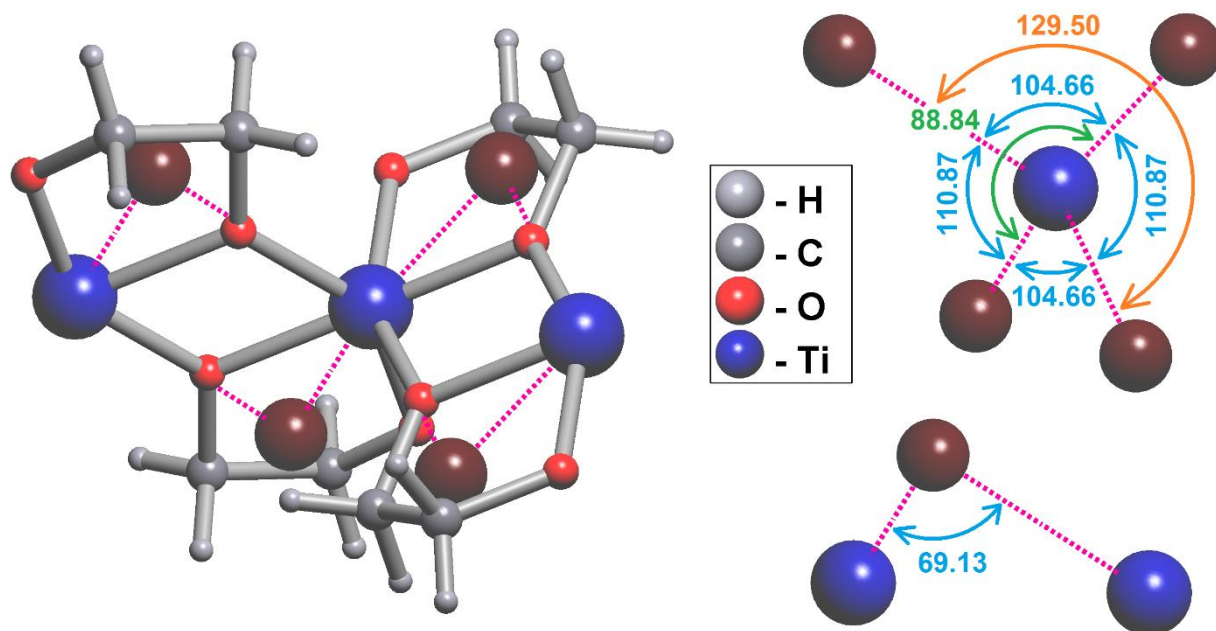


Рис. 28. Фрагмент структурной и базовой сеток (слева), а также координационные фигуры металла T-4{90,105<sup>4</sup>,135} и лиганда A-2{75} (справа) с их плоскими углами (в град.) для цепочечной кристаллической структуры бис(этилендиолата)-титан(IV)  $[\text{Ti}(\text{C}_2\text{H}_4\text{O}_2)_2]$  (KEPFEQ) [46]. Коричневые шары соответствуют центрам тяжести лигандов.

Для определения формы координационной фигуры нами были предложены три метода, основанных на топологии ПВ, максимизации объема пересечения координационных полиэдров, а также сравнении угловых фингерпринтов.

### 2.2.2. Топология полиэдров Вороного

В первом подходе для классификации используется граф ребер упрощенного ПВ центрального атома координационной фигуры. Например, геометрически правильная тригональная бипирамида и квадратная пирамида имеют четко различающиеся формы ПВ: тригональную призму и тетрагональную пирамиду, соответственно (Рис. 29).



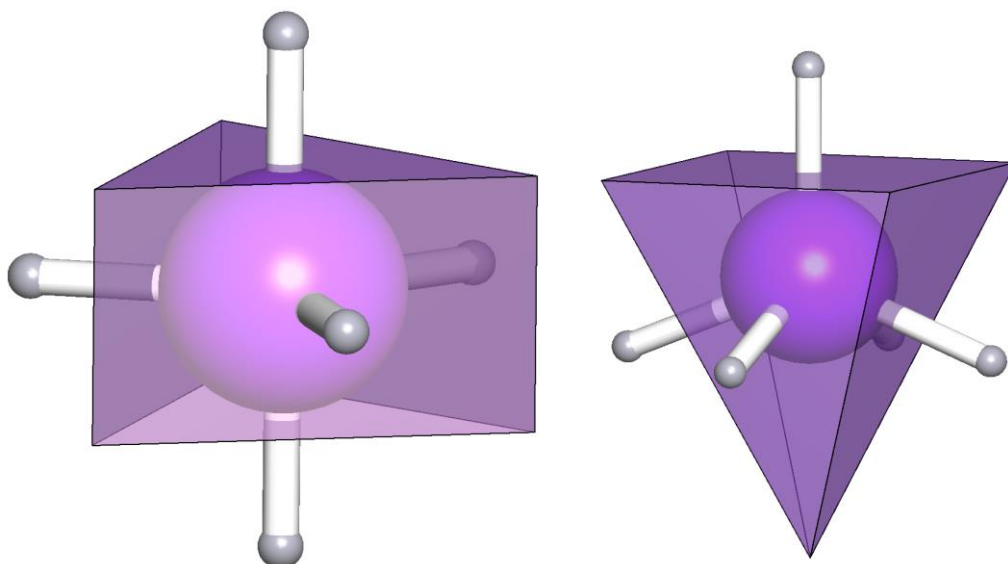


Рис. 29. Координационные фигуры и ПВ для треугольной бипирамиды (слева, топологический тип  $\{5/9/6-1\}$ ) и квадратной пирамиды (справа, топологический тип  $\{5/8/5-1\}$ ).

Преимущество использования ПВ вместо координационных полиэдров для классификации координационных фигур состоит в том, что геометрические искажения координационного полиэдра приводят к изменениям реберного графа ПВ и могут быть легко оценены (Рис. 30 внизу). В этом подходе при построении ПВ учитываются только соседние атомы, валентно-связанные с его центральным атомом. Чтобы сравнить такой ПВ с библиотечными топологическими типами, мы упрощаем его топологию путем последовательного сжатия кратчайших ребер. Размер ребра оценивается углом  $\gamma$ , под которым оно видно из центрального атома (Рис. 30 сверху).

Так, например, треугольная призма топологически отличается от квадратной пирамиды дополнительным ребром (Рис. 30). Анализ распределения реберных углов ПВ атомов Cu или Zn показал, что мы можем стянуть ребро, если  $\gamma < 30^\circ$ , что соответствует первому минимуму распределения (Рис. 31). Этот алгоритм очень быстро сравнивает координационные фигуры, как с одинаковым, так и с разным (даже большим) числом вершин (Рис. 32).

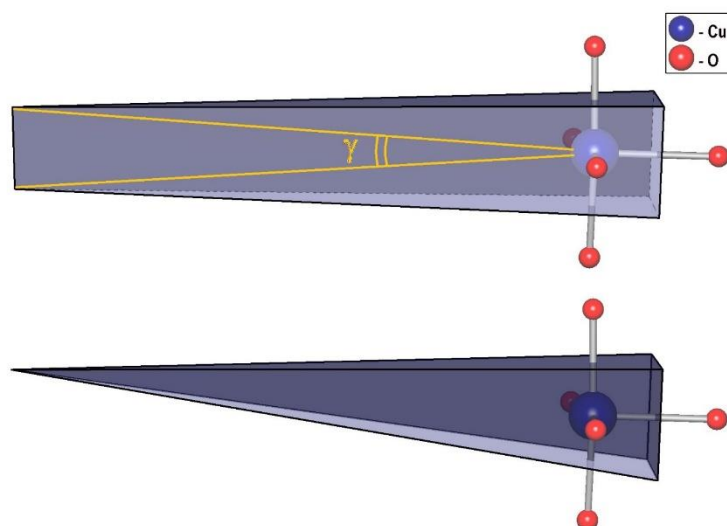


Рис. 30. ПВ (тригональная призма) атома Cu, построенный только с учетом валентно-связанных соседей (вверху), и его упрощенная форма (тетрагональная пирамида), полученная при стягивании ребра многогранника с  $\gamma = 8.07^\circ$  в структуре  $[\text{Cu}\{\text{C}_2(\text{COO})_2\}(\text{H}_2\text{O})_3] \cdot \text{H}_2\text{O}$  ( $\text{C}_2(\text{COO})_2$  = ацетилендикарбоксилат; MACGAZ) [207].

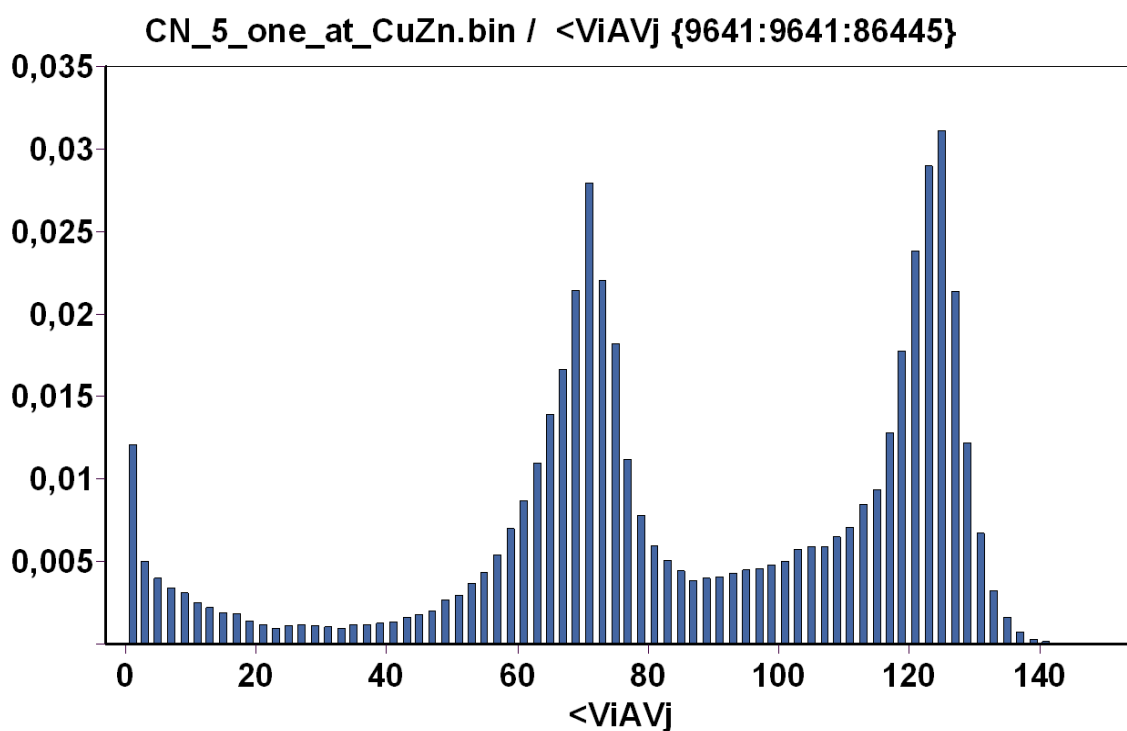


Рис. 31. Распределение углов у атома металла, опирающихся на ребрах ПВ.

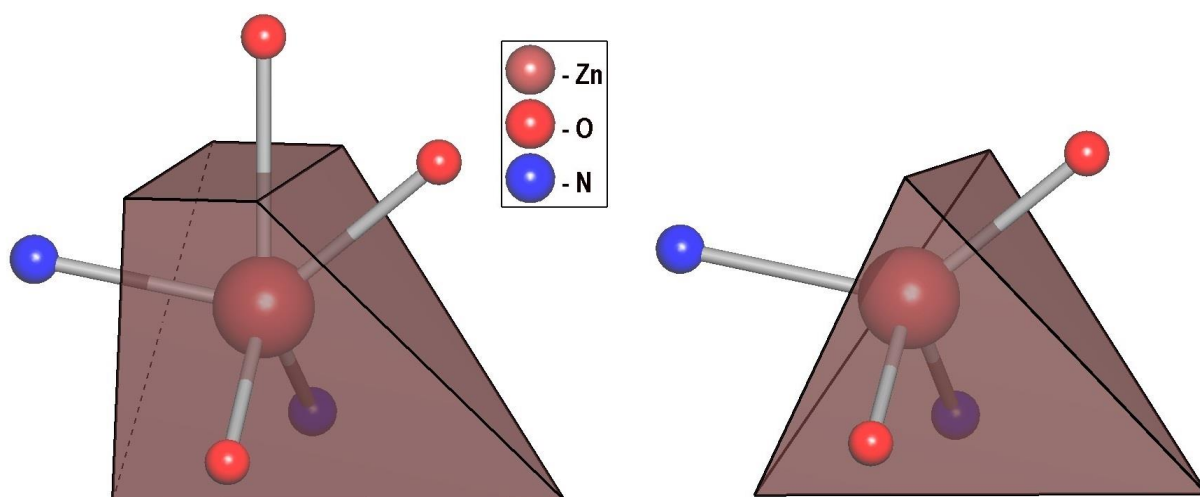


Рис. 32. ПВ, построенный с учетом валентно-связанных соседей (слева), и его упрощенный вариант, полученный путем стягивания ребер многогранника с  $\gamma < 30^\circ$  (справа) для атома цинка в структуре  $\text{Zn(ATIBDC)(bpy)} \cdot 3\text{H}_2\text{O}$  ( $\text{bpy}$  = 4,4'-бипиридин;  $\text{H}_2\text{ATIBDC}$  = 5-амино-2,4,6-тригидроксибензойная кислота; VOMSEW01) [208].

### 2.2.3. Максимизация объема пересечения координационных полиэдров

Подход основан на итеративном наложении координационного и эталонного многогранников для сравнения путем сдвига, поворота или масштабирования координационного полиэдра. В отличие от известных методов наложения [57], в этом подходе в качестве критерия  $\omega$  подобия используется доля максимального пересечения двух многогранников ( $\omega = 0-100\%$ ; чем больше ее значение, тем ближе два многогранника друг к другу, Рис. 33). Самые известные алгоритмы [209, 210] основаны на переборе  $n$  точек сетки  $O(\varepsilon)$  в наборе параметров, определяющих изометрическое трехмерное пространство, и имеют вычислительную сложность не менее  $O(\varepsilon^{-3} n \log^{3.5} n)$ , где  $\varepsilon$  - точность вычисления объема. В результате эти алгоритмы не подходят для обработки больших выборок ( $>1000$  многогранников).



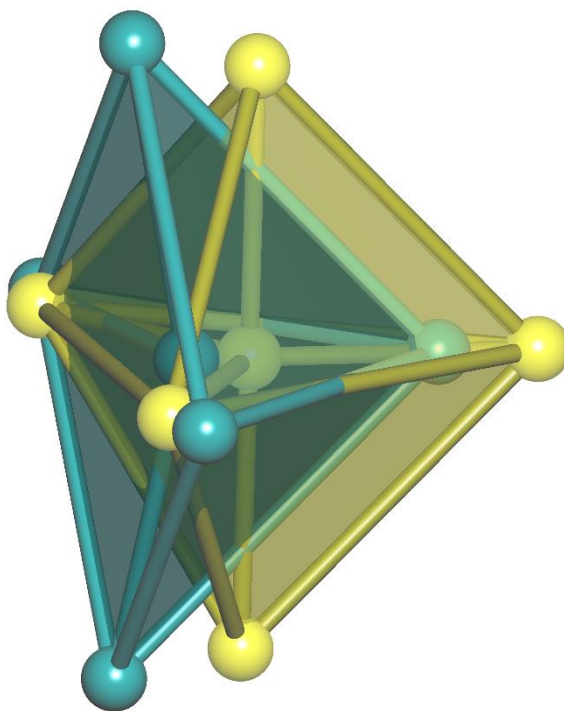


Рис. 33. Наложение треугольной бипирамиды и квадратной пирамиды, в результате чего объем их пересечения  $\omega = 75.4\%$  максимален.

Нами предложен новый алгоритм, основанный на градиентном спуске координат в пространстве изометрий трехмерного пространства и поиске минимума для каждой координаты с помощью метода золотого сечения. Проводится поиск минимума функции  $f(\varphi, \psi, \theta, x, y, z)$ , значение которой равно  $\omega$ . Углы  $\varphi$  и  $\psi$  определяют ось вращения многогранника;  $\theta$  - угол поворота вокруг оси;  $x, y, z$  координаты центроида перемещаемого многогранника относительно центроида неподвижного (опорного) многогранника. Алгоритм включает в себя следующие этапы:

- (i) Выравнивание расстояний от вершин координационной фигуры до ее центрального атома путем умножения векторов, соединяющих вершины с центроидом, на соответствующие множители. Полученный многогранник всегда выпуклый, т. е. эта процедура позволяет классифицировать даже невыпуклые координационные фигуры.
- (ii) Нормализация объема многогранника  $V$  на единицу путем деления координат вершин многогранника на  $V^{1/3}$ .

- (iii) Наложение центроидов многогранников и осей их тензоров инерции. Рассматриваются все шесть возможных ориентаций осей. В результате мы получаем хорошее начальное приближение, чтобы найти требуемый минимум.
- (iv) Перемещение рассматриваемого многогранника относительно эталона путем изменения одного из шести аргументов функции  $f$ . Оптимальное значение каждого аргумента определяется методом золотого сечения [211], а значение функции вычисляется путем построения пересечения двух многогранников с последующим определением объема пересечения. Метод золотого сечения одномерной минимизации очень экономичен. Его применение сокращает время вычислений с десятков минут до нескольких секунд по сравнению с алгоритмами  $\varepsilon$ -перебора [209, 210]. Этот шаг повторяется до тех пор, пока целевая функция  $f$  не отличается менее чем на заданную погрешность после последовательной минимизации всех шести аргументов.

Мы протестировали алгоритм для многогранников с числом вершин до 20 и выявили его высокую эффективность по сравнению с алгоритмами перебора [209, 210].

#### **2.2.4. Сравнение угловых отпечатков**

По аналогии с методами Дасса [62] и Ляхова [63], с помощью которых авторы распознавали лица и кристаллические структуры, соответственно, для определения формы координационной фигуры был разработан и использован алгоритм сравнения угловых отпечатков. Любой координационной фигуре соответствует набор валентных углов со значениями в диапазоне от  $0^\circ$  до  $180^\circ$ . Процедура расчета углового отпечатка по этим данным состоит из следующих этапов:

- диапазон возможных значений делится на равные интервалы (например, если число интервалов равно 10, то их длина  $\delta=18^\circ$ );
- для каждого интервала рассчитывается количество попадающих в него значений валентных углов;
- найденное для каждого интервала число попаданий аппроксимируется гауссовой функцией (8) со стандартным отклонением  $\sigma$ , равным половине интервала (например,  $\sigma=9^\circ$  для  $\delta=18^\circ$ ) и средним значением  $\mu$ , равным середине интервала.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

Используя итоговую функцию (9), являющуюся суммой всех интервальных гауссовых функций, можно посчитать компоненты  $F_i$  фингерпринт-вектора, имеющего размерность, равную числу интервалов.

$$F_i = 100 \frac{\sum_{k=1}^{180/\delta} f_k(\mu_i)}{\sum_{m=1}^{180/\delta} \sum_{k=1}^{180/\delta} f_k(\mu_m)} \quad (9)$$

Чтобы определить степень сходства двух координационных фигур достаточно найти расстояние между их фингерпринт-векторами  $F$  и  $F'$  по формуле (10).

$$\beta = \sqrt{\sum_{i=1}^{180/\delta} \left( \frac{F_i - F'_i}{2} \right)^2} \quad (10)$$

Координационная фигура из библиотеки эталонов, для которой расстояние  $r$  минимально, является искомой. Примеры фингерпринтов для некоторых координационных фигур представлены в **Таблице 6** и на Рис. 34. В рамках данной работы были созданы программный модуль для расчета фингерпринтов и библиотека фингерпринтов правильных пятикоординационных фигур, что позволило проводить машинную классификацию произвольных выборок из таких объектов. Библиотека содержала фингерпринты для тригональной бипирамиды (ТБРУ-5),

квадратной базоцентрированной пирамиды (SPY-5a), квадратной объемоцентрированной пирамиды (SPY-5b) и одношапочного тетраэдра (TF-5).

**Таблица 6.** Угловые фингерпринты для тетраэдра и правильных пятикоординационных фигур при  $\delta=18^\circ$  и  $\sigma=9^\circ$

Интервал углов( $^\circ$ )*	72 $\pm$ 9	90 $\pm$ 9	108 $\pm$ 9	126 $\pm$ 9	144 $\pm$ 9	162 $\pm$ 9	180 $\pm$ 9
Тетраэдр (Т-4)	0	7.6	77.8	14.6	0.1	0	0
Тригональная бипирамида (ТБРУ-5)	6.5	48.1	16.4	19.2	0.7	1.1	8
Базоцентр. квадратная пирамида (SPY-5a)	8.7	64.3	8.7	0	0	2.2	16.1
Объемоцентр. квадратная пирамида (SPY-5b)	7	45.5	26.8	0.8	5.3	14	0.7

\*Для неуказанных диапазонов значения фингерпринт-вектора равно нулю. Обозначения координационных фигур взяты из [205].

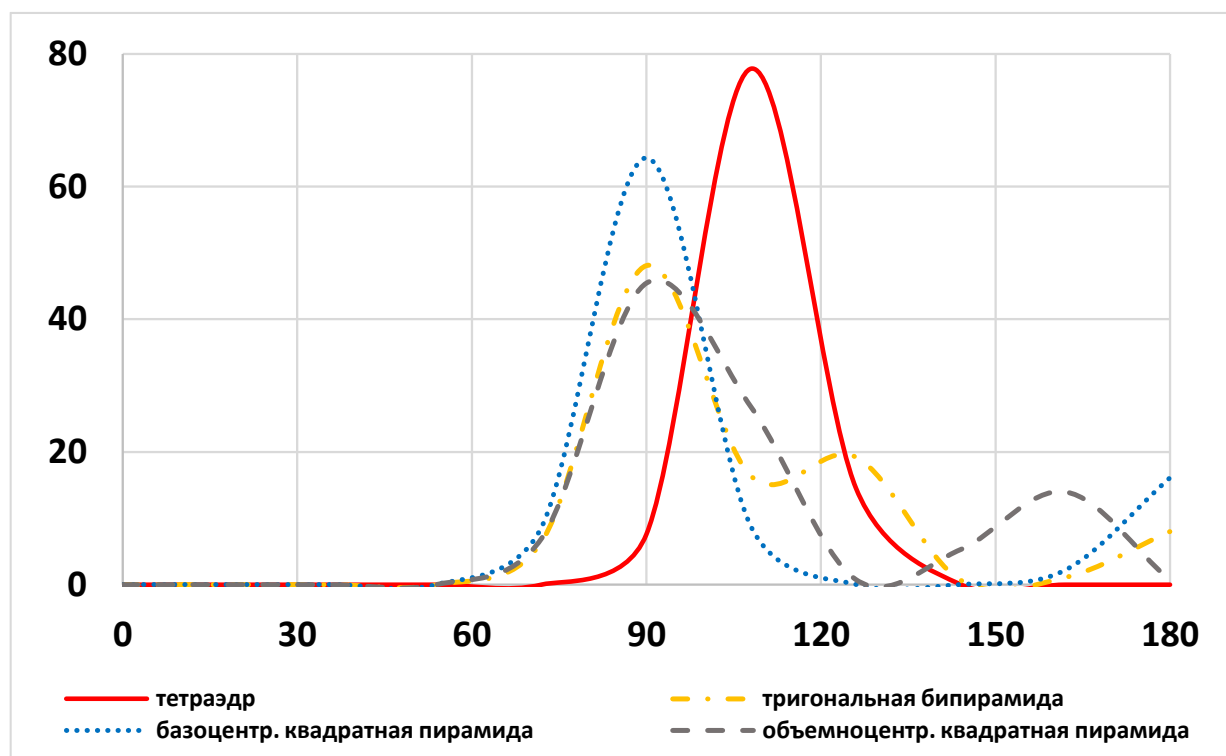


Рис. 34. Фингерпринты правильного тетраэдра, тетрагональной бипирамиды, а также базо- и объемоцентрированных квадратных пирамид.

При определении формы координационной фигуры исследуемого атома, назовем ее образец, необходимо рассчитать для нее отпечаток с теми же параметрами  $\delta$  и  $\sigma$ , а затем найти координационную фигуру в библиотеке эталонов с наиболее схожим отпечатком. Если координационные фигуры образца и эталона идентичны, то расстояние между их отпечатками равно нулю. При максимальном отличии координационных фигур расстояние стремится к 100 % (Таблица 7). Используем абсолютное значение расстояния  $\beta$  между отпечатками образца и эталона, как количественный критерий степени их схожести. Перебирая эталоны из библиотеки мы получим набор  $\beta_k$ , где  $k \in [1..N]$ , если  $N$  - количество эталонов в библиотеке. Для наиболее подходящего  $j$ -эталона значение будет минимальным  $\beta_j = \min(\beta)$ . В то же время библиотека может содержать альтернативный  $i$ -эталон, которому отвечает  $\beta_i$ , близкое по значению к  $\beta_j$ . Для таких случаев выберем разность  $\alpha_{ij} = \beta_i - \beta_j$  в качестве критерия однозначности определения наиболее подходящего эталона, когда величина  $\alpha_{ij}$  более 5 % и близка к расстоянию между отпечатками двух таких эталонов. Если значение  $\alpha_{ij}$  менее 5 %, то следует либо рассматривать оба эталона в качестве КФ образца, либо исключить из библиотеки один из таких эталонов. Так если использовать библиотеку из трех КФ (Таблица 7 без ТВРУ-5) и образцом является ТВРУ-5, то  $\beta_j = 13.2$  для эталона SPY-5b,  $\beta_i = 13.8$  для эталона SPY-5a, а  $\alpha_{ij} = 0.6$ , поэтому в данном случае необходимо учитывать обе КФ. Если добавить в библиотеку и ТВРУ-5, тогда  $\beta_j = 0$ ,  $\beta_i = 13.2$ ,  $\alpha_{ij} = 13.2$ , что даст однозначный ответ.

**Таблица 7.** Расстояния между отпечатками тетраэдра и правильных пятикоординационных фигур

Координационная фигура	T-4	ТВРУ-5	SPY-5a	SPY-5b
T-4	0	37.2	46.2	33.5
ТВРУ-5	37.2	0	13.8	13.2
SPY-5a	46.2	13.8	0	16.5
SPY-5b	33.5	13.2	16.5	0

## 2.3. Определение свободного пространства в структуре кристаллов

### 2.3.1. Оценка доли свободного пространства

Для определения доли свободного пространства кристалла нами введено понятие ван-дер-Ваальсова полиэдра. Если в случае ПВ все межатомные контакты делятся строго пополам, то в измененном алгоритме половинное деление сохраняется лишь в том случае, если межатомное расстояние меньше суммы ван-дер-Ваальсовых радиусов (использовались значения радиусов из [212]). Когда же межатомное расстояние больше этой величины, то контакт делится на удалении, равном ван-дер-Ваальсовому радиусу центрального атома полиэдра (Рис. 3).

### 2.3.2. Генерация подпространства пустот

Расчет центров пустот, в качестве которых принимаются вершины ПВ, осуществляется при помощи программы Dirichlet, входящей в комплекс ToposPro. Как правило, при построении разбиения Вороного учитываются все атомы структуры, однако возможно рассмотрение любого подмножества базисных атомов. Это может быть полезным, например, при изучении пустот в соединениях внедрения, когда часть пустот уже занята, и для их анализа требуется временно не рассматривать находящиеся в них атомы.

Подпространство пустот имеет два принципиальных отличия от подпространства атомов, которые необходимо учитывать при восстановлении его связности:

- расстояния между центрами пустот могут быть сколь угодно малы;
- пустоты априори не дифференцируются по типу, как атомы по химическому сорту.

$N$  близко расположенных центров пустот представляют собой  $N$ -конгломерат, фактически соответствующий одной полости достаточно большого размера, центр которой будет соответствовать геометрическому

центру конгломерата. Такая ситуация является обычной в соединениях, содержащих объемные полости, обрамленные большим количеством атомов (Рис. 35). В определенной степени этот случай подобен статистическому разупорядочению атомов. При необходимости можно осуществлять дополнительную дифференциацию пустот, например, по их размерам и природе атомного окружения или по другим признакам.

Прежде всего отметим, что не все вершины ПВ отвечают физически существующим полостям в кристаллической структуре, даже в случае *I*-конгломерата. В общем случае, любой четверке взаимосвязанных атомов, не лежащих на одной плоскости, соответствует вершина ПВ. Если все или часть этих атомов связаны между собой валентными связями и, в связи с этим, находятся близко друг от друга, объем пустоты, отвечающей такой вершине ПВ, может оказаться мал и недостаточен для внедрения других атомов (возможно, единственное исключение – гидриды металлов). Очевидно, что, если интерес представляют только большие полости и известно, что в структуре существуют дальние невалентные контакты между соседними атомами, такие вершины ПВ желательно исключить из рассмотрения. Так как каждой вершине ПВ, лежащей в общем положении, отвечают шесть контактов между четырьмя окружающими ее атомами, удобно ввести ранг вершины ПВ, равный числу соответствующих ей невалентных контактов и изменяющийся от 0 до 6. Можно ожидать, что вершинам наивысшего ранга будут отвечать полости наибольшего объема.

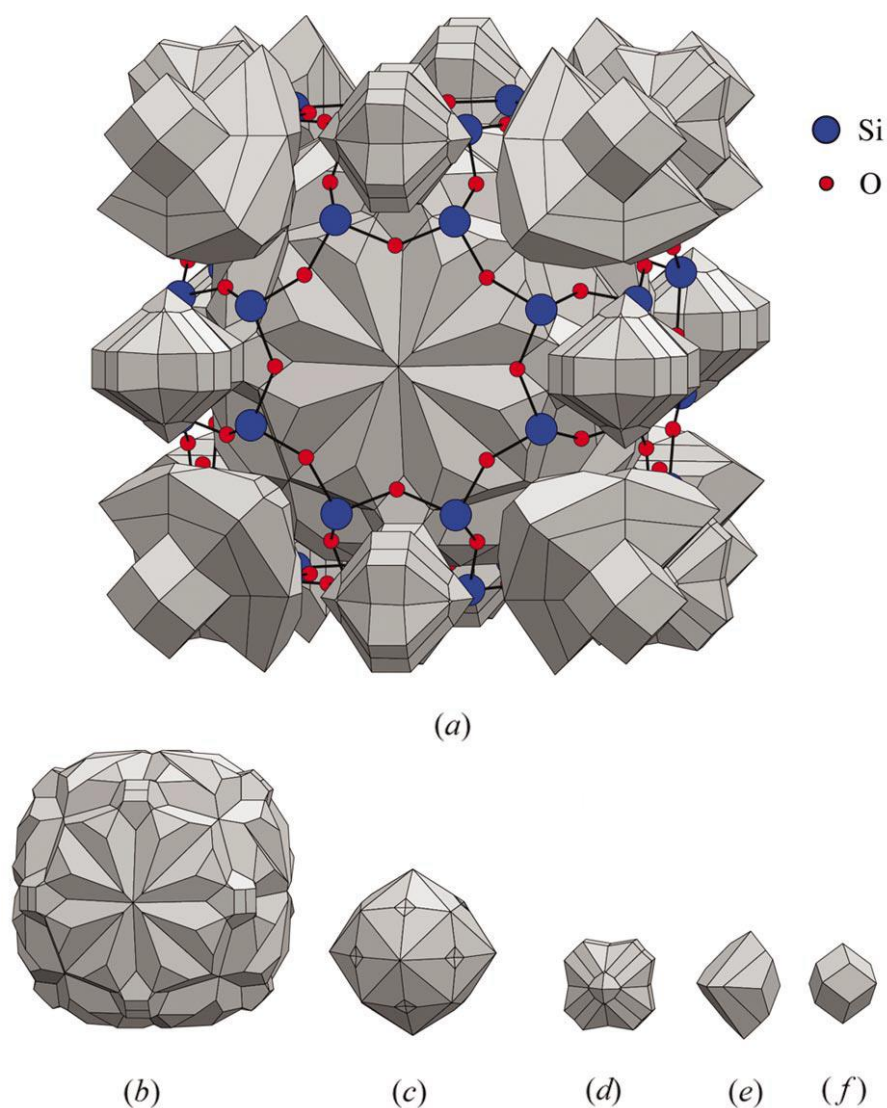


Рис. 35. (a) Представление в виде многогранников пустот в свободной от натрия подрешетке кристаллической структуры цеолита  $\text{NaAlSiO}_4$  [213] и молекулярные ПВ для несвязанных N-конгломератов пустот со следующими значениями N, объема ( $\text{\AA}^3$ ) и  $G_3$ : (b) 60, 799.0, 0.09466; (c) 33, 119.1, 0.0735; (d) 15, 23.8, 0.08151; (e) 3, 16.0, 0.08264; (f) 1, 9.5, 0.07875. Конгломерат (b) содержит конгломерат (c).

По аналогии с критерием наличия межатомного контакта [20] считается, что между пустотами существует связующий канал (Рис. 36), если телесный угол грани ПВ, разделяющей центры пустот, больше определенной величины, в качестве которой обычно принимается 1.5% от полного телесного угла  $4\pi$  стерадиан. Фактически, эта величина соответствует



размерам минимально допустимого сечения канала и отвечает утроенной типичной ошибке в расчете телесного угла, обусловленной погрешностями структурного эксперимента. Отметим, что такой подход к определению системы каналов более прост, чем использование в качестве таковой системы ребер атомных ПВ, так как в последнем случае затруднительно определить четкий критерий для отбраковки слишком узких каналов.

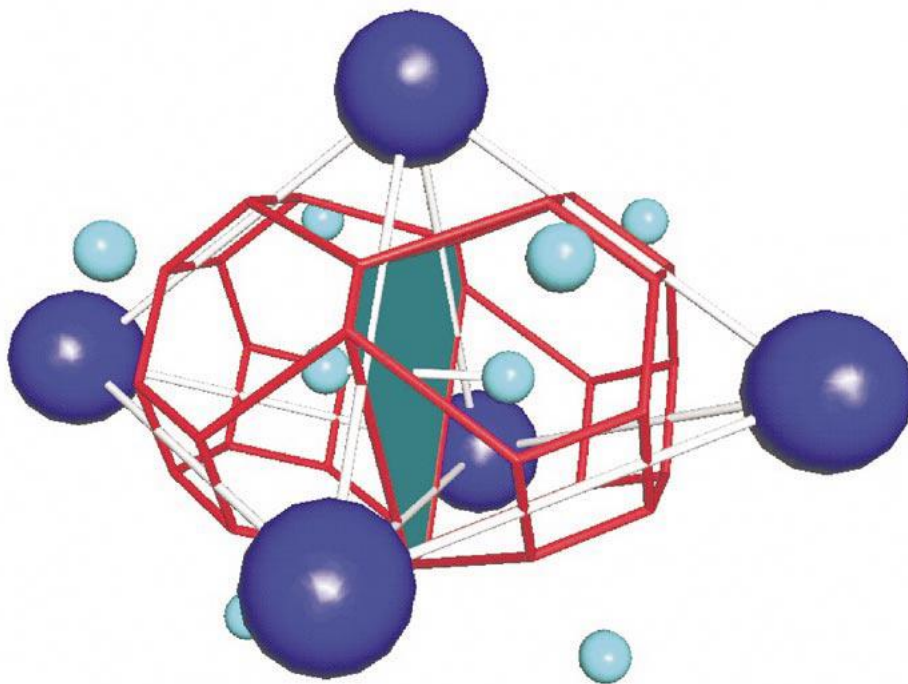


Рис. 36. Фрагмент кристаллической структуры для ГПУ решетки. Атомы и пустоты показаны темными и светлыми шарами, соответственно. ПВ построены для тетраэдрических пустот. Поверхность, соответствующая соединительному каналу, закрашена.

Если подпространство пустот не является односвязным (содержит не связанные друг с другом конгломераты пустот), перед проведением геометрико-топологического анализа необходимо провести его разбиение на односвязные области, что соответствует выделению молекулярных и полимерных группировок при обычном кристаллохимическом анализе подпространства атомов. Эта процедура осуществляется при помощи программы ADS комплекса ToposPro, которая одновременно определяет размерность (0D, 1D, 2D или 3D) односвязных областей и ориентацию

цепочечных и слоистых (1D или 2D) конгломератов относительно кристаллографических направлений (Рис. 37).

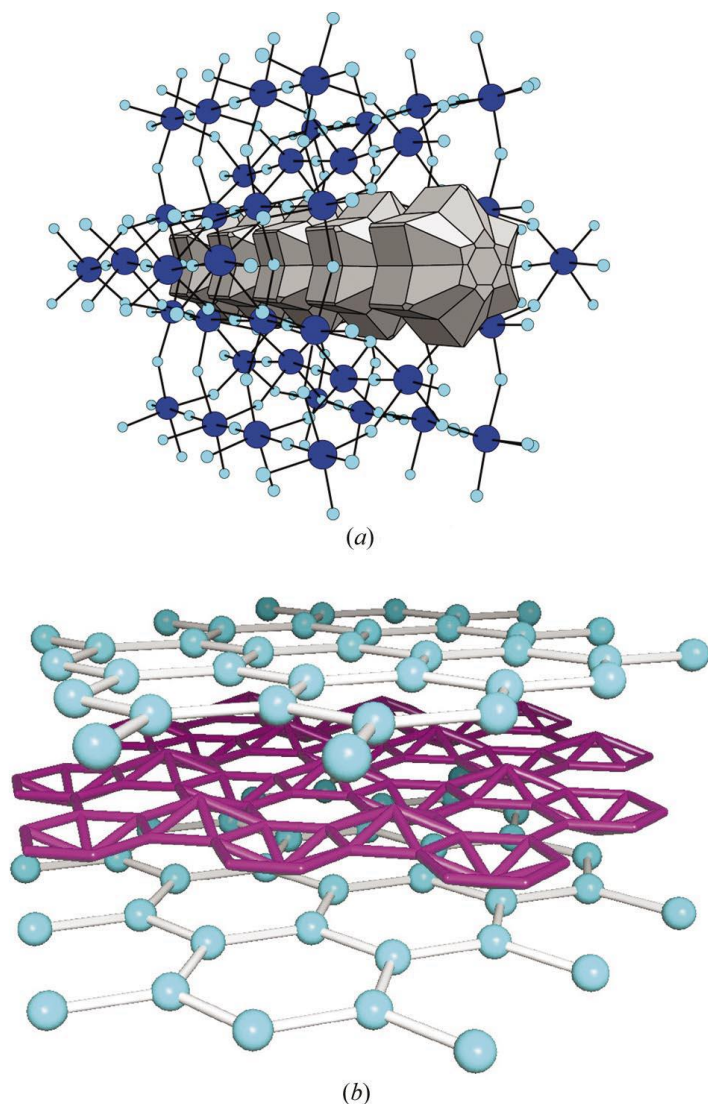


Рис. 37. (а) Полиэдрическое представление 1D цепочечного конгломерата пустот в кристаллической структуре  $\text{Rb}_{0.28}\text{WO}_3$  [214]; (b) графическое представление двумерного слоя конгломерата пустот в кристаллической структуре графита.

### 2.3.3. Определение геометрических характеристик пустот

*Объем пустоты и площадь сечения канала.* Понятие объема имеет смысл применять только для 0D конгломератов пустот. В случае, когда конгломерат состоит только из одной пустоты (*1-конгломерат*), его объем непосредственно равен объему ПВ этой пустоты, построенного с учетом всех соседних атомов, а радиус – значению ее  $R_{sd}$  (т.е. значению радиуса сферы, объем которой равен объему ПВ). Указанные характеристики соответствуют одноименным атомным характеристикам, корректность которых

продемонстрирована на многочисленных примерах [20]. Фактически, объем такого конгломерата равен объему, который имел бы атом, находящийся в данном атомном окружении. Именно поэтому рассчитанные таким образом объемы пустот можно сравнивать с размерами атомов, полученными оценкой объемов их ПВ в соответствующем окружении. При этом исчезает основная проблема использования ПВ в кристаллохимии – проблема выбора коэффициента деления гранями ПВ отрезков, соединяющих атомы [20], так как расчет ПВ как пустот, так и атомов проводится с использованием одного и того же коэффициента (для простоты его значение берется равным 0.5, что соответствует определению ПВ).

При анализе 1D, 2D и 3D конгломератов представляет интерес оценка площади сечения канала, соединяющего между собой полости. Если канал представлен  $N$  пустотами, расположенными на приблизительно равных расстояниях друг от друга (не группирующихся в дополнительные конгломераты), максимальная, минимальная и средняя площади его сечения ( $S_{\max}$ ,  $S_{\min}$ ,  $\langle S \rangle$ ) могут быть рассчитаны по общей формуле:

$$S = \pi R_{sd}^2, \quad (11)$$

где  $R_{sd}$  отвечает соответственно максимальному, минимальному и среднему  $R_{sd}$  для выборки из  $N$  ПВ пустот. Для более точной оценки формы и размеров канала можно рассчитать  $R_{sd}$  для дополнительных точек, расположенных на равных расстояниях друг от друга. ПВ  $N$ -конгломерата непосредственно отражает форму соответствующей полости (Рис. 35).

Для оценки степени отклонения формы  $N$ -конгломерата от геометрически правильного варианта удобно использовать значение нормализованного безразмерного второго момента инерции ПВ ( $G_3$ , см. Раздел 1.3.2). В этом случае суммирование осуществляется для всех симплексов, на которые разбиваются все ПВ, входящие в конгломерат, а за начало координат принимается геометрический центр тяжести конгломерата.

Для оценки геометрической формы полости эти значения следует сравнить с табличными (**Таблица 4**).

Основными характеристиками свободного пространства, которые могут быть рассчитаны в рамках разработанного формализма, являются:

- площадь ван-дер-Ваальсовой поверхности атомов элементарной ячейки ( $S_{vdW}$ ,  $\text{\AA}^2$ );
- объем пространства, ограниченного ван-дер-Ваальсовой поверхностью атомов в элементарной ячейке и элементарной ячейке ( $V_{vdW}$ ,  $\text{\AA}^3$ );
- объем системы каналов, доступной для сферического зонда с заданным радиусом  $R_p$ , ( $V_{channel}$ ,  $\text{\AA}^3$ );
- площадь поверхности системы каналов, доступной для сферического зонда с заданным радиусом ( $S_{channel}$ ,  $\text{\AA}^2$ );
- количество симметрично-независимых систем каналов ( $Z$ );
- периодичность системы каналов ( $P$ );
- вектор направления однопериодичной системы каналов и вектор нормали для двухпериодичной системы каналов ( $D$ );
- доля пустого пространства в структуре, которая рассчитывается по формуле  $\text{Total Porosity} = (V_{cell} - V_{vdw}) / V_{cell} * 100\%$ ;
- радиус самого большого сферического зонда, который может быть включен в структуру без пересечения с ван-дер-Ваальсовой поверхностью структуры ( $R_i$ ,  $\text{\AA}$ );
- радиус самого большого сферического зонда, который может свободно мигрировать в данной системе каналов структуры ( $R_f$ ,  $\text{\AA}$ );
- радиус самого большого сферического зонда, который может быть включен в заданную систему каналов структуры ( $R_{if}$ ,  $\text{\AA}$ ).

### 2.3.4. Топология системы каналов и расположения пустот

Методология топологического анализа кристаллических решеток легко переносится на случай исследования подпространства пустот. При этом аналог графа атомной сетки будет описывать топологию системы каналов, а аналог графа упаковки – топологию совокупности пустот, непосредственно несоединенных друг с другом (как, например, пустот различной формы в плотнейших упаковках). Для построения таких графов-аналогов необходимо после восстановления связности подпространства пустот осуществить только одну дополнительную операцию – стягивание конгломератов пустот в одну вершину графа. Напомним, что стягивание вершины графа означает ее удаление с одновременным замыканием образовавшихся свободных концов ребер [20]. Таким образом, оказывается возможным решение двух важных задач:

- (i) сравнение топологии систем пустот и каналов в двух разных соединениях или системы пустот и каналов в одной структуре с атомной сеткой в другой структуре;
- (ii) определение топологии пространственного расположения пустот.

## 2.4. Топологические библиотеки

### 2.4.1. Библиотеки топологических типов

Топологические базы данных ToposPro были созданы нами путем вычисления топологических дескрипторов для всех координационных соединений из Кембриджской структурной базы данных (CSD). Базы данных, аккумулирующие топологические свойства кристаллических структур, для удобства пользования, объединены в библиотеки (**Таблица 8**). Разработка структуры электронного хранилища библиотек, в виде реляционной базы данных, а также его наполнение данными с помощью специально созданных

для этого программных инструментов были выполнены в рамках данной работы.

Топологическая база данных ToposPro (TTD) — это самая большая библиотека топологических типов для гипотетических и реальных сеток, которые наблюдались или могут быть реализованы в кристаллических структурах. Библиотека TTD содержит топологические индексы для базисных сеток и используется при автоматическом отнесении кристаллической структуры к конкретному топологическому типу.

**Таблица 8.** Топологические электронные библиотеки

Название библиотеки	Сокращение	Содержимое библиотеки	Количество записей <sup>#</sup>
Topological Types Database	TTD	топологические типы сеток	808 445
Topological Types Observed	ТТО	представители топологических типов кристаллических структур	2 675 446
Topological Types Samples	TTS	образцы базовых сеток для часто встречающихся топологических типов	19 792
Topological Types of Polyhedra	TTP	топологические типы полиэдров	4 241
Topological Types of metal Atoms	ТТА	геометрико-топологические дескрипторы атомов металла	1 928 816
Topological Types of Molecules	ТТМ	геометрико-топологические дескрипторы органических молекул	309 353
Topological Types of Ligands	TTL	геометрико-топологические дескрипторы лигандов	269 232
Topological Types of Nanoclusters	TTN	топологические типы нанокластеров	1 234
Topological Types of Tiles	ТТТ	топологические типы тайлов	415

<sup>#</sup>Количество записей указано по состоянию на первый квартал 2024 года.

Библиотека наблюдаемых топологических типов (ТТО) является результатом сканирования базовых сеток кристаллических структур из

кристаллографических баз данных на предмет топологических типов. Если структура допускает разные представления, то все они помещаются в ТТО библиотеку. Комбинируя ресурсы TTD и ТТО, можно определить встречаемость определенного топологического типа в кристаллических структурах.

Библиотека отношений топологических типов (TTR) содержит список наблюдаемых отношений между топологическими типами. Эти отношения обнаруживаются в кристаллических структурах, которые допускают различные представления из-за многовариантного выбора структурных строительных единиц.

Библиотека топологических образцов (TTS) содержит уникальные примеры базовых сеток, которые получены из реальных структур. Представленные образцы встречаются в не менее, чем пяти различных кристаллических структурах.

Библиотека топологических типов полиэдров (TTP) содержит образцы ПВ атомов в реальных структурах и используется для их идентификации. Библиотека автоматически пополняется при работе программы Dirichlet.

Эти базы данных предоставляют широкие возможности для исследования различных структурных аспектов кристаллических координационных соединений. В настоящее время топологические библиотеки используются для определения топологии кристаллических структур по ее CIF-файлу и при поиске представителей конкретного топологического типа (сервис [topcryst.com](http://topcryst.com)).

#### **2.4.2. Библиотеки структурных единиц**

Данные о структурных единицах, выделенных из кристаллических структур по определенному алгоритму, сохранены в базе данных вместе с их дескрипторами. Эта информация может быть использована как для изучения

особенностей поведения конкретных фрагментов, так и в целях дизайна новых кристаллических структур с заданными свойствами. Структурными единицами могут быть атомы металлов (TTA), лиганды (TTL), внешнесферные частицы (TTM) в координационных соединениях, нанокластеры в интерметаллидах (TTN) и тайлы в цеолитах (TTT). С помощью программного комплекса ToposPro в режиме автоматического расчета для каждого из видов вышеперечисленных структурных единиц нами были созданы и периодически обновляются эти топологические библиотеки (Таблица 8). TTA, TTL и TTM библиотеки подготовлены в рамках данной работы. Поиск по топологическим библиотекам структурных единиц доступен через веб-сервисы, размещенные на сайте [topcryst.com](http://topcryst.com).

## 2.5. Модули для обработки и анализа данных

Для удобства проведения редактирования, отбора и анализа топологические библиотеки были сконвертированы в PostgreSQL хранилище, обрабатываемое с помощью языка Structured Query Language (SQL). К этому хранилищу подключены и размещены на сайте [topcryst.com](http://topcryst.com) веб-сервисы, позволяющие определять топологический тип кристаллической структуры по содержимому ее CIF-файла и выполнять через сеть Интернет поиск нужных топологических типов, их представителей и структурных единиц, из которых они сложены.

Дополнительно нами были разработаны вспомогательные компьютерные офф-лайн программы для работы с базами данных и поиска правил и закономерностей в данных. Так программное приложение *Samara Assistant of Crystal Design* (SACryDes, Рис. 38) позволяет формировать сложные запросы как к отдельным, так и нескольким топологическим библиотекам. Новые возможности использованы при подготовке выборок и их машинном анализе, речь о которых пойдет далее.



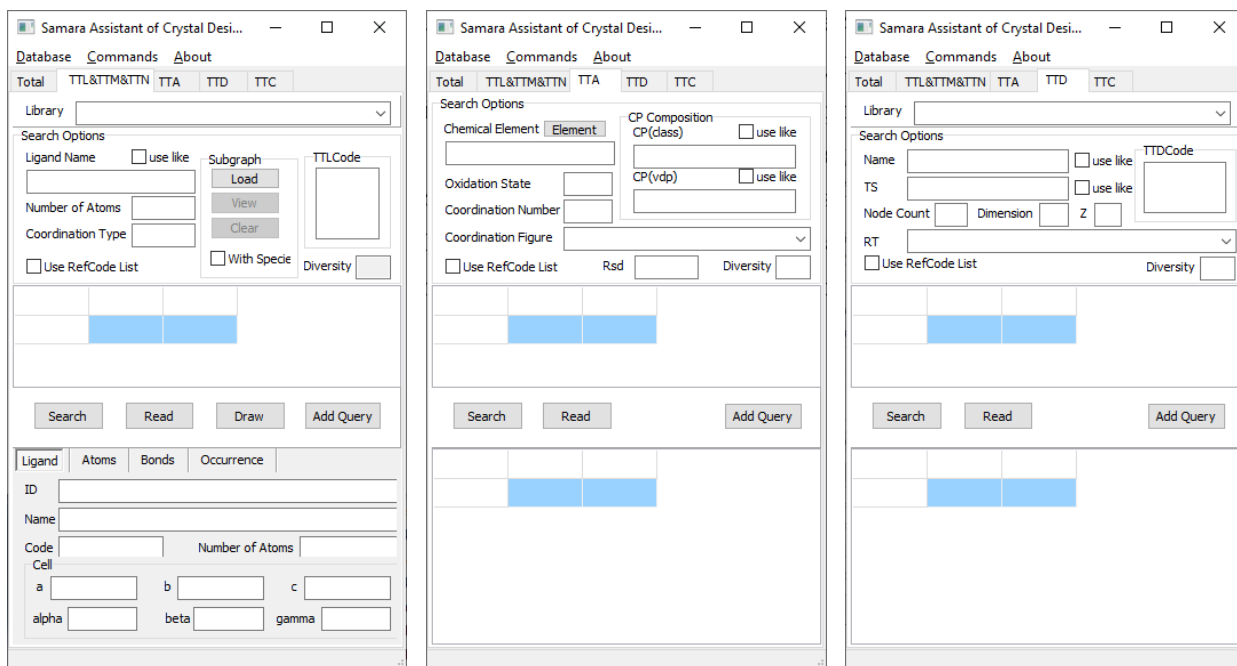
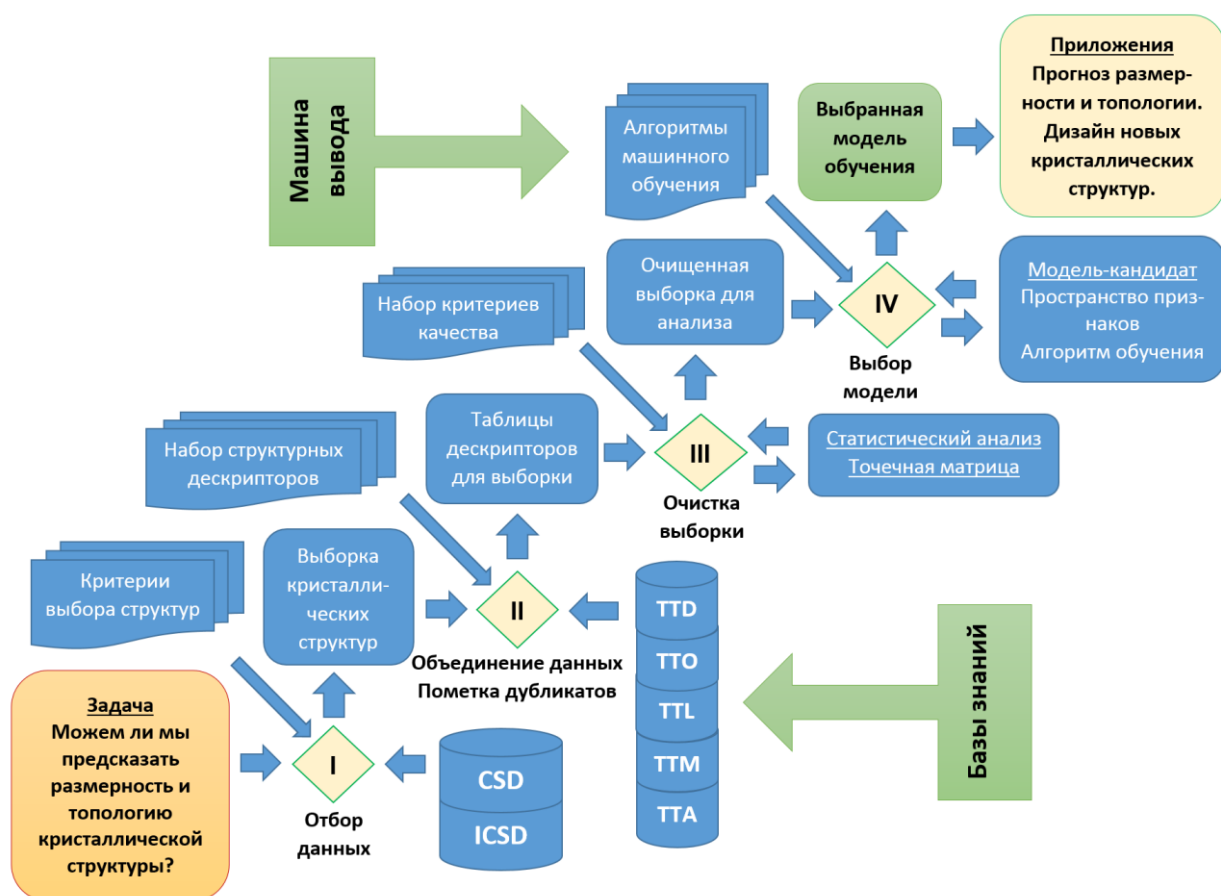


Рис. 38. Окна программы SaCryDes, предназначенной для поиска закономерностей в топологических библиотеках TTL&TTM&TTN (слева), TTA (в центре), TTD&TTO (справа).

## 2.6. Общая схема анализа данных

Общий рабочий процесс анализа данных, который мы используем в этом исследовании, выглядит следующим образом (**Схема 2**):

1. Указываем соответствующую задачу, *целевой дескриптор*, прогноз которого будет выполнять модель, и выбираем класс кристаллических структур для ее решения. Класс должен содержать достаточное количество данных, чтобы искать устойчивые корреляции, а задача должна быть строго и однозначно сформулирована.
2. Формулируем условия отбора данных, необходимых и достаточных для решения задачи. Их правильный выбор позволит сократить количество влияющих факторов и, тем самым, уменьшить количество обрабатываемых дескрипторов, что упрощает задачу и улучшает качество ее решения.



**Схема 2.** Общая схема подготовки и анализа кристаллографических данных и полученных компонентов системы искусственного интеллекта.

3. Выбираем структурные данные из первоисточников по заданным условиям.
4. Подбираем набор дескрипторов, подходящих для решения задачи. Реализуем алгоритмы их вычисления в программном коде, если это необходимо.
5. Готовим данные для анализа. На этом этапе мы исключаем исходные данные с ошибками и низким качеством, помечаем дублирующиеся записи и выбираем представителя для каждого набора идентичных записей.

6. Выполняем статистический анализ значений дескрипторов, чтобы проверить качество данных и исключить записи, отвечающие статистическим выбросам.
7. Используя численные критерии оценки качества результата, выбираем метод машинного обучения, который лучше всего подходит к данной задаче.
8. Выполняем машинное обучение для прогнозирования целевого дескриптора и выбираем значимые дескрипторы из обучающего набора.
9. Прогнозируем целевой дескриптор для кристаллических структур, которые не содержались в обучающей выборке и, сравнивая результаты прогноза с ожидаемыми значениями дескриптора, проверяем качество созданной нами модели.

В результате будет создана модель, которая позволяет пользователю прогнозировать целевые параметры. Для проверки работоспособности и универсальности описанной схемы, нами были сформулированы и решены несколько типичных физико-химических задач (**Таблица 9**), которые детально рассматриваются в следующих главах. Отбор структурных данных для решения этих задач осуществлялся с помощью комплекса программ ToposPro [215].

**Таблица 9.** Физико-химические задачи и выборки кристаллоструктурных данных, использованные для их решения

Физико-химическая задача	Объем выборки	Раздел
Определение формы координационной фигуры	9 642	3.1
Прогноз размерности и топологического типа МОКП	7 690	3.2
Анализ свободного пространства в структуре МОКП	33 790	3.3
Прогноз степени окисления атомов металла в МОКП	33 253	3.4
Анализ размеров атомов в структуре простых веществ	1 249	4.1 4.2
Анализ каналов проводимости в Na-содержащих структурах	402	4.3
Классификация кристаллических структур интерметаллидов	38 000	4.4
Прогноз плотности кристаллов насыщенных углеводородов	147	4.5

## **Глава 3. Корреляции «состав-структура-свойство» в координационных полимерах**

### **3.1. Координационные фигуры атомов меди и цинка**

Как отмечалось ранее, форма КФ атома в кристаллической решетке (узла структурной сетки) является важным фактором, влияющим на ее топологический тип [216]. В литературе мы не обнаружили однозначное описание формы координационной фигуры, подходящее для произвольных случаев, иногда с сильным ее искажением, также отсутствовали программные инструменты для сравнения двух координационных фигур, позволяющие численно охарактеризовать степень их сходства. В связи с этим мы решили частную задачу определения формы пятикоординационных фигур.

#### **3.1.1. Выборка**

В качестве тестовой выборки были рассмотрены комплексные соединения меди и цинка, проявляющие координационное число равное пяти и валентно связанные в структуре кристалла только с атомами кислорода и (или) азота. Такой выбор обусловлен большим разнообразием указанных соединений и сложностью однозначного определения формы пятикоординационной фигуры. Для исключения влияния дополнительных факторов и упрощения анализа данных, были рассмотрены только кристаллические структуры, содержащие один симметрично независимый атом металла. Соединения, удовлетворяющие указанным выше условиям, были отобраны из баз данных ICSD [2] и CSD [1, Версия 5.38]. Объем выборки составил 9642 записи, из которых 7241 и 2401 отвечали комплексами меди и цинка, соответственно. В рамках данной работы была подготовлена выборка, выполнен ее статистический анализ и разработана машинная модель для прогноза размерности и топологии.

### 3.1.2. Определение формы координационной фигуры

Результаты определения формы КФ с использованием угловых отпечатков представлены в **Таблице 10**. Буква рейтинга характеризует близость выбранного значения к максимальному совпадению КФ ( $\beta$ ): А -  $\beta \leq 5\%$ ; В -  $5\% < \beta \leq 10\%$ ; С -  $10\% < \beta \leq 15\%$ ; D -  $15\% < \beta \leq 20\%$ . Чем меньше  $\beta$ , тем точнее совпадение (см. Раздел 2.2.4), поэтому наиболее надежные результаты отвечают рейтингам с буквой А. Цифра рейтинга показывает отличие двух наименьших расстояний между отпечатками ( $\alpha$ ): 1 -  $\alpha \leq 5\%$ ; 2 -  $5\% < \alpha \leq 10\%$ ; 3 -  $10\% < \alpha \leq 15\%$ ; 4 -  $15\% < \alpha \leq 20\%$ . Чем ближе  $\alpha$  к разности для эталонов, тем более достоверны результаты определения формы КФ. Например, при отнесении координационной фигуры образца либо к квадратной пирамиде, либо к треугольной бипирамиде наилучшее значение цифры рейтинга равно 3, так как  $\alpha$  для тригональной бипирамиды (ТВРУ-5) и базоцентрированной (SPY-5a) или объемоцентрированной (SPY-5b) квадратной пирамиды равно 13.2 % или 13.8 %, соответственно (**Таблица 7**). Распределение точек по расстоянию между отпечатками образца и эталона показано на Рис. 39. С учетом сказанного выше, КФ считались нераспознанными для рейтингов 1D, 1E и 2E.

**Таблица 10.** Встречаемость (N) координационных фигур разных типов для Cu или Zn

Рейтинг	SPY-5b		SPY-5a		ТВРУ-5		TF-5		Надежность
	N	w, %	N	w, %	N	w, %	N	w, %	
1A	1	0.0	-	-	19	1.6	-	-	+
1B	2962	46.1	1260	68.0	656	54.8	-	-	+
1C	428	6.7	4	0.2	230	19.2	71	44.4	+
1D	11	0.2	-	-	22	1.8	34	21.3	-
1E	-	-	-	-	-	-	1	0.6	-
2A	1079	16.8	296	16.0	151	12.6	-	-	+
2B	1389	21.6	223	12.0	91	7.6	6	3.8	+

2C	24	0.4	-	-	2	0.2	19	11.9	+
2E	1	0.0	-	-	-	-	-	-	-
3A	532	8.3	71	3.8	27	2.3	-	-	+
3B	3	0.0	-	-	-	-	14	8.8	+
3C	-	-	-	-	-	-	1	0.6	+
4A	-	-	-	-	-	-	8	5.0	+
4B	-	-	-	-	-	-	6	3.8	+
<i>Итого</i>	<i>6430</i>		<i>1854</i>		<i>1198</i>		<i>160</i>		

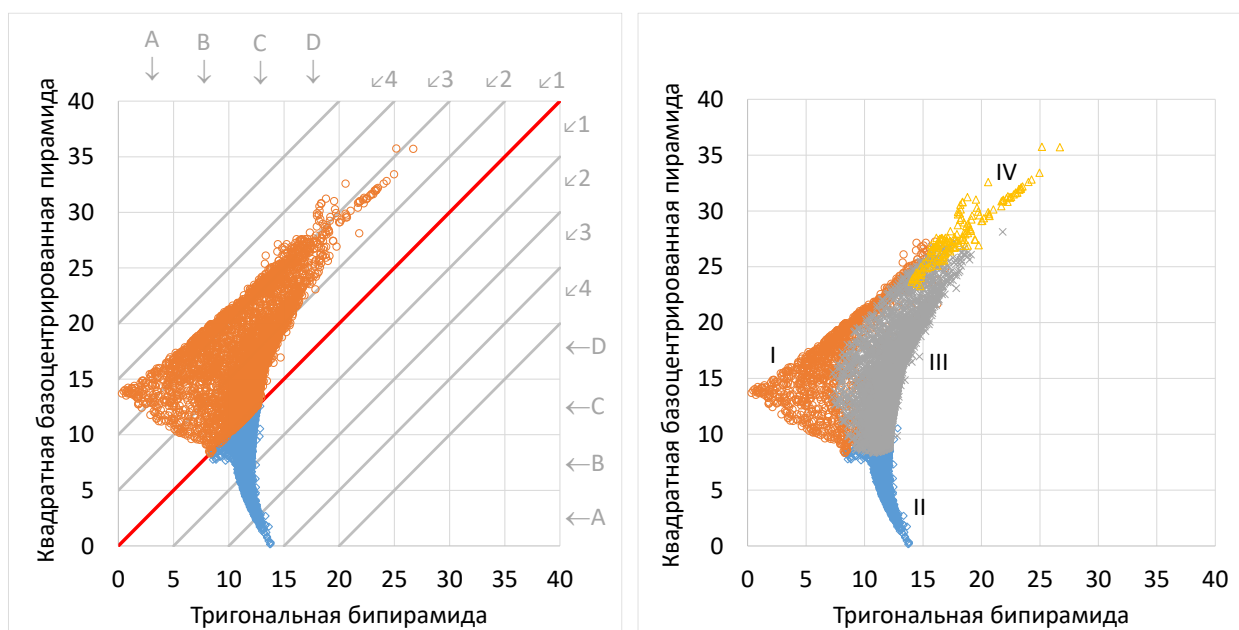


Рис. 39. Распределения расстояний между угловыми фингерпринтами рассматриваемой координационной фигуры и (слева) правильной треугольной бипирамиды или квадратной базоцентрированной пирамиды, а также (справа) тригональной бипирамиды, квадратной базоцентрированной пирамиды, квадратной объемоцентрированной пирамиды и одношапочного тетраэдра (соответственно I, II, III или IV области).

### 3.1.3. Правила и закономерности

Рассчитанный обсуждаемым методом дескриптор формы координационной фигуры использован для поиска новых правил и закономерностей строения координационных соединений. Так, в соединениях Cu и Zn координационные группы чаще изолированы; второе

место по встречаемости занято однопериодическими координационными полимерами, а двух- и трехпериодические координационные группы равновероятны. Это утверждение можно записать в виде правила: *встречаемость координационных групп уменьшается с увеличением их размерности*. Тип координационной фигуры не влияет на размерность комплексной группы, так как все точки диаграммы (Рис. 40, слева) лежат близко к линии равных вероятностей.

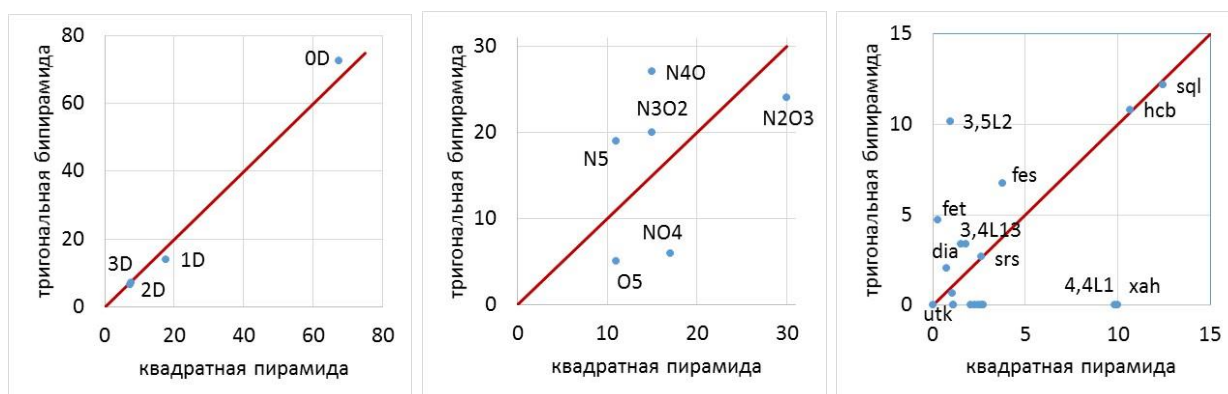


Рис. 40. Диаграмма рассеяния квадратная пирамида – тригональная бипирамида для встречаемости размерности комплексных групп (слева), химического состава (в центре) или общей топологии базовой сетки (справа) в % количества таких КФ.

Первая координационная сфера атомов Cu и Zn может содержать атомы азота и кислорода в любом соотношении. Наиболее предпочтительный состав координационной фигуры является  $N_2O_3$  для обоих металлов. При этом выполняется следующее правило (Рис. 40, в центре): *если атомы азота доминируют в координационном полиэдре ( $n > t$ ), то тригонально-бипирамидальная координация более вероятна, в противном случае преобладает квадратно-пирамидальная координация*. Причиной тому является различное расщепление кристаллическим полем 3d-орбиталей лигандов, координирующихся монодентатно атомами N и O (вариант Cu-I) или бидентатно атомами O карбоксильной группы (вариант Cu-II). Согласно теории кристаллического поля лигандов, для варианта Cu-I неспаренный электрон расположен на  $z^2$ -орбитали, тогда как для варианта Cu-II он



находится на  $x^2-y^2$ -орбитали (Рис. 41). Очевидно, что реализация «китайского фонарика» энергетически выгоднее в варианте Cu-II.

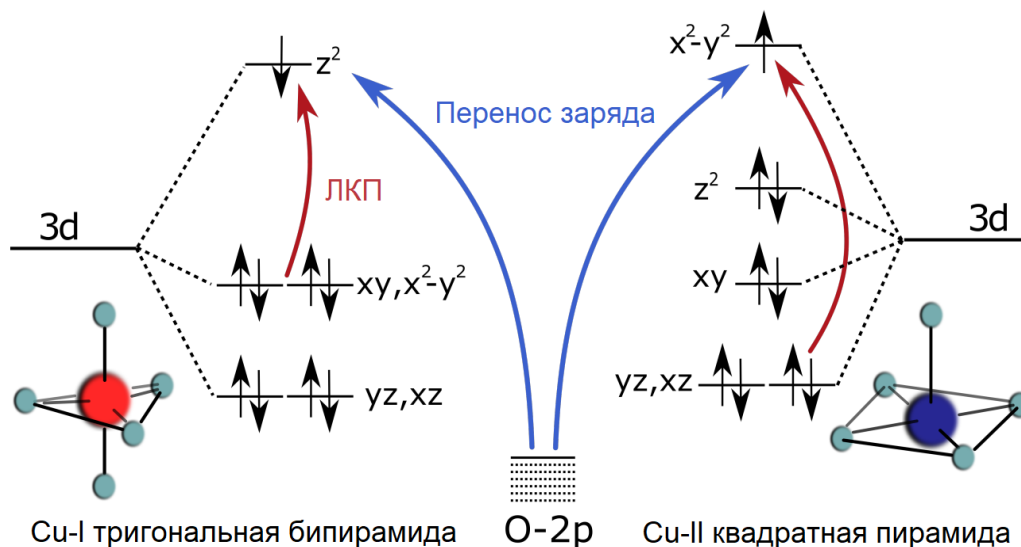


Рис. 41. Трехмерное расщепление кристаллического поля для ионов Cu-I (тригональная бипирамида) и Cu-II (квадратная пирамида) в структуре  $\text{Cu}_2\text{OSeO}_3$ . Дипольные возбуждения локального кристаллического поля (ЛКП) показаны красными стрелками, а возбуждения с переносом заряда - синими стрелками. Использован рисунок из статьи [217].

Форма координационной фигуры влияет на общую топологию сетки, но не во всех случаях (Рис. 40, справа). Такие общие топологии сеток, как 3,5L2, **fet** (Рис. 42 внизу), **fes**, являются предпочтительными для тригонально-бипирамидного типа, тогда как топологические типы **xah** (Рис. 42 вверху), 4,4L1 допустимы только при квадратно-пирамидальной координации (Рис. 43). Топологические типы **sql**, **hcb**, **srs**, **utk**, расположенные близко к линии равных вероятностей, нечувствительны к типу координационной фигуры. Приведенные примеры сеток показывают, что четырехугольная пирамида реализуется в жестко заданных фрагментах, таких как «китайский фонарик» в топологических типах **xah**, 4,4L1. При отсутствии подобных фрагментов либо преобладают треугольно-бипирамидальная форма КФ, либо обе формы являются равновероятными.

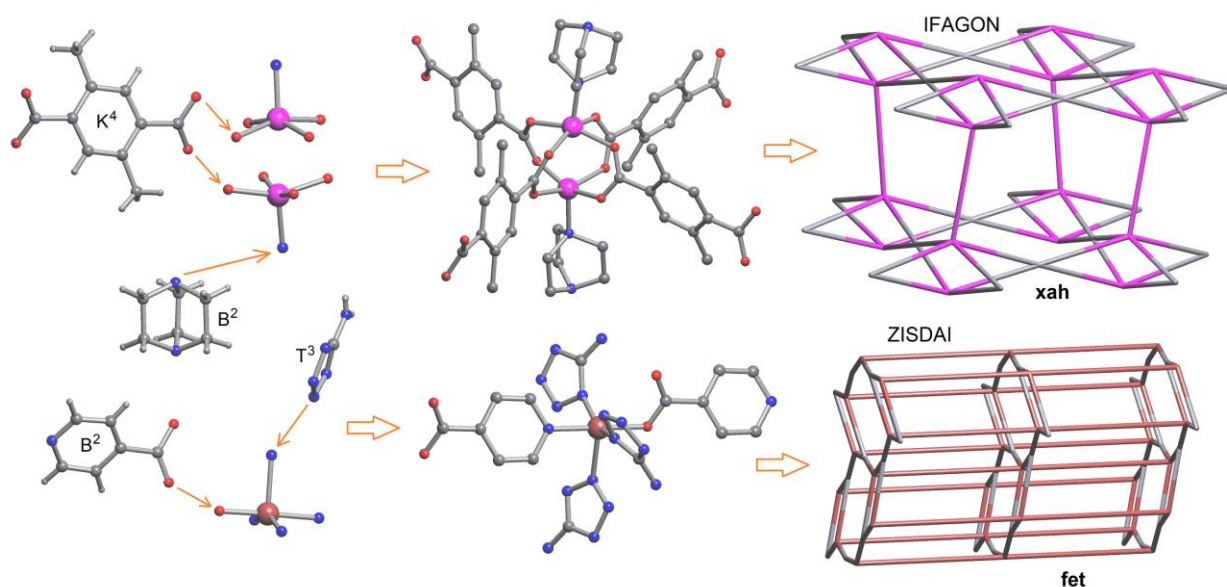


Рис. 42. Влияние природы донорных атомов лигандов на выбор координационной фигуры и топологического типа базовых сеток **fet** (внизу) или **xah** (вверху), реализованных в кристаллических структурах ( $\mu_3$ -5-аминотетразолил)-( $\mu_2$ -изоникотинато)-меди(II) [218] и бис( $\mu_4$ -2,5-диметилтерефталато)-( $\mu_2$ -1,4-дiazобикакло[2.2.2]октан)-дицинка DMF сольвата [219], соответственно.

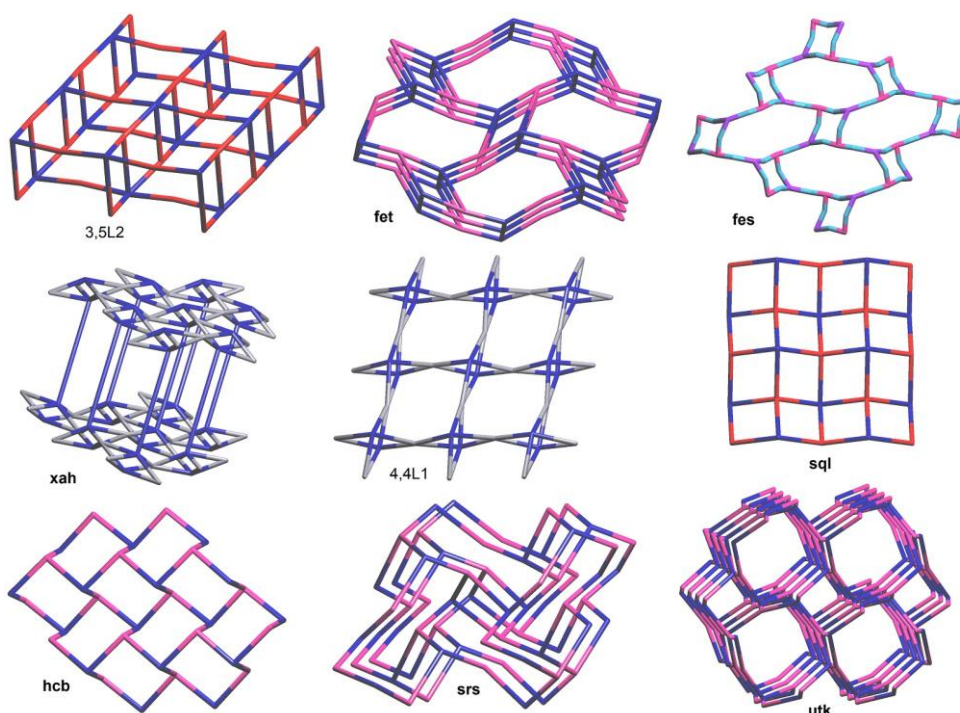


Рис. 43. Некоторые топологии базовых сеток в пятикоординационных соединениях.

Найденные правила, представленные в виде дерева решений (Рис. 44), позволяют прогнозировать вероятное строение кристаллов Cu и Zn на основании знаний о химической природе лигандов. Например, если мы хотим получить МОКП с топологией **xah**, то нужны лиганды с преобладанием координирования атомами кислорода, которые склонны к образованию структурного блока «китайский фонарик», допускающие локальную координацию  $A_2K^4_2B^2$ . Напротив, для получения МОКП с топологией **fet** следует выбирать лиганды с преобладанием координации по атому азота, для которых не характерно образование структурного блока «китайский фонарик», и которые допускают локальную координацию  $AT^3B^2$  (Рис. 42). Вероятность получить топологию **fet** при соблюдении указанных выше условий составит 23% (Рис. 44).

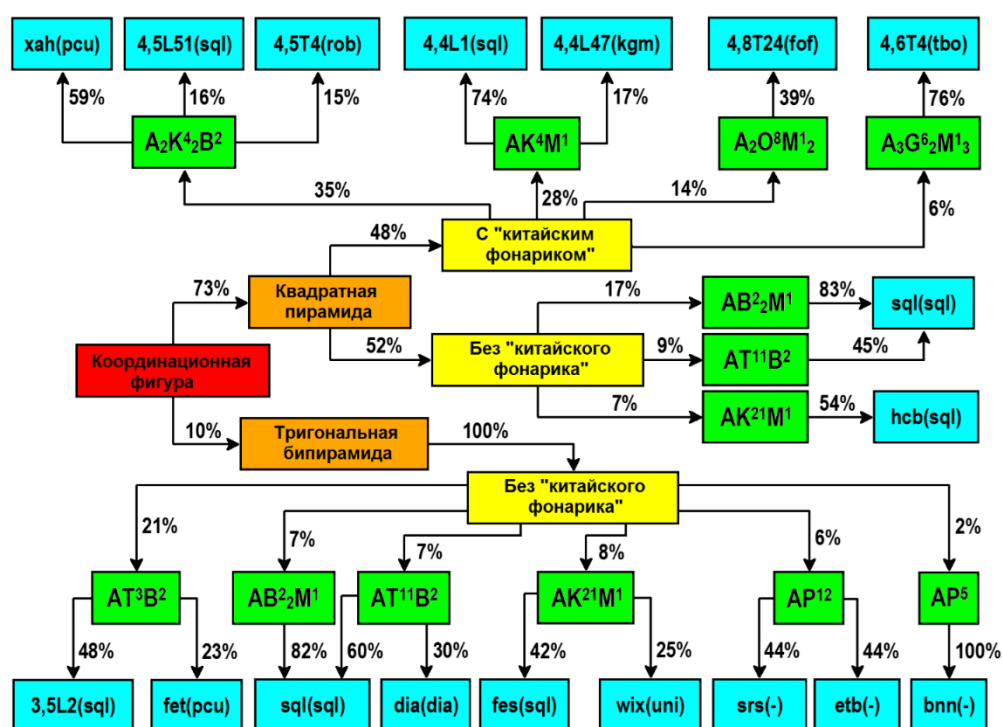


Рис. 44. Множественные корреляции «Координационная фигура (оранжевый) - вторичная строительная единица (желтый) - локальная координация (зеленый) - общая топология (голубой)» для 2D и 3D МОКП Cu или Zn. Топологический тип указан в стандартном и кластерном (в скобках) представлениях.

Рассмотрим один из вариантов моделирования новых кристаллических материалов с применением найденных корреляций и разработанного программного инструментария. Используя данные коллекций TTA и TTL, мы можем получить информацию о том, с какими лигандами предпочитает образовывать соединения тот или иной металл, или, наоборот, с какими металлами наиболее часто реализует комплексы определенный лиганд. Имея эти сведения, можно предположить комбинации металл – лиганд, которые могут быть реализованы, но до сих пор неизвестны, и синтезировать такие соединения. Более того, анализируя предпочтения связывания лиганда или металла, а также известные их комбинации, можно построить новые структуры, например, с заданной топологией или с нужными физическими свойствами.

Связывание атома металла в структуре характеризуют следующие дескрипторы: координационное число, координационная фигура и химический состав первой координационной сферы. Связывание лиганда описывают координационный символ и координационная фигура. Рассчитав такие дескрипторы, мы получим для каждой структурной единицы спектр значений с вероятностями их реализации.

В качестве примера рассмотрим известные кристаллические структуры, содержащие лиганд 3-цианоацетилацетонат  $C_6H_6NO_2^-$  (*acacCN*). В Кембриджской базе данных найдено 98 таких структур (Таблица 11). Анализ данных TTL коллекции показал, что лиганд встречается в структуре либо в одном типе координации  $B^{01}$  (43.0%, LUJCUQ),  $T^{11}$  (39.8%, PEYDAA),  $M^1$  (1.1%, GEKKIQ) и  $T^4$  (1.1%, KIFJOZ) или в двух различных типах  $T^{11} \& B^{01}$  (9.7%, ZAXTUP),  $T^{21} \& T^{11}$  и  $B^{01} \& M^1$  (2.2% каждый, ZAXWOM и EZOYEY, соответственно). В одном случае (структура XOMFIP) координированной (монодентано) является нейтральная молекула 3-цианоацетилацетона. Многие структуры содержат не только лиганд *acacCN*, но и другие лиганды, проявляющие разные типы координации.

**Таблица 11.** Типы координации *асасСN* лиганда, а также дополнительных лигандов в 98 кристаллических структурах.

Лиганд C <sub>6</sub> H <sub>6</sub> NO <sub>2</sub>	N	w, %	Другие лиганды	N	w, %	Лиганд C <sub>6</sub> H <sub>6</sub> NO <sub>2</sub>	N	w, %	Другие лиганды	N	w, %			
<b>0D</b>						<b>2D</b>								
B <sup>01</sup>	31	91.2	M <sup>1</sup>	10	32.3	T <sup>11</sup>	14	66.7	M <sup>1</sup>	7	50			
			B <sup>01</sup>	8	25.8				B <sup>01</sup>	3	21.4			
			-	4	12.9				-	1	7.1			
			T <sup>11</sup>	3	9.7				T <sup>11</sup> &M <sup>1</sup>	1	7.1			
			B <sup>01</sup> &M <sup>1</sup>	2	6.5				M <sup>1</sup>	1	7.1			
			M <sup>3</sup> &M <sup>2</sup>	1	3.2				T <sup>11</sup>	1	7.1			
			M <sup>2</sup>	1	3.2				T <sup>11</sup> &B <sup>01</sup>	3	14.3	-	3	100
			T <sup>11</sup> &B <sup>01</sup>	1	3.2	B <sup>01</sup>	4	19.0	K <sup>4</sup>	2	50			
			T <sup>001</sup> B <sup>2</sup>	1	3.2				T <sup>3</sup>	1	25			
B <sup>01</sup> &M <sup>1</sup>			T <sup>001</sup>	1	50	<b>3D</b>								
			B <sup>01</sup>	1	50									
M <sup>1</sup>	1	2.9	M <sup>1</sup>	1	100	T <sup>11</sup>	10	76.9	-	6	60			
<b>ID</b>									M <sup>2</sup>	2	20			
T <sup>11</sup>	13	52.0	B <sup>01</sup>	5	38.5				T <sup>02</sup> &B <sup>01</sup> &M <sup>2</sup>	1	10			
			M <sup>1</sup>	2	15.4				T <sup>21</sup>	1	10			
			T <sup>11</sup>	2	15.4	T <sup>21</sup> &T <sup>11</sup>	2	15.4	-	2	100			
			T <sup>001</sup>	1	7.7	T <sup>4</sup>	1	7.7	-	1	100			
			B <sup>01</sup> &M <sup>1</sup>	1	7.7									
			M <sup>1</sup>	1	7.7									
			B <sup>11</sup>	1	7.7									
B <sup>01</sup>			T <sup>11</sup>	3	60									
			K <sup>101</sup>	2	40									
T <sup>11</sup> &B <sup>01</sup>	6	24.0	M <sup>1</sup>	2	33.3									
			-	2	33.3									
			B <sup>01</sup> &M <sup>2</sup>	1	16.7									
			T <sup>11</sup> &M <sup>1</sup>	1	16.7									
M <sup>1</sup> *	1	4.0	B <sup>2</sup>	1	100	* 3-цианоацетилацетон C <sub>6</sub> H <sub>7</sub> NO <sub>2</sub>								

\* 3-цианоацетилацетон C<sub>6</sub>H<sub>7</sub>NO<sub>2</sub>

Анализ данных показал, что содержащие серебро соединения встречаются наиболее часто (45 структур). Причина такой особенности состоит в том, что металлы координируют *асасСN* лиганды предпочтительно через атомы кислорода, а не азота, из-за хелатного эффекта и более прочной связи с кислородом. Чтобы получить полимерные структуры, синтетики вводят в состав соединений *асасСN* атомы серебра, для которых устойчивы азотсодержащие комплексы и нет склонности к образованию хелатов. В результате, все соединения, содержащие серебро и лиганд *асасСN*, имеют

полимерное строение, где атомы серебра координируют атомы азота, а другой металл связан с атомами кислорода (Таблица 12).

**Таблица 12.** Встречаемость сочетаний «атом металла – состав координационного полиэдра»

Me_СР лиганд <i>асасCN</i>	N	w, %	Me_СР другой лиганд	N	w, %	Me_СР лиганд <i>асасCN</i>	N	w, %	Me_СР другой лиганд	N	w, %
<b>0D</b>						<b>2D</b>					
Cu_N2O2	6	17.6	Ag_N2	3	50	Cr_O6	8	38.1	Ag_N4	3	37.5
			-	2	33.3				Ag_N3O	2	25
			Ag_N2O2	1	16.7				Ag_N3OAg	1	12.5
Fe_O6	2	5.9	-	2	100				Ag_C2N3	1	12.5
Co_N4O2	2	5.9	-	2	100				Ag_N3FAg&N3OAg	1	12.5
Sc_O7	2	5.9	-	2	100	Co_N2O4	3	14.3	-	3	100
Cr_O6	2	5.9	-	2	100	Al_O6	4	19.0	Ag_N3O	2	50
Cr_N2O4	2	5.9	-	2	100				Ag_N3OAg	1	25
Cr_N4O2	2	5.9	-	2	100				Ag_C2N3	1	25
Cu_O4&O5	1	2.9	-	1	100	Ce_O9	1	4.8	Ag_N4	1	100
Mn_CN3O2	1	2.9	-	1	100	Eu_O9	1	4.8	Ag_N4	1	100
Mn_N4O2	1	2.9	-	1	100	Cr_N2O4	1	4.8	Ag_N4	1	100
Co_N6	1	2.9	-	1	100	Co_N4O2	2	9.5	Ag_N2O2Ag	1	50
Cu_N3O2	1	2.9	-	1	100				Ag_N3	1	50
Mn_O6	1	2.9	-	1	100	Cd_N2O4	1	4.8	-	1	100
Ni_O6	1	2.9	-	1	100						
Cu_O5	1	2.9	-	1	100						
Zn_O5	1	2.9	-	1	100						
Cu_N2O2Br	1	2.9	-	1	100						
Cu_N2O3	1	2.9	-	1	100						
Ce_O9	1	2.9	-	1	100						
Eu_O8	1	2.9	-	1	100						
Yb_O8	1	2.9	-	1	100						
Pd_O4	1	2.9	-	1	100						
Zn_O6	1	2.9	-	1	100						
<b>1D</b>						<b>3D</b>					
Cu_N3O2	5	20.0		5	100	Cd_N2O4	2	15.4	-	2	100
Pd_O4	7	28.0	Ag_N2O2	3	42.9	Eu_O8	2	15.4	Ag_N4	1	50
			Ag_N2O	2	28.6				Ag_N5	1	50
			Ag_N2O4	1	14.3	Yb_O8	2	15.4	Ag_N4	1	50
			Ag_N2O3	1	14.3				Ag_N5	1	50
Cu_N2O2	2	8.0	Ag_N4	2	100	Pd_O4	1	7.7	Ag_NO4	1	100
Mn_N4O2	1	4.0		1	100	Al_O6	2	15.4	Ag_N3	1	50
Fe_O6	2	8.0	Ag_N2O	1	50				Ag_N4	1	50
			Ag_N3O2	1	50	Fe_O6	2	15.4	Ag_N4	1	50
Cu_N2O2 &N2O3	1	4.0	Ag_N3	1	100				Ag_N2O2Ag &N2O4Ag	1	50
Cu_NO4	1	4.0		1	100	Ag_NO3	1	7.7	-	1	100
Co_N2O4	2	8.0	Ni_N6	1	50	Co_N2O4	1	7.7	-	1	100
			Co_N6	1	50						
Sc_NO6	1	4.0		1	100						
Ce_NO8	1	4.0		1	100						
Cu_O5	1	4.0	Ag_N2O2	1	100						
Ag_N2O2	1	4.0		1	100						

В анализируемой выборке лишь немногие отличные от серебра металлы (Sc, Mn, Co, Cu, Cd и Ce) образуют связи с азотом лиганда *asacCN*. Для 70 металлов строение соединений с лигандом *asacCN* неизвестно, что открывает широкие возможности для прогнозирования кристаллических структур с новым химическим составом.

Если изучить известные кристаллические структуры, которые содержат схожий по топологии связей лиганд ацетилацетонат, то можно обнаружить, что с ним образуют соединения практически все металлы периодической системы (за исключением элементов с высокой радиоактивностью). Поэтому весьма вероятно, что могут быть синтезированы устойчивые комплексы металлов Ti, V, Mo, Ru, Rh, Ir, Pt и др. с лигандом *asacCN*. Если нам нужны полимерные структуры, то помимо предложенных металлов, можно вводить в них атомы серебра, как это сделано в некоторых комплексах с *asacCN*. Но можно использовать и другие комбинации, выбрав в качестве атома, связывающего цианидный азот, например, одновалентные медь или золото, стереохимия которых подобна серебру.

### **3.2. Прогноз размерности и топологии кристаллических структур машинными методами**

Решив задачи оценки и сравнения формы КФ, мы использовали этот важный дескриптор для создания машинной модели, прогнозирующей размерность и топологический тип базовой сетки структур КС и МОКП.

#### **3.2.1. Выборка**

Для создания данной прогностической модели мы сформировали выборку из 7690 структурных определений, содержащих 6171 уникальных кристаллических структур координационных соединений и 1279 различных лигандов. Все лиганды этой выборки состояли только из атомов углерода,

водорода и/или кислорода, характеризуясь формулой  $C_nH_mO_k$ , где  $n \geq 0$ ,  $m \geq 0$ ,  $k > 0$ , включая  $H_2O$ ,  $H_3O^+$ ,  $OH^-$ ,  $O^{2-}$  или  $O_2^{2-}$ . Каждый лиганд был связан с атомом металла только через атом кислорода. В анализируемой выборке отсутствовали ограничения на химический сорт атома комплексообразователя, сложность комплексной группировки и химический состав некоординированных молекулярных фрагментов.

Статистический анализ распределений значений дескрипторов этой выборки позволил нам выявить наиболее типичные атрибуты, которые затем можно использовать при поиске корреляций, а также выбросы, которые могут быть исключены после тщательного анализа. Ниже мы обсудим некоторые особенности распределений, играющих важную роль и для других групп КС.

*Химический состав лигандов.* Для анализа разнообразия химического состава лигандов  $C_nH_mO_k$  была предложена *ntk*-диаграмма, аналогичная диаграмме, используемой ван Кревеленом [220] для классификации различных типов природных смесей органических соединений, и треугольнику Гиббса. На Рис. 45 показано содержание легких лигандов (оксалат-, ацетат-, формиат-анионов, тетрагидрофурана и др.). Изомеры попадают в одно и то же положение диаграммы; мы нашли 676 изомеров для 224 составов. Лидером оказался состав  $C_8H_7O_3$ , которому отвечает 14 изомеров (Рис. П1). Как видно из Рис. 45, некоторые группы точек располагаются на прямых линиях, которым, согласно [221], могут быть сопоставлены определенные гомологические ряды соединений.

Классификация по структурным фингерпринтам (см. Раздел 1.3.1) выявляет 45 различных классов; наибольшее количество представителей (236 лигандов или 18.5% от их общего количества) попадают в класс 10100001, что соответствует карбоновым кислотам с одной или несколькими карбоксильными группами (Рис. 46).



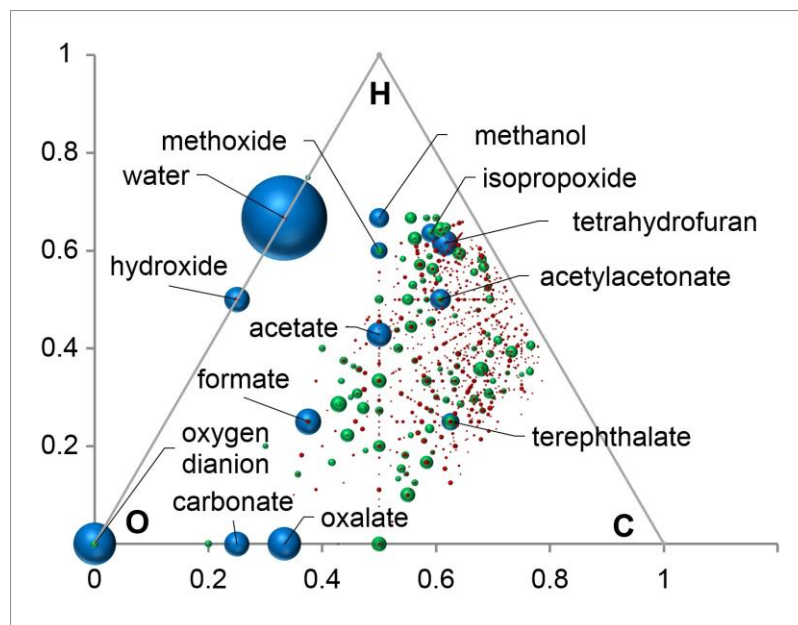


Рис. 45. *nmk*-диаграмма распределения лигандов  $C_nH_mO_k$  по составу. Размер каждой сферы пропорционален встречаемости состава. Точки составов с высокой, средней и низкой встречаемостью окрашены в синий, зеленый и красный цвета, соответственно.

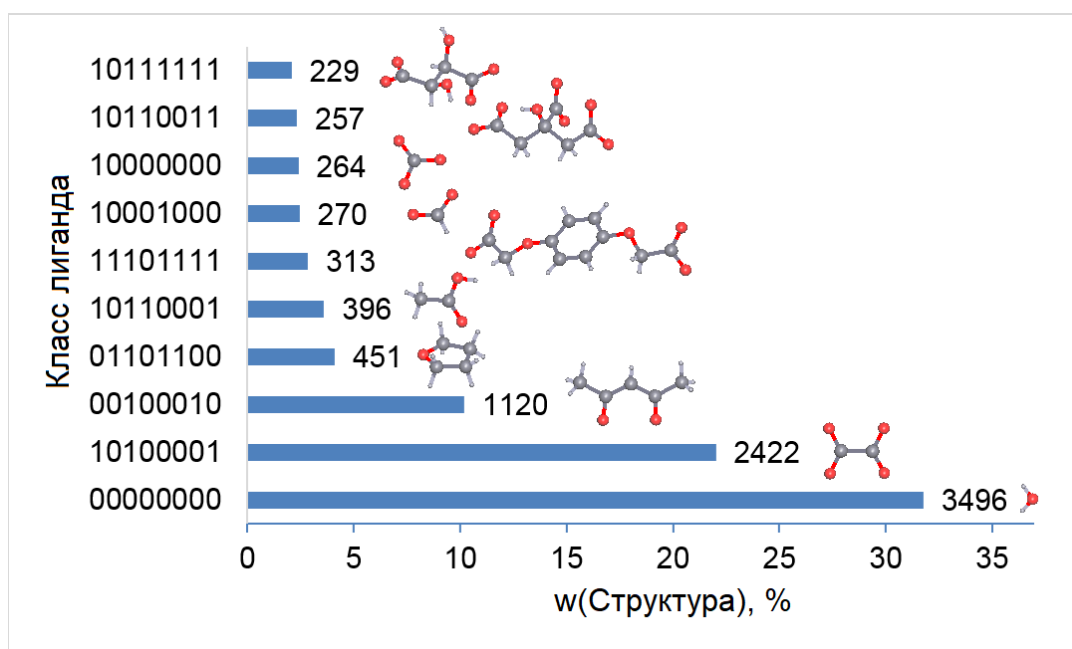


Рис. 46. Первая десятка распространенных топологических фингерпринтов (классов лигандов).

В рассмотренных структурах заряд лигандов ( $q$ ) изменялся в диапазоне от +1 до -8 и имел наиболее вероятное значение -1 (Таблица 13). Выборка

содержит лиганды, образующиеся при диссоциации одной и той же многоосновной органической кислоты и характеризующиеся разной основностью и/или различными положениями атомов водорода. Например, лимонной кислоте соответствует семь топологически отличных лигандов с зарядами в диапазоне от -1 до -4.

**Таблица 13.** Распределение лигандов  $C_nH_mO_k$  по их заряду  $q$

$q$	N	w, %	Примеры
-1	552	47.6	ацетат ( $C_2H_3O_2^-$ , ЕКАВОI); метоксо ( $CH_3O^-$ , АСUZOO);
-2	332	28.6	оксалат ( $C_2O_4^{2-}$ , МIYGIL); терефталат ( $C_8H_4O_4^{2-}$ , IFABIA01);
0	108	9.3	тетрагидрофуран ( $C_4H_8O$ , ZIGKUX); метанол ( $CH_4O$ , HAYGIZ)
-3	86	7.4	тримезат ( $C_9H_3O_6^{3-}$ , LAVTEJ01); цитрат ( $C_6H_5O_7^{3-}$ , МАВВЕZ)
-4	66	5.7	пиромеллитат ( $C_{10}H_2O_8^{4-}$ , WEMGAY); татртрат ( $C_4H_2O_6^{4-}$ , JEDJAE01)
-6	8	0.7	меллитат ( $C_{12}O_{12}^{6-}$ , QAGXUT); циклогексангексакарбоксилат ( $C_{12}H_6O_{12}^{6-}$ , ECUCAN02)
-5	4	0.3	1,2,3,4,5-бензолпентакарбоксилат ( $C_{11}HO_{10}^{5-}$ , WIYZIO)
-8	3	0.3	3,3',5,5'-тетракис(3,5-дикарбоксилатофенил)-4,4'-диметокси-1,1'-бифенил ( $C_{46}H_{22}O_{18}^{8-}$ , HUIGEA)
+1	1	0.1	катион гидроксония ( $H_3O^+$ , WEMBUN)

*Количество атомов* в лиганде  $N_{at}$  изменялось от 1 ( $O^{2-}$ ) до 230 ( $C_{106}H_{118}O_6$ , 930, RAVXOD, [222]) при наиболее вероятном значении равном 18 (Рис. 47), реализующемся в анионах тримезат ( $C_9H_3O_6^{3-}$ ), 2,6-ксиленоксид ( $C_8H_9O^-$ ) или цинамат  $C_9H_7O_2^-$ .

*Координационное число* лигандов ( $CN_{lig}$ ) в рассмотренных структурах изменялось от 1 до 16. Его минимальное значение является наиболее вероятным (Таблица П1), а встречаемость падает при росте  $CN_{lig}$ . Начиная с  $CN_{lig}=6$  при четных значениях наблюдаются микромаксимумы.

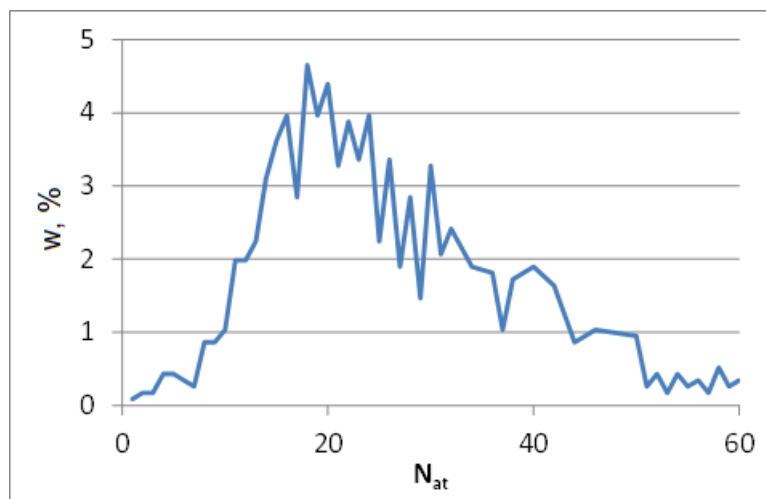


Рис. 47. Распределение лигандов по количеству атомов в них ( $N_{at}$ ). Доля лигандов с  $N_{at} > 60$  равна 8.1%.

Типы координации лигандов [44] весьма разнообразны и характеризуются 386 различными значениями. Распределение лигандов по этому дескриптору представлено в **Таблице П2** и на Рис. 48. Монодентатная координация  $M^1$  оказалась наиболее часто встречающейся.

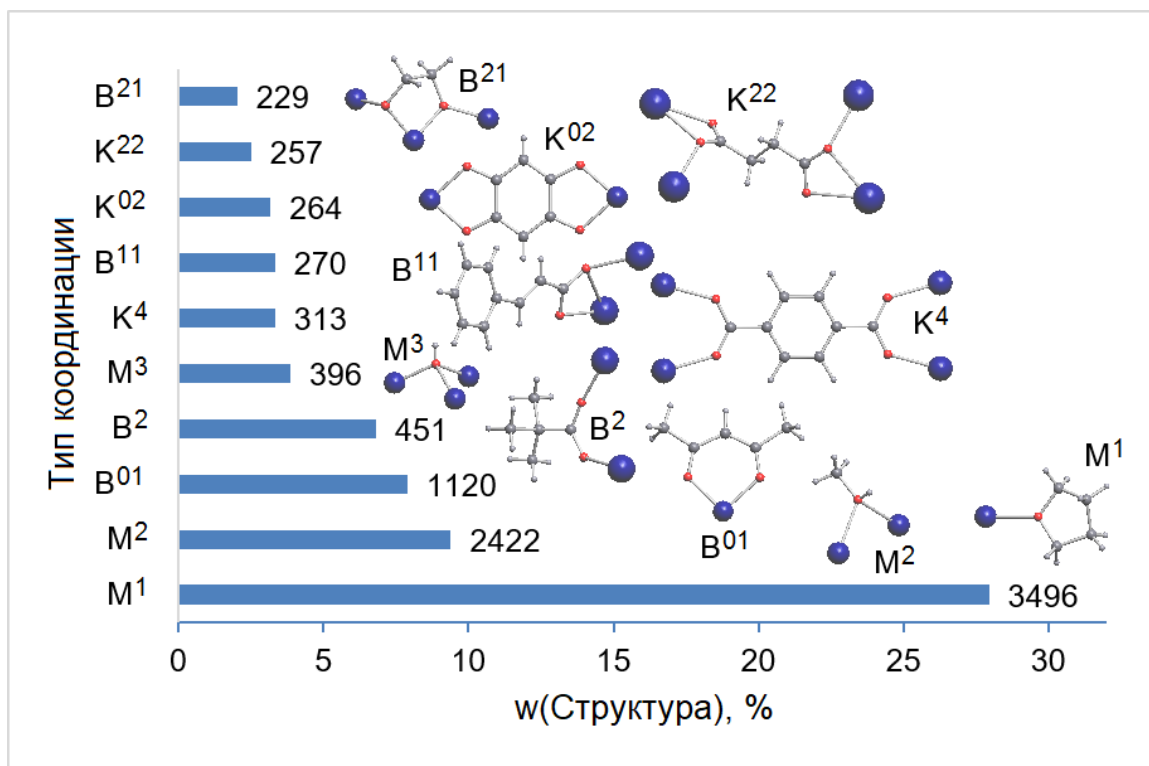


Рис. 48. Первая десятка типов координации лигандов, количества и доли ( $w$ , %) кристаллических структур, в которых они встречаются.

Максимальные наборы различных вариантов координаций отвечали двухзарядным лигандам оксалату  $C_2O_4^{2-}$ , карбонату  $CO_3^{2-}$  и тартрату  $C_4H_4O_6^{2-}$ , имеющих 43, 38, 32 уникальных координаций, соответственно. Такое разнообразие характерно для малых, богатых кислородом молекул, в которых координируемый атом кислорода стерически легко доступен. Тип координации  $K^{02}$  является преобладающим для оксалат-лиганда и реализуется в 47.6% структур. Карбонат- и тартрат-анионы имеют несколько практически равновероятных координаций. Встречаемости 9.9%, 9.5%, 7.9% соответствуют первой тройке координаций  $T^6$ ,  $T^{33}$ ,  $T^{63}$  карбонат-аниона, а 22.8%, 18.4%, 17.6% соответствуют первой тройке координаций  $G^{22}$ ,  $G^{23}$ ,  $K^{02}$  тартрат-аниона.

Очевидно, что большое разнообразие координаций является существенным препятствием для прогнозирования строения соединений, содержащих такие лиганды. Для упрощения построения прогнозов более предпочтительны лиганды с одним возможным типом координации. Хотя их доля в рассматриваемой выборке составляет 58.53%, но часто встречающиеся представители таких лигандов могут быть координированы только к одному атому металла, имея типы координации  $M^1$  для тетрагидрофурана  $C_4H_8O$  (2.36%), этанола  $C_2H_6O$  (0.40%), диэтилового эфира  $C_4H_{10}O$  (0.19%) или  $B^{01}$  для малтолата  $C_6H_5O_3^-$ <sub>459<sup>3</sup></sub> (0.11%) и 4-формил-2-метоксифенолата  $C_8H_7O_3^-$ <sub>163</sub> (ваниллина). Используя лиганды с координациями  $B^{01}$  и  $M^1$  нельзя построить структуры с ненулевой размерностью, для этого необходимо наличие мостиковых лигандов. Только в двух случаях с встречаемостью более 0.05% лиганд реализует мостиковую координацию:  $K^{02}$  для дигидроксималоната  $C_3H_2O_6^{2-}$ <sub>735</sub> и  $B^2$  для диводород-1,2,4,5-бензентетракарбоксилата  $C_{10}H_4O_8^{2-}$ <sub>305</sub>. Однако из-за малого объема

---

<sup>3</sup> Здесь и далее после формулы лиганда указан его код в TTL библиотеке, по которому можно найти этот лиганд через веб-сервисы <https://topcryst.com>

выборки (8 и 6 случаев) выводы о характерности типа координации этих лигандов имеет низкую достоверность.

Комбинации  $M^1 \& M^2$ ,  $V^{01} \& V^{11}$ ,  $V^2 \& V^3$ ,  $M^1 \& V^2$ ,  $M^1 \& V^{01}$  преобладают при наличии в структуре двух типов координации (17.50%) и свойственны метанолу  $CH_4O_6$ , дипивалоилметану  $C_{11}H_{19}O_2^-_{27}$ , 1-нафтилацетату  $C_{12}H_9O_2^-_{742}$ , 1,4-диоксану  $C_4H_8O_2_{238}$ , 4-формилбензоату  $C_8H_5O_3^-_{218}$ , соответственно. За исключением комбинации  $M^1 \& V^{01}$ , перечисленные лиганды могут быть мостиковыми. Далее остановимся более подробно на трех новых дескрипторах, предлагаемых в данном исследовании.

*Координационная фигура.* Мы сформировали библиотеку, которая включает в себя 219 различных классов координационных фигур (**Таблица П3**) и может использоваться для классификации координационных фигур в любых кристаллических структурах. Наиболее распространенные координационные фигуры лигандов для  $CN = 1-4$  представлены на Рис. 49. Рассмотрим задачи классификации на примере координационных фигур для  $CN = 3$ , которые достаточно разнообразны и подходят для визуального представления. Согласно Хартшорну [205], для этой координации можно различить три типа фигур: плоский треугольник TP-3, треугольная пирамида TPY-3 и Т-образная форма TS-3. В нашей классификации координационные фигуры при  $CN = 3$  характеризуются тремя углами  $\{\alpha_{min}, \alpha_{mid}, \alpha_{max}\}$  (Рис. 50).

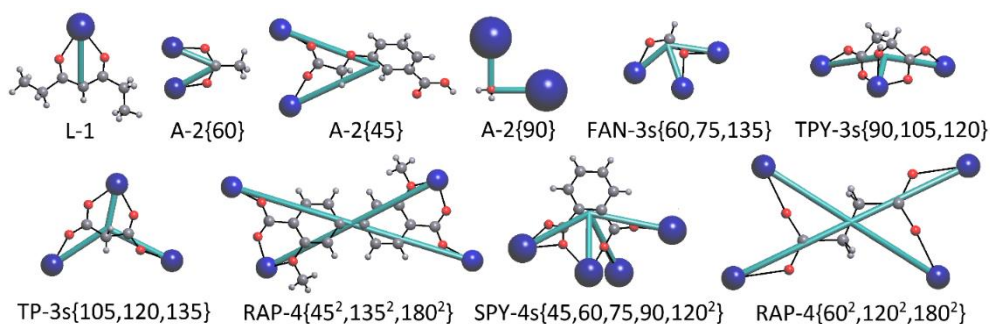


Рис. 49. Часто встречающиеся фигуры координации лигандов (более подробную информацию см. в **Таблице П4**).

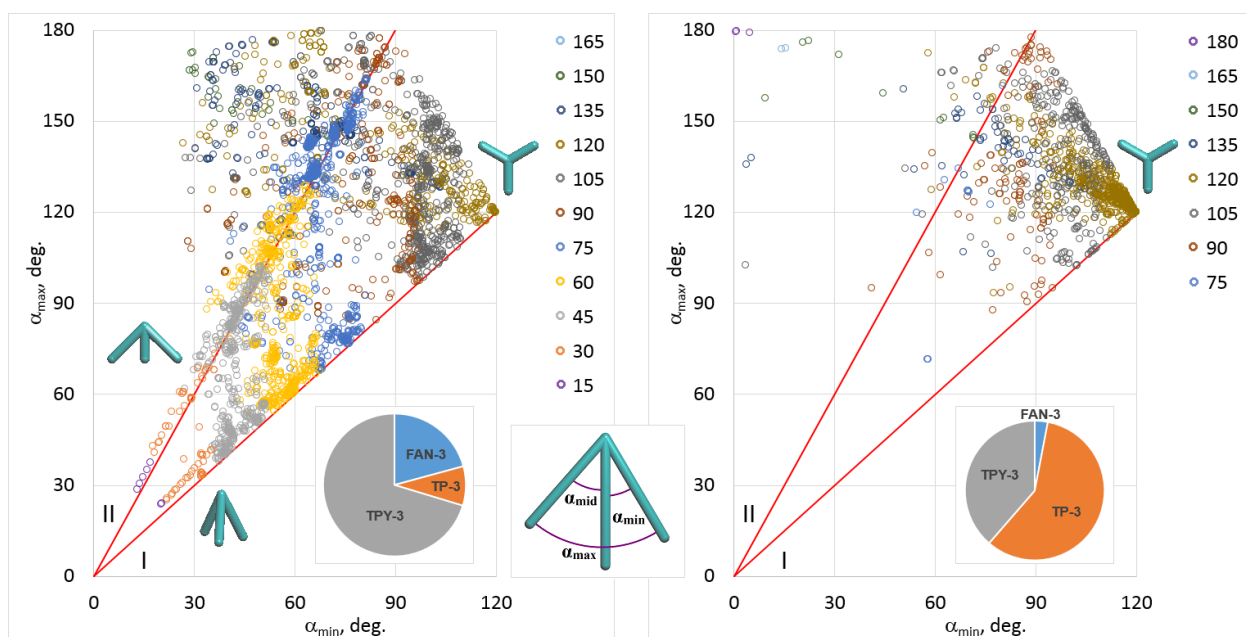


Рис. 50. Конфигурационные пространства углов координационных фигур с  $CN = 3$  для лигандов (слева) и атомов металла (справа). Значение  $\alpha_{mid}$  отображается цветом точек с шагом  $15^\circ$ . Линии I и II соответствуют различным конфигурациям типов  $TPY-3\{\alpha^3\}$  и  $FAN-3\{\alpha^2, 2\alpha\}$ . Круговые диаграммы показывают встречаемость для трех классов координационных фигур.

Анализ конфигурационного пространства углов координационных фигур лигандов (Рис. 50 слева) выявил еще один тип, который является плоским и напоминает веер (линия II). Идеализированные координационные фигуры записаны в наших обозначениях следующим образом:  $TPY-3\{\alpha^3\}$ ,  $TP-3\{120^3\}$  и  $FAN-3\{\alpha^2, 2\alpha\}$ , где первые два типа соответствуют типам Хартшорна  $TPY-3$  и  $TP-3$ . Тип  $TS-3$  Хартшорна является предельным случаем  $FAN-3\{\alpha^2, 2\alpha\}$  при  $\alpha = 90^\circ$ . По очевидным химическим причинам тип  $FAN-3\{\alpha^2, 2\alpha\}$  при малых  $\alpha$  может быть реализован только для лигандов, а не для атомов металла, что приводит к гораздо большей частоте встречаемости этого типа для лигандов (Рис. 50).

*Топологический тип базовой сетки.* Этот общий дескриптор довольно разнообразен; мы нашли 179, 97, 263 и 885 различных топологических типов

для 0D, 1D, 2D и 3D координационных групп, соответственно (Таблица П5, Рис. 51).

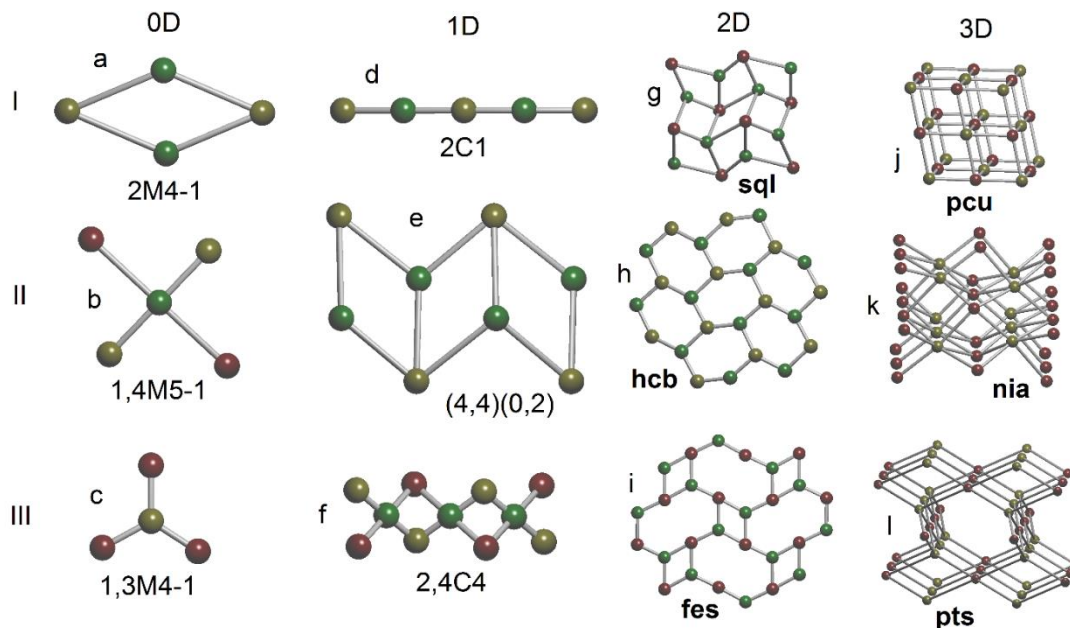


Рис. 51. Первая тройка наиболее распространенных базовых топологий для координационных групп различной размерности.

### 3.2.2. Выбор оптимального машинного метода анализа данных

Перед применением методов выборка должна быть собрана так, чтобы соответствовать разнообразию дескрипторов. Хотя мы использовали ряд ограничений при подготовке данных, разнообразие значений дескрипторов оказалось слишком велико, чтобы все их можно было идеально описать в выборке. Например, тип координации лигандов, координационные фигуры лигандов и металлов и топологический тип базовой сетки имеют 385, 213, 168 и 1439 различных значений, соответственно. При этом некоторые значения прогнозируемого параметра встречаются только в одной кристаллической структуре, что не позволяет использовать метод перекрестной проверки (см. Раздел 1.5.4) и получить качественную модель.

Чтобы уменьшить пространство признаков при сохранении достаточного размера выборки, мы рассмотрели не менее 25 наиболее

распространенных топологических типов для каждой из четырех возможных размерностей (0D – 3D) координационного полимера. Если встречаемость 25-го и последующих топологических типов была одинаковой, то все они были включены в выборку. В результате число различных топологических типов в новом стратифицированной выборке оказалось равным 107, что соответствует более чем половине структур (51.2%) исходной выборки. Количество рассматриваемых типов координации лигандов, координационных фигур металлов и лигандов было сокращено до 162, 149 и 181, соответственно.

Гиперпараметры (параметры схемы прогнозирования) каждого классификатора корректировались с использованием пятикратной перекрестной проверки для каждого набора структурно-топологических дескрипторов. Для этой цели мы использовали класс GridSearchCV из Scikit-learn. Чтобы проверить работоспособность подхода машинного обучения в зависимости от объема обучающих данных, мы вычислили перекрестно проверенные кривые обучения (Рис. П2), исследуя 50 независимых разбиений полного набора данных следующим образом: (i) каждое разбиение делит полный набор данных на обучающий (80% полного набора данных) и тестовый (оставшиеся 20%) наборы, (ii) пять подмножеств различного размера генерируются из обучающего набора и используются для обучения классификатора, (iii) достоверности для каждого тренировочного набора и тестового набора рассчитываются. Наконец, достоверности усредняются по всем 50 разбиениям для каждого размера тренировочного набора.

Следующие особенности полученных графиков (Рис. 52, П2) позволяют сделать важные выводы о моделях прогнозирования:



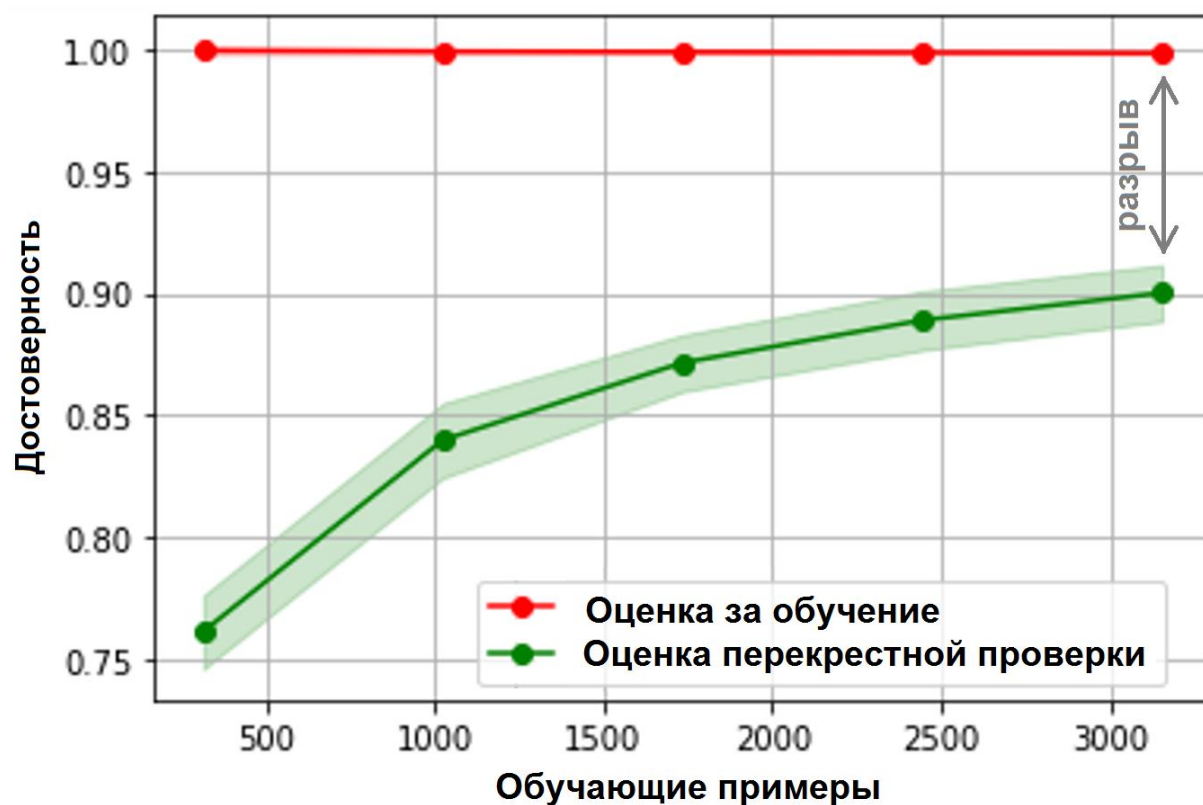


Рис. 52. Кривые обучения, полученные с использованием классификатора случайного леса (RF) для взвешенной схемы CN + CT + CF.

(i) Чем выше кривая тестирования, тем надежнее прогноз. Полученные в результате перекрестной проверки достоверности для тестовых данных, соответствующие максимальному доступному размеру обучающего набора (80% полного набора данных), имеют самые высокие значения для RF-классификатора, а лучший набор дескрипторов -  $CN_{Me} + CT_{Lig} + CF_{Lig}$  (Рис. 53, Таблица П6). Дальнейшее улучшение может быть обеспечено, если мы учтем соотношения структурных единиц и используем взвешенные дескрипторы (CN + CT + CF (взвешенные)), показан на Рис. 53). Вес определяется по стехиометрическим коэффициентам в координационной формуле. Например, цепочечная структура (KEPFEQ) с координационной формулой  $AB^{11}_2$  содержит атомы металла (A) и бидентатно мостиковые лиганды ( $B^{11}$ ) в соотношении 1:2, а все дескрипторы этих структурных единиц имеют вес  $1/3$  и  $2/3$ , соответственно.

(ii) Если кривые обучения и тестирования монотонно сходятся, прогнозирование можно улучшить с помощью большего набора данных. Это означает, что, если мы дополним нашу базу знаний результатами будущих кристаллографических экспериментов, то получим более строгую модель, которая, однако, будет соответствовать текущей модели. Все наши модели подчиняются этому условию.

(iii) Чем меньше разрыв между кривыми обучения и тестирования, тем меньше переобучение классификатора, т. е. тем лучше модель соответствует рассматриваемой задаче. Мы видим, что со временем RF-классификатор будет улучшаться после того, как у нас будет больше кристаллографических данных.

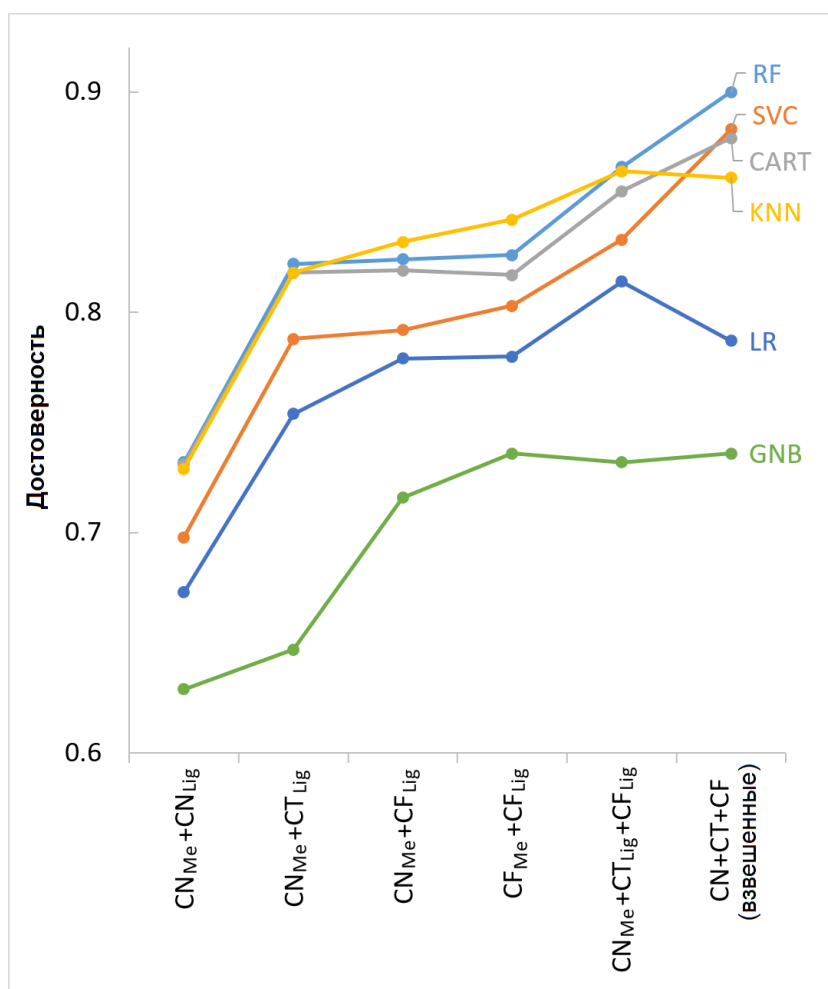


Рис. 53. Значения достоверности для прогнозирования размерности координационных полимеров различными методами машинного обучения.

Этот анализ позволяет нам предположить, что RF-классификатор обеспечивает лучшую модель прогнозирования для общих топологических параметров координационных полимеров, поэтому мы будем использовать далее только этот классификатор.

### 3.2.3. Применение полученных результатов

База знаний и соответствующая машина логического вывода могут служить интерфейсом между ученым и кристаллографическими базами данных, чтобы позволить человеку задавать вопросы для извлечения нужной информации из хранилища. Это способ преобразования экспериментальных данных в инструменты прогнозирования для дизайна новых кристаллических материалов. Ниже мы сформулируем некоторые типичные вопросы, которые возникают, когда химик планирует синтез координационного полимера и интерпретирует возможные ответы системы прогнозирования.

#### **Какая типичная координация у данного металла и лиганда?**

Важный вопрос при планировании синтеза координационного полимера касается локальной координации структурных групп. Например, если мы хотим получить алмазоподобную сетку (**dia**), было бы разумно выбрать металл с тетраэдрическими координационными и мостиковыми лигандами с координационными фигурами A-2 (изогнутая) или L-2 (линейная) (Рис. 54 вверху слева). Подобные лиганды могут быть использованы для сборки примитивной кубической (**pcu**) сетки, но в этом случае металлические центры должны иметь октаэдрическую координацию (Рис. 54 вверху справа). Лиганды также могут играть ту же топологическую роль, что и атомы металлов (Рис. 54 в середине справа).

База знаний может предоставить пользователю исчерпывающую информацию о типичных координационных числах и фигурах как металлов, так и лигандов (**Таблица ПЗ**, Рис. 48, 49). Во многих случаях эта информация нетривиальна и может существенно помочь в дизайне.

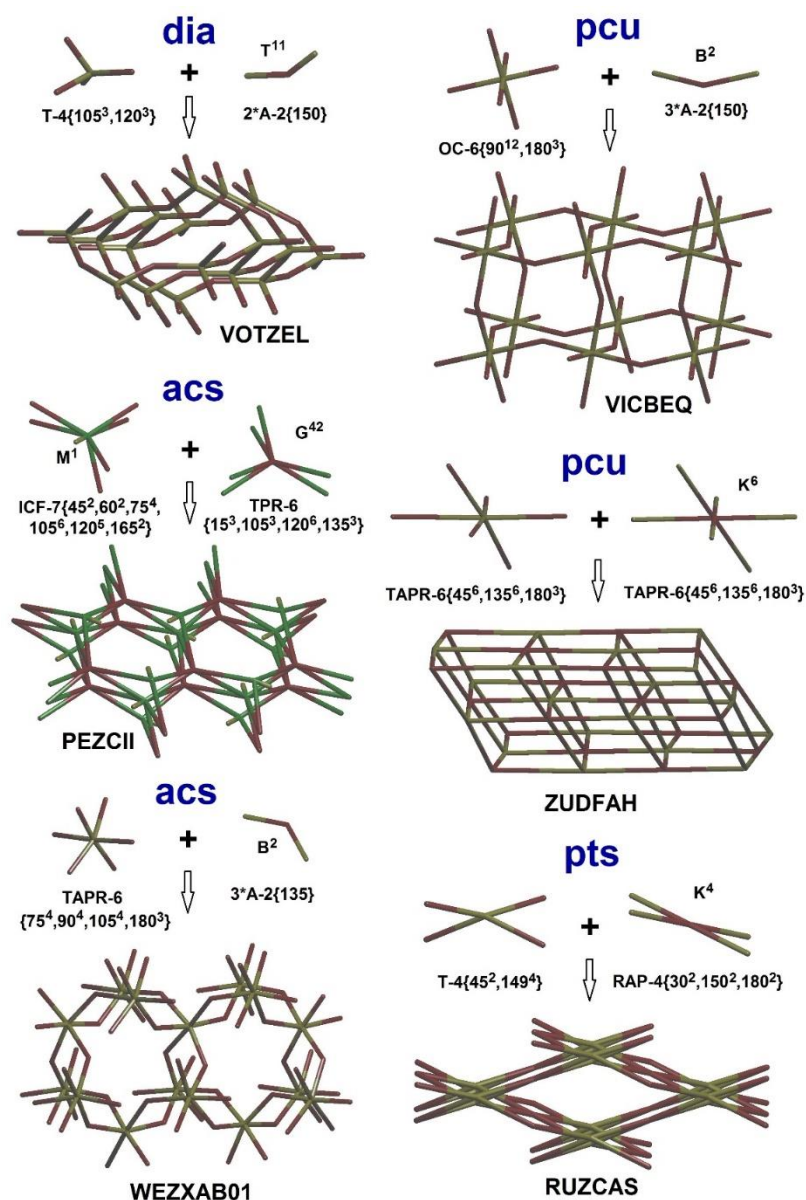


Рис. 54. Сборка координационных полимеров  $[(\mu_2\text{-лапахолато})_2\text{Cd}]$  [223] (**VOTZEL**),  $[(\mu_2\text{-формато})_3\text{Fe}]$  [224] (**VICBEQ**) и  $[(\mu_6\text{-нафталин-2,6-дикарбоксилато})\text{Pb}]$  [225] (**ZUDFAH**),  $[(\mu_6\text{-бензол-1,3,5-трис(4-бензоато))La(H}_2\text{O)}]$  [226] (**PEZCII**) и  $[(\mu_2\text{-формато-O,O'})_3\text{In}]$  [227] (**WEZXAB01**),  $[(\mu_4\text{-4,4'-диметоксибифенил-3,3'-дикарбоксилато})\text{Mn}]$  [228] (**RUZCAS**) с **dia** (вверху слева), **pcu** (справа вверху и в середине), **acs** (слева в середине и внизу) и **pts** (внизу справа) топологией базовой сетки, соответственно. Координационные типы лигандов, а также типы координационных фигур приведены для всех структурных групп.

Например, химикам хорошо известно, что типичная координация атомов Zn является тетраэдрической, а октаэдрические координационные фигуры преобладают для атомов Fe, Co или Ni. Однако какой металл следует выбрать для получения **acs** топологии, где координационная фигура представляет собой тригональную призму?

В соответствии с базой знаний, следует выбрать один из лантаноидов, для которых характерна тригонально-призматическая координационная фигура, если некоторые позиции в координационной сфере заняты монодентатными (например, вода) лигандами (Рис. 54 в середине слева). Координационные фигуры TAPR-6 {75<sup>4</sup>, 90<sup>4</sup>, 105<sup>4</sup>, 180<sup>3</sup>} атомов металла и A-2{135} лигандов также могут обеспечить данную базовую топологию (Рис. 54 внизу слева). Однако, до настоящего времени, характерные координационные фигуры лигандов были неизвестны. В нашей базе знаний впервые аккумулирована такая информация, что позволяет максимально эффективно и полно использовать ее при дизайне новых структур. Например, чтобы получить **pts** топологию (Рис. 51), нужно найти металл и лиганд с тетраэдрической и прямоугольной координационными фигурами (Рис. 54 внизу справа). Для такой цели подходит прямая комбинация часто встречающихся в тетраэдрической координации катионов Zn<sup>2+</sup>, Co<sup>2+</sup> и скворат-аниона C<sub>4</sub>O<sub>4</sub><sup>2-</sup>\_33, у которого координационная фигура квадрат реализуется с вероятностью 30.0%. Возможна и обратная комбинация склонных к квадратной координации катионов Cu<sup>2+</sup> или Ni<sup>2+</sup>, а также лиганда 2,2'-(1,3-фенилен)диацетата C<sub>10</sub>H<sub>8</sub>O<sub>4</sub><sup>2-</sup>\_582, у которого координационная фигура искаженный тетраэдр TDIST-4\_45 реализуется с вероятностью 57.9%.

**Образуют ли данные атомы металла и лиганд координационное соединение в кристаллической форме?** Как мы видим, одни и те же лиганды могут использоваться для конструирования разных топологий, но в этом случае они должны координировать разные атомы-комплексообразователи. Таким образом, информация о совместимости

атомов металла и лигандов также очень важна при планировании синтеза. База знаний может ответить на вопрос, какие металлы и лиганды наиболее часто образуют соединения по экспериментальным данным (Рис. 55). Не во всех случаях эта информация указывает на сильное или слабое сродство металла и лиганда (как, например, для нейтральных лигандов и d-металлов на Рис. 55); в некоторых случаях пик встречаемости / пропуск данных могут существовать благодаря чрезмерному / недостаточному изучению данной группы соединений. Очевидно, что выяснение причины пропуска данных является задачей для исследователя. Однако противоположный случай, то есть частая реализация данной пары металл-лиганд в координационных соединениях, однозначно отвечает высокому взаимному сродству этой пары.

Лиганд\Металл	Li	Na	K	Mg	Ca	Ba	Al	Pb	V	Mn	Fe	Co	Ni	Cu	Zn	Cd	Gd	Eu	U
формиат	5	8	2	13	5	2	-	2	-	8	8	6	3	5	10	4	1	1	-
ацетат	2	2	-	3	2	1	-	3	-	3	1	3	1	7	2	1	3	2	-
бензоат	1	1	2	1	1	1	-	1	-	1	1	1	1	2	2	-	1	1	-
оксалат	2	4	3	1	10	6	1	4	-	6	3	4	2	3	7	4	1	-	1
малонат	1	2	-	1	2	-	-	1	-	2	-	1	1	3	1	3	1	1	-
сукцинат	1	2	2	1	2	-	-	2	-	2	1	3	1	1	1	3	2	-	-
адипинат	-	-	-	1	1	1	-	2	-	2	1	1	1	1	3	1	-	-	-
терефталат	1	2	1	1	2	2	-	2	-	1	1	2	1	3	2	-	-	1	-
фталат	-	-	-	-	3	-	-	-	-	1	-	1	-	3	1	1	1	1	-
тримезат	1	-	-	2	5	2	-	1	-	1	-	4	4	2	2	1	2	-	-
пиромеллитат	-	-	-	4	2	-	-	-	-	1	1	6	-	4	5	3	-	1	-
скворат	2	3	1	1	-	4	-	-	-	2	2	2	1	1	3	8	-	-	-
тарtrat	8	3	2	7	6	4	-	2	-	3	-	4	2	3	-	6	-	1	-
цитрат	4	5	3	1	1	-	-	-	-	3	-	2	-	-	1	1	1	2	-
t-бутоксид	1	1	-	-	-	-	1	1	-	-	1	-	-	1	-	-	-	-	-
i-пропоксид	-	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	-	-	-
метанол	1	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
тетрагидрофуран	3	-	1	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
диметоксиэтан	5	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Рис. 55. Различные комбинации пар металл-лиганд для часто встречающихся структурных единиц в рассматриваемом наборе данных. Целые числа указывают количество структурных исследований.

## Схожа ли роль структурных единиц при сборке кристаллов?

Важным шагом в дизайне кристаллов является подбор структурных единиц, которые обеспечат определенный топологический мотив. Структурные единицы могут быть химически совершенно разными, но обладать схожими характеристиками в архитектуре координационных полимеров. Структурные единицы, которые имеют наименьшую разницу  $\Delta_D$  ( $\Delta_D = 0-1$ ) своих частотных спектров по отношению к дескриптору D, могут считаться обладающими наиболее сходными свойствами, которые характеризуются D. Например, при поиске похожих лигандов можно сравнить частотные спектры их координационных типов и/или координационных фигур (Рис. 56).

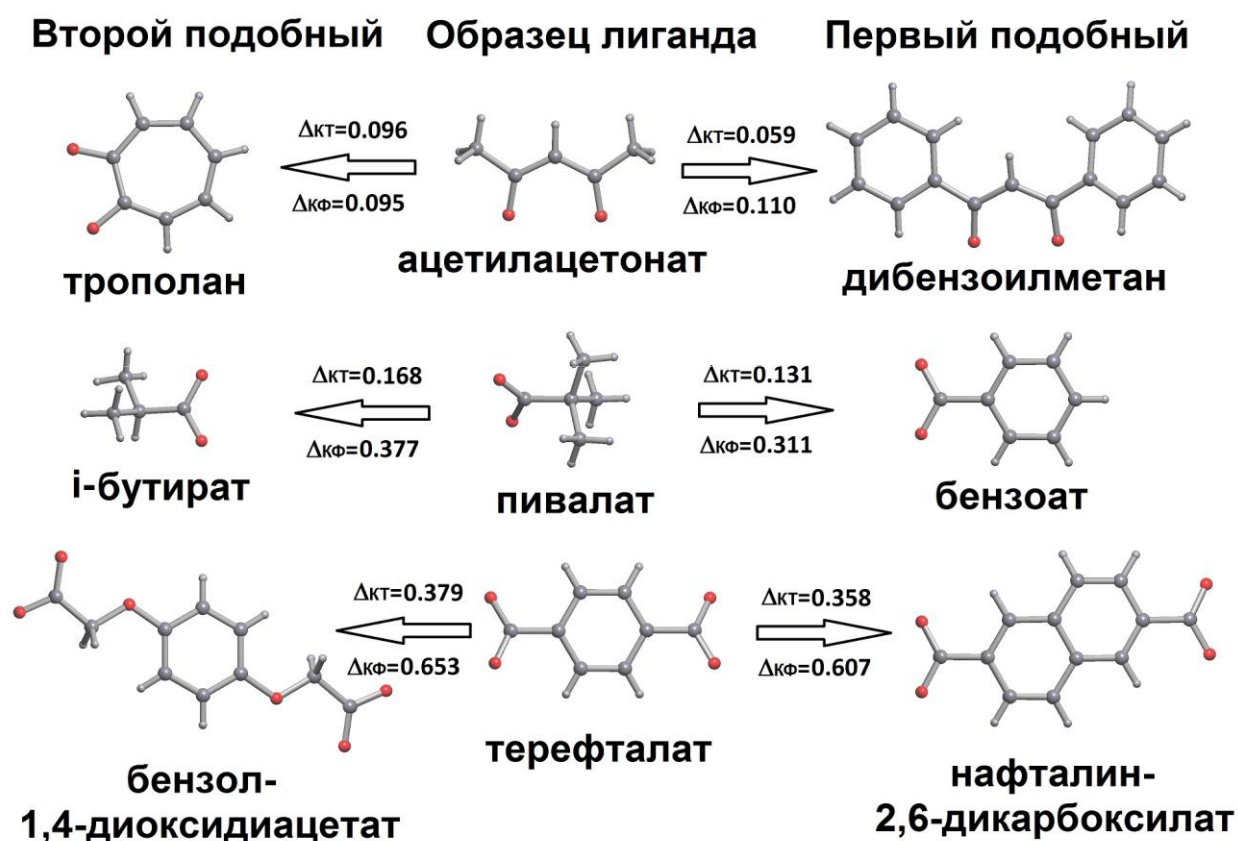


Рис. 56. Сопоставление исходных лигандов (в центре) с другими лигандами, которые наиболее похожи на исходные лиганды по своим частотным спектрам координационных типов (КТ) или координационных фигур (КФ).

В результате, можно установить, что пивалат, как структурная единица, подобен бензоату ( $\Delta_{КТ}=0.131$ ,  $\Delta_{КФ}=0.311$ ) и изобутирату ( $\Delta_{КТ}=0.168$ ,



$\Delta_{\text{KF}}=0.377$ ). На Рис. 56 показано, что некоторые лиганды (например, терефталат) не имеют близких аналогов с малыми  $\Delta_{\text{КТ}}$  и  $\Delta_{\text{KF}}$ , но, тем не менее, можно выбрать наиболее подходящие кандидаты из всего разнообразия лигандов для замены таких лигандов в координационной сетке. Используя информацию о сходстве, можно варьировать структурные единицы, чтобы приспособить их к требуемым свойствам целевого вещества (цена, стабильность, токсичность и т. д.).

### Какова кристаллическая архитектура данного соединения?

Топологический тип координационного полимера является одним из наиболее важных целевых параметров в кристаллическом дизайне. Не все 107 различных топологических типов, включенных в нашу RF-модель, могут быть надежно предсказаны (Рис. 57). Некоторые топологии (**dia**, **flu**,  $6^3\text{Ia}$ ,  $5,5\text{L4}$ ,  $2\text{M8-1}$  и другие) хорошо различимы (Точность = 1), в то время как другие (**pcc**, **bnn**, **bpq**,  $2,4\text{M6-3}$ ,  $2,2,4\text{M7-2}$  и другие) хорошо распознаются (Полнота = 1).

Наилучший прогноз может быть сделан для топологий с максимальными параметрами качества (**bcc**, **fit**, **bex**, **bey**,  $3,6,6\text{T1}$  и другие). Прогноз топологий, расположенных в нижней половине Рис. 57, не является надежным. Причинами могут быть отсутствие каких-либо экспериментальных данных (топологии  $2,2,5\text{C3}$ ,  $4^4\text{Pb}$ ,  $3,4,5\text{T3}$  и другие) или дескрипторов в схеме прогнозирования. Топологии в начале координат ( $4^4(2,2)$ ,  $4,8\text{L15}$ ,  $4,8\text{T1}$  и другие) вообще не могут быть распознаны (система предсказывает нулевую вероятность для них) также из-за недостатка данных. Топологии **fes**, **pts** или **sra** довольно многочисленны (11, 10, 7 представителей в тестовой выборке соответственно), хотя они плохо распознаются (Рис. 57), т. е. другие топологии конкурируют с ними в данном наборе дескрипторов. Например, базовые топологии **sra** и **sql** высоковероятны для квадратно-пирамидальной  $\text{SPY-5}\{90^6, 105^2, 165^2\}$  и прямоугольной  $\text{RAP-4}\{75^2, 105^2, 180^2\}$  координационных фигур атома металла



и лиганда соответственно, если лиганд имеет квадридентатно-мостиковый тип координации  $K^{22}$  (Рис. 58).

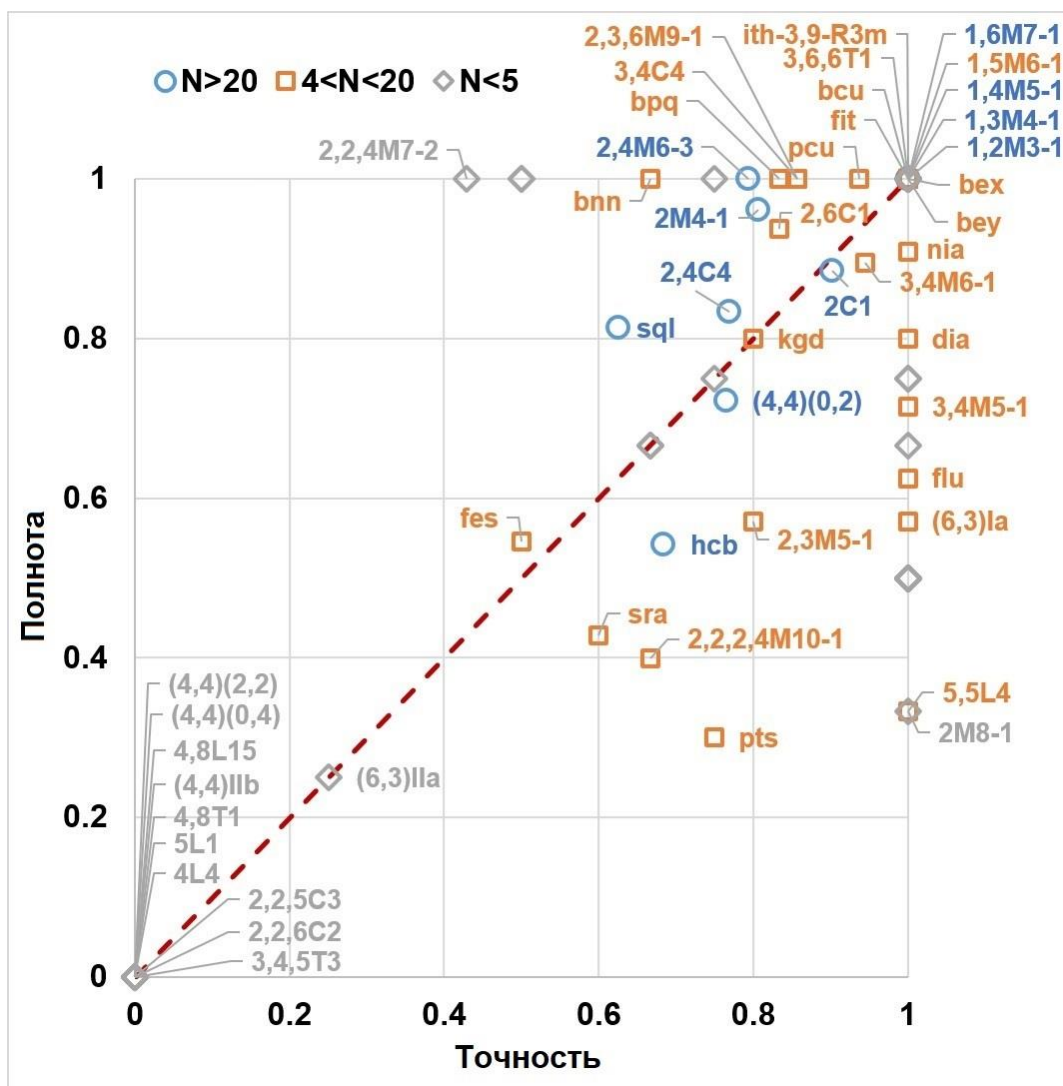


Рис. 57. Диаграмма надежности для прогнозирования топологических типов координационных соединений в рамках модели случайного леса (RF). Топологии, встречаемости которых ( $N$ ) лежат в диапазонах  $N > 20$ ,  $4 < N < 20$  и  $N < 5$ , выделены синим, оранжевым и серым, соответственно. Топологии одно- или двухпериодических сферических упаковок в обозначениях Фишера и Коха [229] приведены в виде обычного текста:  $4^4(0,2) = (4,4)(0,2)$ ;  $4^4(0,4) = (4,4)(0,4)$ ;  $4^4\Pi b = (4,4)\Pi b$ ;  $4^4(2,2) = (4,4)(2,2)$ ;  $6^3\Pi a = (6,3)\Pi a$ .

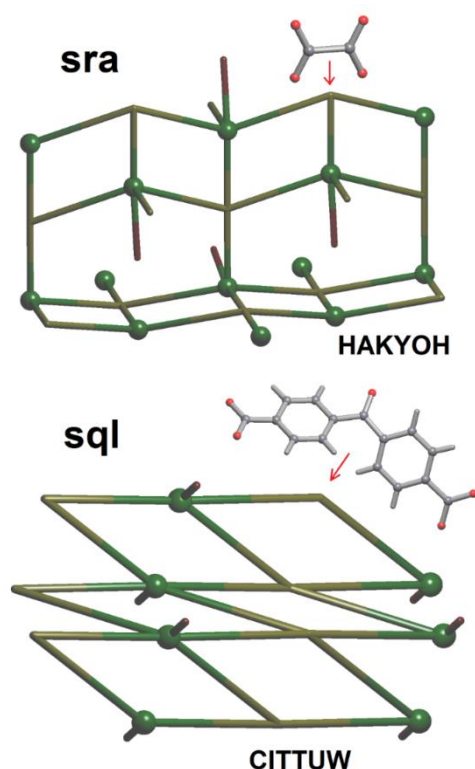


Рис. 58. Координационные полимеры  $[(\mu_3\text{-оксалато})\text{Pb}(\text{H}_2\text{O})]$  [230] и  $[(\mu_4\text{-4,4'-карбонилдибензоато})\text{Zn}(\text{H}_2\text{O})]$  [231] с топологией **sra** и **sql**, искаженные квадридентатно-пирамидальные и прямоугольные координационные фигуры атома металла (зеленый) и органического лиганда (оливковый), соответственно и квадратно-мостиковый тип координации лигандов.

Однако **sql** мотив превалирует над всеми остальными топологиями для октаэдра  $\text{OC-6}\{90^{12}, 180^3\}$  и квазилинейных ( $\text{A-2}\{135\}$  или  $\text{A-2}\{150\}$ ) координационных фигур атома металла и лиганда соответственно, если лиганд является бидентатно-мостиковым (Таблица П5). Это означает, что прогнозирующая модель может быть улучшена двумя способами: путем увеличения набора данных и расширения набора дескрипторов. Быстрый рост кристаллографических баз данных (в первую очередь, CSD) может обеспечить первое условие, в то время как геометрические и топологические дескрипторы, безусловно, должны быть дополнены параметрами электронной плотности. Таким образом, можно ожидать, что число прогнозируемых топологий будет увеличиваться со временем. Между тем, созданная нами база знаний может, по меньшей мере, предоставить пользователю информацию обо всех возможных базовых топологиях, даже если строгий прогноз невозможен при заданном наборе входных параметров.

### 3.3. Прогноз свободного пространства в структуре металлорганических каркасных соединениях

Мы показали, что размерность и топологический тип базовой сетки структуры КС и МОКП можно предсказывать исходя из формы КФ и ряда других дескрипторов атомов металла и лигандов, учитывая соотношение этих структурных единиц в кристалле. Так как МОКП являются подклассом КС, и их важнейшими характеристиками являются топологический тип базовой сетки и пористость, мы изучили взаимосвязи этих двух на первый взгляд независимых параметров кристаллической структуры.

#### 3.3.1. Выборка

Нами впервые была изучена связь между топологическими свойствами МОКП и геометрическими параметрами свободного пространства в их кристаллических структурах. Для этого использованы топологические модели периодической сетки и натурального тайлинга, а также методы их анализа, которые реализованы в программном комплексе ToposPro [215]. Мы применили эти инструменты к 33 790 кристаллическим структурам трехмерных координационных полимеров, которые были отобраны из базы данных CSD [1]. Эта выборка в пять раз больше, чем в последнем обзоре базовых топологий координационных полимеров [60], и более чем в два раза больше, чем в коллекции CoRE MOF 2019 [145]. Мы рассчитали параметры пористости для 17 857 полностью определенных структур из этой выборки. Главный вопрос, на который нужно было ответить: может ли топология каркаса предопределить пористость каркаса и в какой степени? В рамках данной работы была подготовлена выборка МОКП, выполнен ее статистический анализ, а также разработаны алгоритмы определения пористости кристаллических структур и степени искажения тайлов в них для поиска корреляции между топологией структуры и пористостью.

Полученное нами распределение базовых топологий хорошо согласуется с ранее известными данными [60]. 28 747 уникальных координационных структур распределены по 8 555 топологическим типам. Топологии **dia**, **pcu** являются наиболее распространенными (Рис. 59 вверху; Таблица П7).

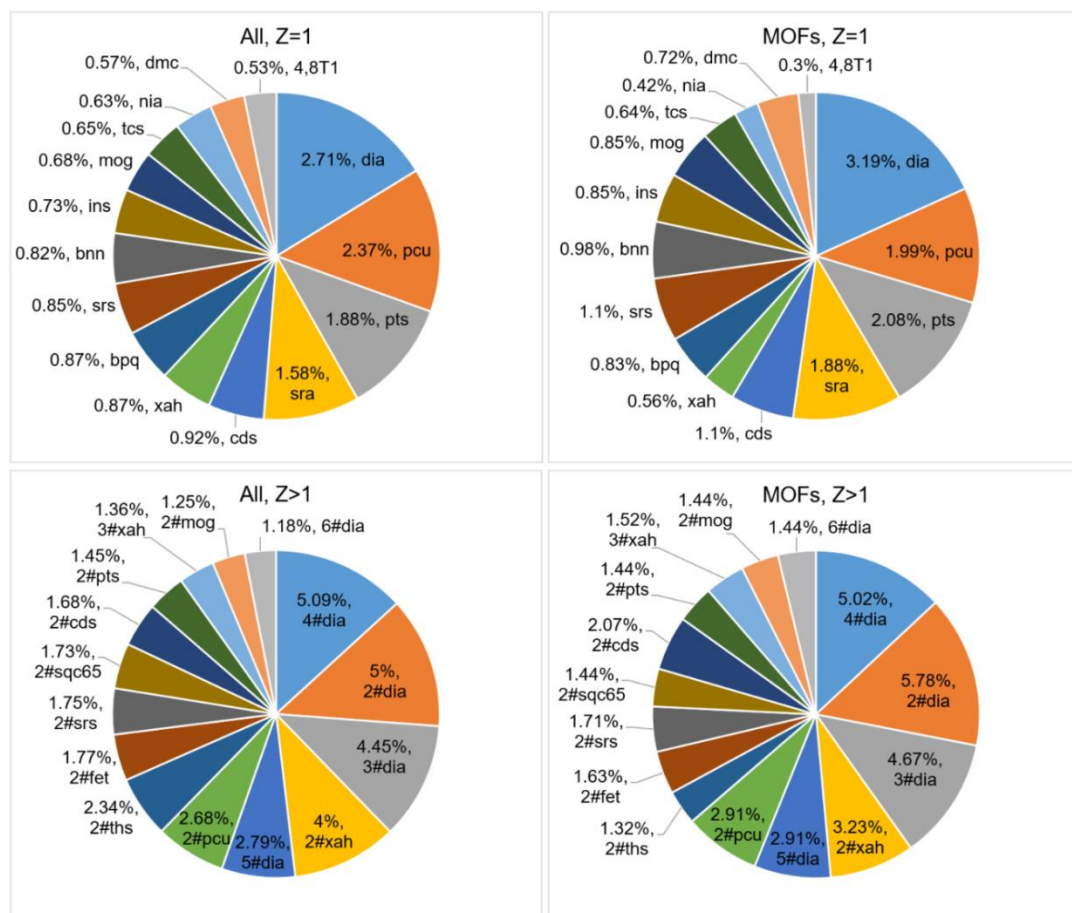


Рис. 59. Наиболее распространенные топологии в координационных полимерах с рассчитанной пористостью (слева) и во всех 3-периодических координационных полимерах (справа) для одиночных сеток ( $Z = 1$ ; вверху) и взаимопроникающих каркасов (значения  $Z > 1$  отделены от символа топологии '#'; внизу).

Разброс распределения очень широк: наиболее часто встречающаяся топология **dia** составляет всего 2.71% от общей выборки, а 7 178 топологических типов встречаются только в одной кристаллической структуре. Следует отметить, что некоторые базовые топологии подходят для

многих комбинаций атомов металла и лигандов, тогда как другие топологические мотивы могут быть реализованы только для специального химического состава. Усилия многих исследовательских групп по поиску новых комбинаций лигандов и металлов приводят к открытию сотен новых топологий ежегодно (Рис. 60).

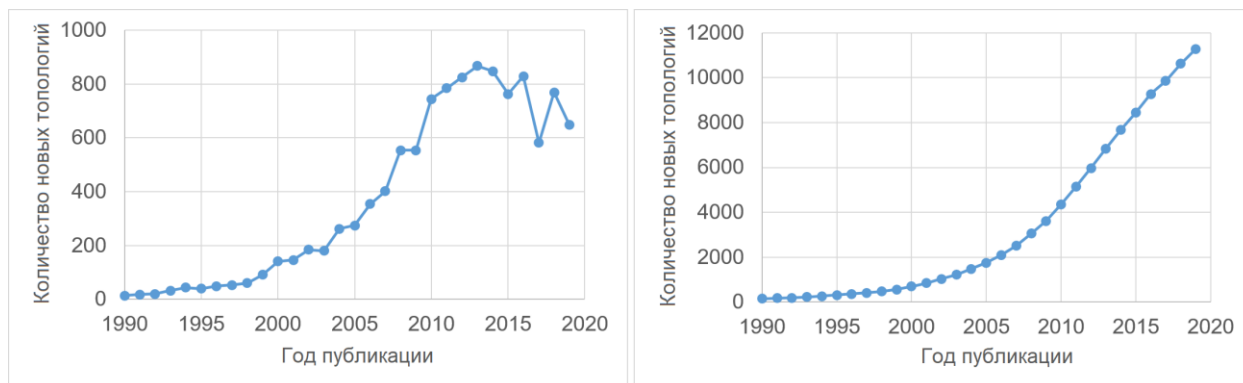


Рис. 60. Зависимость количества новых топологических типов базовых сеток 3D-координационных полимеров от года первой публикации структуры с этой топологией (слева) и соответствующая интегральная зависимость (справа).

Известно, что лишь небольшая часть сеток образует взаимопроникающие каркасы [60], однако нет строгого запрета на взаимопроникновение конкретной сетки. В общей сложности мы обнаружили 5043 взаимопроникающих каркаса, относящихся к 1032 топологическим типам, по сравнению с 975 взаимопроникающими каркасами и 91 топологическим типом в обзоре [60]. Самодуальные сетки, такие как **dia**, **ths**, **pcu**, **cds** или **srs**, особенно часто образуют взаимопроникающие пары (Рис. 59 внизу; Таблицы П7 и П8), поскольку одна сетка строго соответствует системе каналов другой сетки. Однако из 38 самодуальных сеток, содержащихся в базе данных RCSR, только 25 представлены в координационных полимерах (Таблица П9), в то время как 1007 из 1032 взаимопроникающих мотивов не являются самодуальными.

### 3.3.2. Топологическое устройство кристаллов

Существуют ли «пористые» или «плотные» топологии? Пористость является геометрическим свойством и, на первый взгляд, не должна зависеть от топологических свойств координационной сетки. Некоторые авторы сообщали [232] о «деформируемой» и «жесткой» топологиях, но не было предложено никаких строгих критериев для определения деформируемости или жесткости. Мы исследовали корреляции между топологией и пористостью в четырех группах координационных соединений с «низкой» ( $P < 0.25$ ), «нормальной» ( $P = 0.25-0.50$ ), «высокой» ( $P = 0.50-0.75$ ) и «сверхвысокой» пористостью ( $P > 0.75$ ). Можно сделать следующие выводы (Таблица П10):

- (i) Существуют топологии, которые обеспечивают только низкопористые или только высокопористые каркасы, в то время как многие другие топологии могут описывать каркасы любой пористости. Например, наиболее распространенная примитивная кубическая топология **pcu** имеет высокую вероятность реализации во всех группах, тогда как топологии **rht** или **tbo** обнаруживаются только в каркасах с высокой или сверхвысокой пористостью, а базовые сетки с топологией **ith-3,9-R3m** попадают только в «низкопористую» группу.
- (ii) Большинство топологий имеют среднюю пористость около 0.40–0.45 и большое рассеяние значений пористости ( $3\sigma > 0.30$ ), которое охватывает весь диапазон пористости. Такие топологии можно рассматривать как «деформируемые», поскольку они могут сохраняться при сильном геометрическом искажении сетки, обеспечивая тем самым разнообразие свободного пространства. Все наиболее часто встречающиеся топологии относятся к этой группе (Рис. 59), и это может объяснить их разнообразие: действительно, топология должна быть настолько более реализуемой, насколько больше разных геометрических конфигураций сетки она может реализовывать. Топологии с малым рассеянием по пористости ( $3\sigma > 0.06$ )

можно считать «жесткими»; все они встречаются не часто (не более 17 структур данной топологии, **Таблица П10**), что отражает их низкую адаптируемость к геометрическим изменениям сеток различного состава. Большинство таких сеток имеют «нормальную» пористость в диапазоне 0.25–0.50, но есть примеры из других групп пористости, и эти примеры наиболее интересны для практики.

- (iii) Детальный анализ показывает, что все группы структур с малым рассеянием значений пористости представляют собой серии изоструктурных соединений с одним и тем же лигандом (обычно это монолигандные комплексы) или со структурно сходными лигандами и атомами металлов. Это значит, что редкие топологии реализуются при особой комбинации комплексообразующих атомов и лигандов, которые имеют подходящую форму и координацию. Чтобы выбрать соответствующие структурные группы (атомы металлов, многоядерные комплексные группы и лиганды), необходима информация об их структурной роли (типичные координационные числа и фигуры), которую можно получить путем машинного анализа их топологических дескрипторов.

*Как взаимопроникновение влияет на пористость?* При дизайне пористых материалов следует избегать взаимопроникновения в большинстве случаев, поэтому топологии, которые не допускают взаимопроникновения, особенно ценны. Тем не менее, взаимопроникновение также допускает наличие пористого пространства, а некоторые взаимопроникающие каркасы (например,  $n$ -кратная алмазоподобная сетка; Рис. 61) могут даже обеспечить высокую пористость; кроме того, высокопористые мотивы могут быть даже более предпочтительными для них.



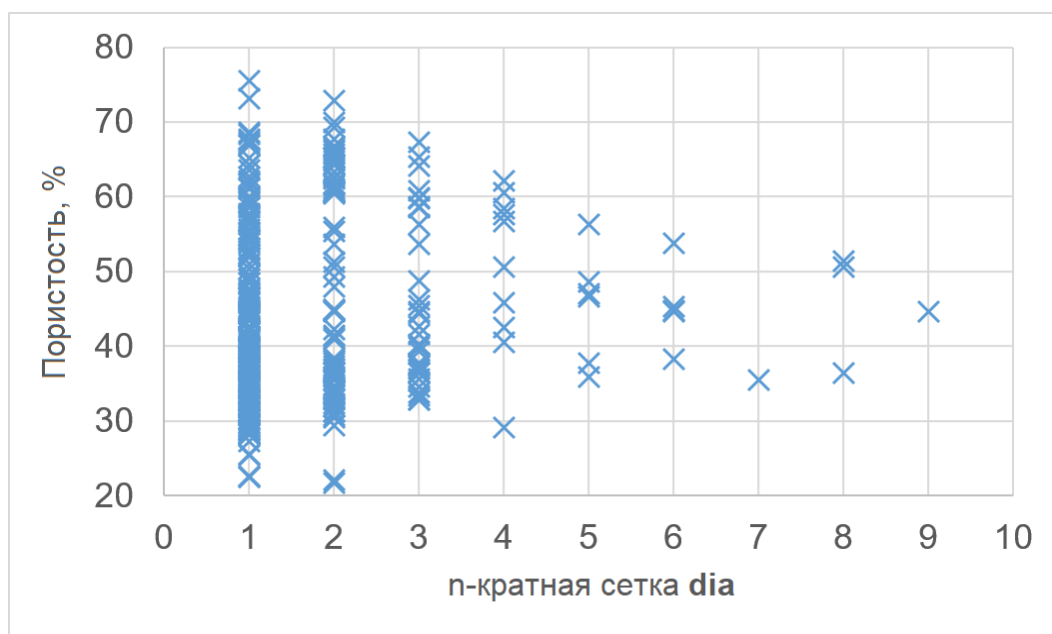


Рис. 61. Зависимость пористости от числа взаимопроникающих алмазоподобных сеток в кристаллической структуре координационного полимера.

*Как топология может описать свободное пространство кристалла?*

Общая пористость характеризует объем свободного пространства кристалла, но не дает информации о деталях его структуры. Такая информация может быть получена из натурального тайлинга, который представляет топологическую модель системы пор. Мы построили натуральные тайлинги для 1966 топологических типов базовых сеток координационных структур и обнаружили 2507 различных топологических типов тайлов, которые не встречаются в других пористых структурах, в частности, в цеолитах. Мы использовали следующие дескрипторы, полученные из натурального тайлинга, чтобы найти корреляции между топологией структуры и пористостью:

- (i) *Топология тайла*. Граф вершин и ребер тайла несет информацию об архитектуре клетки независимо от возможного ее искажения. Однако даже эта чисто топологическая информация может быть полезна при разработке пористых материалов. Для этого можно выбрать тайлы, которые типичны для пористых структур. Таким образом, хотя пористость напрямую не



коррелирует с топологическим размером тайла, то есть с числом его вершин (Рис. 62 вверху), она коррелирует с топологическим типом тайла, а некоторые тайлы обеспечивают высокую пористость благодаря их топологии (Рис. 62 внизу).

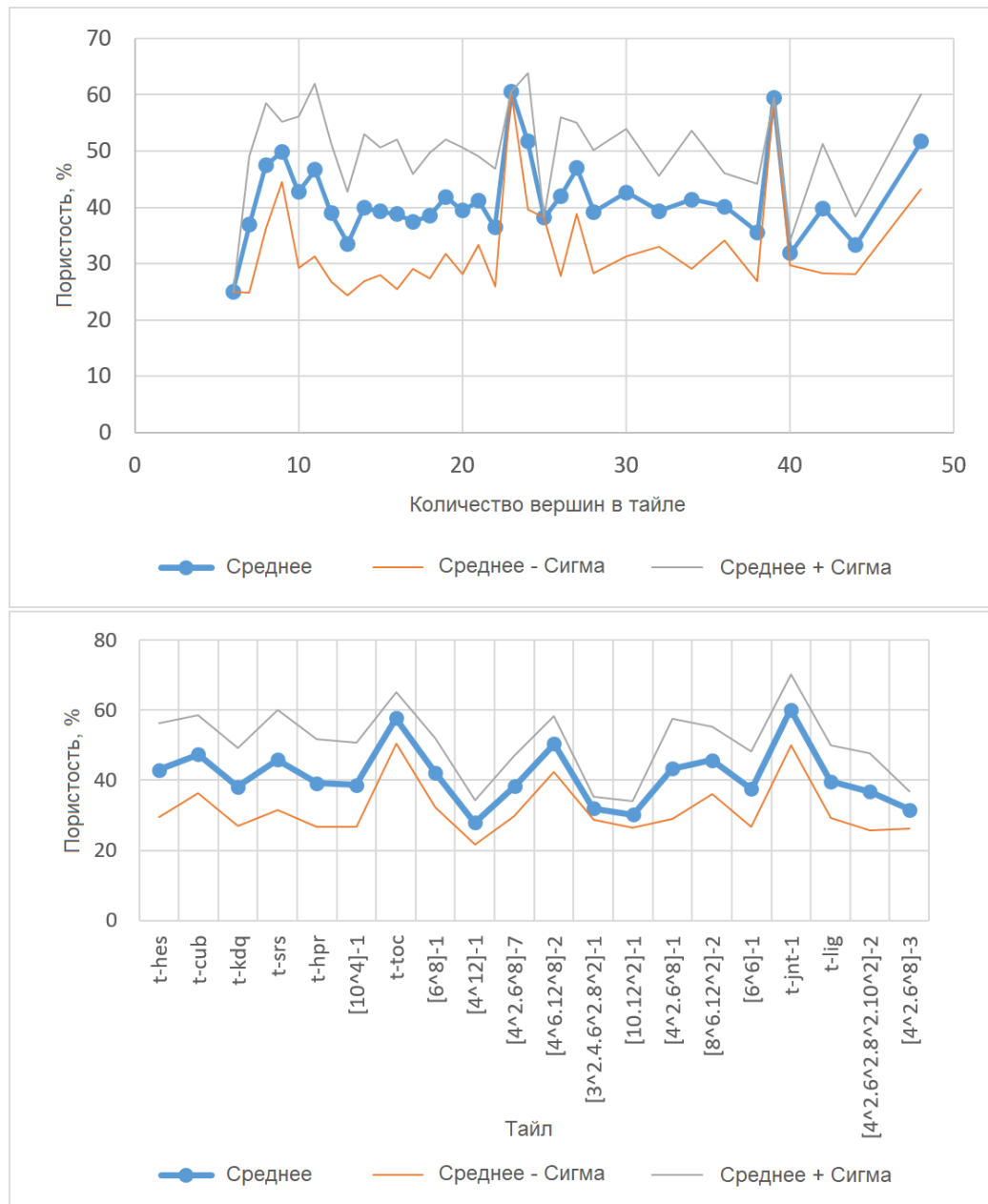


Рис. 62. Зависимость пористости МОКП от числа вершин в тайле (вверху) и его топологического типа (внизу, первые 20 наиболее распространенных тайлов) в базисной сетке тайлинга. Рассмотрены только изоэдральные (состоящие из одного типа тайлов) тайлинги. Границы, соответствующие стандартному отклонению показаны тонкими красными и серыми линиями.

Другой топологический дескриптор, индекс связности (CI), равен отношению чисел ребер тайла (e) и вершин (v):  $CI = e/v$ ; чем больше CI, тем жестче каркас тайла. По определению  $CI > 1$ ;  $CI = 3/2$  соответствует тайлингу, в которой среднее координационное число вершин равно трем, что можно считать минимальным для стабильной конструкции. Например, тайл сетки типа содалита (**sod**) имеет  $CI = 3/2$ , в то время как многие базовые сетки имеют тайлы с  $CI < 3/2$  (Рис. 63). Если координационная сетка собрана только из одного типа тайлов, то есть ее тайлинг изоэдральный, то свойства тайла могут характеризовать всю структуру. Многие часто встречающиеся сетки допускают изоэдральность, и  $CI < 3/2$  является типичным значением для большинства из них (Таблица П11).

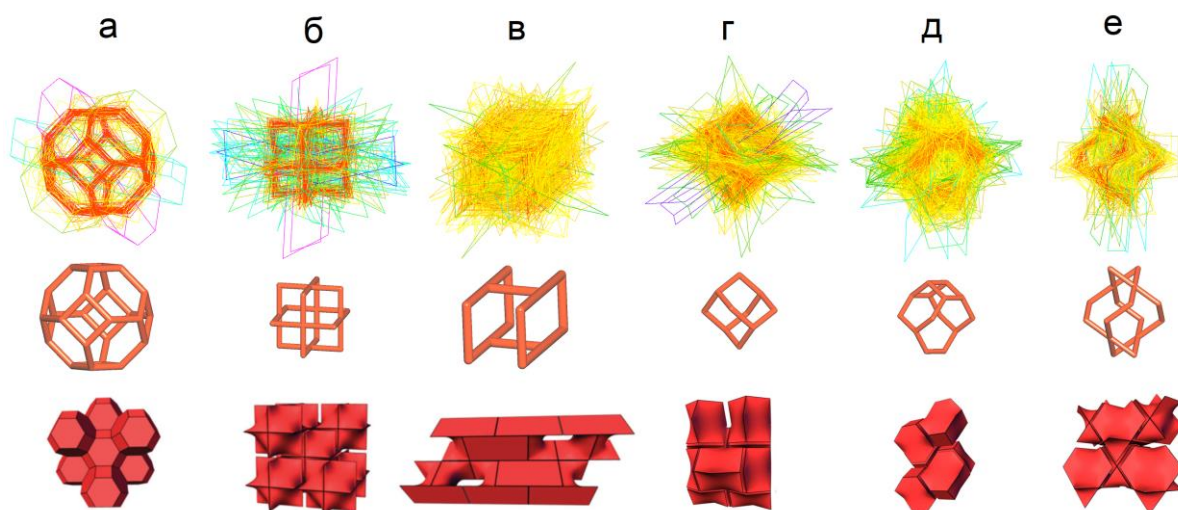


Рис. 63. (вверху) Диаграммы перекрытия для геометрически искаженных тайлов в 165 **sod** (170 тайлов;  $CI = 3/2$ ;  $\Sigma = 0.49 \text{ \AA}^2$ ; а), 86 **nbo** (137 тайлов;  $CI = 4/3$ ;  $\Sigma = 2.33 \text{ \AA}^2$ ; б), 226 **cds** (324 тайлов;  $CI = 7/6$ ;  $\Sigma = 2.64 \text{ \AA}^2$ ; в), 745 **dia** (927 тайлов;  $CI = 6/5$ ;  $\Sigma = 1.15 \text{ \AA}^2$ ; г), 411 **sra** (455 тайлов;  $CI = 9/7$ ;  $\Sigma = 2.52 \text{ \AA}^2$ ; д) и 99 **lvt** (104 тайлов;  $CI = 5/4$ ;  $\Sigma = 3.09 \text{ \AA}^2$ ; е) базовых сеток координационных полимеров по сравнению с наиболее симметричными формами тайлов (в середине) и соответствующими натуральными тайлингами (внизу). Цвет искаженных тайлов зависит от их геометрического искажения ( $\Sigma$ ): от красного (небольшое искажение) до синего (большое искажение).

(ii) *Искажение тайлов ( $\Sigma$ )*. Если базовая сетка в наиболее симметричном вложении принимается в качестве стандарта, все остальные вложения могут рассматриваться как искажения стандарта; то же самое касается натуральных тайлов сеток. Как численный метод оценки гибкости каркаса предложено использовать критерий искажения тайлов  $\Sigma$ , который рассчитывается как усредненный квадрат отклонения вершин эталона  $x$  и образца  $x'$  по формуле  $\Sigma = \frac{1}{n} \sum_{i=1}^n \|x'_i - x_i\|^2$ , где  $n$  - количество вершин в тайле. Мы обнаружили, что некоторые тайлы могут быть сильно искажены, в то время как другие тайлы в большинстве случаев остаются близкими к стандарту (Рис. 63 и 64). Более того, реализация взаимопроникающих сеток требует, как правило, более сильного искажения сетки, чем в одиночных сетках (вставка Рис. 64). Это может указывать на то, что деформируемые топологии сеток более склонны к взаимопроникновению. Если сетка является «жесткой», она может проникать только в том случае, если ее искажение вызывает редкая комбинация факторов, как в единственном примере с двумя взаимопроникающими **sod** сетками (Рис. 65). Мы предполагаем, что этот дескриптор может также распознавать «деформируемые» топологии и может использоваться для дизайна «дышащих» каркасов в комбинации с алгоритмом идентификации дублированных (то есть имеющих одинаковый состав и топологию и, таким образом, образующих одинаковые серии) каркасов [234] и дескрипторы пористости ( $P$ ,  $R_f$ ,  $R_i$ ,  $R_{fi}$  и  $PC$ ) [108] (см. Раздел 1.3.4). Таким образом, мы нашли 4428 дублированных каркасов, сгруппированных в 1306 серий; 1366 из этих каркасов, содержащихся в 479 сериях, существенно отличаются по своим параметрам пористости. Эти 479 серий можно считать потенциально «дышащими»; они распределены по 157 различным базовым топологиям, из которых наиболее часто встречаются **pcu**, **dia** и **bpq**, а также по 27 топологически различным взаимопроникающим каркасам. Многие из этих

топологий деформируются в соответствии с критерием искажения тайлов (Рис. 66). Например, в каркасах **brq** серии MIL-53 гибкость обеспечивается путем сжатия ромбических тайлов вдоль главной диагонали (Рис. 66 слева).

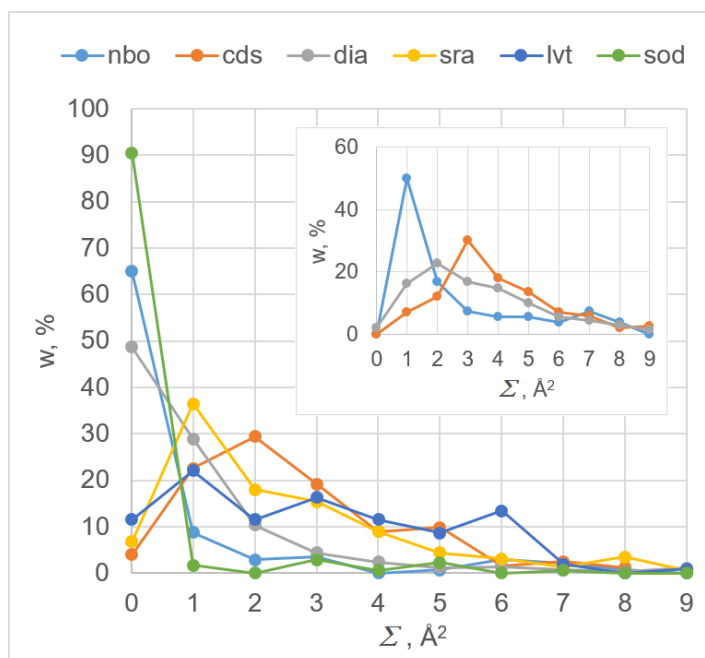


Рис. 64. Зависимость встречаемости тайлов ( $w$ ) от степени их искажения  $\Sigma$  для различных топологий в одиночных базовых сетках координационных полимеров. Вставка содержит те же зависимости, но для 50, 129 и 1034 структур, содержащих взаимопроникающие каркасы сеток с топологиями **nbo**, **cds** или **dia**, соответственно.

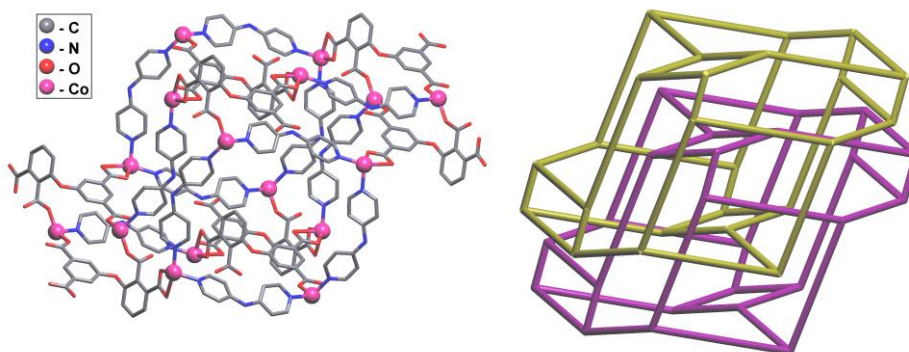


Рис. 65. Фрагмент координационной сетки (слева) и соответствующие ему натуральные тайлы (справа) во взаимопроникающем каркасе двух базовых сеток кристаллической структуры [((3,5-дикарбоксилатофенокси) фталато) (N-(пиридин-4-ил)пиридин-4-амин)Co<sub>2</sub>] (IJOGAR) [233].

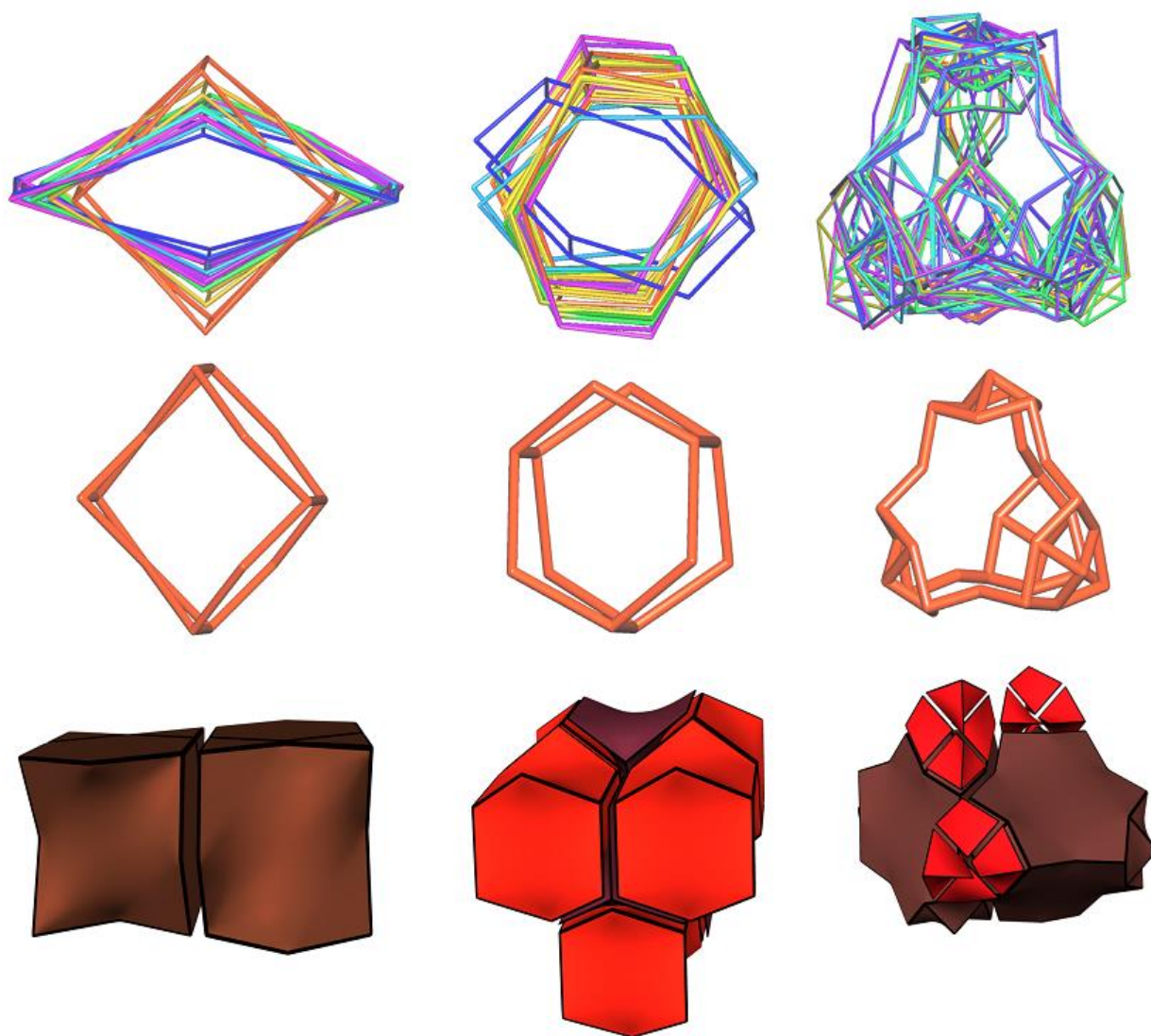


Рис. 66. (вверху) Диаграммы перекрытия для геометрически искаженных тайлов  $[3^4.6^2.8^2]$  ( $CI = 10/7$ ),  $[6^5]$  ( $CI = 5/4$ ) и  $[4^{12}.12^4]$  ( $CI = 24/17$ ) в 14 **brq** (слева)), 11 **lon** (в центре) и 25 **met** (справа) базовых сеток координационных полимеров по сравнению с наиболее симметричными формами тайлов (в центре) и соответствующими натуральными тайлингами (внизу). Цвет искаженных тайлов зависит от их геометрического искажения ( $\Sigma$ ): от красного (небольшое искажение) до синего (большое искажение).

(iii) *Топология системы каналов.* Наши данные о пористости содержат информацию о периодичности системы каналов и количестве независимых систем каналов. В зависимости от радиуса молекулы-зонда можно выбирать разные системы каналов в одной и той же структуре. Если игнорировать слишком узкие каналы, которые не могут содержать даже



такие маленькие молекулы, как молекула водорода с радиусом 1.2 Å, большинство систем каналов остаются 3-периодическими или теряют свою периодичность после расщепления на отдельные тайлы. Однако существуют топологии, которые по своей природе обеспечивают особую периодичность пор. Например, МОКП с топологией 4,8T1 имеют только однопериодичные системы каналов благодаря специальной структуре слоев с большими 12-членными гранями тайлов, выровненных в одном направлении (Рис. 67).

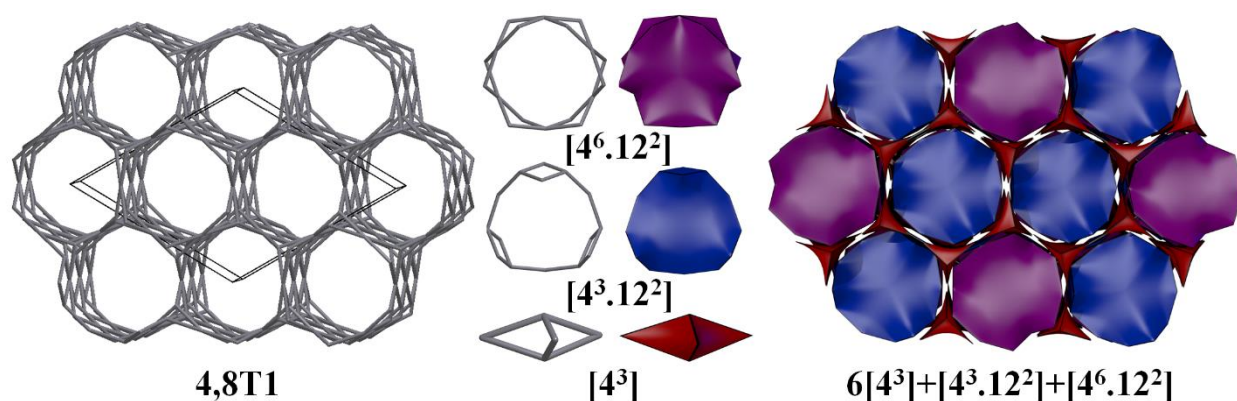


Рис. 67. Натуральные тайлы и тайлинг сетки 4,8T1.

Топологические дескрипторы, рассмотренные выше, вместе с корреляциями между ними и пористостью каркаса образуют базу знаний, которая может быть использована при практическом проектировании новых пористых координационных сеток. Вся эта информация позволяет нам обсудить последний и самый практически важный вопрос.

*Что топология может предсказать в отношении пористости?* Мы видим, что топологическая информация может быть полезна для понимания некоторых деталей структуры пор в координационных полимерах. Какие практические задачи проектирования координационных сеток могут быть решены в рамках топологического подхода, и как топология может помочь в синтезе новых пористых структур? Ниже перечислены некоторые особенности кристаллической структуры координационного полимера, которые можно прогнозировать с помощью топологических дескрипторов.

- (i) *Размерность сетки.* Поскольку низкоразмерные материалы, такие как нанотрубки или наноленты, становятся все более популярными среди координационных полимеров, важно знать условия, которые обеспечивают цепочечные или слоистые мотивы. Напротив, если необходимо получить 3D МОКП, то следует избегать низкоразмерных структур. В обоих случаях предсказание более низкой или более высокой размерности координационной сетки приобретает ценность даже без других топологических деталей. Комбинация локальных топологических дескрипторов, использованная в данной работе, позволяет успешно решить эту задачу (см. Раздел 3.2).
- (ii) *Базовая топология сетки.* Это второй общий топологический параметр, который можно успешно предсказать с помощью методов машинного обучения, используя тот же набор дескрипторов, что и для прогнозирования размерности (см. Раздел 3.2). Как было показано выше, некоторые базовые топологии имеют тенденцию образовывать пористые структуры, в то время как другие топологии являются типичными для плотных сеток. Таким образом, химик-синтетик может попытаться подобрать такой набор структурных единиц, координационные числа, типы и фигуры которых обеспечивают каркасу желаемую пористость.
- (iii) *Каркасные тайлы и каналы.* Натуральный тайл полностью описывает геометрические свойства клеток и каналов; следовательно, базовая топология сетки, которая предопределяет натуральный тайлинг, несет информацию о размере и форме всех пор каркаса.
- (iv) *Взаимопроникновение.* Как уже отмечалось ранее, самодуальность - важное чисто топологическое свойство, способствующее взаимопроникновению; однако это не объясняет существование  $n$ -кратных взаимопроникающих каркасов с  $n > 2$  и

взаимопроникновение несамодуальных сеток. Здесь показано, что еще одним свойством, приводящим к взаимопроникновению, является высокая деформируемость сетки. В целом набор взаимопроникающих структур может служить справочной информацией при планировании дизайна пористых каркасов.

- (v) *«Дышащие» каркасы.* Явление «дыхания», которое особенно важно для создания пористых сорбентов с особыми механическими свойствами, также можно предсказать, опираясь на топологию базовой сетки [235]. Обсуждаемые выше деформируемые топологии могут сыграть важную роль в получении новых «дышащих» каркасов.
- (vi) *Топология пористого пространства.* Топологические свойства системы пор полностью описываются тайлингом каркаса, а тайлинг, в свою очередь, определяется базовой сеткой. Могут быть интересны каркасы, содержащие изолированные клетки, или каналы (поры), распространяющиеся в определенном направлении, или несколько систем непересекающихся каналов.

Этот список не исчерпывает всех структурных особенностей, которые мог бы учесть и, хотя бы частично, предсказать химик-синтетик до начала эксперимента. Разработанная в данной работе база знаний может ответить, как на специальные, так и на более общие вопросы пользователя, касающиеся проблем дизайна пористых МОКП.

### **3.4. Анализ и прогноз степени окисления металлов**

Хотя задача определения степени окисления атомов металла в большинстве кристаллических структур является несложной для химика, однако так не всегда, как будет показано далее, и сложность задачи существенно зависит от объема обрабатываемых данных. Для машинного прогноза состава и строения КС и МОКП нам потребовалось определить



степень окисления атомов металла любого сорта в десятках тысячах записей, но подобные программные инструменты на тот момент отсутствовали.

### 3.4.1. Выборка

В исследовании по прогнозированию размерности и топологии (см. Раздел 3.2) была собрана информация обо всех известных, хорошо определенных кристаллических структурах координационных соединений с любыми комплексообразующими атомами металлов и оксо-лигандами, состоящими только из атомов углерода, водорода и/или кислорода. Выборка содержит кристаллографические данные по 7690 определениям строения 6171 уникальной кристаллической структуры и 13571 кристаллографически неэквивалентным атомам металлов, взятым из CSD [1] и ICSD [2]. В данном исследовании мы использовали этот набор структур как Обучающую Выборку I.

Однако мы исключили из выборки данные о 198 атомах (Таблица 14), которые имели разные неопределенные степени окисления в одной и той же структуре, а средняя степень окисления атома металла по всей структуре была формально дробной, как, например, для атомов Ce в полиоксометаллате  $\text{Ce(III,IV)}_{22}\text{O}_{20}(\text{tme}^{3-})_4(\text{ib}^-)_{26}(\text{OH})_4$  (XALZIV) [236], где  $\text{ib}^-$  =  $\text{i}$ -бутират и  $\text{tme}$  = 1,1,1-трис(оксиметил)этан. В таких соединениях значения степеней окисления не могут быть определены ни по информации базы данных (CSD не содержит информации о степенях окисления атомов во многих структурах), ни по химическому составу. Для повышения качества модели и распространения ее на ионные соединения мы дополнили выборку данными о кислородсодержащих неорганических структурах из ICSD. Как и в Обучающей Выборке I, мы отобрали только кристаллографические данные высокого качества, которые удовлетворяли следующим критериям: (i) положительные и отрицательные заряды сбалансированы; (ii) структура

полностью определена; (iii) нет статистически неупорядоченных атомов; (iv) атомы металла валентно связаны только с атомами кислорода.

**Таблица 14.** Распределения атомов металлов по степеням окисления в обучающих и тестовых выборках

Степень окисления	Обучающая Выборка I <sup>#</sup>	Обучающая Выборка II <sup>\$</sup>	Тестовая Выборка I	Тестовая Выборка II	Тестовая Выборка III
1	3293	17641	589	291	-
2	5946	26716	1540	337	-
3	2660	22416	1213	206	-
4	836	7936	541	45	-
5	194	5032	127	27	-
6	422	5336	444	87	-
7	22	297	7	1	-
Прочее <sup>&amp;</sup>	-	-	-	-	228
Общее число атомов	13373	85374	4461	994	228
Общее количество структур	7646	33253	2359	532	61

<sup>#</sup>Выборка взята из предыдущего исследования (см. раздел 3.2).

<sup>\$</sup>Обучающая Выборка II содержит Обучающую Выборку I.

<sup>&</sup>Данные для атомов металлов со смешанными и дробными степенями окисления.

Мы использовали степени окисления атомов, указанные в ICSD, без каких-либо изменений и исключили шесть соединений, содержащих семь атомов Os(VIII) или Ru(VIII), поскольку такого объема данных недостаточно для машинного прогноза. В результате Обучающая Выборка I была расширена до Обучающей Выборки II с 25 607 структурными исследованиями и 72 001 неэквивалентным атомами металла (Таблица 14, Рис. 68). Кроме того, нами были сформированы Тестовые Выборки I и II, состоящие из 2359 и 532 кристаллических структур координационных соединений, депонированных в 2020–2021 гг. в базы данных CSD и ICSD соответственно (Таблица 14).

Данные по 198 атомам, исключенные из Обучающей Выборки I, были объединены с данными по 30 атомам металлов с дробными средними значениями степеней окисления из обновлений CSD 2020 и 2021 для формирования Тестовой Выборки III; данные обучающих выборок не

содержались в тестовых выборках. Отметим, что в Тестовой Выборке III было всего три соединения, атомы металлов которых действительно находились в дробных степенях окисления, как, например, упомянутое выше соединение  $\text{Co}^{+2.5}_{12}\text{O}_3(\text{crb}^{3-})_8$ .

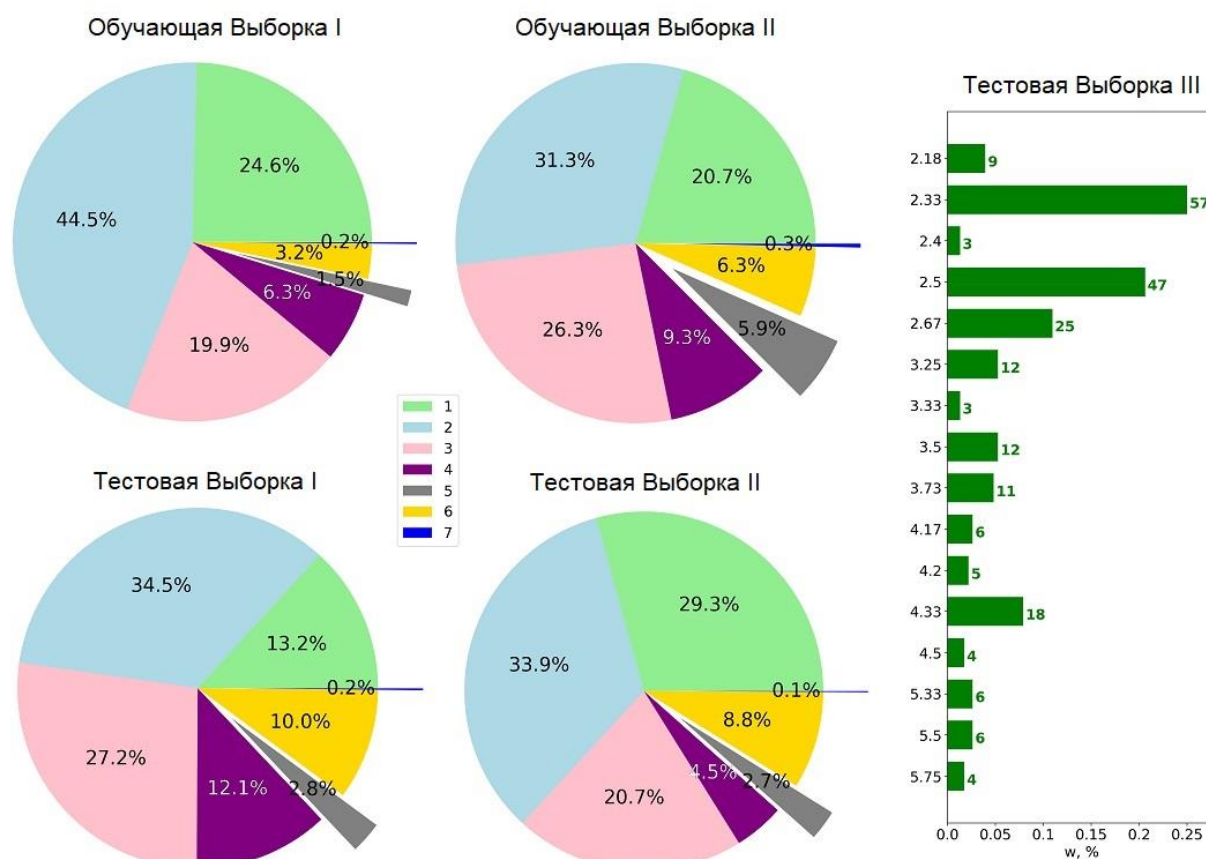


Рис. 68. Распределения значений степени окисления в двух обучающих (вверху) и двух тестовых (внизу) выборках (Таблица 14), а также в тестовой выборке с кристаллическими структурами, содержащими металлы, усредненные по структуре степени окисления которых являются дробными (справа).

Помимо степени окисления, выбранной в качестве целевой переменной, мы включили в пространство признаков 12 дескрипторов, характеризующих атомы и их окружение (Таблица 15). Связность атомов в кристаллических структурах определялась с помощью метода «Domains», реализованного в модуле AutoCN [37].

**Таблица 15.** Список дескрипторов, используемых в схеме прогнозирования степени окисления атомов металлов

Дескриптор	Описание
OxSt	степень окисления атома металла
N <sub>At</sub>	порядковый номер атома металла в Периодической таблице
group	номер группы атома металла в Периодической таблице
mult	кратность кристаллографической позиции
CN	координационное число
WCN	координационное число, взвешенное по телесным углам граней ПВ, соответствующих контактам атома металла с атомами кислорода
V <sub>CP</sub>	объем координационного полиэдра (Å <sup>3</sup> )
G <sub>3</sub>	второй момент инерции ПВ
Δ <sub>CM</sub>	смещение атома металла из геометрического центра тяжести его ПВ (Å)
R <sub>SD</sub>	радиус сферы с объемом ПВ (Å)
V <sub>VP</sub>	объем ПВ (Å <sup>3</sup> )
S <sub>VP</sub>	площадь поверхности ПВ (Å <sup>2</sup> )
ΣΩ(O)	нормированная сумма телесных углов граней ПВ, соответствующих контактам атома металла с атомами кислорода, изменяющаяся в пределах от 0 до 1

Рассматривались только валентные контакты, которые определяют дескрипторы CN и V<sub>CP</sub>. Если окружение описывалось плоской или линейной КФ, то V<sub>CP</sub> присваивался ноль. R<sub>SD</sub> и V<sub>VP</sub> взаимосвязаны:

$$R_{SD} = (3V_{VP}/4\pi)^{1/3} \quad (12)$$

и, строго говоря, в схеме прогнозирования следует использовать только один из этих дескрипторов, но мы сохраняем оба, чтобы позволить пользователю применять любой из них.

Все распределения дескрипторов были проанализированы на предмет возможных ошибок в исходных данных, неправильные данные были исключены из выборок. Однако выбросы значений дескриптора сохранялись, если они были вызваны особенностями кристаллических структур, которые сложно проверить в автоматизированном режиме. В большинстве случаев

проблема заключалась либо в разупорядочении некоторых структурных фрагментов, либо в неполноте определения всех атомов структуры. Например,  $R_{SD}$  и  $G_3$  для некоторых атомов имеют существенно большие значения, чем средние значения по выборке (Рис. 69), что обусловлено

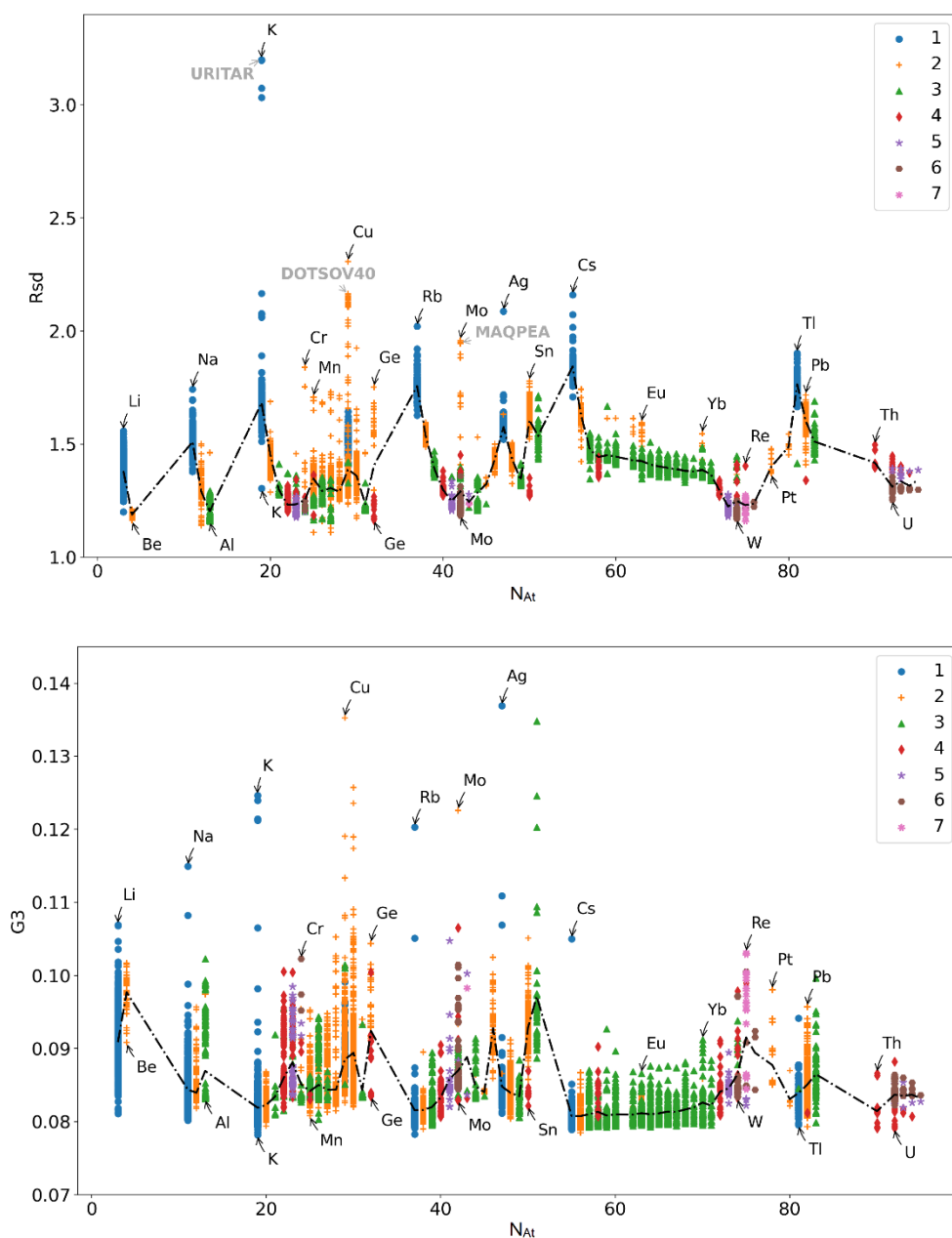


Рис. 69. Распределение значений  $R_{sd}$  (вверху) и  $G_3$  (внизу) атомов металла в Обучающей Выборке I. Значения  $G_3$  для 99 точек, которые варьируются от 0.14 до 0.76, не показаны на графике. Черная пунктирная кривая соответствует средним значениям для всех степеней окисления конкретного металла.

наличием нелокализованных молекул воды и ионов калия (Рис. 70 слева), нелокализованных молекул органических растворителей (Рис. 70 справа) или координационно-ненасыщенных металлических центров (Рис. 70 в центре). В целом количество таких выбросов не превышает 0.8% выборки (104 атома в 86 кристаллических структурах) и не влияет на предсказательную способность схемы.

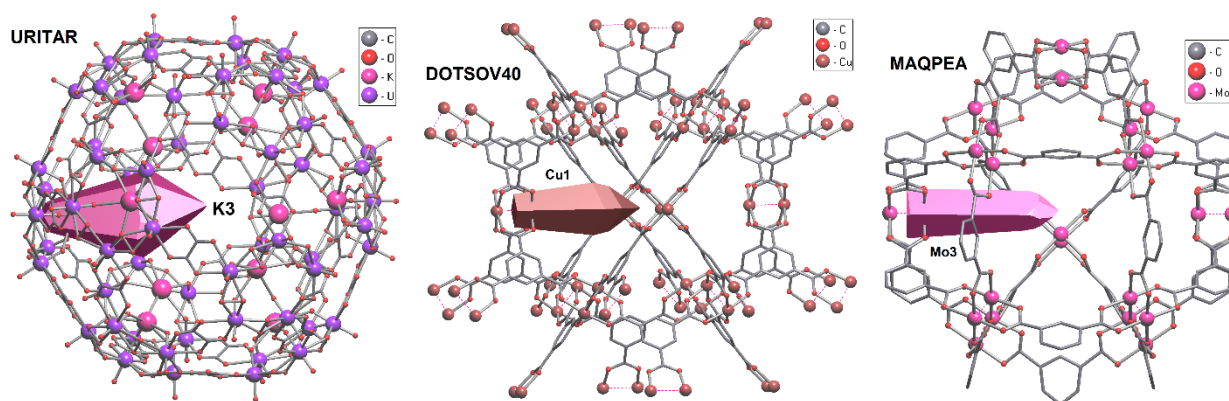


Рис. 70. Примеры ПВ атомов калия (слева), меди (в центре) и молибдена (справа) с завышенными объемами в кристаллических структурах  $K_{60}[\{(UO_2)(O_2)\}_{60}(C_2O_4)_{30}] \cdot nH_2O$ , [237]  $Cu(II)_3btc^{3-}_2$ , [238] and  $Mo(II)bdc^{2-}$  [239], где  $btc = 1,3,5$ -бензолтрикарбоксилат,  $bdc =$  бензол-1,3-дикарбоксилат.

### 3.4.1. Методы машинного анализа

Недавние машинные анализы координационных соединений [121] показали, что алгоритм случайного леса (RF) является оптимальным для предсказания степени окисления атомов, а также топологических особенностей кристаллических структур (см. Раздел 3.2). В данной работе мы использовали исключительно этот алгоритм как для выбора параметров, так и для построения схемы прогнозирования.

Мы разработали схему прогнозирования степени окисления в три этапа. На первом этапе мы выбрали признаки, которые будут использоваться для обучения схемы, и рассмотрели попарные и более сложные корреляции

между ними. На втором этапе мы обучили RF-модель и выбрали оптимальный набор гиперпараметров модели. На заключительном этапе схема была проверена на тестовых выборках путем прогнозирования значений степени окисления и расчета критериев качества. Обучающая Выборка I использовалась на первых двух этапах, поскольку она достаточно большая, чтобы получить оптимальную схему прогнозирования, и не настолько велика, чтобы сделать этот анализ трудоемким. Обучающая Выборка II использовалась для точной настройки гиперпараметров схемы прогнозирования. Чтобы выбрать значимые признаки, мы оценили отношения между степенью окисления и другими дескрипторами с помощью корреляций Пирсона и Спирмена [240] (**Таблица П12**). Первая корреляция устанавливает линейную зависимость между признаком и целевым параметром, в то время как вторая корреляция является более гибкой и оценивает общие значения «больше/меньше» взаимосвязей между ними. По диапазону коэффициента корреляции  $\rho$  различали корреляции разной силы: очень сильные ( $|\rho| > 0.9$ ), сильные ( $0.7 < |\rho| \leq 0.9$ ), умеренные ( $0.4 < |\rho| \leq 0.7$ ), слабые ( $0.2 < |\rho| \leq 0.4$ ) и очень слабые ( $|\rho| \leq 0.2$ ). Процедура выбора дескрипторов, которые будут устойчивыми для прогнозирования степени окисления, включала следующие два этапа: (i) выбор признаков, которые значимо (очень сильно, сильно или умеренно) коррелируют со степенью окисления; (ii) анализ связей между значимыми признаками и выявление коррелирующих или детерминистски связанных признаков.

Для оценки устойчивости схемы прогнозирования использовались численные критерии качества: достоверность, точность, полнота и F1-мера, значения которых были рассчитаны с помощью десятикратной перекрестной проверки (см. Раздел 1.5.4). Для тонкой настройки классификатора мы рассчитали оптимальные гиперпараметры, используя класс GridSearchCV библиотеки scikit-learn [197]. Чтобы проверить производительность модели в

зависимости от объема обучающей выборки, мы также рассчитали перекрестно проверенные кривые обучения.

### 3.4.2. Результаты и обсуждение

По матрице несоответствий Пирсона (**Таблица П12**) можно выделить три группы признаков, которые сильно или очень сильно коррелируют друг с другом: группа метрик ( $R_{sd}$ ,  $V_{VP}$ ,  $S_{VP}$ ), группа окружения (CN, WCN и  $V_{CP}$ ) и группа равномерности ( $G_3$  и  $\Delta_{CM}$ ). При анализе ранговых корреляций по Спирмену (**Таблица П12**) были выделены только две группы ( $R_{SD}$ ,  $V_{VP}$ ,  $S_{VP}$ ) и (CN, WCN,  $V_{CP}$  и  $G_3$ ); признак  $G_3$  попал во вторую группу с высоким отрицательным значением коэффициента корреляции ( $\rho = -0.83$ ).

Мы рассчитали показатели качества для всех 4095 комбинаций 12 признаков. Лучшие комбинации реализовывались в наборах из 4-8 признаков с F1-мерой от 0.951 до 0.955 (**Таблица 16**); все эти наборы содержат дескрипторы  $N_{At}$ , CN и  $G_3$ . Если наличие признака  $N_{At}$  следует из Периодического закона, то признаки CN и  $G_3$  менее очевидны. В итоге мы включили в схему прогнозирования шесть признаков ( $N_{At}$ , Group, CN,  $R_{SD}$ ,  $G_3$  и  $\Sigma\Omega_O$ ), комбинация которых имела наибольшее значение F1-меры, и исключили дескрипторы Mult, WCN,  $V_{CP}$ ,  $\Delta_{CM}$ ,  $V_{VP}$  и  $S_{VP}$ , которые либо слабо коррелировали со степенью окисления, либо принадлежали к тем же корреляционным группам, что и один из шести признаков.

**Таблица 16.** Оптимальные наборы параметров для прогнозирования степени окисления с помощью классификатора «случайный лес»

Количество признаков	$N_{At}$	Group	Mult	CN	WCN	$V_{CP}$	$G_3$	$\Delta_{CM}$	$R_{sd}$	$V_{VP}$	$S_{VP}$	$\Sigma\Omega_O$	Достоверность	Точность	Полнота	F1-мера
6	+	+	-	+	-	-	+	-	+	-	-	+	0.985(27)	0.938(32)	0.941(25)	0.955(26)
4	+	-	-	+	-	-	+	-	+	-	-	-	0.984(3)	0.963(10)	0.938(23)	0.954(20)
	+	-	-	+	-	-	+	-	-	+	-	-	0.983(4)	0.963(10)	0.940(38)	0.951(19)
8	+	+	+	+	-	+	+	-	-	+	-	+	0.985(4)	0.953(19)	0.941(36)	0.951(32)



Выбранные признаки не сильно коррелировали друг с другом и описывали различные свойства атомов металла: химические свойства (NAt, Group и  $\Sigma\Omega_0$ ), геометрический размер ( $R_{SD}$ ), число соседей (CN) и равномерность структурного окружения ( $G_3$ ).

Анализ полученных кривых обучения и матриц несоответствий позволил выявить следующие важные особенности схемы прогнозирования:

- (i) Кривые обучения (Рис. 71) достигают достаточно высоких значений точности для максимального размера обучающих выборок (90% полного набора данных, **Таблица 17**), что доказывает надежность прогноза.
- (ii) Схема прогнозирования не переобучена, поскольку кривые обучения и тестирования сходятся (разрыв между ними сокращается) с увеличением размера выборки (Рис. 71).

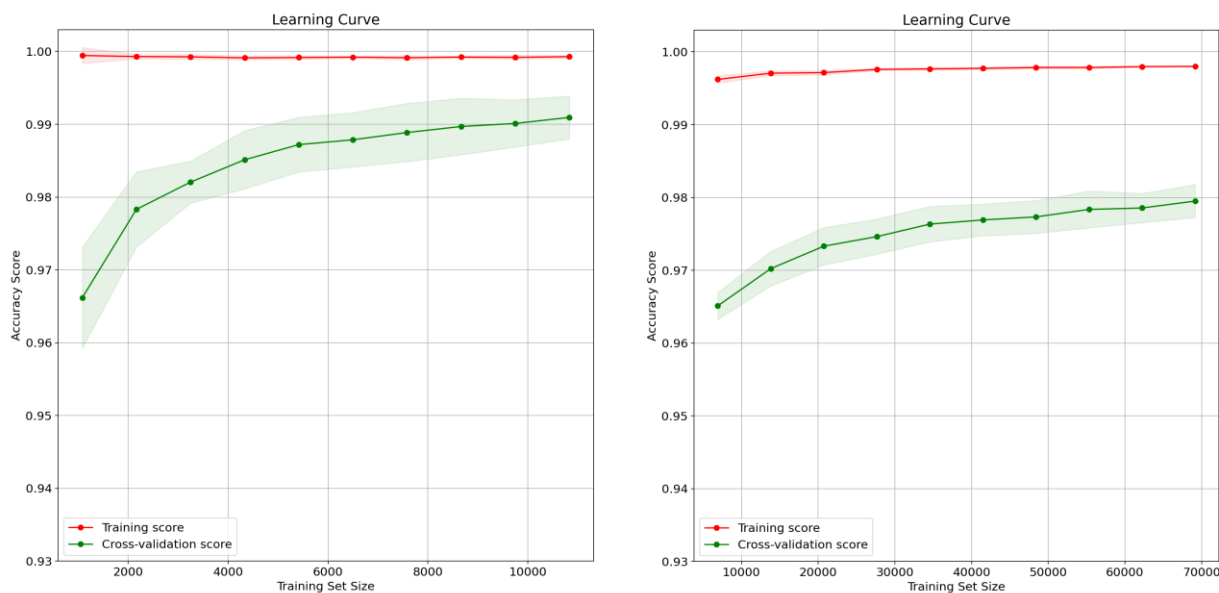


Рис. 71. Кривые обучения, рассчитанные методом «случайный лес» с оптимизированными гиперпараметрами для Обучающей Выборки I (слева) и Обучающей Выборки II (справа).

(iii) Схема идеально предсказывает все значения степени окисления в Обучающих Выборках I и II; совпадение реальных и прогнозируемых значений составляло 0.95 или выше, за исключением класса OxSt=7, который наименее представлен в обучающих выборках (Рис. 72).

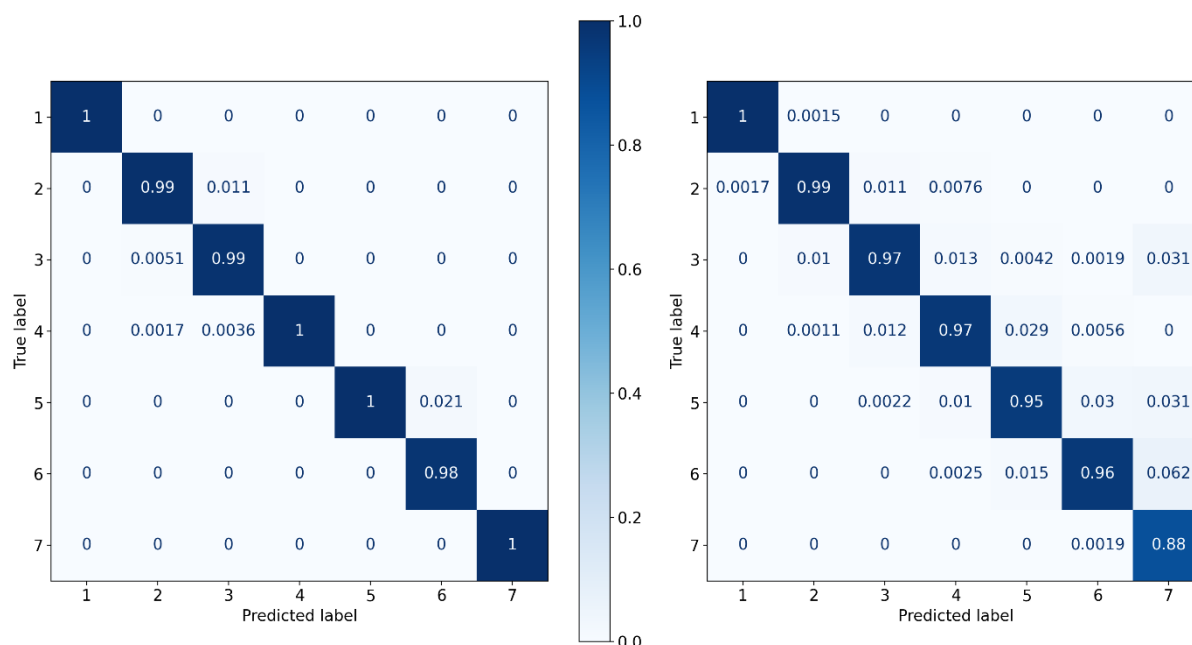


Рис. 72. Матрицы неточностей, рассчитанные для схемы прогнозирования, обученной с помощью Обучающей Выборки I (слева) и Обучающей Выборки II (справа).

(iv) Прогноз с использованием Тестовых Выборок I и II показал наилучшее качество для схемы, обученной с использованием самой большой Обучающей Выборки II. Хотя значения точности были высокими (0.959-0.990) для всех комбинаций испытаний (**Таблица 17**), другие критерии, объясняющие неравномерность распределения данных по классам (значения OxSt), снизили свои значения более чем на 5% для комбинации Обучающая Выборка I и Тестовая Выборка II. Такое снижение качества указывало на нестабильность схемы, обученной на Обучающей Выборке I. Однако модель, обученная на Обучающей Выборке II, обеспечила высокие значения по всем критериям качества (**Таблица 17**). Эта модель

использовалась в качестве окончательной версии нашей схемы прогнозирования.

**Таблица 17.** Оценка достоверности, точности, полноты и F1-меры для прогнозирования значений степени окисления в процедурах кросс-проверки и тестирования

Процедура и условия	Достоверность	Точность	Полнота	F1-мера
Перекрестная проверка для Обучающей Выборки I				
<i>Гиперпараметры по умолчанию</i>	0.991(3)	0.970(33)	0.946(32)	0.952(25)
<i>Оптимизированные гиперпараметры</i>	0.991(3)	0.972(27)	0.949(30)	0.956(28)
Перекрестная проверка для Обучающей Выборки II				
<i>Гиперпараметры по умолчанию</i>	0.979(2)	0.964(9)	0.961(9)	0.962(7)
<i>Оптимизированные гиперпараметры</i>	0.980(2)	0.962(11)	0.962(9)	0.963(8)
Тестирование				
<i>Обучающая Выборка I и Тестовая Выборка I</i>	0.987	0.984	0.964	0.973
<i>Обучающая Выборка I и Тестовая Выборка II</i>	0.959	0.918	0.891	0.903
<i>Обучающая Выборка II и Тестовая Выборка I</i>	0.990	0.985	0.991	0.988
<i>Обучающая Выборка II и Тестовая Выборка II</i>	0.981	0.949	0.952	0.950

(v) Расхождения между реальными и прогнозируемыми значениями наблюдались для 63 атома в 46 структурах (Рис. 73 слева). Более детальный анализ выявил типичные причины этих расхождений.

а) Неправильное определение степени окисления в первоисточнике или ошибка в кристаллографической базе данных. Например,  $\text{Cu}(\text{L1})_2\cdot 4\text{MeCN}$  [241] где  $\text{L1} = [3,5\text{-бис(оксо)гепта-1,6-диен-1,7-диил}]\text{бис(2-метокси-4,1-фенилен)}$  диметилбикарбонат, содержит атомы  $\text{Cu(I)}$ , согласно записи CSD (MOFCOC). Наша схема предсказала  $\text{OxSt}=2$  для  $\text{Cu}$  с вероятностью  $P=0.997$ , что соответствует химическому составу этого соединения и исходным данным [241]. Предлагаемая схема предсказала нереалистичное шестивалентное состояние атомов  $\text{Eu}$  в  $\text{Eu}^{\text{III}}_2(\text{C}_4\text{O}_4)_3(\text{H}_2\text{O})_4$  (SURWIO) [242] из CSD ( $\text{H}_2\text{C}_4\text{O}_4$  = квадратная кислота), поскольку это состояние указано для того же соединения в ICSD (код коллекции 132455) и использовалось

для обучения нашей схемы прогнозирования. После фиксации этого значения в Обучающей Выборке II и переобучения схемы прогноз стал верным.

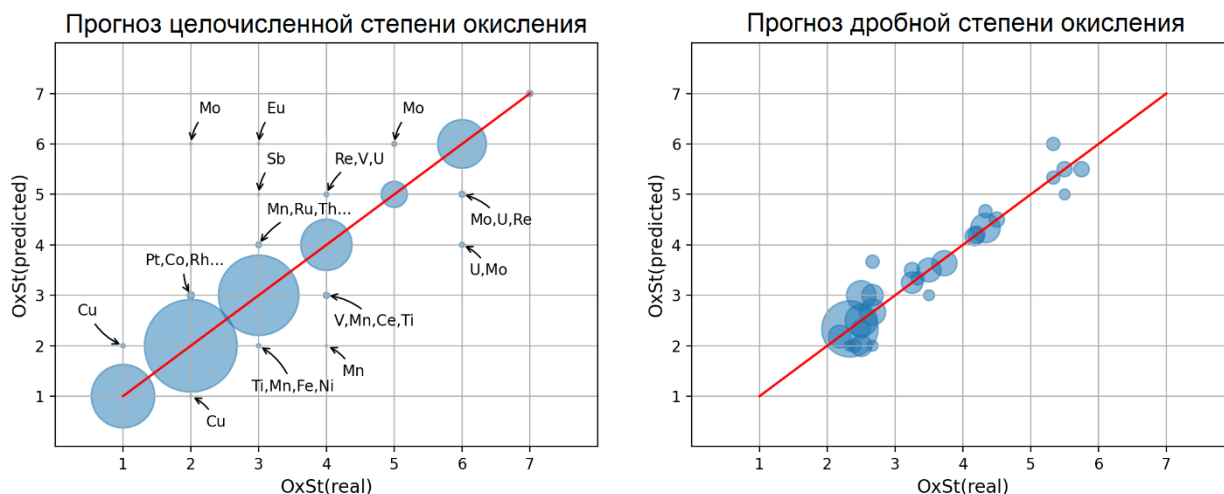


Рис. 73. Расхождения между реальными и прогнозируемыми значениями OxSt для структур из Тестовых Выборках I и II (слева) и III (справа). Размеры кружков пропорциональны возникновению определенной комбинации двух значений OxSt; правильный прогноз соответствует красной диагонали.

(б) Неправильное определение состава отрицательно заряженных частиц, приведшее к ошибкам в определении степеней окисления атомов металлов. Например, для соединения из базы данных CSD указана формула  $\text{Cu}_2(\text{L}^{2-})(\text{HL}^-)(\text{H}_2\text{O})_5 \cdot n\text{H}_2\text{O}$  (НОМQEИ), где  $\text{H}_2\text{L}$  = 5-гидроксиизофталевая кислота, что предполагает смешанную степень окисления Cu(I,II). Мы предсказали двухвалентное состояние атомов Cu и, следовательно, только депротонированные лиганды  $\text{L}^{2-}$ , что соответствовало исходным данным [243]. Соединение  $\text{K}[\text{V}_5\text{O}_8(\text{OH})_3(\text{EtCOO})_3(\text{H}_2\text{O})_2] \cdot 8\text{H}_2\text{O}$  [244] является более сложным примером, в котором смешанная степень окисления не указана в структурной базе данных, и позиции атомов водорода не были определены в структурном эксперименте. Наша схема предсказывала степень окисления +4 для атомов V(1-4) со средней вероятностью 0.93 и степень окисления +5 для атомов V(5) со 100% вероятностью; прогнозируемые

значения полностью согласуются с данными авторов [244]. Еще одним примером является  $\text{Mn(ADBEB)}_2(\text{DMF})_2$  (SUJFOV) [245],  $\text{H}_2\text{ADBEB}$  = 4,4'-(диэтинилантацен-9,10-диил)дibenзойная кислота, где степень окисления Mn должна быть +4 в соответствии с составом, указанным в оригинальной статье и CSD. Однако наша схема предсказала Mn(II), что согласуется с указанной авторами степенью окисления [245]. Это несоответствие можно объяснить, если предположить частичное протонирование карбоксильных групп в форме  $\text{HADBEB}^-$ , что подтверждает координация атомами Mn мостикового лиганда только одним атомом кислорода карбоксильной группы и синтез соединения в кислой среде (1 М HCl).

(в) Отсутствие определенного значения OxSt в обучающих выборках. Наша схема ошибочно предсказала четырехвалентное состояние Th(III) в структуре  $\text{LiTh(L5)}_4(\text{thf})_4 \cdot n\text{Et}_2\text{O}$  (CONTAD) [246], где L5 = 2,6-ди-*t*-бутил-4-метилфенолат, поскольку обучающая выборка содержали только Th(IV) (Рис. 69 вверху). Добавление данных Th(III) в обучающую выборку улучшило схему прогнозирования.

Самой сложной работой было предсказание смешанных степеней окисления для Тестовой Выборки III. Для атомов металла из этой выборки авторы и база данных не указали степени окисления, а усреднение по всей структуре дало дробные значения OxSt. Поскольку наша схема предсказывает только целые степени окисления, мы использовали совпадение средних степеней окисления, как критерий качества предсказания. Результаты нашего прогноза показали, что классификатор работает хорошо, поскольку большая часть кружков расположена на диагональной линии (Рис. 73 справа). Усредненная степень окисления правильно предсказана в 34 случаях из 61. Например, степени окисления V, IV, V и IV были правильно предсказаны для атомов V(1), V(2), V(3) и V(4) в  $(\text{C}_3\text{H}_{12}\text{N}_2)[\text{V}_4\text{O}_{10}]$  (ZOLJEP06) [247] и соответствуют значениям, указанным в оригинальной статье.

Наша схема несколько занизила среднее значение степени окисления церия (3.64 вместо 3.73) для упомянутого выше полиоксометаллата  $\text{Ce(III,IV)}_{22}\text{O}_{20}(\text{tme}^{3-})_4(\text{ib}^-)_{26}(\text{OH})_4$ , где *ib* = *i*-бутирато и *tme* = 1,1,1-трис(оксиметил)этан. Согласно оригинальной работе [236], структура состоит из нейтральных молекулярных комплексов  $\text{Ce}_{22}\text{C}_{124}\text{H}_{222}\text{O}_{88}$ , в которых имеется шесть атомов Ce(III) и 16 атомов Ce(IV). Наша схема предсказала правильную степень окисления для всех атомов Ce(III) с высокой вероятностью 0.876 - 1.000. Степень окисления +4 была предсказана правильно для всех атомов Ce(IV) первой группы, кроме Ce(8), для которого вероятность  $\text{OxSt}=+3$  составила 0.884 из-за завышенного значения  $R_{\text{sd}}$  этого атома (Таблица 18).

**Таблица 18.** Вероятности степеней окисления P(III-V) и особенности схемы прогнозирования для неэквивалентных атомов Ce в структуре  $\text{Ce}_{22}\text{O}_{20}(\text{tme}^{3-})_4(\text{ib}^-)_{26}(\text{OH})_4$  [236]. Вероятности, соответствующие предсказанным значениям  $\text{OxSt}$ , показаны жирным шрифтом.

Atom	OxSt [236]	P(III)	P(IV)	P(V)	CN	$R_{\text{sd}}$	$G_3$	$\Sigma\Omega_{\text{O}}$
Ce(1)	4	0.454	<b>0.526</b>	0.020	7	1.381	0.0834	0.965
Ce(2)	4	0.163	<b>0.837</b>	0.000	9	1.380	0.0807	1.000
Ce(3)	4	0.427	<b>0.573</b>	0.000	8	1.389	0.0824	1.000
Ce(4)	4	0.459	<b>0.526</b>	0.015	9	1.381	0.0806	0.994
Ce(5)	4	0.466	<b>0.515</b>	0.020	7	1.386	0.0833	0.962
Ce(6)	4	0.471	<b>0.516</b>	0.013	8	1.381	0.0821	0.985
Ce(7)	4	0.191	<b>0.809</b>	0.000	8	1.398	0.0843	0.999
Ce(8)	4	<b>0.884</b>	0.116	0.000	8	1.417	0.0819	0.996
Ce(9)	3	<b>1.000</b>	0.000	0.000	9	1.436	0.0809	0.985
Ce(10)	3	<b>0.876</b>	0.124	0.000	8	1.436	0.0831	0.996
Ce(11)	3	<b>1.000</b>	0.000	0.000	9	1.435	0.0807	0.983

Наша схема прогнозирования показывает, что степень окисления атома металла в координационных соединениях может быть удовлетворительно определена на основе положения атома в Периодической таблице и

геометрических дескрипторов атомного ПВ. Эта схема проще и более общая, чем другие подобные модели [170, 121], поскольку она обучена на всех доступных и надежных металлоорганических и неорганических структурах, включая ионные соединения, не имеет ограничений на значения степеней окисления и опирается на наименьшее количество признаков. Хотя текущая версия схемы ограничена структурами с кислородной координацией металла, она может быть легко расширена и на другие виды атомного окружения. Эта расширенная схема будет полезна для повышения точности определения кристаллической структуры, особенно если в эксперименте определяются не все атомы. Еще одно важное приложение касается поиска корреляций между геометрическими свойствами кристаллических структур, закодированных в атомных ПВ, и физическими свойствами, которые зависят от степени окисления атомов, например свойств электридов.

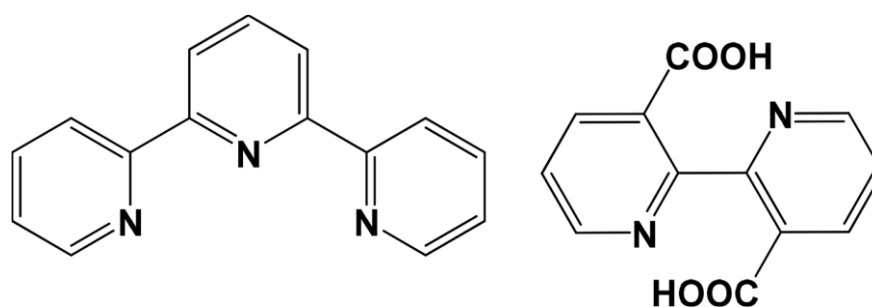
### **3.5. Схема топологической сборки кристаллов**

Исследование механизмов сборки и самосборки кристаллических структур, определение состава и строения структурных единиц, формирующих в итоге определенную кристаллическую структуру, является важной научной задачей современности [104, 139]. Во многом она касается МОКП, которые собираются только при определенных условиях [248], задаваемых параметрами синтеза. Поэтому было интересно использовать геометрико-топологический подход для решения подобной задачи, что и выполнено нами для небольшой выборки из 6 соединений, синтезированных нашими аргентинскими коллегами.

#### **3.5.1. Выборка**

Сольвотермальным способом с использованием лигандов 2,2'-бипиридин-3,3'-дикарбоновой кислоты ( $H_2L$ ) и 2,2':6',2"-терпиридина (ТРУ) (Схема 3) Кристофером Кэхилл и его коллегами из национального

университета Сан-Луиса были синтезированы шесть новых гибридных соединений уранила, которые затем были классифицированы как молекулярные комплексы (связность  $I^0O^0$ ),  $[(UO_2)(L)(TPY)] \cdot H_2O$  (**1**),  $[Ni(TPY)_2][(UO_2)(L)_2] \cdot 3H_2O$  (**2**) и  $[Cu(TPY)_2][(UO_2)(L)_2] \cdot 3H_2O$  (**3**), или МОКП (связность  $I^0O^3$ ),  $[Cu_2(UO_2)_2(OH)(C_2H_3O_2)(L)_3(TPY)_2] \cdot 6H_2O$  (**4**),  $[Zn_2(UO_2)_2(OH)(NO_3)(C_2H_3O_2)(L)_3(TPY)_2] \cdot 4H_2O$  (**5**) и  $Na[Ni(UO_2)_3(OH)(O)(L)_3] \cdot 9H_2O$  (**6**).



**Схема 3.** Органические лиганды  $H_2L$  (справа) и  $TPY$  (слева), использованные в исследовании.

Символами  $I^0$  и  $O^0$  обозначена связанность  $0D$  для органических и неорганических линкеров в молекулярных комплексах, поэтому верхний индекс равен нулю. Строка  $I^0O^3$  обозначает  $3D$  каркас, образованный органическими линкерами  $O^3$ , которыми связаны множественные островные металлические центры  $I^0$ . В проведенном исследовании обсуждено влияния внедрения переходного металла, хелатирующего действия лиганда, и условий синтеза на образование уранильных материалов. В рамках данной работы выполнен топологический анализ обсуждаемых кристаллических структур. Было показано, что особого внимания заслуживает структура соединения **6**, характеризующаяся новой топологией с 9-узловой  $3^7,4,5$ -связной сеткой, далее обозначаемой *geg1*, из-за крупных каналобразных пустот диаметром около  $19.6 \text{ \AA}$ .



### 3.5.2. Топологическое устройство кристаллов

Асимметричная единица структур синтезированных кристаллов содержит два симметрично независимых металлических центра уранила, оба в центрах инверсии, каждый из которых бидентатно координирован к одному симметрично независимому лиганду L (L-1, Рис. 74).

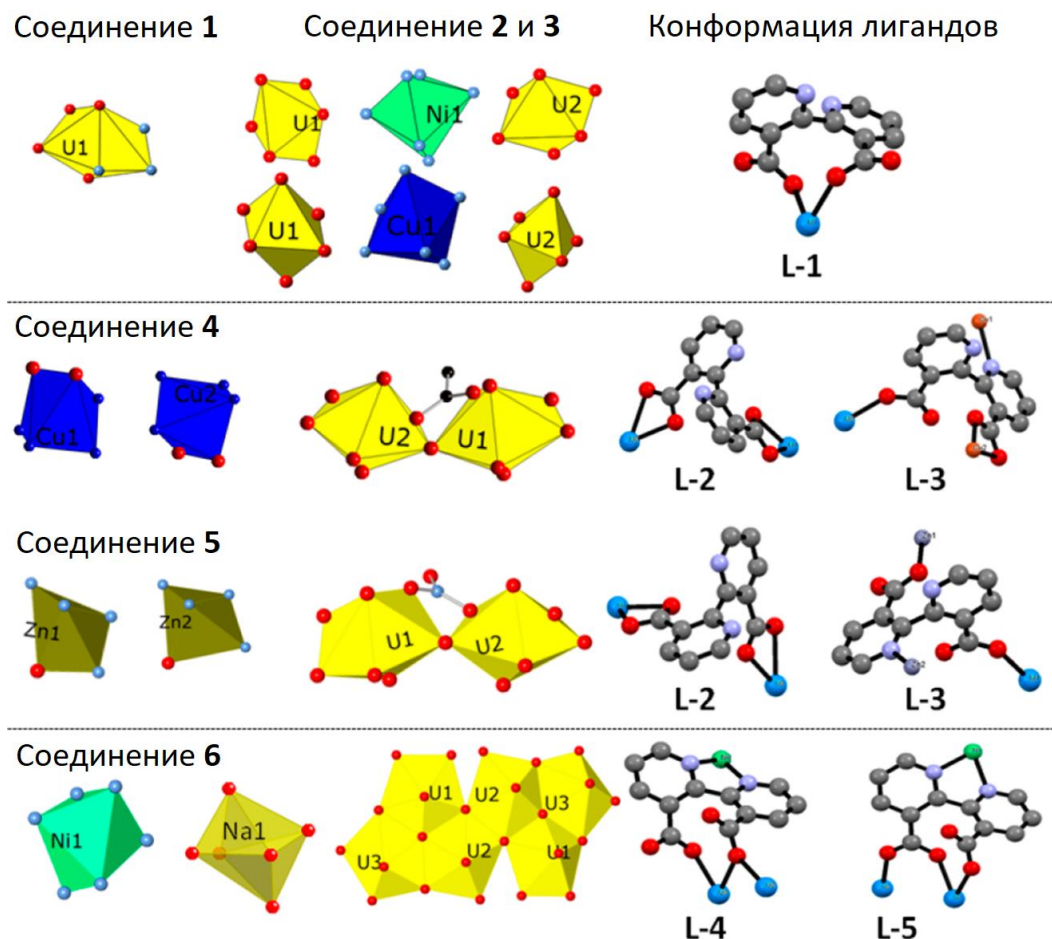


Рис. 74. Краткое описание первичных и вторичных структурных единиц и способов связывания лигандов L в соединениях 1–6. С помощью шариковых моделей показаны конформации лигандов, синие сферы - атомы урана, красные сферы - атомы кислорода, тёмно-серые сферы - атомы углерода, голубые сферы - атомы азота, оранжевые сферы - атомы меди, светло-серые сферы - атомы цинка, а зеленые сферы - атомы никеля.

При проведении топологического анализа с помощью программы ToposPro [215] было обнаружено, что соединения 1, 2, 3 и 4, 5 имеют одинаковую

топологическую структуру соответственно, и что **6** представляет собой новую сетку, обозначенную нами, как *geg1*. Результаты топологии сведены в **Таблицу 19**.

**Таблица 19.** Результаты топологического анализа КС и МОКП на основе актинидов, включая соединения **1–6**

Соединение	Dim	( $\Gamma^n O^m$ )	Сборка	Литература
<b>1</b>	0D	( $\Gamma^0 O^0$ )	1,4M5-1	
<b>2</b>	0D	( $\Gamma^0 O^0$ )	1,4M5-1	
<b>3</b>	0D	( $\Gamma^0 O^0$ )	1,4M5-1	
<b>4</b>	3D	( $\Gamma^0 O^3$ )	<b>ths</b> → <b>tfc</b>	
<b>5</b>	3D	( $\Gamma^0 O^3$ )	<b>ths</b> → <b>tfc</b>	
<b>6</b>	3D	( $\Gamma^0 O^3$ )	<i>geg1</i>	
[UO <sub>2</sub> (L)(H <sub>2</sub> O)]·3H <sub>2</sub> O	1D	( $\Gamma^1 O^0$ )	(4 <sup>2</sup> ,6)	[249, 250]
[UO <sub>2</sub> (L)(DMF)]·0.5H <sub>2</sub> O	1D	( $\Gamma^1 O^0$ )	(4 <sup>2</sup> ,6)	[249, 250]
[UO <sub>2</sub> (L)(H <sub>2</sub> L)]·H <sub>2</sub> O	1D	( $\Gamma^1 O^0$ )	(4 <sup>2</sup> ,6)	[249, 250]
[Ag(UO <sub>2</sub> ) <sub>2</sub> (μ <sub>2</sub> -OH)(L) <sub>2</sub> (H <sub>2</sub> O) <sub>2</sub> ]	3D	( $\Gamma^0 O^3$ )	не указана	[251]
[Ag <sub>3</sub> Th <sub>6</sub> (μ <sub>3</sub> -O)(μ <sub>3</sub> -OH)(μ <sub>2</sub> -OH) <sub>6</sub> (L) <sub>6</sub> (NO <sub>3</sub> ) <sub>6</sub> ·H <sub>2</sub> O]	3D	( $\Gamma^0 O^3$ )	не указана	[251]

По данным кластерного анализа, в структуре соединения **6** присутствуют кластерные фрагменты двух структурных единиц состава U<sub>6</sub>C<sub>2</sub>O<sub>24</sub> и NiC<sub>30</sub>H<sub>17</sub>N<sub>6</sub> (Рис. 75а). Эти фрагменты связаны между собой карбоксилатными группами лиганда L (фрагменты O и CO). Два Ni-содержащих фрагмента связаны с U-содержащим фрагментом тремя координационными связями. Четыре Ni-содержащих фрагмента связаны с U-содержащим фрагментом двумя координационными связями (эти связи показаны на Рис. 75а пунктирной линией). Стягивание этих фрагментов до их геометрических центров тяжести (Рис. 75b) и последующие удаление 0- и 1-координированных узлов и замена 2-координированных узлов ребрами сетки привели к базовой сетке топологического типа *geg1* (Рис. 75с).

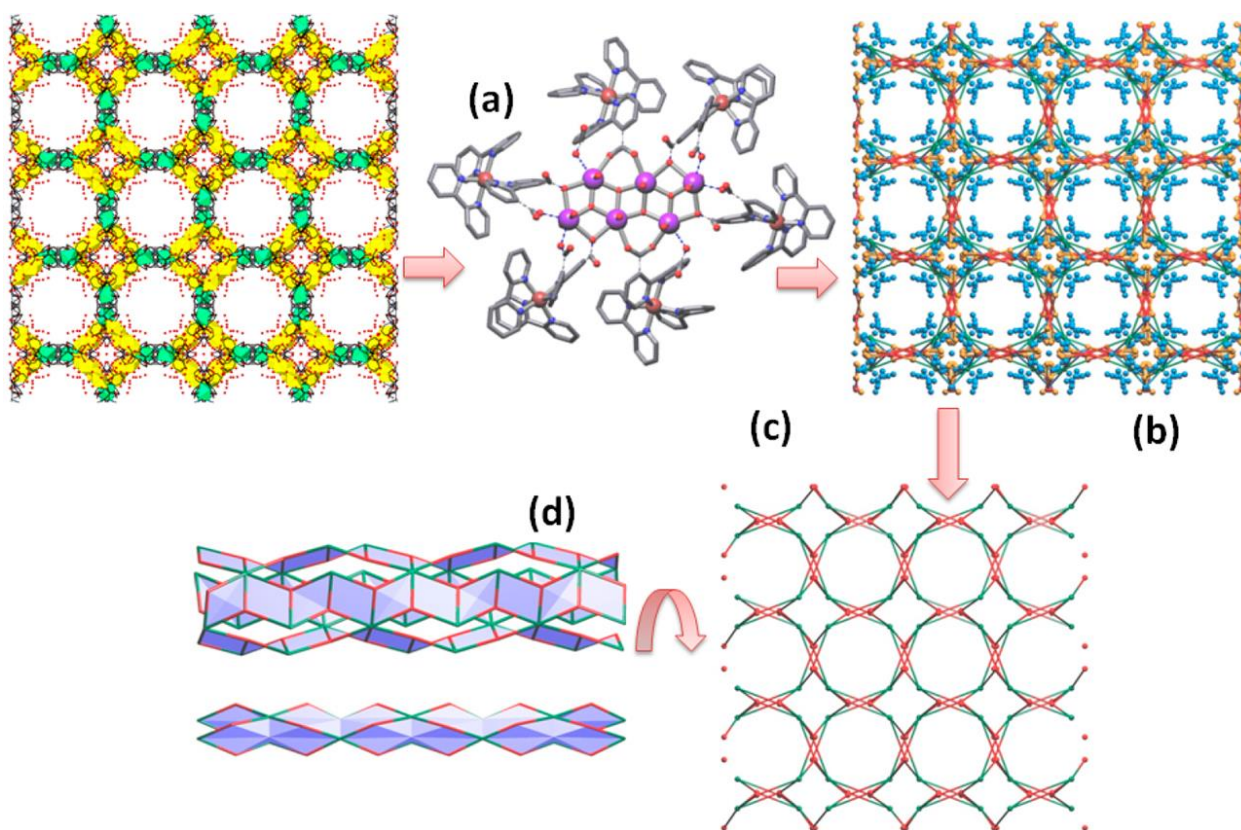


Рис. 75. Кластерные фрагменты  $\text{NiC}_{30}\text{H}_{17}\text{N}_6$  и  $\text{U}_6\text{C}_2\text{O}_{24}$  соединения **6** (a). Атомы водорода не указаны для наглядности. Кластерное представление координационного полимера и сетки, полученное после первичного упрощения (b). Базовая сетка 3,6-с топологического типа *geg1*, полученная после процедуры вторичного упрощения (c). Сетка *geg1* содержит широкие (d, верхний) и узкие (d, нижний) каналы.

Соединение **6** представляет собой первую реализацию типа *geg1* и демонстрирует два отдельных канала, идущих в направлении  $[001]$ . Четыре полосы четырехугольных колец окаймляют широкий канал (Рис. 75d, вверху), тогда как восьмиугольные кольца (Рис. 75d, внизу) образуют узкий канал.

Принимая во внимание различный тип координации лигандов с атомами урана, мы можем собрать несколько фрагментов в один более крупный агрегированный фрагмент или фрагмент «сборки», имеющий состав  $\text{U}_6\text{Ni}_2\text{C}_{64}\text{H}_{34}\text{N}_{12}\text{O}_{28}$  (Рис. 76a). Если такие фрагменты рассматривать как кластеры в процессе упрощения, то мы получим базовую сетку

топологического типа *sqc2077*, состоящую из чередующихся правых и левых спиралей, соединенных между собой (Рис. 76b).

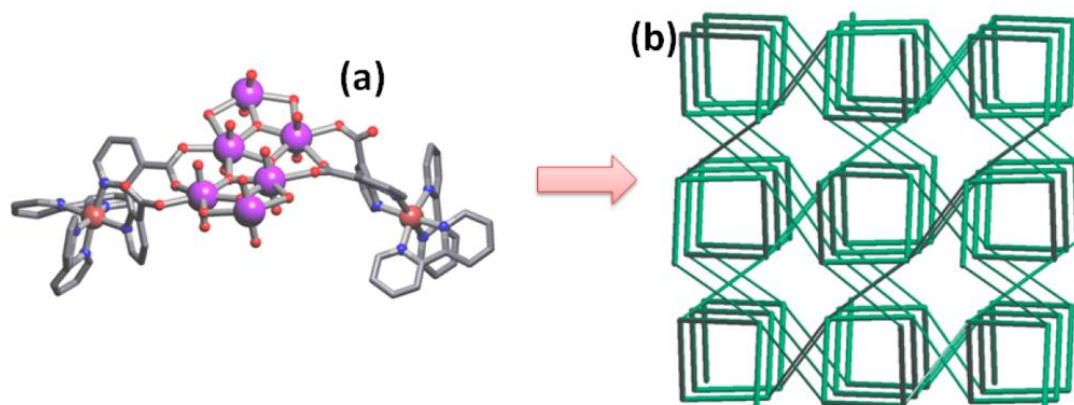


Рис. 76. Фрагмент «сборки» может быть построен из трех фрагментов кластера (a). Атомы водорода не указаны для наглядности. Сетка с топологией *sqc2077* состоит из бесконечных спиралей (b).

Эта сетка встречается редко, но в литературе имеется аналогичный пример в качестве стандартного представления молекулярной структуры, содержащей Н-связь  $[\text{Co}(\text{H}_2\text{PIDC})_2(\text{H}_2\text{O})_2]$  [252], где PIDC представляет собой 2-пропил-5-карбокси-1H-имидазол-4-карбоксилат.

### 3.5.3. Сборка каркасных структур

Линейная природа катиона  $[\text{UO}_2]^{2+}$  и номинально терминальное поведение «ильных» атомов кислорода создают склонность к экваториальной координации и способствуют образованию 1D- и 2D-координационных полимеров [253, 254, 255]. При получении 3D соединений уранила, во многих исследованиях политопные лиганды использовались в качестве стратегии создания архитектур более высокой размерности и новых топологий [256]. Например, Фарха и др. использовали неплоские органические лиганды, такие как псевдочетырехкратный симметричный линкер 4,4',4'',4'''-(пирен-1,3,6,8-тетрил)тетрабензойная кислота ( $\text{H}_4\text{TBApy}$ ) [257] для создания невзаимопроникающих уранилсодержащий МОКП с топологией **tbo** и большими открытыми полостями (17, 24 и 39 Å), а с помощью тритопного

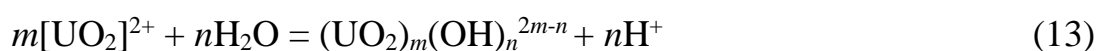
лиганда 5'-(4-карбоксифенил)-2',4',6'-триметил-[1,1':3',1''-терфенил]-4,4''-дикарбоновой кислоты [258] они получили мезопористый U-МОКП, содержащий большие икосододекаэдрические полости с внутренним диаметром 50 и 62 Å. Другой пример, приведенный Ши и др. [259], детализирует сборку актинидов посредством карбоксилатов и ионов  $\text{Ag}^+$  с образованием 3D-соединений  $[\text{Ag}(\text{UO}_2)_2(\mu_2\text{-OH})(\text{L})_2(\text{H}_2\text{O})_2]$  и  $[\text{Ag}_3\text{Th}_6(\mu_3\text{-O})(\mu_3\text{-OH})(\mu_2\text{-OH})_6(\text{L})_6(\text{NO}_3)_6\cdot\text{H}_2\text{O}]$ . При этом атомы азота пиридиновых колец координируются с металлическим центром  $\text{Ag}^+$ , а карбоксилаты координируются с актинидами, обеспечивая дополнительную связь для увеличения размерности каркаса. В настоящем исследовании в качестве политопного лиганда использовалась 2,2'-бипиридин-3,3'-дикарбоновая кислота ( $\text{H}_2\text{L}$ ), поскольку она имеет несколько фрагментов, для которых возможна координация с металлическим центром, карбоксилат-ионами и атомами азота бипиридиновых колец.

Другие исследователи использовали металлические центры с d- или f-блоками для создания каркасов высокой размерности. Ши и его коллеги [251] синтезировали набор биметаллических каркасов с 2,2'-бипиридин-4,4'-дикарбоновой кислотой, охватывающий множество структур, некоторые из которых представляют собой трехмерные каркасы ( $\text{UO}_2\text{-Zn}$ ,  $\text{UO}_2\text{-Cu}$  и  $\text{UO}_2\text{-Co}$ ). Таким образом, наши подходы здесь согласуются с известными подходами, которые привели к созданию 3D-структур, особенно в отношении политипных лигандов и d-металлов.

При соединении катиона  $[\text{UO}_2]^{2+}$  с лигандом  $\text{H}_2\text{L}$  в отсутствие других металлов в работе Тюэри и Маски был получен набор одномерных цепочечных структур [260]:  $[\text{UO}_2(\text{L})(\text{H}_2\text{O})]\cdot 3\text{H}_2\text{O}$ ,  $[\text{UO}_2(\text{L})(\text{DMF})]\cdot 0.5\text{H}_2\text{O}$  и  $[\text{UO}_2(\text{L})(\text{H}_2\text{L})]\cdot \text{H}_2\text{O}$ , что подтверждает склонность  $[\text{UO}_2]^{2+}$  к образованию низкоразмерных соединений. В нашем случае сочетание  $\text{H}_2\text{L}$  и ТРУ с  $[\text{UO}_2]^{2+}$  приводит к 0D молекулярной структуре (1). Добавление  $\text{Cu}^{2+}$  и  $\text{Ni}^{2+}$  в соотношении 0.5:1 по отношению к катиону  $[\text{UO}_2]^{2+}$  при синтезе соединений

**2** и **3** также приводит к образованию 0D-форм за счет хелатирования и инкапсуляции переходных металлов двумя молекулами ТРҮ. Уранильные единицы связаны исключительно лигандами L, что является примером координационных предпочтений уранила и d-металлов в соответствии с предпочтениями теории жестких и мягких кислот и оснований Льюиса, которые, по сути, приводят к «фазовому расслоению» на молекулярном уровне [261]. Когда соотношение  $\text{Cu}^{2+}:[\text{UO}_2]^{2+}$  или  $\text{Zn}^{2+}:[\text{UO}_2]^{2+}$  увеличивается до 1:1, образуются соединения **4** и **5**, и координация  $\text{Cu}^{2+}$  или  $\text{Zn}^{2+}$  с помощью ТРҮ сохраняется, а L-линкеры дополнительно координируют ионы переходного металла и уранила одновременно для создания трехмерных каркасов. Увеличение мольного соотношения  $[\text{UO}_2]^{2+}$  к  $\text{Ni}^{2+}$  до 3:1 приводит к образованию монокристаллов соединения **6**, при этом в кристаллической структуре сохраняется соотношение урана и никеля 3:1.

Гидролиз ионов металлов весьма очевиден в соединениях **4–5** и **6**, где наблюдаются димеры уранила с гидроксидными мостиками и гексамер уранила. Гидролиз катионов  $[\text{UO}_2]^{2+}$  в растворе, представленный в уравнении гидролиза уранила (13), довольно распространен в гидротермальном синтезе и может привести к образованию значительно олигомеризованных единиц за счет образования мостиковых гидроксо- ( $\text{OH}^-$ , гидроксирования) или оксо- ( $\text{O}^{2-}$ , оксоляционные) групп [249, 250].



Согласно уравнению гидролиза иона уранила, на его олигомеризацию в более конденсированные частицы влияют как более высокие концентрации  $[\text{UO}_2]^{2+}$ , так и pH, при этом более крупные звенья преобладают при значениях pH выше 4.5 [262]. pH реакций здесь доводилось примерно до 5, что могло способствовать образованию димеров (**4** и **5**) и гексамеров (**6**). Все соединения, содержащие олигомеризованные формы уранила, также представляли собой трехмерные структуры, что, возможно, позволяет

предположить, что олигомеризация, по крайней мере в этой системе, является предпосылкой для образования МОКП.

### 3.5.4. Сборка молекулярных структур

Различные топологические представления одной и той же структуры позволяют рассматривать разные схемы сборки и выбирать способ, который лучше соответствует существующим теоретическим или экспериментальным данным. Например, кристаллическую структуру 3-метокси-N-(2-метоксифенил)бензол-1-сульфонамида (BOGNAP) [263] можно рассматривать как собранную из молекул или из димеров (Таблица 20), если применить иерархический анализ молекулярной упаковки (см. Раздел 1.3.3).

**Таблица 20.** Размерности (Dim) и топологии базовых скеток для сборки кристаллической структуры 3-метокси-N-(2-метоксифенил)бензол-1-сульфонамида (BOGNAP) из молекул или димеров на разных уровнях интенсивности межмолекулярного взаимодействия, оцениваемого величиной телесного угла молекулярных ПВ ( $\Omega$ ).

Молекула			Димер		
$\Omega$ , %	Dim	Топология	$\Omega$ , %	Dim	Топология
25.94	0D	1M2-1			
10.59	1D	2C1	10.56	1D	2C1
8.84	2D	<b>hcb</b>	10.37	2D	<b>sql</b>
8.50	3D	<b>dia</b>	10.30	3D	<b>pcu</b>
7.06	3D	<b>sxd</b>	6.31	3D	<b>hex</b>
6.40	3D	<b>oob</b>	5.97	3D	<b>bct</b>
4.62	3D	<b>nch</b>	5.74	3D	<b>fcu</b>
3.78	3D	<b>chb</b>	0.75	3D	<b>bcu-x</b>
2.78	3D	<b>bcu-x-12-Cmcm</b>			
2.39	3D	14T2			



Очевидно, что сборка из димеров (Рис. 77) происходит быстрее; более того, на основании уровней интенсивности межмолекулярных взаимодействий, оцениваемой по величине телесных углов молекулярных ПВ (в скобках), можно предположить, что она осуществляется всего в три стадии: димер 1M2-1 (~26%) → **pcu** (~10%) → **fcu** (~6%) → **bcu-x** (~1%). Результирующие топологии молекулярных упаковок (примитивная кубическая, **pcu**; гранецентрированная кубическая, **fcu** и объемно-центрированная кубическая, **bcu-x**) довольно распространены, что соответствует типичной схеме сборки молекулярных кристаллов.

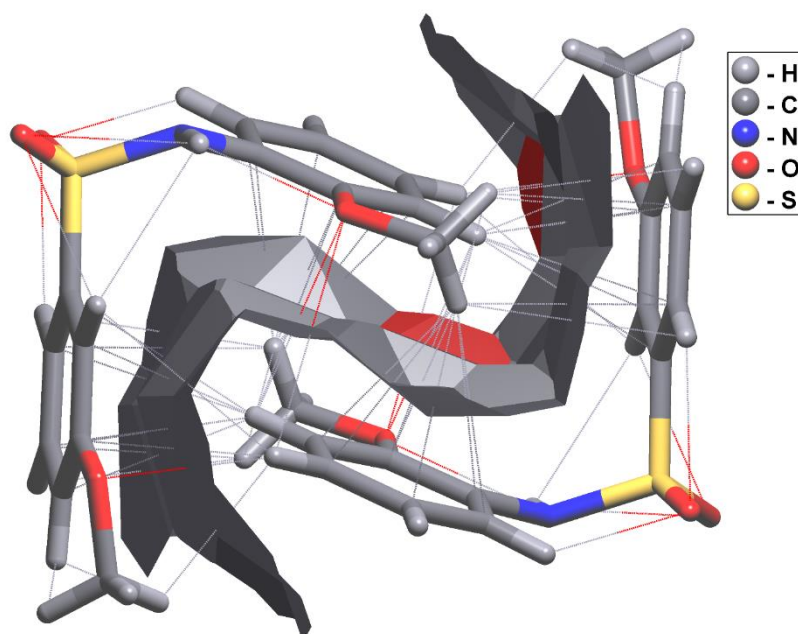


Рис. 77. Молекулярный димер в кристаллической структуре 3-метокси-N-(2-метоксифенил)бензол-1-сульфонамида (BOGNAP). Телесный угол контактной поверхности молекул, состоящий из общих граней молекулярных ПВ, соответствует 25.94% от полного телесного угла всех граней молекулярного ПВ каждой молекулы.



## Глава 4. Геометрический и топологический анализ кристаллических структур с иной природой связывания атомов

### 4.1. Изменение размера атомов инертных газов при фазовых переходах

Поиск взаимосвязей строения вещества с квантово-химическими свойствами составляющих их атомов является важной задачей современности. Универсальность предлагаемого нами геометрико-топологического подхода дает возможность анализировать не только кристаллическое состояние, но и жидкое и даже газообразное, что было продемонстрировано на примере конденсированных состояний инертных газов. Учитывая особенность строения атомов инертных газов, имеющих полностью заполненные все электронные оболочки, представляло интерес сравнить длины волн де Бройля  $\lambda^B_n$  различного порядка (**Таблица 1**) с геометрическими характеристиками атомов инертных газов в твердом, жидком и газообразном состояниях.

При анализе кристаллических структур автор [19] рассматривал только межатомные расстояния. Использование ПВ или атомных доменов [200] расширяет список метрических характеристик кристаллической структуры. В качестве сопоставляемых параметров мы изучили расстояния от ядра центрального атома А до всех особых точек структуры, а именно, до граней  $r(A-f)$ , ребер  $r(A-e)$  и вершин  $r(A-v)$  ПВ (Рис. 78), а также радиусы сферического домена ( $R_{SD}$ ), объем которого равен объему ПВ (**Таблица 21**). Характеристика  $R_{SD}$  оценивает размер атома с точки зрения модели мягких (или деформируемых) сфер [200], допускающей произвольную деформацию атома в кристалле при условии сохранения занимаемого им объема. В отличие от межатомных расстояний, радиус сферического домена (в Å) может быть рассчитан из плотности вещества  $\rho$  (в г/см<sup>3</sup>) по формуле:

$$R_{SD} = \sqrt[3]{\frac{1.245 \cdot M}{\rho \cdot \pi}}, \quad (14)$$

где  $M$  – масса атома  $A$  (в г/моль), что дает возможность оценить размеры атомов не только в кристаллах, но и в жидкостях или газах.

Учитывая, что гелий кристаллизуется лишь при достаточно высоких давлениях, а радон является радиоактивным малоизученным элементом, первоначально рассмотрим метрику атомов неона, аргона, криптона и ксенона (**Таблица 21**). Напомним, что рассматриваемые инертные газы кристаллизуются в ГЦК решетке [264, 265], а их ПВ имеют форму ромбододекаэдра (Рис. 78), все грани и ребра которого одинаково удалены от центра тяжести полиэдра, а расстояния до вершин принимают два значения (**Таблица 21**).

**Таблица 21.** Характеристики ПВ атомов инертных газов ( $A$ ) в кристаллах

Атом $A$	Условия	$R_{SD}, \text{\AA}$	$r(A-f), \text{\AA}$	$r(A-e), \text{\AA}$	$r(A-v), \text{\AA}$	
He	$T=18.76 \text{ K}$ $P=1693 \text{ атм}$	1.66	1.50	1.73	1.84	2.12
Ne	$T=4.2 \text{ K}$	1.73	1.57	1.81	1.92	2.21
Ar	$T=4.2 \text{ K}$	2.05	1.86	2.15	2.28	2.63
Kr	$T=80 \text{ K}$	2.24	2.02	2.34	2.48	2.86
Xe	$T=88 \text{ K}$	2.42	2.19	2.53	2.68	3.10

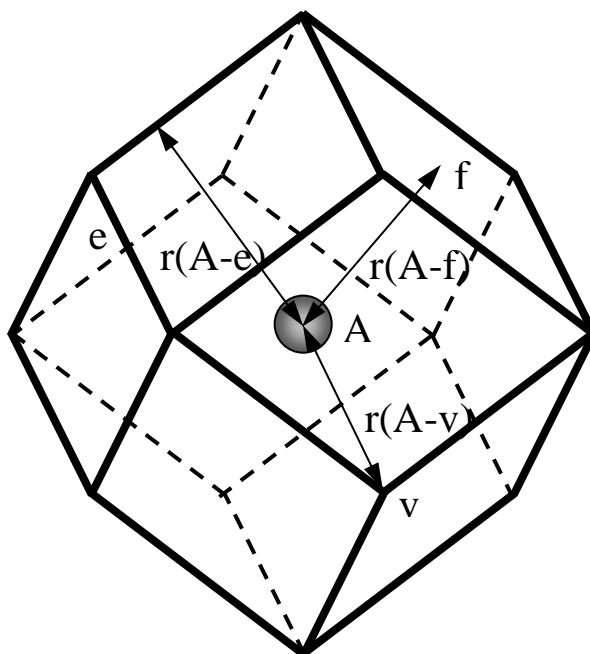


Рис. 78. ПВ атома  $A$  в ГЦК решетке. На рисунке помечены расстояния от атома  $A$  до граней  $r(A-f)$ , ребер  $r(A-e)$  и вершин  $r(A-v)$  ПВ.

Меньшее из них отвечает вершинам, занимающим в ГЦК структуре тетраэдрические пустоты, большее – октаэдрические пустоты.

Сравнение параметров ПВ атомов инертных газов в кристаллах простых веществ с величинами  $\lambda_B^n$  их электронов показало наличие корреляций между этими характеристиками. Как видно из Рис. 79, расстояния  $r(A-f)$  близки по величине к соответствующим  $\lambda_B^{III}$ , а минимальные значения  $r(A-v)$  практически совпадают с  $\lambda_B^{II}$ .

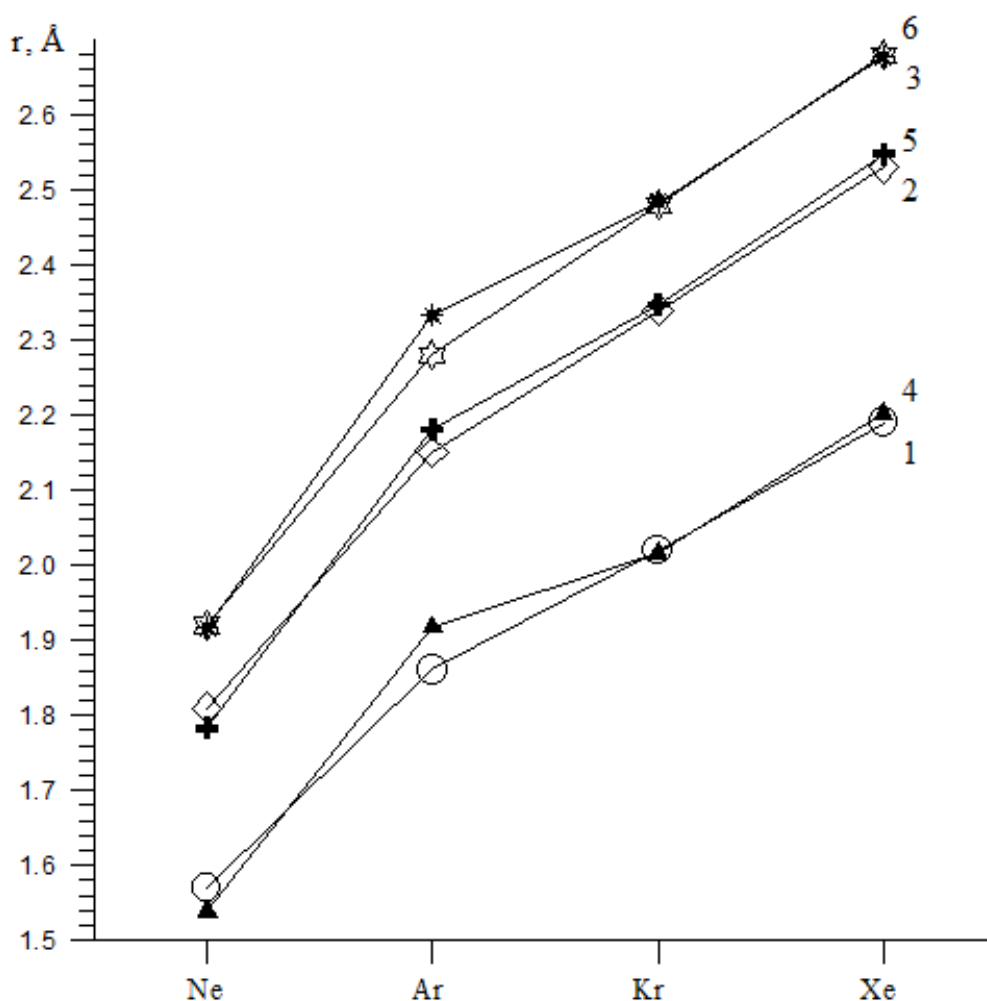


Рис. 79. Зависимости расстояний  $r(A-f)$  (1),  $r(A-e)$  (2) и  $r(A-v)$  (3) в ПВ атомов А и длин волн де Бройля  $\lambda_B^{III}$  (4),  $\lambda_B'^{II}$  (5) и  $\lambda_B^{II}$  (6) от химической природы атома А.

Расстояния до ребер ПВ  $r(A-e)$  в пределах погрешности эксперимента равны длине электронных дебройлевских волн  $\lambda_B'^{II}$  в атомах щелочного металла, следующего за соответствующим инертным газом. Очевидно, что

обнаруженные совпадения величин не могут быть случайными, поэтому следовало ожидать наличие подобных корреляций и в других фазовых состояниях этих элементов.

Дополнительный анализ показал, что длины волн де Бройля взаимосвязаны с метрическими характеристиками атомов инертных газов не только в твердом, но и в жидком или газообразном состояниях. Так половина равновесного межъядерного расстояния  $r(A-A)$  в димерах  $A_2$  (Таблица 22, [23, стр.849]), величина которого рассчитана из спектральных данных, как и соответствующие ван-дер-Ваальсовы радиусы атомов  $A$  ( $R_w$ ), близка по величине к  $\lambda_B^{III}$  (Рис. 80). Это может быть использовано для оценки величины ван-дер-Ваальсовых радиусов атомов инертных газов исходя из значений  $\lambda_B^{III}$  и дает возможность с энергетических позиций трактовать их природу.

**Таблица 22.** Равновесные межъядерные расстояния  $r(A-A)$  в димерах  $A_2$  и ван-дер-Ваальсовы радиусы атомов инертных газов

Атом $A$	$r(A-A)$ , Å [23]	ван-дер-Ваальсовы радиусы $R_w$ , Å		
		[266]	[267]	[19]
He	—	1.22	1.485	1.40
Ne	1.55	1.60	1.54	1.54
Ar	1.88	1.92	1.88	1.88
Kr	2.00	1.98	2.00	2.02
Xe	2.18	2.18	2.18	2.16
Rn	—	2.14	—	—

Как отмечалось выше, размеры атомов  $A$  в жидкости можно охарактеризовать радиусом сферического домена, величина которого рассчитывается из ее плотности (Таблица 23) по формуле (14). Особыми точками на кривой равновесия жидкость-пар являются тройная и критическая. Именно в этих точках можно ожидать корреляцию величины радиуса сферического домена (Таблица 24) с длинами волн де Бройля. Учитывая, что значение радиуса сферического домена атомов  $A$  в жидком состоянии  $R_{SD}^L$  в тройной точке ( $T_t$ ) отличается не более чем на 0.03 Å от  $R_{SD}^L$  в точке кипения ( $T_b$ ) при атмосферном давлении, а также то, что

плотность при  $T_b$  известна также для гелия и радона (Таблица 23), используем для сравнения с  $\lambda_B^n$  значения  $R_{SD}^L$  при  $T_b$ .

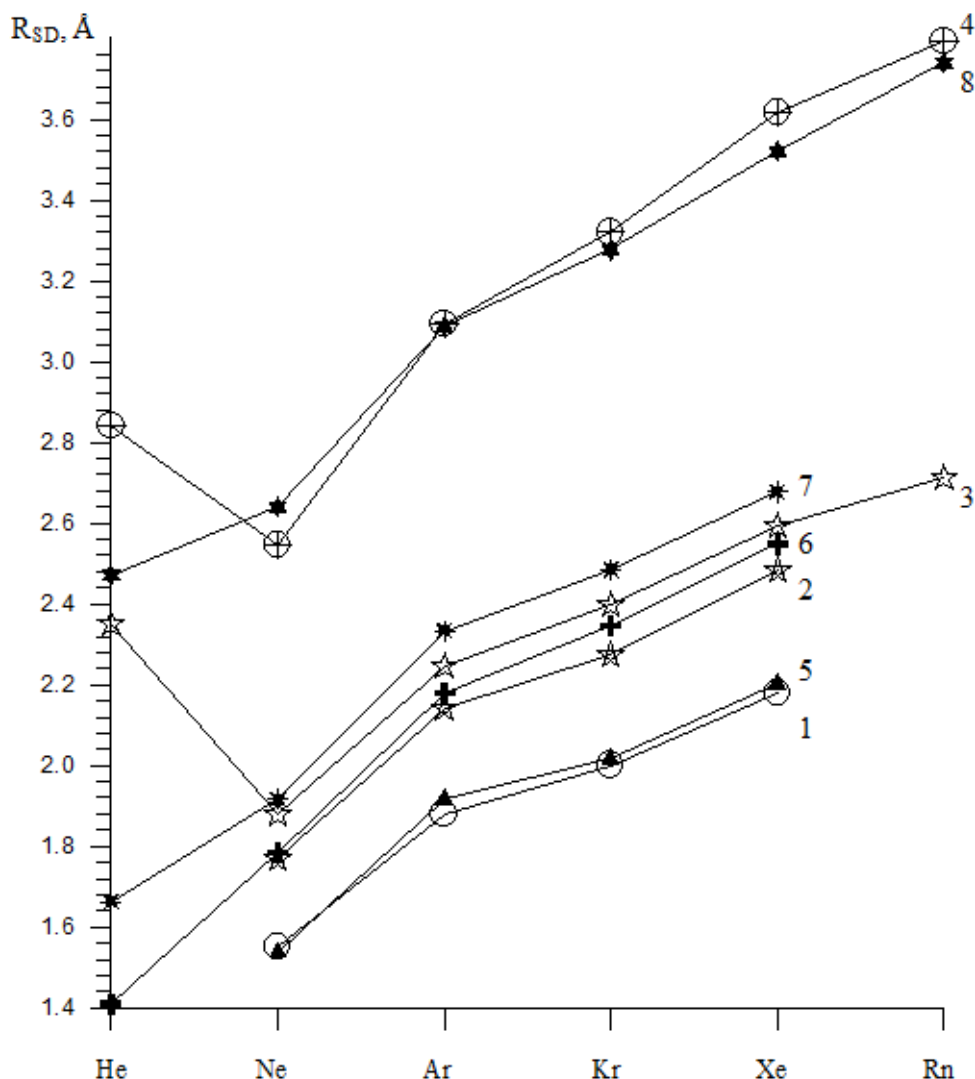


Рис. 80. Зависимости половины межъядерных расстояний  $r(A-A)$  в димерах  $A_2$  (1), величин  $R_{SD}$  атомов  $A$  при  $T_t$  (2) в кристаллах, при  $T_b$  (3) и  $T_c$  (4) в жидкостях, а также длин волн де Бройля  $\lambda_B^{III}$  (5),  $\lambda_B^{II}$  (6),  $\lambda_B^{II}$  (7) и  $\lambda_B^I$  (8) от химической природы атома  $A$ .

Из Рис. 80 хорошо видно, что характеры зависимостей величин  $R_{SD}^L$ ,  $R_{SD}^S$ ,  $\lambda_B^{II}$  и  $\lambda_B^{II}$  от химической природы атома  $A$  подобны друг другу. Причем различия  $R_{SD}^S$  и  $\lambda_B^{II}$  или  $R_{SD}^L$  и  $\lambda_B^{II}$  сравнительно невелики (меньше  $0.1 \text{ Å}$ ), а их разности растут с увеличением порядкового номера атома  $A$  (Рис. 80). Радиус сферического домена атомов  $A$  в критической точке  $R_{SD}^C$  отличается

не более чем на  $0.1\text{\AA}$  от  $\lambda_B^I$  (Рис. 80). Если учесть тот факт, что точность определения плотности вещества, находящегося при критических условиях, невысока, то можно предположить равенство величин  $\lambda_B^I$  и  $R_{SD}^C$  друг другу. Существенно, что близость значений  $R_{SD}^L$  и  $\lambda_B^{II}$ ,  $R_{SD}^C$  и  $\lambda_B^I$  указывает на возможность использования для описания поведения электронных оболочек модели мягких сфер, то есть при некоторых условиях поведение электронных оболочек можно моделировать несжимаемой “жидкостью”, заполняющей все доступное ей пространство.

**Таблица 23.** Некоторые физические характеристики инертных газов [23, 266, 268]\*

Атом А	$\rho_0$ при $T=0\text{ К}$ , $\frac{\text{г}}{\text{см}^3}$	Тройная точка			Точка кипения		Критическая точка	
		$T_t, \text{ К}$	$\rho_t^S, \frac{\text{г}}{\text{см}^3}$	$\rho_t^L, \frac{\text{г}}{\text{см}^3}$	$T_b, \text{ К}$	$\rho_b, \frac{\text{г}}{\text{см}^3}$	$T_c, \text{ К}$	$\rho_c, \frac{\text{г}}{\text{см}^3}$
$^3\text{He}$	—	—	—	—	3.19	0.08235	3.35	0.0413
$^4\text{He}$	0.1484**	—	—	—	4.22	0.1221	5.25	0.0693
Ne	—	24.66	1.444	1.249	27.2	1.205	44.4	0.484
Ar	—	83.81	1.623	1.416	87.3	1.3998	150.85	0.536
Kr	3.100	115.78	2.826	2.451	119.8	2.413	209.35	0.908
Xe	—	161.36	3.399	3.057	164.2	2.987	289.74	1.099
Rn	—	202.3	5.7	—	211.2	4.4	377.5	1.613

\*В таблице использованы следующие обозначения:  $T_t$  – температура тройной точки;  $T_b$  – температура кипения жидкости при атмосферном давлении;  $T_c$  – критическая температура;  $\rho_0$  и  $\rho_t^S$  – плотности кристаллов при абсолютном нуле температур и в тройной точке;  $\rho_t^L$ ,  $\rho_b$  и  $\rho_c$  – плотности жидкости при  $T_t$ ,  $T_b$  и  $T_c$  соответственно.

\*\*Жидкая фаза.

В случае твердого гелия, являющегося квантовым кристаллом, ситуация особая. По известным нам данным, при высоких давлениях он образует три кристаллические модификации: ГЦК при температуре 18.76 К и давлении 1693 атм [269], ГПУ при 3.96 К и 130 атм [270], а также ОЦК при 1.73 К и 29 атм [271]. Обратим внимание, что минимальное из возможных

значений  $R_{SD}$  (**Таблица 21**) атомов гелия идентично длине его электронных волн де Бройля второго порядка  $\lambda_B^{II}$ . Интересно также отметить тот факт, что отвечающие различным модификациям гелия значения  $R_{SD}$  (1.66, 1.89 и 2.02 Å) близки по величине к  $r(A-f)$  в кристаллах Ne (1.57 Å), Ar (1.86 Å) и Kr (2.02 Å) соответственно. Учитывая это, а также то, что при абсолютном нуле температур  $R_{SD}$  атомов в жидком гелии (2.20 Å) практически совпадает с  $r(Xe-f)$  (2.19 Å), можно выдвинуть гипотезу о влиянии различных возбужденных электронных состояний на размеры атомов гелия.

**Таблица 24.** Вычисленные по данным **Таблицы 23** радиусы сферических доменов атомов инертных газов при разных условиях

Атом А	Кристаллы		Жидкость		$R_{SD}^C$ при $T=T_c$ , Å
	$R_{SD}^S$ при $T$ [264, 272], Å	$R_{SD}^S$ при $T=T_t$ , Å	$R_{SD}^L$ при $T=T_t$ , Å	$R_{SD}^L$ при $T=T_b$ , Å	
$^3\text{He}$	—	—	—	2.44	3.07
$^4\text{He}$	2.20 (0 K)	—	—	2.35	2.84
Ne	1.73 (4.2 K)	1.77	1.86	1.88	2.55
Ar	2.05 (4.2 K)	2.14	2.24	2.25	3.09
Kr	2.20 (0 K)	2.27	2.38	2.40	3.32
Xe	2.39 (20 K)	2.48	2.57	2.59	3.62
Rn	—	2.49	—	2.71	3.79

Многозначность величин  $R_{SD}$  атомов в твердом гелии позволяет объяснить заметное различие ван-дер-Ваальсовых радиусов атомов легких элементов первого и второго периода. На наш взгляд, причина тому кроется в определяющем влиянии на размер атомов в кристаллах тех или иных электронных оболочек (или состояний), имеющих конкретную геометрию. Однако обнаружить ее проявление в кристаллических структурах достаточно сложно. Во-первых, в зависимости от интенсивности внешних воздействий на атом, электронная оболочка может вести себя как мягкая или как жесткая сфера, что отражается в разнообразии однотипных межатомных расстояний. Во-вторых, геометрию структуры одновременно могут определять сразу

несколько электронных оболочек, как, например, в кристаллах инертных газов. И, в-третьих, проведение анализа существенно усложняет влияние температурного фактора.

В заключение рассмотрим зависимости  $R_{SD}(T)$  для равновесных состояний инертных газов (Рис. 81), которые построены на основании данных [23, 266, 268, 273]. Эти функции сверху ограничены критической точкой, а минимальное их значение отвечает  $R_{SD}$  атомов А при абсолютном нуле температур (для твердых инертных газов учтены только данные [264, 265, 272], полученные при атмосферном давлении).

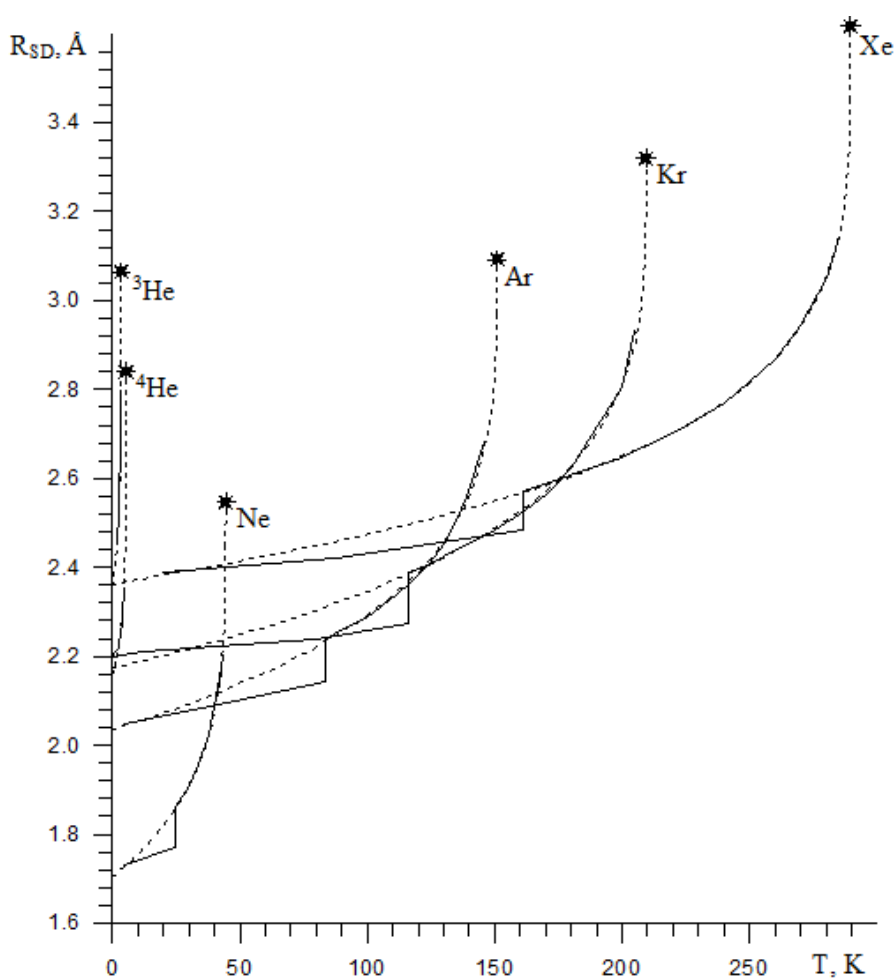


Рис. 81. В виде сплошных линий показаны экспериментальные зависимости  $R_{SD}(T)$  для жидкого и твердого состояний инертных газов. Штриховыми линиями представлены рассчитанные на основании уравнения (15) методом наименьших квадратов графики функций  $R_{SD}(T)$  для жидкого состояния. Коэффициенты В и С для каждой из зависимостей приведены в **Таблице 25**.



**Таблица 25.** Характеристики зависимостей  $R_{SD}(T)$  для жидких инертных газов

Атом А	Коэффициенты в (15)		$R_{SD}$ при $T=0 \text{ К}^*, \text{ \AA}$	Ошибка	
	В	С		$\sigma(\rho), \%$	$\sigma(R_{SD}), \%$
$^3\text{He}$	1.225	-0.0355	2.36	0.27	0.09
$^4\text{He}$	1.463	-0.189	2.16	0.47	0.16
Ne	1.700	0.640	1.70	0.32	0.10
Ar	1.690	0.816	2.04	0.63	0.21
Kr	1.741	0.825	2.17	0.48	0.16
Xe	2.023	0.563	2.36	0.23	0.08

\* Теоретически рассчитанные данные.

В тройной точке наблюдается фазовый переход, отвечающий кристаллизации вещества и приводящий к скачкообразному изменению величины  $R_{SD}$ . Согласно [268] жидкостная ветвь ликвидуса хорошо описывается функцией

$$\frac{\rho}{\rho_c} - 1 = C \left( 1 - \frac{T}{T_c} \right) + B \left( 1 - \frac{T}{T_c} \right)^{\frac{1}{3}}, \quad (15)$$

где  $T_c$  и  $\rho_c$  – критическая температура и плотность жидкости при  $T_c$  соответственно, а  $C$  и  $B$  – константы, зависящие от природы вещества. Другие, подобные (15), функции  $\rho(T)$ , используемые в настоящее время для жидкостей, как правило, более громоздки (например, [274]). Из уравнений (14) и (15) легко получить зависимости  $R_{SD}(T)$ . В **Таблице 25** для всех инертных газов приведены константы  $C$  и  $B$ , подобранные методом наименьших квадратов. Из Рис. 81 и данных **Таблицы 25** хорошо видно, что для всех инертных газов экспериментальные зависимости полностью совпадают с рассчитанными. В **Таблице 25** значения  $\sigma(\rho)$  определяли, как и авторы [274], по формуле:

$$\sigma(\rho) = \frac{100}{n} \sum_i \frac{|\rho_{cal,i} - \rho_{exp,i}|}{\rho_{exp,i}}, \quad (16)$$

где  $n$  – число экспериментальных точек, а  $\rho_{cal}$  и  $\rho_{exp}$  – соответственно теоретически и экспериментально определенные плотности жидкости. Величины  $\sigma(R_{SD})$  рассчитывали по аналогичной (16) формуле.

Обратим внимание, что значения, полученные при экстраполяции обсуждаемых зависимостей к  $T=0$  (без изменения агрегатного состояния жидкости), во всех случаях близки по величине к  $R_{SD}$  атомов инертного газа в кристаллах при абсолютном нуле температур (Рис. 81, **Таблицы 24 и 25**). Эта особенность неоднократно отмечалась и ранее (например, [268]), однако до сих пор ей не найдено удовлетворительного объяснения.

#### **4.2. Динамика изменения атомных дескрипторов и Периодический закон**

Как было показано в Разделе 4.1, имеются взаимосвязи размеров атомов инертных газов в твердом, жидком и газообразном состояниях с длинами волн де Бройля. Представляло интерес выяснить, существуют ли подобные корреляции для других элементов Периодической системы Д.И. Менделеева, поэтому была изучена геометрия известных кристаллических структур элементов второго периода (A), данные по которым взяты из неорганической базы данных [2]. Из рассмотрения были исключены структуры, определенные при высоких давлениях и имеющие разупорядочение атомов.

В результате была создана база данных, которая содержит 723 структурных определений и 314 уникальных кристаллических структур. Как и в Разделе 4.1, атомы элементов второго периода рассмотрены с позиций геометрико-топологического подхода. Помимо кратчайших межатомных расстояний  $r(A-A)$  для атомов A рассчитаны объем ПВ ( $V_{VP}$ ), радиус сферического домена ( $R_{SD}$ ), объем которой равен  $V_{VP}$ , безразмерная характеристика степени сферичности ( $G_3$ ), а также смещение атома из центра тяжести его ПВ ( $D_A$ ). Полученные данные представлены в **Таблице 26**.

Характер их изменения вдоль периода хорошо отражает особенности электронного строения кристаллов. В частности, ненулевая величина  $D_A$  свидетельствует об асимметрии распределения электронной плотности около атома А [275], а превосходящее 0.082 значение второго момента инерции ПВ  $G_3$  подчеркивает ковалентную природу связей, образуемых атомами неметалла [276].

**Таблица 26.** Геометрические характеристики атомов в кристаллах простых веществ

Атом А	Число атомов	$V_{\text{ПВД}}, \text{\AA}^3$	$R_{\text{SD}}, \text{\AA}$	$R_{\text{AT}}, *$ $\text{\AA}$	$r(\text{A-A}), *$ $\text{\AA}$	$G_3$	$D_A, \text{\AA}$	Тип решетки, условия **
Li	1	21.6	1.73	1.52	3.04	0.0785	0	ОЦК, T=298 К
	1	21.3	1.72	1.56	3.11	0.0787	0	ГПУ, T=78 К
Be	2	8.1	1.25	1.12	2.23	0.0788	0	ГПУ, T=293 К
	1	8.3	1.26	1.11	2.21	0.0785	0	ОЦК, T=1528 К
B	4	7.3	1.20	0.83	1.66	0.0965(2)	0.23(3)	$\alpha$ -B, SG: R-3m
	10	7.7(5)	1.22(3)	0.80	1.60	0.099(4)	0.2(1)	SG: P4 <sub>2</sub> /nnm
	7	7.7(7)	1.23(4)	0.83	1.66	0.099(3)	0.2(1)	SG: P-4n2
	15	7.5(11)	1.21(6)	0.80	1.60	0.099(6)	0.2(1)	$\beta$ -B, SG: R-3m
C	6	5.7	1.11	0.77	1.54	0.0912(1)	0	алмаз
	3	8.7(2)	1.28	0.71	1.42	0.111(1)	0	графит, T=293 К
	13	12(2)	1.41(7)	0.71	1.41(4)	0.14(1)	0.5(2)	фуллерен C <sub>60</sub>
N	3	22.5(3)	1.75	0.53	1.06	0.0961(1)	0.47(4)	SG: P2 <sub>1</sub> 3, T=20 и 25 К
O	2	17.6(3)	1.61	0.58	1.16	0.0918(1)	0.37(2)	SG: C2/m, T=23 К
F	1	16.4	1.58	0.78	1.56	0.0858	0.21	SG: C2/c,
Ne	2	22.4(9)	1.75(3)	1.58	3.16	0.0787	0	ГЦК, T $\approx$ 4.3 К

\* $r(\text{A-A})$  - кратчайшее межатомное расстояние в кристаллах,  $R_{\text{AT}}=r(\text{A-A})/2$ .

\*\*SG – пространственная группа симметрии.

Характер зависимости кратчайших межатомных расстояний  $r(\text{A-A})$  или атомных радиусов ( $R_{\text{AT}}$ ) в кристаллах простых веществ от порядкового номера  $n$  (Таблица 26, Рис. 82) хорошо изучен. Известно и то, что для некоторых элементов значения  $R_{\text{AT}}$  близки по величине к орбитальным радиусам Вебера-Кромера ( $R_{\text{WC}}$ ) валентных  $2s^{1/2}$  или  $2p^{1/2}$  оболочек [40] (Таблица 27).

**Таблица 27.** Радиусы основных максимумов орбиталей атомов [22]

Элемент	$1s^{1/2}$	$2s^{1/2}$	$2p^{1/2}$	$2p^{3/2}$
H	0.529	—	—	—
He	0.291	—	—	—
Li	0.186	1.586	—	—
Be	0.138	1.040	—	—
B	0.110	0.769	0.776	—
C	0.091	0.620	0.596	—
N	0.078	0.512	0.487	0.488
O	0.068	0.450	0.413	0.414
F	0.060	0.396	0.359	0.360
Ne	0.054	0.354	0.317	0.318

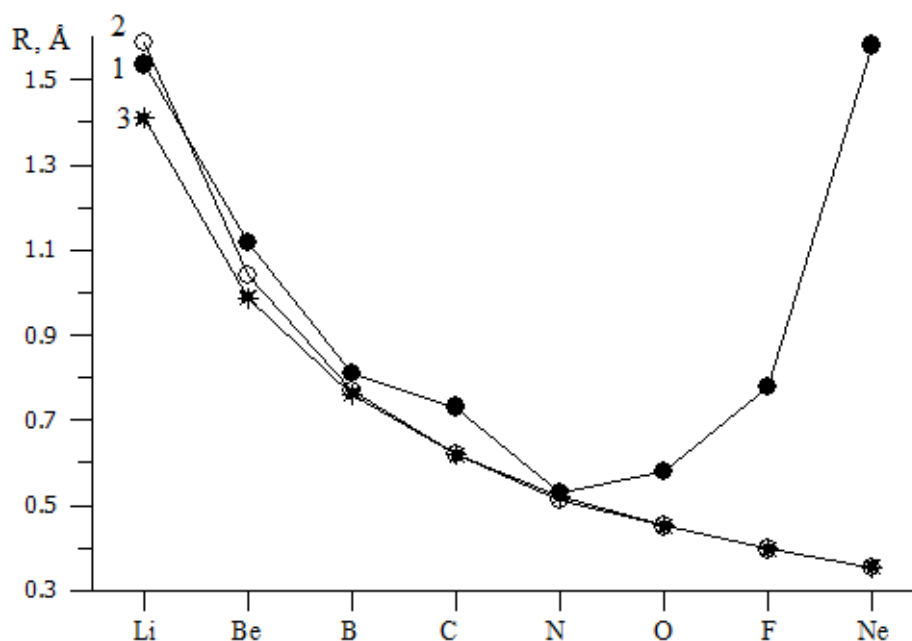


Рис. 82. Зависимости величин  $R_{AT}$ (1),  $R_{WC}$ (2) и  $\lambda_B^{n-1}$ (3) от химической природы атомов А.

Однако нами не обнаружено в литературе упоминаний о корреляции орбитальных радиусов с  $\lambda_B^{n-1}$ , где  $n$  – атомный номер элемента (Рис. 82, Таблица 28). Возможно, что именно близость значений  $\lambda_B^{n-1}$  и  $R_{AT}$  может быть причиной наблюдаемых длин кратчайших межатомных контактов. Как видно из Рис. 82, атомный радиус существенно больше длины дебройлевских волн  $\lambda_B^{n-1}$  только в кристаллах  $O_2$ ,  $F_2$  и Ne.

**Таблица 28.** Рассчитанные по формуле (1) с использованием данных [23] длины волн де Бройля электронов в изолированных атомах

Атом А	$\lambda_B^I$	$\lambda_B^{II}$	$\lambda_B^{III}$	$\lambda_B^{IV}$	$\lambda_B^V$	$\lambda_B^{VI}$	$\lambda_B^{VII}$	$\lambda_B^{VIII}$	$\lambda_B^{IX}$	$\lambda_B^X$
H	3.325	—	—	—	—	—	—	—	—	—
He	2.472	1.662	—	—	—	—	—	—	—	—
Li	5.280	1.410	1.108	—	—	—	—	—	—	—
Be	4.015	2.873	0.988	0.831	—	—	—	—	—	—
B	4.256	2.444	1.991	0.761	0.665	—	—	—	—	—
C	3.654	2.483	1.772	1.527	0.619	0.554	—	—	—	—
N	3.216	2.253	1.780	1.393	1.239	0.522	0.475	—	—	—
O	3.322	2.069	1.654	1.393	1.149	1.043	0.451	0.415	—	—
F	2.937	2.073	1.548	1.313	1.147	0.978	0.901	0.397	0.369	—
Ne	2.640	1.916	1.539	1.244	1.091	0.976	0.852	0.793	0.355	0.332

В Разделе 4.1 отмечена близость значений ван-дер-Ваальсового радиуса атома Ne и величины его  $\lambda_B^{III}$ . Учитывая, что основной максимум радиального распределения  $2s^{-1/2}$  оболочки электронов более удален от ядра, чем у  $2p^{-1/2}$  оболочки (**Таблица 27**), можно предположить, что длина волны  $\lambda_B^{III}$  характеризует радиус р-оболочки. В связи с этим отметим, что  $\lambda_B^{III}$  определяет среднюю длину невалентных контактов  $r(N-N)$ ,  $r(O-O)$  и  $r(F-F)$  (**Таблица 29**), так как для всех элементов от азота до неона схож характер зависимости  $\lambda_B^{III}$  или половины длины невалентных контактов  $r(A-A)$  от  $n$  (Рис. 83). На существенную роль  $\lambda_B^{III}$  в молекулярных кристаллах азота, кислорода и фтора указывает и анализ зависимостей  $R_{SD}$  этих атомов от  $n$ .

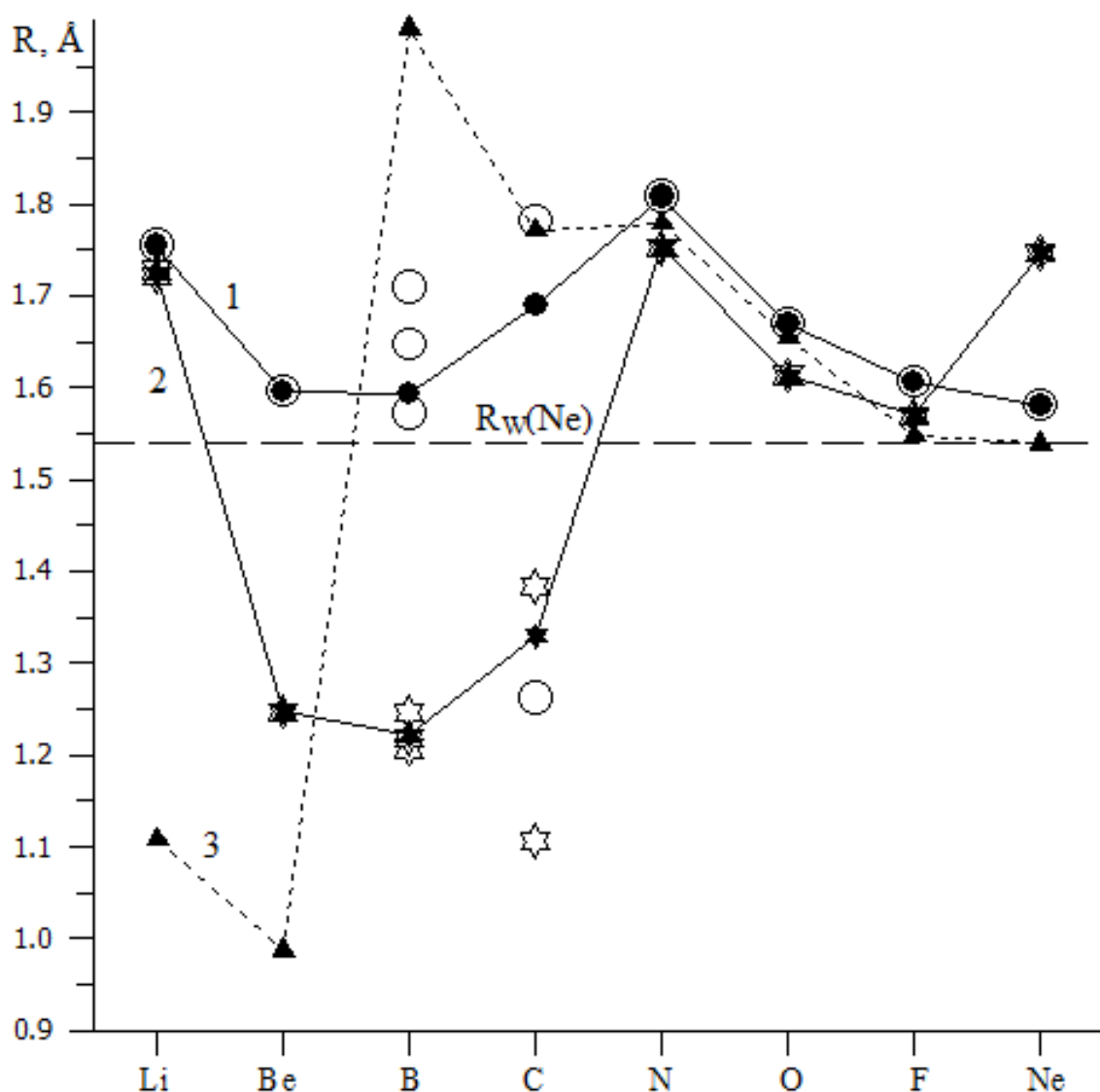


Рис. 83. Зависимости средних значений половины длины невалентных контактов  $r(A-A)/2$  (1),  $R_{SD}(A)$  (2) и  $\lambda_B^{III}$  (3) от химической природы атома А. Незакрашенными символами представлены средние значения соответствующих по форме знаков величин при разных координационных числах атома А.

Из Рис. 83 хорошо видно подобие изменения  $\lambda_B^{III}$  и радиусов сферического домена атомов N, O и F. Как уже отмечалось в [277], значения  $R_{SD}$  этих атомов в кристаллах простых веществ и химических соединений приблизительно постоянны и близки по величине к ван-дер-Ваальсовому радиусу Ne (1.54 Å). Эти факты можно объяснить наличием электронного октета как у Ne, так и у атомов А, причем в последнем случае он сформирован в результате образования химических связей.

**Таблица 29.** Длина и телесные углы контактов, соответствующих граням ПВ атомов азота, кислорода и фтора в кристаллах простых веществ

Структура	$\alpha\text{-N}_2$			$\alpha\text{-N}_2$			$\alpha\text{-O}_2$			$\beta\text{-O}_2$			$\alpha\text{-F}_2$		
Код*	{28179}			{27249}			{18311}			{43430}			{16262}		
Прост. группа	Pa-3			P2 <sub>1</sub> 3			C2/m			R-3m			C2/c		
2R <sub>SD</sub> , Å	3.50			3.50			3.22			3.24			3.16		
r, Å**	3.61(3)			3.6(1)			3.3(1)			3.27(6)			3.2(3)		
Характеристики контактов А-А***															
	m	r, Å	Ω, %	m	r, Å	Ω, %	m	r, Å	Ω, %	m	r, Å	Ω, %	m	r, Å	Ω, %
	1	1.08	35.9	1	1.05	36.2	1	1.15	33.8	1	1.21	33.6	1	1.57	27.5
	6	3.58	5.9	3	3.46	7.2	2	3.18	9.3	3	3.18	9.0	2	2.76	12.2
	6	3.63	4.8	6	3.57	4.7	4	3.20	7.0	6	3.31	6.6	1	3.02	6.3
				3	3.74	4.6	1	3.32	6.9				2	3.20	6.5
							2	3.33	1.1				2	3.20	4.7
							2	3.43	5.1				1	3.21	4.2
							1	4.24	0.4				1	3.22	5.79
				1	3.47	0.8									
				1	3.82	1.1									

\*Код соединения в [2].

\*\*Средняя длина невалентных контактов.

\*\*\*Для каждой группы симметрично равных контактов А-А, образующих грани в полиэдре ВД атома А, указаны их число (m), длина (r) и отвечающий им телесный угол (Ω). Телесные углы выражены в процентах от полного угла 4π ср.

Учитывая соотношение  $R_{SD}(A) \geq R_W(Ne)$ , выполняющееся при условии, что 2р-оболочка изолированного атома А заполнена не менее чем на половину, можно утверждать, что в молекулах  $A_2$  поведение химически сформированного октета электронов описывается моделью мягких деформирующихся сфер, объем которых близок по величине к объему внешней оболочки инертного газа с той же электронной конфигурацией.

Если высказанное предположение верно, то и для атомов водорода в кристаллах  $H_2$  должно выполняться соотношение  $R_{SD}(H) \approx R_W(He)$ . Однако оно не соблюдается, так как эти величины соответственно равны 1.66 Å (рассчитано по данным [278]) и 1.50 Å, что, казалось бы, опровергает

высказанную гипотезу. Однако, для атомов Ne величину  $R_W$  определяет длина волны  $\lambda_B^{III}$ , характеризующая, как предположено выше, размер р-оболочки. Так как у атома гелия таковая отсутствует, то следует ожидать, что в отличие от других инертных газов этот атом будет вести себя не как жесткая, а как мягкая сжимаемая сфера. Примечательно, что радиус  $R_{SD}$  атомов в кристаллическом гелии при температуре 16 К и давлении 1250 бар равен 1.66 Å и тождественен  $R_{SD}$  атомов в молекулярном водороде. Этот факт еще раз подтверждает гипотезу постоянства размера полностью заполненной электронной оболочки.

В заключение остановимся еще на одной интересной особенности, обнаруженной с использованием геометрико-топологического подхода. Как видно из **Таблиц 26-28**, для кристаллов  $F_2$  значение  $\lambda_B^{III}$  (1.55 Å) близко по величине как к  $r(F-F)=1.56$  Å, так и к  $R_{SD}$  атомов фтора (1.58 Å). Эти совпадения позволяют предположить, что поведение р-оболочки одновременно удовлетворяет моделям как жестких, так и мягких сфер. Учитывая, что в модели мягких сфер взаимодействие между двумя атомами А отсутствует если  $r(A-A) > 2R_{SD}(A)$ , а также то, что атом фтора предшествует неону, для которого указанная оболочка ведет себя уже как абсолютно жесткая сфера, такое поведение атома фтора в кристаллах вполне объяснимо. Дуализм поведения оболочек «мягкая сфера – жесткая сфера» свойствен не только атомам фтора. Так в структурах  $N_2$ ,  $O_2$ ,  $C_{60}$  также присутствуют невалентные контакты, длина которых близка или совпадает с удвоенным значением  $R_{SD}$  их образующих атомов (Рис. 83 и **Таблицы 26, 29**). Возможно, что двойственность поведения электронных оболочек связана с корпускулярно-волновым дуализмом электронов и является одной из причин корректного расчета зонной структуры кристалла металлического натрия Вигнером и Зейтцем, которые при решении уравнения Шредингера в качестве граничных условий использовали сферу радиуса  $R_{SD}$  [279].



### 4.3. Использование полиэдров Вороного для построения карты миграции подвижных катионов

Важным свойством ПВ является то, что его вершины и ребра представляют собой наборы точек, наиболее удаленных от атомов каркаса. Это означает, что их можно отнести к центрам пустот и линиям каналов возможной миграции атомов в твердых электролитах, соответственно. Определяя ПВ для центров пустот, можно отбросить слишком мелкие пустоты из-за того, что они недоступны для подвижных катионов, и построить карту миграции.

Описанный выше подход был использован для отбора возможных натриевых твердых электролитов. Чтобы проанализировать с помощью ПВ кристаллическую структуру на предмет возможной ионной проводимости, соединение должно содержать мигрирующий катион -  $\text{Na}^+$  - и группу анионов - здесь  $\text{O}^{2-}$  - а также дополнительные ионы металла  $\text{M}^{n+}$ . В таком материале,  $\text{Na}_x\text{M}_y\text{O}_z$ , вершины ПВ вокруг каждого атома М и О предполагают все характерные для структуры пустоты каркаса  $\text{M}_y\text{O}_z$ . Так как структура должна быть проанализирована для всех позиций, которые могут разместить ионы  $\text{Na}^+$ , имеющиеся в структуре такие ионы должны быть проигнорированы, а все оставшиеся позиции атомов должны быть учтены. Чтобы соединение показало ионную проводимость в соответствии с разбиением Вороного, необходимо выполнение двух условий:

1. Как пустоты, так и каналы необходимо определять только по анионам. Если один из атомов, образующих пустоты, является катионом, кулоновское отталкивание может создать энергетический барьер, слишком большой для другого катиона, чтобы он вошел в эту позицию или канал. Поэтому необходимо пренебречь всеми пустотами, ПВ которых содержат грани между пустотами и катионами с телесным углом больше некоторого значения  $\Omega_{\min}$ .

2. Геометрические данные пустот сравниваются с ожидаемыми значениями:  $R_{SD}$  должно быть, по крайней мере, таким же по величине, как среднестатистическое значение Na в окружении атомов O, равное 1.6 Å.

В итоге, среди 402 выбранных из ICSD кристаллических структур были идентифицированы не только все известные Na-ионные проводники (Na-β-глинозем  $\text{Na}_{1.68}\text{Al}_{11}\text{O}_{17}$ , NaSICON  $\text{Na}_{3.05}\text{Zr}_2\text{Si}_{2.05}\text{P}_{0.95}\text{O}_{12}$ ), но и новые потенциальные твердые электролиты, содержащие подвижные катионы натрия (Рис. 84). Чтобы выделить наиболее многообещающие соединения, все материалы считались перспективными лишь после финального этапа, на котором были рассчитаны кристаллографические и экономические параметры на основе структурных данных и введена схема их взвешивания и оценки. Таким образом, наиболее перспективными оказались алюмосиликаты и оксиды переходных металлов. В рамках данной работы созданы программные средства для расчета дескрипторов структур кандидатов в твердые электролиты.

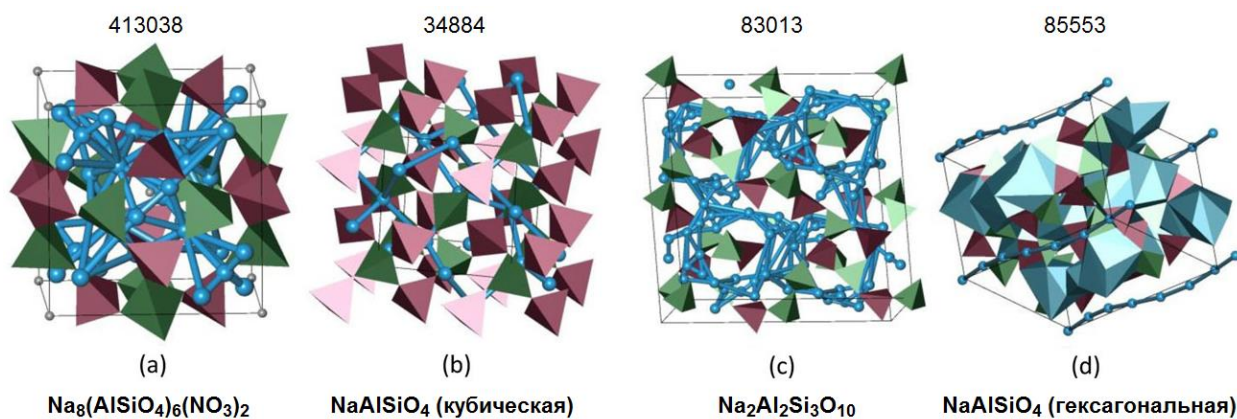


Рис. 84. Четыре наилучших (согласно используемому алгоритму) новых потенциальных Na-ионных проводника. Все они являются алюмосиликатами и содержат КП  $\text{AlO}_4$  (коричневые),  $\text{SiO}_4$  (зеленые),  $\text{NaO}_x$  (синие). Синие шары и цилиндры отвечают выявленным пустотам и каналам.

#### 4.4. Использование полиэдров Вороного для анализа кристаллических структур интерметаллидов

Интерметаллиды существенно отличаются по строению от КС и интерес к ним в последнее время непрерывно растет из-за необычных и полезных свойств таких соединений только из атомов металлов. Интерметаллиды важны для промышленности, как конструкционные, магнитные и с памятью формы материалы, проводники и сверхпроводники. Чтобы лучше понять корреляции между их физическими свойствами и структурой, мы использовали геометрико-топологический подход для классификации более 38 000 интерметаллических соединений по 3708 топологическим типам. При этом интерметаллиды рассматривались на различных уровнях структурной организации: 1) координационные полиэдры; 2) многослойные кластеры; 3) двумерные сетки; 4) трехмерные каркасы.

При анализе координационных полиэдров было показано, что 12-вершинные многогранники являются наиболее частыми (33%) и что почти половина из них имеет форму икосаэдра (46%), за ней следует кубооктаэдр (25%) и, что неожиданно, двухшапочная пятиугольная призма (13%). Для всех 38 000 структур было рассчитано представление интерметаллида в виде сборки из многослойных (луковичных) нанокластеров (Рис. 85), которые были сохранены в TTN коллекции.

Результаты анализа данных коллекции показали, что 41% интерметаллидов можно собрать с помощью одного нанокластера и что 22.4% из них упакованы в соответствии с гранецентрированным кубическим мотивом плотнейшей упаковки сфер в трехмерном пространстве. При анализе топологии слоев установлено, что гексагональная решетка, которая соответствует плотнейшей упаковке сфер на плоскости, существует в более чем 11 тыс. кристаллических структур, подтверждая плотноупакованный характер интерметаллидов.

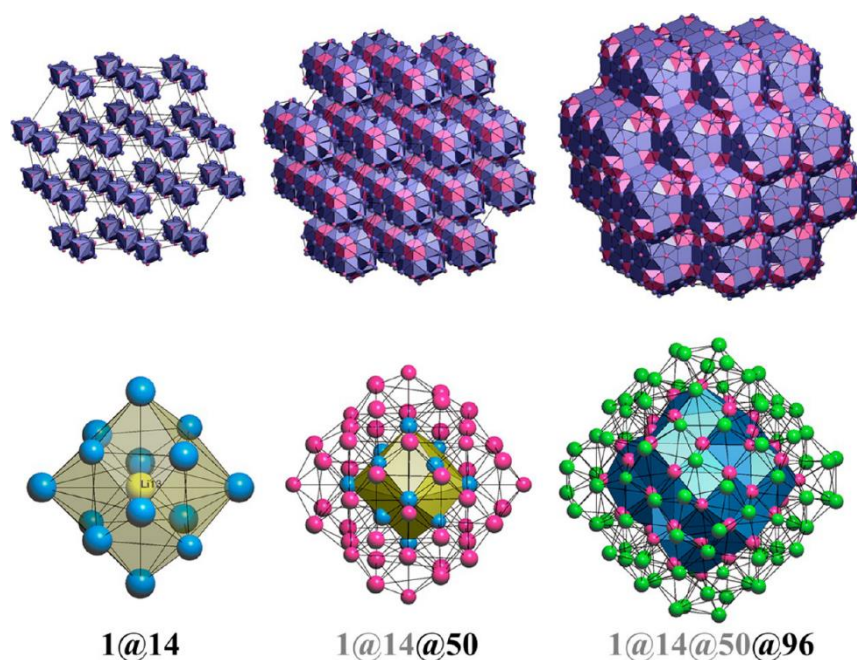


Рис. 85. Послойный рост нанокластеров в кристаллической структуре  $\text{Li}_{17}\text{Ge}_4$ : (слева) первая 14-атомная оболочка; (в центре) вторая 50-атомная оболочка; (справа) третья 96-атомная оболочка.

Дополнительно к анализируемой выборке интерметаллических соединений был применен нанокластерный подход, который выходит за рамки первой координационной сферы и представляет структурные единицы как кластеры с множеством оболочек, собирающие всю структуру в целом. Этот подход показал, что 41% интерметаллидов можно собрать с помощью одного нанокластера и что 22.4% из них упакованы в соответствии с гранецентрированным кубическим мотивом плотнейшей упаковки сфер в трехмерном пространстве. В рамках этой схемы все структурные дескрипторы могут быть связаны с экспериментальными данными и результатами теоретического моделирования, а затем использованы для синтеза новых интерметаллических соединений и для прогнозирования новых материалов.

#### 4.5. Использование полиэдров Вороного для анализа строения кристаллов насыщенных углеводородов

В рамках данной работы выполнен геометрико-топологический анализ структур насыщенных углеводородов, не содержащие статистически разупорядоченных атомов и ошибок в структурных данных. Указанным условиям удовлетворяло 147 соединений, включающих 1931 кристаллографически разный сорт атомов углерода и 3004 сорта атомов водорода или дейтерия (36 атомов в 4-х соединениях). Так как природа изотопа (H или D) практически не изменяет геометрию структуры, то далее они не разделяются и все рассматриваются как атомы водорода.

При первичном анализе межатомных расстояний в восьми структурах (в частности, в циклобутане  $C_4H_8$ , (ZZZWEO02)) были выявлены контакты Н-Н, длина которых не превышала 1.3 Å. Наличие таких аномально коротких контактов является результатом некорректного геометрического определения позиций атомов водорода. Поэтому мы провели повторное определение координат атомов водорода с помощью входящей в состав комплекса ToposPro программы HSite [200], после чего кратчайшее расстояние Н-Н в обсуждаемых структурах оказалось не менее 1.78 Å.

Как известно, в кристаллах насыщенных углеводородов атомы углерода находятся в  $sp^3$ -гибридном состоянии. Если учесть природу атомов окружения, то можно выделить 5 типов атомов С, характеристики которых представлены в **Таблице 30**. Первичный анализ данных показал, что помимо тетраэдрической координации у атома углерода при составе КП  $C_4H_0$  может быть иной тип координации  $C_4H_0(sp^2)$  (**Таблица 30**). Во всех четырех соединениях, содержащих атомы  $C_4H_0(sp^2)$ , имеется один или несколько фрагментов углеродного каркаса в форме тригональной бипирамиды (Рис. 86, **Таблица 31**), апикальные атомы которой не связаны друг с другом (**Таблица 31**,  $\Omega(C2-C4)=8.13\%$ ).

**Таблица 30.** Характеристики ПВ атомов углерода в зависимости от их гибридного состояния и числа атомов водорода в их координационном полиэдре

Состав КП атомов углерода	Число соеди- нений	Число атомов углерода	$R_{SD}(C)$ , Å	$V_{ПВ}(C)$ , Å <sup>3</sup>	$G_3$	Связи C-C*		
						k	r, Å	Ω, %
C0H4	1	9	0.82(1)	2.3(1)	0.104(1)	—	—	—
C1H3	47	175	0.85(3)	2.6(2)	0.108(3)	175	1.53(2)	19(1)
C2H2	132	976	0.93(3)	3.4(3)	0.108(9)	1952	1.52(2)	18(1)
C3H1	89	523	1.02(2)	4.5(3)	0.100(6)	1569	1.54(2)	20(1)
C4H0	79	238	1.15(6)	6.5(1.0)	0.100(10)	952	1.52(4)	20(2)
C4H0(sp <sup>2</sup> )	4	10	1.30(2)	9.3(5)	0.103(2)	30	1.53(2)	19(1)
						10	1.58(1)	8(1)
CnHm**	147	1921	0.98(9)	4.0(1.2)	0.105(9)	4648	1.53(3)	19(2)

\* k – число связей, r – средняя длина связи C-C, а Ω - соответствующая ей величина телесного угла.

\*\* Данные для C4H0(sp<sup>2</sup>) не учтены.

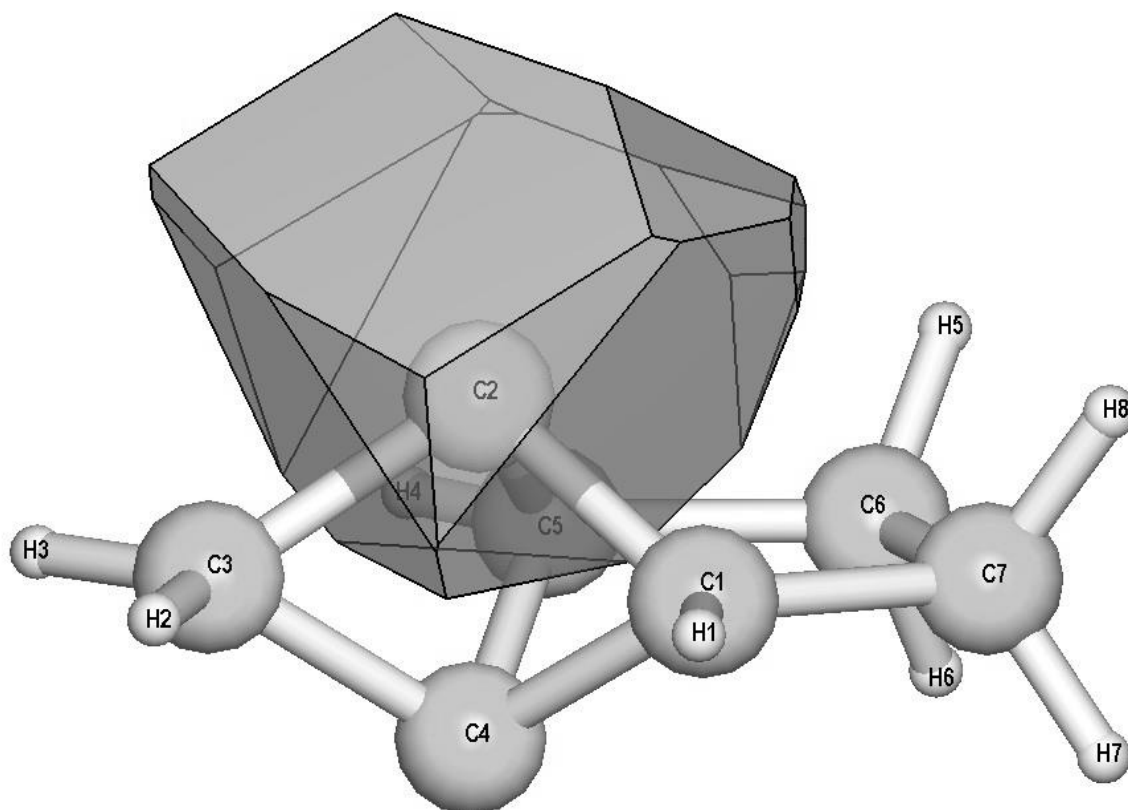


Рис. 86. Молекула тетрацикло(4.1.0.0-1,5-.0-2,6-)гептана [280] (SADHUA).

Очевидно, что апикальные атомы такой бипирамиды находятся не в  $sp^3$ , а  $sp^2$  гибридном состоянии (как и в фуллерене), поэтому 10 обнаруженных сортов атомов углерода типа  $C_4H_0(sp^2)$  в данной работе не обсуждаются. Согласно данным **Таблица 30**, при уменьшении числа атомов водорода, окружающих углерод, среднестатистическая величина его  $R_{SD}$  (или  $V_{ПВ}$ ) возрастает (Рис. 87), что объясняется меньшим размером атома Н по сравнению с атомом С. Величина  $G_3$  максимальна для концевых атомов С.

**Таблица 31.** Характеристики ПВ атома C(2) в структуре тетрацикло (4.1.0.0-1,5-.0-2,6) гептана [280] (SADHUA)\*

Характеристики ПВ: $R_{SD}=1.267 \text{ \AA}$ ; $V_{ПВ}=8.5 \text{ \AA}^3$ ; $G_3=0.1001$									
№	Атом	r, $\text{\AA}$	$\Omega$ , %	$\tau$	№	Атом	r, $\text{\AA}$	$\Omega$ , %	$\tau$
1	C(3)	1.512	18.27	1	10	C(7) #	2.471	0.66	2
2	C(5)	1.519	18.12	1	11	H(1)	2.759	9.18	0
3	C(1)	1.520	18.21	1	12	H(8)	2.824	8.92	0
4	C(4)	1.577	8.13	2	13	H(5) #	2.889	0.76	3
5	H(4) #	2.149	1.97	2	14	H(8) #	2.891	1.07	3
6	H(3) #	2.161	1.82	2	15	H(7)	2.909	7.09	0
7	H(2) #	2.181	1.52	2	16	H(5)	3.372	1.61	0
8	H(1) #	2.182	1.94	2	17	H(2) #	3.689	0.02	0
9	C(6) #	2.469	0.71	2					

\*Обозначения аналогичны указанным в **Таблице 2**.

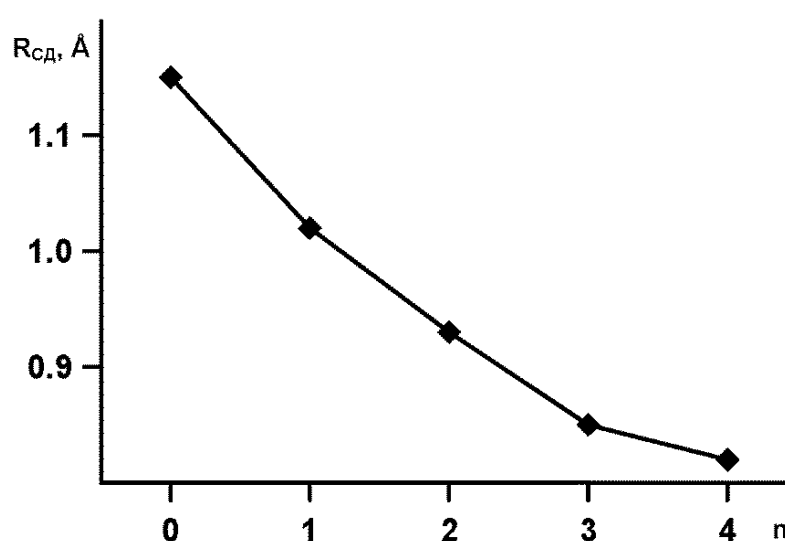


Рис. 87. Зависимость  $R_{SD}$  атомов  $sp^3$ -гибридного углерода от числа связей C-H ( $n$ ).

Межатомное расстояние С-С практически не зависит от типа центрального атома углерода. Для оценки влияния сорта атомов окружения на величину  $R_{SD}$  центрального атома углерода используем отношение ( $\nu$ ) суммы телесных углов, приходящихся на атомы водорода, к полному телесному углу сферы, равному  $4\pi$  ср. Полученная зависимость  $R_{SD}(C)$  от  $\nu$  представлена на Рис. 88. На графике отчетливо выделяются 5 кластеров, отвечающих пяти типам атомов углерода (**Таблица 30**), причем кластер  $C0H4$  представлен на рисунке единственной точкой при  $\nu=1$ . Наличие в кластере  $C4H0$  точек, характеризующихся завышенным значением  $R_{SD}(C)$ , обусловлено присутствием в выборке структур триангуланов (например,  $C_{31}H_{34}$ , (WEXFUA)), которые построены из трехчленных циклов, из-за чего координационные полиэдры атомов углерода типа  $C4H0$  сильно искажены, а объемы их ПВ приблизительно на  $2 \text{ \AA}^3$  больше среднестатистического значения.

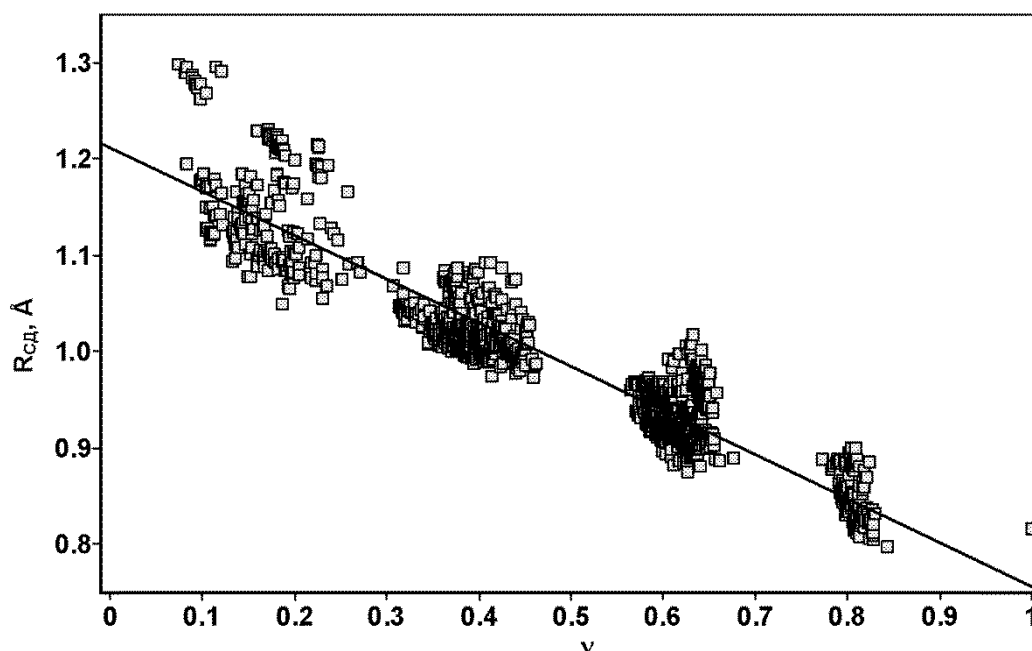


Рис. 88. Зависимость  $R_{SD}$  атомов углерода от величины  $\nu$  ( $\nu$  - отношение суммарного телесного угла, отвечающего граням С-Н, к полному телесному углу).



Зависимость  $R_{SD}(C)$  от  $v$  близка к линейной и, согласно МНК, описывается для 1921 точки уравнением

$$R_{SD}(C) = 1.211(2) - 0.455(4)v \quad (17)$$

с модулем коэффициента корреляции ( $\rho$ ) 0.935. Экстраполяция зависимости (17) до точки пересечения с осью ординат дает значение  $R_{SD}(C)$ , равное 1.21 Å, которое отвечает размеру  $sp^3$ -гибридного атома С в чисто углеродном окружении и близко поэтому к  $R_{SD}$  атомов в структуре алмаза (1.11 Å).

Характеристики ПВ атомов водорода в зависимости от типа атома углерода, с которым они связаны, представлены в **Таблице 32**. Полученные данные свидетельствуют, что изменение типа атома С в целом практически не влияет на величину  $G_3$  и межатомное расстояние  $r(C-H)$ , что неудивительно, так как положение атомов Н в большинстве структур определено на основании геометрических соображений.

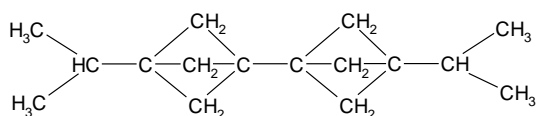
**Таблица 32.** Характеристики ПВ атомов водорода в зависимости от гибридного состояния, связанного с ним атома углерода, и числа атомов водорода в его координационном полиэдре

Состав КП атомов углерода	Число атомов водорода	$R_{SD}(H)$ , Å	$V_{ПВ}(H)$ , Å <sup>3</sup>	$G_3$	Связи С-Н	
					$r$ , Å	$\Omega$ , %
C0H4	24	1.41(3)	11.7(0.7)	0.097(2)	1.10(1)	25(1)
C1H3	520	1.36(5)	10.6(1.2)	0.100(4)	1.01(6)	26(1)
C2H2	1937	1.37(5)	10.7(1.2)	0.098(4)	0.99(5)	29(1)
C3H1	523	1.34(6)	10.2(1.4)	0.097(3)	1.00(5)	30(2)
CnHm	3004	1.36(5)	10.6(1.3)	0.098(4)	1.00(5)	28(2)

Из-за большей длины связи С-С по сравнению с С-Н убыль числа атомов Н в координационном полиэдре атома С приводит к небольшому росту телесного угла  $\Omega(C-H)$ . Величина  $R_{SD}$  (или  $V_{ПВ}$ ) атома Н практически не зависит от

типа связанного с ним атома углерода, что указывает на отсутствие влияния невалентных контактов  $H\cdots C$  или  $H\cdots H$  на размер атома водорода. Вместе с тем, из **Таблицы 32** видно, что из четырех типов атомов H только атомы водорода в структуре метана (тип C0H4) имеют значительно завышенные по сравнению с остальными средние значения  $R_{SD}(H)$ ,  $V_{ПВ}(H)$  и  $r(C-H)$ . В связи с этим отметим, что согласно уравнению (17) при  $v=1$  значение  $R_{SD}(C)$  отвечает  $sp^3$  гибриднему атому углерода в чисто водородном окружении. Существенно, что эта величина ( $0.76 \text{ \AA}$ ) меньше полученной из расчета по структурным данным для кристаллов метана ( $0.82 \text{ \AA}$ , **Таблица 30**). Поскольку все имеющиеся данные для атомов C0H4 опираются на результаты единственного определения структуры  $CD_4$  [281], содержащей 9 кристаллографически разных сортов молекул, можно предположить, что одновременно завышенные значения  $R_{SD}(C)$ ,  $R_{SD}(H)$  и  $r(C-H)$  могут быть обусловлены каким-то приближением, допущенным в работе [281], поэтому независимое повторное определение структуры кристаллов  $CD_4$  представляет, на наш взгляд, несомненный интерес.

По аналогии с результатами [282], полученные средние величины  $V_{ПВД}$  атомов C и H (**Таблицы 30 и 32**) можно использовать для расчета объема молекул ( $V_{\text{мол.}}$ ) и плотности кристаллов ( $d_{\text{ПВД}}$ ) углеводородов. Например, для 3,3'-диизопропил-1,1'-би(бицикло(1.1.1) пентана) [283] (QETVOA):



рассчитанный объем молекулы (для атомов водорода использовали значение  $V_{ПВ}$ , усредненное по всем типам связанного с ним атома C) равен

$$V_{\text{мол.}}(C_{16}H_{26}) = 4 \times 2.5 + 6 \times 3.4 + 2 \times 4.5 + 4 \times 6.5 + 26 \times 10.6 = 341 \text{ \AA}^3, \quad (18)$$

а  $d_{\text{ПВ}} = 1.063 \text{ г/см}^3$ . Экспериментально установленный объем молекулы  $C_{16}H_{26}$ , полученный на основе данных о параметрах элементарной ячейки, равен

360.4 Å<sup>3</sup>, а рентгеновская плотность кристаллов составляет  $d_{\text{рент.}}=1.006$  г/см<sup>3</sup>, что, на наш взгляд, удовлетворительно согласуется с рассчитанными значениями. Подобная процедура вычисления  $d_{\text{пв}}$  была проведена для всех рассмотренных углеводов. Результаты представлены на Рис. 89а в виде ( $d_{\text{рент.}}$ ,  $d_{\text{пв}}$ )-распределения, которое описывается МНК-зависимостью

$$d_{\text{пв}}=1.017d_{\text{рент.}} - 0.004 \quad (|\rho|=0.722). \quad (19)$$

Величины отношения  $d_{\text{рент.}}/d_{\text{пв}}$  лежат в диапазоне от 0.89 до 1.28 г/см<sup>3</sup>. Наибольшие отклонения отвечают структурам, данные для которых представлены в Таблице 33.

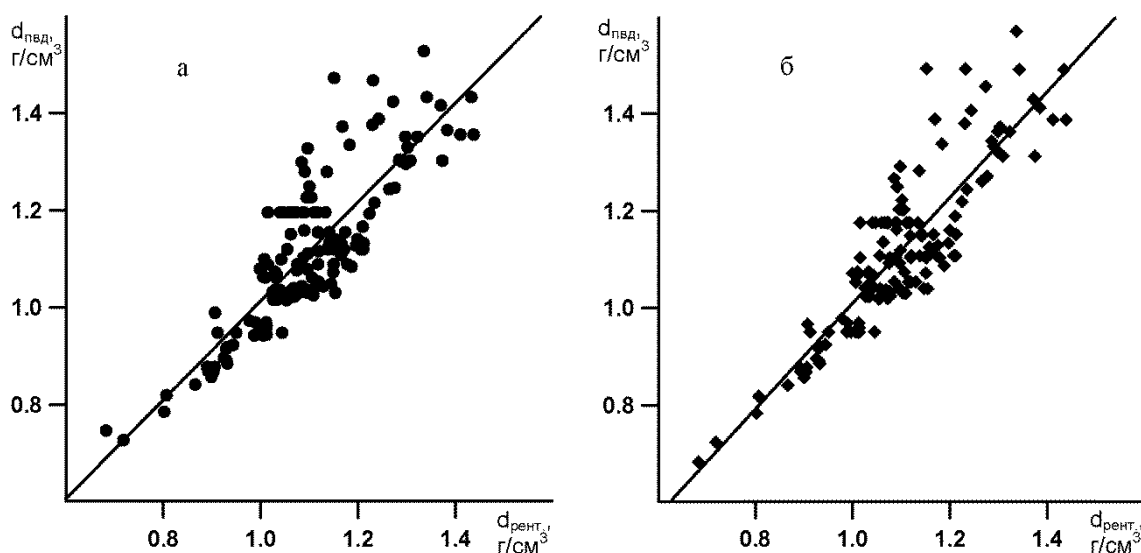


Рис. 89. Корреляция рентгеновской плотности ( $d_{\text{рент.}}$ ) и плотности ( $d_{\text{пв}}$ ), рассчитанной из объемов ПВ атомов С и Н (Таблицы 30, 32, а) или атомов в подрешетках только из атомов С (Таблица 34, б) в 147 кристаллах углеводов.

Заметим, что в [282] используется метод расчета плотности, при котором рассматривают не отдельные атомы, а атомные группировки, включающие неводородный атом и атомы водорода, связанные с ним. По объему таких группировок в итоге и рассчитывают плотность кристаллов. Если использовать подход [282], то на основании данных Таблицы 34

получим схожее с рассмотренным ( $d_{\text{рент.}}$ ,  $d_{\text{пв}}$ )-распределение (Рис. 89б). В этом случае величины отношения  $d_{\text{рент.}}/d_{\text{пв}}$  лежат в диапазоне от 0.90 до 1.30, а МНК-зависимость имеет вид:

$$d_{\text{пв}} = 1.089d_{\text{рент.}} - 0.079. \quad (|\rho| = 0.767). \quad (20)$$

Как видно, оба рассмотренных варианта расчета дают приблизительно одинаковые результаты, однако для отдельных молекул лучшее совпадение  $d_{\text{рент.}}$  и  $d_{\text{пв}}$  получается при использовании первого метода (Таблица 33).

**Таблица 33.** Структуры, характеризующиеся наибольшими отклонениями  $d_{\text{рент.}}$  от  $d_{\text{пв}}$ \*

RefCode	Формула	$d_{\text{рент.}}$ , г/см <sup>3</sup>	Метод I		Метод II	
			$d_{\text{пв}}$ , г/см <sup>3</sup>	$d_{\text{пв}}/d_{\text{рент.}}$	$d_{\text{пв}}$ , г/см <sup>3</sup>	$d_{\text{пв}}/d_{\text{рент.}}$
PIRFIF	C <sub>16</sub> H <sub>28</sub>	1.15	1.03	0.89	1.04	0.90
BAPOCM10	C <sub>18</sub> H <sub>30</sub>	1.15	1.05	0.91	1.04	0.91
BATVOH	C <sub>10</sub> H <sub>20</sub>	1.05	0.95	0.91	0.95	0.91
WEXFUA	C <sub>31</sub> H <sub>34</sub>	1.10	1.33	1.21	1.29	1.18
GICBEZ	C <sub>16</sub> H <sub>18</sub>	1.23	1.47	1.19	1.49	1.21
SADHUA	C <sub>7</sub> H <sub>8</sub>	1.15	1.47	1.28	1.49	1.30

\*Учитываются объемы отдельных атомов (метод I) или атомных группировок (метод II).

**Таблица 34.** Характеристики ПВ атомов в подрешетках из атомов С в зависимости от гибридного состояния и числа атомов водорода в КП атома углерода

Состав КП атомов углерода	$R_{\text{SD}}(\text{C})$ , Å	$V_{\text{VP}}(\text{C})$ , Å <sup>3</sup>	$G_3$	Состав КП атомов углерода	$R_{\text{SD}}(\text{C})$ , Å	$V_{\text{VP}}(\text{C})$ , Å <sup>3</sup>	$G_3$
C0H4	2.27(1)	48.8(0.1)	0.079(1)	C3H1	1.51(6)	14.5(1.8)	0.129(7)
C1H3	2.02(5)	34.5(2.4)	0.096(3)	C4H0	1.22(7)	7.6(1.4)	0.113(13)
C2H2	1.80(5)	24.5(2.1)	0.111(5)	C4H0(sp <sup>2</sup> )	1.54(2)	15.2(6)	0.123(2)
CnHm*	1.67(24)	20.7(8.0)	0.114(12)				

\*Данные для C4H0(sp<sup>2</sup>) не учтены.

Положительные отклонения  $d_{\text{ПВ}}$  от  $d_{\text{рент.}}$  обычно связаны со стерическими затруднениями при размещении атомов в молекуле, проявляющиеся, например, в значительном отклонении формы КП атома углерода от тетраэдрической. Так, в структуре тетрацикло(4.1.0.0-1,5-.0-2,6-)гептана  $\text{C}_7\text{H}_8$  валентные углы С-С-Н и С-С-С лежат в диапазонах  $111-124^\circ$  и  $62-109^\circ$  соответственно. Аналогичные искажения обнаружены в структуре соединений  $\text{C}_{16}\text{H}_{18}$  и  $\text{C}_{31}\text{H}_{34}$ . Напротив, отрицательные отклонения связаны, по-видимому, с более плотной, чем среднестатистическая, упаковкой молекул в кристалле из-за специфики их формы. Поэтому  $d_{\text{ПВ}}$  приблизительно на  $0.1 \text{ г/см}^3$  меньше  $d_{\text{рент.}}$  для кристаллов пергидропарациклофана  $\text{C}_{16}\text{H}_{28}$ , циклодекана  $\text{C}_{10}\text{H}_{20}$  и 1-биапокамфана  $\text{C}_{18}\text{H}_{30}$ .

Исходя из данных **Таблицы 30** и **32**, плотность нормальных насыщенных углеводородов  $\text{C}_n\text{H}_{2n+2}$  (**Таблица 35**) можно оценить по формуле

$$d_{\text{ПВД}} = \frac{1.66 \cdot (n \cdot (M_{\text{C}} + 2M_{\text{H}}) + 2M_{\text{H}})}{n \cdot (V_{\text{C}_2\text{H}_2} + 2V_{\text{H}}) + 2 \cdot (V_{\text{C}_1\text{H}_3} - V_{\text{C}_2\text{H}_2} + V_{\text{H}})} = \frac{1.66 \cdot (14.027n + 2.016)}{24.6n + 19.6}, \quad (21)$$

где  $M_{\text{C}}$  и  $M_{\text{H}}$  – атомные массы углерода и водорода, а  $V$  – объемы ПВ соответствующих атомов. В целом, значения, полученные по уравнению (21), неплохо согласуются с соответствующими им величинами  $d_{\text{рент.}}$  (Рис. 90). На основании зависимости (21) можно рассчитать, что предельное значение плотности при бесконечно большом  $n$  равно  $0.947 \text{ г/см}^3$ . Если плотность углеводородов  $\text{C}_n\text{H}_{2n+2}$  имеет предел, то таковой отсутствует у величины  $G_3$  их молекулярного ПВ, т.е. совокупности ПВ всех атомов, относящихся к одной молекуле. Так,  $G_3$  закономерно увеличивается при росте числа связей С-С в структуре квазилинейных молекул углеводорода (**Таблица 35**).

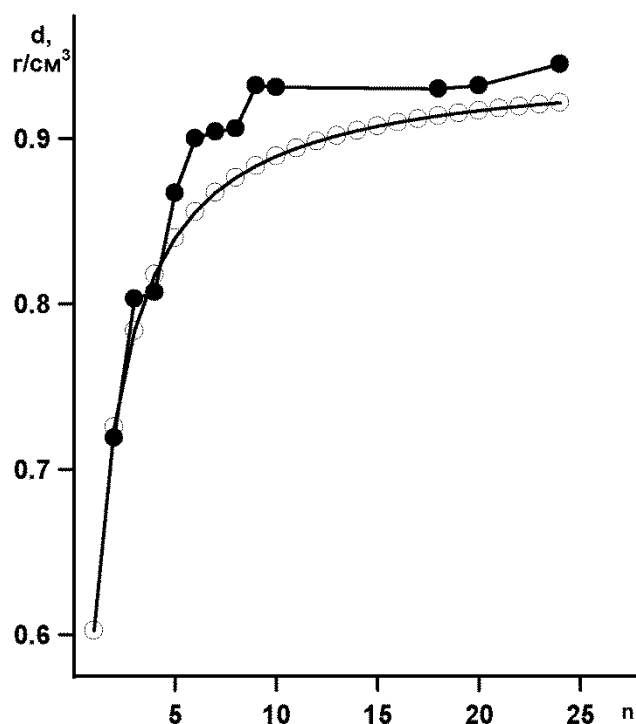


Рис. 90. Зависимость рентгеновской (черные круги) и рассчитанной по уравнению (21) (светлые круги) плотности нормальных насыщенных углеводородов  $C_nH_{2n+2}$  от  $n$ .

**Таблица 35.** Характеристики нормальных насыщенных углеводородов

Состав	$d_{\text{рент.}}, \text{г/см}^3$	Простр. группа	Температура измерения, К	$G_3$	$V_{\text{мол.}}, \text{\AA}^3$	RefCode
$C_2H_6$	0.719	$P2_1/n$	85	0.0845	69.4	ETHANE01
$C_3H_8$	0.803	$P2_1/n$	30	0.0877	91.2	JAYDUI
$C_4H_{10}$	0.807	$P2_1/c$	90	0.0910	119.6	DUCKOB04
$C_5H_{12}$	0.867	$Pbcn$	90	0.1018	138.3	PENTAN01
$C_6H_{14}$	0.900	$P\bar{1}$	90	0.1128	159.1	HEXANE01
$C_7H_{16}$	0.904	$P\bar{1}$	90	0.1258	184.0	HEPTAN02
$C_8H_{18}$	0.906	$P\bar{1}$	90	0.1366	209.4	OCTANE01
$C_9H_{20}$	0.932	$P\bar{1}$	90	0.1536	228.5	QQQFAY01
$C_{10}H_{22}$	0.931	$P\bar{1}$	150	0.1663	253.9	QQQFBG01
$C_{18}H_{38}$	0.930	$P\bar{1}$	—	0.2996	454.4	NOCTDC
$C_{20}H_{42}$	0.932	$P\bar{1}$	—	0.3352	503.2	QQQCIM01
$C_{24}H_{50}$	0.945	$P\bar{1}$	—	0.4200	595.4	ZZZOFC04

Как известно, одной из наиболее используемых характеристик структуры любых кристаллов является набор межатомных расстояний. Если в классических методиках кристаллохимического анализа необходимо искусственно ограничивать рассматриваемые дальние контакты некоторым верхним пределом [154], то при геометрико-топологическом подходе такое ограничение возникает автоматически, так как в качестве соседних учитываются только те атомы, которые имеют общие грани ПВ. Например, в системе “идеальный газ”, которую можно рассматривать и как твердое тело, состоящее из нематериальных точек, во взаимном размещении которых отсутствует не только дальний, но и ближний порядок, распределение расстояний по длине  $(r, N)$  имеет мономодальный вид (Рис. 91а), при этом зависимость телесных углов граней ПВ точек от расстояния между ними  $(r, \Omega)$  имеет характерный вид “хвоста взлетающей ракеты” (Рис. 91б, [20]).

В структуре кристаллов (например, содержащих атомы А и В) из-за наличия дальнего порядка, влияния природы и размера атомов, а также возникновения ближнего порядка, обусловленного совокупностью химических взаимодействий между соседними атомами, вид распределений  $(r, N)$  и  $(r, \Omega)$ , характерных для модельной системы “идеальный газ”, существенно изменяется. Так, из-за наличия у атомов определенных размеров, запрещающих их сближение на расстояние более короткое, чем некоторое  $r_{\min}$ , на таких распределениях (Рис. 91в-е) появляются свободные области в диапазоне от 0 до  $r_{\min}$ , при этом  $(r, N)$  распределения становятся многомодальными, а на зависимостях  $(r, \Omega)$  возникают непересекающиеся области точек, отвечающих определенным контактам (в обсуждаемом случае А-А, А-В или В-В), которые являются либо химическими связями (при больших  $\Omega$  и малых  $r \geq r_{\min}$ ), либо ван-дер-Ваальсовыми взаимодействиями (при малых  $\Omega$  и  $r \gg r_{\min}$ ).

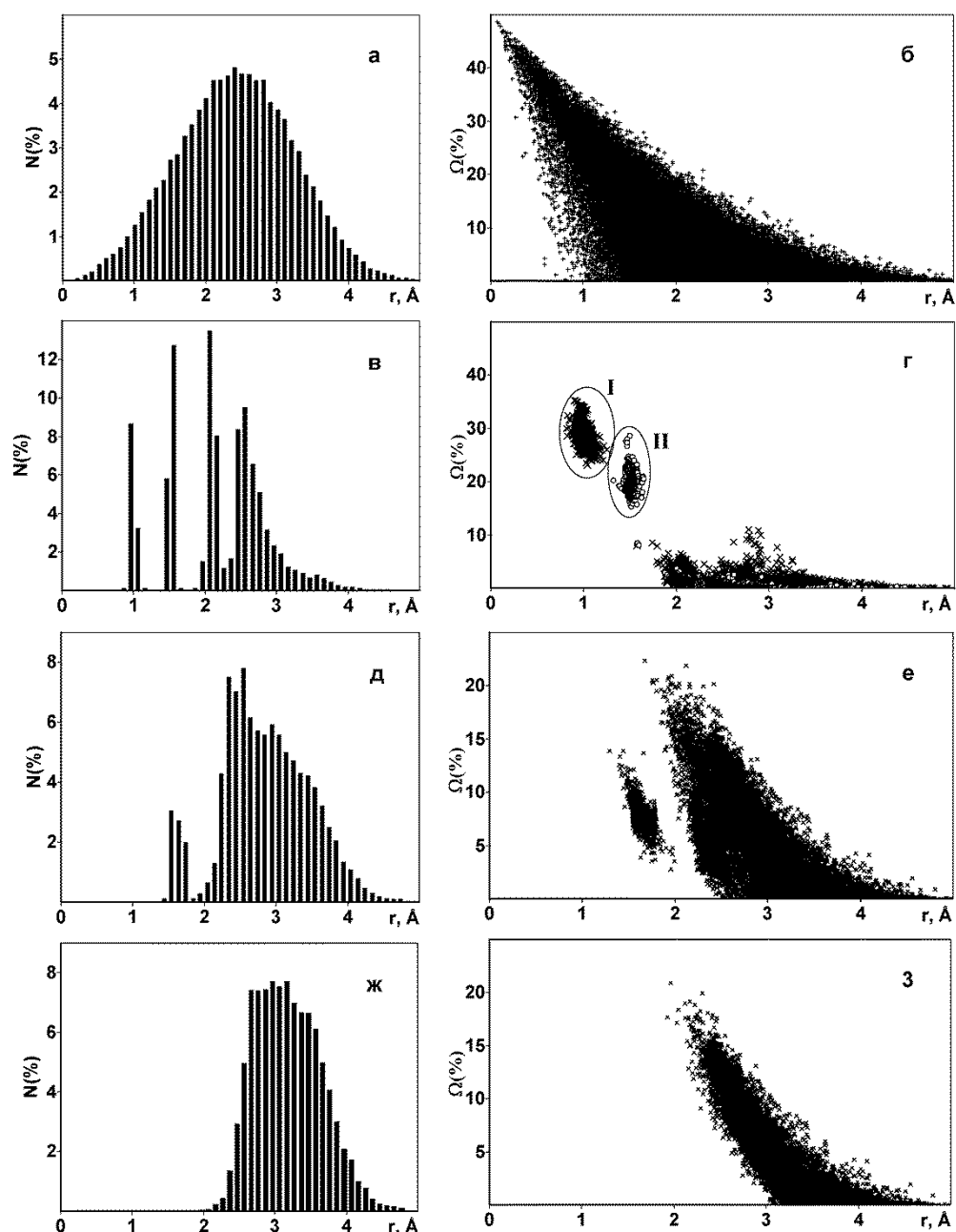


Рис. 91. Гистограммы длин контактов и  $(r, \Omega)$ -распределения в системе «идеальный газ» [20] (а, б, 155584 точки) и в 147 структурах насыщенных углеводородов для С-Н и С-С контактов (в, г, 25137 точек), а также всех Н $\cdots$ Н контактов (д, е, 38427 точек) и Н $\cdots$ Н контактов с нулевым рангом (ж, з, 22014 точек). На рисунке г контакты С-Н обозначены крестами, а контакты С-С – кругами. Область I отвечает связям С-Н, а область II – связям С-С. Остальные точки соответствуют невалентным контактам С $\cdots$ С или С $\cdots$ Н.



Анализ реальных распределений ( $r$ ,  $N$ ) и ( $r$ ,  $\Omega$ ) показывает, что как валентные, так и невалентные контакты, образующие грани в ПВ между парами атомов А и В, достаточно четко разделяются на группы по рангу контакта А-В. Как показали результаты расчетов структуры обсуждаемых углеводородов, наибольшее разнообразие рангов отвечает контактам Н-Н (Таблица 36, Рис. 91 д,е и ж,з). Согласно данным Таблицы 36, все контакты Н-Н с  $\tau=2$  являются непрямыми. Почти полное соблюдение этого правила и для контактов С-Н или С-С с  $\tau=2$  (Таблица 36), по-видимому, указывает на сильный отталкивающий характер таких межатомных взаимодействий.

**Таблица 36.** Зависимость от типа и ранга межатомных взаимодействий их числа, средней длины, а также доли в них прямых контактов в структуре насыщенных углеводородов

Тип	Ранг контактов ( $\tau$ )								
	1	2	3	4	5	6	7	8	0
Число контактов*									
Н-Н	—	3049	5534	4024	2307	908	426	112	22020
С-Н	3036	5779	3186	995	188	25	9	—	1417
С-С	4688	5018	719	54	—	—	—	—	13
Доля прямых контактов (в %)									
Н-Н	—	0	12	79	73	75	56	41	72
С-Н	100	<1	7	11	10	20	0	—	21
С-С	100	2	42	4	—	—	—	—	15
Средняя длина контактов (в Å)*									
Н-Н	—	1.64(9)	2.4(1)	2.6(2)	2.6(5)	2.7(5)	3.0(6)	3.2(4)	3.2(5)
С-Н	1.00(5)	2.09(6)	2.7(1)	2.9(3)	3.0(4)	3.3(8)	3.3(6)	—	3.4(4)
С-С	1.53(3)	2.5(1)	3.0(2)	3.6(3)	—	—	—	—	3.8(2)

\*Учитываются как прямые, так и не прямые контакты соответствующего ранга.

Так как атомы С в обсуждаемых соединениях валентно насыщены, для них характерны валентные прямые контакты с  $\tau=1$ . Интересно, что преимущественно прямыми контактами являются не только межмолекулярные взаимодействия  $H\cdots H$  с  $\tau=0$ , но и внутримолекулярные контакты  $H\cdots H$  с  $\tau=5, 6$  или  $7$ . При увеличении ранга (от  $1$  и выше) средняя длина контактов  $r(H\cdots H)$  возрастает, достигая максимальной величины для контактов с нулевым рангом (**Таблица 36**, Рис. 91 ж,з). Существенно, что найденное значение  $r(H\cdots H)=3.2 \text{ \AA}$  близко по величине к равновесному расстоянию  $R_{eq}$  для потенциала невалентного взаимодействия двух атомов водорода, которое, согласно [284], равно  $3.36 \text{ \AA}$ . Учитывая, что такое же соответствие наблюдается для контактов  $C\cdots H$  и  $C\cdots C$  (**Таблица 36**),  $R_{eq}$  которых равны  $3.29$  и  $3.89 \text{ \AA}$  соответственно, можно предложить использовать гистограммы распределения длин контактов с  $\tau=0$  для расчета параметров потенциальных кривых парных атом-атомных взаимодействий в кристаллах.

## Заключение

1. На основе разработанного геометрико-топологического подхода предложена универсальная методология машинного анализа кристаллических структур, позволяющая прогнозировать их строение и физические свойства, базируясь на имеющейся структурной информации с использованием современных методов интеллектуального анализа данных, а также представления атомов в виде полиэдров Вороного или ван-дер-Ваальсовых полиэдров, кристаллических структур в виде периодических сеток. В рамках данной методологии предложены новые дескрипторы структурных единиц, разработаны машинные модели для прогноза степени окисления атомов металлов, а также размерности и топологии периодических сеток, с точностью прогноза 95%, 90% и 75%, соответственно. Разработанная методология, дескрипторы и машинные модели реализованы в программном обеспечении и веб-сервисах, которые используются во всем мире.
2. В рамках предложенной методологии структурного анализа разработаны и реализованы в виде компьютерной программы три новых алгоритма определения формы координационной фигуры атома или структурной группировки, основанные на топологии полиэдров Вороного, максимизации объема пересечения координационных полиэдров и сравнении угловых фингерпринтов, которые позволили найти ряд правил и закономерностей строения пятикоординационных соединений меди и цинка в N,O-окружении.
3. При помощи разработанных программных средств созданы информационные хранилища структурных единиц и их дескрипторов для 1 928 816 атомов металла, 269 232 лигандов в структуре координационных соединений, 309 353 органических молекул, 1 234 нанокластеров, а также 808 445 топологических типов кристаллических структур и 2 675 446 их представителей. Все обучающие и тестовые выборки для машинного обучения были сформированы на основании

этих данных. Доступ к хранилищу через Интернет реализован в виде веб-сервисов на сайте <https://topcryst.com>.

4. Предложен новый метод оценки пористости кристаллических структур с помощью ван-дер-Ваальсовых полиэдров атомов. Метод использован для изучения взаимосвязей доли свободного пространства в кристаллической структуре с ее топологическим типом в 33 790 каркасных координационных соединениях. Показано, что существуют топологии периодических сеток, которые обеспечивают только низкую или только высокую пористость соответствующих атомных каркасов, в то время как многие другие топологии могут описывать каркасы любой пористости. На основании разнообразия формы полостей атомного каркаса, представленных в виде натуральных тайлов, предложен численный метод оценки гибкости каркаса.
5. Показано, что геометрико-топологические методы являются универсальными и могут быть применены не только для структур координационных соединений, но и других классов химических веществ. На основании предлагаемого подхода осуществлен отбор 402 потенциальных Na-ионных твердых электролитов, а также подтвержден плотноупакованный характер интерметаллидов по результатам анализа коллекции нанокластеров и топологий слоев кристаллических структур. Впервые, на примере структур насыщенных углеводородов, введен ранг контакта, который образует грань в полиэдре Вороного атома и равен количеству валентных связей в кратчайшей цепи, соединяющей центральный атом с атомом окружения. Обнаружены корреляции радиуса сферического домена атомов в простых веществах с длинами электронных волн де Бройля, рассчитанных из потенциалов ионизации изолированных атомов, что подтвердило правомерность модели «мягких сфер», то есть представления атомов в кристаллах и жидкостях, как деформируемых сфер, полностью заполняющих все пространство.

## Библиографический список

1. Groom C. R., Bruno I. J., Lightfoot M. P., Ward S. C. The Cambridge Structural Database. *Acta Cryst.* 2016, B72, 171-179.
2. Belsky A., Hellenbrandt M., Karen V. L. and Luksch P., New Developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in Support of Materials Research and Design. *Acta Crystallogr., Sect. B: Struct. Sci.* 2002, 58, 364-369. <https://icsd.fiz-karlsruhe.de>
3. International Centre for Diffraction Data, 12 Campus Blvd., Newtown Square, PA 19073-3273 U.S.A. База данных PDF-4 доступна по адресу <https://www.icdd.com/pdf-4/>
4. Rose P. W., Prlic A., Bi C. et al. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education, *Nucleic Acids Res.* 2015, 43, D345-56. База данных PDB доступна по адресу <http://www.rcsb.org/pdb/>.
5. Villars P., Cenzual K. *Pearson's Crystal Data: Crystal Structure Database for Inorganic Compounds*, Release 2017/18 (ASM International, Materials Park, OH, 2017); <http://www.crystalimpact.com/pcd/>.
6. Villars P., Cenzual K., Daams J. L. C., Hulliger F., Massalski T. B., Okamoto H., Osaki K., Prince A., Iwata S. (Eds.) PAULING FILE, Binaries Edition, Materials Park: ASM International, 2002; <http://paulingfile.com/>
7. Gražulis S., Daškevič A., Merkys A. et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucl. Acids Res.*, 2012, 40, D420-D427. <http://www.crystallography.net/cod/>.
8. Hall S. R., Allen F. H., Brown I. D. The Crystallographic Information File (CIF): a new standard archive file for crystallography. *Acta Cryst.* 1991, A47(6), 655–685. doi:10.1107/S010876739101067X.

9. Brown I. D., McMahon B. CIF: the computer language of crystallography. *Acta Cryst.* 2002, B58 (Pt 3 Pt 1), 317–324.
10. Taylor R., Wood P. A. A Million Crystal Structures: The Whole Is Greater than the Sum of Its Parts. *Chem. Rev.* 2019, 119(16), 9427-9477. doi: 10.1021/acs.chemrev.9b00155
11. Ziletti A., Kumar D., Scheffler M. et al. Insightful classification of crystal structures using deep learning. *Nat Commun* 2018, 9, Article number: 2775.
12. Ryan K., Lengyel J., Shatruk M. J. Crystal Structure Prediction via Deep Learning. *Am. Chem. Soc.*, 2018, 140, 32, 10158-10168; doi: 10.1021/jacs.8b03913
13. Белов Н. В. Структура ионных кристаллов и металлических фаз. М.: Изд-во АН СССР, 1947. 238 с.
14. Пирсон У. Кристаллохимия и физика металлов и сплавов. Ч. 1 / У. Пирсон; Пер. с англ. С. Н. Горина. – М.: Мир, 1977. - 419 с.
15. Gilbert T. L. Soft-Sphere Model for Closed-Shell Atoms and Ions. *J. Chem. Phys.* 1968, 49, 2640–2642.
16. Lang P. F., Smith B. C. Ionic radii for Group 1 and Group 2 halide, hydride, fluoride, oxide, sulfide, selenide and telluride crystals. *Dalton Transactions.* 2010, 39, 7786–7791.
17. Блатов В. А., Шевченко А. П., Сережкин В. Н. Правило четырнадцати соседей и структура координационных соединений. *Доклады АН.* 1994, 335(6), 742–744.
18. Friedrichs O. D., O'Keeffe M., Yaghi O. M. Three-periodic nets and tilings: semiregular nets. *Acta Cryst.* 2003, A59, 515-525.
19. Bondi A. Van der Waals Volumes and Radii. *J. Phys. Chem.* 1964, 68(3), 441-451. <https://doi.org/10.1021/j100785a001>

20. Блатов В. А., Сережкин В. Н. Стереoaтомная модель структуры неорганических и координационных соединений. *Журн. неорган. химии*. 2000, 45, 105–222.
21. Rahm M., Hoffmann R., Ashcroft N. W. Atomic and Ionic Radii of Elements 1–96. *Chem. Eur. J.* 2016, 22, 14625–14632.
22. Waber J. T., Cromer D. T. Orbital Radii of Atoms and Ions. *J. Chem. Phys.* 1965, 42(12), 4116–4123.
23. Физические величины: справочник. Под ред. И. С. Григорьева, Е. З. Мейлихова. - М.: Энергоатомиздат. 1999. 1232 с.
24. Conway, J. H. & Sloane, N. J. A. (1993). *Sphere Packings, Lattices and Groups.* , 2nd ed. New York: Springer-Verlag.
25. Галиулин Р. В. Кристаллографическая геометрия. «Наука», Москва, 1984, 135 с.
26. Blatov V. A. Voronoi–dirichlet polyhedra in crystal chemistry: theory and applications. *Cryst. Rev.* 2004, 10, 249–318.
27. Blatov V. A., Shevchenko A. P. Analysis of voids in crystal structures: the methods of 'dual' crystal chemistry. *Acta Cryst.* 2003, A59, 34–44.
28. Alinchenko M. G., Anikeenko A. V., Medvedev N. N., Voloshin V. P., Mezei M., Jedlovszky P. Morphology of Voids in Molecular Systems. A Voronoi–Delaunay Analysis of a Simulated DMPC Membrane. *J. Phys. Chem. B* 2004, 108, 19056–19067.
29. Blatov V. A., Ilyushin G. D., Blatova O. A., Anurova N. A., Ivanov-Schits A. K., Dem'yanets L. N. An analysis of migration paths in fast-ion conductors by means of Voronoi-Dirichlet partition. *Acta Cryst.* 2006, B62, 1010–1018.
30. Klepov V. V., Vologzhanina A. V., Alekseev E. V. et al. Structural diversity of uranyl acrylates. *CrystEngComm* 2016, 18, 1723–1731.

31. Boese R., Weiss H.-C., Bläser D. The Melting Point Alternation in the Short-Chain n-Alkanes: Single-Crystal X-Ray Analyses of Propane at 30 K and of n-Butane to n-Nonane at 90 K. *Angew. Chem.* 1999. 38. 988-992.
32. Medvedev N. N., Voloshin V. P., Luchnikov V., Gavrilova M. L. Algorithm for three-dimensional Voronoi S-network. *J. Comput. Chem.* 2006, 27(14), 1676-1692.
33. Anurova N. A., Blatov V. A., Ilyushin G. D., Blatova O. A., Ivanov-Schitz A. K., Dem'yanets L. N. Migration maps of Li<sup>+</sup> cations in oxygen-containing compounds. *Solid State Ionics* 2008, 179, 2248–2254.
34. Medvedev N. N., Voloshin V. P., Naberukhin Yu. I. Local environmental geometry of atoms in the Lennard-Jones systems. *Mater. Chem. Phys.* 1986, 14(6), 533-548.
35. Lyssenko K. A., Vologzhanina A. V., Torubaev Yu. V., Nelyubina Y. V. A comparative study of a mixed-ligand copper(II) complex by the theory of atoms in molecules and the Voronoi tessellation. *Mendeleev Commun.* 2014, 24(4), 216-218.
36. Сережкин В. Н., Михайлов Ю. Н., Буслаев Ю. А. Метод пересекающихся сфер для определения координационного числа атомов в структуре кристаллов. *Журн. неорганич. химии.* 1997, 42(12), 2036-2077.
37. Blatov V. A. A method for topological analysis of rod packings. *Structural Chemistry* 2016, 27(6), 1605–1611.
38. O'Keeffe M., A proposed rigorous definition of coordination number. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* 1979, 35(5), 772-775.
39. Zolotarev P. N., Arshad M. N., Asiri A. M., Alamshany Z. M., Blatov V. A. A Possible Route toward Expert Systems in Supramolecular Chemistry: 2-Periodic H-Bond Patterns in Molecular Crystals. *Cryst Growth Des* 2014, 14, 1938-1949.



40. Slater J. C. Atomic Radii in Crystals. *J. Chem. Phys.* 1964, 41(10), 3199-3204.
41. Chung S. J., Hahn Th., Klee W. E. Nomenclature and generation of three-periodic nets: the vector method. *Acta Crystallogr.* 1984, A40, 1, 42-50.
42. Klee W. E. The topology of crystal structures: Invariants. *Z. Kristallogr.* 1987, 179, 67-76.
43. Блатов В. А., Сережкин В. Н. Метод анализа топологии кристаллической решетки с помощью теории графов. *Кристаллография* 1992, 37(1), 51-62.
44. Serezhkin V. N., Vologzhanina A. V., Serezhkina L. B., Smirnova E. S., Grachova E. V., Ostrovac P. V., Antipin M. Yu. Crystallochemical formula as a tool for describing metal-ligand complexes - a pyridine-2,6-dicarboxylate example. *Acta Cryst.* 2009, B65, 45-53.
45. Tahir A. A., Hamid M., Mazhar M., Zeller M., Hunter A. D. Dioxobis(pentane-2,4-dionato)(tetrahydrofuran)uranium(VI). *Acta Crystallogr., Sect.E: Struct. Rep. Online*, 2006, 62(8), m1780- m1781.
46. Wang D., Yu R., Kumada N., Kinomura N. Hydrothermal Synthesis and Characterization of a Novel One-Dimensional Titanium Glycolate Complex Single Crystal:  $\text{Ti}(\text{OCH}_2\text{CH}_2\text{O})_2$ . *Chem. Mater.*, 1999, 11(8), 2008-2012.
47. Yoneyama S., Kodama T., Kikuchi K. et al. Deuterium substitution effects on the structural and magnetic phase transitions of a hydrogen-bonded coordination polymer, bis(glycolato)copper(ii). *Dalton Trans.* 2016, 45, 16774-16778.
48. Wang X., Samarasekere P., Jacobson A. J. Synthesis and single crystal structures of  $\text{V}(\text{OH})\text{ndc}\cdot\text{H}_2\text{O}$ ,  $\text{V}(\text{OH})\text{ndc}$ , and  $\text{VOndc}$ . *Microporous and Mesoporous Materials* 2018, 267, 20-23.
49. Crowe J. E., Lynch M. F., Town W. G. Analysis of structural characteristics of chemical compounds in a large computer-based file. Part I. Non-cyclic fragments. *J. Chem. Soc. C* 1970, 990-996.

50. Sturm N., Sun J., Vandriessche Y. et al. Application of Bioactivity Profile-Based Fingerprints for Building Machine Learning Models. *J. Chem. Inf. Model.* 2019, 59(3), 962-972.
51. Garcia-Hernandez C., Fernández A., Serratosa F. Ligand-Based Virtual Screening Using Graph Edit Distance as Molecular Similarity Measure. *J. Chem. Inf. Model.* 2019, 59, 4, 1410–1421.
52. Yang S. H. Learning over Molecules: Representations and Kernels, Bachelor's thesis, Harvard College, 2014.
53. Landrum G. RDKit Documentation, Release 2017.09.1.dev1; <https://www.rdkit.org>
54. Бацанов С. С. Ван-дер-Ваальсовы радиусы элементов. *Неорганические материалы* 2001, 37(9), 1031-1046.
55. Уэллс А. Структурная неорганическая химия: В 3-х т. Т. 1: Пер с англ. - М.: Мир, 1987. - 408 с., ил.
56. Ockwig N. W., Delgado-Friedrichs O., O’Keeffe M., and Yaghi O. M. Reticular Chemistry: Occurrence and Taxonomy of Nets and Grammar for the Design of Frameworks. *Acc. Chem. Res.* 2005, 38, 176-182.
57. Alvarez S., Avnir D., Llunellbd M., and Pinskyce M. Continuous symmetry maps and shape classification. The case of six-coordinated metal compoundsy, *New J. Chem.* 2002, 26, 996–1009.
58. Klein H.-J., Mennerich C.: Searching similar clusters of polyhedra in crystallographic databases, *Int. J. on Advances in Systems and Measurements* 2009, (2&3), 192-203.
59. Terrones H., Mackay A. L. The characterisation of coordination polyhedra by invariants. *J. Math. Chem.* 1994, 15, 157-181.
60. Alexandrov E. V., Blatov V. A., Kochetkov A. V., Proserpio D. M. Underlying nets in three-periodic coordination polymers: topology, taxonomy and prediction

from a computer-aided analysis of the Cambridge Structural Database. *CrystEngComm* 2011, 13(12), 3947–3958.

61. Waroquiers D., Gonze X., Rignanese G.-M. et al. Statistical Analysis of Coordination Environments in Oxides. *Chem. Mater.* 2017, 29, 8346-8360.

62. Dass R., Rani R., Kumar D. Face Recognition Techniques: A Review, *Int. J. Eng. Research and Development* 2012, 4, 70-78.

63. Lyakhov A. O., Oganov A. R., Valle M. How to predict very large and complex crystal structures, *Comp. Phys. Comm.* 2010, 181, 1623-1632.

64. Lee J., Shin J., Ko T.-W., Lee S., Chang H., Hyon Y. Descriptors of atoms and structure information for predicting properties of crystalline materials. *Mater. Res. Express* 2021, 8, 026302.

65. Ward L., Liu R., Krishna A., Hegde V. I., Agrawal A., Choudhary A., Wolverton C. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Physical Review B* 2017, 96, 024104.

66. Blatov V. A. Topological analysis of ionic packings in crystal structures of inorganic sulfides: the method of coordination sequences. *Z. Kristallogr.* 2001, 216, 165-171.

67. Peresypkina E. V., Blatov V. A. Search for structure-forming components in molecular crystals of binary compounds: a topological approach. *Z. Kristallogr.* 2002, 217, 91-102.

68. Hoppe R., Köhler J. SCHLEGEL projections and SCHLEGEL diagrams – new ways to describe and discuss solid state compounds. *Z. Kristallogr. B* 1988, 183(1), 77–111.

69. Blatov V. A., O’Keeffe M., Proserpio D. M. Vertex-, face-, point-, Schlafli-, and Delaney-symbols in nets, polyhedra and tilings: recommended terminology. *CrystEngComm* 2010, 12, 44–48.

70. Blatov V. A., Proserpio D. M. Periodic-Graph Approaches in Crystal Structure Prediction. Modern Methods of Crystal Structure Prediction; Oganov, A. R., Ed.; John Wiley & Sons, Ltd: Weinheim, Germany, 2011; pp 1–28.
71. Müller P., Grüner R., Bon V. et al. Topological control of 3,4-connected frameworks based on the Cu<sub>2</sub>-paddle-wheel node: tbo or pto, and why? *CrystEngComm* 2016, 18, 8164-8171.
72. Kalmutzki M. J., Hanikel N., Yaghi O. M. Secondary building units as the turning point in the development of the reticular chemistry of MOFs. *Sci. Adv.* 2018, 4(10).
73. Bonneau C., O’Keeffe M., Proserpio D. M., Blatov V. A., Batten S. R., Bourne S. A., Soo Lah M., Eon J.-G., Hyde S. T., Wiggins S. B., Öhrström L. Deconstruction of Crystalline Networks into Underlying Nets: Relevance for Terminology Guidelines and Crystallographic Databases. *Cryst. Growth Des.* 2018, 18, 3411–3418.
74. Cheng-Peng Li, Hang Zhou, Yue Ju, Miao Du, Water-Mediated Structural Transformations of Cu<sup>II</sup> 5-Halonicotinate Coordination Networks with Distinct Mechanisms. *Chemistry-A European Journal*, 2017, 23, 12985-12990.
75. Hursthouse M. B., Hughes D. S., Gelbrich T. and Threlfall T. L. Describing hydrogen-bonded structures; topology graphs, nodal symbols and connectivity tables, exemplified by five polymorphs of each of sulfathiazole and sulfapyridine. *Chemistry Central Journal* 2015, 9, Article number: 1.
76. Peresypkina E. V., Blatov V. A. Topology of molecular packings in organic crystals. *Acta Cryst.* 2000, B56, 1035-1045.
77. Aman F., Asiri A. M., Siddiqui W. A. et al. Multilevel topological description of molecular packings in 1,2-benzothiazines. *CrystEngComm* 2014, 16, 1963-1970.
78. Rabinovich D. Molecular packing modes. Part I. Crystal and molecular structures of p-chlorocinnamide. *J. Chem. Soc. A* 1969, 2361-2366.

79. Peresypkina E. V., Blatov V. A. Molecular coordination numbers in crystal structures of organic compounds. *Acta Cryst.* 2000, B56, 501-511.
80. Vologzhanina A. V., Ushakov I. E., Korlyukov A. A. Intermolecular Interactions in Crystal Structures of Imatinib-Containing Compounds. *Int. J. Mol. Sci.* 2020, 21, 8970.
81. Yaghi O. M., O’Keeffe M., Ockwig N. W., Chae H. K., Eddaoudi M., Kim J. Reticular synthesis and the design of new materials. *Nature* 2003, 423, 705-714.
82. Yaghi O. M., Kalmutzki M. J. and Diercks C. S. Introduction to Reticular Chemistry: Metal-Organic Frameworks and Covalent Organic Frameworks, Wiley-VCH, Weinheim, Germany, 2019.
83. Kuznetsova E. D., Blatova O. A., Blatov V. A. Predicting new zeolites: a combination of thermodynamic and kinetic factors. *Chem. Mater.* 2018, 30, 2829-2837.
84. Hill A. R., Cubillas P., Gebbie-Rayet J. T., Trueman M., de Bruyn N., al Harthi Z., Pooley R. J. S., Attfield M. P., Blatov V. A., Proserpio D. M., Gale J. D., Akporiaye D., Arstade B., Anderson M. W. CrystalGrower: a generic computer program for Monte Carlo modelling of crystal growth. *Chem. Sci.*, 2021, 12, 1126. DOI: 10.1039/d0sc05017b
85. Lee H., Kim M.-Y., Lee H. Structural and anionic effects of microcrystalline Zn-CPs on 4-nitrophenol sensing performances. *RSC Adv.* 2022, 12, 12957-12966.
86. MacGillivray L. R., Subramanian S., Zaworotko M. J. Interwoven two- and three-dimensional coordination polymers through self-assembly of Cu<sup>I</sup> cations with linear bidentate ligands. *J. Chem. Soc., Chem. Commun.* 1994, 1325-1326.
87. Carlucci L., Ciani G., Proserpio D. M. Polycatenation, polythreading and polyknotting in coordination network chemistry. *Coord. Chem. Rev.* 2003, 246, 247–289.

88. Alexandrov E. V., Blatov V. A., Proserpio D. M. A topological method for classification of entanglements in crystal networks. *Acta Crystallogr., Sect. A: Found. Crystallogr.* 2012, 68, 484-493.
89. Alexandrov E. V., Blatov V. A., Proserpio D. M. Interpenetration of three-periodic networks in crystal structures: Description and classification methods, geometrical-topological conditions of implementation. *J. Struct. Chem.* 2015, 55, 1308-1325.
90. Withersby M. A., Blake A. J., Champness N. R., Cooke P. A., Hubberstey P., Schroder M. Parallel interpenetration in novel herringbone sheets formed by Co(II) and Cd(II) complexes with trans-4,4'-azobis(pyridine). *New J. Chem.* 1999, 23, 573-575.
91. Chatterjee T., Ansary D., Chanthapally A. et al. Elusive Network Topology of a 5-Connected Self-Catenated 3D Coordination Polymer: Featuring Ligand Formation via the In Situ S–S Bond. *Cryst. Growth Des.* 2024, 24, 1, 539–544.
92. Langlet G. A., Figueiredo M. O., Lima-de-Faria J. Determination of spherical voids in layered structures (void program). *J. Appl. Cryst.* 1977, 10, 21-23.
93. Alard P., Wodak S. J. Detection of cavities in a set of interpenetrating spheres. *J. Comp. Chem.* 1991, 12(8), 918-922.
94. McArdle P., Cunningham D. VOID: A PC program for the location and display of voids in crystal structures. *J. Appl. Cryst.* 2000, 33, 993.
95. Медведев Н. Н. Метод Вороного-Делоне в исследовании структуры некристаллических систем. Новосибирск: СО РАН НИЦ ОИГТМ, 2000. – 209 с.
96. Barbour L. J. Crystal porosity and the burden of proof. *Chem. Commun.* 2006, 1163-1168. DOI:10.1039/B515612M
97. Willems T. F., Rycroft C. H., Kazi M., Meza J. C., Haranczyk M. Algorithms and tools for high-throughput geometry- based analysis of crystalline porous

- materials. *Microporous and Mesoporous Materials* 2012, 149, 134-141.  
<http://www.maciejharanczyk.info/Zeopp/about.html>
98. Li Y., Yu J. New Stories of Zeolite Structures: Their Descriptions, Determinations, Predictions, and Evaluations. *Chem. Rev.* 2014, 114 (14), 7268–7316.
99. Xiang Z., Cao D., Dai L. Well-defined two dimensional covalent organic polymers: rational design, controlled syntheses, and potential applications. *Polym. Chem.* 2015, 6, 1896–1911.
100. Lu W., Yuan D., Zhao D. et al. Porous Polymer Networks: Synthesis, Porosity, and Applications in Gas Storage/Separation. *Chem. Mater.* 2010, 22, 5964–5972.
101. Evans J. D., Huang D. M., Haranczyk M., Thornton A. W., Sumbly C. J., Doonan C. J. Computational identification of organic porous molecular crystals. *CrystEngComm*, 2016, 18, 4133-4141.
102. Evans J. D., Jelfs K. E., Day G. M., Doonan C. J. Application of computational methods to the design and characterisation of porous molecular materials. *Chem. Soc. Rev.* 2017, 46, 3286-3301.
103. O’Keeffe M., Yaghi O. M. Deconstructing the Crystal Structures of MetalOrganic Frameworks and Related Materials into Their Underlying Nets. *Chem. Rev.* 2012, 112, 675–702.
104. Cook T. R., Zheng Y.-R., and Stang P. J. Metal–Organic Frameworks and Self-Assembled Supramolecular Coordination Complexes: Comparing and Contrasting the Design, Synthesis, and Functionality of Metal–Organic Materials. *Chem. Rev.* 2013, 113, 734–777.
105. Halder R., Sikdar N., and Maji T. K. Interpenetration in coordination polymers: structural diversities toward porous functional materials. *Materials Today* 2015, 18(2), 97-116.

106. Jianga H.-L., Makal T. A., Zhou H.-C. Interpenetration control in metal–organic frameworks for functional applications. *Coord. Chem. Rev.* 2013, 257, 2232–2249.
107. Niekel F., Lannoeye J., Reinsch H., Munn A. S. et al. Conformation-Controlled Sorption Properties and Breathing of the Aliphatic Al-MOF [Al(OH)(CDC)]. *Inorg. Chem.* 2014, 53 (9), pp 4610–4620.
108. Blatova O. A., Golov A. A., Blatov V. A. Natural tilings and free space in zeolites: models, statistics, correlations, prediction. *Z. Kristallogr.* 2019, 234 (7-8), 421–436.
109. Wei Q., Nieuwenhuyzen M., Meunier F., Hardacre C., James S. L. Guest sorption and desorption in the metal–organic framework [Co(INA)<sub>2</sub>] (INA = isonicotinate) – evidence of intermediate phases during desorption. *Dalton Transactions* 2004, 1807–1811.
110. Delgado-Friedrichs O., O’Keeffe M., Yaghi O. M. Taxonomy of periodic nets and the design of materials. *Phys. Chem. Chem. Phys.* 2007, 9, 1035–1043.
111. Motherwell W. D. S. The CSD–450,000 answers ... but what are the questions? *Cryst. Rev.*, 2008, 14, 97–116.
112. Бушмаринов И. С., Лысенко К. А., Антипин М. Ю. Энергия атомов в теории "атомы в молекулах" и ее использование для решения химических задач. *Успехи химии* 2009, 78, 307–327.
113. Bruno I. J., Cole J. C., Kessler M., Luo J., Motherwell W. D. S., Purkis L. H., Smith B. R., Taylor R., Cooper R. I., Harris S. E., Guy Orpen A. Retrieval of Crystallographically-Derived Molecular Geometry Information. *J. Chem. Inf. Comput. Sci.* 2004, 44, 6, 2133–2144. doi: 10.1021/ci049780b
114. Blatov V. A., Alexandrov E. V., Shevchenko A. P. Topology: ToposPro. In *Comprehensive Coordination Chemistry (Reference Module in Chemistry, Molecular Sciences and Chemical Engineering)*, 2019.



115. Kong C. S., Luo W., Arapan S. et al. Information-Theoretic Approach for the Discovery of Design Rules for Crystal Chemistry. *J. Chem. Inf. Model* 2012, 52, 7, 1812-1820.
116. Cerqueira T. F. T., Sarmiento-Pérez R., Amsler M., Nogueira F., Botti S., Marques M. A. L. Materials Design On-the-Fly. *J. Chem. Theory Comput.* 2015, 118, 3955-3960.
117. Eremin R. A., Zolotarev P. N., Golov A. A., Nekrasova N. A. and Leisegang T. Ionic Transport in Doped Solid Electrolytes by Means of DFT Modeling and ML Approaches: A Case Study of Ti-Doped KFeO<sub>2</sub>. *J. Phys. Chem. C* 2019, 123(49), 29533.
118. Bialon A. F., Hammerschmidt T., Drautz R. Three-Parameter Crystal-Structure Prediction for sp-d-Valent Compounds. *Chem. Mater.* 2016, 28, 2550-2556. doi: 10.1021/acs.chemmater.5b04299
119. J. Graser, S. K. Kauwe, T. D. Sparks, Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons. *Chem. Mater.* 2018, 30, 11, 3601-3612. doi: 10.1021/acs.chemmater.7b05304
120. Oliynyk A. O., Adutwum L. A., Harynuk J. J., Mar A. Classifying Crystal Structures of Binary Compounds AB through Cluster Resolution Feature Selection and Support Vector Machine Analysis. *Chem. Mater.* 2016, 28, 18, 6672-6681.
121. Jablonka K. M., Ongari D., Moosavi S. M., Smit B. Using collective knowledge to assign oxidation states. *Nat. Chem.* 2021, 13, 771-777. <https://oximachine.materialscloud.io/>
122. Tan J. C., Cheetham A. K. Mechanical properties of hybrid inorganic–organic framework materials: establishing fundamental structure–property relationships. *Chem. Soc. Rev.* 2011, 40, 1059-1080.

123. Fernandez M., Boyd P., Daff T., Aghaji M. Z., Woo T. K. Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO<sub>2</sub> Capture. *J. Phys. Chem. Lett. Submitted*. 2014, 3056-3060.
124. Moosvai S. M., Chidambaram A., Talirz L., Haranczyk M., Stylianou K. C., Smit B. Capturing chemical intuition in synthesis of metal-organic frameworks. *Nat. Commun.* 2019, 10, Article number: 539.
125. Kitchin J. R. Machine learning in catalysis. *Nature Catalysis* 2018, 1, 230–232.
126. Li H., Zhang Z., Liu Z. Application of Artificial Neural Networks for Catalysis: A Review. *Catalysts* 2017, 7, 306.
127. Meyer B., Sawatlon B., Heinen S., von Lilienfeld O. A., Corminboeuf C. Machine learning meets volcano plots: computational discovery of cross-coupling catalysts. *Chem. Sci.* 2018, 9, 35, 7069-7077.
128. Eremin R. A., Kabanova N. A., Morkhova Ye. A., Golov A. A., Blatov V. A. High-throughput search for potential potassium ion conductors: A combination of geometrical-topological and density functional theory approaches. *Solid State Ionic* 2018, 326, 188-199.
129. Нейлор К. Как построить свою экспертную систему: Пер. с англ. – М.: Энергоатомиздат, 1991. – 286 с.
130. Jackson P. Introduction to Expert Systems, Addison-Wesley, 3rd edn, 1999.
131. Hemmer M. C. Expert Systems in Chemistry Research, CRC Press, Boca Raton, FL, 2008.
132. Judson P. Knowledge-Based Expert Systems in Chemistry: Not Counting on Computers, RSC Publishing, Cambridge, UK, 2009.
133. ASM Engineered Materials Reference Book, Second Edition, Michael Bauccio, Ed. ASM International, Materials Park, OH, 1994.

134. Hicks D., Toher C., Ford D. C. et al. AFLOW-XtalFinder: a reliable choice to identify crystalline prototypes. *npj Comput. Mater.* 2021, 7, 30.
135. Bradlyn B., Elcoro L., Cano J. et al. Topological quantum chemistry. *Nature* 2017, 547, 298-305.
136. O'Keeffe M., Peskov M. A., Ramsden S. J., Yaghi O. M. The Reticular Chemistry Structure Resource (RCSR) database of, and symbols for, crystal nets. *Acc. Chem. Res.* 2008, 41, 1782-1789. База данных RCSR доступна по адресу <http://rcsr.anu.edu.au>.
137. Ramsden S. J., Robins V., Hyde S. T. *Acta Cryst.* 2009, A65, 81. Database of Zeolite Structures, Structure Commission of the International Zeolite Association, <http://www.iza-structure.org/databases/>.
138. Hu J., Stefanov S., Song Y. et al. MaterialsAtlas.org: a materials informatics web app platform for materials discovery and survey of state-of-the-art. *npj Comput. Mater.* 2022, 8, Article number: 65.
139. Pétuya R., Durdy S., Antypov D. et al. Machine-Learning Prediction of Metal–Organic Framework Guest Accessibility from Linker and Metal Chemistry. *Angew. Chem. Int. Ed.* 2022, 61, 1-6, e202114573.
140. García A. G. and Haranczyk M. Toward crystalline porosity estimators for porous molecules. *CrystEngComm* 2020, 22, 7242-7251. DOI: 10.1039/C9CE01753D.
141. Koch E., Fischer W., Sowa H. Interpenetration of homogeneous sphere packings and of two-periodic layers of spheres. *Acta Crystallogr.* 2006, A62, 152–167.
142. Wilmer C. E., Kim K. C., Snurr R. Q. An Extended Charge Equilibration Method. *J. Phys. Chem. Lett.* 2012, 3(17), 2506-2511.
143. Lin L.-C., Berger A. H., Martin R. L. et al. In silico screening of carbon-capture materials. *Nat. Mater.* 2012, 11, 633-641.

144. Goldsmith J., Wong-Foy A. G., Cafarella M. J., Siegel D. J. Theoretical Limits of Hydrogen Storage in Metal–Organic Frameworks: Opportunities and Trade-Offs. *Chem. Mater.* 2013, 25, 3373-3382.
145. Chung Y. G., Haldoupis E., Bucior B. J. et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* 2019, 64, 5985-5998.
146. First E. L., Gounaris C. E., Wei J., Floudas C. A. Computational characterization of zeolite porous networks: an automated approach. *Phys. Chem. Chem. Phys.* 2011, 13, 17339-17358.
147. First E. L., Floudas C. A. MOFomics: Computational pore characterization of metal–organic frameworks. *Microporous Mesoporous Mater.* 2013, 165, 32-39.
148. Moghadam P. Z., Li A., Wiggin S. B., Tao A. et al. Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future. *Chem. Mater.* 2017, 29, 2618.
149. Kumar N., Rajagopalan P., Pankajakshan P., Bhattacharyya A., Sanyal S., Balachandran J., Waghmare U. V. Machine Learning Constrained with Dimensional Analysis and Scaling Laws: Simple, Transferable, and Interpretable Models of Materials from Small Datasets. *Chem. Mater.* 2019, 31, 314-321. doi: 10.1021/acs.chemmater.8b02837
150. Mitina T. G., Blatov V. A. Topology of 2-Periodic Coordination Networks: Toward Expert Systems in Crystal Design. *Cryst. Growth Des.*, 2013, 13, 1655.
151. Корлюков А. А., Антипин М. Ю. Исследование строения кристаллов органических и элементоорганических соединений с помощью современных квантово-химических расчетов в рамках теории функционала плотности. *Успехи химии* 2012, 81, 105–129.
152. Nandy A., Duan C., Taylor M. G., Liu F., Steeves A. H., Kulik H. J. Computational Discovery of Transition-metal Complexes: From High-throughput Screening to Machine Learning. *Chem. Rev.* 2021, 121, 16, 9927–10000.

153. Li Pengfei, Merz K. M. Jr. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* 2017, 117, 1564–1686.
154. Rowland R. S., Taylor R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *J. Phys. Chem.* 1996. V.100. P.7384-7391.
155. Baur W. H. Variation of mean Si-O bond lengths in silicon-oxygen tetrahedra. *Acta Cryst.* 1978, B34, 1751-1756.
156. Кулешова Л. Н., Антипин М. Ю. Кембриджский банк структурных данных как инструмент изучения общих закономерностей строения органических молекулярных кристаллов. *Успехи химии* 1999. Т.68. №1. С.3.
157. Nyburg S. C., Faerman C. H. A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br and I bonded to carbon. *Acta Cryst.* 1985, B41, 274-279.
158. Киселев Ю. М., Добрынина Н. А. "Химия координационных соединений" Издательство М.: "Академия", 2007. - 352 с.
159. Davies D. W., Butler K. T., Isayev O., Walsh A. Materials discovery by chemical analogy: role of oxidation states in structure prediction, *Faraday Discuss.* 2018, 211, 553-568.
160. Greenwood N. N., Earnshaw A. Chemistry of the elements. — 2-е изд. — Oxford: Butterworth-Heinemann, 1997. — С. 28. — ISBN 0080379419.
161. Langmuir I. The arrangement of electrons in atoms and molecules. *J. Am. Chem. Soc.* 1919, 41, 868-934.
162. Zuo C.-Y., Lu Z.-Y., Bai J.-F., Zheng B.-S., Li Y.-Z. Systematic Investigation of the Influences of Solvent and Temperature on the Formation of Three Co-Cluster-Based Metal-Organic Frameworks: Structures and Gas Adsorption Properties. *Chin. J. Inorg. Chem.* 2012, 28, 1799–1808.

163. Farrugia L. J., Evans C. Experimental X-ray Charge Density Studies on the Binary Carbonyls  $\text{Cr}(\text{CO})_6$ ,  $\text{Fe}(\text{CO})_5$ , and  $\text{Ni}(\text{CO})_4$ . *J. Phys. Chem.* 2005, 109, 8834-8848.
164. Huang R. H., Huang S. Z., Dye J. L. Syntheses and structures of six compounds that contain the sodium anion. *J. Coord. Chem.* 1998, 46, 13-31.
165. Banger S., Nayak V., Verma U.P. Hydrogen storage in lithium hydride: A theoretical approach. *Journal of Physics and Chemistry of Solids* 2018, 115, 6-17, ICSD RefCode: 680196.
166. Wang J., Hanzawa K., Hiramatsu H., Kim J., Umezawa N., Iwanaka K., Tada T., Hosono H. Exploration of Stable Strontium Phosphide-Based Electrides: Theoretical Structure Prediction and Experimental Validation. *J. Am. Chem. Soc.* 2017, 139(44), 15668-15680.
167. Vajenine G. V., Simon A.  $\text{NaBa}_2\text{O}$ : A Fresh Perspective in Suboxide Chemistry. *Angew. Chem. Int. Ed. Engl.* 2001, 40(22), 4220-4222.
168. Junjie Wang, Qiang Zhu, Zhenhai Wang, Hosono H. Ternary inorganic electrides with mixed bonding. *Phys. Rev.* 2019, B99, 064104(1-10).
169. Ganose A., Jain A. Robocrystallographer: Automated crystal structure text descriptions and analysis. *MRS Communications* 2019, 9(3), 874-881.
170. Reeves M. G., Wood P. A., Parsons S. Automated oxidation-state assignment for metal sites in coordination complexes in the Cambridge Structural Database. *Acta Cryst.* 2019, B75, 1096-1105.
171. Mellot-Draznieks C., Newsam J. M., Gorman A. M. et al. De Novo Prediction of Inorganic Structures Developed through Automated Assembly of Secondary Building Units (AASBU Method). *Angew. Chem. Int. Ed.* 2000, 39, 2270-2275.
172. Mellot-Draznieks C., Dutour J., Férey G. Hybrid Organic–Inorganic Frameworks: Routes for Computational Design and Structure Prediction. *Angew. Chem. Int. Ed.* 2004, 43, 6290-6296.

173. Foster M. D., Delgado Friedrichs O., Bell R. G. et al. Structural Evaluation of Systematically Enumerated Hypothetical Uninodal Zeolites. *Angew. Chem. Int. Ed.* 2003, 42, 3896-3899.
174. Férey G., Serre C., Mellot-Draznieks C. et al. A Hybrid Solid with Giant Pores Prepared by a Combination of Targeted Chemistry, Simulation, and Powder Diffraction. *Angew. Chem. Int. Ed.* 2004, 43, 6296-6301.
175. Férey G. A chromium terephthalate-based solid with unusually large pore volumes and surface area. *Science* 2005, 309, 2040-2042.
176. Baburin I. A., Leoni S., Seifert G. Enumeration of not-yet-synthesized zeolitic zinc imidazolate MOF networks: a topological and DFT approach. *J. Phys. Chem. B* 2008, 112, 9437-9443.
177. Trouselet F., Boutin A., Coudert F.-X. Novel porous polymorphs of zinc cyanide with rich thermal and mechanical behavior. *Chem. Mater.* 2015, 27, 4422-4430.
178. Bureekaew S., Schmid R. Hypothetical 3D-periodic covalent organic frameworks: exploring the possibilities by a first principles derived force field. *CrystEngComm* 2013, 15, 1551-1562.
179. Eddaoudi M., Kim J., Rosi N. et al. Systematic Design of Pore Size and Functionality in Isorecticular MOFs and Their Application in Methane Storage. *Science* 2002, 295, 469-472.
180. Yaghi O. M., Li H., Eddaoudi M., O’Keeffe M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* 1999, 402, 276-279.
181. Surblé S., Serre C., Mellot-Draznieks C., Millange F., Férey G. A new isorecticular class of metal-organic-frameworks with the MIL-88 topology. *Chem. Commun.* 2006, 284-286.

182. Martin R. L., Lin L.-C., Jariwala K., Smit B., Haranczyk M. Mail-Order Metal–Organic Frameworks (MOFs): Designing Isorecticular MOF-5 Analogues Comprising Commercially Available Organic Molecules. *J. Phys. Chem. C* 2013, 117, 12159-12167.
183. Farha O. K., Yazaydın A. Ö., Eryazici I. et al. De novo synthesis of a metal-organic framework material featuring ultrahigh surface area and gas storage capacities. *Nat. Chem.* 2010, 2, 944-948.
184. Gomez-Gualdron D. A., Gutov O. V., Krungleviciute V. et al. Computational Design of Metal–Organic Frameworks Based on Stable Zirconium Building Units for Storage and Delivery of Methane. *Chem. Mater.* 2014, 26, 5632-5639.
185. Bao Y., Martin R. L., Simon C. M. et al. In silico discovery of high deliverable capacity metal organic frameworks. *J. Phys. Chem. C* 2015, 119, 186-195.
186. Schoedel A., Li M., Li D., O’Keeffe M., Yaghi O. M. Structures of Metal-Organic Frameworks with Rod Secondary Building Units. *Chem. Rev.* 2016, 116, 12466-12535.
187. Boyd P. G., Woo T. K. A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. *CrystEngComm* 2016, 18, 3777-3792.
188. Addicoat M. A., Coupry D. E., Heine T. AuToGraFS: Automatic topological generator for framework structures. *J. Phys. Chem. A* 2014, 118, 9607–9614.
189. Henle E. A., Gantzler N., Thallapally P. K. et al. PoreMatMod.jl: Julia Package for in Silico Postsynthetic Modification of Crystal Structure Models. *J. Chem. Inf. Model* 2022, 62, 423-432.
190. Turcani L., Tarzia A., Szczypiński F. T., Jelfs K. E. stk: An extendable Python framework for automated molecular and supramolecular structure assembly and discovery. *J. Chem. Phys.* 2021, 154, 214102.



191. Китайгородский А. И. Органическая кристаллохимия. – М.: Акад. Наук СССР, 1955, 561 с.
192. Китайгородский А. И. Молекулярные кристаллы. – М.: Наука, 1971, 424 с.
193. Belsky V. K., Zorkaya O. N., Zorky P. M. Structural classes and space groups of organic homomolecular crystals: new statistical data. *Acta Crystallogr., Sect. A: Found. Crystallogr.* 1995, A51, 4, 473-481.
194. Belov N. V., Ocherki po strukturnoi mineralogii (Essay on Structural Mineralogy), Moscow: Nedra, 1976. *Russian Journal Of Coordination Chemistry*, Vol. 25, No. 7, 1999
195. Borisov S. V., Podberezskaya N. V. Stabil'nye kationnye karkasy v strukturakh ftoridov i oksidov (Stable Cationic Frameworks in Structures of Fluorides and Oxides), Novosibirsk: Nauka, 1984.
196. Aslanov L. A. Crystal-chemical model of atomic interactions. 1. The cubic system. *Acta Crystallogr., Sect. B: Struct. Sci.* 1988, vol. 44, p. 449-458.
197. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* 2011, 12, 2825–2830.
198. Bruce P. and Bruce A. Practical Statistics for Data Scientists: 50 Essential Concepts. O'Reilly, 2017, 318 p.
199. Препарата Ф., Шеймос М. Вычислительная геометрия: Введение: Пер. с англ. мо. М.: Мир, 1989. - 478 с.
200. Блатов В. А., Шевченко А. П., Сережкин В. Н. Автоматизация кристаллохимического анализа - комплекс компьютерных программ TOPOS. *Координац. химия* 1999. Т.25. № 7. С.483-497.
201. Anipa V., Tarzia A., Jelfs K. E., Alexandrov E. V., Addicoat M. A. Pore topology analysis in porous molecular systems. *R. Soc. open sci.* 2023, 10:220813. <http://doi.org/10.1098/rsos.220813>

202. Jieren Shao, Jingfeng Zheng, Lei Qin, Songwei Zhang, Yang Ren, Yiying Wu.  $\text{K}_3\text{SbS}_4$  as a Potassium Superionic Conductor with Low Activation Energy for K–S Batteries. *Angew. Chem.* 2022, e202200606 (1 of 6).
203. Xiaonan Gao, Ai-Yun Fu, Bo Liu, Jun-cheng Jin, Long-Tao Dou, Li-Xia Chen. Unique Topology Analysis by ToposPro for a Metal–Organic Framework with Multiple Coordination Centers. *Inorg. Chem.* 2019, 58, 5, 3099–3106.
204. Hughes D. S., Delori A., Rehman A., Jones W. Using crystallography, topology and graph set analysis for the description of the hydrogen bond network of triamterene: a rational approach to solid form selection. *Chemistry Central Journal* 2017, 11, Article number: 63.
205. Hartshorn R. M., Hey-Hawkins E., Kalio R., Leigh G. J. Representation of configuration in coordination polyhedra and the extension of current methodology to coordination numbers greater than six (IUPAC Technical Report). *Pure Appl. Chem.* 2007, 79, 1779–1799.
206. Python: A dynamic, open-source programming language. Python Software Foundation; <https://www.python.org>
207. Billetter H., Hohn F., Pantenburg I., Ruschewitz U.  $[\text{Cu}\{\text{C}_2(\text{COO})_2\}(\text{H}_2\text{O})_3]\cdot\text{H}_2\text{O}$ , the First Copper Complex of Acetylenedicarboxylic Acid. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* 2003, 59, m130–m131.
208. Dai F., He H., Sun D. A metal – organic nanotube exhibiting reversible adsorption of  $(\text{H}_2\text{O})_{12}$  cluster. *J. Am. Chem. Soc.* 2008, 130, 43, 14064–14065.
209. Ahn H.-K., Cheng S.-W., Kweon H. J., Yon J. Overlap of convex polytopes under rigid motion, *Computational Geometry: Theory and Applications* 2014, 47, 15–24.
210. Vigneron A. Geometric optimization and sums of algebraic functions, in Proceedings of the 21<sup>st</sup> Annual ACM-SIAM Symposium on Discrete Algorithms, 2010, pp. 906-917.

211. Gill P. E., Murray W., Wright M. H. Practical optimization; Academic Press. London. 1981.
212. Alvarez S. A Cartography of the Van der Waals Territory, *Dalton Trans.* 2013, 42, 8617-8636.
213. Smith, J. V., Dowell, L. G. Revised crystal structure of dehydrated Na-type A-zeolite. *Z. Kristallogr.* 1968, 126, 135-142.
214. Labbé P., Goreaud M., Reveau B., Monier J. C. Etude comparative des structures  $M_xWO_3$  de type bronze hexagonal. I. Analyse structurale des bronzes de composition  $M_{0,30}WO_3$ . Stéréochimie des éléments  $M = Rb^I$ ,  $Tl^I$  et  $In^I$ . J. C. *Acta Cryst.* 1978, B34, 1433-1438.
215. Blatov V. A., Shevchenko A. P., Proserpio D. M. Applied Topological Analysis of Crystal Structures with the Program Package ToposPro. *Cryst. Growth Des.* 2014, 14, 3576–3586.
216. Çamkıran J., Parsch F., Hibbard G. D. On the topology of the space of coordination geometries. *Eur. Phys. J.* 2023, B96, 72.
217. Versteeg R. B. et al. Optically probed symmetry breaking in the chiral magnet  $Cu_2OSeO_3$  *Phys. Rev. B* 94, 094409.
218. Liu D.-S., Yu J.-Z., Chen W.-T. et al. Synthesis, structural and magnetic investigation of copper(II) coordination polymer based on 5-amino-tetrazolate and isonicotinate mixed ligands. *Inorganica Chim. Acta* 2014, 410, 131-135.
219. Burtch N. C., Jasuja H., Dubbeldam D., Walton K. S. Molecular-level Insight into Unusual Low Pressure  $CO_2$  Affinity in Pillared Metal–Organic Frameworks. *J. Am. Chem. Soc.* 2013, 135, 19, 7172–7180.
220. Van Krevelen D. Graphical-statistical method for the study of structure and reaction process of coal. *Fuel* 1950, 29, 269–284.

221. Sleighter R. L., Hatcher P. G. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *J. Mass Spectrom.* 2007, 42, 559-574.
222. Deng H., Grunder S., Cordova K. E. et al. Large-pore apertures in a series of metal-organic frameworks. *Science* 2012. 336. 1018-1023.
223. Farfan R. A., Espindola J. A., Gomez M. I., de Jimenez M. C. L., Piro O. E., Castellano E. E., Martinez M. A. Crystal structure, spectroscopic and thermal properties of  $[\text{Zn}(\text{Lap})_2(\text{DMF})(\text{H}_2\text{O})]$  and isomorphous  $[\text{M}(\text{Lap})_2]_n$  (M: Cd, Mn) complexes. *J. Mol. Struct.* 2015, 1087, 80-87.
224. Paredes-Garcia V., Rojas I., Madrid R. et al. Structural and magnetic characterization of the tridimensional network  $[\text{Fe}(\text{HCO}_2)_3]_n \cdot n\text{HCO}_2\text{H}$ . *New J. Chem.* 2013, 37, 2120.
225. Hazari D., Jana S. K., Fleck M., Zangrando E., Dalai S. Two three-dimensional coordination polymers of lead(II) with iminodiacetate and naphthalene-dicarboxylate anions: Synthesis, characterization and luminescence behavior. *J. Solid State Chem.*, 2014, 219, 43-48.
226. Jingui D., Higuchi M., Horike S. et al. High  $\text{CO}_2/\text{CH}_4$  and  $\text{C}_2$  Hydrocarbons/ $\text{CH}_4$  Selectivity in a Chemically Robust Porous Coordination Polymer. *Adv. Funct. Mater.* 2013, 23, 3525-3530.
227. Jie Su, Yingxia Wang, Sihai Yang, Guobao Li, Fuhui Liao, Jianhua Lin. New series of indium formates: hydrothermal synthesis, structure and coordination modes. *Inorg. Chem.* 2007, 46, 8403-8409.
228. Lundvall F., Vajeeston P., Wragg D. S., Dietzel P. D. C., Fjellvag H. Two New Series of Coordination Polymers and Evaluation of Their Properties by Density Functional Theory. *Cryst. Growth Des.* 2016, 16, 339-346.
229. Koch E., Fischer W. Types of sphere packings for crystallographic point groups, rod groups and layer groups. *Z. Kristallogr.* 1978, 148, 107-152.

230. Virovets A. V., Naumov D. Yu., Boldyreva E. V., Podberezskaya N. V. Structure of lead(II) oxalate dehydrate. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* 1993, 49, 1882.
231. Lan Qin, Lei Yu, Xiao-Zhi Yan, Wen-Na Zhao, Lei Han. Two sinusoidal wavelike 2D coordination polymers with benzophenone-4,4'-dicarboxylate. *Inorg. Chim. Acta* 2014, 409, 233-237.
232. Amombo Noa F. M., Grape E. S., Brülls S. M., Cheung O., Malmberg P., Ken Inge A., McKenzie C. J., Mårtensson J., Öhrström L. Metal–Organic Frameworks with Hexakis(4-carboxyphenyl)benzene: Extensions to Reticular Chemistry and Introducing Foldable Nets. *J. Am. Chem. Soc.* 2020, 142(20), 9471–9481.
233. Dong X.-Y., Si C.-D., Fan Y. et al. Effect of N-Donor Ligands and Metal Ions on the Coordination Polymers Based on a Semirigid Carboxylic Acid Ligand: Structures Analysis, Magnetic Properties, and Photoluminescence. *Cryst. Growth Des.* 2016, 16, 2062.
234. Barthel S., Alexandrov E. V., Proserpio D. M., Smit B. Distinguishing Metal–Organic Frameworks. *Cryst. Growth Des.* 2018, 18, 1738-1747.
235. Alexandrov E. V., Goltsev A. V., Eremin R. A., Blatov V. A. Anisotropy of Elastic Properties of Metal-Organic Frameworks and the Breathing Phenomenon. *J. Phys. Chem. C* 2019, 123, 24651–24658.
236. Malaestean I. L., Ellern A., Baca S., Kogerler P. Cerium oxide nanoclusters: commensurate with concepts of polyoxometalate chemistry? *Chem. Comm.* 2012, 48, 1499-1501.
237. Ling J., Wallace C. M., Szymanowski J. E. S., Burns P. C. Hybrid uranium-oxalate fullerene topology cage clusters. *Angew. Chem., Int.Ed.* 2010, 49, 7271-7273.
238. Yue W., Kobayashi A., Halder G. J., Peterson V. K., Chapman K. W., Lock N., Southon P. D., Kepert C. Negative thermal expansion in the metal-organic

framework material  $\text{Cu}_3(1,3,5\text{-benzenetricarboxylate})_2$ . *J. Angew. Chem., Int. Ed.* 2008, 47, 8929-8932.

239. Ke Y., Collins D. J., Zhou, H.-C. Synthesis and Structure of Cuboctahedral and Anticuboctahedral Cages Containing 12 Quadruply Bonded Dimolybdenum Units. *Inorg. Chem.* 2005, 44, 4154-4156.

240. <https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Correlation.pdf>

241. Lorenz V., Liebing P., Suta M. et al. Synthesis, structure, complexation, and luminescence properties of the first metal-organic curcumin compound bis(4-triphenylsiloxy)curcumin. *J. Lumin.* 2019, 211, 243-250.

242. Brenner N., Sperling J. M., Poe T. N. et al. Trivalent f-Element Squarates, Squarate-Oxalates, and Cationic Materials, and the Determination of the Nine-Coordinate Ionic Radius of Cf(III). *Inorg. Chem.* 2020, 59, 9384-9395.

243. Lal G., Gelfand B. S., Lin J.-B., Banerjee A., Trudel S., Shimizu G. K. H. Three Sequential Hydrolysis Products of the Ubiquitous  $\text{Cu}_{24}$  Isophthalate Metal–Organic Polyhedra. *Inorg. Chem.* 2019, 58, 9874-9881.

244. Wang K., He Y., Zhao Y., Ma P., Wang J. A propionate-functionalized polyoxovanadate  $\text{K}_2[\text{V}_{10}\text{O}_{16}(\text{OH})_6(\text{CH}_3\text{CH}_2\text{CO}_2)_6]\cdot 20\text{H}_2\text{O}$ : As catalyst for degradation of methylene blue. *J. Mol. Struct.* 2019, 1195, 184-188.

245. Liu H., Guo Z., Lv H., Liu X., Che Y., Mei Y., Bai R., Chi Y., Xing H. Visible-light-driven self-coupling and oxidative dehydrogenation of amines to imines *via* a Mn(II)-based coordination polymer. *Inorg. Chem. Front.* 2020, 7, 1016-1025.

246. Huh D. N., Roy S., Ziller J. W., Furche F., Evans W. J. Isolation of a Square-Planar Th(III) Complex: Synthesis and Structure of  $[\text{Th}(\text{OC}_6\text{H}_2^t\text{Bu}_2\text{-2,6-Me-4})_4]^{1-}$ . *J. Am. Chem. Soc.* 2019, 141, 12458-12463.

247. Yao Z.-Y., Qian Y., Ren Q., Zhang G.-Q., Zhang J., Liu W.-L., Luo H.-B., Ren X.-M. Order–disorder transformation of intercalated cations triggering huge negative thermal expansion, switchable dielectrics and ion conduction near room temperature in a 2D vanadium oxide hybrid. *J. Mater. Chem. C* 2019, 7, 13243–13252.
248. Zheng Zhiling, Zhang O., Borgs C., Chayes J. T., Yaghi O. M. ChatGPT Chemistry Assistant for Text Mining and Prediction of MOF Synthesis. *J. Am. Chem. Soc.* 2023, 145, 32, 18048–18062.
249. Knope K. E., Soderholm L. Solution and Solid-State Structural Chemistry of Actinide Hydrates and Their Hydrolysis and Condensation Products. *Chem. Rev.* 2013, 113, 944–994.
250. Baes C. F., Mesmer, R. E. The Hydrolysis of Cations; John Wiley and Sons: New York, NY, 1976.
251. Zhao R., Mei L., Hu K. Q., Tian M., Chai Z. F., Shi W. Q. Bimetallic Uranyl Organic Frameworks Supported by Transition-Metal-Ion-Based Metalloligand Motifs: Synthesis, Structure Diversity, and Luminescence Properties. *Inorg. Chem.* 2018, 57(10), 6084–6094.
252. Yang Z., Chen N., Wang C., Yan L., Li G. Syntheses, Crystal Structures, and Properties of Four Complexes Constructed From 2-Propyl-1H-Imidazole-4,5-Dicarboxylic Acid. *Synth. React. Inorg., Met.- Org., Nano-Met. Chem.* 2012, 42(3), 336–344.
253. Thuéry P., Harrowfield J. Chiral one- to three-dimensional uranyl–organic assemblies from (1R,3S)-(+)-camphoric acid. *CrystEngComm* 2014, 16, 2996–3004.
254. Mei L., Wu Q.-Y., An S.-W., Gao Z.-Q., Chai Z.-F., Shi W.-Q. Silver Ion-Mediated Heterometallic Three-Fold Interpenetrating Uranyl-Organic Framework. *Inorg. Chem.* 2015, 54, 10934–10945.

255. Cahill C. L., de Lill D. T., Frisch M. Homo- and heterometallic coordination polymers from the f-elements. *CrystEngComm* 2007, 9, 15–26.
256. Thuéry P., Harrowfield J. Recent advances in structural studies of heterometallic uranyl-containing coordination polymers and polynuclear closed species. *Dalton Trans.* 2017, 46, 13660–13667.
257. Cheetham A. K., Rao C. N. R., Feller R. K. Structural diversity and chemical trends in hybrid inorganic–organic framework materials. *Chem. Commun.* 2006, 4780–4795.
258. Li P., Vermeulen N. A., Malliakas C. D. et al. Bottom-up construction of a superstructure in a porous uranium-organic crystal. *Science* 2017, 356, 624–627.
259. Zhao R., Mei L., Hu K.-Q., Wang L., Chai Z.-F., Shi W.-Q. Two Three Dimensional Actinide Silver Heterometallic Coordination Polymers Based on 2,2'-Bipyridine 3,3'-dicarboxylic Acid with Helical Chains Containing Dimeric or Trimeric Motifs. *Eur. J. Inorg. Chem.* 2017, 2017, 1472–1477.
260. Thuéry P., Masci B. Uranyl–organic one- and twodimensional assemblies with 2,2'-bipyridine-3,3'-dicarboxylic, biphenyl-3,3',4,4'-tetracarboxylic and bicyclo[2.2.2]oct-7-ene-2,3,5,6-tetracarboxylic acids. *CrystEngComm* 2012, 14, 131–137.
261. Liu K., Zhou J.-M., Li H.-M., Xu N., Cheng P. A Series of Cu<sup>II</sup>–Ln<sup>III</sup> Metal–Organic Frameworks Based on 2,2'-Bipyridine-3,3'-dicarboxylic Acid: Syntheses, Structures, and Magnetic Properties. *Cryst. Growth Des.* 2014, 14, 6409–6420.
262. Grenthe I., Fuger J., Konings R. J. M., Lemire R. J., Nguyen-Trun C., Wanner H. Chemical Thermodynamics of Uranium; Organization for Economic Cooperation and Development: Issy-les Moulineaux, France, 2004.
263. Kikkawa S., Masu H., Katagiri K. et al. Characteristic Hydrogen Bonding Observed in the Crystals of Aromatic Sulfonamides: 1D Chain Assembly of



- Molecules and Chiral Discrimination on Crystallization. *Cryst. Growth Des.* 2019, 19, 5, 2936-2946.
264. Henshaw D. G. Atomic distribution in liquid and solid neon and solid argon by neutron diffraction Locality: synthetic Sample: at T = 4.2 K. *Phys. Rev.* 1958, 111, 1470-1475.
265. Эмсли Дж. Элементы. М.: Мир, 1993. 256 с.
266. Химическая энциклопедия. М.: Большая Российская энциклопедия. Т.1-4. 1988-1995.
267. Бацанов С. С. Структурные аспекты ван-дер-Ваальсовых комплексов. *Координац. химия.* 1998. Т.24. №7. С.483-487.
268. Филиппов Л. П. Методы расчета и прогнозирования свойств веществ. М.: Изд-во МГУ, 1988. 252 с.
269. Schuch A. F., Mills R. L. New Allotropic Form of He<sup>3</sup>. *Phys. Rev. Lett.* 1961, 6, 596.
270. Wyckoff R. W. G. Crystal structures. New York: Interscience. 1964. V.1.
271. Тонков Е. Ю. Фазовые диаграммы элементов при высоких давлениях. М.: Наука, 1979. 192 с.
272. Физическая энциклопедия. М.: Большая Российская энциклопедия. Т.1-4. 1988-1994.
273. Фастовский В. Г., Ровинский А. Е., Петровский Ю. В. Инертные газы. Атомиздат, 1972. 352 с.
274. Nasrifar K., Ayatollahi S., Moshfeghian M. An extended saturated liquid density equation. *Fluid Phase Equilibria.* 1999, 166, 163-181.
275. Серёжкин В. Н., Буслаев Ю. А. Стереозффект неподеленной электронной пары во фторидах сурьмы. *Журн. неорган. химии* 1997, 42(7), 1178-1185.

276. Блатов В. А., Полькин В. А., Сережкин В. Н. Полиморфизм простых веществ и принцип равномерности. *Кристаллография* 1994, 39(3), 457-463.
277. Блатов В. А., Шевченко А. П., Сережкин В. Н. Области действия анионов в структуре кристаллов. *Доклады АН* 1997, 354(3), 336-339.
278. Ишмаев С. Н., Садиков И. Р., Чернышев А. А. и др. Нейтронно-структурные исследования твердого параводорода при давлениях до 24 кбар. *Журн. exper. и теорет. физики* 1983, 84(1), 394-403.
279. Ашкрофт Н., Мермин Н. Физика твердого тела. М.Мир, 1979. с.201.
280. Seiler P., Belzner J., Bunz U., Szeimies G. Crystal Structure and Electron-Density Distribution of Two [1.1.1] Propellane Derivatives at 81 K. *Helv. Chim. Acta* 1988, 71, 2100-2110.
281. Prokhvatilov A. I., Isakina A. P. An X-ray powder diffraction study of crystalline  $\alpha$ -methane- $d_4$ . *Acta Cryst.* 1980, B36(7), 1576.
282. Tsai J., Taylor R., Chothia C., Gerstein M. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 1999, 290, 253-266.
283. Rehm J. D. D., Ziemer B., Szeimies G. A Facile, Palladium-Catalyzed Synthesis of 1,1'-Bi(bicyclo[1.1.1]pentanes). *Eur. J. Org. Chem.* 2001, 1049-1052.
284. Dunitz J. D., Gavezzotti A. Attractions and Repulsions in Molecular Crystals: What Can Be Learned from the Crystal Structures of Condensed Ring Aromatic Hydrocarbons? *Acc. Chem. Res.* 1999, 32, 677-684.

## Приложение

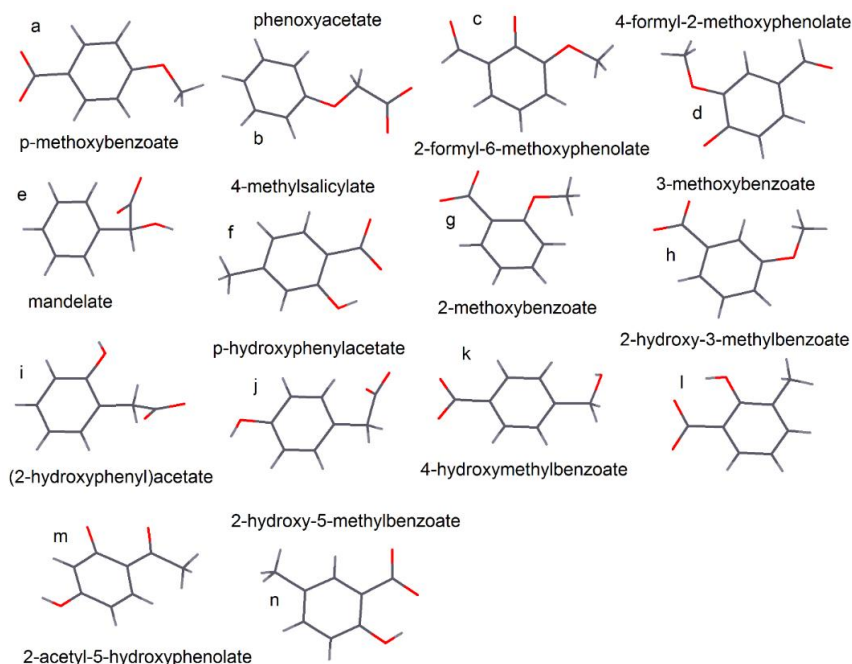


Рис. П1. Структура 14 изомеров состава  $C_8H_7O_3$ .

Таблица П1. Встречаемость координационных чисел лигандов ( $CN_{lig}$ )

$CN_{lig}$	N	w, %	Примеры
1	11188	46.74	диэтиловый эфир ( $C_4H_{10}O_{215}$ , YISLUJ); пивалиновая кислота ( $C_5H_{10}O_{2389}$ , FUSCIG)
2	5927	24.76	1,4-бензохинон ( $C_6H_4O_{2807}$ , POWYIL); дигидроксималонат ( $C_3H_2O_6^{2-735}$ , MAQVIL)
3	2728	11.40	цис,цис-циклогексан-1,3,5-трикарбоксилат ( $C_9H_9O_6^{3-956}$ , RORQEW); 2,6-диметоксикарбонилфенолят ( $C_{10}H_9O_5^{3-816}$ , NUFWUG)
4	2038	8.51	4,4'-диметоксибифенил-3,3'-дикарбоксилат ( $C_{16}H_{12}O_6^{2-964}$ , RUZCAS); бифенил-2,3',5,5'-тетракарбоксилат ( $C_{16}H_6O_8^{3-1063}$ , VUJLET)
5	546	2.28	гидрокси(окси)малонат ( $C_3HO_6^{3-141}$ , AZESON); 5-(6-карбоксилатонафталин-2-ил)бензол-1,3-дикарбоксилат ( $C_{19}H_9O_6^{3-777}$ , NAJTAW)
6	812	3.39	дигидроген-трансаконитат ( $C_6H_5O_6^{3-645}$ , KHTACN); 4,4',4''-(бензол-1,3,5-триилтрис(окси))трибензоат ( $C_{27}H_{15}O_9^{3-852}$ , PODGEW)
7	202	0.84	мезо-2,2'-оксидисукцинат ( $C_8H_6O_9^{4-171}$ , BILKIR); трис(4-карбоксилатофенил)метан ( $C_{22}H_{13}O_6^{3-276}$ , DAWCOV)
8	281	1.17	1,1':4',1''-терфенил-2',3,3'',5'-тетракарбоксилат ( $C_{22}H_{10}O_8^{4-117}$ , EKIIEP); галактар-1,2,3,4,5,6-гексат ( $C_6H_4O_8^{6-550}$ , HUQPAK)
9	69	0.29	карбонат ( $CO_3^{2-310}$ , XUNCAM); циклобутан-1,2,3,4-тетракарбоксилат ( $C_8H_4O_8^{4-489}$ , YUQYIU)
10	73	0.30	1,4-дикарбоксициклогексан-2,3,5,6-тетракарбоксилат ( $C_{12}H_8O_{12}^{4-687}$ , LAWSEJ); 1,2,3,4,5-бензолпентакарбоксилат ( $C_{11}HO_{10}^{5-979}$ , WIYZIO)
11	25	0.10	циклогексан-1,2,3,4,5,6-гексакарбоксилат ( $C_{12}H_6O_{12}^{6-335}$ , ROGRUB); оксалат ( $C_2O_4^{2-74}$ , WUWTOX)
12	37	0.15	родизонат ( $C_6O_6^{2-1160}$ , VOQXEE); 1,1',1''-(1,3,5-бензолтриил)трис(3,5-дикарбоксилатобензол) ( $C_{30}H_{12}O_{12}^{6-844}$ , ONIXOZ)
13	2	0.01	1,1',1''-(1,3,5-бензолтриил)трис(3,5-дикарбоксилатобензол) ( $C_{12}H_6O_{12}^{6-335}$ , ROGRAN); гидроцитрат ( $C_6H_5O_7^{3-177}$ , MABBEZ)
14	4	0.02	меллитат ( $C_{12}O_{12}^{6-291}$ , QAGXUT); гидроцитрат ( $C_6H_5O_7^{3-177}$ , ORONOA)
16	3	0.01	3,3',5,5'-тетракис(3,5-дикарбоксилатофенил)-4,4'-диметокси-1,1'-бифенил ( $C_{46}H_{22}O_{18}^{8-547}$ , HUGGEA); бензол-1,2,4,5-тетракарбоксилат ( $C_{10}H_2O_8^{4-36}$ , AVOGAT)

**Таблица П2.** Часто встречающиеся координационные типы лигандов. Преобладающий координационный тип  $M^1$  не включен в представленные данные. Его доля составляет 31.12% от всех имеющихся 11 891 различных сочетаний RefCode $\leftrightarrow$ Coord.type.

Рейтинг	Тип коорд.	N	w, %	Примеры
1	$M^2$	1155	14.10	метоксо ( $CH_3O^-$ _7, ACUZOO); мезитилоксо ( $C_9H_{11}O^-$ _54, SUKGIP)
2	$B^{01}$	978	11.94	ацетилацетонат ( $C_5H_7O_2^-$ _9, ACACIN); 1,2-диметоксиэтан ( $C_4H_{10}O_2$ _55, ZEQUFUW)
3	$B^2$	775	9.46	ацетат ( $C_2H_3O_2^-$ _10, EKABOI); бензоат ( $C_7H_5O_2^-$ _61, EJEDAA)
4	$M^3$	459	5.60	метоксо ( $CH_3O^-$ _7, BAFQEE); t-бутоксо ( $C_4H_9O^-$ _77, BAFQEE)
5	$B^{11}$	402	4.91	ацетат ( $C_2H_3O_2^-$ _10, JIPMEB); салицилальдегидат ( $C_7H_5O_2^-$ _8, WOXXOX)
6	$K^4$	397	4.85	терефталат ( $C_8H_4O_4^{2-}$ _81, IFABIA01); фуран-2,5-дикарбоксилат ( $C_6H_2O_5^{2-}$ _308, DIZQAG)
7	$K^{02}$	339	4.14	оксалат ( $C_2O_4^{2-}$ _74, MIYGIL); тартрат ( $C_4H_4O_6^{2-}$ _19, XEDBOZ)
8	$K^{22}$	249	3.04	оксалат ( $C_2O_4^{2-}$ _74, HAKYOH); сукцинат ( $C_4H_4O_4^{2-}$ _51, YAWZUS)
9	$B^{21}$	216	2.64	ацетат ( $C_2H_3O_2^-$ _10, XAJWAN); бинафтил-2,2'-диолат ( $C_{20}H_{12}O_2^{2-}$ _159, OKENEZ)
10	$B^3$	170	2.08	формиат ( $CHO_2^-$ _1, CUFORM02); ацетат ( $C_2H_3O_2^-$ _10, CEKNEG)
11	$K^{31}$	149	1.82	сукцинат ( $C_4H_4O_4^{2-}$ _51, HOTPOW); изофталат ( $C_8H_4O_4^{2-}$ _207, CEJBIE)
12	$K^{21}$	121	1.48	бензол-1,3,5-трикарбоксилат ( $C_9H_3O_6^{3-}$ _119, EKEXUO); 4,4'-оксидибензоат ( $C_{14}H_8O_5^{2-}$ _317, KUSDAE)
13	$K^{12}$	111	1.36	оксалат ( $C_2O_4^{2-}$ _74, DOGYAA); фталат ( $C_8H_4O_4^{2-}$ _132, NUYSOQ)
14	$M^4$	109	1.33	оксо ( $O^{2-}$ _12, EBOVOH); t-бутоксо ( $C_4H_9O^-$ _77, VIZCOY)
15	$T^{11}$	100	1.22	гидроксиацетат ( $C_2H_3O_3^-$ _46, EQINOI); малонат ( $C_3H_2O_4^{2-}$ _13, FIZJEF)
16	$B^4$	87	1.06	формиат ( $CHO_2^-$ _1, ONOCOK); ацетат ( $C_2H_3O_2^-$ _10, YUWKUX)
17	$K^{42}$	81	0.99	оксалат ( $C_2O_4^{2-}$ _74, YEPLAG02); сукцинат ( $C_4H_4O_4^{2-}$ _51, RISXAT)
18	$T^3$	77	0.94	4-карбоксибензол-1,2-дикарбоксилато ( $C_9H_4O_6^{2-}$ _108, TIFREN); 3,3'-(1,3-фенилен)бисакрилат ( $C_{12}H_8O_4^{2-}$ _500, LOZXIJ)
19	$K^6$	68	0.83	сукцинат ( $C_4H_4O_4^{2-}$ _51, EYUGEL); скворат ( $C_4O_4^{2-}$ _33, FIJBUV01)
20	$T^{21}$	64	0.78	скворат ( $C_4O_4^{2-}$ _33, ODATOE); оксидоацетат ( $C_2H_2O_3^{2-}$ _383, RUXMII)
21	$G^6$	56	0.68	4,4',4''-(1,3,5-трифенилен)-трибензоат ( $C_{27}H_{15}O_6^{3-}$ _259, OTEDAU); бензол-1,2,4,5-тетракарбоксилат ( $C_{10}H_2O_8^{4-}$ _36, MURBOQ01)
22	$G^{22}$	55	0.67	тартрат ( $C_4H_4O_6^{2-}$ _19, IFIWIE); бензол-1,4-диоксиацетат ( $C_{10}H_8O_6^{2-}$ _57, DOYYEW)
23	$K^{41}$	48	0.59	4,4'-оксидибензоат ( $C_{14}H_8O_5^{2-}$ _317, ZEFSUA); малонат ( $C_3H_2O_4^{2-}$ _13, JETVAF)
24	$T^{02}$	45	0.55	салицилат ( $C_7H_4O_3^{2-}$ _235, FAQMUG); 2-оксиизофталат ( $C_8H_3O_5^{3-}$ _21, ADICIA)
25	$T^{001}$	42	0.51	2,2-оксидиацетат ( $C_4H_4O_5^{2-}$ _205, YOMVOL); диглим ( $C_6H_{14}O_3$ _147, HEYVIS)
26	$T^{31}$	41	0.50	карбонат ( $CO_3^{2-}$ _310, FABZOZ); гидрофталат ( $C_8H_5O_4^-$ _198, YUHTAW)
27	$K^{32}$	39	0.48	оксалат ( $C_2O_4^{2-}$ _74, OKOJOP); диметилмалонат ( $C_5H_6O_4^{2-}$ _98, POJYOE)
28	$K^5$	39	0.48	бензол-1,3,5-трикарбоксилат ( $C_9H_3O_6^{3-}$ _119, KURVID); циклогексан-1,2-дикарбоксилат ( $C_8H_{10}O_4^{2-}$ _527, TOQRON)
29	$G^{000001}$	39	0.48	18-краун-6 эфир ( $C_{12}H_{24}O_6$ _85, LAPPEX); дициклогексил-18-краун-6 эфир ( $C_{20}H_{36}O_6$ _324, VUYYPAP)
30	$M^5$	36	0.44	оксо ( $O^{2-}$ _12, HULWAO); метоксо ( $CH_3O^-$ _7, AJACUK);
31	$K^8$	33	0.40	скворат ( $C_4O_4^{2-}$ _33, NAZDAU); бензол-1,4-дикарбоксилат ( $C_8H_4O_4^{2-}$ _81, MAQDUD)
Иное	326	1611	19.67	

**Таблица ПЗ.** Типы координационных фигур и их обозначение

CN	Символ полиэдра	Координационная фигура
0	S-0	Одиночный атом
1	L-1	Линейная
2	L-2	Угловая
	A-2_165	
	A-2_150	
	A-2_135	
	A-2_120	
	A-2_105	
	A-2_90	
	A-2_75	
	A-2_60	
	A-2_45	
	A-2_30	
	A-2_15	
	A-2_0	
3	TRIG-3	Плоский треугольник
	ТРУ-3_105	Некомпланарный треугольник
	ТРУ-3_90	
	ТРУ-3_75	
	ТРУ-3_60	
	ТРУ-3_45	
	ТРУ-3_30	
	ТРУ-3_15	
	ТРУ-3_0	
	TS-3	Т-образная
4	T-4	Правильный тетраэдр
	TDIST-4_90	Искаженный тетраэдр
	TDIST-4_75	
	TDIST-4_60	
	TDIST-4_45	
	TDIST-4_30	
	TDIST-4_15	
	TDIST-4_0	
	QUAG-4	Плоский квадрат
	SPY-4_75	Некомпланарный квадрат
	SPY-4_60	
	SPY-4_45	
	SPY-4_30	
	SPY-4_15	
	SPY-4_0	
	SS-4	Качели
5	PENG-5	Пентагон
	SPY-5a	Базоцентрированная квадратная пирамида
	SPY-5b	Объемоцентрированная квадратная пирамида
	TBPY-5	Треугольная бипирамида
5	TF-5	Одношпичный тетраэдр
	TE-5	Тетраэдр с вершиной на ребре
	TAPRV-5	Тригональная антипризма без одной вершины
	SSE-5	Качели с вершиной на ребре
	OC-6	Октаэдр
	TPR-6	Треугольная призма

6	TPR-6_Mg	Треугольная призма Mg
	TAPR-6_CPU&HEXG-6	Треугольная призма Cu (60 град.)
	TAPR-6_CPU_75	Треугольная призма Cu (75 град.)
	PPY-6	Пятиугольная пирамида
	HEXG-6	Шестиугольник
7	PBPY-7	Пятиугольная бипирамида
	TPRS-7	Треугольная призма с шапкой на квадратной грани
	TPRT-7	Треугольная призма с шапкой на треугольной грани
	OCF-7	Одношапочный октаэдр
	HEPG-7	Семиугольник
8	CU-8	Куб
	SAPR-8	Квадратная антипризма
	TPRS-8	Двухшапочная треугольная призма на квадратных гранях
	TPRT-8	Двухшапочная тригональная призма на треугольных гранях
	DD-8	Тригональный додекаэдр
	HEXBPY-8	Шестиугольная бипирамида
	OCBPA-8	Парабиаугментированный октаэдр
	OCOVA-8	Ортобиаувеличенный октаэдр
	OCTG-8	Восьмиугольник
9	TPRS-9	Трехшапочная треугольная призма
	HEBPY-9	Семиугольная бипирамида
	ICTD-9	Трехмерный икосаэдр
	SAPRS-9	Одношапочная квадратная антипризма на квадратной грани
	CUS-9	Одношапочная квадратная призма на квадратной грани
	OCF-9	Трехшапочный октаэдр
	NONG-9	Девятиугольник
10	PPR-10	Пятиугольная призма
	PAPR-10	Пятиугольная антипризма
	SAPRS-10	Двухшапочная квадратная антипризма на квадратной грани
	CUS-10	Двухшапочная квадратная призма
	ICMBD-10	Метабидиминированный икосаэдр
	DECG-10	Десятиугольник
11	VESTA-11	Модель отталкивания электронных пар - 11
	PPRP-11	Одношапочная пятиугольная призма на пятиугольной грани
	PPRS-11	Одношапочная пятиугольная призма на квадратной грани
	TPRS-11	Пятишапочная треугольная призма
12	IC-12	Икосаэдр
	PPRP-12	Двухшапочная пятиугольная призма на пятиугольной грани
	TRT-12	Усеченный тетраэдр
	CUOC-12	Кубооктаэдр
	CUOCH-12	Гексагональный кубооктаэдр
	TOBC-12	Треугольный ортобикупол
	HPR-12	Шестиугольная призма
12	HAPR-12	Шестиугольная антипризма
	PPRS-12	Двухшапочная пятиугольная призма на квадратной грани
13	VESTA-13	Модель отталкивания электронных пар - 13
	HPRS-13	Одношапочная шестиугольная призма на квадратной грани
14	VESTA-14	Модель отталкивания электронных пар - 14
	RDD-14	Ромбододекаэдр
	BB-14	Билунабиротонда
	HSPMC-14	Гебесфеномегакорона
	HPRMBA-14	Двухшапочная шестиугольная призма на квадратной грани
	HPRPBA-14	Парабиаугментированная шестиугольная призма
	VESTA-15	Модель отталкивания электронных пар - 15

15	ETC-15	Вытянутый треугольный купол
	HPRS-15	Триаугментированная шестиугольная призма
16	VESTA-16	Модель отталкивания электронных пар - 16
	OAPR-16	Восьмиугольная антипризма
	DPHC-16	Дисфеноцингулум
	SAPRSN-16	Курносая квадратная антипризма
	SGBC-16	Гиробикупол квадратный
	SOBC-16	Квадратный ортобикупол
	OPR-16	Восьмиугольная призма
17	VESTA-17	Модель отталкивания электронных пар - 17
18	VESTA-18	Модель отталкивания электронных пар - 18
	ETGBC-18	Гиробикупол вытянутой треугольной формы
	ETOBC-18	Удлиненный треугольный ортобикупол
	GETBC-18	Гироудлинённый треугольный бикупол
19	VESTA-19	Модель отталкивания электронных пар - 19
20	VESTA-20	Модель отталкивания электронных пар - 20
	DD-20	Додекаэдр
	DPR-20	Десятиугольная призма
21	VESTA-21	Модель отталкивания электронных пар - 21
	DDS-21	Аугментированный додекаэдр
22	VESTA-22	Модель отталкивания электронных пар - 22
	DDMBA-22	Метабиаугментированный додекаэдр
	DDPBA-22	Парабиаугментированный додекаэдр
23	VESTA-23	Модель отталкивания электронных пар - 23
	DDTA-23	Триаугментированный додекаэдр
24	VESTA-24	Модель отталкивания электронных пар - 24
	RCUOC-24	Ромбокубооктаэдр
	SNCU-24	Курносый куб
	TRCU-24	Усеченный куб
	TROC-24	Усеченный октаэдр
	GESBC-24	Гироудлинённый квадратный бикупол
25	VESTA-25	Модель отталкивания электронных пар - 25
	PGCR-25	Пятиугольная гирокуполаротонда
	POCR-25	Пятиугольная ортокуполаротонда
26	VESTA-26	Модель отталкивания электронных пар - 26
	DELICT-26	Дельтоидный икоситетраэдр
	DISDD-26	Додекаэдр Дисдякиса
27	VESTA-27	Модель отталкивания электронных пар - 27
28	VESTA-28	Модель отталкивания электронных пар - 28
29	VESTA-29	Модель отталкивания электронных пар - 29
	EPGBC-29	Удлиненный пятиугольный гиробикупол
30	VESTA-30	Модель отталкивания электронных пар - 30
	ICDD-30	Икосододекаэдр
	EPOBC-30	Удлиненный пятиугольный ортобикупол
	GEPBC-30	Гироудлинённый пятиугольный бикупол
	POBR-30	Пятиугольная ортобиротонда
31	VESTA-31	Модель отталкивания электронных пар - 31
32	PKDD-32	Додекаэдр Пентакиса
33	VESTA-33	Модель отталкивания электронных пар - 33
34	VESTA-34	Модель отталкивания электронных пар - 34
35	VESTA-35	Модель отталкивания электронных пар - 35
	EPGCR-35	Вытянутая пятиугольная гирокуполаротонда
	EPOCR-35	Вытянутая пятиугольная ортокуполаротонда
36	VESTA-36	Модель отталкивания электронных пар - 36

**Таблица П4.** Распределение первых десяти наиболее распространенных координационных фигур лигандов по координационным числам (CN) в диапазоне 1–4 в базовых сетках. Первые три координационные цифры для каждого координационного номера показаны на Рис. 49.

Координационная фигура	N	w, %	Пример (название лиганда, формула, CSD RefCode)
<b>CN=1</b>	<b>4253</b>	<b>31.7</b>	
L-1	4253	100	гептан-3,5-дионато, $C_7H_{11}O_2^-$ , CIKSAT
<b>CN=2</b>	<b>3602</b>	<b>26.8</b>	
A-2{60}	622	17.3	ацетат, $C_2H_3O_2^-$ , CUAQAC01
A-2{45}	557	15.5	3-карбоксифеноксияцетат, $C_9H_7O_5^-$ , JAQPUN
A-2{90}	472	13.1	вода, $H_2O$ , OQOGUY
L-2 = A-2{180}	420	11.7	терефталат, $C_8H_4O_4^{2-}$ , DIKQET04
A-2{75}	376	10.4	манделат, $C_8H_6O_3^-$ , INIROM
A-2{105}	351	9.7	вода, $H_2O$ , XADLIX
A-2{120}	205	5.7	5-карбоксибензол-1,3-дикарбоксилат, $C_9H_4O_6^{2-}$ , VUXBUL01
A-2{135}	179	5.0	гидроксиацетат, $C_2H_3O_3^-$ , GDHOAC11
A-2{30}	156	4.3	бензоат, $C_7H_5O_2^-$ , NEXZAS
A-2{150}	131	3.6	4-гидроксифенол, $C_6H_4O_2^{2-}$ , NIVFUT
<i>Иных 3</i>	<i>133</i>	<i>3.7</i>	
<b>CN=3</b>	<b>1950</b>	<b>14.5</b>	
FAN-3{60,75,135}	72	3.7	формиат, $CHO_2^-$ , NIHBEM
TPY-3{90,105,120}	63	3.2	2-гидроксипутан-1,4-диоат, $C_4H_4O_5^{2-}$ , FAYXEJ02
TP-3{105,120,135}	62	3.2	малонат, $C_3H_2O_4^{2-}$ , MALMND01
FAN-3{60 <sup>2</sup> ,120}	61	3.1	циклобутан-1,1-дикарбоксилат, $C_6H_6O_4^{2-}$ , JUPJIP
TPY-3{90,105 <sup>2</sup> }	59	3.0	гидроксид, $OH^-$ , UWANEH
TPY-3{45,135,165}	59	3.0	бензол-1,3-диоксидацетат, $C_{10}H_8O_6^{4-}$ , BIPJAM
TPY-3{45,60,90}	57	2.9	фуран-3-карбоксилат, $C_5H_3O_3^-$ , DESKIW
FAN-3{45 <sup>2</sup> ,90}	54	2.8	9Н-ксантен-9-карбоксилат, $C_{14}H_9O_3^-$ , NIDDEL
TP-3{75,135,150}	52	2.7	формиат, $CHO_2^-$ , LOSKUA
TPY-3{45 <sup>3</sup> }	50	2.6	феноксо, $C_6H_5O^-$ , JODDAH01
<i>Иных 45</i>	<i>1361</i>	<i>69.8</i>	
<b>CN=4</b>	<b>1754</b>	<b>13.1</b>	
RAP-4{45 <sup>2</sup> ,135 <sup>2</sup> ,180 <sup>2</sup> }	115	6.6	3,3'-диметоксибифенил-4,4'-дикарбоксилат, $C_{16}H_{12}O_6^{2-}$ , WEDNUQ
SPY-4{45,60,75,90,120 <sup>2</sup> }	108	6.2	бензол-1,2-дикарбоксилат, $C_8H_4O_4^{2-}$ , LEJPIA
RAP-4{60 <sup>2</sup> ,120 <sup>2</sup> ,180 <sup>2</sup> }	106	6.0	сукцинат, $C_4H_4O_4^{2-}$ , MAQZIP
RAP-4{30 <sup>2</sup> ,135,150,180 <sup>2</sup> }	94	5.4	терефталат, $C_8H_4O_4^{2-}$ , IXODUV
RAP-4{75 <sup>2</sup> ,105 <sup>2</sup> ,180 <sup>2</sup> }	71	4.0	бутан-1,2,3,4-тетракарбоксилат, $C_8H_6O_8^{4-}$ , IZEGEA
SS-4{75,90,105,120 <sup>2</sup> ,150}	69	3.9	тартарат, $C_4H_4O_6^{2-}$ , ZOMREA
RAP-4{60 <sup>2</sup> ,90,120,135,150}	59	3.4	сукцинат, $C_4H_4O_4^{2-}$ , IHILUI
FAN-4{60,75 <sup>2</sup> ,135 <sup>2</sup> ,150}	57	3.2	фуран-2,5-дикарбоксилат, $C_6H_2O_5^{2-}$ , VAZLAL
SP-4{90 <sup>4</sup> ,180 <sup>2</sup> }	56	3.2	циклобутан-1,2,3,4-тетрон, $C_4O_4^{2-}$ , RISXAU
T-4{90 <sup>2</sup> ,105,120 <sup>2</sup> ,135}	51	2.9	2,3-дигидроксипутандиоат, $C_4H_4O_6^{2-}$ , JIFXIG
<i>Иных 47</i>	<i>968</i>	<i>55.2</i>	
<b>CN&gt;4</b>	<b>1877</b>	<b>14.0</b>	
<b>Всего</b>	<b>13436</b>		



**Таблица П5.** Распределение первых десяти наиболее распространенных топологических типов (ТТ) базовых сеток МОКП в стандартном представлении для разных размерностей

ТТ	N	w, %	Примеры				
			CSD RefCode	Лиганд(ы)	Металл(ы)	Сольваты	Рисунок*
0D	2087	33.8					
2M4-1	323	15.5	HEXVOY	C <sub>10</sub> H <sub>11</sub> O <sub>5</sub> <sup>-</sup>  C <sub>15</sub> H <sub>11</sub> O <sub>2</sub> <sup>-</sup>	Dy	-	a
1,4M5-1	310	14.9	CEVNEZ	C <sub>13</sub> H <sub>7</sub> O <sub>2</sub> <sup>-</sup>  C <sub>4</sub> H <sub>8</sub> O	Ni	C <sub>4</sub> H <sub>8</sub> O	b
1,3M4-1	222	10.6	ACACGA03	C <sub>5</sub> H <sub>7</sub> O <sub>2</sub> <sup>-</sup>	Ga	-	c
1,2M3-1	176	8.4	CUBEAC02	C <sub>10</sub> H <sub>9</sub> O <sub>2</sub> <sup>-</sup>	Cu	-	
1,6M7-1	172	8.2	GETDIV	C <sub>9</sub> H <sub>4</sub> O <sub>6</sub> <sup>2-</sup>  H <sub>2</sub> O	Mg	H <sub>2</sub> O	
2,4M6-3	132	6.3	MEQPAC	C <sub>28</sub> H <sub>21</sub> O <sub>2</sub> <sup>-</sup>  C <sub>4</sub> H <sub>10</sub> O	Rh	C <sub>4</sub> H <sub>10</sub> O	
3,4M6-1	78	3.7	WEXBUZ	C <sub>8</sub> H <sub>13</sub> O <sub>3</sub> <sup>-</sup>	Li Ni	-	
1,5M6-1	73	3.5	CIKSAT	C <sub>7</sub> H <sub>11</sub> O <sub>2</sub> <sup>-</sup>  H <sub>2</sub> O	Yb	-	
3M8-1	34	1.6	RIGNAZ	C <sub>14</sub> H <sub>20</sub> O <sub>2</sub> <sup>-</sup>  C <sub>4</sub> H <sub>8</sub> O	Na	-	
2,3M5-1	31	1.5	RIGNED	C <sub>14</sub> H <sub>20</sub> O <sub>2</sub> <sup>-</sup>  C <sub>4</sub> H <sub>8</sub> O	Li	C <sub>4</sub> H <sub>8</sub> O	
Иные 169	536	25.7					
1D	903	14.6					
2C1	280	31.0	FOZHUX12	C <sub>2</sub> O <sub>4</sub> <sup>2-</sup>  H <sub>2</sub> O	Mn	-	d
(4,4)(0,2)	136	15.1	BOVQAF02	C <sub>8</sub> H <sub>7</sub> O <sub>2</sub> <sup>-</sup>  H <sub>2</sub> O	Pb	-	e
2,4C4	120	13.3	BADRAZ01	CHO <sub>2</sub> <sup>-</sup>  HO <sup>-</sup>	Cu	-	f
2,6C1	94	10.4	CEXBOZ	C <sub>9</sub> H <sub>9</sub> O <sub>2</sub> <sup>-</sup>	La	-	
3,4C4	40	4.4	AFEHOK01	C <sub>8</sub> H <sub>7</sub> O <sub>2</sub> <sup>-</sup>	Cu	-	
2,2,5,6C1	21	2.3	VORHOB	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> <sup>-</sup>  H <sub>2</sub> O	Co Gd	H <sub>2</sub> O	
3,5C2	14	1.6	CATPAL04	C <sub>8</sub> H <sub>4</sub> O <sub>4</sub> <sup>2-</sup>  H <sub>2</sub> O	Ca	-	
2,2,3C6	10	1.1	HORTIS	C <sub>14</sub> H <sub>4</sub> O <sub>8</sub> <sup>4-</sup>  H <sub>2</sub> O	Ni	H <sub>2</sub> O	
3,4C3	10	1.1	DAYZUB	C <sub>4</sub> H <sub>5</sub> O <sub>2</sub> <sup>-</sup>  O <sup>2-</sup>  H <sub>2</sub> O	Na U	-	
3,4,5C2	10	1.1	DEHGEF	C <sub>6</sub> H <sub>8</sub> O <sub>4</sub> <sup>2-</sup>  C <sub>8</sub> H <sub>4</sub> O <sub>5</sub> <sup>2-</sup>  H <sub>2</sub> O	Pr	H <sub>2</sub> O	
Иные 87	168	18.6					
2D	1036	16.8					
sql	201	19.4	CAPHTH03	C <sub>8</sub> H <sub>4</sub> O <sub>4</sub> <sup>2-</sup>  H <sub>2</sub> O	Ca	-	g
hcb	114	11.0	GUQXIZ06	C <sub>9</sub> H <sub>4</sub> O <sub>6</sub> <sup>2-</sup>  H <sub>2</sub> O	Cd	-	h
fes	58	5.6	OWEZEW01	C <sub>9</sub> H <sub>4</sub> O <sub>6</sub> <sup>2-</sup>  H <sub>2</sub> O	Ca	-	i
bey	51	4.9	CETHER	C <sub>6</sub> H <sub>8</sub> O <sub>4</sub> <sup>2-</sup>  C <sub>8</sub> H <sub>4</sub> O <sub>4</sub> <sup>2-</sup>  H <sub>2</sub> O	Gd	-	
kgd	40	3.9	PAXMOT	C <sub>11</sub> H <sub>21</sub> O <sub>2</sub> <sup>-</sup>	Pb	-	
bex	32	3.1	WEWHEO	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub> <sup>2-</sup>  H <sub>2</sub> O	Na	-	
(6,3)Ia	27	2.6	XEQKEM	C <sub>10</sub> H <sub>3</sub> O <sub>8</sub> <sup>3-</sup>  H <sub>2</sub> O	La	-	
3,5L60	18	1.7	DAQRUK	C <sub>4</sub> H <sub>4</sub> O <sub>6</sub> <sup>2-</sup>  C <sub>8</sub> H <sub>4</sub> O <sub>4</sub> <sup>2-</sup>  H <sub>2</sub> O	Na Sm	-	
(6,3)IIa	15	1.4	IGIVIF	C <sub>12</sub> H <sub>6</sub> O <sub>4</sub> <sup>2-</sup>	Cd	-	
5,5L4	15	1.4	YAMFAW	C <sub>10</sub> H <sub>14</sub> O <sub>4</sub> <sup>2-</sup>	Pb	-	
Иные 253	465	44.9					
3D	2145	34.8					
pcu	84	3.9	ETAFOU06	CO <sub>3</sub> <sup>2-</sup>	Fe	-	j
nia	50	2.3	UDOREM02	CHO <sub>3</sub> <sup>-</sup>	Na	-	k
pts	45	2.1	RUZBUL	C <sub>16</sub> H <sub>12</sub> O <sub>6</sub> <sup>2-</sup>	Zn	-	l
bnn	41	1.9	YETQOG	C <sub>9</sub> H <sub>5</sub> O <sub>6</sub> <sup>-</sup>  H <sub>2</sub> O	Na	-	
dia	35	1.6	ESADUZ	C <sub>14</sub> H <sub>8</sub> O <sub>5</sub> <sup>2-</sup>  H <sub>2</sub> O	Cd	-	
flu	31	1.4	NEZQAN	C <sub>7</sub> H <sub>4</sub> O <sub>3</sub> <sup>2-</sup>  CH <sub>4</sub> O	Li	-	
3,6,6T1	29	1.4	ZZZSDW03	C <sub>2</sub> H <sub>3</sub> O <sub>2</sub> <sup>-</sup>	Mg	-	
sra	24	1.1	CARGEK	C <sub>9</sub> H <sub>3</sub> O <sub>6</sub> <sup>3-</sup>  H <sub>2</sub> O	Bi	H <sub>2</sub> O	
tcs	18	0.8	MEJPEZ	C <sub>2</sub> O <sub>4</sub> <sup>2-</sup>  C <sub>8</sub> H <sub>4</sub> O <sub>5</sub> <sup>2-</sup>  H <sub>2</sub> O	Nd	H <sub>2</sub> O	
mab	18	0.8	NIFORM02	CHO <sub>2</sub> <sup>-</sup>  H <sub>2</sub> O	Ni	-	
Иные 875	1770	82.5					
Всего	6171						

\* см. Рис. 51, а-1 слева направо.

**Таблица П6.** Точность данных тестирования (20 % от полного набора данных), основанная на перекрестной проверке, соответствует максимальному доступному размеру обучающего набора (80 % от полного набора данных). Полученные точности указаны со стандартными отклонениями, основанными на 50 независимых разбиениях тренировочный/тестовый полного набора данных.

Классификатор/ набор признаков	RF	KNN	CART	SVC	LR	GNB
CN <sub>Me</sub> +CN <sub>Lig</sub>	0.732 ± 0.013	0.698 ± 0.017	0.731 ± 0.012	0.729 ± 0.013	0.673 ± 0.013	0.629 ± 0.016
CN <sub>Me</sub> +CT <sub>Lig</sub>	0.822 ± 0.013	0.788 ± 0.016	0.818 ± 0.012	0.818 ± 0.014	0.754 ± 0.014	0.647 ± 0.016
CN <sub>Me</sub> +CF <sub>Lig</sub>	0.824 ± 0.010	0.792 ± 0.013	0.819 ± 0.011	0.832 ± 0.009	0.779 ± 0.013	0.716 ± 0.013
CF <sub>Me</sub> +CF <sub>Lig</sub>	0.826 ± 0.013	0.803 ± 0.012	0.817 ± 0.012	0.842 ± 0.010	0.780 ± 0.013	0.736 ± 0.013
CN <sub>Me</sub> +CT <sub>Lig</sub> +CF <sub>Lig</sub>	0.866 ± 0.009	0.833 ± 0.009	0.855 ± 0.010	0.864 ± 0.010	0.814 ± 0.013	0.732 ± 0.011
CN+CT+CF взвешенная схема	0.900 ± 0.011	0.883 ± 0.012	0.879 ± 0.012	0.861 ± 0.010	0.787 ± 0.015	0.736 ± 0.013

**Таблица П7.** Встречаемость топологических типов всех 3D-структур (Все) и пористых координационных полимеров (МОКП) для одиночных (Z=1) и взаимопроникающих (Z>1) отделяется от символа топологии «#») сеток в стандартном представлении

Топология	Все, Z=1		МОКП, Z=1		Топология	Все, Z>1		МОКП, Z>1	
	N	w, %	N	w, %		N	w, %	N	w, %
<b>dia</b>	640	2.71	401	3.19	<b>4#dia</b>	224	5.09	126	5.02
<b>pcu</b>	560	2.37	250	1.99	<b>2#dia</b>	220	5.00	145	5.78
<b>pts</b>	444	1.88	262	2.08	<b>3#dia</b>	196	4.45	117	4.67
<b>sra</b>	373	1.58	236	1.88	<b>2#xah</b>	176	4.00	81	3.23
<b>cds</b>	217	0.92	138	1.10	<b>5#dia</b>	123	2.79	73	2.91
<b>xah</b>	205	0.87	71	0.56	<b>2#pcu</b>	118	2.68	73	2.91
<b>bpq</b>	205	0.87	105	0.83	<b>2#ths</b>	103	2.34	33	1.32
<b>srs</b>	201	0.85	138	1.10	<b>2#fet</b>	78	1.77	41	1.63
<b>bnn</b>	195	0.82	123	0.98	<b>2#srs</b>	77	1.75	43	1.71
<b>ins</b>	172	0.73	107	0.85	<b>2#sqc65</b>	76	1.73	36	1.44
<b>mog</b>	160	0.68	107	0.85	<b>2#cds</b>	74	1.68	52	2.07
<b>tcs</b>	154	0.65	80	0.64	<b>2#pts</b>	64	1.45	36	1.44
<b>nia</b>	149	0.63	53	0.42	<b>3#xah</b>	60	1.36	38	1.52
<b>dmc</b>	135	0.57	90	0.72	<b>2#mog</b>	55	1.25	36	1.44
<b>4,8T1</b>	125	0.53	38	0.30	<b>6#dia</b>	52	1.18	36	1.44
<i>Иные</i>	<i>19712</i>	<i>83.36</i>	<i>10378</i>	<i>82.52</i>	<i>Иные</i>	<i>2708</i>	<i>61.49</i>	<i>1542</i>	<i>61.48</i>
Всего	23647		12577		Всего	4404		2508	

**Таблица П8.** Встречаемость топологических типов всех 3D-структур (Все) и пористых координационных полимеров (МОКП) для одиночных ( $Z=1$ ) и взаимопроникающих ( $Z>1$ ) сеток в кластерном представлении

Топология	Все, $Z=1$		МОКП, $Z=1$		Топология	Все, $Z>1$		МОКП, $Z>1$	
	N	w, %	N	w, %		N	w, %	N	w, %
<b>pcu</b>	1698	16.30	916	16.64	<b>2#pcu</b>	673	32.80	360	32.70
<b>bcu</b>	589	5.65	312	5.67	<b>3#pcu</b>	179	8.72	98	8.90
<b>dia</b>	335	3.22	202	3.67	<b>2#dia</b>	133	6.48	73	6.63
<b>rtl</b>	329	3.16	201	3.65	<b>4#pcu</b>	56	2.73	23	2.09
<b>flu</b>	205	1.97	112	2.03	<b>3#dia</b>	48	2.34	35	3.18
<b>nbo</b>	187	1.79	84	1.53	<b>2#fsc</b>	38	1.85	23	2.09
<b>tfz-d</b>	181	1.74	104	1.89	<b>2#tfz-d</b>	34	1.66	21	1.91
<b>cds</b>	175	1.68	109	1.98	<b>2#rtl</b>	32	1.56	16	1.45
<b>fcu</b>	171	1.64	48	0.87	<b>2#sit</b>	31	1.51	17	1.54
<b>flu-3,6-C2/c</b>	158	1.52	75	1.36	<b>4#dia</b>	22	1.07	10	0.91
<b>fsc</b>	138	1.32	68	1.24	<b>2#hex</b>	21	1.02	10	0.91
<b>hex</b>	135	1.30	82	1.49	<b>2#lvt</b>	21	1.02	10	0.91
<b>rob</b>	130	1.25	88	1.60	<b>2#the</b>	20	0.97	11	1.00
<b>pts</b>	127	1.22	88	1.60	<b>2#hms</b>	17	0.83	9	0.82
<b>acs</b>	126	1.21	74	1.34	<b>2#pto</b>	17	0.83	11	1.00
<i>Иные</i>	<i>5734</i>	<i>55.04</i>	<i>2941</i>	<i>53.44</i>	<i>Иные</i>	<i>710</i>	<i>34.60</i>	<i>374</i>	<i>33.97</i>
Всего	10418		5504		Всего	2052		1101	

**Таблица П9.** Встречаемость самодуальных сеток в структурах координационных полимеров

Символ	Стандартное		Кластерное		Символ	Стандартное		Кластерное	
	N(Z=1)	N(Z>1)	N(Z=1)	N(Z>1)		N(Z=1)	N(Z>1)	N(Z=1)	N(Z>1)
<b>bbr</b>	-	-	-	-	<b>pte</b>	-	-	-	-
<b>cbs</b>	-	-	-	-	<b>pyr</b>	16	1	39	10
<b>cdq</b>	3	-	3	7	<b>qtz-x</b>	-	-	7	1
<b>cds</b>	217	121	175	32	<b>rtw</b>	-	-	-	-
<b>ctn</b>	20	1	15	2	<b>sda</b>	-	-	3	-
<b>dia</b>	619	898	335	236	<b>smt</b>	-	-	-	-
<b>est</b>	1	-	-	-	<b>srs</b>	194	111	26	14
<b>ete</b>	-	1	-	-	<b>sto</b>	-	-	-	-
<b>fsf</b>	1	-	-	-	<b>svn</b>	-	-	-	-
<b>ftw</b>	1	-	32	-	<b>swl</b>	-	-	-	-
<b>gsi</b>	4	-	-	-	<b>sxd</b>	3	1	10	1
<b>hms</b>	28	37	3	17	<b>tfa</b>	5	11	3	-
<b>hst</b>	-	-	1	-	<b>tfc</b>	6	1	6	4
<b>lcy</b>	1	-	21	7	<b>ths</b>	122	182	1	5
<b>mab</b>	33	-	120	14	<b>tph</b>	-	-	1	-
<b>mcf</b>	-	-	-	-	<b>ttv</b>	-	-	-	-
<b>mco</b>	-	-	1	-	<b>unj</b>	1	-	5	-
<b>mgc</b>	-	-	-	-	<b>vck</b>	-	-	-	-
<b>pcu</b>	498	145	1698	915	<b>vtx</b>	-	-	-	-

**Таблица П10.** Распределение часто встречающихся топологий (стандартных и кластерных) однокаркасных координационных по квартилям пористости

Топология*	Q1	Q2	Q3	Q4	Топология*	Q1	Q2	Q3	Q4
<b>dia</b>	2.06	3.00	3.57	0.39	<b>alb-4,8-Imma→lvt</b>	-	-	0.04	1.97
<b>pcu</b>	2.79	1.56	3.22	0.79	<b>tbo</b>	-	-	0.04	1.97
<b>pts</b>	1.47	1.73	2.29	0.79	<b>4,4,4,8T13→cor</b>	-	-	-	1.97
<b>srs</b>	0.59	0.61	0.82	3.15	<b>3,3,6,8,8T1→csq</b>	-	-	-	1.97
<b>3,4,8T15→fcu</b>	0.15	-	0.79	3.15	<b>tcs</b>	0.88	0.59	0.36	-
<b>fff→pcu</b>	-	-	0.14	3.15	<b>4,16T2→scu</b>	-	-	0.64	1.18
<b>4,12T1→nts</b>	-	-	0.39	2.76	<b>sod</b>	-	0.17	1.11	0.39
<b>3,6,8,8T1→flu</b>	-	-	0.32	2.76	<b>mog</b>	0.88	0.53	0.25	-
<b>alb-4,8-P4<sub>2</sub>/mmc→pts</b>	-	0.01	0.18	2.76	<b>3,5T1</b>	1.03	0.53	0.04	-
<b>cds</b>	0.74	1.23	0.43	0.39	<b>fsc</b>	0.15	0.42	0.89	-
<b>4,4,4,4,4,6,6T1→muo</b>	-	-	-	2.76	<b>flu</b>	1.03	0.27	0.04	-
<b>bnn</b>	1.32	0.92	0.50	-	<b>lvt</b>	0.29	0.52	0.43	-
<b>sra</b>	0.74	1.15	0.46	-	<b>ths</b>	0.15	0.45	0.64	-
<b>4,8T24→nbo</b>	-	-	0.75	1.57	<b>4,4,6T19</b>	0.88	0.19	0.14	-
<b>bpq</b>	0.15	0.71	1.47	-	<b>4,8T1</b>	-	0.08	1.11	-
<b>nts→tbo</b>	-	-	0.25	1.97	<b>fet→pcu</b>	-	0.47	0.64	-
<b>rtl</b>	1.62	0.46	0.11	-	<b>fsx-4,5-C2/c</b>	1.03	-	-	-
<b>nia</b>	1.32	0.33	0.54	-	<b>fsy</b>	-	0.50	0.43	-
<b>5,8T16→fsc</b>	-	0.01	0.11	1.97	<b>ins</b>	0.29	0.52	0.11	-
<b>ith-3,9-R3m</b>	2.06	-	-	-	<b>srs→dia</b>	0.29	0.50	0.11	-

\* Топология кластерного представления (если есть) указана после стрелки.

**Таблица П11.** Наиболее распространенные базовые сетки координационных полимеров с изоэдрическими тайлами и их индексы связности (CI).

Сетка	CI	Сетка	CI	Сетка	CI	Сетка	CI
<b>dia</b>	6/5	<b>srs</b>	15/14	<b>dmp</b>	11/10	<b>hex</b>	3/2
<b>pcu</b>	3/2	<b>bnn</b>	3/2	<b>ths</b>	10/9	<b>lvt</b>	5/4
<b>sra</b>	9/7	<b>bpq</b>	10/7	<b>bbf</b>	4/3	<b>qtz</b>	7/6
<b>cds</b>	7/6	<b>tcs</b>	14/9	<b>nbo</b>	4/3		

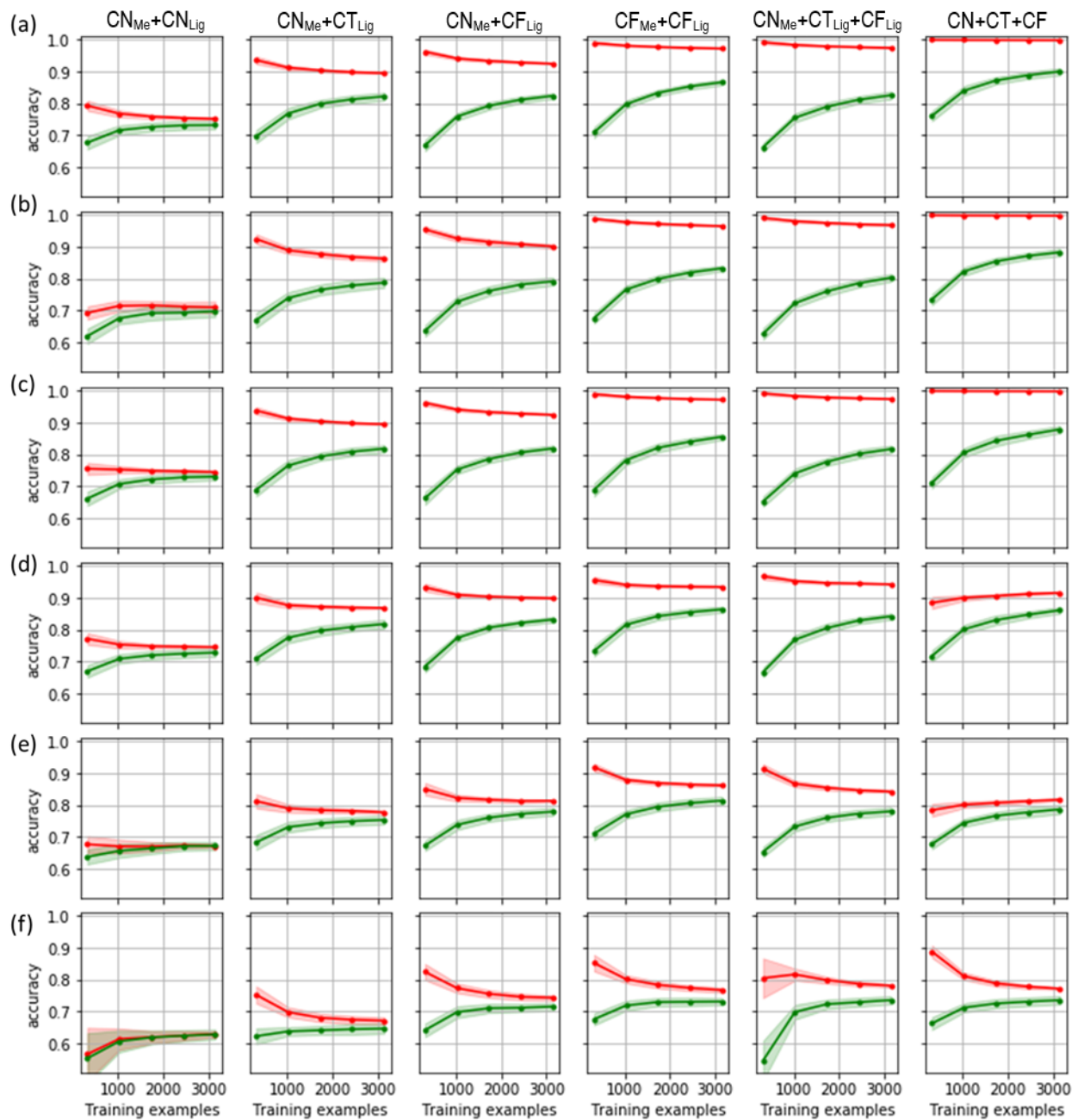


Рис. П2. Кривые обучения на основе перекрестной проверки для обученного (а) классификатора случайного леса - RF, (b) классификатора К-ближайших соседей - KNN, (c) классификатора дерева решений - CART, (d) метода опорных векторов - SVC, (e) классификатора логистической регрессии - LR и (f) гауссовского наивного байесовского классификатора - GNB, полученных для 50 независимых стратифицированных перетасованных разбиений полного набора данных (обучающий набор - 80%, тестовый набор - 20%) и 5 различных размеров обучающих подмножеств, выбранных из обучающего набора. Красные и зеленые линии представляют зависимости точности классификации обученных моделей от размера обучающего и тестового наборов данных, соответственно. Полупрозрачные области вдоль линий соответствуют одному стандартному отклонению полученных оценок.

**Таблица П12.** Матрицы парной корреляции Пирсона и Спирмена значений признаков для Тестовой Выборки I, используемой для прогнозирования степени окисления. Описание признаков приведено в **Таблице 15**. Жирным черным шрифтом выделены диагональные элементы матрицы, которые всегда равны единице. Жирный зеленый шрифт соответствует парам переменных, которые умеренно коррелируют со степенью окисления. Если признаки сильно или очень сильно коррелируют друг с другом, они объединяются в группы, указанные в строках CorrGr. Две переменные  $R_{sd}$  и  $V_{VP}$ , данные для которых выделены синим цветом, функционально зависят друг от друга, поэтому коэффициент корреляции Спирмена для них равен единице.

Признак	OxSt	N <sub>At</sub>	Group	Mult	CN	WCN	V <sub>CP</sub>	G <sub>3</sub>	Δ <sub>CM</sub>	R <sub>sd</sub>	V <sub>VP</sub>	S <sub>VP</sub>	ΣΩ <sub>O</sub>
Матрица парных корреляций Пирсона													
OxSt	1.000	0.528	0.465	-0.023	0.171	-0.220	-0.098	-0.027	-0.086	-0.525	-0.448	-0.473	0.470
N <sub>At</sub>	0.528	1.000	0.344	-0.017	0.401	0.264	0.299	-0.057	0.035	0.120	0.099	0.047	0.141
Group	0.465	0.344	1.000	-0.045	0.283	0.119	0.069	-0.074	-0.108	-0.348	-0.321	-0.350	0.419
Mult	-0.023	-0.017	-0.045	1.000	-0.079	-0.106	-0.069	0.558	0.516	0.199	0.243	0.330	0.012
CN	0.171	0.401	0.283	-0.079	1.000	0.748	0.903	-0.217	-0.183	0.201	0.166	0.068	0.464
WCN	-0.220	0.264	0.119	-0.106	0.748	1.000	0.796	-0.205	-0.113	0.451	0.387	0.274	0.005
V <sub>CP</sub>	-0.098	0.299	0.069	-0.069	0.903	0.796	1.000	-0.207	-0.137	0.491	0.429	0.339	0.198
G <sub>3</sub>	-0.027	-0.057	-0.074	0.558	-0.217	-0.205	-0.207	1.000	0.922	0.303	0.398	0.519	-0.068
Δ <sub>CM</sub>	-0.086	0.035	-0.108	0.516	-0.183	-0.113	-0.137	0.922	1.000	0.470	0.571	0.646	-0.197
R <sub>sd</sub>	-0.525	0.120	-0.348	0.199	0.201	0.451	0.491	0.303	0.470	1.000	0.959	0.955	-0.446
V <sub>VP</sub>	-0.448	0.099	-0.321	0.243	0.166	0.387	0.429	0.398	0.571	0.959	1.000	0.972	-0.392
S <sub>VP</sub>	-0.473	0.047	-0.350	0.330	0.068	0.274	0.339	0.519	0.646	0.955	0.972	1.000	-0.386
ΣΩ <sub>O</sub>	0.470	0.141	0.419	0.012	0.464	0.005	0.198	-0.068	-0.197	-0.446	-0.392	-0.386	1.000
CorrGr	-	-	-	-	окружение			равномерность		метрика			-
Матрица парных корреляций Спирмена													
OxSt	1.000	0.563	0.668	-0.032	0.224	-0.131	-0.018	-0.072	-0.288	-0.573	-0.573	-0.616	0.649
N <sub>At</sub>	0.563	1.000	0.426	-0.007	0.324	0.194	0.239	-0.219	-0.045	0.062	0.062	-0.005	0.223
Group	0.668	0.426	1.000	-0.023	0.261	0.145	0.126	-0.244	-0.264	-0.331	-0.331	-0.381	0.489
Mult	-0.032	-0.007	-0.023	1.000	0.021	-0.002	0.040	0.008	0.169	0.101	0.101	0.115	0.045
CN	0.224	0.324	0.261	0.021	1.000	0.739	0.935	-0.830	-0.217	0.218	0.218	0.087	0.447
WCN	-0.131	0.194	0.145	-0.002	0.739	1.000	0.803	-0.833	-0.169	0.462	0.462	0.327	0.060
V <sub>CP</sub>	-0.018	0.239	0.126	0.040	0.935	0.803	1.000	-0.833	-0.119	0.449	0.449	0.333	0.250
G <sub>3</sub>	-0.072	-0.219	-0.244	0.008	-0.830	-0.833	-0.833	1.000	0.362	-0.223	-0.223	-0.077	-0.275
Δ <sub>CM</sub>	-0.288	-0.045	-0.264	0.169	-0.217	-0.169	-0.119	0.362	1.000	0.426	0.426	0.470	-0.446
R <sub>sd</sub>	-0.573	0.062	-0.331	0.101	0.218	0.462	0.449	-0.223	0.426	1.000	1.000	0.978	-0.474
V <sub>VP</sub>	-0.573	0.062	-0.331	0.101	0.218	0.462	0.449	-0.223	0.426	1.000	1.000	0.978	-0.474
S <sub>VP</sub>	-0.616	-0.005	-0.381	0.115	0.087	0.327	0.333	-0.077	0.470	0.978	0.978	1.000	-0.506
ΣΩ <sub>O</sub>	0.649	0.223	0.489	0.045	0.447	0.060	0.250	-0.275	-0.446	-0.474	-0.474	-0.506	1.000
CorrGr	-	-	-	-	окружение + равномерность			-		метрика			-

*Автор выражает глубокую благодарность и признательность профессору Блатову Владиславу Анатольевичу за всестороннюю поддержку, ценные советы и идеи, помощь и консультации на разных этапах выполнения работы. Большая благодарность от автора Сереежкину Виктору Николаевичу за знакомство с удивительным симметричным миром кристаллохимии и строения вещества, а также Балаковскому Сергею Львовичу и Антошкиной Светлане Миленитиновне за направление в самые дебри информационных технологий и за помощь в их успешном прохождении.*

*Автор также выражает благодарность проф. Давиде М. Прозертио (Миланский университет, Италия), Фалку Мейцнеру (Фрайбергская горная академия, Германия), Джорджу Костакису (Университет города Сассекс, Великобритания) за новые идеи и полезные рекомендации при разработке программных продуктов.*

*За помощь в выполнении диссертационной работы автор отдельно благодарит своих коллег из МНИЦТМ Самарского университета: Е. В. Александрова, А. А. Шабалина, А. А. Голова, И. В. Медриш, М. В. Пескова, Т. Г. Ахметшину, Н. А. Кабанову, П. Н. Золотарева.*

*Автор выражает сердечную благодарность своей жене Татьяне Александровне Шевченко за понимание и поддержку во время выполнения диссертационной работы.*

*Работа выполнена при финансовой поддержке Мегагранта №14.B25.31.0005 Министерства образования и науки Российской Федерации, а также грантов Российского фонда фундаментальных исследований №№ 01-07-90092, 17-43-630619, 18-07-00183, 18-29-04010, 20-13-50007 и Российского Научного Фонда №№ 16-13-10158, 19-73-10026, 23-23-00387.*