

# Фингерпринт-вектор как дескриптор координационной фигуры

А.П. Шевченко<sup>1</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. Академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

**Аннотация.** Предложен новый метод определения геометрической формы локального атомного окружения в кристаллах, представленного в виде координационной фигуры. Метод, основанный на сравнении угловых фингерпринтов полиэдров, был внедрен в пакет программ ToposPro и протестирован на 9642 кристаллических структурах пятикоординированных комплексов меди и цинка. Полученные результаты использованы для поиска корреляций типа координационной фигуры с другими структурными дескрипторами.

## 1. Введение

При анализе данных, как правило, рассматриваются выборки из чисел. Методы их исследования общеизвестны, хорошо отлажены и реализованы во многих библиотеках статистических функций (R, NumPy, Scipy) или программах обработки данных (Excel, Statistica, MathCAD). Реальные выборки часто содержат не только числовые данные, но и текстовую информацию, категориальные величины и даже графические объекты. Возникает вопрос: каким образом проводить анализ нечисловых данных, как осуществлять их классификацию.

В подобных случаях используют некоторые дескрипторы информации, называемыми фингерпринтами [1], вид которых зависит от анализируемого объекта. В качестве примера рассмотрим механизм расчета фингерпринтов формы координационной фигуры атома в кристаллической структуре. Согласно [2], координационной фигурой называют объект, содержащий центральный атом и соседние атомы, валентно-связанные с ним. Определение формы координационной фигуры атома металла является важной задачей кристаллографии, так как эта характеристика оказывает существенное влияние на топологию кристаллической решетки в целом [3]. Чтобы решить ее необходимо сопоставить анализируемой координационной фигуре некоторый эталон (или образец) из библиотеки правильных координационных фигур. Для этих целей был использован алгоритм сравнения угловых фингерпринтов.

## 2. Описание метода

Любой координационной фигуре соответствует набор валентных углов со значениями в диапазоне от  $0^\circ$  до  $180^\circ$ . Расчета углового фингерпринта 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 (1) со стандартным отклонением  $\sigma$  и средним значением  $\mu$ , равным середине интервала (например,  $\sigma=9^\circ$ ).

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

Используя итоговую функцию (2), являющуюся суммой всех интервальных гауссовых функций, можно посчитать компоненты  $F_i$  отпечаток-вектора, имеющего размерность равную числу интервалов.

$$F_i = 100 \frac{\sum_{k=1}^{180/\delta} f_k(\mu_i)}{\sum_{m=1}^{180/\delta} \sum_{k=1}^{180/\delta} f_k(\mu_m)} \tag{2}$$

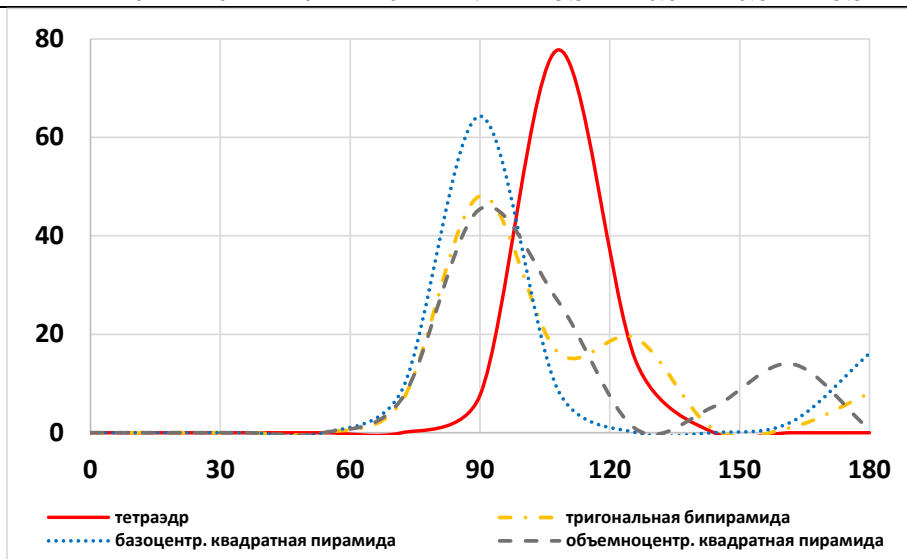
Чтобы определить степень сходства двух координационных фигур достаточно найти расстояние между их отпечаток-векторами  $F$  и  $F'$  по формуле (3).

$$r = \sqrt{\sum_{i=1}^{180/\delta} \left(\frac{F_i - F'_i}{2}\right)^2} \tag{3}$$

Координационная фигура из библиотеки эталонов, для которой расстояние  $r$  минимально, является искомой. Примеры отпечатков для некоторых правильных координационных фигур представлены в табл.1 и на рис.1. Нами были созданы программный модуль для расчета отпечатков и библиотека отпечатков правильных координационных фигур, что позволило проводить машинную классификацию произвольных выборок из таких объектов. Библиотека содержала отпечатки для тетраэдра (Т-4), тригональной бипирамиды (ТВРУ-5), квадратной базоцентрированной пирамиды (SPY-5a), квадратной объемцентрированной пирамиды (SPY-5b) и одношапочного тетраэдра (TF-5).

**Таблица 1.** Угловые отпечатки для правильных пятикоординационных фигур и тетраэдра при  $\delta=18^\circ$  и  $\sigma=9^\circ$ .

Интервал углов(°)	0±9	18±9	36±9	54±9	72±9	90±9	108±9	126±9	144±9	162±9	180±9
Т-4	0	0	0	0	0	7.6	77.8	14.6	0.1	0	0
ТВРУ-5	0	0	0	0	6.5	48.1	16.4	19.2	0.7	1.1	8
SPY-5a	0	0	0	0	8.7	64.3	8.7	0	0	2.2	16.1
SPY-5b	0	0	0	0	7	45.5	26.8	0.8	5.3	14	0.7



**Рисунок 1.** Отпечатки правильного тетраэдра, тригональной бипирамиды, а также базо- и объемцентрированных квадратных пирамид.

Для определения координационной фигуры некоего атома необходимо посчитать для него фингерпринт с теми же параметрами  $\delta$  и  $\sigma$ , а за тем найти правильную координационную фигуру с наиболее схожим фингерпринтом. Как было оговорено ранее, схожесть фингерпринтов, определяется по расстоянию между ними. Если координационные фигуры идентичны, то расстояние между их фингерпринтами равно нулю. При максимальном отличии координационных фигур расстояние близко к 100 (табл.2).

**Таблица 2.** Расстояния между фингерпринтами правильных координационных фигур.

Координационная фигура	T-4	TBPY-5	SPY-5a	SPY-5b
T-4	0	37.2	46.2	33.5
TBPY-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

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. Как уже отмечалось ранее, 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.

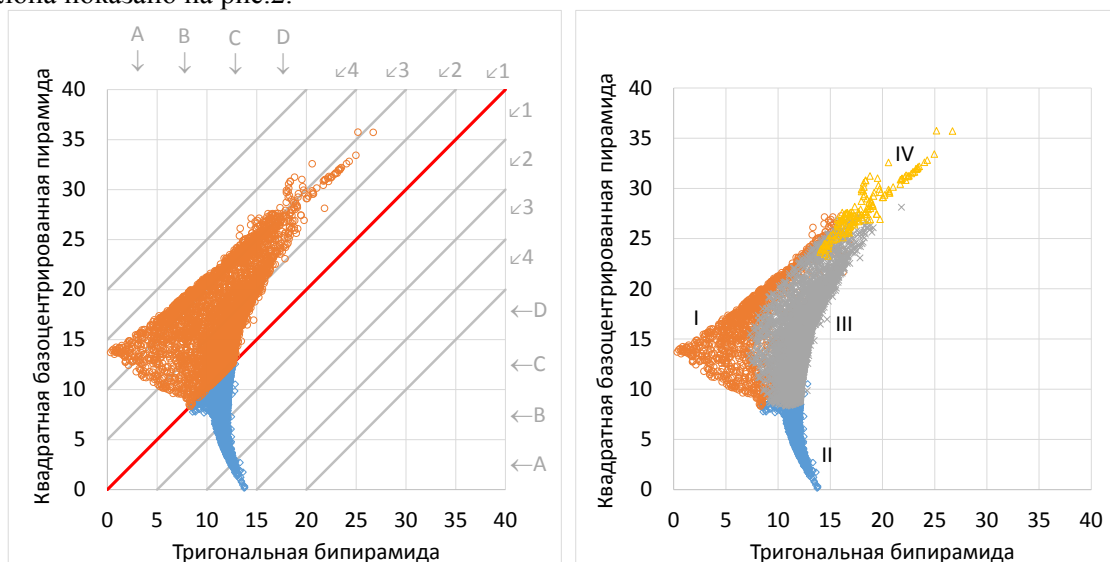
Соединения, удовлетворяющие указанным выше условиям, отобраны из баз данных ICSD [4] и CSD [5]. Объем выборки составил 9642, из которых 7241 и 2401 являлись комплексами меди и цинка, соответственно. Отбор структурных данных осуществлялся с помощью комплекса ToposPro [6]. Расчет матрицы смежности атомов проводили методом Domains.

**Таблица 3.** Встречаемость координационных фигур Cu или Zn. Координационные фигуры считались нераспознанными для рейтингов 1D, 1E и 2E.

Рейтинг	SPY-5b		SPY-5a		TBPY-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>		

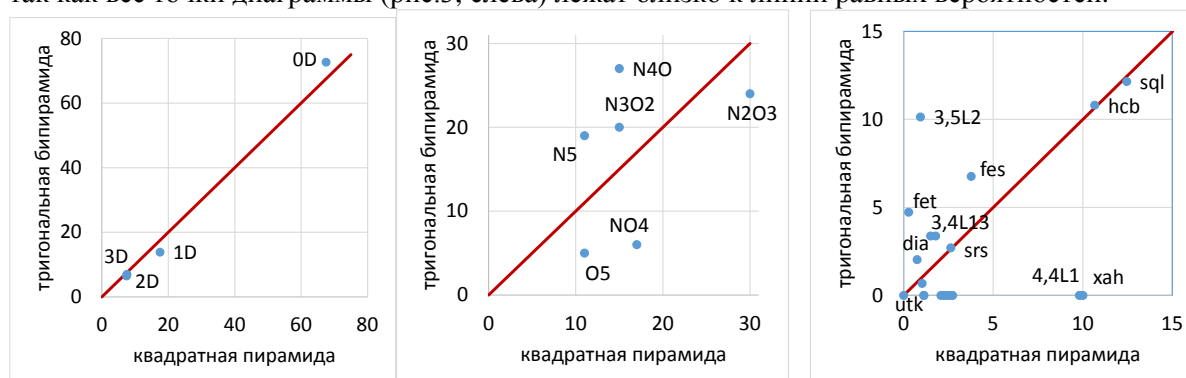
Результаты определения формы координационной фигуры с использованием фингерпринтов представлены в табл.3. Цифра рейтинга показывает отличие между двумя наименьшими расстояниями ( $\alpha$ ): 1 -  $\alpha \leq 5\%$ ; 2 -  $5\% < \alpha \leq 10\%$ ; 3 -  $10\% < \alpha \leq 15\%$ ; 4 -  $15\% < \alpha \leq 20\%$ . Чем ближе  $\alpha$  к разности для эталонов, тем более достоверны результаты определения координационной

фигуры. Например, при отнесении координационной фигуры образца либо к квадратной пирамиде, либо к треугольной бипирамиде наилучшее значение  $\alpha$  равно 3 (табл.2). Буква рейтинга характеризует близость выбранного значения к максимальному совпадению ( $\beta$ ): А -  $\beta \leq 5\%$ ; В -  $5\% < \beta \leq 10\%$ ; С -  $10\% < \beta \leq 15\%$ ; D -  $15\% < \beta \leq 20\%$ . Наиболее надежные результаты отвечают букве А. Распределение точек по расстоянию между фингерпринтами образца и эталона показано на рис.2.



**Рисунок 2.** Распределения расстояний между угловыми фингерпринтами рассматриваемой координационной фигуры и (слева) правильной треугольной бипирамиды или квадратной базоцентрированной пирамиды, а также (справа) треугольной бипирамиды, квадратичной базоцентрированной пирамиды, квадратная объемоцентрированной пирамиды и одношапочного тетраэдра (соответственно I, II, III или IV области).

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.3, слева) лежат близко к линии равных вероятностей.



**Рисунок 3.** Диаграмма рассеяния квадратная пирамида – треугольная бипирамида для встречаемости координационных фигур по размерности комплексных групп (слева), химическому составу координационной фигуры (в центре) или общей топологии сетки (справа).

Первая координационная сфера атомов Cu и Zn может содержать азот и кислород в любом соотношении. Наиболее предпочтительный состав координационной фигуры является  $N_2O_3$  для обоих металлов. При этом выполняется следующее правило (рис.3, в центре): *если атомы азота доминируют в координационном полиэдре ( $n > t$ ), то тригональная бипирамидная координация более вероятна, в противном случае преобладает квадратичная пирамидальная координация.*

### 3. Заключение

Ожидается, что форма координационной фигуры влияет на общую топологию сетки, но не во всех случаях. Такие общие топологии сеток, как 3,5L2, **fet**, **fes**, являются предпочтительными для тригонально-бипирамидного типа, тогда как топологические типы **xah**, 4,4L1 допустимы только для квадратно-пирамидального типа. Топологические типы (**sql**, **hcb**, **srs**, **utk**, рис.3, справа), расположенные близко к линии равных вероятностей, нечувствительны к типу координационной фигуры.

Таким образом, создание новых надежных дескрипторов является важной проблемой проектирования материалов: каждый новый дескриптор позволяет находить множество новых корреляций с другими параметрами вещества и существенно расширять соответствующую базу знаний. Здесь мы показали, насколько сложным и неочевидным может быть такое исследование даже для кажущегося простым геометрического параметра, каковым является координационная фигура.

### 4. Литература

- [1] Dass, R. Face Recognition Techniques: A Review / R. Dass, R. Rani, D. Kumar // Int. J. Eng. Research and Development. – 2012. – Vol. 4. – P. 70-78.
- [2] Alexandrov, E.V. Underlying nets in three-periodic coordination polymers: topology, taxonomy and prediction from a computer-aided analysis of the Cambridge Structural Database / E.V. Alexandrov, V.A. Blatov, A.V. Kochetkov, D.M. Proserpio // Cryst. Eng. Comm. – 2011. – Vol. 13(12). – P.3947-3958.
- [3] Ockwig, N.W. Reticular Chemistry: Occurrence and Taxonomy of Nets and Grammar for the Design of Frameworks / N.W. Ockwig, O. Delgado-Friedrichs, M. O’Keeffe, O.M. Yaghi // Acc. Chem. Res. – 2005. – Vol. 38. – P. 176-182.
- [4] Belsky, A. New Developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in Support of Materials Research and Design / A. Belsky, M. Hellenbrandt, V.L. Karen, P. Luksch // Acta Crystallogr., Sect. B: Struct. Sci. – 2002. – Vol. 58. – P. 364-369.
- [5] Allen, F.H. The Cambridge Structural Database: a quarter of a million crystal structures and rising / F.H. Allen // Acta Crystallogr., Sect. B: Struct. Sci. – 2002. – Vol.58. – P. 380-388.
- [6] Blatov, V.A. Applied Topological Analysis of Crystal Structures with the Program Package ToposPro / V.A. Blatov, A.P. Shevchenko, D.M. Proserpio // Cryst. Growth Des. – 2014. – Vol.14. – P. 3576-3586.

# Fingerprint-vector as a descriptor of the coordination figure

A.P. Shevchenko<sup>1</sup>

<sup>1</sup>Samara National Research University, Moskovskoye shosse, 34, Samara, Russia, 443086

**Abstract.** A new method for determining the geometric shape of the local atomic environment in crystals, presented in the form of a coordination figure, is proposed. The method based on the comparison of angular fingerprints of polyhedra was introduced into the ToposPro software package and tested on 9642 crystal structures of five-coordinated copper-zinc complexes. The results obtained were used to find correlations of the coordination figure type with other structural descriptors.

**Keywords:** coordination figure, local atomic environment, crystal design, big data.