

АЛГОРИТМ ПОИСКА ТОЧЕЧНЫХ ПОДМНОЖЕСТВ И ЕГО ПРИМЕНЕНИЕ ДЛЯ АНАЛИЗА АТОМНОЙ СТРУКТУРЫ МОДЕЛЬНЫХ КЛАСТЕРОВ

Д.С. Крупянский, А.Д. Фофанов

В настоящей статье представлены результаты по разработке метода исследования атомной структуры кластеров, формируемых при компьютерном моделировании. Данный метод основан на поиске координационных многогранников в исследуемых кластерах и построении графа, описывающего их взаимное расположение. Далее метод предполагает расчет ряда топологических индексов для полученного графа с целью их дальнейшего сопоставления с физико-химическими свойствами соответствующих кластеров. Для нахождения координационных многогранников предложен алгоритм поиска подмножеств в конечных точечных множествах по шаблону. В ходе работы было исследовано несколько различных по форме, структуре и составу кластеров. Также было предложено несколько простейших инвариантов графа, отражающих особенности структуры исследуемых кластеров. Представленный алгоритм реализован в компьютерной программе, позволяющей производить поиск координационных многогранников, строить соответствующий граф и рассчитывать предложенные инварианты.

Ключевые слова: поиск точечных подмножеств; моделирование атомной структуры; анализ структуры.

Введение

В задачах структурного анализа атомную структуру вещества часто интерпретируют как множество точек трехмерного пространства. Также с каждой точкой такого множества сопоставлено некоторое целое число t , соответствующее типу атома или иона. При таком подходе модель структуры вещества описывается множеством наборов вида $\{t_i; x_i; y_i; z_i\}$, где t_i – тип частицы, а x_i , y_i и z_i – ее координаты. Временную эволюцию такой системы можно отследить с помощью молекулярно-динамического эксперимента, в результате проведения которого будут известны не только термодинамические параметры этой системы, но и координаты всех составляющих ее частиц.

Для оценки степени упорядоченности проэволюционировавшей системы необходимы алгоритмы, позволяющие найти в этой системе координационные многогранники и получить данные об их взаимном расположении такие, как информация о соседствующих многогранниках и об объединении таких многогранников в группы. Под соседством в данном случае понимается наличие у двух многогранников общей вершины, ребра или грани. В зависимости от количества соседей можно выделить несколько типов координационных многогранников. В частности, для описания структуры силикатных расплавов и стекол в настоящее время широко используют [1, 2] структурные единицы Q^n – кремнево-кислородные тетраэдры SO_4 , где n – число мостиковых (принадлежащих двум атомам кремния) атомов кислорода. Такой подход может быть распространен на любые типы координационных многогранников. В соответствии с этим, многогранники, не имеющие соседей, будем называть многогранниками типа Q^0 , имеющие одного соседа – Q^1 , двух соседей – Q^2 и так далее.

В данной статье представлен результат исследования по разработке метода анализа атомной структуры кластеров, формируемых при компьютерном моделировании неупорядоченных систем. Данный метод основывается на анализе информации о взаимном расположении координационных многогранников, описываемых конечным набором точек. Для

поиска этих многогранников предложен алгоритм, позволяющий в произвольном конечном множестве точек выявить подмножества, которые совпадают при наложении с некоторым заранее определенным набором точек (шаблоном). При этом поиск производится с учетом заданного допустимого отклонения расстояний между точками. Также рассмотрены простейшие способы анализа результатов поиска, в основе которых лежит рассмотрение графа, вершины которого соответствуют найденным координационным многогранникам, а ребра – их соединениям посредством общих вершин, ребер, граней.

Реализованная программа позволяет производить поиск координационных многогранников в компьютерной модели атомной структуры кластера, строить соответствующий граф и рассчитывать численные параметры, характеризующие его структуру. Программа состоит из двух модулей: модуля поиска точечных подмножеств и модуля анализа результатов поиска.

1. Существующие методы анализа структуры вещества

Решению задачи исследования структуры компьютерных моделей систем атомов посвящено множество работ. В работе [3] описан алгоритм сопоставления точечных множеств, позволяющий выявить степени различия или подобия упорядоченных структур. Для анализа неупорядоченных систем часто используется функция радиального распределения атомов $g(r)$, определяющая плотность вероятности нахождения какой-либо частицы на расстоянии r от некоторой исходной частицы. Однако, для аморфных тел функция $g(r)$ затухает при некоторых значениях r_k , называемых дальностью корреляции, и поэтому удобна для описания только ближнего порядка. В работах [4, 5] для исследования компьютерных моделей жидкости и аморфных тел применялась разностная радиальная функция распределения, позволяющая оценить характер дальних осцилляций.

Также удачным средством для исследования структуры жидкостей оказался подход, основанный на геометрических методах Вороного-Делоне [6]. Многогранник Вороного является непосредственным геометрическим образом ближайшего окружения атома, вершины этого многогранника являются центрами симплексов Делоне, заполняющих пространство без зазоров и наложений. При использовании этих методов для анализа структуры рассматриваются различные метрические характеристики многогранников Вороного и симплексов Делоне. В работе [7] в качестве количественной характеристики формы симплекса используется длина его самого длинного ребра; в работах [5, 8] используются индексы тетраэдричности и октаэдричности, характеризующие форму симплекса; в монографии [6] в качестве метрических характеристик многогранника предлагается использовать его объем, площадь поверхности и коэффициент сферичности.

В теоретической химии для изучения молекулярной структуры применяются такие математические теории, как топология и теория графов. С формальной точки зрения структура молекулы интерпретируется как граф, вершины которого соответствуют атомам, а ребра – связям. Различные инварианты (топологические индексы) этого графа зависят от структуры, поэтому производятся попытки увязать их значения с теми или иными физико-химическими свойствами различных веществ [9, 10].

2. Постановка задачи

Формально задача поиска точечных подмножеств состоит в следующем.

Пусть в n -мерном евклидовом пространстве \mathbb{E}^n заданы два конечных множества точек X и Y , также задано некоторое конечное множество натуральных чисел $M \subset \mathbb{N}$. Причем каждой точке множеств X , Y поставлено в соответствие некоторое натуральное число $t \in M$,

обозначающее тип этой точки. Множество Y будем называть шаблоном поиска. Задача состоит в том, чтобы найти все такие подмножества $\tilde{X} \in X$, равномощные с Y , чтобы при наложении друг на друга множества \tilde{X} и Y совпадали с точностью до некоторого заранее заданного положительного вещественного числа ε . При этом, будем считать равномощные множества \tilde{X} и Y совпадающими с точностью до ε , если из элементов множества \tilde{X} можно составить последовательность $\{\tilde{x}_i\}$, а из элементов множества Y – последовательность $\{y_i\}$, так, чтобы $t_{\tilde{x}_i} = t_{y_i}$, а $|\rho(\tilde{x}_i, \tilde{x}_j) - \rho(y_i, y_j)| \leq \varepsilon$ для любых допустимых i, j .

3. Поиск точечных подмножеств

Решение поставленной задачи представляет собой некоторый набор искомым подмножеств исходного точечного множества. Расположение этих подмножеств в пространстве может быть произвольным: они могут иметь общие точки, либо не пересекаться. Однако, для поиска координационных многогранников требуются дополнительные ограничения на взаимное расположение найденных подмножеств в пространстве. Например, для этой задачи недопустимо их перекрывание (рис. 1, г), т.к. в этом случае возникает неопределенность в том, какое из найденных подмножеств является координационным многогранником.

Определение 1. Будем говорить, что два точечных подмножества $X, Y \subset \mathbb{E}^n$ перекрываются, если пересечение относительных внутренних точек их выпуклых оболочек не пусто, то есть $riConvX \cap riConvY \neq \emptyset$.

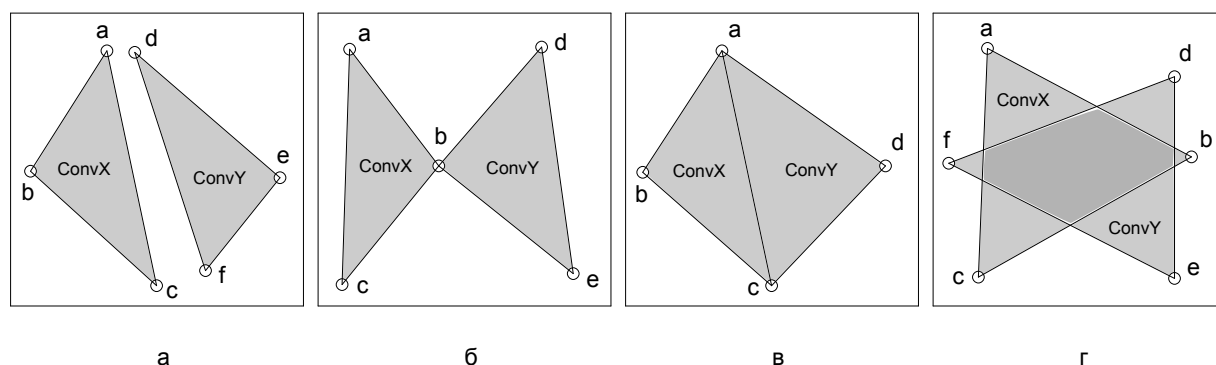


Рис. 1. Варианты взаимного расположения точечных множеств на плоскости

На рис. 1 представлены некоторые варианты взаимного расположения двух точечных множеств X и Y на плоскости: (а) – множества $X = \{a, b, c\}$ и $Y = \{d, e, f\}$ не пересекаются; (б) – множества X и Y и их выпуклые оболочки $ConvX$ и $ConvY$ пересекаются в точке b ; (в) – множества X и Y имеют две общие точки a и c , а их выпуклые оболочки – отрезок $[a, c]$; (г) – множества X и Y не пересекаются, но перекрываются в смысле определения 1.

Частично решить проблему перекрывающихся подмножеств можно путем введения некоторого условия, позволяющего выделить из всех перекрывающихся подмножеств одно, «наиболее подходящее». Например, при поиске треугольников на плоскости таким критерием может служить радиус описанной окружности. Так, если будет найдено несколько пересекающихся треугольников, то в соответствии с этим критерием выбрать нужно будет тот, радиус описанной окружности которого наиболее близок к радиусу описанной окружности шаблонного треугольника. Для точечных множеств более сложной конфигурации вместо радиуса описанной окружности, например, можно использовать диаметр множества, определяемый как $\sup_{x,y \in M} \rho(x, y)$, где $M \subset \mathbb{E}^n$.

Алгоритм поиска.

Входные данные: массив X , содержащий координаты и тип точек, среди которых будет производиться поиск; массив Y , содержащий координаты и тип точек, составляющих шаблон поиска; вещественное число ε , определяющее допустимое отклонение расстояний между точками.

Выходные данные: двумерный массив R , каждая строка которого содержит номера точек входного массива X , совпадающих с шаблоном поиска с максимальным отклонением расстояний ε .

Шаг 1. Формируем матрицу расстояний (r_{ij}) , где r_{ij} – расстояние между i -ой и j -ой точками массива Y , и определяем пустой массив R для хранения результатов поиска;

Шаг 2. Выбираем из массива Y первую точку y_k ;

Шаг 3. Выбираем из массива X первую точку x_i ;

Шаг 4. Если тип точки x_i совпадает с типом точки y_k , переходим на Шаг 6;

Шаг 5. Выбираем из массива X следующую точку x_i и переходим на Шаг 4;

Шаг 6. Формируем массив F , содержащий один элемент – индекс точки x_i в массиве X ;

Шаг 7. Выбираем из массива Y следующую точку y_k ;

Шаг 8. Выбираем из массива X все точки x_j , такие, что: индекс точки x_j не содержится в массиве F , тип x_j совпадает с типом y_k , а расстояния от точки x_j до всех точек из X , номера которых содержатся в F , совпадают с соответствующими элементами матрицы расстояний (r_{ij}) с точностью до ε ;

Шаг 9. Если таких точек нет, переходим к Шагу 5. Иначе для каждой точки x_j формируем новый массив F'_j , содержащий все индексы точек из массива F , а также индекс точки x_j в массиве X ;

Шаг 10. Если все точки из массива Y были выбраны, то в массивах F'_j содержатся результаты поиска – записываем номера точек из каждого массива F'_j в конец массива R и переходим к Шагу 11. Иначе, последовательно обозначая каждый из получившихся массивов F'_j за F , продолжаем выполнение алгоритма с Шага 7;

Шаг 11. Выводим содержимое массива R и завершаем выполнение алгоритма.

Предложенный алгоритм разработан на основе алгоритма сопоставления точечных множеств [3] и реализован в программе, предназначенной для анализа атомной структуры кластеров, формируемых при компьютерном моделировании. Атомные кластеры и координационные многогранники описываются наборами точек в трехмерном пространстве, поэтому разработанный алгоритм может быть применен к решению данной задачи.

4. Реализация

Для применения предложенного алгоритма к задаче исследования структуры кластеров была разработана программа, состоящая из двух модулей: модуля поиска точечных подмножеств и модуля анализа результатов.

Модуль поиска точечных подмножеств.

Данный модуль предназначен для поиска в массиве, описывающем некоторую систему атомов, заранее определенных структур. В большинстве проводимых экспериментов такими структурами являлись координационные многогранники.

Входные параметры: массив координат и типов атомов, в котором будет осуществляться поиск; массив координат и типов атомов, описывающий шаблон поиска; положительное вещественное число ε , определяющее допустимые отклонения расстояний.

Суть работы модуля заключается в рекурсивном построении подсистем атомов исходного массива, совмещаемых с системой атомов, определяемой шаблоном поиска, операциями параллельного переноса, поворота и зеркального отражения, в соответствии с алгоритмом, изложенным выше.

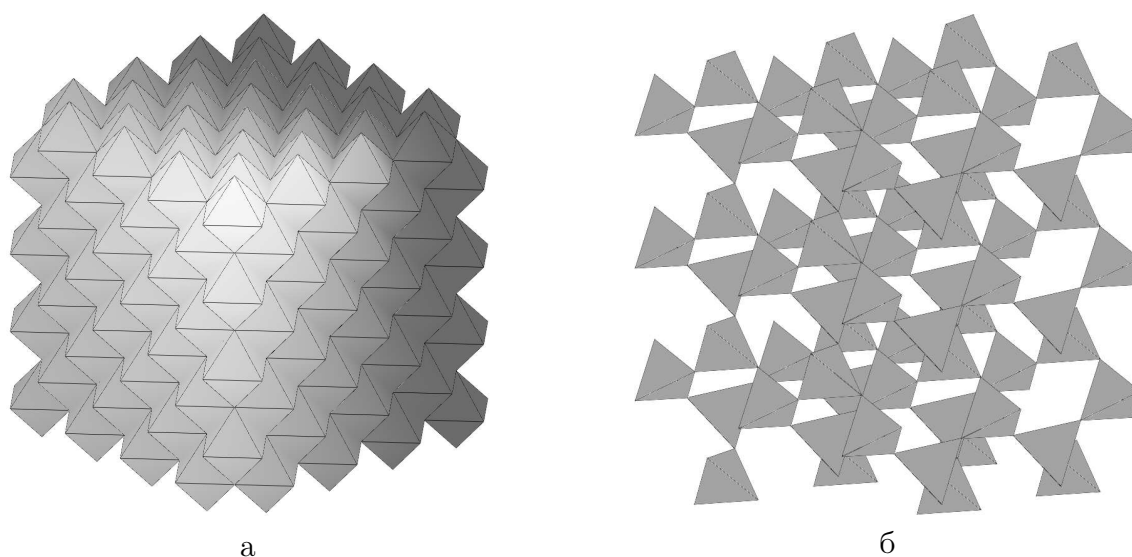


Рис. 2. Результаты поиска: (а) – кобальт-кислородных октаэдров в структуре оксида кобальта; (б) – кремниво-кислородных тетраэдров в структуре α -кварца

Результатом работы модуля является двумерный массив, каждая строка которого содержит номера атомов, образующих одно из найденных подмножеств. На рис. 2 представлена визуализация результата работы программы. В данном случае производился поиск: (а) – кобальт-кислородных октаэдров в упорядоченной структуре оксида кобальта (CoO); (б) – кремниво-кислородных тетраэдров в кристалле α -кварца. Также работоспособность реализованного алгоритма была проверена на некристаллических структурах (Рис. 3).

В ходе отладки была проведена проверка результатов поиска на корректность. Также оценивалось время выполнения алгоритма в зависимости от размеров обрабатываемых массивов и наличие перекрываемости оболочек найденных подмножеств в зависимости от значения параметра ε .

Оптимальное значение параметра ε определяется целью поиска. Так, например, при меньших значениях параметра можно обнаружить структуры, наиболее близкие к идеальному шаблону, при бóльших – более деформированные и зачастую перекрывающиеся.

На рис. 3 приведен результат поиска кремниво-кислородных тетраэдров (SiO_4): (а) – в структуре жидкого стекла легированного кобальтом $\text{Na}_2\text{Si}_4\text{O}_{10}\text{Co}$; (б) – в структуре аморфного оксида кремния (SiO_2). Длина ребра (связь $\text{O}-\text{O}$) идеального тетраэдра (шаблона) составляет $\sim 2,66 \text{ \AA}$, а значение параметра ε принято за $0,5 \text{ \AA}$. При таком значении параметра ε лишь незначительная часть найденных тетраэдров перекрывают друг друга, при этом сравнение радиусов описанных сфер перекрывающихся тетраэдров с радиусом описанной сферы шаблонного тетраэдра позволяет отбросить «лишние» тетраэдры и получить более полную информацию о структуре системы атомов по сравнению с информацией, полученной при меньшем значении ε .

Модуль анализа результатов

Данный модуль предназначен для построения графа, описывающего структуру исходной системы атомов на уровне найденных подмножеств (координационных многогранников), а также для расчета инвариантов этого графа, которые, по нашему мнению, могут быть использованы для количественного описания структуры кластеров, формируемых при компьютерном моделировании. Вершины этого графа соответствуют найденным подмножествам, а каждое его ребро – одной общей точке двух пересекающихся подмножеств.

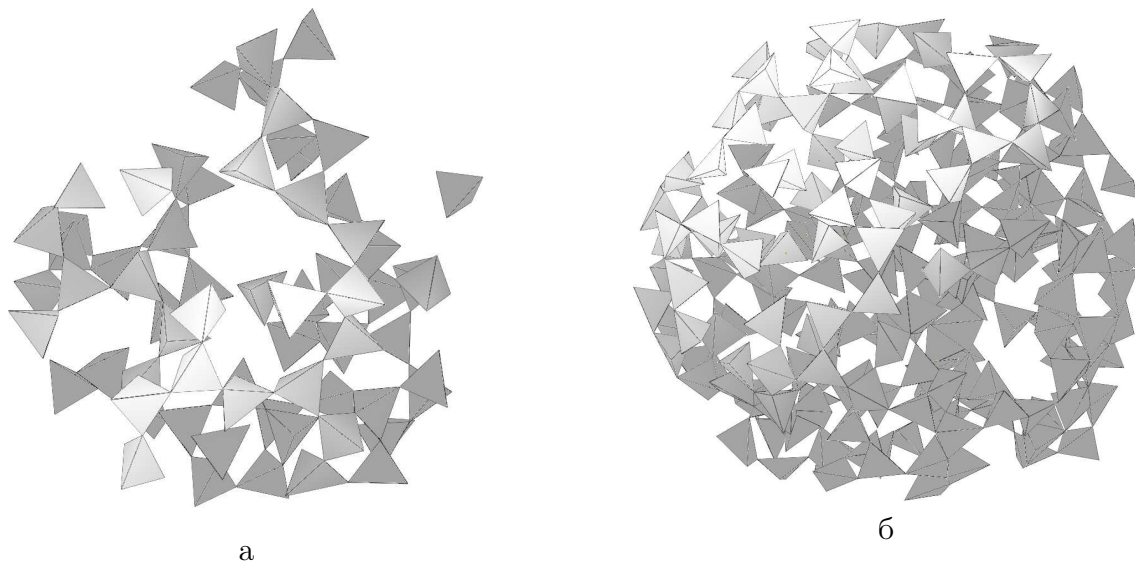


Рис. 3. Результат поиска кремнево-кислородных тетраэдров: (а) – в структуре жидкого стекла, легированного кобальтом; (б) – в структуре аморфного оксида кремния

В общем случае полученную с помощью модуля поиска систему подмножеств можно описать некоторым несвязным мультиграфом. Анализ структуры этого графа с учетом дополнительной информации о найденных подмножествах (например, индекс тетраэдричности [5, 7] для четырехточечных подмножеств), а также расчет различных топологических индексов может быть использован для поиска аналитических соотношений типа структура–свойство.

Входные данные: массив координат и типов атомов, в котором производился поиск (необходим для определения степени искаженности); двумерный массив результатов поиска, полученный с помощью модуля поиска.

Основным результатом работы модуля является массив вершин графа и массив его ребер. На основе этих массивов формируется двумерный массив, описывающий компоненты связности полученного графа. Выводится информация о количестве компонент связности, их средней и максимальной длине. Также для многогранников выводятся данные об их распределении по типам $Q^0 - Q^n$ [1].

В таблице приведены характеристики четырех исследуемых кластеров (α -кварца, жидкого стекла, легированного кобальтом ($\text{Na}_2\text{Si}_4\text{O}_{10}\text{Co}$), аморфного оксида кремния (SiO_2) и кристаллического оксида кобальта (CoO)) и результаты их анализа, полученные с помощью данного программного модуля. Из таблицы видно, что численные характеристики такие как, количество компонент связности и среднее количество вершин в компоненте, сильно разнятся для кластеров в кристаллическом и аморфном состоянии. Для кластеров в кристаллическом состоянии характерно наличие у соответствующего графа только одной компоненты связности, поэтому среднее количество вершин в компоненте совпадает с общим количеством вершин. Графы неупорядоченных кластеров имеют множество компонент связности, многие из которых состоят из одной изолированной вершины. Таким образом, даже простейшие численные характеристики графа, построенного на основе информации о взаимном расположении координационных многогранников, отражают особенности атомной структуры кластера. Расчет и объединение подобных характеристик в более сложные интегральные топологические индексы может быть использован для построения корреляций типа структура–свойство.

Таблица

Результаты анализа различных кластеров

	α -кварц	$Na_2Si_4O_{10}Co$	SiO_2	CoO
Количество частиц в кластере	1125	1190	1200	1000
Форма кластера	паралл-д	шар	шар	куб
Диаметр кластера, Å	45,6	35,3	33,6	33,1
Площадь поверхности кластера, 1000Å^2	3,2	3,1	2,6	2,2
Объем кластера, 1000Å^3	14,1	17,0	11,2	6,6
Доля атомов в поверхностном слое толщиной $2,5\text{Å}$, %	59,4	45,3	53,0	78,4
Тип координационного многогранника	тетраэдр	тетраэдр	тетраэдр	октаэдр
Отклонение расстояний ε , Å	0,4	0,4	0,4	0,5
Количество найденных многогранников	280	186	185	256
Среднее количество смежных вершин по всем вершинам	3,20	1,80	1,83	9,19
Количество компонент связности	1	33	29	1
Среднее количество вершин в компоненте	280,00	5,64	6,38	256,00
Количество изолированных вершин (Q^0)	0	16	20	0
Порядок наибольшей компоненты связности	280	43	140	256

5. Заключение

В результате данной работы предложен метод исследования атомной структуры кластеров, формируемых при компьютерном моделировании, основанный на анализе взаимного расположения координационных многогранников. Также предложен алгоритм поиска подмножеств в конечных точечных множествах, лежащий в основе данного метода. Он описан в самом общем виде и может быть использован для решения широкого круга задач. Проверка программной реализации этого алгоритма была выполнена на кластерах различной формы, структуры и состава. Анализ результатов работы программы показал, что предложенный метод может быть использован для исследования неупорядоченных атомных систем.

Работа выполнена при поддержке Программы стратегического развития (ПСР) ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности на 2012 – 2016 гг.

Литература

1. Анфилогов, В.Н. Силикатные расплавы / В.Н. Анфилогов, В.Н. Быков., А.А. Осипов. – М.: Наука, 2005. – 357 с.

2. Королева, О.Н. Физико-химическая модель натриевосиликатного расплава и термодинамика Q^n -единиц / О.Е. Королева, А.А. Тупицын, В.А. Бычинский // Вестник ЮУрГУ. Серия: Химия. – 2012. – № 36. – С. 39–44.
3. Тарачева, И.А. Решение задачи сопоставления точечных множеств для выявления общих подмножеств / И.А. Тарачева, Б.М. Щедрин // Кристаллография. – 1994. – Т. 39, № 4. – С. 586–589.
4. Волошин, В.П. Радиальные функции распределения атомов и пустот в больших компьютерных моделях воды / В.П. Волошин, Н.Н. Медведев, Ю.И. Наберухин, А. Гайгер, М. Клене // Журнал структурной химии. – 2005. – Т. 46, № 3. – С. 451–458.
5. Наберухин, Ю.И. Структура больших некристаллических леннард-джонсовских моделей / Ю.И. Наберухин, В.П. Волошин // Журнал структурной химии. – 2006. – Т. 47, № 7. – С. 129–143.
6. Медведев, Н.Н. Метод Вороного-Делоне в исследовании структуры некристаллических систем / Н.Н. Медведев. – Новосибирск: Изд-во СО РАН, 2000. – 214 с.
7. Anikeenko, A.V. Polytetrahedral Nature of the Dense Disordered Packings of Hard Spheres / A.V. Anikeenko, N.N. Medvedev // Physical review letters. – 2007. – 98(23), 235504.
8. Anikeenko, A.V. Shapes of Delaunay Simplexes and Structural Analysis of Hard Sphere Packings / A.V. Anikeenko, M.L. Gavrilova, N.N. Medvedev // in book: Generalized Voronoi Diagram: A Geometry-Based Approach to Computational Intelligence. – 2008. – SCI Vol.158 – pp. 13–45.
9. Зефирова, Н.С. Применение теории графов в химии / Н.С. Зефирова, С.И. Кучанов. – Новосибирск: Наука, 1988. – 306 с.
10. Кинг Р. Химические приложения топологии и теории графов / Р. Кинг. – Москва: Мир, 1987. – 560 с.

Дмитрий Сергеевич Крупянский, аспирант, кафедра «Физика твердого тела», Петрозаводский государственный университет (г. Петрозаводск, Российская Федерация), krupjanski@rambler.ru.

Анатолий Дмитриевич Фофанов, доктор физико-математических наук, профессор, кафедра «Физика твердого тела», Петрозаводский государственный университет (г. Петрозаводск, Российская Федерация), afofanov@psu.karelia.ru.

Поступила в редакцию 16 декабря 2013 г.

Bulletin of the South Ural State University.
Series "Mathematical Modelling, Programming & Computer Software",
2014, vol. 7, no. 2, pp. 46–54.

MSC 68U01

DOI: 10.14529/mmp140204

An Algorithm Searching for Point Subsets with Applications to the Analysis of the Atomic Structure of Modelled Clusters

D.S. Krupjanskiy, Petrozavodsk State University, Petrozavodsk, Russian Federation, krupjanski@rambler.ru,

A.D. Fofanov, Petrozavodsk State University, Petrozavodsk, Russian Federation, afofanov@psu.karelia.ru

