  
Министерство образования и науки российской федерации  
  
федеральное государственное АВТОНОМНОЕ   
образовательное учреждение высшего образования  
«САМАРСКИЙ национальный исследовательский   
УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»  
(Самарский университет)

Факультет информатики

Кафедра программных систем

**На правах рукописи**

**Шашов Кирилл Владимирович**

**Автореферат магистерской диссертации**

**(выпускной квалификационной работы магистра)**

**Эвристический алгоритм поиска   
глобально-оптимальной конформации   
атомного кластера Морса**

Направление подготовки магистров – «020402.68 Фундаментальная информатика и информационные технологии»

Магистерская программа – «Инженерия программного обеспечения»

Научный руководитель

зав. каф. ПС, А.Н. Коварцев

Магистрант

К.В. Шашов

Самара, 2017 г.

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы диссертации.** Задачи глобальной оптимизации доказали свою актуальность во многих прикладных областях математики, информатики, физики, биологии, химии и смежных с ними отраслями знаний. Широко обсуждаются разновидности важной для нефтехимической отрасли задачи о транспортных сетях с пулами, различные формулировки задачи об оптимальном распределении ресурсов, планировании ресурсов, составлении расписаний, задачи об оптимальном расположении фигур на плоскости, задачи, связанные с расчетом электронной плотности внутри элементарной ячейки белкового кристалла, геометрические задачи восстановления трехмерных структур сцены по нескольким изображениям.

Кроме того, задачи глобальной оптимизации так же рассматривают в пределах интересов конформационного анализа, занимающегося поиском оптимальных конформаций (структур) молекул, исследованием их химических и физических свойств и т.д. Совместные усилия ученых, направленные на решение этих задач, имеют своей целью открыть новые возможности интерпретации данных, полученных на моделирования новых биологически активных соединений, в том числе для целей создания лекарственных препаратов.

В последние годы геномные исследования заметно увеличили число установленных аминокислотных последовательностей белков, опережая тем самым возможности экспериментальных методов определения пространственных структур белковых молекул. Поэтому возникший дефицит потенциально возможных пространственных структур белков стимулирует современные методы компьютерного моделирования.

**Цель исследования**. Целью выпускной квалификационной работы является разработка эвристического алгоритма поиска атомных структур, обладающих минимальными значениями потенциала Морса.

**Задачи исследования.** Для достижения указанной цели были поставлены следующие задачи:

1. Рассмотреть особенности постановки задачи глобальной оптимизации при поиске атомных структур.
2. Проанализировать существующие методы поиска оптимальных атомных структур.
3. Исследовать особенности формирования конформаций кластеров Морса, выявить закономерности для их дальнейшего использования в алгоритме при построении начальных конформаций.
4. Разработать эвристический алгоритм глобальной оптимизации атомных кластеров Морса.
5. Разработать программную реализацию алгоритма, провести тестирование и отладку.
6. Провести вычислительные эксперименты.

**Практическая значимость полученных результатов.** В ходе работы были разработаны алгоритм оптимизации атомных кластеров и его программная реализация, позволяющие отыскивать глобальные оптимумы больших размерностей. Опыт практических расчетов стационарных точек на потенциальных поверхностях сложной структуры может быть использован, например, в химии для исследования структуры биомолекул.

**Публикации по теме диссертации.** По итогам работы была опубликована 1 статья в сборнике международной конференции.

**Структура и объем диссертации.** Работа состоит из введения, 4 глав и заключения. Диссертация изложена на 79 страницах машинописного текста, библиография включает 44 наименований.

# Основное содержание работы

**Во введении** раскрывается актуальность, определяется степень научной разработки темы, цель исследования, теоретическая и практическая значимость работы.

Знание деталей формирования кластерной структуры является определяющим для повышения эффективности различных способов производства нанокластеров с фиксированными физическими свойствами, что открывает широкие перспективы для их практического применения при создании новых веществ с заданными механическими, электрическими, магнитными и оптическими свойствами.

Поскольку сложно идентифицировать структуру свободных кластеров в молекулярных пучках, обычно оценивают параметры, зависящие от структуры, и используют модели с прогнозируемыми предпочтительными геометрическими конфигурациями. Одной из основных задач при формировании кластерной структуры является минимизация энергии взаимодействия входящих в нее атомов.

Было доказано, что сложность задачи определения минимума потенциальной энергии кластера, в котором между атомами учитывается парное взаимодействие, принадлежит к классу NP-полных задач, и до сих пор не известен не экспоненциальный алгоритм ее решения. На практике наивысших результатов добились генетические и эвристические алгоритмы, учитывающие известные закономерности формирования кластерных структур [1]. Тем не менее на сегодняшний день из-за высокой вычислительной сложности данная задача остается актуальной и слабо проработанной для структур размерностью более 200 атомов.

**В первой главе** работы рассмотрена задача глобальной оптимизации, в частности, особенности ее постановки и решения при работе с атомными структурами. Приведен обзор основных существующих методов поиска конформаций атомных кластеров. Рассмотрены генетические методы, методы семейства BH (basin hopping), методы поиска на динамической сетке с конструированием ядра.

***Задача глобальной оптимизации атомных кластеров***

Задача поиска оптимальных конформаций атомных кластеров может быть сформулирована как следующая задача глобальной оптимизации:

, (1)

где – число атомов кластера, – координаты центра - ого атома, – евклидова норма, – парный потенциал взаимодействия.

Наиболее распространенные в исследованиях потенциальные функции – потенциал Леннарда-Джонса:

(2)

и потенциал Морса:

, (3)

где – параметр, характеризующий физические свойства кластера.

Обе приближенные модели потенциала часто используются для описания кластеров с определенным типом атомов, например, некоторых газов. На рисунке 1 изображены графики потенциалов кластеров Морса и Леннарда-Джонса.

Задача глобальной оптимизации кластеров Морса часто используется в роли проверочной системы для алгоритмов глобальной оптимизации [2]. Однако до сих пор ни один алгоритм не сумел обнаружить все глобальные минимумы потенциала Морса без привлечения дополнительных знаний о предметной области – о структурных, термодинамических закономерностях и т.д.

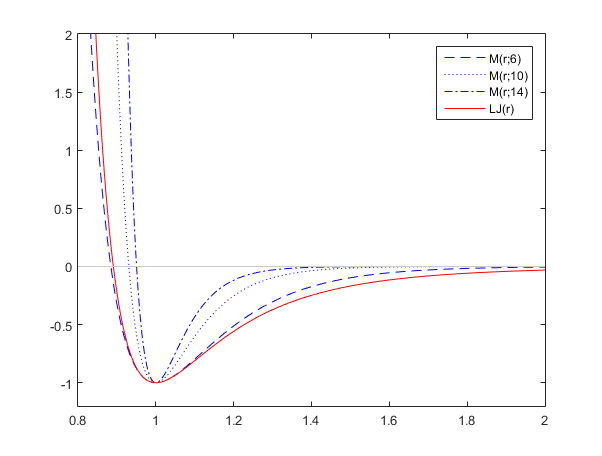


Рисунок 1 – Потенциалы кластеров Морса и Леннарда-Джонса

***Обзор существующих методов оптимизации***

На текущий момент прямые методы способны находить решение лишь для функций с десятками переменных, однако в задачах оптимизации атомных кластеров участвует более 900 оптимизируемых переменных. Таким образом, основной трудностью при поиске глобального минимума является сильное усложнение поверхности потенциальной энергии при увеличении размеров кластера.

Наиболее популярные методы оптимизации можно разделить на три группы. Первая группа включает эволюционные и генетические алгоритмы, методы быстрого отжига (fast annealing evolutionary algorithm, FAEA), метод роя частиц (particle swarm optimization, PSO), алгоритм адаптивной свободной оптимизации (adaptive immune optimization algorithm, AIOA), популяционный алгоритм (populationbased algorithm).

Вторая группа алгоритмов, основанная на идее трансформирования энергетической поверхности, включает метод случайного спуска (basin-hopping, BH) и его модификации, алгоритм скачкообразных минимумов (minima hopping), алгоритм перемещения по воронкам (funnel hopping method).

Алгоритмы третьего класса осуществляют поиск на так называемой динамической сетке (dynamic lattice, DL), перемещая атомы с наивысшей энергией на вакантные узлы решетки с наименьшей энергией, тем самым позволяя отыскивать возможные местоположения исходного минимума с помощью построенных сеток.

Существуют и другие эффективные алгоритмы, такие как методы большого взрыва (big-bang method) и непрерывной экстремальной оптимизации (continuous extremal optimization, CEO).

**Во второй главе** описана методика решения поставленной задачи, которая лежит в основе разработанного алгоритма оптимизации.

***Формирование плотной упаковки атомов***

Во избежание экспоненциального роста сложности в предложенном алгоритме осуществляется переход от задачи пространственной оптимизации 3N переменных к задаче оптимизации одной переменной. Такой переход становится возможным благодаря использованию заранее выгодной геометрической структуры.

Практикой доказано, что икосаэдрические, додекаэдрические и гранецентрированные решетки являются хорошим начальным приближением для поиска глобального минимума потенциальной энергии. Благодаря этой закономерности можно заранее построить плотную упаковку атомов, и уже в ней осуществлять поиск наиболее выгодной конформации.

В предложенным алгоритме используется геометрически обоснованный метод [3]. Основу данного метода составляет алгоритм формирования конфигураций для размещения центров атомов, с помощью которых строятся пространственные слои плотной упаковки шаров.

Использование заранее сформированной решетки атомов позволяет перейти к более простой задаче структурной оптимизации, а именно к выбору наиболее выгодных N атомов из M доступных.

***Оптимизация функции одной переменной***

Полученную задачу структурной оптимизации можно привести к классической задаче глобальной оптимизации функции одной переменной:

, (4)

где – численное представление конформации, – потенциальная энергия кластера. Имея решетку атомов, конфигурацию атомов можно представить в виде двоичного числа длиной , имеющее ровно единиц.

Такой переход к новой постановке задачи позволил бы использовать существующие методы глобальной оптимизации, однако целевая функция не отвечает требованиям таких методов, например, непрерывности, поэтому требуется разработка собственного метода оптимизации.

Хотя методы глобальной оптимизации неприменимы или неэффективны для поставленной задачи, тем не менее можно использовать их адаптированные версии. В предложенном алгоритме ключевое место занимает модифицированный метод Стронгина [4]. Данный статистический метод старается размещать точки испытаний либо в окрестностях локальных минимумов оптимизируемой функции, либо в интервалах неопределенности, размеры которых велики по сравнению с другими участками функции. Метод Стронгина определяет некоторое количество атомов, из которых составляется конфигурация, в дальнейшем достраиваемая алгоритмом развития до необходимого размера .

В процессе работы метода Стронгина сохраняется несколько конфигураций, которым соответствовали наименьшие значения потенциальной энергии. Причем, для каждой малой окрестности алгоритм хранит не более одной конфигурацию, тем самым обеспечивая разнообразие найденных структур. Все найденные структуры подвергаются процедуре локальной оптимизации

Таким образом, искомая конформация атомов образуется поэтапно, благодаря нескольким алгоритмам и эвристикам, а значение N можно представить в виде суммы:

, (5)

где – количество атомов начальной конфигурации, – количество атомов, определяемых методом Стронгина, – количество «достраиваемых» атомов алгоритмом развития.

Подобное использование алгоритмов напоминает по своей структуре генетические методы. В данном случае родителями являются начальная структура, метод Стронгина, и алгоритм развития.

**В третьей главе** описаны теоретические и вычислительные методы, адаптации известных алгоритмов оптимизации и прочие эвристики, использованные в работе для определения оптимальных структур кластеров.

***Метод Стронгина***

Метод Стронгина позволяет находить абсолютный минимум функции на отрезке. Проводится ряд испытаний, в результате которых начальный интервал поиска делится на несколько меньших интервалов. Очередная точка испытаний выбирается внутри интервала, на котором наиболее вероятен абсолютный минимум, добавляется в список известных значений, после чего происходит переход к следующей итерации. Положение точки в интервале соответствует математическому ожиданию положения минимума. Алгоритм останавливается, когда расстояние между точками очередного делимого отрезка становится меньше заданного критерия.

Для использования метода Стронгина в предложенной методике была использована целевая функция, использующая специально разработанный алгоритм развития.

В качестве начального интервала используется отрезок, составленный в зависимости от размеров искомой конформации и решетки атомов .

Поскольку в нашем случае метод Стронгина из-за низкой скорости сходимости используется для определения лишь первых атомов, и полная его работа не требуется, критерий остановки может быть существенно упрощен.

***Алгоритм развития***

Идея алгоритма имеет много общего с подходом, лежащим у истоков детерминированной глобальной оптимизации атомных кластеров – подход схемы развития (growth scheme) [5]. Согласно этой схеме оптимальные кластеры Леннарда-Джонса искались путем присоединения дополнительного атома к кластерам с плотной структурой (выгодной конфигурации с точки зрения вкладов атомов в общую сумму потенциальной энергии).

Основными критериями для выбора нового атома являются число соседних атомов из предыдущей конформации и вклад в итоговое значение потенциальной энергии кластера.

Данный алгоритм ориентирован на поиск связей между глобальными оптимумами и оптимумами с "сильным" ядром – т.е. структур с таким расположением атомов, которое доказало свою оптимальность для кластеров меньших размерностей.

***Локальная оптимизация***

Большинство существующих методов глобальной оптимизации атомных кластеров являются двухэтапными, используя технику локальной оптимизации кластера из близкого к оптимальному решению положения, поскольку сложность алгоритмов локальной оптимизации имеет полиномиальный характер. В предложенном алгоритме процедура локальной оптимизации применяется ко всем отобранным в метода Стронгина структурам.

В данной работе в качестве метода локальной оптимизации использовался метод BFGS (Broyden, Fletcher, Goldfarb, Shanno) с ограниченной памятью. Версия алгоритма BFGS с ограниченной памятью была опубликована Liu и Nocedal в их совместной работе [6]. Преимущество модификации заключается в эффективной работе с большим числом оптимизируемых переменных.

**В четвертой главе** приведены экспериментальное исследование параметров разработанного алгоритма, результаты вычислительных экспериментов по поиску оптимальных конформаций кластеров Морса, а также описание программной реализации алгоритма, с помощью которой проводились данные эксперименты.

***Программная реализация***

В процессе работы было реализовано приложение (рисунок 2) для поиска глобально-оптимальных структур кластеров Морса. Программа написана языке программирования Java в среде разработки IntelliJ IDEA 14. Имеется поддержка арифметики произвольной точности и запуск вычислений на нескольких вычислительных потоках.

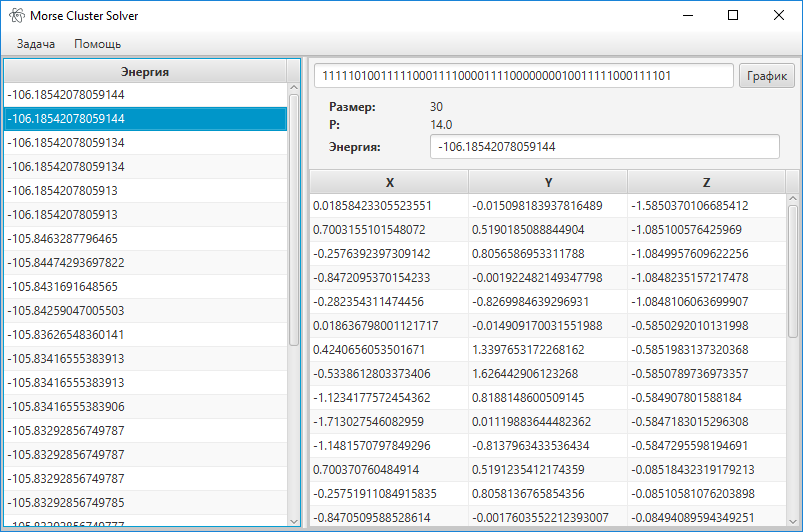


Рисунок 2 – Отображение подробной информации об атомной структуре

***Вычислительные эксперименты***

В качестве первых испытаний алгоритма и его программной реализации были успешно решены задачи глобальной оптимизации кластеров размерностью менее 40 атомов. С помощью данных вычислительных экспериментов были проверены генерация решетки атомов, метод Стронгина, алгоритм развития, локальная оптимизация.

Работа алгоритма была проверена и на значительно более трудной   
задаче – оптимизации кластера Морса размером 209 атомов при ρ=14. Алгоритм нашел сразу две структуры, которым соответствует глобальный минимум E=-990.5870484495.

**В заключении** подводятся итоги исследования, формируются окончательные выводы по рассматриваемой теме.

В процессе выполнения работы автором были рассмотрены особенности формирования атомных структур, сделан обзор существующих методов их оптимизации. Предложена методика поиска глобально-оптимальных конформаций кластеров Морса на основе использования заранее сформированной решетки атомов, модификации алгоритма Стронгина, алгоритма развития, процедур локальной оптимизации. Разработано программное приложение глобальной оптимизации атомных структур. С помощью него проведен ряд экспериментальных исследований, подтверждающих на практике применимость предложенного алгоритма для решения поставленной задачи.

# Список опубликованных работ

1. Коварцев А.Н., Шашов К.В. Эвристический алгоритм поиска оптимальных конформаций атомного кластера Морса // Сборник трудов международной конференции «Инновационные исследования: проблемы внедрения результатов и направления развития». 2017. Т. 2. С. 8-12.

# Список основной литературы

1. Коварцев А.Н. Эволюционный детерминированный алгоритм глобальной оптимизации атомных кластеров Морса // Компьютерная оптика. 2015. №39. Т. 2. С. 234-240.
2. Leary R.H. Global optimization on funneling landscapes // Journal of Global Optimization. 2000. V. 18. P. 367-383.
3. Коварцев А.Н. Геометрически обоснованный метод формирования атомных кластеров Морса больших размеров // Компьютерная оптика. 2017. №41. Т. 1. C. 118-125.
4. Стронгин Р.Г., Гергель В. П., Баркалов К. А. Параллельные методы решения задач глобальной оптимизации // Изв. вузов. Приборостроение. 2009. №52. Т. 10. С. 25-33.
5. Hoare M.R., Pal P. Physical cluster mechanism: statistic and energy surface for monotonic systems // Advances in Physics. 1971. V. 20. P. 161-196.
6. Liu D.C., Nocedal J. On the limited memory BFGS method for large scale optimization // Mathematical Programming. 1989. V. 45. P. 503-528.