

Параллельный алгоритм структурной эволюции кластеров Морса больших размеров

А.Н. Коварцев¹, Д.А. Попова-Коварцева¹, Г.П. Климашова¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Формирование атомных кластеров Морса больших размеров представляет собой сложную и трудоёмкую задачу, особенно в том случае, когда жёсткость кластера равна 14. В настоящее время известны оптимальные конформации кластеров размером только до 240 атомов. В статье предлагается параллельный стохастический алгоритм структурной эволюции кластеров Морса, позволяющий находить оптимальные конформации для кластеров больших размеров. Значительное число проблем, связанных со сложностью задачи, решаются с использованием предлагаемого алгоритма оптимизации структуры кластера. Приводятся результаты оптимизации кластеров Морса для $N > 240$.

1. Введение

Кластер Морса представляет собой довольно точную математическую модель реально существующего кластера, под которым обычно понимают объединение однотипных элементов (атомов/ионов/молекул), связанных между собой по принципу «каждый с каждым». Модель кластера, которой посвящена статья далеко не является единственной. Так, например, в литературных источниках часто используется потенциал Леннарда-Джонса. Однако в отличие от кластера Леннарда-Джонса кластер Морса обладает более сложным потенциалом и более разнообразным набором структурных конформаций. Для сравнения, приведём потенциалы Леннарда-Джонса (1) и Морса (2) [1]:

$$v(r) = LJ(r) = \varepsilon \left(\frac{r_0}{r} \right)^{12} \left(\left(\frac{r_0}{r} \right)^6 - 2 \right), \quad (1)$$

$$v(r_{ij}) = M(r, \rho) = e^{\rho(1-r)} \left(e^{\rho(1-r_{ij})} - 2 \right), \quad (2)$$

где ρ – коэффициент жёсткости кластера (в настоящей статье исследуются только «жёсткие» кластеры с коэффициентом жёсткости $\rho = 14$); r_{ij} – расстояние между атомами i и j ; $\varepsilon = r_0 = 1$.

Энергию взаимодействия всех атомов кластера можно вычислить как сумму энергий парных взаимодействий

$$v(X) = \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (i \neq j)}}^N v(r_{ij}), \quad (3)$$

где $X = (x_1, \dots, x_N)'$, $x_i \in R^3$ – координаты центров атомов кластера

При этом оптимальным кластером считается такая конформация атомов кластера, которая обеспечивает глобальный минимум потенциальной энергии кластера (3). В общем случае ещё в 1979 году учёными А. Немировским и Д. Юдиным [2] было доказано что, задача глобальной оптимизации (ГО) относится к классу NP -полных задач. С учетом большой размерности, что характерно для задач оптимизации атомных конформаций, задачи ГО фактически становятся труднорешаемыми. Именно поэтому при разработке методов глобальной оптимизации в качестве тестового примера часто использовалась задача минимизации потенциальной энергии кластеров Морса.

Интересно, что своей популярностью упомянутые модели кластеров обязаны задачам глобальной оптимизации, направленных на предсказание структуры биологических молекул, сворачивания и стыковки белков. Именно они спровоцировали исследования в направлении предсказания структуры *атомных кластеров* в качестве испытательной площадки для методов предсказания структуры биомолекул. И те, и другие часто рассматриваются в пределах интересов *конформационного анализа*, занимающегося поиском оптимальных конформаций молекул, исследованием их химических и физических свойств.

С термодинамической точки зрения, сворачивание белка представляет собой переход белковой молекулы к наиболее статистически вероятной конформации наименьшей потенциальной энергии. С кинетикой сворачивания связывают парадокс Левинталя [3], согласно которому если бы молекула небольшого белка (около 100 аминокислотных остатков) для достижения нативной формы «перебирала» все возможные конформации, этот процесс потребовал бы времени, превышающего время существования Вселенной. В то же время из эмпирических данных следует, что обычное время сворачивания белка составляет миллисекунды. Это объясняется тем, что молекула белка не принимает всех теоретически возможных конформаций – молекула находит кратчайший путь на гиперплоскости потенциальной энергии к точке, соответствующей нативной конформации белка.

Изучение кластеров имеет важнейшее значение для различных областей человеческой деятельности. Данные, полученные в ходе исследований кластеров, применяются в медицине (в том числе, при разработке новых лекарств), в металлургии (при моделировании металлов), а также при разработке новых материалов. Указанные данные также необходимы для понимания процессов катализа, конденсации паров воды в облаках, при расчёте электронных и динамических характеристик наноматериалов, создании новых источников света и т.д. Основной задачей данного направления является обнаружение такой геометрической структуры атомного кластера (конформации), которая соответствует минимуму энергии взаимодействия входящих в него атомов [4], т.е. минимальному значению потенциальной энергии кластера.

Исследование топографии поверхности потенциальной энергии признано важным по той причине, что даже в случае упрощённого допущения об энергии взаимодействия, задача нахождения минимума суммарной потенциальной энергии является крайне сложной, так как она представляет собой случай невыпуклой оптимизации с большим числом локальных минимумов [5].

Важно, что основную проблему при формировании оптимальных конформаций атомных кластеров составляет «проклятие размерности» глобальной оптимизации потенциальной функции, которую задают большое количество переменных. Теоретические исследования методов ГО приводят к выводу о том, что практически значимые прямые методы ГО для функций с числом переменных больше 30 не существуют [6]. В то время как под кластерами понимаются объединения от 5 до 2000 элементов, и на данное время исследовано всего 240 кластеров, зарегистрированных в Кембриджских базах данных [7]. В связи с этим, существующая практика поиска оптимальных конформаций основывается на использовании стохастических алгоритмов ГО. Успех их применения связан с использованием большого количества эвристик о свойствах объекта исследования. В настоящее время глобально-оптимальные конформации для кластеров Морса построены вплоть до кластеров, состоящих из 240 атомов [6]. Однако построение атомных кластеров Морса больших размеров ($N > 240$) до сих пор представляет серьёзные трудности из-за значительных вычислительных затрат.

В статье рассматриваются вопросы, связанные с формированием глобально-оптимальных конформаций атомных кластеров Морса больших размеров ($N > 240$). Предлагаемый подход основывается на поиске рациональных эвристик для формирования начальных приближений атомных кластеров Морса и использовании локальных методов оптимизации функций большого количества переменных. И, фактически, настоящая статья является продолжением статьи [8] о правилах формирования начальных приближений конформаций атомных кластеров Морса на основе геометрически обоснованного метода.

2. Постановка задачи

При поиске конформаций, которые гарантируют глобальный минимум потенциальной энергии кластера, на сегодняшний день применяются такие методы оптимизации, которые используют дополнительную информацию о свойствах кластера. Примером таких дополнительных сведений является информация о том, что решётки расположения атомов в конформации кластера Морса тяготеют к симметричной сферической форме [9]. Что, в свою очередь, привело к созданию геометрически-обоснованных методов ГО кластеров Морса.

Стоит отметить, что к идее учёта геометрических свойств конформации кластера для минимизации его потенциальной энергии часто относились с недоверием, однако впоследствии к ней не раз возвращались [10,11,12]. В данной работе рассматривается задача поиска оптимальных конфигураций кластера Морса в наиболее сложной постановке, когда $\rho = 14$ и атомы могут лишь касаться друг друга. В этом случае задача поиска оптимальных конформаций (т.е. конформаций с минимальной потенциальной энергией) атомных кластеров Морса напрямую связана с проблемой плотной упаковки шаров [13]. По умолчанию предполагается, что для кластеров наиболее важными факторами являются

- сферичность формы конформации (чем ближе к сфере, тем лучше);
- достижимость плотной «упаковки» атомов.

Для «жестких» шаров наиболее плотная упаковка получается для гранцентрированной кубической решётки. Хотя самой популярной конформацией, как свидетельствуют неоднократно проведённые эксперименты, является декаэдрическая решётка (рисунок 1). В этом случае шары заполняют чуть больше 74% объёма пространства [14]. В то же время для «мягких» атомов кластера Морса более плотные упаковки формируются для икосаэдрических и додекаэдрических сеток [11, 15, 16, 17].

Нахождение оптимальной конформации кластера заключается в поиске такого расположения атомов кластера, при котором все элементы, с учётом их взаимодействия по схеме «каждый с каждым» и игнорировании химических свойств кластера, будут расположены в пространстве оптимальным образом. Таким образом, для достижения оптимальной конформации кластера, состоящего из 30 атомов необходимо решать задачу ГО с 90 переменными, что, как уже было отмечено выше, просто не под силу современным «прямым» алгоритмам ГО.

В результате анализа [4] глобально-оптимальных конформаций кластеров Морса для $\rho = 14$, зарегистрированных в Кембриджской в базе данных [7], было выяснено, что всегда можно найти такое положение одной из координатных осей, например оси z , когда атомы кластера будут располагаться слоями, образуя нечёткую слоистую структуру (рисунок 1). Под «нечётким» слоем в данном случае понимается расположение группы атомов (шаров) таким образом, чтобы их центры находились не на плоскости, а в достаточно «тонком» слое $z_j \pm \Delta z_j$.

При этом оказалось, что конформации кластеров следующего поколения, с большим количеством атомов, «наследуют» геометрию своих предшественников: от «родительских» кластеров к «дочерним».

Однако в определенные моменты происходят качественные изменения геометрической структуры глобально-оптимальных конформаций. Из рисунка 2 видно, что для $N=201$ оптимальной становится смещенная икосаэдрическая структура. В работе [18] такая структура называется плотно упакованной (close-packed). Тем не менее, и здесь наблюдается наследование родительских свойств предшествующих кластеров.

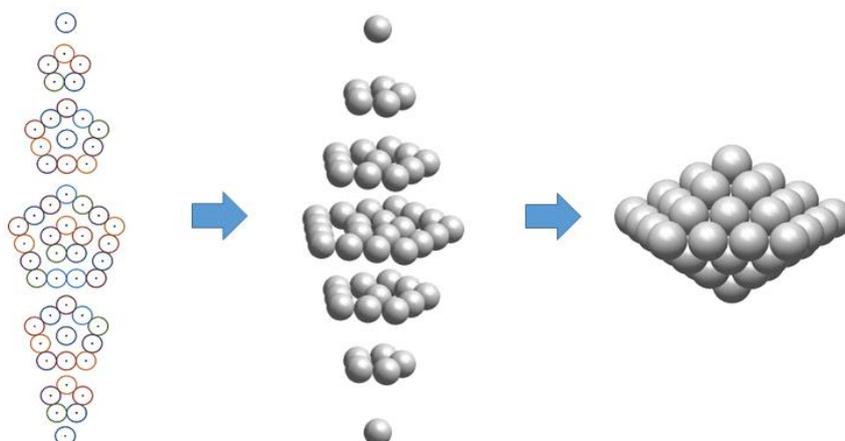


Рисунок 1. Послойное формирование структуры кластера.

Слой	N=192	N=196	N=200	N=201	N=203	N=209
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						

Рисунок 2. Слои глобально-оптимальных кластеров Морса $\rho = 14$.

Анализ зарегистрированных глобально-оптимальных конформаций атомных кластеров Морса [7] позволил обнаружить, что выбор оптимальных расположений атомов кластера происходит из ограниченной области компактного расположения атомов икосаэдрической решётки. Причём часть атомов, образующих ядро кластера, содержится во всех глобально-оптимальных кластерах икосаэдрического типа [8].

3. Алгоритм структурной эволюции кластеров Морса

Идея, предлагаемого в данной статье алгоритма, достаточно проста и базируется на использовании плотноупакованных решеток – заранее известных местах расположения атомов формируемого кластера. На рисунке 3 показана несимметричная «косая» икосаэдрическая решетка, содержащая 14 слоев и пригодная для поиска оптимальных конформаций кластеров размером до 270 атомов. Очевидно, что оптимальные конформации кластеров Морса будут компактно вписываться внутри используемой решётки, пока хватает её размеров. Остается «правильно» выбрать подходящие атомы, образующие конфигурацию минимальной потенциальной энергии. Данная эвристика позволяет задачу параметрической глобальной оптимизации заменить задачей дискретной структурной оптимизации, которая значительно проще.

Основу предлагаемого метода Монте-Карло поиска глобально-оптимальной конформации кластера составляет алгоритм динамического его расширения. Суть которого, заключается в рационально организованном наращивании ядра кластера атомами до заданных размеров. При этом ядро кластера может формироваться на основе статистических данных оптимальных кластеров меньших размеров.

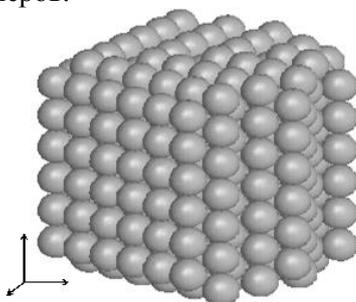


Рисунок 3. «Косая» плотно упакованная решетка.

Алгоритм динамического расширения кластера состоит из следующих шагов:

Шаг 1. Формируется матрица смежности расположения реальных или фантомных атомов в позициях их рационального расположения в узлах поисковой решетки. При этом два атома считаются связанными сильной потенциальной связью, если расстояние между атомами $\rho < 1.08$. Граф связей описывается в системе типологически – описанием связей всех вершин графа. Для этого используется массив структур:

$$Gr = \begin{cases} from & - \text{номер текущей вершины,} \\ to & - \text{список смежных вершин,} \\ deg & - \text{степень вершины} \end{cases}$$

Фактически каждая позиция решетки описывается с указанием все смежных узлов решетки.

Шаг 2. В работе [5] было показано, что кластеры больших размеров наследуют типовые конфигурации внутренних слоев кластера меньших размеров. Наследуемую часть атомов назовем ядром кластера. Ядро кластера условно сворачивается в одну вершину с номером 0. При построении кластера до нужных размеров нам теперь нужны связи ядра кластера (вершины 0) со свободными (не занятыми) вершинами поисковой решетки.

Далее формируются матрица смежности свободных вершин графа $IGr0$ и список вершин ядра кластера $Top0$. Вычисляется потенциал ядра кластера;

Шаг 3. Основная идея предлагаемого алгоритма сводится к динамическому пошаговому расширению кластера, путем присоединения наиболее перспективных узлов решетки (атомов кластера) из числа свободных вершин.

Шаг 3.1. Пусть $Top0$ – список вершин ядра кластера; $Top1$ – список присоединенных вершин на данном шаге алгоритма. Остальные считаются свободными. Тогда множество текущих вершин кластера: $Top1 = Top0 \cup Top1$ и недостроенный кластер

имеет вид $XS = \bigcup_{i \in Top11} X_i$. Его потенциальная энергия вычисляется по формуле

$$v11 = v(XS).$$

Шаг 3.2. Далее для всех свободных вершин решетки подсчитывается их возможный вклад в энергию кластера XS , если таковую $X_j \notin Top11$ присоединить к кластеру. На основании чего можно построить матрицу эффективности свободных вершин

$$TopG = \begin{pmatrix} T_1 & T_2 & \dots & T_{n1} \\ v(X_1) & v(X_2) & \dots & v(X_{n1}) \\ v(XS \cup X_1) & v(XS \cup X_2) & \dots & v(XS \cup X_{n1}) \end{pmatrix},$$

где T_j - номер свободной вершины; $v(X_j)$ - потенциальный вклад свободной вершины в общую энергию кластера; $v(XS \cup X_j)$ - общая энергия кластера с вершиной T_j .

Шаг 3.3. Матрица $TopG$ сортируется по второму показателю и получается новый список свободных вершин, упорядоченный по критерию вклада потенциальной энергии:

$$TopS = \begin{pmatrix} \tilde{T}_1 & \tilde{T}_2 & \dots & \tilde{T}_{n1} \\ v(\tilde{X}_1) & v(\tilde{X}_2) & \dots & v(\tilde{X}_{n1}) \\ v(XS \cup \tilde{X}_1) & v(XS \cup \tilde{X}_2) & \dots & v(XS \cup \tilde{X}_{n1}) \end{pmatrix}$$

Матрица $TopS$ усекается до первых m членов, удовлетворяющих условию:

$$|v(\tilde{X}_j) - v(\tilde{X}_1)| \leq 0.02$$

В результате получается матрица $TopMax$, состоящая из лучших претендентов на присоединение к кластеру

Шаг 3.4. Далее возможны два варианта:

1. Если $\|TopMax\| = 1$, то выбирается первая вершина списка $TopS$.
2. Если $\|TopMax\| = n_t > 1$, тогда равновероятно разыгрывается одна из вершин множества $TopMax$.

Далее «выигршная» вершина Tr_j присоединяется к кластеру, а из множества свободных вершин она исключается.

Шаг 3.5. Этот итерационный процесс продолжается до тех пор, пока кластер не приобретет нужные размеры.

Шаг 4. Процесс динамического пошагового расширения кластера (шаги 1-3) повторяются многократно. Число попыток фиксировано. Исходная ситуация одинакова для всех попыток. Рекордная конформация по значению критерия (2) запоминается.

4. Параллельный алгоритм структурной эволюции кластеров Морса

В нашем случае организация параллельного алгоритма достаточно проста. Последовательный алгоритм динамического расширения кластера, как показали вычислительные эксперименты над кластерами, зарегистрированными в Кембриджской в базе данных, обладает полиномиальной сложностью для класса вероятностных алгоритмов *BPP* (bounded-error probabilistic polynomial). Фактически последовательного алгоритма достаточно для формирования оптимальных кластеров размера до 300 атомов на обычных компьютерах. Однако учитывая стохастический характер алгоритма для обеспечения с высокой доверительной вероятностью точного решения необходимо увеличивать число испытаний. Чего несложно добиться, если растажирировать программу, предложенного алгоритма, на большом количестве процессоров. В результате, лучшая конформация, полученная на одном или нескольких процессорах, выбирается в качестве решения поставленной задачи.

5. Результаты вычислительных экспериментов

Адекватность предлагаемого метода подтверждается результатами, представленными в таблице 1. Как видно из представленной таблицы, с помощью предложенного алгоритма авторами были рассчитаны потенциальные энергии кластеров, размерности которых больше, чем у кластеров, зарегистрированных в Кембриджских базах данных [7]. При этом на несимметричной «косоугольной» решётке сформировались оптимальные кластеры, что соответствует тенденциям, наблюдаемым для кластеров опубликованным в базе данных [7]. Что говорит об эффективности предлагаемого алгоритму.

Таблица 1. Результаты.

Количество атомов в кластере	Плотно	
	упакованная структура	Икосаэдрическая структура
241	-1155,08652	-
242	-1160,10568	-
243	-1166,11859	-1162,41193
244	-1172,1300	-1171,1000
245	-1176,1500	-1175,1000
246	-1181,14472	-
247	-1186,1930	-1184,3192
248	-1192,1703	-1190,7000
249	-1197,1831	-1195,0000
250	-1203,19575	-1196,36683
251	-1207,2207	-
252	-1212,2456	-
...
260	-1254,299	-
266	-1284,4123	-1283,9000
...
318	-	-1560,11539
...
389	-	-1935,87887

6. Заключение

В данной работе представлен алгоритм параллельной стохастической структурной эволюции кластеров Морса, позволяющий находить оптимальные начальные приближения конформации атомных кластеров Морса больших размеров. Предложенный способ позволяет достаточно просто организовать параллельные вычислительные процессы, что существенно расширяет возможности поиска оптимальных конформаций кластеров Морса больших размеров.

В дальнейшем авторы планируют исследовать возможности совершенствования предложенного алгоритма.

7. Литература

- [1] Morse, P. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels / P. Morse // Phys. Rev. – 1929. – Vol. 34. – P. 57-64.
- [2] Немировский, А.С. Сложность задач и эффективность методов оптимизации / А.С. Немировский, Д.Б. Юдин. – М.: Наука, 1979. – 384 с.
- [3] Levinthal, C. How to Fold Graciously / C. Levinthal // Mossbauer Spectroscopy in Biological System: Proceedings of a meeting held at Allerton House. – 1969. – Vol. 67(41). – P.22-24.
- [4] Коварцев, А.Н. Геометрически-обоснованный метод формирования атомных кластеров Морса больших размерностей / А.Н.Коварцев // Компьютерная оптика. – 2017. – Т. 41, № 1. – С. 118-125. DOI: 10.18287/2412-6179-2017-41-1-118-125.

- [5] Коварцев, А.Н. Применение информационных технологий при выяснении правил формирования начальных приближений конформаций атомных кластеров Морса на основе геометрически обоснованного метода / А.Н. Коварцев, Г.П. Климашова // «Перспективные информационные технологии 2018». – Издательство СНЦ РАН), 2018.
- [6] Коварцев, А.Н. К вопросу об эффективности параллельных алгоритмов глобальной оптимизации функций многих переменных / А.Н. Коварцев, Д.А. Попова-Коварцева // Компьютерная оптика. – 2011. – Т. 35, № 2. – С. 256-261.
- [7] Table of the global minima of Morse clusters for N 81 – 240 [Electronic resource]. – Access mode: <http://staff.ustc.edu.cn/~clj/morse/table.html> (07.10.2016).
- [8] Коварцев, А.Н. Правила формирования начальных приближений конформаций атомных кластеров Морса на основе геометрически обоснованного метода / А.Н. Коварцев, Г.П. Климашова [Электронный ресурс]. – Режим доступа: <http://repo.ssau.ru/handle/Informacionnye-tehnologii-i-nanotehnologii/Pravila-formirovaniya-nachalnyh-priblizhenii-konformacii-atomnyh-klasterov-Morsa-na-osnove-geometricheski-obosnovannogo-metoda-69251> (03.12.2018).
- [9] Ren, L. Structural prediction of (Au 20) N (N=240) clusters and their building-up principle / L. Ren, L. Cheng // Computational and Theoretical Chemistry. – 2012. – Vol. 984. – P. 142-147.
- [10] Locatelli, M. Global Optimization: Theory, Algorithms and Applications / M. Locatelli, F. Schoen. – SIAM, 2013. – 433 p.
- [11] Cheng, L. Funnel hopping: searching the cluster potential energy surface over the funnels / L. Cheng, Y. Feng, J. Yang, Ji. Yang // Journal of Chemical Physics. – 2009. – Vol. 130(21). – P. 214112.
- [12] Shao, N. Probing the structural evolution of medium-sized gold clusters: Au n- (n = 27-35) / N. Shao, W. Huang, Y. Gao, L.M. Wang, X. Li, L.S. Wang, X.C. Zeng // Journal of the American Chemical Society. – 2010. – Vol. 132(18). – P.6596-6605.
- [13] Бери, Р.С. Конкуренция кубической гранцентрированной и икосаэдральной кластерных структур / Р.С. Берри, Б.М. Смирнов, А.Ю. Стрижев // ЖЭТФ. – 1997. – Т 112, № 3. – С 1082-1090.
- [14] Hales, T.C. A proof of the Kepler conjecture / T.C. Hales // Annals of Mathematics. – 2005. –Vol. 162. – P. 1065-1185.
- [15] Doye, J.P.K. Structural consequences of the range of the interatomic potential: a menagerie of clusters Unabridged version of this paper: html, postscript, gzipped postscript / J.P.K. Doye, D.J. Wales // J. Chem. Soc., Faraday Trans. – 1997. –Vol. 93. – P. 4233-4244.
- [16] Pullan, W. Unbiased Geometry Optimisation of Morse Atomic Clusters / W. Pullan // WCCI 2010 IEEE World Congress on Computational Intelligence, 2010. – P. 4496-4502.

Благодарности

Работа выполнена при государственной поддержке Министерства образования и науки РФ в рамках реализации мероприятия Программы повышения конкурентоспособности СГАУ среди ведущих мировых научно-образовательных центров на 2013-2020 годы, также гранта РФФИ №16-41-630637.

Parallel algorithm of structural evolution of large Morse clusters

A.N. Kovartsev¹, D.A. Popova-Kovartseva¹, G.P. Klimashova¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. The formation of large Morse atomic clusters is a complex and time-consuming task, especially when the cluster stiffness equals 14. Currently, optimal cluster conformations are known for sizes up to 240 atoms. The article proposes a parallel stochastic algorithm for the structural evolution of Morse clusters, which makes it possible to find the optimal conformations of large atomic clusters. A significant number of problems associated with the complexity of the task are resolved using the proposed algorithm for optimizing the cluster structure. The results of optimization of Morse clusters for $N > 240$ are given.