

УДК 519.6: 519.8

**МЕТОДЫ И РАСПРЕДЕЛЕННАЯ ПРОГРАММНАЯ ИНФРАСТРУКТУРА
ДЛЯ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧИ ПОИСКА
МОЛЕКУЛЯРНЫХ КЛАСТЕРОВ С МИНИМАЛЬНОЙ ЭНЕРГИЕЙ***

© 2010 г.

М.А. ПосыпкинИнститут системного анализа РАН,
Москва

mposypkin@mail.ru

Поступила в редакцию 15.08.2009

Предлагается распределенная программная инфраструктура для реализации стохастических методов в GRID-среде, позволяющая концентрировать находящиеся в глобальной сети разрозненные вычислительные ресурсы в целостную вычислительную среду, обладающую неограниченно наращиваемой производительностью. Приводятся результаты расчетов, подтверждающие эффективность предложенного подхода при решении задачи поиска конформации молекулярного кластера с минимальной энергией. Исследуется влияние различных параметров на скорость и точность расчетов.

Ключевые слова: методы оптимизации, распределенные вычисления, молекулярный кластер.

Введение

Кластерами принято называть группы близких расположенных, тесно связанных друг с другом атомов, молекул, ионов. Изучение кластеров имеет важнейшее значение для понимания процессов конденсации, расчета электронных и динамических характеристик наноматериалов, создания новых источников света и многих других областей [1]. Одной из фундаментальных задач данного направления является определение геометрической структуры, или, как иногда говорят, *конформации* кластера, соответствующей минимальной энергии взаимодействия входящих в него частиц. Такие конформации наиболее часто наблюдаются в веществе при определенных условиях, например во время перехода от молекулярного к конденсированному состоянию.

Широко используется модель, где учитываются только парные взаимодействия частиц, входящих в кластер. Энергия взаимодействия двух частиц задается *парным потенциалом* $v(r)$: функцией одного аргумента, которая определяет зависимость энергии взаимодействия от расстояния между частицами. При этом энергия взаимодействия всех частиц кластера определяется как сумма энергий парных взаимодействий N входящих в него частиц:

$$E(x) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N v(r_{ij}), \quad (1)$$

где r_{ij} – расстояние между частицами i и j , а вектор x содержит $3N$ декартовых координат этих частиц. Так как эта модель обычно применяется для моделирования химических веществ, состоящих из одинаковых атомов, то частицы кластера будем также называть атомами. При исследовании различных веществ и процессов применяются различные потенциалы. В данной работе ограничимся рассмотрением трех пространственных потенциалов: Леннарда – Джонса, Морзе и Дзугутова, характеристики которых приведены в таблице 1. Эти потенциалы применяются для моделирования структуры различных химических веществ. Кластеры Леннарда – Джонса встречаются в инертных газах, Дзугутова – в структурах металлов и стекол. Параметр ρ в потенциале Морзе позволяет моделировать различные вещества. Обычно рассматривают вариации этого параметра в диапазоне 3–14.

Нахождению глобального минимума функции (1) посвящено значительное число работ. Существует база данных [2], в которой перечислены наименьшие найденные значения энергии и соответствующие конформации для различного числа атомов и разных потенциалов взаимодействия. Для нахождения минимума функции (1) применялись различные методы, которые можно условно разделить на две категории. К первой относятся подходы, не исполь-

* Статья рекомендована к печати программным комитетом Международной научной конференции «Параллельные вычислительные технологии 2009» (<http://agora.guru.ru/pavt>).

Таблица 1

Парные потенциалы взаимодействия

Потенциал	Формула
Леннард – Джонс	$v_{LJ}(r) = r^{-12} - 2r^{-6}$
Морзе	$v_M(r) = e^{\rho(1-r)}(e^{\rho(1-r)} - 2)$
Дзугутов	$v_D(r) = A(r^{-m} - B)e^{c/(r-a)}\Theta(a-r) + Be^{d/(r-b)}\Theta(b-r)$, где $\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$ $A = 5.82, B = 1.28, c = 1.10, m = 16, a = 1.87, b = 1.94, d = 0.27$

зующие специфичные для данной задачи свойства целевой функции, т.е. неспециализированные методы оптимизации. Во вторую категорию входят методы, использующие специфику задачи. Начнем с рассмотрения второй категории. Эти методы основаны на общих геометрических закономерностях, наблюдаемых для конформаций с минимальной энергией. Например, было замечено, что подавляющее большинство кластеров в модели Леннарда – Джонса имеют икосаэдральную структуру [3]. Изучение геометрических свойств кластеров Морзе показало, что минимумы могут соответствовать решеткам различных типов [4]. Подобные наблюдения привели к созданию геометрически обоснованных методов минимизации функции (1), которые на данный момент признаны наиболее эффективными для этой задачи.

Несмотря на высокую эффективность геометрически обоснованных методов, они имеют ряд существенных недостатков. Во-первых, данные методы ориентированы на достаточно узкий класс задач, поэтому для каждой рассматриваемой модели взаимодействия требуется разрабатывать свой метод. В частности, методы, основанные на икосаэдральных решетках, показавшие высокую эффективность при отыскании энергетически оптимальных конформаций кластеров Леннарда – Джонса, не позволяют столь же эффективно находить минимумы для потенциала Морзе. Во-вторых, априорные предположения о геометрической природе кластера могут стать причиной ошибок в нахождении минимумов, что было продемонстрировано в работе [5]. Поэтому большое внимание уделяется также решению данной задачи неспециализированными оптимизационными методами. Считается [6], что такие методы обязательно должны применяться для верификации результатов проблемно-ориентированной оптимизации, основанной на геометрических соображениях.

К сожалению, попытки решения рассматриваемой задачи методами, гарантирующими глобальную оптимальность, не увенчались успехом для кластеров из 8 и более атомов [7]. Поэтому

в качестве наиболее перспективных подходов к решению данной задачи рассматриваются различные эвристические алгоритмы. Одним из наиболее известных алгоритмов, хорошо зарекомендовавших себя при решении задачи поиска конформации с минимальной энергией взаимодействия, является локально-стохастический метод Monotonic Sequence Basin-Hopping (MSBH), предложенный в [5]. Суть подхода состоит в комбинации незначительных сдвигов в пространстве допустимых решений и локального поиска. Более подробно этот алгоритм рассматривается далее. MSBH позволил найти все известные энергетически оптимальные конформации кластеров Леннарда – Джонса в диапазоне 3–150 атомов. При этом удалось улучшить некоторые ранее найденные минимумы [8]. Несомненным достоинством данного алгоритма является его простота и применимость для широкого класса задач, основным недостатком – высокая ресурсоемкость. В частности, до ближайшего времени непосредственное применение этого алгоритма не позволяло получить известные минимумы для кластеров Морзе при $\rho = 14$ в диапазоне 40–80 атомов [9].

В последнее время в связи с развитием многопроцессорной техники и Грид-систем появились возможности для качественного повышения эффективности расчетов, в результате которого нередко удается расширить границы применимости различных методов. В данной работе представлена реализация распределенного варианта метода MSBH в рамках программной инфраструктуры BNB-Grid [10], предназначенной для решения задач конечномерной оптимизации в среде параллельных и распределенных вычислений. Для проведения вычислительного эксперимента были задействованы суперкомпьютеры различных организаций (МЦ РАН, ВЦ РАН, ИСП РАН, ТГУ). Результаты расчетов показали, что объединение ресурсов распределенной вычислительной среды позволяет существенно повысить эффективность вычислений и расширить

АЛГОРИТМ MSBH**ВХОДНОЙ ПАРАМЕТР:** точка x_0 **ВЫХОДНОЙ ПАРАМЕТР:** точка x А. Поместить в список точку x_0 .В. Повторять в цикле до выполнения условия останова (число итераций $< N_{\max}$):

1. сгенерировать точку $\Phi(x)$;
2. применив к точке $\Phi(x)$ алгоритм локальной минимизации, получить точку $\tilde{x} = L_f(\Phi(x))$;
3. если $f(\tilde{x}) < f(x)$, то $x = \tilde{x}$.

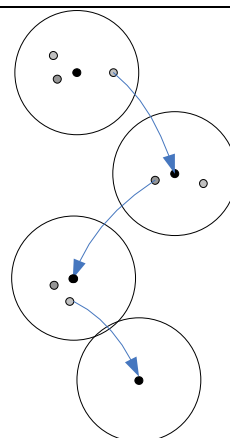
С. Вернуть точку x 

Рис. 1. Локально-стохастический алгоритм MSBH

область применимости данного метода. В частности, с использованием разработанного программного комплекса были получены известные энергетически оптимальные конформации кластеров Морзе для 70, 75, 80, 85, 90, 100 атомов, что не удавалось сделать ранее на однопроцессорных системах.

1. Постановка задачи и метод решения

Задача поиска структуры кластера с минимальной энергией взаимодействия формулируется как задача безусловной оптимизации: требуется найти минимум функции (1) в евклидовом пространстве размерности $n = 3N$, где N – число атомов в кластере. Аналитические выражения для градиента и гессиана целевой функции могут быть легко получены.

Рассматриваемый в данной работе алгоритм MSBH работает по следующей схеме (рис. 1). В качестве начальной выбирается точка x_0 локального минимума функции. На каждом шаге алгоритма текущая точка локального минимума x подвергается возмущению Φ , к результату которого применяется локальный алгоритм L_f : $\tilde{x} = L_f(\Phi(x))$. Если $f(\tilde{x}) < f(x)$, то точка x заменяется на \tilde{x} . Далее процесс повторяется. Алгоритм останавливается, когда после заданного числа N_{\max} итераций не удастся улучшить найденный локальный минимум. Оператор возмущения изменяет координаты точки в пределах некоторой окрестности: $\Phi(x) = x + \xi$, $\xi_i \in [-r, r], i = 1, \dots, n$. Величины ξ_i генерировались как равномерно распределенные случайные числа в диапазоне $[-r, r]$. Возможно применение других законов распределения. Основными параметрами алгоритма MSBH являются радиус окрестности про-

смотра r и количество N_{\max} итераций при генерации случайных точек. Радиус окрестности просмотра r должен быть достаточно небольшим, чтобы допускать эффективный перебор с помощью оператора возмущения. Эти параметры подбираются экспериментально.

На практике алгоритм MSBH чаще всего применяется вместе с каким-либо алгоритмом, генерирующим начальные приближения, например методом Монте-Карло. Мы применяли вариант, предложенный в работе [5], в котором некоторое количество начальных приближений генерируется случайно в кубе $[-1, 1]^n$. Далее сгенерированные точки используются в качестве начальных для MSBH. Такой алгоритм идеально подходит для реализации в среде параллельных и распределенных вычислений, так как различные начальные приближения могут обрабатываться независимым образом.

2. Реализация алгоритма MSBH в системе BNB-Grid

2.1. Управление вычислительным пространством в системе BNB-Grid

В настоящее время в распоряжении отдельных исследователей и научных коллективов, как правило, имеются ресурсы следующих типов: рабочие станции, небольшие многопроцессорные комплексы, доступные в монопольном режиме, суперкомпьютеры коллективного доступа. Программный комплекс BNB-Grid [10] предназначен для решения задач оптимизации на распределенных системах, состоящих из узлов перечисленного типа. BNB-Grid запускает и организует взаимодействие параллельных приложений, решающих задачу с помощью библиотеки BNB-Solver [11] на различных вычислительных узлах. В результате формируется

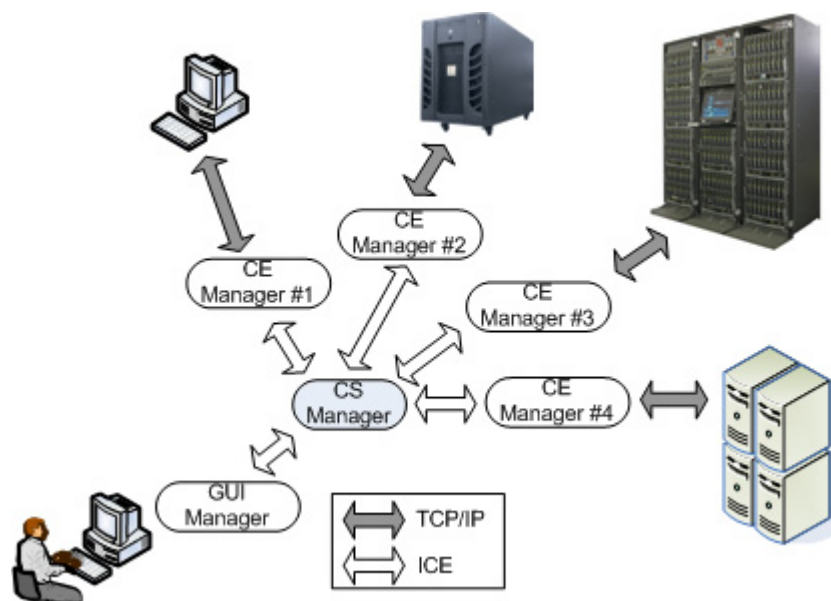


Рис. 2. Организация вычислений в BNB-Grid

иерархическая распределенная система: на верхнем уровне части работы распределяются между параллельными приложениями, а далее они распределяются по процессорам средствами библиотеки BNB-Solver.

Ядро системы реализовано на языке программирования Java с использованием промежуточного программного обеспечения Internet Communication Engine (ICE) [12] – аналога CORBA. Каждый вычислительный узел представлен в системе распределенным экземпляром объекта типа CE-Manager – Computing Element Manager (рис. 2). Этот объект предоставляет интерфейс для запуска и остановки приложений на вычислительном узле. Координация работы объектов типа CE-Manager осуществляется с помощью другого распределенного объекта типа CS-Manager – Computing Space Manager. Графический интерфейс пользователя также реализован как ICE-объект, обозначенный на рисунке GUI Manager.

ICE-объекты размещаются либо на одном, либо на нескольких компьютерах в пределах локальной сети. Доступ к удаленному вычислительному узлу осуществляется по следующей схеме (рис. 3). Объект CE-Manager устанавливает SSH-соединение с внешним управляющим модулем узла, запускает на нем процесс CE-Server и устанавливает с ним TCP/IP-соединение. Если удаленный узел защищен сетевыми экранами, то применяется механизм туннелирования сетевого соединения через SSH-канал. Процесс CE-Server информирует CE-Manager о состоянии вычислительного узла. При обрыве

соединения или перезагрузке узла CE-Server перезапускается. При получении запроса на запуск параллельного приложения CE-Manager создает объект APP-Manager, CE-Server создает процесс APP-Proxy, который запускает параллельное приложение BNB-Solver на узле. При этом CE-Manager взаимодействует с APP-Manager средствами ICE, а APP-Manager, в свою очередь, устанавливает TCP/IP-соединение с BNB-Solver через APP-Proxy. Процесс APP-Proxy необходим, так как на суперкомпьютерах общего доступа непосредственный доступ из внешней сети к вычислительным модулям, как правило, невозможен.

На одном вычислительном узле может быть запущено несколько экземпляров приложения BNB-Solver. Такой подход часто оказывается целесообразным для суперкомпьютеров коллективного доступа, работающих под управлением систем пакетной обработки. На таких системах приложение, запросившее достаточно большое число процессоров, может быть надолго помещено в очередь в ожидании соответствующего «окна» в расписании заданий. В то же время приложение, запросившее существенно меньшее число процессоров, может быть запущено ранее. Это объясняется тем, что системе пакетной обработки проще эффективно размещать небольшие задания. Поэтому в наших экспериментах на суперкомпьютерах коллективного доступа запускалось от пяти до двадцати приложений.

Выделение графического интерфейса пользователя в отдельный ICE-объект преследует

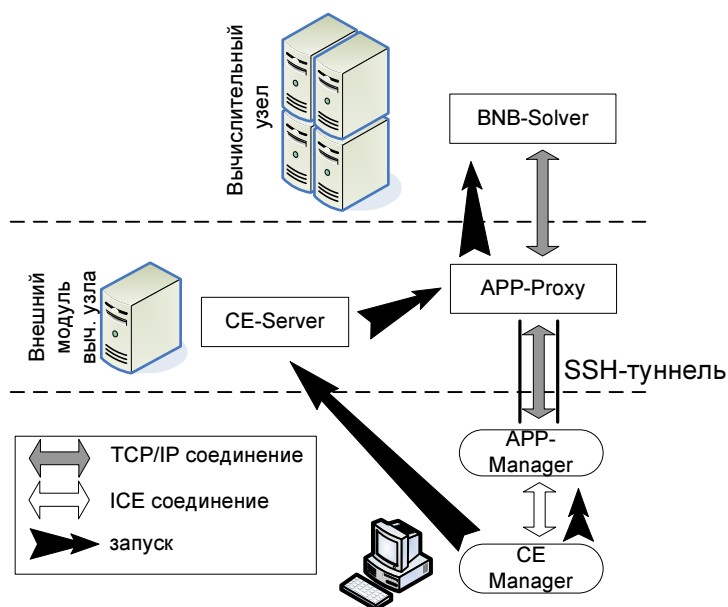


Рис. 3. Запуск приложений на удаленном вычислительном узле и установление соединения

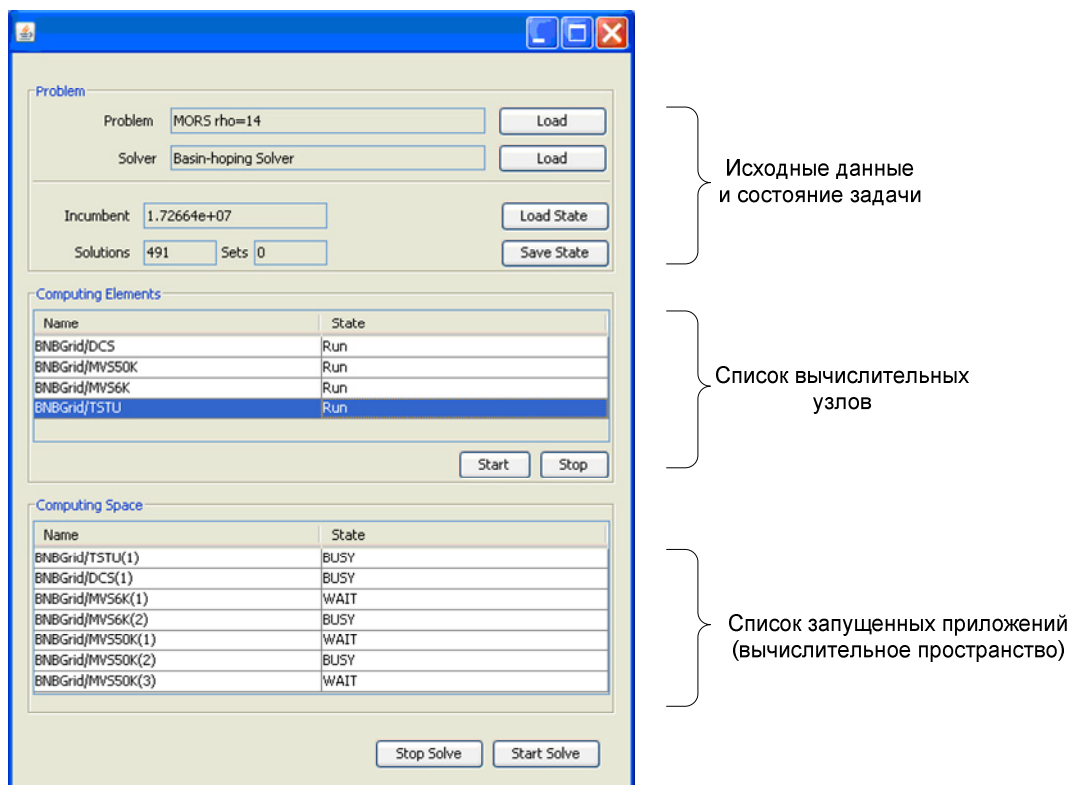


Рис. 4. Графический интерфейс пользователя системы BNB-Grid

две цели. Во-первых, для работы в системе BNB-Grid пользователю достаточно установить на своем персональном компьютере только этот компонент. Во-вторых, при таком подходе графический интерфейс становится легко заменяемым компонентом, который не является неотъем-

лемой частью системы. Общий и специализированный для рассматриваемой задачи варианты графического интерфейса пользователя представлены на рис. 4 и рис. 5 соответственно. Общий графический интерфейс предоставляет возможности для загрузки исходных данных

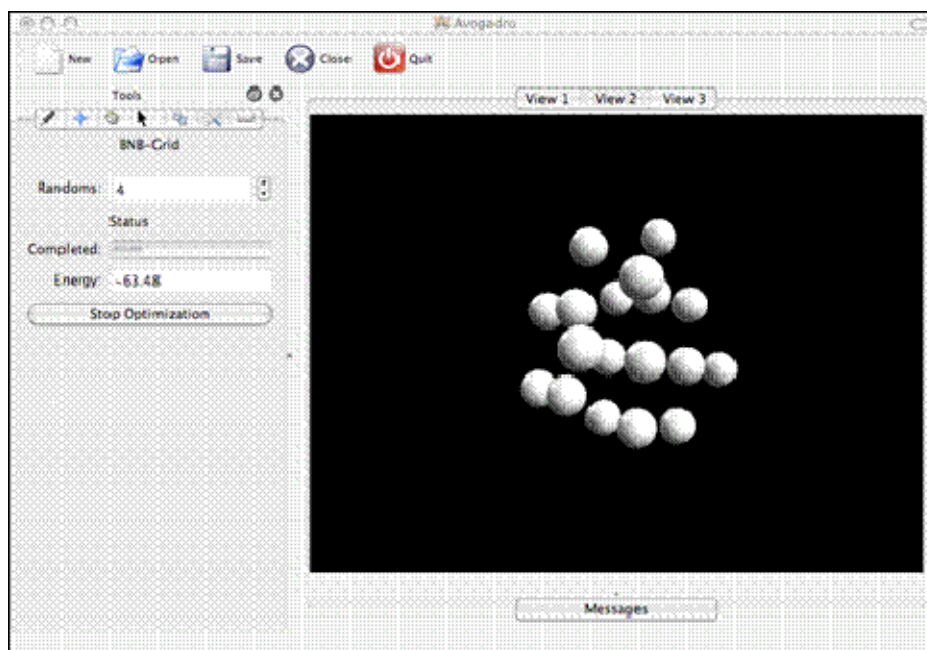


Рис. 5. Версия графического интерфейса пользователя системы BNB-Grid для решения задачи поиска энергетически минимальных конформаций молекулярных кластеров

задачи, управления вычислительным пространством, загрузки и сохранения состояния решения задачи. Специализированный графический интерфейс, представленный на рис. 5, позволяет просматривать и модифицировать исходные и полученные в результате расчетов кластеры. Такие возможности необходимы для исследования геометрических свойств энергетически оптимальных кластеров. Вместе с системой хранения конформаций этот интерфейс образует развитую оболочку для исследования данной задачи. Более подробно эта система рассмотрена в работе [13].

2.2. Иерархическое управление распределением вычислительной нагрузки

Управляющие модули BNB-Grid и запущенных экземпляров библиотеки BNB-Solver образуют иерархическую систему управления распределением вычислительной нагрузки (рис. 6). На верхнем уровне части работы распределяются ICE-объектом CS-Manager между экземплярами приложений BNB-Solver, а на нижнем – управляющим процессом (Master Process) приложения BNB-Solver между его рабочими процессами (Worker Process). Понятие части работы зависит от применяемого алгоритма: для методов ветвей и границ частями будут подмножества пространства допустимых решений, для эвристических алгоритмов – допустимые решения.

Рассмотрим управление вычислительной нагрузкой в системе BNB-Grid на примере

рассматриваемого алгоритма MSBH. На первом шаге на CS-Manager загружается некоторое число начальных точек. ICE-объекты App-Manager запрашивают точки у CS-Manager и передают их на обработку связанным с ними экземплярам приложения BNB-Solver. Каждому из экземпляров передается число точек, не превосходящее числа рабочих процессов в приложении. Внутри параллельного приложения управляющий процесс библиотеки BNB-Solver распределяет точки по рабочим процессам. Как только все переданные параллельному приложению точки обработаны, приложение запрашивает новое задание с CS-Manager.

Из-за принудительного завершения приложения системой пакетной обработки, сетевого или аппаратного сбоя экземпляр BNB-Solver может не завершить обработку выделенной ему части работы. Для предотвращения потери данных в такой ситуации CS-Manager сохраняет резервные копии передаваемых заданий для каждого из экземпляров запущенных приложений. В случае аварийного завершения любого из них соответствующая резервная копия возвращается в общий пул.

Для методов ветвей и границ балансировка усложняется, так как новые части работы генерируются в процессе вычислений. Методы балансировки нагрузки для методов ветвей и границ в программных комплексах BNB-Grid и BNB-Solver подробно рассмотрены в работах [10, 14].

маломощные узлы 4 и 5 практически отпала.

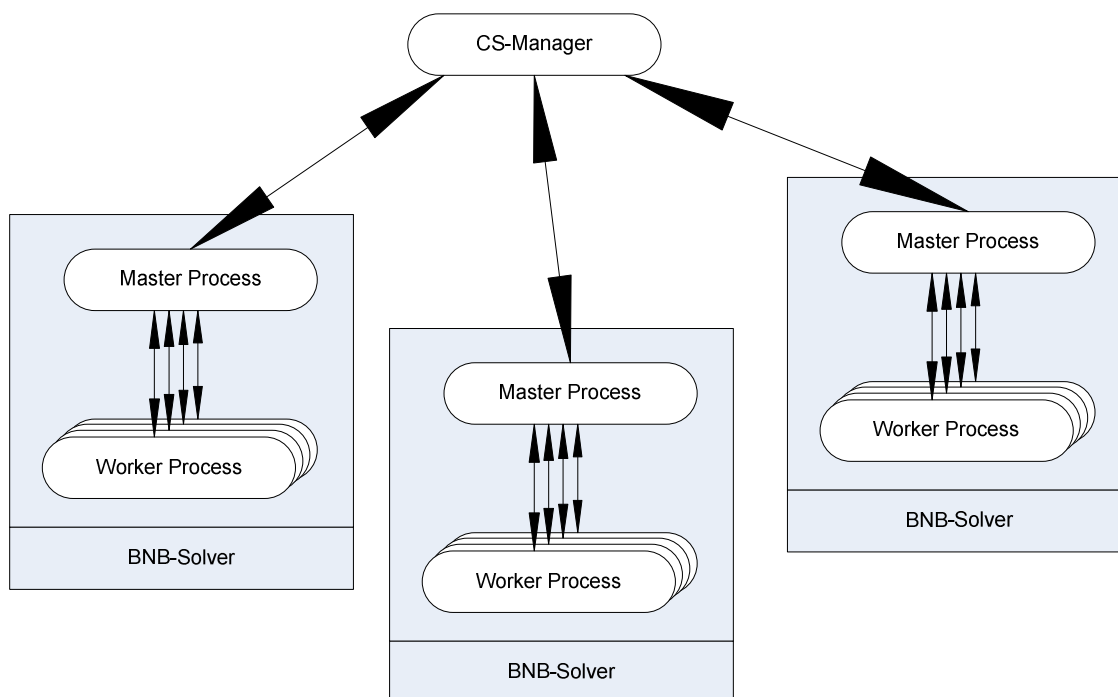


Рис. 6. Иерархическая структура BNB-Grid

Таблица 2

Характеристики узлов распределенной системы

№	Название	Архитектура процессора	Число вычислительных ядер	Местоположение	Наличие системы пакетной обработки
1	MVS100K	Intel Xeon 5365, 3 GHz	7856	Межведомственный суперкомпьютерный центр (Москва)	+
2	MVS6K	Intel Itanium II, 2.2 GHz	256	Вычислительный центр РАН (Москва)	+
3	OMEGA	Intel Xeon 5355, 2.66 GHz	88	Институт системного программирования РАН (Москва)	+
4	TSTU	Pentium IV, 3.2 GHz	8	Тамбовский государственный университет	–
5	DCS	Pentium IV, 3.2 GHz	1	Институт системного анализа РАН (Москва)	–

3. Результаты вычислительных экспериментов

Вычислительные эксперименты проводились для потенциалов Леннарда – Джонса, Дзугутова и Морзе (таблица 1). В вычислениях были задействованы вычислительные узлы, приведенные в таблице 2.

На начальной стадии экспериментов узлы 4, 5 были востребованы, так как из-за загруженности суперкомпьютеров коллективного доступа 1, 3 не удавалось оперативно получить большое количество вычислительных ядер. В дальнейшем, в связи с появлением доступа на кластер 3 и увеличением числа доступных ядер на MVS 100K, необходимость использовать в расчетах

Приводимые ниже результаты были получены на распределенной системе, состоящей из узлов 1–3. При этом общее число задействованных процессоров достигало 512. Среднее число колебалось в интервале приблизительно 200–512 и зависело от текущей загрузки используемых кластеров.

Алгоритм MSBH управляется следующими параметрами: количеством начальных приближений, максимальным числом сдвигов N_{\max} в алгоритме MSBH до момента улучшения и максимальным радиусом возмущения r . Сначала для каждого из рассматриваемых потенциалов проводилась небольшая серия предварительных экспериментов для установления наилучшего значения параметра r . Как отмечается в работе

Таблица 3

Наилучшие значения r для различных потенциалов

Потенциал	Значение r
Морзе $\rho = 6$	0.4
Морзе $\rho = 14$	0.35
Леннард – Джонс	0.4
Дзугутов	0.4

Таблица 4

Результаты вычислительного эксперимента

№	Потенциал	Число атомов	Число начальных приближений	N_{\max}	Общее время расчетов (мин)	Количество «попаданий»	Найденный минимум	Наилучший известный минимум
1	Леннард – Джонс	98	512	8192	43	8	-543.665361	-543.665361
2	Морзе $\rho = 6$	85	512	8192	63	3	-405.246158	-405.246158
3	Морзе $\rho = 6$	90	512	8192	91	74	-433.355380	-433.355380
4	Морзе $\rho = 6$	100	512	8192	96	98	-488.675685	-488.675685
5	Морзе $\rho = 14$	70	1024	8192	174	9	-292.462856	-292.462856
6	Морзе $\rho = 14$	75	1024	8192	205	2	-318.407330	-318.407330
7	Морзе $\rho = 14$	80	1024	8192	244	3	-340.811371	-340.811371
8	Морзе $\rho = 14$	85	1024	8192	188	5	-363.891261	-363.893075
9	Морзе $\rho = 14$	85	512	32768	372	2	-363.893075	-363.893075
10	Морзе $\rho = 14$	90	1024	8192	266	2	-388.401652	-388.401652
11	Морзе $\rho = 14$	100	1024	8192	232	8	-439.070547	-439.070547
12	Дзугутов	50	1024	8192	175	2	-104.366189	-104.366189
13	Дзугутов	100	1024	8192	175	1	-218.678229	-219.523265
14	Дзугутов	100	1024	32768	371	1	-218.744395	-219.523265

[5], это значение практически не зависит от числа атомов, а определяется только видом потенциала. В таблице 3 приведены наилучшие значения параметра r для различных потенциалов.

В первом вычислительном эксперименте отыскивался минимум энергии кластера Леннарда – Джонса из 98 атомов. Известно, что данная конформация является одной из наиболее сложных для отыскания [8, 9]. Обнаружение нового глобального минимума для 98 атомов стало заметным событием в свое время, послужившим основанием для публикации [8]. Результат вычислительного эксперимента приведен в строке 1 таблицы 4. N_{\max} полагалось равным 8192. Для обработки 512 начальных при-

ближений потребовалось 43 минуты, на протяжении которых наименьшее известное на настоящее время значение целевой функции было найдено 8 раз. На основании этого можно сделать вывод о том, что применяемый метод позволяет уверенно находить оптимальную конфигурацию.

Вторая серия экспериментов проводилась для потенциала Морзе при $\rho = 6$. Было выбрано также 512 начальных приближений, N_{\max} полагалось равным 8192. Результаты расчетов для 85, 90 и 100 атомов приведены в строках 2–4 таблицы 4. Удалось улучшить минимум для 85 атомов, ранее найденный с применением параллельных вычислений [15]. Интересно отметить очень большое число попаданий в точку минимума для $N = 90$,

100 и то, что нахождение минимума для $N = 85$ атомов оказалось более сложной задачей по сравнению с $N = 90$ и $N = 100$.

Третья серия экспериментов проводилась для потенциала Морзе при $\rho = 14$, который рассматривается как наиболее сложный с точки зрения отыскания минимума по сравнению с другими значениями параметра. Результаты экспериментов для 70, 75, 80, 85, 90, 100 атомов приведены в строках 5–11 таблицы 4. Для 70, 75, 80, 90 и 100 атомов найденные минимумы совпали с найденными с помощью геометрически обоснованного алгоритма Dynamic Lattice Search [4]. Удалось улучшить минимумы для 85 и 90 атомов, ранее найденные с применением параллельной реализации MSBH [15].

Следующая серия экспериментов проводилась для потенциала Дзугутова (таблица 4, строки 12–14). Известные значения минимумов были получены с использованием алгоритма [16], предложенного Дойе, который отличается от MSBH тем, что модификация решений проводится на основе геометрических соображений (сдвигаются атомы), а также для получения конфигурации из $N+1$ атома используется конфигурация из N атомов, полученная при предыдущих расчетах. Были проведены расчеты для 50 и 100 атомов. Для 50 атомов те же параметры, что и для кластеров Морзе, дали хорошие результаты – минимум был найден. Для 100 атомов при первоначальном расчете с $N_{\max} = 8192$ было получено значение, отличающееся от известного минимума. При пересчете с $N_{\max} = 32768$ было получено несколько лучшее значение, но все равно превосходящее известный минимум. При этом время расчетов составило приблизительно 12 часов. Полученные результаты позволяют сделать вывод о том, что кластеры Дзугутова являются более трудными для обнаружения по сравнению с кластерами Морзе и Леннарда – Джонса и для успешного отыскания минимумов в диапазоне 50–100 атомов требуется консолидация значительных вычислительных ресурсов.

Заключение

В работе рассмотрена распределенная программная инфраструктура для решения задач конечномерной оптимизации, которая позволяет объединять разрозненные разнородные суперкомпьютерные ресурсы в единое вычислительное пространство. В рамках этой инфраструктуры был реализован известный стохастический алгоритм оптимизации MSBH. Результаты экспериментов показали, что суммарная вы-

числительная мощность нескольких суперкомпьютеров позволяет существенно раздвинуть границы применимости этого метода и получать оптимальные либо близкие к оптимальным конформации за приемлемое время для различных потенциалов межатомного взаимодействия. Были получены минимальные значения энергии взаимодействия и соответствующие им пространственные конфигурации для размерностей $N > 70$ для потенциала Морзе, что ранее не удавалось сделать с помощью данного метода.

Полученный результат показывает, что консолидация большой вычислительной мощности нескольких суперкомпьютеров позволяет успешно решать задачи оптимизации структуры молекулярных кластеров для размерностей и потенциалов, которые ранее удавалось решить только с применением геометрически обоснованных подходов. Этот факт представляется важным, так как применение неспециализированных методов требует меньше усилий по созданию новых алгоритмов, основанных на априорных предположениях о геометрической структуре химического соединения. Кроме того, неспециализированные методы могут рассматриваться как проверочные для проблемно-ориентированных подходов [6].

В дальнейшем планируется подключить к расчетам вычислительные ресурсы Московского государственного университета [16], Южно-Уральского государственного университета [17], компьютеры Грид-сегментов Европейского Грид-сообщества EGEE [18]. Предполагаемое увеличение постоянно доступных вычислительных ядер до 2000–4000 позволит выйти на новый уровень точности вычислений и верифицировать существенно больший спектр известных минимумов. При этом, возможно, будут улучшены многие ранее найденные значения.

Другим путем увеличения производительности алгоритма является улучшение алгоритма поиска минимума. Представляется интересным использовать идеи генетического подхода, примененного в работе [19] для задач докинга лигандов. Принципы комбинирования свойств различных решений и предотвращения обхода одинаковых областей множества допустимых решений, предложенные в [19], могут существенно повысить эффективность стохастического поиска.

Автор выражает глубокую благодарность академику РАН Ю.Г. Евтушенко за постановку данной задачи и плодотворные обсуждения результатов работы. Автор благодарен профессору А.П. Афанасьеву за ценные советы по реали-

зации распределенных алгоритмов и внимание к работе, а также зам. директора МСЦ РАН Б.М. Шабанову, зам. директора ВЦ РАН Г.М. Михайлову, сотруднику ИСП РАН О.И. Самоварову за предоставление суперкомпьютерных вычислительных ресурсов и оперативное решение технических проблем в процессе расчетов.

Работа выполнена при поддержке программы П-15 президиума РАН.

Список литературы

1. Елецкий А.В. «Экзотические» объекты атомной физики // Соросовский образовательный журнал. 1995. № 4. С. 86–95.
2. The Cambridge Cluster Database. URL: <http://www-wales.ch.cam.ac.uk/CCD.html>.
3. Northby J.A. Structure and binding of Lennard-Jones clusters: $13 \leq N \leq 147$ // Journal of Chemical Physics. 1987. Vol. 87. P. 6166–6178.
4. Cheng L., Yang J. Global Minimum Structures of Morse Clusters as a Function of the Range of the Potential: $81 \leq N \leq 160$ // Journal of Physical Chemistry. 2007. 111. P. 5287–5293.
5. Leary R.H. Global Optima of Lennard–Jones Clusters // Journal of Global Optimization. 1997. 11. 1. P. 35–53.
6. Doye J.P., Wales D.J., Simdyankin S.I. Global optimization and the energy landscapes of Dzugutov clusters // Faraday Discuss. 2001. Vol. 118. P. 159–170.
7. Maranas C., Floudas C. A Global Optimization Approach for Lennard–Jones Microclusters // Journal of Chemical Physics. 1992. Vol. 97. P. 7667–7677.
8. Leary R.H. and Doye J.P.K., Tetrahedral global minimum for the 98-atom Lennard–Jones cluster // Phys. Rev. E. 1999. Vol. 60. № 6. P. 6320–6322.
9. Grosso A., Locatelli M., Schoen F. A population-based approach for hard global optimization problems based on dissimilarity measures // Mathematical Programming. 2007. Vol. 110. P. 373–404.
10. Афанасьев А.П., Посыпкин М.А., Сигал И.Х. Проект BNB-Grid: решение задач глобальной оптимизации в распределенной среде // Труды Второй Международной конференции «Системный анализ и информационные технологии» САИТ-2007. Том 2. С. 177–181.
11. Посыпкин М.А. Архитектура и программная организация библиотеки для решения задач оптимизации методом ветвей и границ на многопроцессорных вычислительных комплексах // Проблемы вычислений в распределенной среде: распределенные приложения, коммуникационные системы, математические модели и оптимизация // Труды ИСА РАН. М.: КомКнига, 2006. С. 18–25.
12. Henning M. A New Approach to Object-Oriented Middleware // IEEE Internet Computing. 2004.
13. Смирнов С.А. Распределенный программный комплекс для моделирования структуры молекулярных соединений // Труды Третьей Международной научно-практической конференции «Современные информационные технологии и ИТ-образование» Москва, 6–9 декабря 2008 г. М.: МАКС ПРЕСС, 2008. С. 521–528.
14. Посыпкин М.А., Сигал И.Х. Комбинированный параллельный алгоритм решения задачи о ранце // Известия РАН. Теория и системы управления. 2008. № 4. С. 50–58.
15. Посыпкин М.А. Параллельный эвристический алгоритм глобальной оптимизации // Труды ИСА РАН. 2008. Т. 32. С. 166–179.
16. Суперкомпьютерный комплекс МГУ. URL: <http://parallel.ru/cluster/>.
17. Суперкомпьютерный центр Южно-Уральского государственного университета URL: <http://supercomputer.susu.ru/>.
18. Enabling Grids for E-science (EGEE) URL: <http://www.eu-egee.org/>.
19. Романов А.Н., Кондакова О.А., Григорьев Ф.В. и др. Компьютерный дизайн лекарственных средств: программа докинга SOL // Вычислительные методы и программирование. 2008. Т. 9. С. 213–233.

SEARCHING FOR MINIMUM ENERGY MOLECULAR CLUSTERS: METHODS AND DISTRIBUTED SOFTWARE INFRASTRUCTURE FOR NUMERICAL SOLUTION OF THE PROBLEM

M.A. Posypkin

The problem is considered of searching for a geometrical structure of a cluster of identical atoms, where the interaction between the atoms is described by pair potentials. Stochastic optimization methods applied to calculations of this problem are highly time-consuming which hampers their use in the analysis of structures with a large number of atoms. A distributed software infrastructure which realizes stochastic methods in a Grid environment is proposed. This infrastructure enables the concentration of globally distributed computational resources in an integrated computing environment with unlimited scalability. Computing experiment results confirming the efficiency of the proposed approach of searching for a minimum energy molecular cluster conformation are presented. The influence of different parameters on calculation speed and accuracy is investigated.

Keywords: optimization methods, distributed computing, molecular cluster.