

Оганов А.Р.^{1,2,3,4}, Посыпкин М.П.⁵, Ройзен В.В.², Самцевич А.И.¹, Сухомлин В.А.⁶, Храпов Н.П.⁷

¹ Сколковский институт науки и технологий, Инновационный центр Сколково, г. Москва, Россия

² Лаборатория компьютерного дизайна материалов, Московский физико-технический институт, г. Москва, Россия

³ Геологический факультет, Университет Стони Брук, штат Нью-Йорк, США

⁴ Школа материаловедения и инженерии, Северо-западный университет, г. Сиань, Шанкси, КНР

⁵ Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, г. Москва, Россия

⁶ Московский государственный университет им. М.В. Ломоносова, г. Москва, Россия

⁷ Институт Проблем Передачи Информации РАН, г. Москва, Россия

СИСТЕМА ДОБРОВОЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ КОМПЬЮТЕРНОГО ДИЗАЙНА МАТЕРИАЛОВ

АННОТАЦИЯ

Кристаллическая структура вещества является наиболее важным носителем информации о веществе – зная структуру, можно предсказать огромное число его свойств. К настоящему моменту стало возможно предсказывать стабильные кристаллические структуры неорганических соединений при заданном химическом составе. Одним из мощных и популярных методов для решения этой задачи является эволюционный алгоритм USPEX. В настоящей работе мы описываем его адаптацию для системы добровольных вычислений на платформе BOINC.

КЛЮЧЕВЫЕ СЛОВА

Дизайн материалов; распределённые вычисления; программирование; грид-системы; оптимизация; успех.

Oganov A.^{1,2,3,4}, Posypkin M.⁵, Rozen V.², Santsevich A.¹, Sukhomlin V.⁶, Khrapov N.⁷

¹ Skolkovo Institute of Science and Technology, Moscow, Russia

² Computational Materials Discovery Lab MIPT, Moscow, Russia

³ Stony Brook University, Computational materials discovery laboratory (Oganov's Lab), USA

⁴ Northwest Normal University, Xian, China

⁵ Dorodnicyn Computing Centre, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

⁶ Institute for Information Transmission Problems of RAS, Moscow, Russia

VOLUNTEER COMPUTING FOR COMPUTATIONAL MATERIALS DESIGN

ABSTRACT

The problem of crystal structure prediction is very old and does, in fact, constitute the central problem of theoretical crystal chemistry. In this paper, we talk about the USPEX evolutionary algorithm. Here we present the distributed computing implementation of the USPEX based on a popular BOINC volunteer computing platform. In the talk we present experimental results and discuss project performance.

KEYWORDS

Distributed computing; evolutionary algorithm.

Введение

Компьютерный дизайн материалов – это новая, революционная область в науке. Вместо того, чтобы полагаться на экспериментальные методы проб и ошибок или же на случай – а большинство материалов до настоящего момента были найдены именно этими способами – теперь можно открывать новые материалы на компьютере, задавая направление для экспериментальных работ [1].

Эволюционный алгоритм USPEX [2] является одним из самых эффективных методов для

решения этой задачи, что подтверждается «слепыми» тестами по предсказанию органических и неорганических кристаллов и рядом исследований [1, 3]. Кроме того, USPEX успешно применялся для предсказания структур наночастиц [4] и полимеров [5].

Тем не менее, эффективное решение задачи компьютерного дизайна материалов требует мощных вычислительных ресурсов. Перспективным способом для решения этой проблемы являются добровольные вычисления, которые уже успешно применяются для решения близкой задачи предсказания структуры белков в рамках проекта Rosetta@HOME [7].

В настоящей работе мы описываем реализованную для алгоритма USPEX систему распределённых вычислений, основанную на популярной платформе BOINC [6]. Принцип работы системы можно кратко описать следующим образом: алгоритм USPEX запускается на отдельном сервере и в процессе работы создаёт задания, которые затем загружаются на сервер проекта добровольных вычислений. Затем эти задания отправляются на машины добровольцев с помощью инфраструктуры BOINC. Задания представляют из себя входные файлы для расчётов с помощью программы GULP [8]. После выполнения задания оно собирается сервером проекта BOINC. После чего полученные результаты пересылаются на USPEX-сервер. Далее в работе мы приводим подробное описание принципов работы системы и полученные с её помощью результаты.

Компьютерный дизайн материалов

Кристаллическая структура вещества является наиболее важным носителем информации о нём – зная структуру, можно предсказать широкий набор его свойств. Это позволяет создавать материалы с нужными характеристиками путём подбора нужной конфигурации атомов. С точки зрения математики это является задачей глобальной многопараметрической оптимизации, численное решение которой требует разработки надёжного и эффективного метода.

Рассмотрим в качестве примера задачу предсказания кристаллической структуры. Из простых соображений комбинаторики [2] легко получить оценку числа потенциально возможных кристаллических структур:

$$C = \binom{V/\beta^3}{N} \prod_i \binom{N}{n_i},$$

где N - это полное число атомов в элементарной ячейке объёмом V , β - разумный параметр дискретизации (например, 1 ангстрем) и n_i число атомов типа i в элементарной ячейке.

Даже для небольших систем ($N \approx 10-20$), C принимает астрономическое значение (приблизительно 10^N).

Используя особенности изучаемых систем (кристаллов, полимеров, наночастиц), реально понизить размерность решаемой задачи на несколько порядков. Тем не менее, даже в этом случае нужны большие масштабные вычислительные ресурсы. Поэтому актуальна проблема создания вычислительных систем нового поколения, адаптированных для компьютерного дизайна материалов.

Эволюционный алгоритм

USPEX - это эволюционный алгоритм [2], первоначально разработанный для предсказания кристаллических структур. В настоящее время он также адаптирован для предсказания нанокластеров [4] и полимеров [5]. Его эффективность обусловлена сочетанием глобальной и локальной оптимизации, что позволяет эффективно исследовать пространство поиска и избегать "залипания" в локальных минимумах.

Общая схема работы алгоритма представлена на рис. 1. Ниже приводится краткое описание его работы.

Весь расчёт разбит на поколения, состоящие из структур. Первое поколение генерируется случайным образом с помощью операторов симметрии, характерных для изучаемой системы (например, пространственных групп симметрии для кристаллов). Затем все структуры релаксируются в несколько этапов для приведения их к локальным минимумам. Для этого используются сторонние программы для квантово-химических расчётов или для классических расчётов методами потенциалов силового поля. В данной работе применялась программа General Utility Lattice Program (GULP) [8]. Это свободное ПО, позволяющее проводить расчёты с помощью различных потенциалов силового поля.

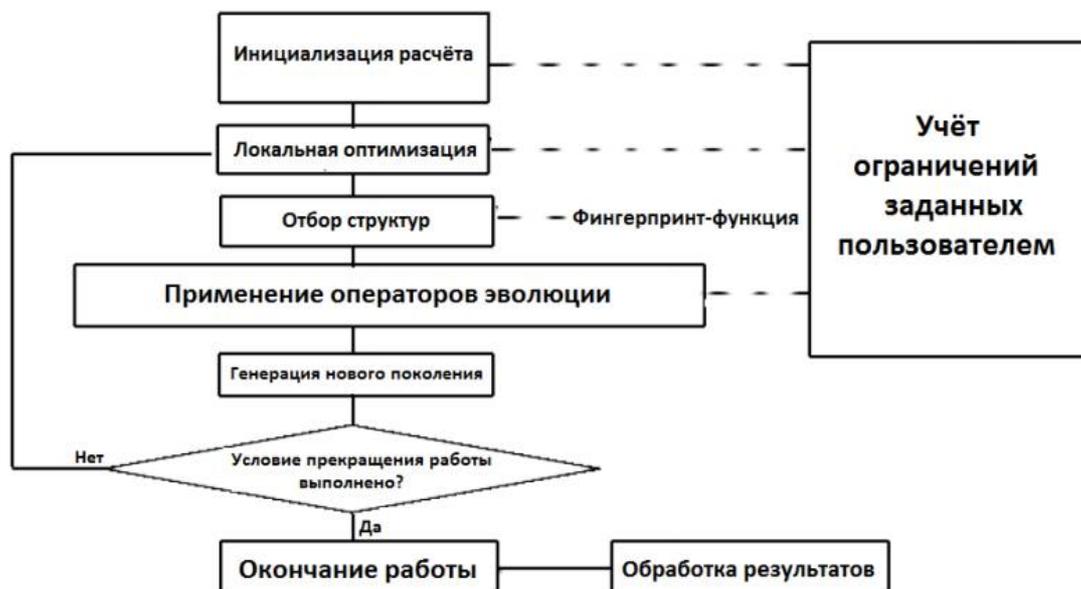


Рис.1. Схема работы алгоритма USPEX

После релаксации заданный процент лучших структур в поколении отбирается для создания следующего поколения с помощью эволюционных операторов. В коде USPEX реализовано несколько функций соответствия, использование конкретной из них определяется при инициализации расчёта. Детали работы эволюционных операторов описаны ранее [2, 9]. Для сохранения разнообразия в популяции, критично важного для эффективной работы эволюционного алгоритма, также добавляется некоторое количество случайных структур. После чего все итерации повторяются ещё раз пока не выполняются критерий остановки расчёта (по умолчанию это неизменность лучшей структуры в течение фиксированного числа поколений).

Наиболее затратным с точки зрения вычислительных ресурсов является этап релаксации структур. Использование системы добровольных вычислений позволяет ускорить расчёты в несколько раз. Далее мы подробно опишем принципы работы разработанной системы.

Добровольные вычисления

Если сравнивать принцип функционирования ГСПК с сервисными грид-системами и кластерами, то можно выделить основные отличия:

1. Обмен данными осуществляется только между вычислительным узлом и сервером проекта, прямого обмена данных между вычислительными узлами нет.
2. Большое время на инициализацию задания, включающее в себя подготовку, копирование данных на вычислительный узел и получение результатов обратно;
3. Нестабильность работы узлов – любой узел может в любой момент на неопределённое время прекратить выполнять расчёты. Это приводит к очень существенной разнице во времени обработки расчетных блоков.

Для исследования влияния перечисленных факторов на производительность распределенной реализации метода ветвей и границ и выработки оптимальной стратегии такой реализации была разработана программный прототип для платформы BOINC. BOINC (Berkeley Open Infrastructure for Network Computing) представляет собой платформу с открытым кодом для организации проектов добровольных вычислений. Разработка системы ведется в U.C. Berkeley Spaces Sciences Laboratory (США) исследовательской группой, которая также разрабатывала проект SETI@home. Работа над BOINC была начата в 2002 году с целью создания универсальной программной платформы для проектов подобного рода, которая бы упростила процесс развертывания необходимой инфраструктуры и разработки приложений. Первый проект добровольных вычислений на основе BOINC был запущен в 2004 году. В настоящее время насчитывается более 80 публичных проектов на основе BOINC, делая платформу стандартом де-факто в данной области.

Все программное обеспечение BOINC можно разделить на две компоненты: клиентскую и серверную части программного обеспечения. Клиентская часть устанавливается на вычислительном узле. В её задачи входит:

1. Подключиться к одному из проектов, к какому именно указывает владелец машины;
2. Запрашивать задания у центрального сервера;
3. Скачивать задания с сервера, если они там есть;
4. Запускать у себя скачанные задания;
5. Результаты работы заданий отсылать обратно на сервер.

Серверная часть программного обеспечения BOINC выполняет следующие действия:

1. Создает задания для пересылки на вычислительные узлы;
2. Отвечает на клиентские запросы, отправляет задания на вычислительные узлы;
3. Получает результаты работы задания и передает их для дальнейшей обработки;
4. Содержит в себе web-сервер для получения информации о проекте через web-интерфейс.

Клиентская часть распределённого приложения и есть исполняемый файл, запускаемый на вычислительном узле. Она выполняет основную вычислительную нагрузку. Серверная часть распределённого приложения создает задания (расчетные блоки) для вычислительных узлов. Как правило, расчётный блок состоит из исполняемого файла клиенткой части, объединённый со специфическим для конкретного задания входным файлом с данными. После отправки задания в вычислительную инфраструктуру серверная часть распределённого приложения ждёт результатов задания. Получив из инфраструктуры все результаты заданий, серверная часть производит их обработку, и создает единый результат работы распределённого приложения.

Интеграция систем BOINC и USPEX

Принцип работы алгоритма USPEX хорошо сочетается с системами добровольных вычислений. Система USPEX предоставляет возможность добавлять интерфейсы для интеграции с различными системами параллельных и распределённых вычислений. Таким образом можно осуществить интеграцию систем, используя встроенные возможности алгоритма USPEX.

Одной из основных проблем интеграции является балансировка нагрузки между вычислительными узлами. Время выполнения отдельного USPEX-задания может варьироваться от 10-ти секунд, до нескольких часов. Таким образом, получается, что генерация одного задания BOINC на основе одного задания USPEX неэффективна для малых заданий, т.к. издержки на инициализацию превышают полезную работу. По этой причине при интеграции систем использовался подход на основе агрегации нескольких заданий USPEX в одно задание BOINC. Такой подход позволяет снизить издержки на инициализацию одного задания. Параметры для агрегации подбирались эмпирическим способом.

Для осуществления интеграции был разработан набор программных компонентов, осуществляющий взаимодействие между USPEX и BOINC-сервером. На стороне USPEX запускаются скрипты, осуществляющие отправку заданий в инфраструктуру BOINC, а на стороне BOINC-сервера были реализованы скрипты, обрабатывающие запросы системы USPEX.

Обсуждение

Для тестирования стабильности разработанной системы было проведено несколько расчётов. В данной работе мы представим результаты расчёта стандартного примера USPEX по предсказанию кристаллической структуры $MgAl_2O_4$ при давлении 100 ГПа. Это вычисление с переменными параметрами ячейки с помощью потенциала Букингема, реализованного в коде GULP.

Для расчёта были использованы следующие настройки. Число атомов в элементарной ячейке было выставлено равным 28. Структуры первого поколения генерировались случайным образом с помощью специального оператора. Структуры во всех следующих поколениях генерировались с помощью операторов наследования (50%), перестановки (10%), атомной мутации (20%) и случайным образом с помощью специального оператора (20%). После первого поколения процентное соотношение между операторами определялось USPEX самостоятельно.

Для проверки стабильности работы системы процедуры предсказания были проведены 5 раз. Во всех случаях верная структура была найдена в течение первых 5 поколений. Предсказанная структура показана на рис. 2.

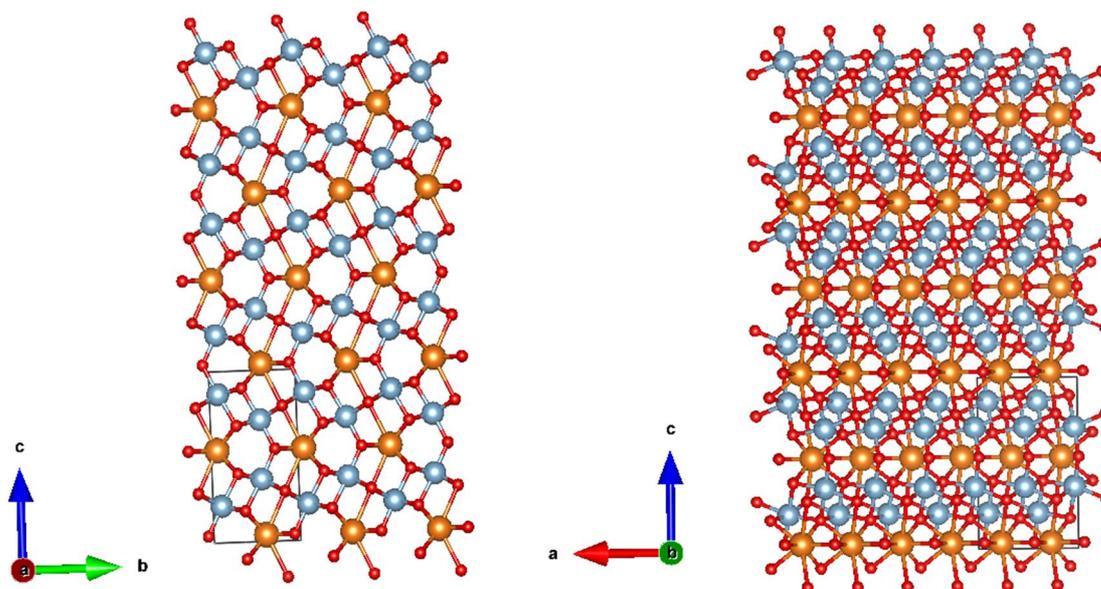


Рис.2. Кристаллическая структура $MgAl_2O_4$ в разных проекциях, предсказанная USPEX. Пространственная группа: C_{2v} Параметры ячейки: $a = 2.63$ $b = 8.78$ $c = 8.72$; $Mg(0.00,0.12,0.75)$, $Si(0.50,0.13,0.076)$, $O(0.50,0.05,0.25)$, $O(0.00,0.27,0.11)$, $O(0.00,0.00,0.00)$

Полученные результаты позволяют заключить, что разработанная система успешно работает. Однако для раскрытия полного потенциала добровольных вычислений потребуются дальнейшая работа.

Во-первых, адаптация самого алгоритма USPEX, изначально разработанного для проведения расчётов на ПК и суперкомпьютерных кластерах. Данные системы имеют стабильную архитектуру и обеспечивают высокую скорость передачи данных, что важно для работы эволюционного алгоритма, который не может перейти в генерации нового поколения без полного завершения текущего.

Кроме того, необходима разработка BOINC-интерфейса для программ для проведения квантово-химических расчётов, позволяющих проводить вычисления с недостижимой для других методов точностью.

Работа выполнена при поддержке грантов РФФИ 16-07-00458 А, 16-07-00873 А, гранта государственной поддержки ведущих научных школ НШ-8860.2016.1.

Литература

- Oganov A.R., Schön J.C., Jansen M., Woodley S.M., Tipton W.W., Hennig R.G. (2010). Modern Methods of Crystal Structure Prediction (ed. A.R. Oganov), pp. 223-231. Berlin: Wiley-VCH.
- Oganov A.R., Glass C.W. (2006). Crystal structure prediction using ab initio evolutionary techniques: principles and applications. J. Chem. Phys. 124, art. 244704.
- Zhang W.W., Oganov A.R., Goncharov A.F., Zhu Q., Boulfelfel S.E., Lyakhov A.O., Stavrou E., Somayazulu M., Prakupenka V.B., Konopkova Z. (2013). Unexpected stoichiometries of stable sodium chlorides. Science 342, 1502-1505.
- Matsko N. L., Tikhonov E. V., Baturin V. S., Lepeshkin S. V., Oganov A. R. (2016). The impact of electron correlations on the energetics and stability of silicon nanoclusters. J. Chem. Phys. 145, 074313.
- Sharma, V., Wang, C., Lorenzini, R. G., Ma, R., Zhu, Q., Sinkovits, D. W., Pilania, G., Oganov, A. R., Kumar, S., Sotzing, G. A., Boggs S.A., Ramprasad R. (2014). Rational design of all organic polymer dielectrics. Nat. Commun. 5, art. 4845.
- Anderson, D. P. (2004, November). Boinc: A system for public-resource computing and storage. In Grid Computing, 2004. Proceedings. Fifth IEEE/ACM International Workshop on (pp. 4-10). IEEE.
- Simons K, Kooperberg C, Huang E, Baker D (1997) Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions. J. Mol. Biol. 268:209-225.
- J.D. Gale and A.L. Rohl. The General Utility Lattice Program (GULP). Mol. Simul., 29:291-341, 2003.
- Oganov A.R., Lyakhov A.O., Valle M. (2011). How evolutionary crystal structure prediction works - and why. Acc. Chem. Res. 44, 227-237.

References

- Oganov A.R., Schön J.C., Jansen M., Woodley S.M., Tipton W.W., Hennig R.G. (2010). Modern Methods of Crystal Structure Prediction (ed. A.R. Oganov), pp. 223-231. Berlin: Wiley-VCH.

2. Oganov A.R., Glass C.W. (2006). Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* 124, art. 244704.
3. Zhang W.W., Oganov A.R., Goncharov A.F., Zhu Q., Bouffefel S.E., Lyakhov A.O., Stavrou E., Somayazulu M., Prakapenka V.B., Konopkova Z. (2013). Unexpected stoichiometries of stable sodium chlorides. *Science* 342, 1502-1505.
4. Matsko N. L., Tikhonov E. V., Baturin V. S., Lepeshkin S. V., Oganov A. R. (2016). The impact of electron correlations on the energetics and stability of silicon nanoclusters. *J. Chem. Phys.* 145, 074313.
5. Sharma, V., Wang, C., Lorenzini, R. G., Ma, R., Zhu, Q., Sinkovits, D. W., Pilania, G., Oganov, A. R., Kumar, S., Sotzing, G. A., Boggs S.A., Ramprasad R. (2014). Rational design of all organic polymer dielectrics. *Nat. Commun.* 5, art. 4845.
6. Anderson, D. P. (2004, November). Boinc: A system for public-resource computing and storage. In *Grid Computing, 2004. Proceedings. Fifth IEEE/ACM International Workshop on* (pp. 4-10). IEEE.
7. Simons K, Kooperberg C, Huang E, Baker D (1997) Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions. *J. Mol. Biol.* 268:209-225.
8. J.D. Gale and A.L. Rohl. The General Utility Lattice Program (GULP). *Mol. Simul.*, 29:291-341, 2003.
9. Oganov A.R., Lyakhov A.O., Valle M. (2011). How evolutionary crystal structure prediction works - and why. *Acc. Chem. Res.* 44, 227-237.

Поступила 21.10.2016

Об авторах:

Оганов Артём Романиевич, профессор Сколковского института науки и технологии, доктор технических наук, A.Oganov@skoltech.ru;

Посыпкин Михаил Анатольевич, зав. отделом, Федеральный исследовательский центр "Информатика и управление" РАН, доктор физико-математических наук, mposypkin@gmail.com;

Ройзен Валерий Валерьевич, младший научный сотрудник и аспирант лаборатории компьютерного дизайна материалов МФТИ, valmipt@gmail.com;

Самцевич Артём, аспирант Сколковского института науки и технологии;

Сухомлин Владимир Александрович, ФГБОУ ВО "Московский государственный университет имени М.В. Ломоносова", доктор технических наук, профессор, заведующий лабораторией открытых информационных технологий факультета вычислительной математики и кибернетики, sukhomlin@mail.ru;

Храпов Николай Павлович, инженер Института передачи информации им. А.А. Харкевича Российской академии наук, nkhrapov@gmail.com.