

ОРГАНИЗАЦИЯ КВАНТОВОХИМИЧЕСКИХ РАСЧЕТОВ С ИСПОЛЬЗОВАНИЕМ ПАКЕТА FIREFLY В ГЕТЕРОГЕННОЙ ГРИД НА БАЗЕ VOINC

Е.Е. Ивашко, Н.Н. Никитина

Введение. Успешное выполнение современных научных и прикладных исследований требует, как правило, проведения большого числа ресурсоемких вычислительных экспериментов. Среди отраслей наук суперкомпьютеринг является наиболее востребованным в квантовой химии, в особенности, в следующих ее направлениях:

- изучение строения и структуры вещества;
- создание веществ и материалов с заранее заданными свойствами;
- механизмы протекания сложных химических реакций;
- и др.

Одной из наиболее популярных форм организации вычислительных ресурсов для проведения квантовохимических расчетов является Грид. Согласно определению [1], Грид-вычисления — это форма организации распределенных вычислений, в которой участвуют территориально распределенные разнородные компьютеры, объединенные с помощью сети передачи данных и работающие совместно для решения задач, требующих значительных вычислительных ресурсов. С точки зрения сетевой организации Грид представляет собой согласованную, открытую и стандартизованную среду, которая обеспечивает гибкое, безопасное, скоординированное разделение вычислительных ресурсов и ресурсов хранения данных [2]. Наиболее эффективно использование Грид при выполнении следующих работ:

- анализ и обработка независимых наборов данных;
- решение задач, обладающих хорошей степенью параллелизации по данным.

Существует большое количество систем промежуточного программного обеспечения (СППО), предназначенного для организации, сопровождения и управления Грид. По назначению СППО можно условно разделить на две группы. Первая группа содержит системы, предназначенные для объединения относительно небольшого числа высокопроизводительных вычислителей (кластеров). К таким системам относятся, например, Condor [3], Globus Toolkit [4], Unicore [5] и другие. Вторая группа СППО содержит системы, цель которых заключается в объединении в Грид большого числа (до сотен тысяч) вычислителей, каждый из которых обладает относительно невысокой производительностью. Специфика таких Грид заключается, в частности, в высокой вероятности недоступности отдельных вычислительных узлов. Наиболее популярной СППО этой группы является платформа VOINC [6]; другой пример такой СППО — разработка НИВЦ МГУ система X-Com [7].

Грид-сегмент КарНЦ РАН. В августе 2010 г. на базе ЦКП КарНЦ РАН "Центр высокопроизводительной обработки данных" [8] был создан и запущен в тестовую эксплуатацию гетерогенный Грид-сегмент, объединяющий вычислительные ресурсы кластера ЦКП КарНЦ РАН, серверов институтов и рабочих персональных компьютеров сотрудников КарНЦ РАН. К февралю 2011 г. в состав Грид были включены управляющий и десять вычислительных узлов кластера, два сервера и один персональный компьютер.

В качестве системы промежуточного программного обеспечения Грид-сегмента КарНЦ РАН используется разработанная в университете Беркли платформа VOINC (Berkeley Open Infrastructure for Network Computing) — активно развивающаяся СППО с открытым исходным кодом, ставшая основой множества независимых проектов добровольных вычислений (наиболее популярные из них — [9–11]). Платформа VOINC отличается простотой в установке, настройке и администрировании, а также обладает хорошими возможностями по масштабируемости, простоте подключения новых вычислительных узлов, использованию дополнительного ПО, интеграции с другими Грид-системами и т.д.

Платформа VOINC имеет архитектуру "клиент-сервер", при этом клиентская часть может работать на компьютерах с различными аппаратными и программными характеристиками. Центральным понятием VOINC является *проект* — автономная сущность, которая производит распределенные вычисления. VOINC-сервер поддерживает одновременную работу большого числа независимых проектов. Один вычислительный узел также может одновременно производить вычисления для нескольких VOINC-проектов. Проект однозначно идентифицируется своим URL-адресом. VOINC предоставляет возможность гибкой настройки клиентской части, регулируя максимальный размер загружаемых файлов, время выполнения рабочих заданий, загрузку CPU или GPU, выделяемый объем оперативной памяти и дискового пространства.

Рабочий процесс в Грид-системе, основанной на платформе VOINC, организован следующим образом. Вычислительные узлы, имеющие свободные ресурсы, обращаются к серверу для получения новых рабочих заданий. Сервер VOINC рассылает клиентским приложениям экземпляры рабочих заданий, клиенты выполняют вычисления и отсылают обратно результаты. После получения результатов сервер проверяет и обрабатывает их, например, заносит в базу данных или автоматически создавая на их основе новые рабочие задания.

Платформа BOINC позволяет выполнять в Грид-системе специализированные приложения, в которых, в частности, может быть реализован механизм сохранения промежуточных результатов вычислений, графическое отображение прогресса выполнения приложения, автоматическое распараллеливание вычислительного процесса в зависимости от количества доступных на клиенте вычислительных ядер и т.п. Однако унаследованные (неадаптированные для работы в Грид) приложения с помощью специальных приложений-"обертки" также могут быть перенесены на платформу BOINC без необходимости изменения их исходного кода и перекомпиляции [12]. Приложения-"обертки" берут на себя взаимодействие с ядром программы-клиента, запуская исходное приложение как свой дочерний процесс.

Квантовохимические расчеты с использованием системы Firefly. Программный пакет Firefly [13], частично основанный на исходном коде системы GAMESS (US) [14], является одним из наиболее популярных средств проведения квантовохимических расчетов. Firefly предназначен для выполнения расчетов *ab initio* и DFT на Intel-совместимых процессорах x86, AMD64 и EM64T. Вычислительные методы *ab initio*, или неэмпирические, предназначены для расчета с максимально возможной точностью физических и химических свойств заданного химического соединения (как многоатомной системы) на основе представлений и методов квантовой механики.

В феврале 2011 г. на базе Грид-сегмента КарНЦ РАН была проведена серия расчетов энергий многоатомных систем с использованием пакета Firefly. Вычислительные эксперименты проводились магистранткой Петрозаводского государственного университета Кременецкой О. В. в рамках проекта по прогнозированию состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов на основе квантовохимических расчетов модельных систем. Проект ведется ПетрГУ совместно с лабораторией высокотемпературной электрохимии Института химии Кольского научного центра РАН с использованием вычислительных ресурсов ЦКП КарНЦ РАН «Центр высокопроизводительной обработки данных» и при финансовой поддержке РФФИ (грант №08-03-00397-а). Основные результаты, полученные в рамках проекта, опубликованы в работах [15,16].

Объектом исследования в проекте являются расплавы фторидов и хлоридов щелочных металлов Na, K, Cs, содержащих небольшие (~1 моль%) добавки фторидных и хлоридных комплексов переходных металлов. Интерес к данным объектам обусловлен тем, что они широко используются в промышленности для проведения таких технологических процессов, как:

- получение чистых и высокочистых металлов;
- получение защитных и каталитически активных покрытий;
- синтез соединений, которые невозможно получить из водных и неводных сред при низких температурах.

Знание состава комплексов в расплаве позволяет гораздо точнее моделировать химические реакции и перенос заряда в расплаве. При этом знание механизма реакции и влияния на него разных факторов — состава, температуры, поверхности и др. — дает возможность управления реакцией.

Расчеты проводятся квантовохимическими методами HF, MP2, DFT с помощью программного пакета Firefly. Экспериментально установлено, что процесс переноса заряда в исследуемых объектах протекает по-разному в зависимости не только от катионного, но и от анионного состава электролита расплава. Это обстоятельство делает необходимым при расчете различных характеристик включение в модельную систему не только второй, но и третьей координационной сферы (КС) комплексов переходных металлов. Под первой КС понимается совокупность атомов или групп атомов, непосредственно присоединенных к рассматриваемому. Все атомы, с которыми связан центральный посредством атомов первой КС, образуют его вторую КС и т. д.; *k*-я КС образована атомами, которые связаны с атомами *k*-1-й КС, без учета всех атомов *k*-2-й КС.

Выбор состава второй КС основан на том факте, что расчетные энергетические параметры, определяющие устойчивость комплексных частиц (комплекс + катионы первой КС), имеют экстремальный характер зависимости от числа внешнесферных катионов.

Процедура нахождения наиболее вероятного состава окружения комплекса в расплаве включает:

1. расчет энергии образования частицы при разных значениях числа катионов во второй координационной сфере и определение наиболее устойчивого состава;
2. расчет энергии активации переноса заряда (E_{act}) частиц различного состава;
3. сопоставление результатов с экспериментальными данными по константам скорости переноса заряда.

В рамках проекта исследуется процесс присоединения электрона к комплексу, в результате которого происходит изменение не только валентности переходного металла, но и положения атомов как входящих в состав комплекса, так и окружающих его. Поэтому для реализации указанной выше процедуры требуется рассчитать энергию частиц с разным числом катионов (1–7) во второй КС комплекса до и после присоединения электрона, а также энергию неравновесных частиц, в которых переходной металл имеет валентность, соответствующую состоянию до присоединения электрона, но геометрическую структуру состояния после присоединения (эта энергия нужна для вычисления энергии активации переноса электрона и сравнения с экспериментальными данными).

Для учета влияния анионного состава расплава в модельную систему вводилось некоторое количество электролита MX . Производилась оптимизация геометрической структуры систем вида $3\text{MCr(III)F}_6+18\text{MCl}$ и $3\text{MCr(II)F}_6+18\text{MCl}$ ($\text{M} = \text{Na}, \text{K}, \text{Cs}$), а также расчет энергии неравновесной структуры $3\text{MCr(III)F}_6+18\text{MCl}$ в геометрии $3\text{MCr(II)F}_6+18\text{MCl}$. Для каждой из систем были рассчитаны энергии «псевдочастиц» вида ${}^*(n\text{MCrF}_6)$, $n=1, \dots, 7$, где верхний индекс звездочки обозначает, что геометрическая структура и базис волновых функций для таких частиц соответствуют всей системе.

Таким образом, требовалось рассчитать по 7 энергий частиц для трех систем и трех видов катионов, то есть 63 энергии, не включая еще некоторые дополнительные менее затратные по времени расчеты. При квантовохимических расчетах время счета пропорционально четвертой степени числа волновых функций (атомных орбиталей), составляющих молекулярные орбитали. В данном случае расчет энергий частиц $n\text{MCrF}_6$ производился в базисе волновых функций всей системы ($3\text{MCr(III)F}_6+18\text{MCl}$, например), следовательно, число атомных орбиталей очень велико. Кроме того, все исследуемые параметры составляют сравнительно небольшую разность двух больших по величине значений энергии, что предъявляет дополнительные требования к точности вычислительного эксперимента. Для этого необходимо было использовать широкие базисные наборы, т.е. раскладывать каждую атомную орбиталь на большое количество математических функций (например, Гаусса) и при расчете подбирать большое количество коэффициентов разложения, что также увеличивает время счета.

Указанные обстоятельства приводят к необходимости задействования больших вычислительных ресурсов. Однако требование организации большого числа однотипных расчетов при вычислении энергий частиц дало возможность использовать Грид-сегмент КарНЦ РАН для проведения вычислительных экспериментов в рамках данной задачи.

Грид-сегмент использовался для проведения расчетов энергий частиц, состоящих из комплекса CrF_6 и от одного до семи катионов Na^+ или Cs^+ , содержащих трехвалентный хром, но имеющих расположение атомов, соответствующее системе с двухвалентным хромом ($3\text{MCr(II)F}_6+18\text{MCl}$, $\text{M} = \text{Na}, \text{Cs}$). Значения энергии частиц использовались для оценки энергии активации переноса электрона и сравнения с экспериментальными данными. Кроме того, рассчитывались энергии комплексов Cr(III)F_6 и частиц, состоящих только из катионов Na^+ или Cs^+ (от одного до семи) в геометрии систем $3\text{MCr(II)F}_6+18\text{MCl}$, $\text{M} = \text{Na}, \text{Cs}$. Такие расчеты позволяют оценить вклад энергии, требующейся для перестройки электронных оболочек атомов, изменения геометрической структуры комплекса и сближения катионов между собой, в общую энергию образования частицы, состоящей из комплекса и катионов.

Исходные данные для проведения вычислительных экспериментов представляют собой наборы параметров, регулирующих работу программы, в том числе координаты атомов, метод расчета и базис волновых функций. Выходные файлы должны содержать информацию о полной энергии рассчитываемой системы, ее энергетических уровнях, параметрах молекулярных орбиталей и т.п.

Реализация вычислительных экспериментов в Грид-сегменте КарНЦ РАН. На момент проведения вычислительных экспериментов в состав Грид-сегмента КарНЦ РАН входили 14 вычислительных узлов с различными аппаратными и программными характеристиками, а также разными настройками, связанными с организацией вычислений. В частности, два узла обслуживали проекты BOINC в монопольном режиме, а на десяти вычислительных узлах кластера расчеты в рамках Грид регулярно приостанавливались для выполнения расчетов пользовательских задач, запускаемых на кластере с помощью системы управления заданиями. Суммарная пиковая производительность Грид составила 1,04 TFLOPS.

Все рабочие задания выполнялись в однопроцессорном режиме. С учетом настроек ограничения времени работы BOINC на вычислительных узлах, а также накладных расходов на передачу данных между сервером и клиентами, максимальное время выполнения одного задания составило 66,0 часов, среднее — 17,7 часов. Суммарное процессорное время выполнения всего набора рабочих заданий составило 383,7 часов, а время между отправкой первого рабочего задания и получением последнего результата составило 66,5 часов.

Адаптация пакета Firefly для выполнения на платформе BOINC была проведена при помощи программы-"обертки" GenWrapper, которая позволяет в дополнение к основному приложению выполнять сценарии командной оболочки, соответствующие стандарту POSIX. Сценарии использовались для установки значений переменных окружения и составления параметров командной строки, необходимых для запуска Firefly, а также для определения статуса завершившегося рабочего задания по содержимому выходного файла.

Заключение. В рамках Грид-сегмента на базе ЦКП КарНЦ РАН было организовано проведение квантовохимических расчетов с использованием программного пакета Firefly. Пакет Firefly был адаптирован для выполнения в Грид на платформе BOINC при помощи программы-"обертки" GenWrapper. Использование Грид позволило снизить нагрузку и повысить эффективность использования вычислительного кластера ЦКП КарНЦ РАН «Центр высокопроизводительной обработки данных».

По итогам проведения вычислительных экспериментов был разработан ряд изменений, которые необходимо внести в процесс обработки данных в Грид-системе для повышения эффективности и снижения суммарного времени расчетов:

- использовать на узлах Грид сборки Firefly, оптимизированные под архитектуру и ПО

вычислителя;

- проверять необходимость передачи пакета Firefly на вычислительный узел;

Помимо квантовохимических расчетов, Грид-система Центра высокопроизводительной обработки данных используется также для организации вычислительных экспериментов в интересах других научных и прикладных исследований, проводимых пользователями ЦКП КарНЦ РАН «Центр высокопроизводительной обработки данных».

ЛИТЕРАТУРА:

1. Грид-вычисления: Wikipedia. URL: <http://ru.wikipedia.org/wiki/Грид>
2. Foster I. The Grid: Blueprint for a New Computing Infrastructure. — Morgan Kaufmann Publishers. — ISBN ISBN 1-55860-475-8.
3. Condor. High Throughput Computing. URL: <http://www.cs.wisc.edu/condor>
4. The Globus Toolkit. URL: <http://www.globus.org/toolkit>
5. UNICORE — Distributed Computing and Data Resources. URL: <http://www.unicore.eu>
6. BOINC: Программное обеспечение с открытым исходным кодом для организации добровольных распределённых вычислений и распределённых вычислений в сети. URL: <http://boinc.berkeley.edu/index.php>
7. Семейство программ X-Com. URL: <http://x-com.parallel.ru/>
8. Центр высокопроизводительной обработки данных ЦКП КарНЦ РАН / Институт прикладных математических исследований Карельского научного центра РАН. URL: <http://cluster.krc.karelia.ru>
9. Проект добровольных вычислений Climateprediction.net. URL: <http://climateprediction.net>
10. Проект добровольных вычислений SETI@home. URL: <http://setiathome.berkeley.edu>
11. Проект добровольных вычислений Einstein@home. URL: <http://einstein.phys.uwm.edu>
12. GenWrapper. A generic BOINC wrapper for legacy applications. URL: <http://genwrapper.sourceforge.net>
13. A.A. Granovsky, Firefly version 7.1.G, URL: <http://classic.chem.msu.su/gran/firefly/index.html>
14. M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery. General Atomic and Molecular Electronic Structure System // J. Comput. Chem. — 1993. — V.14. — P. 1347-1363.
15. Кременецкий В. Г., Кременецкая О. В., Кузнецов С. А., академик Калинин В. Т. Квантово-химический подход к оценке состава устойчивых комплексных частиц в расплавах галогенидов щелочных металлов // Доклады Академии наук, 2011, том 437, №6, с. 782–784.
16. Stulov Yu. V., Kremenetsky V. G., Kremenetskaya O. V., Fofanov A. D., Kuznetsov S. A. Standart rate constants of charge transfer for the redox couple Cr(III)/Cr(II) in chloride melts: experiment and calculation // 9th International Frumkin Symposium. Electrochemical technologies and materials for XXI century. Abstracts, Moscow 2010. P. 31.