

КРУПНОМАСШТАБНЫЕ ЗАДАЧИ ХИМИИ НА ПАРАЛЛЕЛЬНЫХ И РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ ПОЛИГОНАХ: СОВРЕМЕННОЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ

В.М. Волохов, Д.А. Варламов, А.В. Пивушков

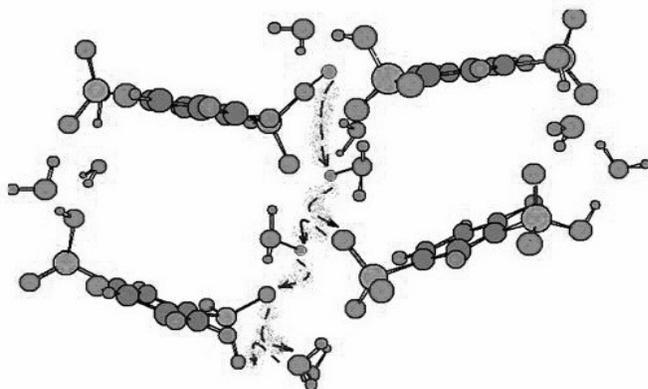
Квантово-химические расчеты являются важнейшим звеном при проведении исследований в области строения вещества, наноматериалов, физики твердого тела, биофизики и всех научных дисциплин, связанных с исследованием электронной структуры вещества и его строения.

В настоящее время в мире реализуется несколько проектов в области распределенных вычислений, близких по направлению к нашим исследованиям и организационно представленных виртуальными организациями. Отметим два из них. Первый проект представлен ВО GEMS (Grid Enabled Molecular Simulator) и выполняется в рамках EGEE. Координационный центр находится в Перуджио (Италия). В проекте решается амбициозная задача формализации и автоматизации расчета химических процессов, начиная от исследования электронной структуры и кончая получением кинетических и вероятностных характеристик процесса. Второй проект представлен порталом Nimrod (<http://www.csse.monash.edu.au/~davida/nimrod>), специализирующимся на многопараметрическом моделировании, и решает широкий круг задач вычислительной химии, астрофизики, геофизики и т.д.

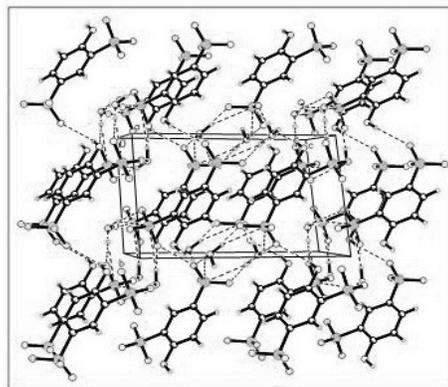
Крупномасштабные квантово-химические расчеты - основное направление деятельности ВЦ ИПХФ РАН [1,2]. Эти расчеты выполняются в основном с использованием наиболее популярных программ Gaussian, GAMESS, Dalton и CPMD, являющихся плодом многолетнего труда интернациональных коллективов физиков, химиков, программистов, специалистов в области вычислительных методов и многих других, что делает их уникальным достоянием научного сообщества и требует специализированной вычислительной базы. Институт располагает богатейшей в России и постоянно пополняемой библиотекой параллельных квантово-химических и молекулярно-динамических программ. Несколько лет в ИПХФ функционирует ресурсный GRID-узел, являющийся составной частью сообщества RDIG, и использующийся для проведения экспериментов в области крупномасштабных параллельных и квантово-химических расчетов [1,3]. Работы с системами распределенных и параллельных вычислений в ИПХФ РАН проводились в 2005-2007 гг. в рамках программы № 21 фундаментальных исследований Президиума РАН «Разработка фундаментальных основ создания научной распределенной информационно-вычислительной среды на основе технологий GRID», в 2005-м году в рамках Федеральной целевой научно-технической программы "Исследования и разработки по приоритетным направлениям развития науки и техники" по проекту "Создание комплекса пакетов прикладных программ для моделирования сложных научных и промышленных задач на суперкомпьютерных системах терафлопного уровня и в распределенных вычислительных средах", а также в 2007-2008 годах в рамках программы Союзного Государства «СКИФ-ГРИД».

Многолетний опыт проведения подобных расчетов [1-4] позволил разделить квантово-химические задачи на два основных класса:

- 1) задачи, распадающиеся на совокупность независимых заданий (пример - рис. 1),
- 2) задачи, представляющие собой единый вычислительный процесс (пример - рис. 2).



А – изолированная молекула



Б – наноструктура

Рис.1 Исследование протонной проводимости в наноматериале, используемом для производства мембран твердых топливных элементов. Использовались программы GAMESS и GAUSSIAN-D03. Одна точка расчета энергии протона на пути миграции (А, размытая линия) требует около месяца расчетного времени на 8 процессорах класса Intel em64t Xeon 3,6 ГГц. Для расчета 500 точек в среде GRID (Б) необходимо около 4000 CPU и месяц расчетов.

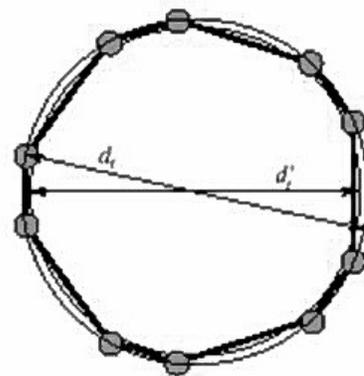
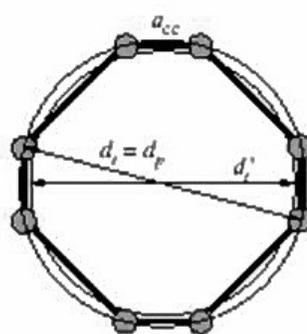
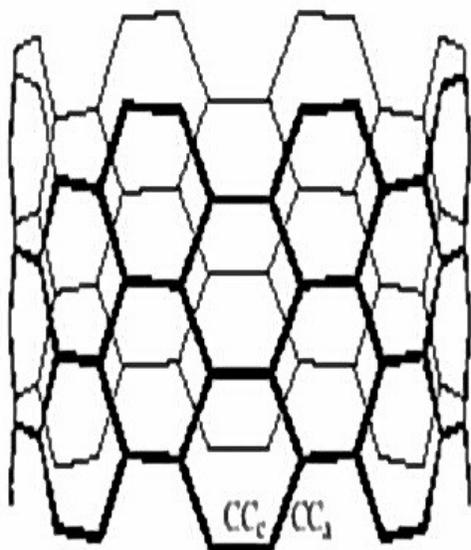


Рис.2 Расчеты нанотрубок как наноструктурных наполнителей, упрочнителей и композитов конструкционного назначения и процессов, проходящих на их поверхности. Использовались программы GAMESS и Gaussian-D03. Расчет нанотрубки (200 атомов) с атомами Pt на поверхности требует около 3-4 недель расчетного времени на 8 процессорах класса Intel EM64T Xeon 3,6 ГГц. Для расчета нанотрубки, содержащей 2000 атомов C + Pt необходимо месяц времени и ориентировочно 800 подобных CPU.

Примеры типичных задач первого и второго класса приведены соответственно на рис.1 и 2.

Характерная особенность задач первого класса состоит в том, что, распределяя независимые задания на множество небольших кластеров (каждый кластер – 10-20 процессоров, задание исполняется на нем как параллельное) можно добиться высокой эффективности использования вычислительных ресурсов. При этом возможно использование весьма больших вычислительных полигонов (до 10^3 процессоров) как в локальном варианте (в гетерогенной вычислительной среде), так и в условиях распределенных сред (совокупность удаленных кластеров).

Задачи второго класса представляют собой существенную проблему, т.к. эффективность их решения непосредственно связана с эффективностью распараллеливания вычислительного процесса и высокими требованиями к ресурсам узла. Для программы Gaussian известно эмпирическое правило: масштабируемость пропорциональна кубическому корню из числа процессоров. Для программы GAMESS масштабируемость существенно лучше и для нескольких десятков процессоров остается практически линейной. Таким образом, для характерных задач исследования наноструктур и молекулярных кристаллов необходимо несколько тысяч процессоров и процессорное время порядка месяца.

Нами были проведены масштабные квантово-химические расчеты ряда программ на локальных параллельных ресурсах ИПХФ (с использованием до 130 процессоров) и в распределенной среде GRID на основе middleware gLite на вычислительных ресурсах BO RGSTest (с использованием в ряде экспериментов до 400 процессоров на 4-5 ресурсных узлах), которые показали значительную эффективность применения параллельных и/или распределенных технологий для решения большинства классов химических задач.

Для проведения этих экспериментов была проведена адаптация прикладных пакетов вычислительной химии к работе в параллельных и/или распределенных средах. Были изучены различные методы параллелизации прикладных пакетов и авторских программ: сокетная параллелизация, работа в средах Mpiich 1 и 2, применение методов формирования «пучков» независимых заданий (до $n \cdot 10^4$ заданий). Были разработаны низкоуровневые интерфейсы для работы параллельных приложений в распределенных вычислительных средах типа gLite, X-Com, Upisoge. Основные методы адаптации химических задач к работе в различных вычислительных обстановках более подробно описаны в представленном на этой же конференции докладе авторов «ВАРИАНТЫ РЕШЕНИЯ ПРИКЛАДНЫХ ЗАДАЧ ХИМИИ В РАСПРЕДЕЛЕННЫХ И ПАРАЛЛЕЛЬНЫХ СРЕДАХ».

Проведенные крупномасштабные вычислительные эксперименты продемонстрировали огромные возможности применения параллельных и/или распределенных технологий в области вычислительной химии. Существенные трудности, связанные с установкой соответствующего ПО различного уровня многократно перекрываются новыми возможностями и перспективами в решении крупномасштабных химических задач.

ЛИТЕРАТУРА:

1. Д.А. Варламов, В.М. Волохов, Г.А. Покатович, Н.Ф. Сурков, А.В. Пивушков Труды Международной научной конференции “Параллельные вычислительные технологии (ПаВТ’2008)” ; Санкт-Петербург, 28 января–1 февраля 2008 г., с.318-322
2. Д.А. Варламов, В.М. Волохов, А.В. Пивушков, Н.Ф. Сурков Труды Всероссийской научной конференции, "Научный сервис в сети ИНТЕРНЕТ: технологии параллельного программирования. 15 лет РФФИ", Новороссийск, 2007 г., М.:, изд-во МГУ, 2007, с.230-233
3. С.М. Алдошин, Д.А. Варламов, В.М. Волохов, Г.А. Покатович, Н.Ф. Сурков, А.И. Станиловский Труды Всероссийской научной конференции «Научный сервис в сети Интернет: технологии параллельного программирования», Новороссийск, 2006, М.:, изд-во МГУ, с.91-93
4. Д.А. Варламов, В.М. Волохов, Г.А. Покатович, Н.Ф. Сурков, А.В. Пивушков Труды второй международной конференции «Распределенные вычисления и Грид-технологии в науке и образовании», Дубна, 2006, с.243-254