

ГРИД-ЦЕНТР ИПХФ РАН ДЛЯ КОМПЕТЕНЦИИ ПРИКЛАДНОГО ПО И ПРОВЕДЕНИЯ КВАНТОВО-ХИМИЧЕСКИХ И МОЛЕКУЛЯРНО-ДИНАМИЧЕСКИХ РАСЧЕТОВ НА ОСНОВЕ «ГИБРИДНЫХ» И РАСПРЕДЕЛЕННЫХ ТЕХНОЛОГИЙ

В.М. Волохов¹, Д.А. Варламов^{1,2}, А.В. Пивушков¹, А.В. Волохов¹, Г.А. Покатович¹

¹*Институт проблем химической физики РАН, Черноголовка*

²*Институт экспериментальной минералогии РАН, Черноголовка*

Описаны перспективы создания в ИПХФ РАН проблемно-ориентированного грид-центра квантово-химического и молекулярно-динамического прикладного ПО, а также проведения расчетов с широким использованием «гибридных» и распределенных технологий вычислений. Центр предназначен для компетенции широкого спектра прикладных пакетов в области химии и смежных наук и проведения ресурсоемких расчетов с их использованием, а также организации упрощенного доступа пользователей через клиентские интерфейсы высокого уровня к высокопроизводительным вычислительным ресурсам. Помимо общей компетенции большого количества прикладных пакетов будут созданы доступные веб- и грид-ориентированные вычислительные сервисы для квантовой химии и молекулярной динамики.

Введение

Вычислительная (прежде всего квантовая) химия, молекулярная динамика и молекулярное моделирование, а также сопряженные с ними области знаний являются наиболее заинтересованными в высокопроизводительных вычислениях отраслями науки. Исследования, проводимые в области химии и смежных наук, в настоящее время, как правило, неэффективны без использования сверхмощных параллельных и распределенных вычислительных ресурсов для решения задач самых разных классов [1].

Современное мировое состояние вычислительной химии характеризуется тотальным использованием мощных вычислительных ресурсов – как параллельных (включая «гибридные» на основе графических процессоров и мультядерных сопроцессоров), так и распределенных – для решения задач различных классов. В настоящее время в крупнейших мировых научных центрах, специализирующихся в теоретической химии и теории химических превращений, ведутся активные работы по разработке наиболее эффективных вычислительных процедур для параллельных, «гибридных» и распределенных вычислительных сред, позволяющих исследовать самые разнообразные химические процессы.

Потенциально задачи в области молекулярного моделирования и многоуровневого иерархического моделирования материальных объектов (от квантово-механического уровня до уровня сплошных сред и конструкций) могут выходить на уровень востребованности многих петафлопс. Некоторые задачи оптимизации крупных молекулярных структур требуют выполнения до 10^9 отдельных расчетов. Типичный докинг белковых лигандов с размерностью 200 атомов \times 300 000 конфигураций \times 1000 CPU-часов требует до 300 Пф вычислительных мощностей. Для построения многомерных потенциальных поверхностей, адекватно описывающих химические реакции, нужно провести 10^2 – 10^N независимых весьма ресурсоемких *ab initio* расчетов. Детальное моделирование даже отдельных этапов каталитических реакций в топливных элементах на основе водорода требует до 60000 CPU-часов. Для исследования динамики не очень сложной

химической реакции с использованием метода классических траекторий зачастую нужно рассчитать до 10^7 - 10^9 независимых траекторий. Численное исследование многопараметрических функций $F(x_1, x_2 \dots x_n)$ в области изменения параметров $x_1, x_2 \dots x_n$ при разбиении диапазона изменения каждого параметра на 10 ячеек требует проведения 10^N независимых расчетов F , и так далее.

Востребованность вычислительной химией все возрастающих вычислительных ресурсов подтверждается тем, что большинство суперкомпьютерных центров США и Европы предоставляют до 40% вычислительных мощностей для нужд биохимии, молекулярного моделирования, квантовой химии, нанотехнологических расчетов. При этом существует устойчивая тенденция на создание проблемно-ориентированных суперкомпьютерных центров (как вариант – выделение фиксированных пулов ресурсов в рамках петафлопсных суперкомпьютеров), специализирующихся исключительно на квантово-химических и молекулярно-динамических расчетах. При этом эти ресурсы часто входят в различные грид-полигоны и образуют достаточно обширные виртуальные организации (типа BO Gaussian в консорциуме EGI), объединяющие на разных уровнях ресурсы и пользователей.

К сожалению, при наличии в Российской Федерации достаточно мощных вычислительных суперкомпьютерных центров (МГУ, МСЦ, Саров, ЮжУрГУ и т.п.) подобные проблемно-ориентированные центры достаточной мощности в настоящее время практически отсутствуют. Данная публикация предназначена осветить их актуальность, а также предлагаемые технологии создания и аспекты эффективной эксплуатации.

1. Основы для создания грид-центра компетенции

Ранее, в результате проводимых в 2004-2012 годах в ИПХФ РАН (Научный центр Черноголовка) исследовательских работ в области распределенных вычислений был сформирован прототип предлагаемого к созданию проблемно-ориентированного вычислительного грид-центра. Работы с квантово-химическими прикладными пакетами в условиях суперкомпьютерных установок и грид-полигонов распределенных вычислений в ИПХФ РАН осуществлялись в течение более чем десятилетия по программам Президиума РАН, федеральным целевым научно-техническим программам, грантам РФФИ и Министерства науки и образования, программам Союзного Государства РФ-РБ [2, 3]. Исследования включали: 1) создание ресурсных сайтов ряда грид-полигонов (EGEE-RDIG/EGI-[RU-NGI], СКИФ-Полигон, Национальная нанотехнологическая сеть, Российская грид-сеть для высокопроизводительных вычислений); 2) адаптацию прикладного ПО в области вычислительной химии к работе на суперкомпьютерах и в различных грид-инфраструктурах, включая создание высокоуровневых грид-сервисов; 3) создание и апробацию новых методов вычислений на суперкомпьютерах и в распределенных средах, связанных со спецификой используемого ПО [4]; 4) проведение с использованием созданной инфраструктуры широкомасштабных вычислений в интересах пользователей грид-полигонов.

В 2012-2013 годах на базе созданного пула «гибридных» вычислительных узлов была проведена первичная компетенция квантово-химических и молекулярно-динамических пакетов, адаптированных на проведение расчетов с использованием GPU сопроцессоров, а также проведен ряд расчетов в области моделирования топливных элементов, молекулярной динамики, расчета кристаллических структур [2, 4].

В 2013 году в распоряжение авторов поступил вычислительный кластер с пиковой производительностью около 15 Тф (176 двухпроцессорных узлов HP Proliant на основе 4- и 6-ядерных процессоров Intel Xeon 5450 и 5670 частотой 3 ГГц и оперативной памятью 8 и 12 Гбайт – всего 1472 ядра; коммуникационная сеть Infiniband DDR на основе коммутаторов Voltaire и Cisco – скорость двунаправленного обмена данными 1400

Мбайт/с, латентность 3.2-4.5 мкс; транспортная и управляющая сети – Gigabit Ethernet; жесткие диски – не менее 36 Гбайт на узел), а также мини-кластер на базе двух высокопроизводительных «гибридных» узлов (1 узел – 12 вычислительных ядер [2x6 3.46GHz Intel® Xeon® X5675], 48 Гб RAM, 3 Тб HDD/SSD, 2 GPU Nvidia Tesla C2075, 3 сети Gigabit Ethernet, пиковая производительность узла – до 1.3 Тф для вычислений двойной точности и 2.3 Тф для вычислений ординарной точности). Пиковая производительность данного пула составляет около 2.6 Тф.

На базе данного оборудования начато создание (с использованием существующего прототипа) высокопроизводительного проблемно-ориентированного грид-сайта, направленного на решение «тяжелых» задач квантовой химии и молекулярной динамики с применением новейших технологий параллельных, распределенных и «гибридных» вычислений, а также на разработку и внедрение новых оригинальных методов вычислений для решения и оптимизации подобных задач.

2. Задачи центра

В процессе исследований давно стала очевидной потребность в создании высокопроизводительного ресурсного центра, узкоспециализированного на использовании в интересах вычислительной химии и смежных с ней областей науки.

На основе опыта предыдущих работ предложены следующие направления и методы, используемые для создания и успешной работы такого центра:

- 1) Отбор наиболее востребованных пользователями квантово-химических и молекулярно-динамических программных прикладных пакетов (ППП) – свободно распространяемых и лицензируемых, их жесткая конкуренция между собой, определение возможностей работы в качестве грид- и веб-ориентированных вычислительных сервисов, а также уровня реализации в них новейших вычислительных технологий;
- 2) Установка и отладка максимально широкого спектра выбранного ПО, эксплуатация его в интересах пользователей в рамках локального ресурса для решения научно-практических задач;
- 3) Разработка и внедрение в практику квантово-химических расчетов новейших вычислительных технологий («гибридные» вычисления на базе графических ускорителей и мультиядерных сопроцессоров, виртуализация вычислений и т.п.), использование адаптированных к данным технологиям версий ППП;
- 4) Интеграция центра в доступную грид-среду (в том числе с созданием соответствующих ресурсных сайтов и виртуальных организаций), адаптация прикладного ПО в качестве грид-сервисов для решения входящих задач, создание пользовательских интерфейсов для запуска ППП в качестве исходящих заданий грид-полигона;
- 5) Создание на базе центра веб-портала с интегрированными в него высокоуровневыми грид- и веб-интерфейсами к прикладным пакетам, предназначенными для формирования, запуска и мониторинга локальных и грид-заданий, и последующей доставки результатов пользователям. Портал обеспечивает корректную авторизацию пользователей, хранение их данных, учет и выбор ресурсов, а также поддерживает многие другие функции, максимально повышающие эффективность работы конечных пользователей.

Основным подходом является комбинированное использование новейших общепотребительных технологий и методов проведения высокоинтенсивных расчетов (в первую очередь «гибридных» вычислений) совместно с ранее разработанными авторами методами расчетов в грид-средах. Результатом этого должно стать создание сбалансированного проблемно-ориентированного ресурса, объединяющего комплекс разработанных методов расчетов, центр компетенции ПО, собственно высокопроизводительный вычислительный ресурс (ориентированный на решение задач квантовой химии и

молекулярной динамики) с поддержкой «гибридной» модели вычислений, полноценный грид-сайт и высокоуровневые веб-интерфейсы.

3. Преимущества создания центра и проблемы его эксплуатации

Преимущества создания подобных центров очевидны:

- 1) создание пулов узлов, ориентированных именно на «тяжелые» квантово-химические вычисления (большие объемы RAM и HDD на ресурсных узлах, специфика коммуникационной среды и т.п.);
- 2) возможность установки и тестирования очень широкого спектра прикладных пакетов в выбранной области;
- 3) квалифицированная поддержка пользователей системными специалистами, хорошо знакомыми с особенностями конкретного прикладного ПО, а также ориентирующимися в весьма сложных системах входных данных и многозначности стратегии вычислений;
- 4) создание специализированных высокоуровневых интерфейсов к прикладному ПО для работы с локальными и грид-заданиями;
- 5) легкость (по сравнению с суперкомпьютерами общего назначения) интеграции в различные грид-полигоны с уже готовыми прикладными вычислительными сервисами, адаптируемыми к грид-средам.

Отдельно следует указать, что стремительное развитие «гибридных» (CPU+GPU или CPU+ускорители типа Intel Phi) вычислительных технологий достигло точки, когда множество существующих приложений в различных областях реализуются с их использованием и работают значительно быстрее, чем на обычных многоядерных системах. В настоящее время резко интенсифицировался перевод программного кода прикладных пакетов на «гибридные» технологии вычислений. При этом программирование обновленного или создаваемого заново «гибридного» кода для квантово-химических и молекулярно-динамических пакетов занимает одно из ведущих позиций в этом направлении. В течение последних лет существенно упростилось программирование соответствующей модели параллельных вычислений с использованием среды программирования CUDA™, которая обеспечивает набор абстракций, позволяющих выражать как параллелизм данных, так и параллелизм задач. Это позволяет проводить высокоинтенсивные вычисления с применением встраиваемых в расчетные узлы кластеров высокопроизводительных графических ускорителей (типа Nvidia Tesla и Kepler) и многоядерных сопроцессоров Intel Xeon Phi, что значительно повысит эффективность использования параллельных методов расчетов и снизит как требования к расчетным центрам, так и операционную стоимость расчетов, особенно при использовании большого числа расчетных узлов. Применение методов параллелизации с использованием GPU-устройств или ускорителей Intel Xeon Phi для прикладных пакетов в области квантовой химии и молекулярной динамики в большинстве случаев ведет к повышению эффективности расчетов от десятков процентов до нескольких порядков (PetaChem, Gromacs, NAMD, VMD) либо (как вариант) дает возможность существенного повышения точности или детальности расчетов.

К сожалению, в России при наличии в настоящее время достаточного количества ресурсов с технической поддержкой «гибридных» технологий пока гораздо меньше времени уделяется адаптации и тестированию прикладного ПО (в том числе с исходными кодами) к работе в подобных условиях, что в значительной мере снижает эффективность использования дорогостоящих ускорителей.

Заключение

Создаваемый проблемно-ориентированный грид-центр (базирующийся на достаточно мощном, минимум первые десятки терафлопс, вычислительном ресурсе) направ-

лен на разработку и внедрение новейших вычислительных (параллельных, распределенных, «гибридных») технологий с использованием широкого спектра прикладных пакетов. Он позволит решать на принципиально новом вычислительном уровне многочисленные «тяжелые» задачи в следующих областях: строение молекул и структура твердых тел, кинетика и механизмы сложных химических реакций, химическая физика процессов горения и взрыва, процессов образования и модификации полимеров, биологических процессов и систем, общие проблемы химической физики. Специализированный центр позволяет также разрабатывать новые оригинальные методы проведения локальных и распределенных вычислений.

Исследование выполнено при поддержке Министерства образования и науки Российской Федерации, соглашение №8026 от 10.07.2012 г.

Литература

1. Суперкомпьютерные технологии в науке, образовании, промышленности. Вып. 3 (под ред. В.А.Садовниченко). – М.: Изд-во МГУ, 2012. – 232 с.
2. Волохов В.М., Варламов Д.А., Пивушков А.В., Волохов А.В. Применение GRID технологий в области вычислительной химии // Известия Академии наук. Серия химическая. 2011. № 7. С. 1483-1490.
3. Волохов В.М., Варламов Д.А., Волохов А.В., Пивушков А.В., Покатович Г.А., Сурков Н.Ф. Грид-сервисы в вычислительной химии: достижения и перспективы // Вестник Уфимского государственного авиационно-технического университета. Серия Управление, вычислительная техника и информатика. 2011. Т.15, № 5 (45). С. 161-169.
4. Пивушков А.В., Волохов В.М., Варламов Д.А., Волохов А.В. Применение новых вычислительных технологий для повышения эффективности расчетов в грид-средах // Вестник Уфимского государственного авиационно-технического университета. Серия Управление, вычислительная техника и информатика. 2013. Т.17, № 2 (55). С. 117-124.