

УДК 004.272.2+004.75+544.18

ТЕХНОЛОГИИ ГРИД В ВЫЧИСЛИТЕЛЬНОЙ ХИМИИ

В. М. Волохов¹, Д. А. Варламов^{1,2}, А. В. Пивушков¹,
Г. А. Покатович¹, Н. Ф. Сурков¹

Рассмотрены основные варианты применения ГРИД-технологий в области вычислительной химии, а также основные достижения авторов в использовании подобных технологий на базе ресурсного ГРИД-сайта Института проблем химической физики РАН. Описаны основные типы задач, использованные технологии, а также основные методы решения задач в распределенных средах, включая разработку прикладных программных интерфейсов различного уровня для запуска в распределенных вычислительных средах, методы формирования и запуска “пучков” независимых задач на распределенных ресурсах, создания пользовательских WWW-интерфейсов, методы динамического формирования среды выполнения и др. Статья подготовлена по материалам доклада авторов на международной конференции “Параллельные вычислительные технологии” (ПаВТ-2010; <http://agora.guru.ru/pavt>).

Ключевые слова: распределенные вычисления, квантовая и вычислительная химия, ГРИД.

1. Введение. Вычислительная химия в своем современном состоянии невозможна без использования сверхмощных параллельных и распределенных вычислительных ресурсов, которые требуются для решения задач самых разных классов. Вычислительная и квантовая химия являются одними из наиболее заинтересованных в ГРИД-вычислениях отраслей науки.

Следующие основные тематические научные направления наиболее заинтересованы в использовании ГРИД-технологий:

- изучение строения вещества;
- строение молекул и структура твердых тел;
- создание материалов с заранее заданными свойствами;
- создание биологически активных веществ и лекарственных препаратов, биотехнологии;
- кинетика и механизм сложных химических реакций;
- химическая физика процессов горения и взрыва;
- газодинамика экстремальных состояний;
- химическая физика процессов образования и модификации полимеров;
- химическая физика биологических процессов и систем;
- предсказательное моделирование наноструктур;
- нанотехнологии;
- общие проблемы химической физики

и многое другое.

Для выполнения крупномасштабных вычислений в области вычислительной и квантовой химии и сопряженных областей науки требуется проведение *высокоинтенсивных* параллельных и распределенных расчетов. Например, некоторые задачи оптимизации молекулярных структур требуют выполнения до 10^9 отдельных расчетов. Подобные расчеты требуют вычислительных ресурсов, которые не может предоставить ни один из доступных вычислительных центров. Для таких задач необходимо развитие и применение ГРИД-технологий в области вычислительной и квантовой химии для организации распределенных вычислений.

Крупномасштабные квантово-химические расчеты — одно из основных научных направлений работы вычислительного центра Института проблем химической физики РАН (ИПХФ РАН) [1, 2]. Эти расчеты выполняются с использованием авторских программ, “open source” пакетов (GAMESS-US, CPMD,

¹ Институт проблем химической физики РАН, просп. акад. Семенова, д. 1, 142432, г. Черноголовка; В. М. Волохов, зав. отделом, e-mail: vvm@pro.icp.ac.ru; А. В. Пивушков, ст. науч. сотр., e-mail: pav@icp.ac.ru; Н. Ф. Сурков, ст. науч. сотр., e-mail: surkov@icp.ac.ru; Г. А. Покатович, зав. сектором, e-mail: pga@icp.ac.ru

² Институт экспериментальной минералогии РАН, ул. Институтская, д. 4, 142432, г. Черноголовка; Д. А. Варламов, ст. науч. сотр., e-mail: dima@iem.ac.ru

Dalton-2, NAMD, ABINIT и др.), а также лицензионных программ (Gaussian-98, Gaussian-03, VASP, Морас2002, MolPro). Институт располагает богатейшей в России библиотекой параллельных квантово-химических и молекулярно-динамических программ. Работы с системами распределенных вычислений в ИПХФ РАН были начаты в 2004–2008 гг. по программам Президиума РАН и федеральным целевым научно-техническим программам и продолжаются в настоящее время в рамках Программы фундаментальных исследований Президиума РАН № 1 на 2009–2011 гг. “Проблемы создания национальной научной распределенной информационно-вычислительной среды на основе развития ГРИД-технологий и современных телекоммуникационных сетей”, а также в рамках программы СКИФ–ГРИД.

Основными задачами авторов стало развитие двух основных направлений:

1) адаптация наиболее востребованного прикладного ПО в области вычислительной (прежде всего квантовой) химии к работе в инфраструктуре ГРИД и обеспечение широкого доступа пользователей к работе с ним с использованием самых различных методов и технологий;

2) развитие ресурсного ГРИД-сайта (для нескольких распределенных сред), выступающего как в роли полигона для проведения вычислительных экспериментов в данной области, так и в роли средства для решения реальных фундаментальных и научно-практических задач.

Выбор данных направлений был обусловлен основной стратегией развития инфраструктуры ГРИД как в России, так и в мире, и позволяет наилучшим образом “приблизить” конечного пользователя (прежде всего — ученого-химика) к широкомасштабному использованию распределенных вычислительных ресурсов и обеспечить возможность решения задач, принципиально трудно разрешимых в настоящее время на единичных вычислительных комплексах.

2. Основные типы и классы задач вычислительной химии. Почему же так важны квантово-химические расчеты и насколько велики могут быть востребованные ими вычислительные ресурсы? Подобные 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.

Наш многолетний опыт проведения подобных расчетов [1–4] позволяет разделить большинство квантово-химических задач на два основных *вычислительных типа*:

— задачи, распадающиеся на совокупность практически независимых заданий, число которых зависит обычно от количества параметров задачи или от “сетки” разбиения искомой области данных;

— задачи, представляющие собой единый вычислительный процесс, как правило, требующий единовременного выделения большого количества ресурсов (количество CPU, оперативная память на ядро, дисковые массивы).

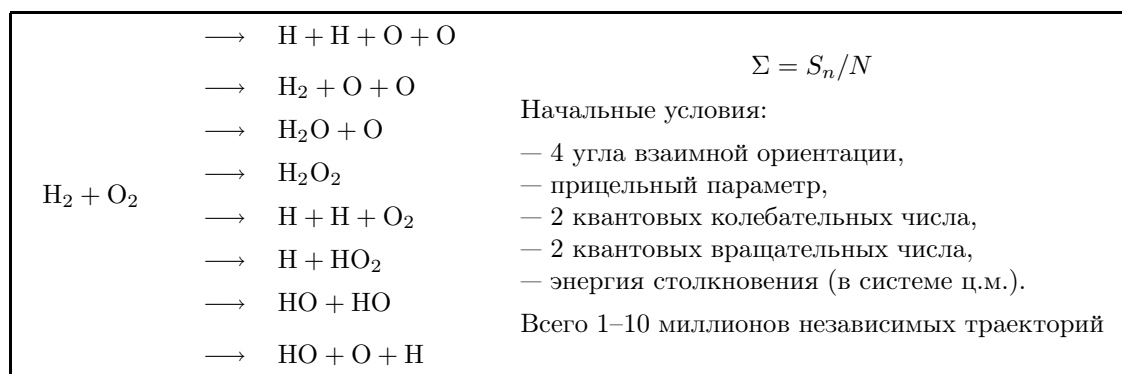


Рис. 1. Исследование траекторных расчетов сечений химических реакций (кислород + водород).

Время расчета одной траектории — от долей минуты до нескольких часов, общее время расчета — до нескольких лет работы единичного CPU)

Задачи первого типа наиболее применимы к работе в ГРИД-средах, поскольку, распределяя независимые задания на множество небольших кластеров (каждый кластер — 10–20 процессоров, задание выполняется на нем как параллельное), можно добиться высокой эффективности использования вычислительных ресурсов. При этом возможно использование весьма больших вычислительных полигонов (до 10³–10⁴ процессоров) как в локальном варианте (в гетерогенных вычислительных средах типа Condor), так и в условиях распределенных сред (совокупность удаленных кластеров — ресурсных узлов). Хорошим примером служат траекторные расчеты химических реакций. Характерной системой для подобных расчетов является реакция H₂ + O₂ (рис. 1). Расчет представляет собой компьютерное моделирование с помо-

пью классических траекторий элементарного акта столкновения. Как правило, расчет одной траектории занимает не более нескольких минут. Для расчета полного сечения подобной реакции для набора необходимой статистики следует разыграть: по два угла взаимной ориентации для каждой молекулы, начальные колебательные и вращательные квантовые числа, параметр столкновения, относительную энергию столкновения. Последовательный перебор указанных параметров приводит к необходимости расчета десятков миллионов траекторий, что приводит к нереальности решения подобных задач на локальных ресурсах и требует перехода к их решению на распределенных полигонах. Аналогичным способом могут решаться многочисленные многопараметрические задачи химии, для которых свойственен перебор многомерных “сеток” входных параметров, что ведет к увеличению числа независимых расчетов до 10^8 – 10^9 .

Задачи второго типа представляют собой существенную проблему, так как эффективность их решения непосредственно связана с эффективностью распараллеливания вычислительного процесса и высокими требованиями к ресурсам узла. Например, 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. Аналогично возможно масштабирование программ типа VASP. Таким образом, для характерных задач исследования наноструктур и молекулярных кристаллов возможно использование до нескольких тысяч процессоров с потреблением процессорного времени порядка месяца, т.е. желательно использование кластеров терафлопного уровня. При этом возможно использование ГРИД-сред для запуска подобных задач на удаленных ресурсах, что с точки зрения пользователя отличается от обычного удаленного запуска заданий (например, через SSH) в сторону упрощения выбора необходимого (и доступного!) вычислительного ресурса.

Характерные примеры задач первого и второго вычислительных классов приведены на рис. 1 и 2.

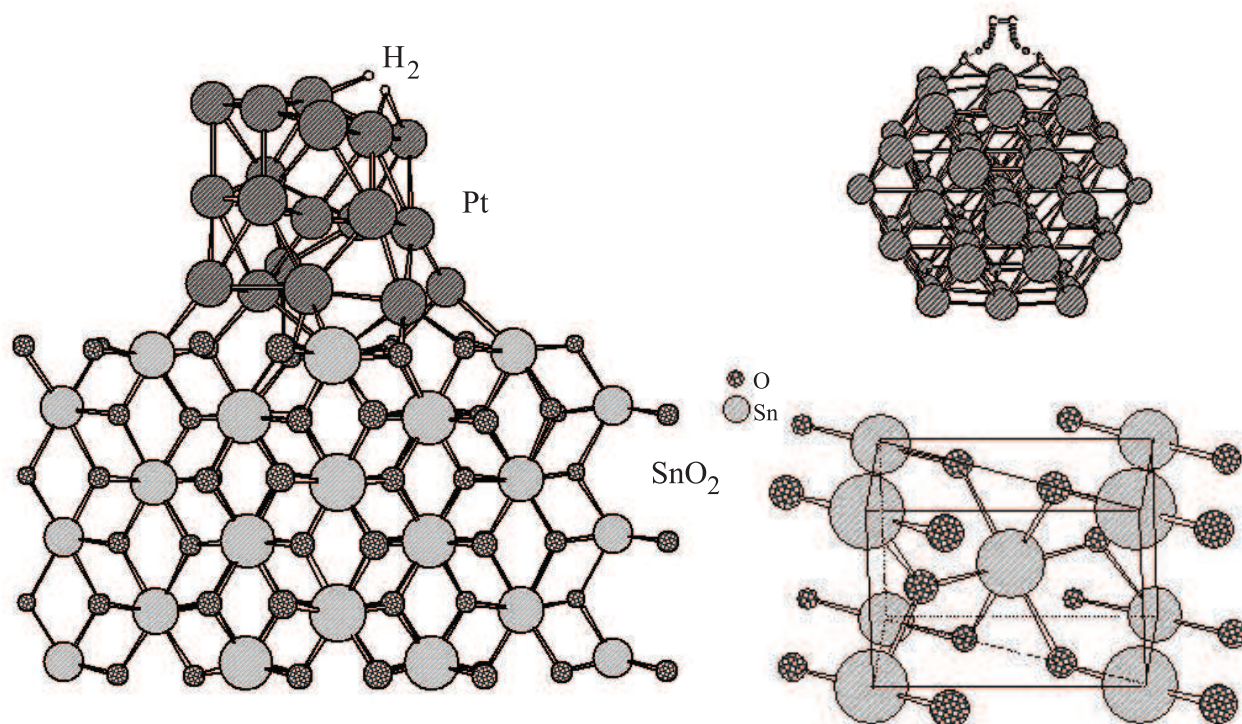


Рис. 2. Адсорбция водорода (H₂) на октаэдрический кластер Pt₁₉, лежащий на поверхности SnO₂. При расчетах на СКИФ-МГУ “Чебышев” (программа VASP) использовано до 200 CPU, затрачено 15 часов, рассчитано всего 10 шагов оптимизации — при необходимых 100–200 шагах

С теоретической же точки зрения большинство задач вычислительной химии можно разбить на два класса.

1. Стационарные и важнейшие из них квантово-химические задачи и задачи на нахождение собственных функций и значений уравнения Шредингера (задача Штурма–Лиувилля).

2. Нестационарные задачи, исследующие временное поведение молекулярных систем или связанные

с использованием траекторных методов расчета сечений реакций.

Первый класс достаточно обычен с точки зрения вычислений и, большей частью, относится ко второму типу вычислений. В среде ГРИД данные задачи могут запускаться как отдельные задания на удаленных узлах в зависимости от требуемых ресурсов (память, количество CPU и др.), а также наличия предустановленных необходимых прикладных пакетов квантово-химического ПО на этих узлах.

Второй же класс куда более интересен с точки зрения применимости ГРИД-технологий. Нестационарные задачи в области теоретической химии и физики, безусловно, являются передним краем современной науки, и все исследования в этой области являются новыми. Как показывают последние конференции по современным методам в области вычислительной химии, все ведущие научные центры, проводящие теоретические исследования химических превращений, сфокусировали свое внимание на проводящихся в ГРИД-среде вычислениях подобного типа. Задачи этого класса представляют собой передний край мировой науки. В настоящее время в крупнейших мировых научных центрах, специализирующихся в теоретической физике и теории химических превращений, ведутся активные работы по разработке вычислительных процедур в распределенных и параллельных вычислительных средах, позволяющих исследовать развитие химического превращения во времени. Кратко физическую суть этих работ можно описать так. Волновая функция химической системы в начальный момент времени представляется в виде разложения по полному набору функций, например по функциям Гаусса. В результате создается пакет, во многом аналогичный волновой функции, строящейся в квантово-химической программе Gaussian. Однако, в отличие от Gaussian, где построение такой волновой функции является конечным результатом, в нестационарном подходе возможно развитие ее во времени вдоль классических траекторий. При этом вдоль траекторий распространяются компоненты пакета с учетом интерференции между ними. Таким образом, строится волновая функция молекулярной системы в любой момент времени. Иными словами, компьютерные вычисления позволяют построить пропагатор, описывающий развитие квантово-химической системы во времени. Основная вычислительная процедура, которая используется для построения пропагатора, — это метод Монте-Карло, идеально подходящий для распределенных ГРИД-систем, так как, опять же, представляет собой большое количество независимых заданий.

3. Ресурсный центр ИПХФ РАН, структура и возможности. Для проведения всех работ с распределенными средами основой стал реализованный ранее на основе вычислительного комплекса ИПХФ РАН ресурсный ГРИД-сайт, включающий узлы распределенных сред gLite и Unicore. Целью создания ресурсных узлов стало (помимо постоянного проведения текущих расчетов) формирование опытного полигона по проведению вычислительных экспериментов в российском ГРИД-сегменте по вычислительной и квантовой химии. Основными задачами в рамках этого направления стали:

- развитие инфраструктуры центра, его вычислительных мощностей и функциональности;
- разработка и использование системы запуска *исходящих* задач (т.е. запускаемых пользователями на удаленных ресурсных узлах) различного типа в распределенных средах;
- обеспечение проведения вычислительных экспериментов и расчета реальных *входящих* задач на собственном ресурсном узле ИПХФ (выступающем в роли удаленного распределенного ресурса и тестового полигона).

Существующий ресурсный сайт ИПХФ был достаточно детально описан авторами ранее [1, 2, 4], поэтому лишь коротко охарактеризуем его здесь. В настоящее время интегрированная вычислительная мощность распределенной (локальной) вычислительной сети в Институте достигла 1.5 TFLOPS (пиковая) с совокупным дисковым пространством до 20 ТВ. Количество процессоров (ядер) основного вычислительного ресурса, кластера на процессорах Intel Xeon (с поддержкой 64-битных приложений), достигло 130, что позволяет апробировать работу программного обеспечения (системного и прикладного) в условиях работы достаточно больших вычислительных систем и предоставить значительную долю ресурсов для распределенных вычислений. Более детально собственно вычислительный комплекс ИПХФ описан здесь: <http://cc-icp.icp.ac.ru>.

В состав ресурсного ГРИД-сайта входят:

- ресурсный узел консорциума EGEE-RDIG (Enable GRID for E-sciencE и Russian Data Intensive GRID (<http://www.egee-rdig.ru>)) на основе распределенной среды gLite (<http://glite.web.cern.ch>), работы ведутся в рамках виртуальной организации (ВО) RGSTEST;
- ресурсный сайт категории “А” для работы в рамках созданного в России крупномасштабного вычислительного полигона СКИФ-Полигон (<http://skif-grid.botik.ru>) на базе промежуточного ПО Unicore (<http://www.unicore.eu>);
- для облегчения доступа пользователей сформирован WWW-портал (<http://grid.icp.ac.ru>, GECР — Grid Enabled Chemical Physics), включающий в себя высокоуровневые пользовательские WWW-интерфей-

сы для работы с рядом задач (GAMESS, многопараметрические задачи) в условиях распределенных сред.

В состав ресурсного сайта входит комплекс интерфейсов различных уровней для взаимодействия прикладного ПО с распределенными ГРИД-средами, позволяющий запускать целый ряд задач вычислительной химии на распределенных вычислительных ресурсах с возможностью формирования, запуска на удаленных ресурсах, мониторинга заданий и сбора результатов и статистики.

Работа в рамках ВО RGSTEST обеспечивает доступ к вычислительным мощностям порядка до 800 процессоров (без учета их возможной многоядерности) и дисковым массивам порядка 8–15 терабайт в нескольких географических зонах (Москва, Протвино, Харьков, Черноголовка и др.). Разнородность узлов данной ВО позволяет достаточно легко варьировать параметры запускаемых задач, ориентируясь на различные типы ресурсов. Использование подобного полигона обеспечивает проведение достаточно масштабных вычислительных экспериментов как научного, так и прикладного характеров.

Созданный ресурсный сайт для среды Unicore позволяет выполнять входящие задачи сертифицированных пользователей СКИФ-Полигона, производить мониторинг задач и передавать полученные результаты пользователям. Обеспечена возможность мониторинга состояния сайта извне. Настроен клиентский интерфейс, проведены успешные запуски исходящих задач через внешний брокер ресурсов на собственном ресурсном сайте ИПХФ в роли удаленного ресурса (<https://unicorgw.icp.ac.ru:8080>) и ресурсных узлах СКИФ-Полигона — ИПС РАН, Cyberia (Томский ГУ, <https://cyberia.tsu.ru>), СевКазГУ (<https://skif-poligon.ncstu.ru>), Нижегородского ГУ (<https://85.143.3.48>) и др., что продемонстрировало работоспособность как ресурсного сайта, так и клиентского интерфейса среды Unicore. Сайт позволяет проводить вычислительные эксперименты на полигонах, поддерживающих среду Unicore, с использованием входящих и исходящих распределенных задач вычислительной химии.

Составная часть ресурсного центра — ГРИД-портал, объединение ГРИД- и web-сервисов. Создан наиболее современный интерфейс, позволяющий более эффективно использовать все преимущества ГРИД, продуктивно работать, экономить время и ресурсы. Это среда, которая позволяет пользователям получить доступ к ГРИД-ресурсам и сервисам, вызывать и настраивать их с помощью web-браузера. Портальный интерфейс также очень важен потому, что с его помощью пользователь даже начальной подготовки и уровня знаний может без особых проблем начать работу в ГРИД. Архитектура ГРИД-портала основана на идее, что порталная система является контейнером для пользовательских интерфейсов (инструментов, клиентов), обеспечивающих работу с ГРИД-службами. Преимущество данной архитектуры в том, что она достаточно легко позволяет встраивать в портал интерфейсы новых ГРИД-служб и изменять существующие. Портальные сервисы контролируют и визуализируют пользовательский интерфейс.

В настоящее время Портал объединяет интерфейсы приложений двух видов.

1. Квантово-химический комплекс GAMESS для теоретического исследования свойств химических систем, *ab initio*³ методы которого могут использовать параллельные вычисления.

2. Вычисление многопараметрических функций, под каждой из которых следует понимать целый класс нераспределенных задач химической физики, обладающих свойством параллелизма по данным (Data Parallel).

Данные интерфейсы позволяют определять входные параметры и условия (включая загрузку данных и конфигурационных файлов), формировать сложные первичные файлы запуска, производить (при условии сертификации пользователя) запуск данного ПО в распределенной среде, осуществлять мониторинг выполнения заданий и сбор результатов. Интегрирована также технология работы через web-интерфейс с “пучками” независимых заданий на “нарезаемых” областях данных. Заметим, что основная часть программного кода web-интерфейсов не связана напрямую с выбранной распределенной средой, поэтому они могут быть подключены к нескольким вариантам таковых сред. Данные интерфейсы значительно снижают трудоемкость работы пользователя в части формирования задач и работы с первичными данными. Они также значительно облегчают работу с пакетами в распределенных средах, особенно для неподготовленного пользователя.

4. Основные варианты использования ГРИД-технологий в химии.

4.1. Адаптация к работе в распределенных средах прикладных пакетов ПО вычислительной химии и авторских программ. Нами проводилась экспериментальная проверка и апробация возможности использования ГРИД-ресурсов для реальных расчетов на стандартных прикладных пакетах программ (в том числе и параллельных), используемых в вычислительной химии, а также различных авторских программ, разработанных в ИПХФ и ИЦЧ РАН. Особый интерес имеет адаптация этих программ для распределенных вычислений на максимуме доступных ресурсов российской и международной

³Ab initio (лат. от начала): в физике — решение задачи из первых основополагающих принципов без привлечения дополнительных эмпирических предположений. Обычно подразумевается прямое решение уравнений квантовой механики.

ГРИД-инфраструктуры.

Для адаптации в распределенных вычислительных средах (см. выше) были выбраны следующие, наиболее востребованные пользователями ИПХФ прикладные программные пакеты:

— GAMESS-US (<http://www.msg.ameslab.gov/GAMESS>) — одна из самых популярных программ для теоретического исследования свойств химических систем, уступает по известности лишь комплексу Gaussian, позволяет рассчитывать энергию, структуры молекул, частоты их колебаний, а также разнообразные свойства молекул в газовой фазе и в растворе, как в основном, так и в возбужденных состояниях. Основное направление — развитие методов расчета сверхбольших молекулярных систем;

— VASP (Vienna University, <http://cms.mpi.univie.ac.at/vasp>) предназначен для моделирования объема и поверхности твердых тел в рамках неэмпирических подходов, основанных на применении функционалов плотности с использованием периодических граничных условий с базисами на плоских волнах. VASP позволяет проводить оптимизацию структуры и выполнять моделирование в рамках молекулярной динамики. Программный комплекс VASP необходим для моделирования процессов на поверхности и в объеме твердых тел (прежде всего катализа и ионной проводимости);

— Gaussian-03 (<http://www.gaussian.com>) — самое популярное средство выполнения квантово-химических расчетов среди основной массы химиков. Основные причины этой популярности — широта охвата реализованных квантово-химических методик, высокая эффективность и удобный интерфейс пользователя. Современные версии комплекса программ Gaussian отличаются расширением спектра поддерживаемых квантово-химических методов и их модификаций. Комплекс программ Gaussian позволяет рассчитывать энергию, структуру молекул, частоты их колебаний, а также разнообразные свойства молекул в газовой фазе и в растворе, как в основном, так и в возбужденных состояниях. Основное направление, в котором развиваются версии, — это развитие методов расчета сверхбольших молекулярных систем. Однако использование пакета в распределенных средах затруднено лицензионными ограничениями;

— Dalton-2 (<http://www.kjemi.uio.no/software/dalton/dalton.htm>) позволяет рассчитывать синглет-синглетные возбуждения, а также электронные структуры, вращательные и колебательные спектры молекул, учитывать релятивистские эффекты и эффект сольватации;

— CPMD (<http://www.cpmid.org>) — это расчеты в области молекулярной динамики;

— NAMD (<http://www.ks.uiuc.edu/Research/namd>, University of Illinois at Urbana-Champaign, Computational Biophysics Group) — хорошо масштабируемая молекулярно-динамическая программа, причем одна из наиболее быстрых при параллельном вычислении на большом числе процессоров. Программа активно используется в ИПХФ РАН для расчетов мицеллы (micelle — коллоидная частица, несущая электрический заряд и объединяющая в себе несколько крупных молекул);

— авторские программы (разработки ИПХФ), включающие многопараметрические задачи из области квантовой химии и молекулярной динамики.

Для всего выбранного ПО был проведен детальный анализ модульной структуры квантово-химического кода и изучены особенности работы различных реализаций однопроцессорных и параллельных версий, определены стратегии реализации выбранных типов квантово-химических вычислений применительно к распределенным средам.

Для большинства выбранных прикладных пакетов созданы и протестированы на реальных задачах низкоуровневые интерфейсы для запуска их в распределенных вычислительных средах (в основном для среды gLite, в меньшей степени — для среды Unicore). Данные интерфейсы включают набор скриптов по формированию исходящих заданий, запуску через брокер ресурсов на удаленных узлах, мониторингу выполнения задач, возвращению полученных результатов с удаленных ресурсов и “сборку” окончательных результатов на интерфейсе пользователя. Реализованы интерфейсы для однопроцессорных и параллельных (SMP, сокетные, MPI-1,2) вариантов указанного ПО. На ресурсном ГРИД-узле ИПХФ, использованном в качестве удаленного распределенного ресурса, проведены запуски указанного прикладного ПО через инфраструктуры BO RGSTEST (EGEE-RDIG) и СКИФ-Полигона. Запуски всего адаптированного ПО проводились в разных режимах и конфигурациях (с разным количеством востребованных процессоров и использованием разных вариантов параллельных расчетов). Были изучены варианты совмещения различных вариантов распараллеливания (например, SMP+MPI) вычислений применительно к некоторым прикладным пакетам (пакеты Dalton-2 и CPMD). После ряда вычислительных экспериментов была проведена коррекция созданных низкоуровневых интерфейсов и окончательная оптимизация их для распределенных сред. Были скорректированы проблемы запуска и работы параллельных (SMP, сокетные, MPI-1,2) вариантов указанного ПО на различных типах ресурсных узлов (разные пакетные системы PBS и параллельные среды).

Следует отметить, что большинство указанных прикладных пакетов вычислительной химии отлича-

ются сложностью конфигураций и повышенными требованиями к среде выполнения, особенно для проведения параллельных расчетов. Обычно эта проблема решается путем создания *виртуальных организаций*, т.е. объединением через распределенные среды во многом однотипных (по установленному программному обеспечению и настройкам) вычислительных ресурсов. Для них выбранные прикладные пакеты (вместе со средствами конфигурирования и настройки) распространяются из единого репозитория (как, например, для прикладных пакетов ЦЕРНа — Atlas, CMC, Alice и т.п.). В большинстве же случаев неподготовленный ресурсный сайт не имеет нужного заранее установленного прикладного ПО или хотя бы даже не сконфигурирован должным образом, поэтому запуск непредустановленных сложных прикладных пакетов обычно для таких ресурсов оканчивается неудачей. Поэтому в общем случае необходима ручная или полуавтоматическая перенастройка ресурсных узлов распределенных сред, включающая установку собственно пакетов, конфигурирование центрального узла и расчетных узлов (настройка переменных окружения, общих NFS-ресурсов, PBS-очереди), установка дополнительных системных библиотек и исполняемых файлов (включая параллельные среды типа Mprich-2). При таком условии возможны запуски пакетов на распределенных узлах.

Для частичного решения данной проблемы авторами был разработан метод создания виртуальных перемещаемых программных “контейнеров”. “Контейнер”, включающий собственно прикладной пакет, набор необходимых системных файлов и библиотек, скрипты по развертыванию и настройке среды исполнения, файлы данных и конфигурационные файлы, доставляется на удаленный ресурсный узел ГРИД-среды стандартными средствами распределенного middleware. Применение таких “контейнеров” позволяет передавать заранее настроенную среду как единое задание, не требующее дополнительного конфигурирования и сложной процедуры установки и настройки, производимых, как правило, вручную администратором кластеров. “Контейнер” по прибытии на ресурсный узел производит развертывание пакета и необходимых системных библиотек, настройку среды исполнения (включая параллельную среду) и запуск задания; по его окончании проводится отправка результатов на пользовательский интерфейс и “очистка” среды исполнения, т.е. приведение ресурса в первоначальное состояние. Так могут быть решены проблемы установки, настройки, несовместимости с операционной системой и другими программами, разрешаются конфликты одинаковых приложений.

4.2. Работа с “пучками” формально независимых заданий. Для решения части задач “первого” вычислительного типа (например, широкого класса многопараметрических задач вычислительной химии) с использованием ГРИД-технологий был создан метод запуска “пучков” независимых заданий для использования всех доступных ресурсов распределенной среды. Как уже говорилось, в области химической физики существует класс задач, требующих перебора большого количества параметров. При этом полная задача разбивается на огромное количество независимых подзадач (каждая определяется группой значений совокупности параметров). Задача автоматизации процесса разбиения полной задачи на фрагменты важна и определяет удобство пользования системой. Типичный пример — фундаментальная задача в теории элементарных химических процессов: туннельные реакции под воздействием электромагнитного излучения. Параметрами являются частота и амплитуда излучения. Задача имеет высокую вычислительную сложность, однако вычисления в каждой точке сетки в ней происходят независимо друг от друга, поэтому оказалось возможным разбить область вычислений на множество непересекающихся подобластей и для каждой из них запускать задачу на различных процессорах.

Была разработана методика запуска задач и получения результатов методом запуска “пучков” заданий на всех доступных ресурсах выбранной распределенной среды. На языке Perl написан комплекс программ для запуска “пучков” заданий и получения результатов счета с использованием пользовательских интерфейсов (UI) сред gLite и Unicore. Для решения многопараметрических задач квантовой химии были разработаны методы формирования “пучков” независимых заданий с варьирующими параметрами — до 10^4 , в перспективе до 10^7 “атомарных” заданий на задачу. Для выбранных областей данных авторскими скриптами производится “нарезка” областей данных, формирование пулов независимых заданий, создание очередей запуска и отправки заданий на брокер ресурсов. После запуска периодически запускаемые (средствами ОС, например по cron) скрипты ведут мониторинг выполнения заданий, контроль таймаутов, перезапуск неудачных заданий и сбор результатов выполненных заданий (с использованием базы данных и таблиц в ней, контролирующих состояние заданий — “ожидание”, “запуск”, “выполнение” и т.д.). По окончании расчетов проводится сборка “атомарных” результатов в единый выходной файл. Для части задач (требующих значительного числа параллельных независимых расчетов) дополнительно созданы авторские механизмы по разбиению областей данных (или расчетов) на большие независимые подсетки или независимые задания, передаче всех их интерфейсам распределенных сред с последующим запуском на параллельных узлах и “сборкой” финальных результатов из множества полученных независимых. Были сформированы и направлены на распределенные ресурсы ВО RGSTEST и СКИФ-Полигона

“пучки” заданий, осуществлен мониторинг их выполнения (с использованием комплекса скриптов и базы данных MySQL), “сборка” результатов с различных ресурсов. Для ВО RGSTEST было задействовано до 400 процессоров на различных ресурсных узлах (Москва, Протвино, Харьков, Черногоровка), для СКИФ-Полигона — доступные CPU узлов, указанных выше (до 120). С использованием данных методов был решен ряд реальных научных задач, включая:

- а) поведение низкотемпературных химических реакций под сильным электромагнитным воздействием;
- б) первичный расчет установок по росту кристаллов;
- в) ряд газодинамических задач.

5. Заключение. Авторами описаны использование некоторых технологий ГРИД-вычислений применительно к приложениям вычислительной химии. Наши работы позволили создать в рамках технологий ГРИД вычислительную среду для проведения крупномасштабных расчетов в области вычислительной химии. Это позволило достигнуть нового уровня расчетов в области вычислительной химии:

— создан комплекс адаптированных к различным ГРИД-средам (gLite, Unicore) прикладных программных пакетов вычислительной химии с интерфейсами различного уровня (от низкоуровневых интерфейсов вплоть до web-портала);

— разработаны новые методики вычислений (методы формирования “пучков” независимых заданий, метод “виртуальных контейнеров” и т.д.) в распределенных и параллельных средах применительно к прикладному ПО вычислительной химии;

— создан ресурсный центр (включающий ресурсные узлы полигонов EGEE-RDIG и СКИФ-Полигона, а также web-портал) для проведения вычислительных экспериментов в этой предметной области, объединяющий как ресурсы для решения входящих заданий в средах gLite и Unicore, так и пользовательские интерфейсы к этим распределенным средам для решения исходящих задач.

В результате выполнения всего проекта создан вычислительный центр, позволяющий проводить масштабные расчеты в области вычислительной химии в распределенных средах на крупномасштабных полигонах (в перспективе до 10^4 CPU на узлах многотерафлопного масштаба). На ряде реальных задач продемонстрирована применимость созданных ресурсов для решения крупномасштабных химических задач на высокопроизводительных вычислительных полигонах. Это позволяет ставить и решать вычислительные задачи фундаментального и прикладного характера в области химических наук, ранее недоступные из-за ограниченности возможностей вычислительных ресурсов. Основные научные области применения — химическая физика, квантовая химия, исследование наноструктур, молекулярная динамика, фармацевтика, разработка топливных элементов и прочие близкие отрасли наук.

СПИСОК ЛИТЕРАТУРЫ

1. Волохов В.М., Варламов Д.А., Пивушков А.В., Сурков Н.Ф., Покатович Г.А. GRID и вычислительная химия // Вычислительные методы и программирование. 2009. **10**, № 1. 78–88.
2. Варламов Д.А., Волохов В.М., Пивушков А.В., Сурков Н.Ф., Покатович Г.А. Распределенные и параллельные вычисления в области химии на ресурсном узле ГРИД ИПХФ РАН // Distributed computing and GRID technologies in science and education. 3rd Int. Conf. Дубна: Изд-во ОИЯИ, 2008. 120–122.
3. Волохов В.М., Варламов Д.А., Пивушков А.В. Крупномасштабные задачи химии на параллельных и распределенных вычислительных полигонах: современное состояние и перспективы // Научный сервис в сети Интернет: решение больших задач. Труды Всероссийской научной конференции. Новороссийск. М.: Изд-во Моск. ун-та, 2008. 210–212.
4. Алдошин С.М., Волохов В.М., Варламов Д.А., Пивушков А.В. Вычислительная химия в среде GRID: параллельные и распределенные вычисления // Труды Второй международной конференции “Суперкомпьютерные системы и их применение SSA’2008”. Минск: ОИПИ НАН Беларуси, 2008. 114–118.

Поступила в редакцию
20.03.2010