

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

## **Создание грид-центра для высокопроизводительных квантово-химических и молекулярно-динамических расчетов на основе гибридных и распределенных технологий вычислений<sup>1</sup>**

Аннотация. В 2013 году в ИПХФ РАН на базе высокопроизводительного (до 15 Tflops) вычислительного кластера начато создание проблемно-ориентированного грид-центра для компетенции квантово-химического и молекулярно-динамического прикладного ПО и проведения высокопроизводительных расчетов с широким использованием гибридных и распределенных технологий вычислений. Центр предназначен для адаптации широкого спектра прикладных пакетов в области химии и смежных наук и проведения расчетов с использованием новейших гибридных и распределенных вычислительных технологий, а также организации упрощенного доступа пользователей через web- и грид-интерфейсы высокого уровня к вычислительным ресурсам.

Создаваемый центр помимо общей компетенции большого количества прикладных пакетов позволит создавать доступные web- и грид-ориентированные вычислительные сервисы, которые смогут решать на новом уровне многочисленные ресурсоемкие задачи в следующих областях: строение молекул и структур твердых тел, кинетика и механизмы сложных химических реакций, химическая физика процессов горения и взрыва, образования и модификации полимеров, биологических процессов и систем, а также в области общих проблем химической физики.

*Ключевые слова и фразы:* ресурсоемкие квантово-химические расчеты, молекулярная динамика, параллельные и гибридные вычисления, распределенные вычисления и грид-среды

---

<sup>1</sup> (Рекомендована к публикации.... Поддержана...!)

## Введение

Вычислительная (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, так и распределенных, для решения задач различных классов. В настоящее время в крупнейших мировых научных центрах, специализирующихся в теоретической физике 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. Для исследования динамики не очень сложной химической реакции с использованием метода классических траекторий зачастую нужно рассчитать до  $10^7-10^9$  независимых траекто-

рий. 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  $F$ .

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 (или выделение фиксированных пулов ресурсов в рамках петафлопсных суперкомпьютеров), специализирующихся исключительно на квантово-химических и молекулярно-динамических расчетах, при этом часто входящих в различные грид-полигоны и образующих достаточно обширные виртуальные организации, объединяющие на разных уровнях ресурсы и пользователей.

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

## 1. Предыстория

Использование систем высокопроизводительных вычислений в области химии велось в ИПХФ на протяжении более чем 50 лет. Работы в области распределенных вычислений в ИПХФ РАН были начаты в 2004 году по программам Президиума РАН, научно-техническим программам Союзного Государства, Федеральным целевым научно-техническим программам и в какой-то мере продолжают в настоящее время (Программа фундаментальных исследований Президиума РАН, программы развития Национальной Нанотехнологической Сети (ГридННС), лоты федеральных целевых программ и т.д.)

Были сформулированы три направления исследований: 1) адаптация прикладного ПО в области вычислительной химии к работе в условиях параллельных сред и различных грид инфраструктурах, включая обеспечение возможностей запуска задач на распределенных ресурсах; 2) развитие ресурсных грид сайтов (для нескольких распределенных сред), выступающих как в роли полигона для проведения вычислительных экспериментов, так и в роли средства для решения реальных задач; 3) создание новых методик вычислений в параллельных и распределенных средах, связанных со спецификой используемого ПО и основанных на внедрении новейших технологий вычислений («гибридные» модели, технологии виртуализации ресурсов и приложений, собственные авторские разработки и т.п.).

На протяжении 2004-2012 годов ИПХФ РАН был полноправным членом ряда распределенных вычислительных полигонов:

- 1) ВО RGSTEST консорциума EGEE-RDIG / EGI-[RU-NGI] (<http://www.egee-rdig.ru>), среда gLite;
- 2) СКИФ-Полигон (<http://skif-grid.botik.ru>), сайт категории "А" (ICSP-SKIF-SITE), среда Unicore;
- 3) ВО Nanochem Национальной Нанотехнологической Сети (ГридННС, <http://www.ngrid.ru>), среда – модифицированный Globus Toolkit 4 и 5;
- 4) Российская грид-сеть для высокопроизводительных вычислений (<https://grid-russia.ru>)

На базе грид-кластера ИПХФ были сформированы и успешно функционировали ресурсные сайты всех вышеуказанных полигонов. Были также настроены полнофункциональные пользовательские интерфейсы для ряда прикладных программных пакетов в области квантовой химии и молекулярной динамики, позволяющие работать с вычислительными мощностями вплоть до нескольких тысяч CPU в условиях различных грид сред. Кроме того, был обеспечен доступ к входящим в состав грид-полигонов суперкомпьютерным установкам типа СКИФ-МГУ («Чебышёв»). Был разработан web портал <http://grid.icp.ac.ru>, объединяющий несколько

высокоуровневых web-интерфейсов для запуска широкого класса задач химической физики и квантово-химических пакетов в различных распределенных средах [2].

Использование этих полигонов позволило как провести достаточно масштабные вычислительные эксперименты научного и прикладного характеров в области вычислительной химии, так и разработать ряд авторских методик проведения расчетов, а также обеспечить решение ряда практических задач на грид-полигонах. Таким образом, в ИПХФ был сформирован прототип предлагаемого к созданию проблемно-ориентированного вычислительного грид-центра.

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 [3].

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; 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 2-х 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 про-

блемно-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. Направления работы центра

В процессе исследований давно стала очевидной потребность в создании высокопроизводительного ресурсного центра, узкоспециализированного на использование 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 прикладных пакетов (свободно-распространяемых и лицензируемых), их жесткая компетенция между собой, определение возможностей работы в качестве грид- и web-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 мультядерных 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 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 web-портала с интегрированными в него высокоуровневыми грид- и web-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 общеупотребительных технологий и методов проведения высокоинтенсивных расчетов (в первую очередь «гибридных» вычислений) совместно с ранее разработанными авторами методами расчетов в грид-средах, результатом чего должно стать создание сбалансированного проблемно-ориентированного ресурса, объединяющего комплекс разработанных методов расчетов, центр компетенции ПО, собственно высокопроизводительный вычислительного ресурса (ориентированного на решение задач квантовой химии и молекулярной динамики) с поддержкой «гибридной» модели вычислений, полноценного грид-сайта и объединяющих их в единое целое высокоуровневых web-интерфейсов.

### **3. Преимущества и проблемы создания центра**

Преимущества создания подобных центров очевидны:

- а) создание пулов узлов, ориентированных именно на «тяжелые» квантово-химические вычисления (большие объемы RAM и HDD на ресурсных узлах, т.п.);
- б) возможность установки и тестирования очень широкого спектра прикладных пакетов в выбранной области;

- в) квалифицированная поддержка пользователями системными специалистами, хорошо знакомыми с особенностями конкретного прикладного ПО, а также ориентирующимися в весьма сложных системах входных данных и многозначности стратегии вычислений;
- г) создание специализированных высокоуровневых интерфейсов к прикладному ПО для работы с локальными и грид-заданиями;
- д) 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 параллельные технологии с использованием технологий программирования CUDA™ и аналогичных. 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 (Nvidia – Tesla и Kepler, AMD – ATI Radeon), что значительно повышает эффективность использования параллельных методов расчетов и снижает

требования к расчетным центрам и операционную стоимость расчетов, особенно при использовании большого числа расчетных узлов. Применение методов параллелизации с использованием GPU устройств (или ускорителей Intel 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, 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 мощных, но доступных вычислительных ресурсах) направлен на разработку и внедрение новейших вычислительных (параллельных, распределенных, «гибридных») технологий с использованием широкого спектра прикладных пакетов, позволит решать на принципиально новом вычислительном уровне многочисленные "тяжелые" задачи в следующих областях: строение молекул и структур твердых тел, кинетика и механизмы сложных химических реакций, химическая физика процессов горения и взрыва, процессов образования и модификации полимеров, биологических процессов и систем, общих проблем химической физики. Специализированный центр позволяет также разрабатывать новые оригинальные методы проведения локальных и распределенных вычислений на основе.

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

### Список литературы

- [1] Суперкомпьютерные технологии в науке, образовании, промышленности, вып.3 (под ред. В.А.Садовниченко // М., изд-во МГУ, 2012, 232с.
- [2] В.М.Волохов, Д.А.Варламов, А.В.Волохов, А.В.Пивушков, Г.А.Покаатович, Н.Ф.Сурков *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.
- [3] 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

*Об авторах:*



#### **Вадим Маркович Волохов**

Зав. отд. вычислительных и информационных ресурсов Ин-та проблем химической физики РАН (ИПХФ). Д-р физ.-мат. наук по хим. физике, физике горения и взрыва (2007). Иссл. в обл. квантовой химии, математических методов в химии, параллельных и распределенных вычислений, грид-технологий, суперкомпьютинга.)

*e-mail:* [vvm@icp.ac.ru](mailto:vvm@icp.ac.ru)

**Варламов Дмитрий Анатольевич**

с.н.с. ИПХФ и Ин-та экспериментальной минералогии РАН, геолог (МГУ, 1987). Параллельные и распределенные вычислительные системы, грид-технологии, web-технологии, виртуализация, минералогия, кристаллография

**Пивушков Александр Викторович**

с.н.с. ИПХФ

Кандидат физ.-мат.наук по химической физике, «физика горения и взрыва» (2006).

Квантовая химия, математические методы в химии, параллельные и распределенные вычислительные системы, грид-технологии, виртуализация, web-технологии

**Волохов Александр Вадимович**

ведущий инженер ИПХФ, управление и информатика технических систем (МГОУ, 2005). Параллельные и распределенные вычислительные системы, грид-технологии, виртуализация ресурсов и приложений, распределенные среды

**Покатович Геннадий Александрович**

ведущий инженер ИПХФ. Диплом: инженер-электромеханик АСУ (МВТУ, 1975) Математические методы в химии, вычислительная химия, суперкомпьютинг, параллельные и распределенные вычисления