

# Квантово-химические прикладные пакеты на Российских гид-полигонах\*

В.М. Волохов<sup>1</sup>, Д.А. Варламов<sup>1,2</sup>, А.В. Волохов<sup>1</sup>,  
А.В. Пивушков<sup>1</sup>, Г.А. Покатович<sup>1</sup>, Н.Ф. Сурков<sup>1</sup>

Институт проблем химической физики РАН<sup>1</sup>,  
Институт экспериментальной минералогии РАН<sup>2</sup>

В статье сделан обзор текущего состояния размещения и использования прикладных пакетов в области квантовой химии и молекулярной динамики на существующих российских гид-полигонах. Охарактеризованы основные компоненты адаптированных прикладных программных пакетов и обслуживающих их гид-сервисов. Описаны основные прикладные пакеты вычислительной химии, адаптированные в настоящее время для работы в качестве вычислительных гид-сервисов. Сформулированы перспективы применения прикладных пакетов с использованием гид-вычислений.

## 1. Введение

Вычислительная химия и сопряженные с ней области знаний являются одними из наиболее заинтересованных в гид-вычислениях (в том числе на входящих в состав гид-полигонов суперкомпьютерах) отраслями науки [1-3]. Исследования, проводимые в области химии и смежных наук, зачастую абсолютно неэффективны без использования сверхмощных параллельных и распределенных вычислительных ресурсов для решения задач самых разных классов. Например, некоторые задачи оптимизации крупных молекулярных структур требуют выполнения до  $10^9$  отдельных расчетов. Подобные расчеты требуют вычислительных ресурсов, которые не может предоставить ни один из вычислительных центров, что приводит к необходимости использования мощностей крупных распределенных гид-полигонов либо суперкомпьютеров высокой производительности.

Данная статья опирается на опыт использования прикладных программных пакетов (далее – «ППП») вычислительной химии в гид-средах, что является одним из основных направлений работы вычислительного центра Института Проблем Химической Физики в Черноголовке (ИПХФ РАН, <http://www.icp.ac.ru>). Адаптация прикладных пакетов проводилась авторами для различных распределенных сред, основанных на middleware gLite, Unicore, Globus Tools, в условиях основных российских гид-полигонов (Национальная Нанотехнологическая Сеть - ГидННС, СКИФ-Полигон, EGEE(EGI)-RDIG).

Институт располагает богатейшей в России библиотекой параллельных квантово-химических и молекулярно-динамических программ (авторских, «open source» и лицензионных), что позволяет проводить обширные эксперименты по адаптации подобных программ к гид-средам в интересах собственных пользователей. В течение года в институте проводится расчет от 3 до 4 тысяч вычислительных задач высокой сложности с публикацией более чем 400 печатных работ с использованием результатов проведенных расчетов. Работы с использованием гид-вычислений в интересах квантовой химии и молекулярной динамики в ИПХФ ведутся с 2004 года, и в настоящее время осуществляются под эгидой нескольких государственных программ (включая Федеральные Целевые Программы, Программы Президиума РАН, гранты РФФИ). ИПХФ является инициатором по использованию квантово-химических пакетов в ряде создаваемых в настоящее время российских гид-полигонов разного масштаба.

Данная статья может представлять интерес как для конечных пользователей-химиков, заинтересованных в интенсификации своих расчетов, так и для администраторов ресурсных гид-сайтов, заинтересованных в увеличении функциональности своих ресурсов. Отметим, что здесь не рассматриваются процедуры адаптации ППП, установки и настройки как клиентских интерфейсов пользователей гид-среды, так и собственно ресурсных сайтов гид-сред (см.[1,3]).

\* Работа выполнена при поддержке гранта РФФИ № 11-07-00686-а

## 2. Основные компоненты адаптированных к грид-средам ППП

### 2.1 Собственно прикладной пакет

Использование стандартных (с точки зрения химика, да и любого конечного пользователя) ППП в грид-средах обуславливается несколькими параметрами:

- востребованность прикладных пакетов научно-техническими работниками, т.е. конечными пользователями (нет смысла проводить весьма трудоемкую адаптацию и последующую настройку грид-среды для малоиспользуемых пакетов);
- лицензионные ограничения (в том числе существующие и для свободно распространяемых пакетов); для коммерческих ППП условия лицензирования могут либо крайне ограничить возможность реализации их в грид-среде, либо напрямую запретить это;
- доступность исходных кодов ППП, что весьма желательно для некоторых модулей пакетов (например, коммуникационных или отвечающих за реализацию параллельных протоколов);
- возможность работы без интерактивного взаимодействия с пользователями и графического вывода информации (что пока маловероятно реализовать в условиях грид-среды).

Как правило, установка прикладных пакетов (из исходных текстов или бинарных дистрибутивов) на ресурсные грид-сайты почти не отличается от таковой для обычных локальных вычислительных кластеров. Некоторые особенности связаны с настройкой управляющих очередей PBS (или аналогичных систем), Как правило, основные проблемы связаны с ограничениями прав псевдопользователей («mapped users») по сравнению с локальными пользователями, ограничениями на время исполнения программ (что весьма актуально для ППП вычислительной химии, где время выполнения может достигать месяцев) и ограничениями в ресурсах. Также, при наличии выбора, рекомендуется вместо сокетных параллельных версий ППП использовать MPI или OpenMP версии, поскольку они гораздо лучше отвечают требованиям грид-сред в части детального мониторинга ресурсов и управления заданиями.

### 2.2 Настройка грид-шлюза для обслуживания ППП

Грид-шлюз ресурсного сайта предназначен для информирования брокера ресурсов о подчиненных ресурсах, поддерживаемых виртуальных организациях (ВО) и входящих в их состав ППП, обработки входящих грид-заданий и передачи их вычислительным элементам, мониторинга подчиненных ресурсов и выполнения грид-задач и многих других функций, описание которых не входит в задачи данной публикации. Для конкретного ППП в зависимости от используемой грид-среды вносится информация о доступности ППП различным Виртуальным Организациям (ВО), ACL листах, размещении исполняемых файлов ППП, допустимых лимитах использования ресурсов сайта.

### 2.3 Низкоуровневые клиентские интерфейсы ППП

Низкоуровневые клиентские интерфейсы (или интерфейсы командной строки – ИКС) включают набор shell-скриптов по формированию исходящих заданий, запуску через грид-инфраструктуру на удаленных ресурсных узлах (с возможностью выбора ресурсов), мониторингу выполнения задач, возвращению полученных результатов и «сборку» окончательных результатов на интерфейсе пользователя. Как правило, для ППП включают в свой состав модернизированные скрипты локальных запусков, а также установки необходимых переменных окружения на удаленных ресурсах.

### 2.4 Высокоуровневые web-интерфейсы (ПОИ, ВИГ)

Высокоуровневые web-интерфейсы (ПОИ – «Проблемно-Ориентированный Интерфейс», ВИГ – «Веб-интерфейс ГридННС») к адаптированным ППП позволяют более эффективно использовать все преимущества грид-расчетов, особенно для неподготовленного пользователя. Эта среда позволяет пользователям получить доступ к грид-ресурсам и сервисам, вызывать и настраивать их с помощью web-браузера. Web-интерфейсы являются «контейнером» для низ-

коуровневых пользовательских интерфейсов, обеспечивающих полнофункциональную работу с грид-службами. ПОИ предоставляют средства работы с файлами данных и конфигурационными файлами ППП, обеспечивая их загрузку, хранение, редактирование, удаление. Как базовые компоненты ПОИ содержит средства формирования грид-задания (с возможностью выбора доступного (для выбранной ВО) ресурса грид-полигона или использования произвольного ресурса), средства запуска сформированного задания «прозрачно» для пользователя через брокер/менеджер ресурсов грид-полигона с учетом требований безопасности и контроля передачи информации, принятым в рамках грид-полигона. Web-интерфейсы предоставляют возможность постоянного мониторинга выполнения задания (включая его останов и перезапуск) и получения конечных результатов выполнения задачи, лог-файлов и (при необходимости) прочих файлов на портал, хранение данных файлов и доставку их по запросу на компьютер пользователя.

В ряде случаев для ППП могут быть созданы дополнительные модули по интерактивному формированию пользователем сложных конфигурационных файлов и файлов данных для запуска адаптированных ППП в грид-среде. Модули на основе многоуровневых меню и системы «деревьев» выбора обеспечивают корректный выбор из большого массива альтернатив методов вычислений и различных параметров расчетов, ввод и коррекцию численных данных, тестовую проверку файлов конфигурации и данных перед запуском задания. Они могут также включать возможность работы пользователя с использованием типовых шаблонов задач вычислительной химии (с минимумом изменяемых входных параметров) для решения стандартизированных задач предметной области. В ИПХФ РАН пилотный web-интерфейс с такими возможностями реализован для пакета GAMESS, что позволяет создавать конфигурационные файлы ППП, содержащие описания нескольких сотен параметров и методов вычислений.

### **3. Квантово-химические и молекулярно-динамические пакеты, адаптированные для работы в составе российских грид-полигонов**

Ниже приведены краткие описания функциональности прикладных пакетов вычислительной химии, адаптированных для проведения вычислений в рамках Национальной Нанотехнологической Сети (<http://www.ngrid.ru>), в том числе на ресурсном сайте ИПХФ РАН (<http://grid.icp.ac.ru> и субдомены <http://nanogrid.icp.ac.ru> и <https://webgrid.icp.ac.ru>). Все указанные ППП в разной степени (в зависимости от лицензий) доступны в рамках ВО «NanoChem», а также ВО, относящимся к конкретным ППП (Gamess, AbInit и т.п.), через портал ГридННС <https://ui.ngrid.ru>. Отметим, что часть ППП была также ранее реализована авторами на пространстве грид-полигонов EGEE(EGI)-RDIG (<http://www.egee-rdig.ru>) и СКИФ-Полигона (<http://skif-grid.botik.ru>). Как правило, адаптация пакетов для разных грид-сред различается особенностями реализации низкоуровневых интерфейсов между middleware и ППП. Большинство нижеперечисленных пакетов установлены на кластерах ресурсных центров, входящих в ГридННС (в том числе ИПХФ) и доступны в качестве вычислительных грид-сервисов, для части сделаны тестовые инсталляции (в основном из-за проблем с лицензиями). Для пакетов после установки и тестирования на кластерах настроены шлюзы приема входящих грид-заданий и работы с GridFTP ресурсами, установлены высокоуровневые интерфейсы (различной степени сложности) и проведено массовое тестирование (включая стресс-тесты) на различных задачах. ИПХФ выступал в качестве установщика ряда адаптированных пакетов (Gamess, Gaussian), предоставлял ресурсы своего грид-центра и производил тестирование созданных грид-сервисов через низкоуровневые и высокоуровневые web-интерфейсы грид-среды.

- **GAMESS-US** (<http://www.msg.chem.iastate.edu/GAMESS>, General Atomic and Molecular Electronic Structure System) – свободно распространяемый прикладной программный пакет, позволяет рассчитывать энергию, геометрию и 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.
- **Gaussian** (<http://www.gaussian.com>) предназначен для расчета структуры и свойств молекулярных систем разной размерности. Выделяется широким спектром реализованных квантово-химических методов расчетов, высокой эффективностью и удобным интерфейсом поль-

зователя (всего реализовано более 80 методов расчетов). 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. Молекулярные модели, рассчитанные в Gaussian, могут быть применены как к стабильным соединениям, так и к тем, которые трудно или невозможно наблюдать экспериментально. Gaussian может использоваться в SMP и параллельных вариантах расчетов (с использованием пакета TCP Linda). Пакет Gaussian существенно ограничен лицензионными правами, однако, допускает получение лицензии на использование в грид-средах.

- **AbInit** (<http://www.abinit.org>) используется для решения ряда задач квантово-механических расчетов с высокой вычислительной сложностью и относительной независимостью параметров от результатов расчетов предыдущих задач. Поэтому, помимо встроенной поддержки параллельных вычислений (на базе MPI), возможна параллелизация выполняемых задач по данным. ППП обеспечивает решение теоретических задач – вычисление полной энергии, плотности заряда и электронной структуры атомных систем (молекулы и периодические твердые тела). Вычисления в ППП AbInit основаны на применении теории функционала электронной плотности, псевдопотенциалов, базиса из присоединенных и расширенных плоских волн. ППП Abinit позволяет использовать теорию возмущения многих тел (GW-приближение) и теорию функционала, зависящую от времени. Он обеспечивает режим оптимизации атомной геометрии под действием межатомных DFT-сил и внешнего давления, а также позволяет проводить моделирование молекулярной динамики, используя эти силы, с определением колебательно-фононных, диэлектрических, механических и термодинамических свойств твердых тел. AbInit позволяет оптимизировать геометрию исследуемой системы, минимизируя силы или напряжения, проводить молекулярно-динамическое моделирование, вычислять распределение электронной плотности, определять динамическую матрицу, эффективный заряд и многое другое. Распространяется на основе GPL.
- **GROMACS** (<http://www.gromacs.org>, GROningen MACHine for Chemical Simulations), пакет молекулярной динамики для моделирования физико-химических процессов, в том числе - динамики крупных молекулярных систем ( $10^3$ - $10^6$  частиц). Представляет собой набор программ, предназначенных для расчета траекторий движения отдельных частей молекулы, аппроксимированной механической системой физических материальных точек, связанных набором сил. Пакет предназначен в основном для моделирования крупных молекул, в том числе биомолекул (белки и липиды), имеющих много связанных взаимодействий между атомами. GROMACS является пакетом прикладного программного обеспечения для расчетов как классической молекулярной динамики различных систем, так и с использованием сторонних квантово-механических пакетов таких как Gaussian, Morac, GAMESS-UK, ORCA, гибридных расчетов (QM/MM). GROMACS обладает высоким уровнем параллелизации, рассчитанным на использование на высокопроизводительных кластерах и суперкомпьютерах с разделяемой памятью. Имеет две реализации параллельных алгоритмов: с использованием MPI-2 и с использованием POSIX threads/NPTL. Из дополнительных возможностей – поддержка выполнения части вычислений на GPU. GROMACS является программным обеспечением с открытым исходным кодом, выпущенным под лицензией GPL.
- **OpenMX** (<http://www.openmx-square.org>, Open source package for Material eXplorer) – квантово-механический программный пакет для моделирования наноструктур, основанный на использовании метода DFT (Density Functional Theories, теория функционала плотности), с использованием псевдопотенциалов для известных типов атомов. Данная особенность позволяет пакету хорошо масштабироваться при увеличении размерности системы примерно как  $O(N)$  и  $O(N \log N)$ . Также ППП позволяет выполнять квантово-механические расчеты молекулярной динамики структур с использованием DFT. OpenMX может использовать несколько различных схем параллелизации для различных типов вычислительных систем: стандартный MPI v2.0, OpenMP и гибридное использование MPI+OpenMP, что позволяет использовать OpenMX на широком классе параллельных вычислительных систем. OpenMX является мощным средством исследования наноматериалов широкого спектра, таких как биоматериалы, углеродные нанотрубки, магнитные материалы и нанопроводники. Распространение пакета и его исходных кодов соответствует лицензии GPLv2.
- **LAMMPS** (<http://www.lammps.sandia.gov>, Large-scale Atomic/Molecular Massively Parallel

Simulator) предназначен для моделирования физико-химических процессов с использованием уравнений классической молекулярной динамики (уравнений Ньютона). Атомы в молекулах рассматриваются как материальные точки, взаимодействующие посредством потенциальных полей. Данный пакет распространяется в виде исходного кода по лицензии GPL и создан специально для работы на высокопроизводительных параллельных системах. Он использует MPI для работы на системах с распределенной памятью, написан на языке C++. LAMMPS является многофункциональным продуктом, который позволяет моделировать физические процессы и химические реакции, происходящие в самых разных системах, таких как атомные системы, жидкости, кристаллы металлов и полупроводников, полимеры, белки, ДНК, гранулярные материалы, эллипсоидальные частицы, точечные диполи, крупнозернистые мезомасштабные модели, а также комбинации всего вышеперечисленного.

- **VASP** (<http://cms.mpi.univie.ac.at/vasp>) обеспечивает вычисление полной энергии, плотности заряда и электронной структуры атомных систем (молекулы и периодические твердые тела), оптимизацию атомной геометрии в статических условиях и под внешним давлением, а также позволяет проводить моделирование методом молекулярной динамики с определением важных физических свойств твердых тел (колебательно-фононных, диэлектрических, механических, термодинамических). Предназначен для моделирования процессов в объеме и на поверхности твердых тел (прежде всего катализа и ионной проводимости) в рамках неэмпирических подходов, основанных на применении функционалов плотности с использованием периодических граничных условий с базами на плоских волнах. Подход, реализованный в программе VASP, основан на приближении локальной плотности (при конечных температурах), при этом свободная энергия считается вариационным параметром, и на каждом шаге выполняется точная оценка мгновенного электронного основного состояния. В программе VASP значительно улучшены процедуры сходимости процессов ССП и оптимизации. Имеется процедура статистического «размывания» краев запрещенной зоны и оптимизации спинового состояния для моделирования металлов и узкозонных полупроводников. Распространяется на коммерческой основе, подразумевает обязательное лицензирование, для определенных типов лицензий допускается использование в качестве грид-сервиса.
- **PWscf** (<http://www.pwscf.org>, <http://www.quantum-espresso.org>, Plane-Wave Self-Consistent Field) построен на базе теории функционала электронной плотности (DFT) и методе псевдопотенциала (PAW-метод). Представляет собой мощный инструмент для энергетических расчетов многоэлектронных систем и предназначен для моделирования на квантово-механическом уровне малых кластеров с числом атомов 10-100, определяющих существование возможных в материале фаз. Во многом является аналогом коммерческого квантово-химического ППП VASP. Описание моделируемого объекта строится на языке волновых функций и заданного гамильтониана системы. Целевыми функциями являются электронный энергетический спектр, собственные функции и плотность состояний изолированного кластера при фиксированном положении ядер, потенциальная энергия системы с учетом электронно-ядерных подсистем. С помощью ППП PWscf можно прогнозировать плотности электронных состояний произвольных кристаллических материалов и их свойств, исходя из основ квантовой теории строения вещества. Расчеты с применением ППП PWscf оптимизированы для использования сотен процессоров, поэтому большой интерес вызывает возможность использования этих пакетов на большом числе процессоров в грид-системах (на основе MPI). ППП PWscf распространяется в исходных текстах по лицензии GNU GPL, что позволяет свободно использовать его в качестве грид-сервиса.
- **NWChem** (<http://www.nwchem-sw.org>, New Wave Chemistry) – используется для проведения смешанного квантово-механического и молекулярно-динамического моделирования, когда определенные локальные области материала моделируются на квантово-механическом, а другие – на молекулярно-динамическом уровнях. Позволяет проводить расчеты геометрии молекулярных структур, расстояний между атомами, сил взаимодействия, свободных энергий поверхностей и т.п. Варианты применения NWChem фокусируются на предоставлении возможностей научных расчетов в области кинетики и динамики химических превращений, химических взаимодействий на границах фаз и в конденсированных (твердых и жидких) фазах. Расчеты с применением ППП NWChem оптимизированы для параллельного использования сотен и тысяч процессоров (на основе MPI). ППП NWChem

свободно распространяется под собственной лицензией в виде бинарных кодов и исходных текстов, что облегчает использование его в качестве грид-сервиса.

- **NAMD** (<http://www.ks.uiuc.edu/Research/namd>, NAnoscale Molecular Dynamics) – масштабируемая программа для молекулярной динамики, написанная с использованием модели параллельного программирования Charm++, обладает высокой эффективностью распараллеливания. Используется для моделирования больших и сверхбольших структур и кластеров с числом атомов  $10^3$ - $10^6$  (вплоть до расчета ДНК структур) вблизи состояния равновесия, и в рамках задач, относящихся к неравновесной молекулярной динамике. Моделирование молекулярных систем может производиться в различных приближениях в зависимости от доступных вычислительных ресурсов и сложности рассчитываемой системы. К числу таких приближений относятся: моделирование молекул как твердых тел, моделирование внутримолекулярных потенциалов взаимодействия как гармонических, моделирование только локального кулоновского поля в периодических граничных условиях, или учет дальнедействующей составляющей. Может быть использована для выполнения расчета траекторий движения атомов заданной системы за счет интегрирования уравнений движения с использованием эмпирических потенциалов взаимодействия. Программа активно используется для расчетов мицелл (micelle – коллоидная частица, несущая электрический заряд и объединяющая в себе несколько крупных молекул) и подобных органических и неорганических молекулярных структур. ППП NAMD является ПО с открытым исходным кодом и свободно распространяется под собственной лицензией в виде бинарных кодов или исходных текстов, что позволяет использовать его на ресурсных грид-сайтах.
- **Firefly** (<http://classic.chem.msu.su/gran/firefly>, ранее известен как PC GAMESS) - одна из самых популярных и высокопроизводительных программ для теоретического исследования свойств химических систем, позволяет рассчитывать энергию, геометрию и 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 – расчет больших и сверхбольших молекулярных систем. Достоинством пакета является широкомасштабный охват основных вычислительных квантово-химических алгоритмов и реализация для большого количества процессорных архитектур и параллельных сред. Основные возможности программы Firefly близки таковым для ППП GAMESS-US.

#### 4. Заключение

Применение адаптированных к грид-средам прикладных программных пакетов в области квантовой химии и молекулярной динамики позволяет ставить и решать вычислительные задачи фундаментального и прикладного характера в области химических наук, ранее не доступные из-за ограниченности возможностей вычислительных ресурсов. Основные научные области применения – химическая физика, квантовая химия, исследование наноструктур, молекулярная динамика, биохимия, фармацевтика, разработка топливных элементов и близкие отрасли наук.

#### Литература

1. В.М.Волохов, Д.А.Варламов, А.В.Пивушков, Н.Ф.Сурков, Г.А.Покатович GRID и вычислительная химия // «Вычислительные методы и программирование», М.: МГУ, 2009, т.10, № 1, с.224-235
2. В.М.Волохов, Д.А.Варламов, А.В.Пивушков, А.В.Волохов «Применение GRID технологий в области вычислительной химии» // «Известия Академии наук. Серия химическая», 2011, № 7, с.1483-1490
3. В.М.Волохов, Д.А.Варламов, А.В.Волохов, А.В.Пивушков, Г.А.Покатович, 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.