В.М. Волохов, Д.А. Варламов,
А.В. Волохов, А.И. Прохоров

**Веб-инструментарий для проведения
массовых вычислительных
экспериментов в области
суперкомпьютерного
квантово-химического
моделирования нанокompозитных
материалов**

Рекомендуемая форма библиографической ссылки

Волохов В.М., Варламов Д.А., Волохов А.В., Прохоров А.И. Веб-инструментарий для проведения массовых вычислительных экспериментов в области суперкомпьютерного квантово-химического моделирования нанокompозитных материалов // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18-23 сентября 2017 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2017. — С. 98-105. — URL: <http://keldysh.ru/abrau/2017/66.pdf> doi:[10.20948/abrau-2017-66](https://doi.org/10.20948/abrau-2017-66)

Размещена также [презентация к докладу](#)

Веб-инструментарий для проведения массовых вычислительных экспериментов в области суперкомпьютерного квантово-химического моделирования нанокompозитных материалов

В.М. Волохов¹, Д.А. Варламов^{1,2}, А.В. Волохов¹, А.И. Прохоров¹

1 Институт проблем химической физики РАН, г. Черноголовка, РФ

2 Институт экспериментальной минералогии РАН, г. Черноголовка, РФ

Аннотация. В ИПХФ РАН на основе высокопроизводительных ресурсов и комплекса программных и инфраструктурных средств создана и развивается веб-ориентированная вычислительная среда, предназначенная для проведения массовых вычислительных экспериментов в области квантово-химического и молекулярно-динамического моделирования структур и свойств нанокompозитов и процессов в них в интересах широкого круга исследователей. Высокоуровневые проблемно-ориентированные веб-интерфейсы к вычислительным и инфраструктурным сервисам позволяют разрешить многие проблемы проведения массовых ресурсоемких вычислительных экспериментов в области моделирования нанокompозитов и предоставить облегченный доступ к суперкомпьютерным ресурсам.

Ключевые слова: высокопроизводительные вычисления, нанокompозитные материалы, вычислительная среда, веб-интерфейсы, компьютерное моделирование

Опыт работы авторов над проектами в области вычислительной химии при моделировании нанокompозитов и наноструктурированных объектов [1-6], показал, что для проведения массовых вычислительных экспериментов помимо использования суперкомпьютерных центров класса «Ломоносов-1,2» или МВС-10000 весьма перспективным является создание достаточно узкоспециализированных проблемно-ориентированных ресурсов среднего размера. Подобные ресурсы направлены на решение определенных классов задач, и при условии оснащения внешними интуитивными дружелюбными (по отношению к конечному пользователю, часто малоискушенному в суперкомпьютерных расчетах) веб-интерфейсами позволяют во многом решить многочисленные проблемы суперкомпьютерного моделирования:

- 1) объединить в рамках единого пользовательского интерфейса все стадии моделирования – от формирования (загрузка + редакция) массивов исходных данных и выбора методов расчетов до получения интерактивных компьютерных моделей и формирования баз данных результатов моделирования;

- 2) предоставить доступ к средствам моделирования как можно более широкому кругу пользователей через организацию интуитивно понятных и дружелюбных пользователю веб-сервисов;
- 3) вовлекать в расчеты большинство доступных вычислительных ресурсов – локальные кластеры, «гибридные» установки, ресурсные сайты распределенных сред (опционально);
- 4) по возможности оказывать интеллектуальную поддержку пользователю при выборе стратегии моделирования и использования методов расчетов;
- 5) соединять в рамках единого комплекса прикладные пакеты, ориентированные на разные аспекты моделирования (нано)объектов – создание своего рода «вычислительных последовательностей», при этом выбирая наиболее приемлемые варианты установки данных пакетов и наиболее подходящие пулы ресурсов для них.
- 6) давать возможность пользователю выбирать формы представления полученных результатов – от наборов «сырых» результирующих и лог-файлов размером в сотни мегабайт до интерактивных 3-D образов моделируемых структур.

При этом может быть также во многом решен целый ряд вопросов, специфичных для квантово-химических расчетов и молекулярного моделирования и обычно существенно ограничивающих эффективность работы исследователей:

- подавляющее большинство прикладных пакетов квантовой химии и молекулярной динамики в качестве исходных данных требует введения нескольких файлов сложного синтаксиса, содержащих описания до нескольких сот параметров (в том числе весьма неявно связанных между собой) и методов расчетов, многих сотен и даже тысяч первичных координат атомов, полевых потенциалов, определенной иерархии используемых констант, данных и вычислительных процедур, многовариантности построения моделей и т.п. Их подготовка трудоемка, требует детального знания синтаксиса входных файлов, тщательной проверки введенных данных и практически затруднена для пользователя (часто ошибки несогласованности данных обнаруживаются на этапе непосредственно выполнения расчетов, при уже большом затраченном времени счета);
- зачастую требуется выполнять многочисленные расчеты задач, отличающихся друг от друга лишь небольшими изменениями параметров («шаги» по температуре, сдвиги координат отдельных атомов, замена сортов атомов и т.п.). На основе хранимых данных весьма желательна возможность создавать массовые «шаблоны» для последующих проектов;
- для построения и хранения мультипараметрических моделей требуется создание и поддержка достаточно объемных структурированных баз данных и хранилищ «сырой» расчетной информации;
- большие временные затраты на расчет моделей (например, до 40-50 часов на расчетный шаг на СК «Ломоносов» при необходимых 150-200 шагах для

расчета транспорта единичного Li-иона через интерфейс «Si-C электрод – полимерный электролит» - до $n \cdot 10^5 - 1 \cdot 10^6$ процессоро-часов, т.е. месяцы расчетов на средних пулах) требуют применения на ресурсах новейших методов проведения расчетов, основанных на использовании «гибридных» и распределенных моделей вычислений;

- прикладные пакеты квантовой химии в результате вычислений создают огромные (вплоть до многих гигабайтов) «сырые» (raw) файлы результатов, требующие постпроцессинга данных, генерализации, упрощения и перевода в другие форматы для наглядной интерактивной визуализации и сопряжения с другими прикладными пакетами;
- как правило, на основе результатов расчетов требуется построение интерактивных 3-D моделей молекулярных структур, позволяющих проводить различные манипуляции с ними и получать наиболее эффективное представление о данных структурах. Кроме того, требуется формировать «на лету» большое количество текстово-табличной и графической информации (бинарные и более сложные диаграммы), позволяющей адекватно оценивать качество созданных моделей в зависимости от входных параметров.

Большинство этих проблем плохо решаемы непосредственно в зоне ответственности суперкомпьютерных центров и даже вычислительных центров конкретных организаций ввиду «всеядности» последних и незаинтересованности их персонала, поэтому очевидна необходимость построения «внешних» проблемно-ориентированных пользовательских «шлюзов» в форме веб-ориентированных высокоуровневых интерфейсов.

В области собственно химии (исключая биохимию и фармацевтику) сейчас почти отсутствуют комплексные решения, позволяющие снять эти проблемы, поэтому возникла задача разработки современного комплексного компьютерного инструментария для упрощения работы с квантово-химическими и молекулярно-динамическими прикладными пакетами. Это по нашим расчетам позволит значительно интенсифицировать проведение вычислительных экспериментов и во многом улучшить ситуацию в области молекулярного моделирования нанокompозитных материалов и нанообъектов на их основе.

Построение вычислительной среды

В последние годы тенденция на создание проблемно-ориентированных вычислительных RIA-сервисов с дистанционным доступом пользователей является одним из основных трендов развития прикладных вычислений в разных областях науки и должна стать эффективным обязательным научным инструментом для российских исследователей. Развитие же программных и инфраструктурных средств для создания высокоэффективных веб- и грид-сервисов в настоящее время позволяет создавать инструменты, контролирующие все стадии процесса моделирования: от ввода первичных данных до интерактивной 3-D визуализации объектов или процессов.

Подобные программно-аппаратные комплексы позволят решать на принципиально новом вычислительном уровне многочисленные «тяжелые» задачи в следующих областях: строение молекул и структур твердых тел, кинетика и механизмы сложных химических реакций, электрохимические превращения, описание процессов образования и модификации полимеров, биологических процессов и систем, а также в области общих проблем химической физики.

Опыт участия авторов в течение 2004-2013 годов в ряде российских грид-проектов [1, 2], как-то: EGEE-RDIG и EGI-RU-NGI, СКИФ-Грид, ГридННС («Национальная нанотехнологическая Сеть»), «Развитие пилотной зоны российской грид-системы для высокопроизводительных вычислений...», различных программ Президиума РАН, ориентированных в том числе на создание проблемно-ориентированных прикладных веб-сервисов в грид-вычислениях, позволил сформировать достаточно оптимальную структуру [7] прикладной компьютерной среды (программно-аппаратного комплекса), направленной (в нашем случае) на эффективное проведение массовых вычислений в области моделирования наноматериалов через веб-сервисы, обеспечивающие дистанционный доступ (рис.1)

Рис. 1. Структура среды компьютерного моделирования нанокompозитных материалов на базе высокопроизводительных вычислительных ресурсов и веб-ориентированных вычислительных сервисов (интерфейсов) к ним [7]

После разработки структуры и интеграции начальных модулей происходило постепенное наращивание функциональности комплекса за счет

ввода отдельных модулей и улучшения пользовательских характеристик интерфейсов.

Подобный комплекс способен работать как в «локальном» режиме (с вовлечением локальных ресурсов – кластеров, «гибридных» GPU- или мультитядерных установок, многоядерных высокопроизводительных станций), так (опционально) и с удаленными ресурсами (при условии соблюдения достаточных мер безопасности) и (при установке требуемого грид-полигоном промежуточного Grid-middleware – шлюзы, ресурсные грид-сайты, грид-хранилища) в составе грид-полигонов. При этом доступ к ресурсам и проведение расчетов осуществляется через высокоуровневые, дружелюбные конечному пользователю веб-интерфейсы. Кроме того, для администрирования комплексом предусмотрен ряд управляющих интерфейсов. Среда объединяет следующие основные компоненты: (а) вычислительные сервисы, созданные на высокопроизводительных ресурсах на основе прикладных квантово-химических и молекулярно-динамических пакетов (пакеты VASP, CPMD), в том числе с применением «гибридных» технологий расчетов на основе GPU- и мультитядерных ускорителей, и низкоуровневых средств доступа к ним от HTTPD сервера; (б) интернет-ориентированный инструментарий для пользовательского доступа к этим сервисам в виде комплекса высокоуровневых веб-интерфейсов для формирования и запуска задач на базе Apache HTTPD-сервера; (в) средства размещения, хранения и обработки первичных данных и полученных результатов (SQL сервер баз данных MariaDB, сетевые хранилища; (г) средства визуализации смоделированных структур и процессов и обработки полученных данных (на основе пакета JMol).

Связь между модулями осуществляется по SSHv2 протоколу (+SFTP), обращения к модулям извне – по защищенному https протоколу (используются сгенерированные собственные ключи). На настоящем этапе использование грид-протоколов не поддерживается, хотя в ранних версиях поддерживался Globus и модифицированные среды на его основе типа GridNNN.

В состав высокоуровневых веб-сервисов помимо стандартных приемов дистанционной работы с прикладными пакетами вводятся упрощенные для пользователя средства по формированию (загрузка и редакция) и тестированию сложных конфигурационных файлов и массивов входящих данных, инструменты анализа и предварительной обработки полученных результатов. Также доступны средства запуска и мониторинга выполняемых заданий (включая упрощенный мониторинг ресурсов).

В настоящее время в ИПХФ РАН реализован работоспособный вариант подобной системы, основанный на локальных ресурсах (кластер на 176 двухпроцессорных узлов HP Proliant производительностью до 15 Тфлопс и 2 «гибридные» установки на базе GPU Tesla C2075 – 2,6 Тфлопс), с поддержкой прикладных пакетов VASP (Vienna Ab initio Simulation Package, <https://www.vasp.at>) и CPMD (Car-Parrinello Molecular Dynamics, <http://cpmd.org>). Система многоуровневых веб-интерфейсов позволяет

авторизовать пользователя комплекса, проводить мониторинг доступных ресурсов, подготовить проект вычислительного задания (включая создание и редакцию начальных данных и конфигурационных файлов, в том числе на основе имеющихся шаблонов задач и предыдущих проектах); запустить прикладной пакет в инфраструктуре комплекса на выбранном ресурсе; вести мониторинг выполнения задания (включая останов и перезапуск); по завершении – получить результаты счета, провести их пост-обработку (генерализацию, архивацию, конвертацию в другие форматы) и отправку в хранилище, а также обеспечить визуализацию смоделированных объектов и структур.

Основным подходом к решению поставленной задачи является комбинированное использование большинства современных вычислительных технологий для компьютерного моделирования в области квантовой химии и молекулярной динамики в рамках единой компьютерной среды, дружественной конечному пользователю-химику, позволяющей эффективно проводить расчеты в удаленном режиме. Создаваемая вычислительная среда позволяет более эффективно решать "тяжелые", многостадийные задачи в областях молекулярного моделирования наноматериалов и наноструктур, делая их доступными более широкому кругу исследователей. Модульная структура комплекса (вычислительной среды) позволяет достаточно просто вводить новые методы расчетов, выбирать различные варианты сочетания вычислительных сервисов, вводить новые функциональные возможности (в том числе новые прикладные пакеты) без ущерба общей работоспособности комплекса.

Результаты применения и сложность задач

С использованием комплекса проведено более 60 вычислительных экспериментов в области компьютерного моделирования компонентов низкотемпературных топливных элементов [4, 8] (нанокompозитные электроды на основе диоксидов переходных металлов и платины и мембраны на основе сложных гетерополикислот, полные модели элементов) и более 200 – при моделировании литий-ионных аккумуляторов новых типов [3, 5, 6, 9] (композитные структуры на основе углерода и кремния – нанотрубки, наностержни, структурные кластеры кремния размерностью до 300 атомов, «нанобумага», твердые и полимерные электролиты, процессы зарядки/разрядки, транспортные модели ионного переноса). Продолжительность некоторых расчетов на кластере ИПХФ достигала 1 месяца. Примеры смоделированных наноструктур и процессов приведены на рис.2.

В целом, опыт применения созданной вычислительной среды показал, что происходит интенсификация труда химиков-исследователей (повышается количество проведенных расчетов) и повышение качества интерпретации

полученных результатов. Существуют определенные проблемы в обеспечении корректной авторизации и безопасности для доступных ресурсов и пользовательской информации, но они постепенно разрешаются с развитием созданной вычислительной среды.

Рис. 2. Примеры результатов компьютерного моделирования с использованием веб-ориентированных вычислительных сервисов: (а) транспортная модель функционирования топливного (водородного) элемента с участием кластеров платины Pt_{19} на поверхности сверхструктуры диоксида олова $(SnO_2)_{64}$ [9]; (б) 3-D модель комплекса полимерного электролита LiNafion*8DMSO и слоев кремний-углеродной «нанобумаги» $Si_{32}C_{38}$. [10]

Проведение работ по данной тематике поддержано грантом РФФИ № 15-07-07867-а, рук. Волохов В.М.

Литература

1. 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
2. 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

3. Volokhov V.M., Varlamov D.A., Zyubina T.S., Zyubin A.S., Volokhov A.V. Supercomputer simulation components and processes in the Li-ion power sources of new types // Казань:: 3-7 апреля 2017 г.: Казанский (Приволжский) федеральный университет, "Параллельные вычислительные технологии (ПаВТ) 2017" / "Parallel computational technologies (PCT) 2017". In: Communications in Computer and Information Science (in press)
4. В.М. Волохов, Д.А. Варламов, Т.С. Зюбина, А.С. Зюбин, А.В. Волохов, Г.А. Покатович Квантово-химическое моделирование наноструктурированных компонентов низкотемпературных электрохимических топливных элементов // "Параллельные вычислительные технологии (ПаВТ'2014)", труды международной научной конференции (1–3 апреля 2014 г., г. Ростов-на-Дону). Челябинск: Издательский центр ЮУрГУ, 2014, (389 с.), с.235-242
5. В.М. Волохов, Д.А. Варламов, Т.С. Зюбина, А.С. Зюбин, А.В. Волохов Суперкомпьютерное моделирование процессов взаимодействия кремний-углеродных наноструктурированных электродов и твердых электролитов в Li-ионных источниках тока новых типов // Суперкомпьютерные дни в России: Труды международной конференции (26-27 сентября 2016 г., г. Москва). – М.: Изд-во МГУ, 2016. (1128 с.) с.690-699
6. В.М. Волохов, Д.А. Варламов, Т.С. Зюбина, А.С. Зюбин, А.В. Волохов, Г.А. Покатович Суперкомпьютерное моделирование транспортных и энергетических процессов в нанокompозитных материалах на основе углерода и кремния // Параллельные вычислительные технологии (ПаВТ'2016): труды X международной научной конференции (28 марта – 1 апреля 2016 г., г. Архангельск, САФУ). Челябинск: Издательский центр ЮУрГУ, 2016. с.105-117
7. В.М. Волохов, Д.А. Варламов, А.В. Волохов, А.И. Прохоров, Г.А. Покатович Построение среды компьютерного моделирования нанокompозитных материалов на базе комплекса вычислительных сервисов и высокоуровневых веб-интерфейсов // Научный сервис в сети Интернет: труды XVII Всероссийской научной конференции (21-26 сентября 2015 г., г.Новороссийск). - М.: ИПМ им. М.В.Келдыша, 2015. (336 с.) с. 63-68
8. Т. С. Зюбина, А. С. Зюбин, Ю. А. Добровольский, В.М. Волохов Водород на композитном катализаторе Pt₂₉/SnO₂. Квантово-химическое моделирование // Известия Академии наук. Серия химическая, 2015, № 4, С.752-758.
9. Зюбин А.С., Зюбина Т.С., Добровольский Ю.А., Волохов В.М. Анодные материалы на основе кремния и углерода: квантово-химическое моделирование // Журнал неорганической химии, 2016, 61, 1, с.51–58, DOI:10.7868/S0044457X16010256