

ВЫСОКОПРОИЗВОДИТЕЛЬНЫЙ ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ КВАНТОВО-МЕХАНИЧЕСКИХ РАСЧЕТОВ И МОДЕЛИРОВАНИЯ НАНОРАЗМЕРНЫХ АТОМНО-МОЛЕКУЛЯРНЫХ СИСТЕМ

А.В. Бухановский, В.Н. Васильев, В.Г. Маслов, Н.Н. Розанов

В докладе обсуждаются ключевые проблемы и реализованные для их решения подходы, методы и технологии, воплощенные в ходе разработки отечественного высокопроизводительного программного комплекса для квантово-механических расчетов и моделирования наноразмерных атомно-молекулярных систем. Программный комплекс предназначен для проведения расчетов из первых принципов общего характера и для компьютерного моделирования и расчета наноструктур и наноматериалов с наперед заданными свойствами в различных областях, включая основные приоритетные направления развития работ в области нанотехнологий (нанoeлектроника, наноматериалы для сверхпроводящих, магнитных и оптических систем, диагностика и контроль свойств наноструктур, моделирование наноструктур конструкционного назначения и пр.).

Современный этап развития программного обеспечения для квантово-химических расчетов примечателен тем, что, несмотря на многообразие предлагаемых решений, с классическими задачами устойчиво ассоциирован ряд эталонных пакетов (например, GAMESS, Molpro, Jaguar и пр.), которые создавались в течение десятилетий распределенными научными коллективами из десятков разработчиков. Эти пакеты отличаются, прежде всего, наличием развитого сообщества пользователей, что обеспечивает возможность их дальнейшей эволюции как независимых приложений с высокой надежностью получаемых результатов за счет обратной связи. Вместе с тем, типовой современный пользователь на практике использует обычно лишь один-два таких пакета. Лицензионная специфика здесь не является определяющей; существенно большая проблема заключается в сопоставлении данных и задач, реализуемых различными пакетами. Сопоставление пакетов на уровне задач может проводить только высококвалифицированный эксперт предметной области, имеющий большой опыт выполнения таких расчетов, их физической интерпретации и дальнейшего использования. Таким образом, на настоящий момент видится целесообразным создание программного инструментария нового поколения, который мог бы не только интегративно реализовывать широкий спектр возможностей, предоставляемых уже существующими пакетами, но и обеспечивать пользователю интеллектуальную поддержку по их применению для решения различных классов задач. В первую очередь, к ним относятся задачи компьютерного моделирования наноразмерных структур, а также материалов и устройств на их основе.

Компьютерное моделирование в нанотехнологиях ставит своей целью решение обратной (с точки зрения производства) задачи – определение макрохарактеристик объекта или явления, используя для этого знания об его атомно-молекулярной структуре, полученные из первых принципов. Традиционные квантово-химические пакеты не решают поставленной задачи; требуется их сопряжение с проблемно-ориентированными пакетами компьютерного моделирования макрообъектов. При этом следует учитывать, что математические модели физических процессов в макрообъектах используют различные (полу)эмпирические параметры, которые могут быть определены различными способами, в том числе, из экспериментов, справочной литературы или квантово-химических расчетов в разных постановках. При этом невозможно априори сформулировать адекватный способ выбора. Потому в большинстве задач не может существовать однозначных рекомендаций, как именно нужно получать входные данные посредством квантово-химических расчетов, и как нужно выполнять преобразование выходных данных квантово-химических расчетов во входные данные того или иного пакета моделирования макрообъекта. Как следствие, для создания вычислительных цепочек, реализующих расчеты макрообъектов, используя данные первопринципных вычислений, необходимо применение специальных компонентов сопряжения, осуществляющих пересчет, реорганизацию и форматирование данных для бесшовного взаимодействия пакетов (что является задачей и технологической, и содержательной, поскольку метод пересчета должен быть согласован с математическими моделями, реализованными как в пакете- «доноре», так и в пакете-«реципиенте»). При этом специфика организации бесшовного взаимодействия осложняется тем, что современное программное обеспечение квантово-химических расчетов и компьютерного моделирования в нанотехнологиях (в силу широты предметной области) выполнено коллективами разработчиков разной квалификации с использованием различных технологий программирования. Оно рассчитано на работу под различными операционными системами, имеет различные формы организации ввода- вывода (не говоря о форматах входных и управляющих данных). Совмещение такого программного обеспечения в рамках единого комплекса на основе традиционного компонентного подхода (например, инкапсуляция вычислительных компонентов в форму объектов с использованием паттерна проектирования bridge) не представляется целесообразным.

Отдельная проблема состоит в том, что квантово-химические расчеты наноразмерных структур и комплексов являются крайне ресурсоемкими в силу степенной сложности применяемых алгоритмов. Однако в большинстве случаев их функциональное распараллеливание (исключая отдельные случаи, например, методы квантового Монте- Карло) даже на кластерах с быстрой коммуникационной сетью эффективно на весьма

небольшом количестве вычислителей (16-32). Как следствие, распараллеливание на значительном количестве вычислителей (и даже на нескольких суперкомпьютерах) реализуется в подавляющем большинстве по данным, т.е. путем независимых запусков пакета для разных наборов параметров (которые обычно имеют неодинаковое время выполнения). Именно потому проблема построения сбалансированных параллельных алгоритмов для квантово-химических расчетов на суперкомпьютерах терафлопной производительности является не технологической, а предметно-ориентированной. Механизмы распараллеливания задачи по данным могут быть сформулированы только экспертом предметной области на уровне сценариев запусков пакета. При этом внутреннее распараллеливание также может быть применено для более эффективного использования вычислительных ресурсов (иерархическая, или гибридная схема). Следовательно, программный комплекс должен сам обеспечивать возможности реализации таких сценариев распараллеливания (хранящихся как соответствующие знания) при запуске задания, автоматически выбирая режим запуска (способ декомпозиции, количество узлов, потоков на узле и пр.), оптимальный по производительности.

Рассмотренные выше проблемы и предлагаемые решения учтены при проектировании и разработке рабочего образца высокопроизводительного программного комплекса для квантово-механических расчетов и моделирования наноразмерных атомно-молекулярных систем. Программный комплекс реализует концепцию iPSE (intelligent Problem Solving Environment) [1,2], которая расширяет известную концепцию создания проблемно-ориентированных сред PSE [3] за счет введения следующих положений: - iPSE реализует функции интеллектуальной (советующей) системы поддержки принятия решений исследователя при формулировке вычислительных задач и построении расчетных сценариев. - iPSE ориентирована на выполнение высокопроизводительных вычислений на неоднородных вычислительных системах в рамках модели метакомпьютинга с обеспечением максимальной производительности расчетов на основе априорных знаний о предметной области и мониторинга характеристик целевых систем. - iPSE предоставляет единый интерфейс взаимодействия для предметно-ориентированных программных модулей и компонентов, которые могут разрабатываться различными коллективами, могут быть написаны на разных языках и иметь различные условия распространения и использования. Указанные положения, направленные на повышение масштабируемости, гибкости и эргономичности комплекса в целом, принципиально отличают его от традиционных квантово-химических PSE (например, Molecular Science Software Suite, MS3 [4]), которые, по сути, являются оболочками виртуализации и интеграции вычислительных ресурсов, оставляя вопросы формирования заданий и запуска расчетов на рассмотрение пользователю в меру его квалификации.

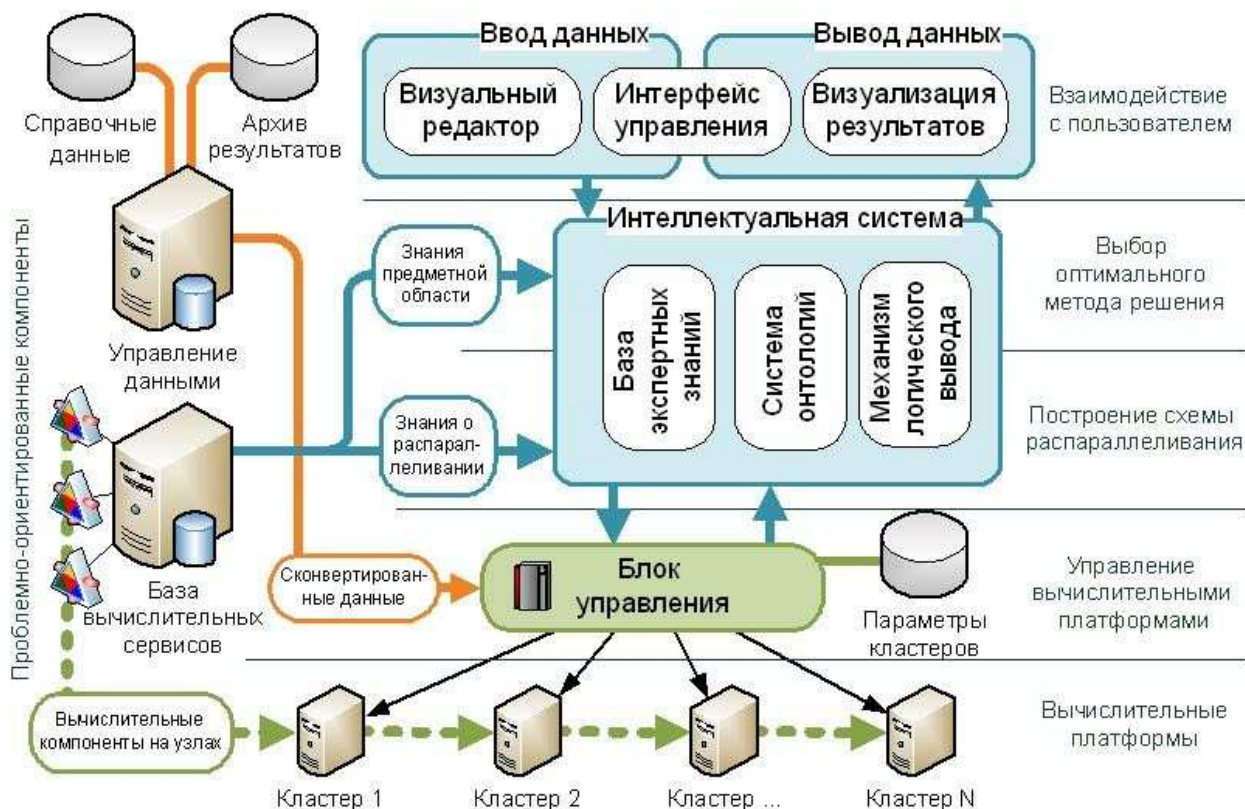


Рис. 1.

На рис. 1 приведена общая архитектура программного комплекса.

Программный комплекс представляет собой сложную распределенную программную систему, которая состоит из системных компонентов (формирующих управляющее ядро), проблемно-ориентированных

компонентов, установленных на целевых вычислительных системах, а также одной или нескольких клиентских частей, обслуживающих конкретных пользователей. В качестве модели организации компонентов программного комплекса рассматривается подход SaaS (Software as a Service), реализуемый посредством SOA (Service Oriented Architecture). Это дает возможность объединять вместе программные компоненты (как отдельные «черные ящики» - сервисы), разработанные разными авторами, реализованные на различных технологиях и имеющие различную форму распространения и поддержки, интерпретируя их как web-сервисы.

Проблемно-ориентированные сервисы представлены программными компонентами, которые размещаются на различных суперкомпьютерах; при этом никакие связи между компонентами в рамках одного суперкомпьютера не устанавливаются. При запуске на своем суперкомпьютере каждый из программных компонентов допускает внутреннее (функциональное) распараллеливание в соответствии с заложенным в нем алгоритмом и доступными ресурсами суперкомпьютера. Подготовка и передача входных данных, запуск пакетов и сбор результатов осуществляются централизованно, на основе команд, поступающих от управляющего ядра. Следует отметить, что целевые суперкомпьютеры и работающие на них программные компоненты могут функционировать под различными операционными системами и иметь различные системы управления; их объединение выполняется посредством кроссплатформенной управляющей оболочки (платформы исполнения). Она осуществляет управление и мониторинг текущего состояния целевых суперкомпьютеров, исполняемых на них задач, а также общей коммуникационной сети, обеспечивающей функционирование гиперкластера в целом.

Стратегия управления (правила взаимодействия сервисов) формируется управляющим ядром комплекса, как набором сервисов, установленных на отдельном управляющем сервере. Основой ядра является главный управляющий сервис (компонент управления параллельным исполнением). Управляющий сервис, используя информацию о текущих заданиях, знания (в форме параметрических моделей) об их производительности и результаты мониторинга состояния целевых систем, осуществляет планировку совместного исполнения цепочки задач наиболее эффективно по пользовательским критериям (времени выполнения, стоимости, надежности и пр.). Управляющий сервис является, по сути, генератором команд и централизующим звеном для остальных сервисов, заменяя, таким образом, традиционную сервисную шину (что необходимо для достижения наибольшей производительности комплекса).

Доступ пользователей к возможностям комплекса осуществляется через графический интерфейс, который является «толстым клиентом» для управляющего ядра. Пользователь имеет возможность в привычном оконном интерфейсе формулировать задачу на языке предметной области (получая при необходимости интерактивную справку), подготавливать и вводить файлы данных, давать команду на выполнение задания, осуществлять мониторинг выполнения. По окончании вычислений пользователь может получить требуемые ему результаты расчетов, а также визуализировать их посредством инструментария, установленного на его клиентском компьютере. Данные вычислений доступны постфактум; в реальном времени передачи данных и визуализации расчетов не производится в силу (а) несоответствия физических масштабов реальному времени вычислений и (б) разнородности применяемых пакетов без возможностей их полной унификации.

Пользователь может взаимодействовать с комплексом в трех режимах работы. Мануальный режим позволяет пользователю через графический клиентский интерфейс готовить данные и удаленно запускать конкретный программный компонент из доступных в комплексе. При этом пользователь должен быть достаточно квалифицирован в области возможностей и форматов данных выбранного им пакета. Автоматический режим (режим экспертной системы), напротив, дает возможность пользователю, проходя интервью в терминах предметной области, выбрать наиболее подходящие для его задач программные компоненты, корректно подготовить данные и описать последовательность запусков. Полуавтоматический режим (режим сценариев) дает возможность пользователю выбрать уже готовый сценарий сопряжения компонентов для расчета тех или иных физических характеристик. Автоматический и полуавтоматический режимы реализуются посредством компонента интеллектуальной поддержки пользователя, также установленном как сервис на управляющем сервере.

Результаты исследований масштабируемости программного комплекса и некоторые его практические приложения будут наглядно проиллюстрированы в докладе.

Работа выполнена в рамках проекта «Разработка высокопроизводительного программного комплекса для квантово-механических расчетов и моделирования наноразмерных атомно-молекулярных систем и комплексов» (№ ГР 01.2.009.50865) в рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2012 годы».

ЛИТЕРАТУРА:

1. Высокопроизводительный программный комплекс моделирования атомно-молекулярных наноразмерных систем / В.Н. Васильев [и др.] // Научно-технический вестник СПбГУ ИТМО. Выпуск 54. Технологии высокопроизводительных вычислений и компьютерного моделирования. – СПб.: СПбГУ ИТМО, 2008. – С. 3-12
2. Ковальчук С.В. и др. Особенности проектирования высокопроизводительных программных комплексов для моделирования сложных систем // Информационно-управляющие системы, 3(34), 2008, с. 10-18.

3. "Computer as Thinker/Doer: Problem-Solving Environments for Computational Science" // S.Gallopoulos, E. Houstis, J. Rice, IEEE Computational Science and Engineering, Summer 1994.
4. www.emsl.pnl.gov:2080/mscf/about/descr_ms3.htm.