

СПОСОБЫ РЕШЕНИЯ ПРИКЛАДНЫХ МНОГОПАРАМЕТРИЧЕСКИХ ЗАДАЧ ВЫЧИСЛИТЕЛЬНОЙ ХИМИИ НА ПЕТА- И ЭКЗАФЛОПНЫХ СИСТЕМАХ

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

Развитие вычислительных мощностей и перспективных технологий уже позволили [1] достичь на суперкомпьютерных установках уровня производительности в первые петафлопы (10^{15} оп/с - петафлопный рубеж был преодолен в 2008 году), а в обозримом будущем (2014-2018 гг.) позволят как сделать подобные системы во многом рядовыми, так и создать первые системы следующего, экзафлопного (10^{18}) уровня производительности. Все ведущие в области ИТ страны (США, Япония, Китай, Европейский Союз, Россия) сформировали проекты по построению систем сверхвысокой производительности и созданию системного и прикладного программного обеспечения для их эффективного использования: IESP (International Exascale Software Project), EESI (European Exascale Software Initiative), Национальная суперкомпьютерная платформа и др. Наряду с массой уже озвученных технических проблем (энергопотребление, коммуникации, доступ к памяти, объемы хранения, надежность) уже сейчас перед программистами и системными администраторами стал весьма актуальный вопрос: как эффективно использовать в прикладных областях подобные суперкомпьютерные вычислительные системы сверхвысокого уровня производительности? Очевидно, что будет необходимо решить ряд принципиально новых проблем, вставших как перед разработчиками системного ПО, так и перед программистами, работающими с прикладным ПО.

Встала необходимость перестройки всего ПО, работающего на современных суперкомпьютерах (как правило, терафлопной мощности), пересмотреть парадигму развития всех программных компонент будущих вычислительных систем, включая средства управления заданиями и разработки программ, модели и методы параллельного программирования сверхбольших систем, анализ производительности и эффективности прикладного и системного ПО, обеспечение надежности и отказоустойчивости, обеспечение высокой степени масштабируемости прикладных пакетов – перед разработчиками ПО стоит, на первый взгляд, необъятное поле деятельности.

К сожалению, приходится констатировать, что большинство прикладных пакетов вычислительной химии пока откровенно не готовы для работы в подобных условиях. Несмотря на очень высокие требования к ресурсам и большие требуемые объемы вычисления для многих задач вычислительной химии [2], степень параллелизации и, соответственно, масштабируемость большинства используемых химиками прикладных пакетов не очень высока. Как правило, эффективная параллелизация обычно наблюдается на первых сотнях процессоров, исключение составляют только некоторые пакеты, как, например, NAMD. NAMD – Nanoscale Molecular Dynamics [3] – программа для молекулярной динамики, написанная с использованием модели параллельного программирования Charm++, обладающей высокой эффективностью распараллеливания (до нескольких тысяч ядер) и часто используемой для симуляции больших систем (миллионы атомов). Таким образом, для нужд подобных пакетов вполне достаточен уровень существующих суперкомпьютерных ресурсов.

Как же (в первом приближении) могут быть использованы сверхпроизводительные системы для нужд вычислительной химии? Авторам представляется возможным следующий путь: разработать и подготовить к использованию технологии для решения ресурсоемких многопараметрических задач из различных областей науки на суперкомпьютерных системах сверхвысокого уровня производительности. Эти задачи работают с большими наборами последовательно изменяемых входных параметров, а также решаются не только для одного фиксированного набора первичных данных, а на создаваемой детальной «сетке» входных данных. Большинство многопараметрических задач решается разбиением первичной задачи (например, моделирование поведения системы или процесса) на значительное количество совершенно независимых друг от друга заданий, либо не требующих обмена данными и сигналами между собой, либо сводящих подобный обмен к минимуму. Сами по себе задания имеют, как правило, небольшую вычислительную сложность и внутреннее распараллеливание на относительно небольшом количестве процессоров – от SMP задач на нескольких ядрах до первых десятков MPI процессов, поэтому не требуют кардинальной перестройки парадигмы программирования. В результате общее количество заданий, на которые может быть разбита задача, экспоненциально растет в зависимости как от количества используемых изменяемых параметров расчета, так и детальности разбиения областей входных данных и определяются как требованиями достижимой точности решения и общего масштаба задачи, так и главным образом – доступными вычислительными мощностями. В настоящее время большинство таких задач решается созданием «пучков» заданий с количеством 10^4 - 10^6 на терафлопных установках или на распределенных ГРИД полигонах, однако, требуемая точность и детальность сеток входящих данных для большого диапазона актуальных многопараметрических задач оценивается в 10^8 - 10^{10} заданий (и даже выше). Только при этом обеспечивается точность вычислительного эксперимента, либо сопоставимая с натурными, весьма дорогостоящими экспериментами, либо в принципе недостижимая другими способами (что особенно характерно для квантовой химии и моделирования наноструктур). Применение таких новых принципов

вычислений позволит решать конкретные задачи, интересные для промышленного применения и позволяющие создавать реальные рыночные продукты. На существующих вычислительных мощностях исследователи вынуждены либо существенно занижать количество изменяемых параметров задачи, что ведет к неадекватному учету ряда физико-химических эффектов, либо «загрублять» сетки входных данных, что ведет к утрате части немонотонных результатов (краевые эффекты, локальные «потенциальные ямы» и т.п.). Повышение точности расчетов часто ведет к выявлению непредсказуемых заранее теорией результатов и эффектов.

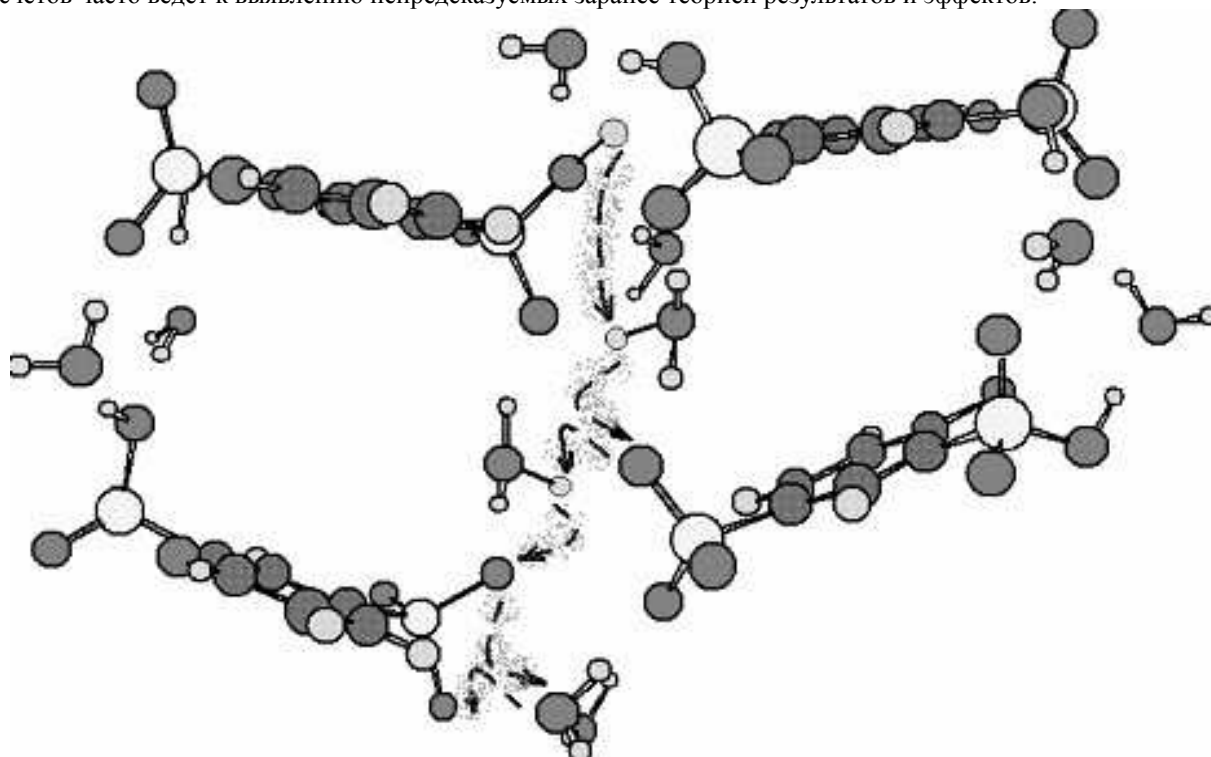


Рис. 1

Рис.1 Пример одной из многопараметрических задач квантовой химии – протонная проводимость в фенол-сульфокислотах, используемых для производства мембран твердых топливных элементов. Используемые пакеты – GAMESS, GAUSSIAN-03. В 2010 году – одна точка требует до месяца расчетного времени на 8 ядрах. Для построения точной модели распределения протонов нужно до 5000 точек, соответственно, до 40000 CPU в течение месяца

Основная проблема решения подобных задач состоит в разработке устойчиво работающих систем генерации, запуска, мониторинга выполнения и сбора результатов решения, способных справиться со сверхбольшим (с нынешней точки зрения) количеством используемых заданий, на которые разбиваются подобные задачи. При росте их количества более 10^7 - 10^8 на задачу сама система генерации и управления, а также сбора результатов, становится весьма ресурсоемкой (первые терафлопы процессорного времени и терабайты памяти) и требующей новых механизмов управления и сохранности данных/результатов.

Дополнительными преимуществами подобных технологий являются:

- эволюционное развитие в процессе роста доступных ресурсов от нынешних терафлопных ресурсов через петафлопные к эксафлопным. На всех стадиях существуют задачи, отвечающие уровню развития вычислительных технологий на текущем этапе, что позволяет развивать указанные методы расчета постепенно и использовать для решения научно-практических задач разных типов;
- возможность использования решения подобных задач для масштабного «заполнения» рабочего пространства пета- и эксапроизводительных ресурсов, невостребованного параллельным проведением сверхресурсоемких расчетов, что также возможно на всем протяжении проекта в соответствии с ростом доступных ресурсов.

Результаты наших исследований могут послужить основой дальнейшей разработки комплекса программных средств и инструментов, способных эффективно решать многопараметрические задачи различных типов на компьютерах сверхвысокой производительности, а также позволят решить ряд научно-практических задач в области вычислительной химии и смежных областей науки, решение которых ныне не производится или проводится с очень грубыми приближениями как по количеству используемых параметров, так и на очень разреженных областях входных данных.

ЛИТЕРАТУРА:

1. <http://www.top500.org> - TOP500 Supercomputing Sites
2. Волохов В.М., Варламов Д.А., Пивушков А.В., Покатович Г.А., Сурков Н.Ф. Технологии ГРИД в вычислительной химии // "Вычислительные методы и программирование", М.: МГУ, 2010, т.11, № 1, с.175-182
3. <http://www.ks.uiuc.edu/Research/namd> - NAMD, Scalable Molecular Dynamics