

# МЕТОДЫ ВИЗУАЛЬНОЙ ПОДДЕРЖКИ ДЛЯ ЗАДАЧ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ НА СУПЕРКОМПЬЮТЕРЕ BLUE GENE /P

Н.Н. Попова, О.В. Джосан, Г.Н. Шумкин

Задачи молекулярного моделирования — это один из основных классов задач, которые в настоящее время эффективно решаются на суперкомпьютерах серии Blue Gene [1]. Среди наиболее популярных кодов, которые инсталлируются и используются на вычислителях этой модели, можно отметить CPMD[2], NUMD[3], LAMMPS[4], AMBER[5]. Для прикладных исследований важной составляющей является визуальное представление информации. Поэтому визуализация результатов многопроцессорных параллельных вычислений — это важная функциональность для пользователей, решающих прикладные задачи. Разработка набора таких средств, оптимизированных для использования на вычислителе Blue Gene и предоставляющих оптимальный сценарий использования при моделировании молекулярной динамики, является в настоящее время актуальной задачей.

Для кодов моделирования молекулярной динамики существуют специализированные системы визуализации. Наиболее распространенной среди них является VMD[6]. VMD обладает широким набором параметров моделирования и возможностью интерактивного управления визуализацией. Однако основные проблемные места этой и других подобных систем — отсутствие средств параллельной визуализации, а так же сложности интеграции с системами моделирования, а вследствие этого невозможность осуществлять визуализацию в режиме счета.

Еще одной проблемой для систем визуализации научных данных является формат представления входных данных. В системах для визуальной поддержки работы с молекулярной динамикой обычно используется специализированные форматы данных, которые основываются на описании молекулярной структуры как набора объектов с некоторыми признаками. Одним из таких форматов является .cube.

В данном докладе рассматривается специализированное программное средство для визуальной поддержки анализа данных при решении задач молекулярной динамики на многопроцессорных вычислительных комплексах, в частности на машинах семейства Blue Gene. Разработана специализированная параллельная библиотека, позволяющая проводить визуализацию данных в процессе расчета, т.е. в режиме реального времени. Еще одной важной характеристикой предложенных методов визуализации является использование параллельных вычислений для улучшения визуального качества получаемого изображения. Также преимуществом является широкий набор характеристик, которые возможно визуально показать, а также возможность визуализировать несколько характеристик различной природы на одном кадре.

Предложен эффективный сценарий работы системы визуализации, который позволяет специалистам-прикладникам осуществлять анализ данных, полученных в результате крупномасштабного вычислительного эксперимента, на многопроцессорном вычислительном комплексе в режиме реального времени, не передавая огромные массивы информации на локальную машину. Это существенно экономит время анализа, а также стоимость трафика.

Работа предложенной системы визуализации возможна в двух режимах: режиме постобработки, т.е. когда анализ данных происходит после окончания эксперимента, и режиме визуализации, совмещенной со счетом. В режиме визуализации совмещенной со счетом также возможно два сценария работы: режим отображения текущего состояния эксперимента и режим просмотра истории эксперимента. Это позволяет анализировать данные в процессе долгосрочного эксперимента и корректировать его условия при необходимости до окончания счета.

В системе визуализации реализована иерархическая структура распределения функций по вычислительным узлам, что позволяет эффективно распределять вычислительное время и оптимизировать общую стоимость вычислений.

Апробация работы системы была проведена на примере интеграции с кодом моделирования молекулярной динамики CPMD. Тестирование было проведено в рамках вычислительных экспериментов на комплексе Blue Gene /P, установленном в Московском Государственном Университете. Система включает в себя две стойки с вычислительными узлами и узлами ввода-вывода: 1024 четырехядерных вычислительных узла в каждой из стоек и 16 узлов ввода-вывода в стойке. Вычислительный узел включает в себя четырехядерный процессор, 2 ГБ общей памяти и сетевые интерфейсы. Микропроцессорное ядро - PowerPC 450 с рабочей частотой 850 MHz. Доступ к вычислительным ресурсам осуществляется через фронтенд IBM pSeries55A.

Реальный эксперимент был проведен на данных, получаемых при моделировании переключения в структурах молекул. Использовалась модель из 82 атомов, 129 орбиталей, с временным шагом 0.1 fs на решетке 320x320x192. Объем данных одного шага моделирования составляет в среднем 30Mb. Общий объем данных эксперимента при этом составил около 200Gb. Для счета использовалось 512 процессоров, при этом время расчета составляет около суток. Визуализация данных проводилась как в режиме постобработки, так и в режиме

реального времени, что позволило существенно ускорить процесс подбора оптимальных параметров для решения задачи моделирования переключения.

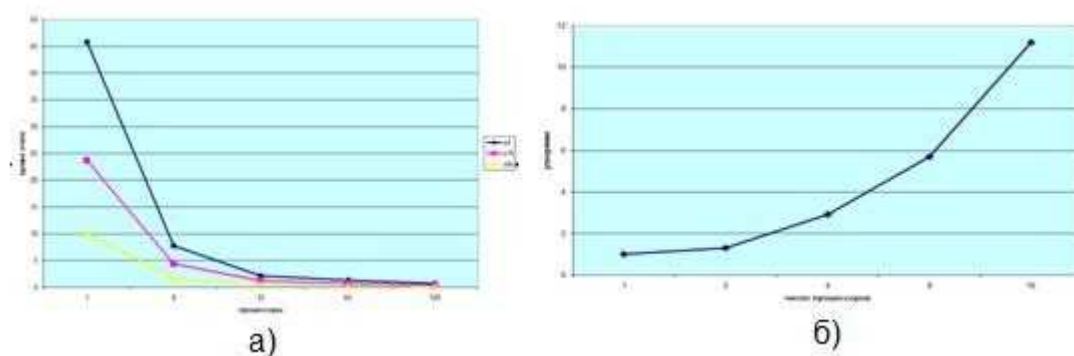


Рис. 1 Эффективность применения системы визуализации

На Рис.1 показаны численные результаты использования системы визуализации. На рисунке Рис.1а показана эффективность использования алгоритма увеличения разрешения изображения (уменьшение времени счета кадра) в зависимости от коэффициента интерполяции и числа процессоров, которые используются при построении изображения. На Рис.2а проиллюстрировано ускорение алгоритма параллельного сжатия в зависимости от количества процессоров.

На Рис.2 показаны примеры использования разработанной системы визуализации при интеграции с моделированием в коде CPMD. На Рис.2а показана структура молекулы, а также значение электронной плотности вокруг атомов. На Рис.2б проиллюстрирована структура молекулы, где размер атома показан в

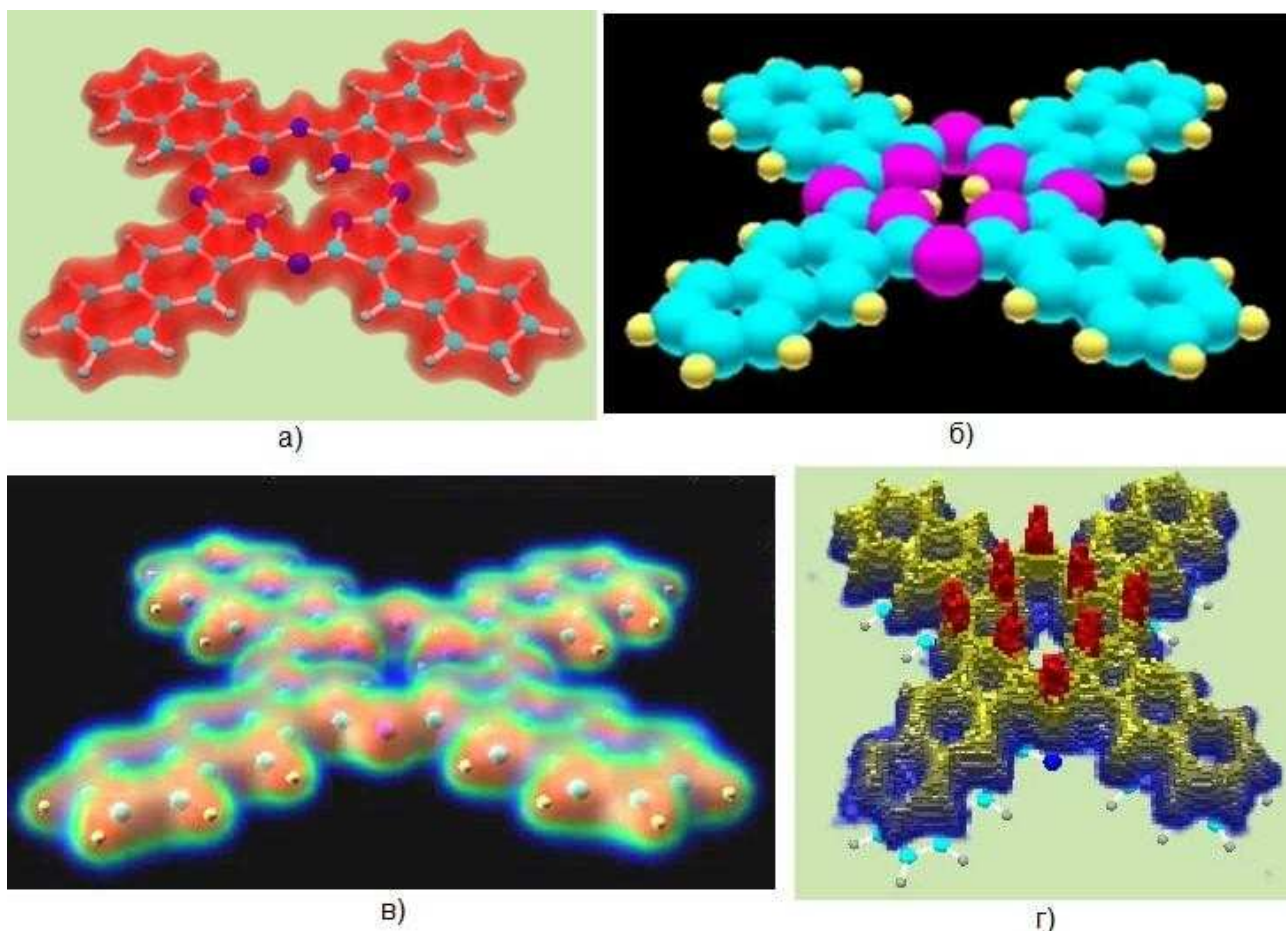


Рис. 2 Пример визуализации данных молекулярной динамики

соответствии с его атомарной массой. На Рис.2в различными цветами показаны изоповерхности электронной плотности. Рис.2г иллюстрирует структуру молекулы, а также двумерную диаграмму значений плотностей на срезе.

Предложенная система визуализации имеет ряд особенностей по сравнению с существующими аналогами. В частности в данной системе визуализации реализована модель визуализации данных в процессе эксперимента, т.е. режим совмещения со счетом. В этом режиме возможен как просмотр текущего состояния эксперимента по моделированию молекулярной динамики, так и анализ истории эксперимента с помощью просмотра видео по данным предыдущих шагов. Важной особенностью системы является применение эффективных параллельных алгоритмов для построения изображения и для обработки видео перед передачей на локальную машину. Реальный эксперимент по интеграции предложенной системы визуализации и кода молекулярной динамики CPMD показал ее высокую эффективность и необходимость дальнейших исследований в направлении развития сценариев взаимодействия при крупномасштабном вычислительном эксперименте, а также формализации интерфейсов этого взаимодействия для дальнейшей интеграции с другими системами моделирования молекулярной динамики.

Данная работа выполняется при поддержке гранта РФФИ № 08-07-12081-офи.

#### ЛИТЕРАТУРА:

1. IBM Blue Gene team, Overview of the IBM Blue Gene/P project // IBM Journal of Research and Development, Vol. 52, No. 1/2, 2008, pp. 199-220
2. Hutter J, Curioni A, Dual-level parallelism for ab initio molecular dynamics: Reaching teraflop performance with the CPMD code // journal PARALLEL COMPUTING , vol. 31 (1), 2005
3. James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten, Scalable molecular dynamics with NAMD // Journal of Computational Chemistry, Vol. 26, 2005, pp.1781-1802.
4. C. Grindon, S. Harris, T. Evans, K. Novik, P. Coveney, C. Laughton, Large-scale molecular dynamics simulation of DNA: implementation and validation of the AMBER98 force field in LAMMPS // Philosophical Transactions of Royal Society A, vol. 362, 2004, pp. 1373-1386.
5. D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods. The Amber biomolecular simulation programs. // journal Computational Chemistry, Vol. 26, 2005, pp. 1668-1688.
6. Jen Hsin, Anton Arkhipov, Ying Yin, John E. Stone, and Klaus Schulten, Using VMD - an introductory tutorial // Current Protocols - Bioinformatics, 5:Unit 5.7, 2008.
7. О.В. Джосан, Н.Н. Попова, «Система визуальной поддержки высокопроизводительных вычислений» // тезисы конференции «Научный сервис в сети Интернет», Россия, Новороссийск, сен. 2008,с.с. 160-162
8. О.В. Джосан «О визуализации научных данных для высокопроизводительных параллельных приложений» // тезисы конференции ПАВТ 2009, Россия, Нижний Новгород, март 2009,с.с. 449-456
9. Lev N. Korolev, Nina N. Popova, Oxana V. Dzhosan «Image and video processing methods for high performance data management systems » // conf. pros. INFORMATION SYSTEMS & GRID TECHNOLOGIES, Bulgaria, Sofia, May 2009