

# МАСШТАБИРУЕМОСТЬ ПРОГРАММ МОЛЕКУЛЯРНОЙ ДИНАМИКИ

А.В. Комолкин

Моделирование классической молекулярной динамики в настоящее время является одним из важных методов исследования молекулярных систем. В настоящее время открываются новые перспективы в исследовании полимеров [1], биологических молекул [2], наносистем [3] и других сложных систем в полноатомном представлении, содержащие сотни тысяч атомов. Поэтому особое внимание при разработке компьютерного кода для моделирования молекулярной динамики необходимо уделять масштабируемости программ. В данном докладе описывается реализация параллельного алгоритма в программе AKMD [4-6].

В классической молекулярной динамике все атомы рассматриваются как материальные точки, взаимодействующие на основе классических потенциалов (не квантово-механических) и подчиняющиеся уравнениям движения Ньютона. Моделирование заключается в численном решении уравнений движения и в нахождении траекторий всех атомов. Алгоритм моделирования можно представить так [4]:

1. в момент времени  $t$  по координатам всех атомов вычисляют энергию их взаимодействия и силы, действующие на каждый атом со стороны остальных атомов;
2. дифференциальные уравнения второго порядка, описывающие движение частиц, интегрируют численно на отрезке времени  $\Delta t$ , т.е. вычисляют скорости и координаты атомов в момент  $t+\Delta t$ ;
3. в ансамблях NVT или NpT (с постоянной температурой или постоянными температурой и давлением) корректируют температуру (скорости атомов) и давление (объем системы);
4. записывают координаты и скорости атомов в файл для последующего анализа;
5. момент времени  $t+\Delta t$  делают текущим ( $t$ ) и выполняют следующий шаг с пункта 1.

Очевидно, что параллелизм программы моделирования может быть заключен только внутри одного шага моделирования (пункты 1-4), при этом между шагами 1 и 2, внутри шага 3 и между шагами 3 и 4 выполняется обмен данными между всеми параллельными ветвями программы.

Межмолекулярная энергия складывается обычно из энергии невалентных (Ван-дер-Ваальсовских) взаимодействий, кулоновских взаимодействий и энергии водородной связи. Все эти вклады являются парными, то есть полная энергия зависит аддитивно от энергии взаимодействия каждой пары атомов  $i$  и  $j$ , которые должны принадлежать разным молекулам:  $F_{ij} = F(r_{ij})$ ,  $F_{ij} = F_{ji}$ , где  $r_{ij}$  — расстояние между взаимодействующими атомами. Кулоновские взаимодействия являются дальнедействующими и вычисляются методом Эвальда [7].

Внутримолекулярная энергия имеет несколько компонент, которые зависят от координат двух, трех или четырех атомов (энергия растяжения валентных связей, искажения валентных углов, энергия торсионных взаимодействий) [5]. Важно отметить, что все внутримолекулярные взаимодействия могут быть вычислены независимо для каждой молекулы.

Вычисление энергии всех типов взаимодействий, таким образом, разделяется на три задачи — вычисление межмолекулярных взаимодействий, внутримолекулярных взаимодействий и взаимодействий по методу Эвальда, что может быть реализовано в парадигме *параллелизма задач*. Важным фактором является балансировка нагрузки на процессоры, так как в типичных задачах время вычисления энергии и сил межмолекулярных взаимодействий занимает приблизительно 70-75%, по методу Эвальда — 20-25%, а внутримолекулярных — 5-10% общего времени вычисления энергии и сил. Как отмечалось ранее [4], новая эффективная реализация метода Эвальда в программе AKMD обладает ограниченной масштабируемостью. Поэтому каждая из этих задач дополнительно подразделяется на подзадачи и выполняется на разных процессорах. В программе на данном этапе ее разработки осуществляется статическая балансировка нагрузки, что не требует обмена данными между процессорами и отдельно работающим диспетчером задач во время вычислений, но предполагает предварительный прогон программы для данной моделируемой системы для определения времени работы каждой из ее подзадач. Предполагается, что время вычислений каждой подзадачи не зависит от значений координат атомов и на каждом шаге моделирования одинаково. Это позволяет назначить различные наборы подзадач каждому процессору на следующем прогоне программы в зависимости от имеющегося числа процессоров. Назначение подзадач процессорам осуществляется последовательно, начиная с самой длительной подзадачи из еще не распределенных подзадач, наименее загруженному процессору.

Перед началом вычислений (перед пунктом 1 алгоритма) координаты всех атомов должны быть известны всем процессорам. В результате выполнения этого пункта алгоритма координаты каждый процессор вычисляет частичные силы, действующие на атомы, в соответствии с распределенными ему подзадачами. С точки зрения молекулярной системы в этом пункте алгоритма реализован параллелизм по парным атом-атомным взаимодействиям и по ячейкам в методе Эвальда. Движение атомов (пункты 2 и 3) должно быть реализовано параллельно по всем атомам с использованием парадигмы *параллелизма данных*. К сожалению, невозможно реализовать параллелизм вычислений по атомам в пункте 1. Поэтому, перед пунктом 2 необходимо вычислить полные силы, действующие на каждый атом, что требует передачи данных между всеми процессорами, и

распределить полученные силы между процессорами в соответствии с теми атомами, движение которых вычисляет данный процессор в пунктах 2 и 3.

Необходимо отметить, что пункт 3 при использовании ансамблей NVT и NpT содержит в себе обмен данными между всеми процессорами. Для поддержания температуры и давления итерационно выполняется их коррекция, что требует четырех или пяти итераций. Процессоры для этого должны знать среднюю температуру и давление по всему ансамблю, т.е. по всем атомам. Хотя на каждой итерации объем передаваемых данных составляет всего одно (для NVT) или два (для NpT) вещественных числа, но обмен заключается в суммировании данных со всех процессоров и передаче суммы всем процессорам.

Проведенный анализ быстродействия программы показывает, что использование всех имеющихся процессоров кластера для перемещения атомов может оказаться нецелесообразным из-за увеличения времени обмена силами и координатами в пунктах 1 или 5 алгоритма и из-за вычисления средних температуры и давления в пункте 3. В зависимости от числа атомов в системе использование процессоров только одного узла кластера для перемещения атомов может оказаться более эффективным, чем использование всех процессоров, поскольку полностью исключается обмен данными по межузловым каналам связи.

**Заключение.** Основным вкладом в уменьшение масштабируемости программы АКМД моделирования молекулярной динамики вносит обмен данными. Сокращение обмена путем уменьшения числа параллельных ветвей на этапе перемещения атомов может увеличить эффективность программы. Количество процессоров необходимо подбирать (оптимизировать) для каждого типа молекулярных систем и для каждой вычислительной системы. Например, на параллельных многопроцессорных системах, возможно, будет эффективным использование всех процессоров, в то время как на параллельных кластерах может быть оптимальным использование процессоров только одного или нескольких узлов.

В программе АКМД реализован новый эффективный алгоритм распределения подзадач вычисления частичных сил меж- и внутримолекулярных взаимодействий и сил в методе Эвальда и способ статической балансировки нагрузки, что улучшает масштабируемость программы.

Работа выполнялась частично в рамках проектов РФФИ (гранты номер 09-03-01105-а и 09-02-00010-а) и программы "Развитие научного потенциала высшей школы", проект РНП.2.2.2820.

#### ЛИТЕРАТУРА:

1. S.D.O. Costa, A.J.F. Romero, J.J.L. Cascales "Physicochemical study of the acetonitrile insertion into polypyrrole films" // *Journal of Chemical Physics*, 2010, Vol. 132, No. 14, Article No. 144702, p. 1-7.
2. S. Keten, M.J. Buehler "Atomistic model of the spider silk nanostructure" // *Applied Physics Letters*, 2010, Vol. 96, No. 15, Article No. 153701, p. 1-3.
3. S. Chelvaraja, P. Ortoleva "Thermal nanostructure: An order parameter multiscale ensemble approach" // *Journal of Chemical Physics*, 2010, Vol. 132, No. 7, Article No. 075102, p. 1-9.
4. А.В. Комолкин, В.С. Неверов "Параллельные алгоритмы молекулярной динамики" // Труды Всероссийской суперкомпьютерной конференции "Научный сервис в сети Интернет: масштабируемость, параллельность, эффективность" (г. Новороссийск, 21-26 сентября 2009 г.). — М.: Изд-во МГУ, 2009. — С. 67-70.
5. Е.Э. Эльц, А.В. Комолкин, Л.А. Кутуля, Н.С. Пивненко "Взаимное влияние ЖК среды и растворенных в ней хиральных добавок 2-арилиден-производных 1R,4R-п-ментан-3-она" // *Журнал физической химии*, 2007, том 81, No. 11, с. 2011-2016.
6. А.В. Комолкин "Новые подходы к моделированию молекулярной динамики жидких кристаллов" // В кн.: *Органические и гибридные наноматериалы: монография* / под ред. В.Ф. Разумова и М.В. Клюева. — Иваново: Иван. гос. ун-т, 2009. — С. 202-225.
7. M.P. Allen, D. Tildesley. *Computer Simulation of Liquids*. — Oxford: Clarendon Press, 1987