

Ускорение молекулярно-динамического моделирования многофазных систем при помощи GPU¹

Марьян Д.Ф.

Центр «Микро- и наномасштабной динамики дисперсных систем», БашГУ, Уфа
Институт механики им. Р.Р. Мавлютова УНЦ РАН, Уфа

В работе представлены результаты по ускорению моделирования методами молекулярной динамики при помощи графических процессоров (GPU). Рассматривалась полярная жидкость на примере воды. Межмолекулярное взаимодействие осуществлялось на основе потенциала Кулона и обрезанного потенциала Леннарда-Джонса. Проведены вычислительные эксперименты по ускорению и производительности.

1. Введение

Задачи динамики дисперсных систем в микро- и наномасштабе возникают во многих отраслях науки и промышленности: механической, химической, нефтяной, экологической и др. Экспериментальные исследования процессов, происходящих на микро- и наноуровнях, сильно затруднены тем фактом, что размеры исследуемых элементов и структур зачастую оказываются на порядок меньше размеров длины волны видимого света. Это означает, что фиксация происходящих процессов либо достаточно сложна и требует сложного дорогостоящего оборудования, либо просто невозможна, так как коротковолновое рентгеновское и гамма излучения характеризуются высокой энергией квантов излучения, то есть их использование может в значительной степени искажать реальную картину наблюдаемых процессов.

Теоретические исследования в этой области ограничены применимостью классических континуальных моделей многофазных систем, и актуальность приобретают кинетические модели, используемые в методах молекулярной динамики.

На фоне описанных сложностей вычислительный эксперимент при помощи методов молекулярной динамики, позволяющий описывать и измерять мельчайшие детали, становится незаменимым.

Однако и при проведении вычислительного эксперимента имеется ряд проблем, которые связаны с тем, что при достаточно подробном матема-

тическом описании проблемы, учитывающем многомерность и многопараметричность, а также с использованием при моделировании большого числа частиц, серьезно возрастают требования к производительности как используемого программного кода, так и вычислительной системы в целом.

Существует два пути для уменьшения необходимого времени для проведения вычислительного эксперимента: использование других алгоритмов и реализация на высокопроизводительном оборудовании. Первый заключается в снижении вычислительной сложности используемых алгоритмов. Так сложность прямого алгоритма пропорциональна числу всех парных взаимодействий в системе размера N и равна $O(N^2)$, то есть время выполнения данного алгоритма возрастает квадратично с ростом размера системы. Использование алгоритмов, которые имеют меньшую вычислительную сложность, является задачей чрезвычайно важной и актуальной.

Наряду с использованием алгоритмов с низкой вычислительной сложностью существует еще одно направление повышения производительности вычислений — использование высокопроизводительных вычислительных систем. В настоящее время наиболее эффективными для задач динамики многих тел являются гетерогенные системы, представляющие собой вычислительные кластеры, узлы которых содержат как CPU (центральный процессор), так и GPU (графический процессор). Однако эффективное использование описанных вычислительных систем требует как значительной моди-

¹Работа выполнена при поддержке Министерства образования и науки РФ (грант 11.G34.31.0040).

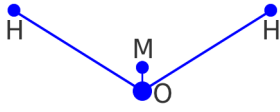


Рис. 1. Схематическое представление модели TIP4P

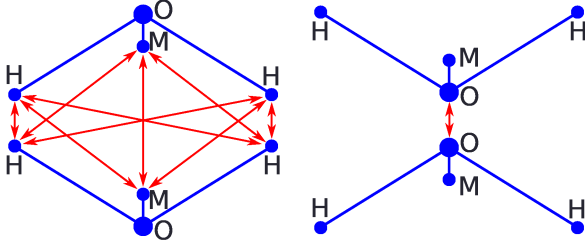


Рис. 2. Дальнее (слева) и ближнее (справа) взаимодействия

фикации существующих алгоритмов, так и разработки новых.

Таким образом, проведение исследований процессов динамики дисперсных систем, происходящих на микро- и наноуровнях, требует реализации молекулярно-динамических моделей с использованием гетерогенных вычислительных систем.

2. Математическая модель

Для моделирования воды, которая представляет собой полярную жидкость, использовалась модель TIP4P (рис. 1) [1]. Эта модель описывает воду как жесткую молекулу и является расширением традиционной трехточечной модели TIP3P путем добавления дополнительного безмассового узла, в котором располагается заряд кислорода. Этот узел располагается на фиксированном расстоянии от атома кислорода на биссектрисе угла НОН. Такое расположение атомов, узлов, масс и зарядов позволяет корректно моделировать электростатическое поле вокруг молекулы и воспроизводить экспериментально установленные свойства воды.

Так как молекула воды полагается жесткой, то внутри молекулы атомы не взаимодействуют. Также эта модель воды полагается только на нехимическое взаимодействие атомов, а именно на дальнее (Кулоновское) и ближнее (Леннард-Джонсовское) взаимодействия. Потенциал, описывающий взаимодействие между атомами i и j различных молекул, описывается формулой

$$U_{ij} = U_{ij}^{Coulomb} + U^{LJtrunc}(r_{ij}), \quad (1)$$

Первое слагаемое (1) представляет собой электростатическое взаимодействие и описывается при

помощи потенциала Кулона:

$$U_{ij}^{Coulomb} = \frac{k_c q_i q_j}{r_{ij}},$$

где k_c — электростатическая константа; q — электрический заряд; r_{ij} — расстояние между атомами i и j .

Второе слагаемое в формуле (1) описывает Леннард-Джонсовское взаимодействие. Для ускорения расчетов потенциал Леннарда-Джонса обрывается на расстоянии $r_c = 2,5\sigma$. И, чтобы избежать нефизичной ситуации, такой, что при пересечении сферы радиуса r_c какой-то молекулой энергия системы меняется скачкообразно, потенциал сдвигается, чтобы выполнялось условие $U^{LJtrunc}(r_c) = 0$. Таким образом обрезанный потенциал Леннарда-Джонса принимает следующий вид

$$U^{LJtrunc}(r) = \begin{cases} U^{LJ}(r) - U^{LJ}(r_c), & r \leq r_c, \\ 0, & r > r_c. \end{cases}$$

Сам потенциал Леннарда-Джонса имеет вид

$$U^{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right],$$

где r — расстояние между частицами; ϵ — глубина потенциальной ямы; σ — расстояние, на котором энергия взаимодействия становится равной нулю.

Потенциал Кулона применяется при расчете взаимодействия между атомами водорода H и дополнительным узлом M , а потенциал Леннарда-Джонса для расчета взаимодействия между атомами кислорода O (см. рис. 2).

Кинетические уравнения движения атомов следуют из второго закона Ньютона:

$$m\ddot{\mathbf{r}}_i = \mathbf{f}_i = \sum_{\substack{j=1 \\ (j \neq i)}}^N \mathbf{f}_{ij},$$

где $\mathbf{f}_{ij} = -\nabla U_{ij}$.

Макроскопические параметры (температура, давление, плотность среды и др.) могут быть получены, исходя из положений молекулярно-кинетической теории.

3. Численное моделирование

Для интегрирования уравнений движения используется численная схема — метод предиктор-корректор (Adams-Bashforth-Moulton), который имеет вид

$$x(t+h) = x(t) + h\dot{x}(t) + h^2 \sum_{i=1}^{k+1} \alpha_i f(t + [l-i]h),$$

$$h\dot{x}(t+h) = x(t+h) - x(t) + h^2 \sum_{i=1}^{k+1} \beta_i f(t + [l-i]h),$$

где x — координаты атома; t — текущий временной шаг; h — шаг по времени; k — показатель глубины схемы по временным шагам; α_i, β_i, l — коэффициенты схемы для определенного k и шага (предиктор или корректор). Метод заключается в первоначальном вычислении значений координат и скоростей и их последующей корректировке. Данная схема имеет локальную ошибку порядка $O(h^{k+1})$.

Аналогично описанному выше двухэтапному методу вычисления поступательного движения молекул, вычисляется и их вращение. В этом случае в качестве координат частиц выступают кватернионы.

В качестве граничных условий использовались периодические граничные условия. В качестве начальных условий молекулы равномерно распределялись по кубу. Количество молекул выбиралось исходя из размеров куба и заданной плотности.

4. Методы ускорения

Для проведения физически достоверных расчетов необходимо использование большого числа молекул. В общем, для проведения расчетов требуется $O(N^2)$ операций, где N — число частиц (атомов). Ускорить расчет ближнего взаимодействия позволяет схема обрезания с использованием списка соседей или специализированной структуры данных, которая позволяет снизить число операций до порядка $O(N)$. Учитывая быстро убывающую природу $1/r^6$, эта схема удовлетворительна для потенциала Леннарда–Джонса.

Но использование схемы обрезания для расчета потенциала Кулона ведет к значительной нелинейности, так как величина потенциала убывает пропорционально $1/r$. Как результат, метод обрезания не подходит для расчета дальнедействующих взаимодействий и неэффективен для их моделирования.

Ускорение расчета дальнедействующих взаимодействий было достигнуто путем использования GPU. Реализация расчета ближних взаимодействий на GPU также позволяет значительно ускорить вычисления.

4.1. Структура данных

Использование структуры данных позволяет снизить вычислительную сложность всего алгоритма расчета ближнего взаимодействия с $O(N^2)$ до $O(NM)$, где N — общее число частиц; M — среднее число частиц в соседних боксах. Построение структуры данных основано на использовании гистограммы распределения частиц по боксам и bucket-сортировке частиц [2]. Вычислительная сложность алгоритма построения структуры данных равна $O(N)$. Таким образом, общая вычис-

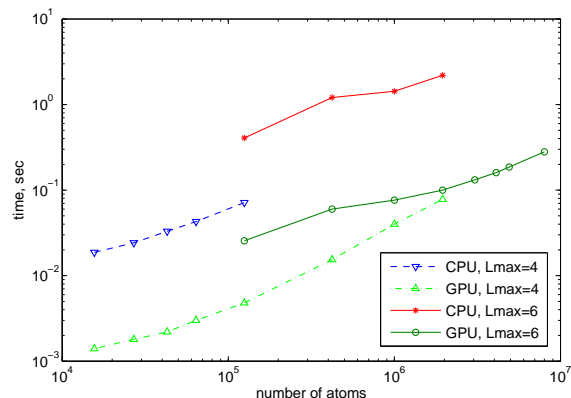


Рис. 3. Время генерации структуры данных в зависимости от числа частиц

лительная сложность снижается при оптимальном подборе числа боксов для данного N .

5. Результаты

Тестовые расчеты проводились на вычислительной системе с CPU Intel Xeon 5660, 2.8GHz, GPU NVidia Tesla C2075, операционной системой Linux 64bit, компиляторами GCC v.4.4, CUDA v.4.0. Расчеты проводились для чисел с плавающей точкой двойной точности. Размер блока при проведении расчетов выбирался исходя из оптимальности и, начиная с некоторого числа частиц, равнялся 256 потокам на блок. В качестве начального условия использовалось равномерное распределение частиц по моделируемой области.

На рис. 3 показано время генерации структуры данных на CPU и на GPU в зависимости от числа частиц для двух максимальных уровней структуры данных ($L_{max} = 4$ и $L_{max} = 6$). Ускорение, благодаря использованию GPU, составляет порядка 40 раз.

На рис. 4 показано время необходимое для расчета обрезанного потенциала Леннарда–Джонса. Реализация brute-force метода на GPU позволяет достичь ускорения порядка 300 по сравнению с реализацией на CPU. Использование структуры данных позволяет снизить вычислительную сложность с $O(N^2)$ до $O(N)$. Однако, если не увеличивать максимальный уровень структуры данных (L_{max}) с ростом числа частиц, то график перестает расти линейно и начинает стремиться к квадратичному росту в связи со значительным увеличением числа частиц в каждом боксе структуры данных, что можно наблюдать на рисунке в виде прерывистой линии ($L_{max} = 4$). Выбор максимального уровня структуры данных основан на оптимальном соотношении между временем генерации структуры

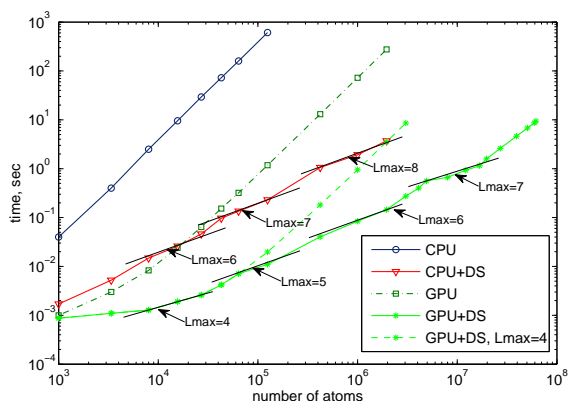


Рис. 4. Время расчета ближнего взаимодействия в зависимости от числа частиц

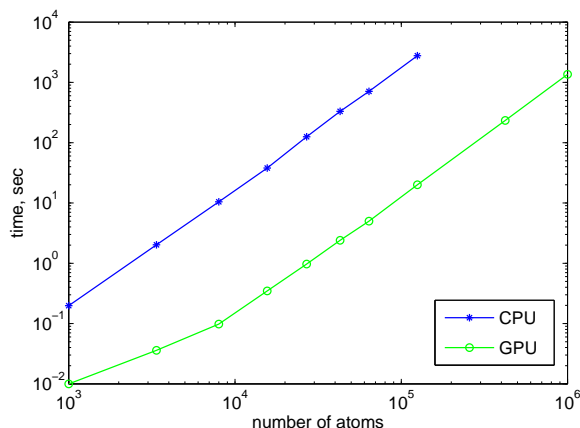


Рис. 5. Время расчета дальнего взаимодействия в зависимости от числа частиц

данных и временем расчета потенциала Леннарда–Джонса. Ввиду того, что ускорение на GPU генерации структуры данных меньше ускорения расчета потенциала Леннарда–Джонса, реализация расчета потенциала Леннарда–Джонса с использованием структуры данных на GPU уменьшает необходимое для расчетов время в 40 раз.

На рис. 5 показано время расчета потенциала Кулона на CPU и на GPU. Использование GPU позволяет ускорить расчет почти в 200 раз.

6. Заключение

Использование GPU позволяет значительно ускорить расчеты, в которых необходимо проводить ряд одинаковых операций над массивом данных, и сделать возможным проведение моделирования больших систем на персональных суперкомпьютерных станциях, оснащенных GPU. GPU ускоряет генерацию структуры данных в 40 раз, расчет

потенциала Леннарда–Джонса — в 300 раз и расчет потенциала Кулона — в 200 раз. Использование структуры данных для расчета обрезанного потенциала Леннарда–Джонса позволяет снизить вычислительную сложность с $O(N^2)$ до $O(N)$.

Дальнейшим этапом является применение алгоритмических методов ускорения расчета потенциала Кулона. Одним из таких методов является быстрый метод мультиполей (Fast Multipole Method, FMM), который позволяет снизить вычислительную сложность расчета далекодействующих взаимодействий с $O(N^2)$ до $O(N)$.

Список литературы

- [1] Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., Klein M.L. Comparison of simple potential functions for simulating liquid water // J. Chem. Phys. 1983. Vol. 79. P. 926–935.
- [2] Hu Q., Gumerov N.A. and Duraiswami R. Scalable fast multipole methods on distributed heterogeneous architectures // SC'11, International Conference for High Performance Computing, Networking, Storage, and Analysis, Seattle, WA, November 12–18, 2011.
- [3] Rapaport D.C. The art of molecular dynamics simulation. Cambridge University Press, 2004. P. 400.
- [4] NVIDIA Corporation. NVIDIA CUDA Compute Unified Device Architecture Programming Guide. Version 4.0. 2011.