

ИСПОЛЬЗОВАНИЕ ГИБРИДНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ ДЛЯ ЗАДАЧ МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОГО МОДЕЛИРОВАНИЯ НЕИДЕАЛЬНОЙ ПЛАЗМЫ

Р.Г. Быстрый, И.В. Морозов

1. Введение

Неидеальная электрон-ионная плазма образуется при электровзрыве проводников, в прикатодной области высоковольтных разрядов, на стенках ускорителей под воздействием сильных электромагнитных полей, в твердых телах под действием мощных пучков частиц и др. Важным случаем неидеальной плазмы является так называемая лазерная плазма, т.е. плазма образующаяся при воздействии интенсивных лазерных импульсов фемто- и пикосекундной длительности на металлические мишени. Исследования лазерной плазмы имеют широкое применение в микроэлектронике (напыление тонких пленок, микрошлифовка), в рентгеновской микроскопии и микролитографии, а также для создания новых источников энергии (генерация рентгеновского излучения для инициирования термоядерных реакций).

Для теоретического исследования неидеальной плазмы широко применяется метод молекулярной динамики (МД). Он позволяет на уровне движения электронов и ионов исследовать элементарные процессы в плазме, такие как плазменные волны, электрон-ионная релаксация, взаимодействие излучения с плазмой и др. [1]. В отличие от метода Монте-Карло, МД может применяться для описания неравновесных систем. В частности, он успешно применяется для описания неидеальной плазмы, образующейся при ионизации металлических нанокластеров короткими лазерными импульсами [2]. Число атомов в одном кластере как правило невелико (до 1млн.), что позволяет напрямую моделировать динамику электронов и ионов, учитывая взаимодействия отдельных частиц друг с другом, без применения приближения сплошной среды.

В то же время метод МД предъявляет высокие требования к производительности вычислительных систем и эффективности распараллеливания программ моделирования. В наибольшей степени на скорость работы программ влияют число частиц в системе (размер моделируемой области) и объем статистического усреднения.

Следует отметить, что статистическое усреднение является неотъемлемой частью моделирования неравновесных сред. Согласно подходу, изложенному в [3], расчет любых релаксационных процессов, заключается в генерации набора МД траекторий из начальных условий, которые идентичны с точки зрения макроскопических параметров, однако представляют собой различные статистически независимые микросостояния. Результаты анализа каждой из данных траекторий усредняются для получения окончательного результата. Это дает возможность получить статистически достоверный результат для случайных во времени событий, таких как фазовый переход из метастабильного состояния. То же относится и к равновесным системам, где в соответствии с эргодической теоремой усреднение по времени заменяется на усреднение по ансамблю.

Статистическое усреднение дает возможность простого распараллеливания программ, поскольку расчет траекторий для различных начальных микросостояний может производиться независимо. Требования к коммуникационной сети для соответствующей параллельной вычислительной системы являются, очевидно, минимальными, поэтому данная задача идеально подходит не только для традиционных кластеров, но и для кластеров на гибридной архитектуре, а также для Грид-систем. Кроме того, статистическое усреднение позволяет уменьшить требования к надежности системы, так как сбой оборудования на одном из вычислительных узлов, лишь несильно уменьшит точность результата, но не приведет к полной остановке программы. В случае использования гибридного кластера предполагается, что расчет каждой траектории производится на отдельном графическом ускорителе.

На более низком уровне возможно распараллеливание расчета отдельной траектории методами декомпозиции по пространству (domain decomposition) или декомпозиции по частицам (force decomposition). Здесь эффективность напрямую зависит от скорости обмена между процессами, поэтому этот тип распараллеливания должен применяться на кластерах с высокоскоростной внутренней сетью. Альтернативой здесь является применение графических ускорителей, о которых далее пойдет речь.

Среди современных высокопроизводительных систем в последнее время все чаще используются гибридные системы, включающие графические ускорители (ГПУ – графическое процессорное устройство). Этому во многом способствовало появление сред программирования, таких как CUDA и OpenCL, предназначенных для создания и выполнения на ГПУ программ, не связанных с обработкой изображений. Гибридные системы на основе ГПУ зачастую показывают быстроедействие, в десятки или сотни раз превышающее быстроедействие систем на традиционных процессорах (ЦПУ – центральной процессорное устройство). Следует, однако, отметить, что создание программ, эффективно использующих вычислительные возможности ГПУ, это достаточно трудоемкий процесс, требующий учета специфики аппаратной

архитектуры, без которого производительность на конкретной задаче может оказаться существенно меньше теоретического предела для данного устройства.

Одним из условий использования ГПУ является хорошая распараллеливаемость программы в рамках модели общей памяти. Метод МД отвечает этому требованию, и за последние несколько лет появились пакеты программ, позволяющие проводить МД моделирование на ГПУ, такие как LAMMPS [4], NAMD/VMD [5], HOOMD [6] и др. Эта область активно развивается, однако в настоящее время набор потенциалов взаимодействия частиц, реализованных в указанных пакетах на ГПУ, очень невелик и ограничивается, как правило, простейшими потенциалами типа Леннарда-Джонса или гармонической связи между атомами в молекуле. В то же время для моделирования динамики электронов, в ионизированных лазером кластерах, требуются более сложные потенциалы, в частности, потенциал электрон-ионного взаимодействия с квантовыми поправками на близких расстояниях, использованный в работе [2]. Опыт реализации данного потенциала на ГПУ описывается в настоящей работе.

Мы приводим также анализ производительности пакетов LAMMPS и HOOMD, как наиболее подходящих для задач моделирования конденсированных сред. Пакет LAMMPS имеет достаточно большую историю и успешно применялся для моделирования конденсированного состояния вещества [7]. Его преимуществом является хорошая эффективность распараллеливания на традиционных кластерах с большим количеством ЦПУ. С 2009г. в него включена возможность проводить вычисления на ГПУ для нескольких простых потенциалов взаимодействия.

С другой стороны, HOOMD — это сравнительно новая программа, изначально разработанная для применения на ГПУ. Она основана на технологии Nvidia CUDA с применением специальной оптимизации для графических ускорителей Nvidia. Основные этапы алгоритма МД, такие как построение списка ближайших соседей, расчет взаимодействия между частицами и интегрирование уравнений движения, выполняется на ГПУ без копирования промежуточных данных на ЦПУ. Это позволяет уменьшить обмен данными между ГПУ и ЦПУ по шине PCI Express, гораздо более медленной чем внутренняя шина памяти ГПУ.

Во втором разделе обсуждаются тесты производительности LAMMPS и HOOMD для ГПУ и ЦПУ на примере моделирования равновесной леннарда-джонсоновской жидкости. Третий раздел посвящен особенностям реализации потенциала электрон-ионного взаимодействия на ГПУ. В четвертом разделе обсуждаются результаты тестирования быстродействия полученных программ моделирования.

2. Производительность гибридных и традиционных систем для потенциала Леннарда-Джонса

В качестве основы для создания программ был выбран свободно распространяемый пакет МД моделирования HOOMD (Highly Optimized Object-oriented Many-particle Dynamics). Этот пакет изначально создавался для использования на гибридных вычислительных системах с видеоускорителями Nvidia и включает ряд оптимизационных решений с учетом архитектуры ГПУ. Для сравнения использовался также свободно распространяемый пакет LAMMPS.

Для оценки эффективности применения ГПУ для задач МД было проведено тестовое моделирование равновесной леннарда-джонсоновской (LJ) жидкости при температуре $T = 1.0$, плотности $\rho = 0.19$, отсечке потенциала $r_{cut} = 3$ (в единицах LJ). Предварительные результаты были опубликованы в [8].

Результаты для гибридной системы, включающей два ЦПУ Intel Xeon E5520 и два ГПУ Nvidia GeForce 480GTX, показаны на рис. 1. Видно, что при увеличении числа частиц ГПУ становится все более эффективным за счет более полной загрузки параллельных вычислительных блоков. На рис. 2 показано, что максимальное отношение производительности ГПУ к двум ЦПУ Intel Xeon E5520 (750%) достигается на $N_{part} = 5 \cdot 10^5$ частицах при использовании ГПУ Nvidia GeForce 480GTX. При этом пакет HOOMD показывает примерно в 4 раза большую производительность, чем пакет LAMMPS, при максимальном числе частиц.

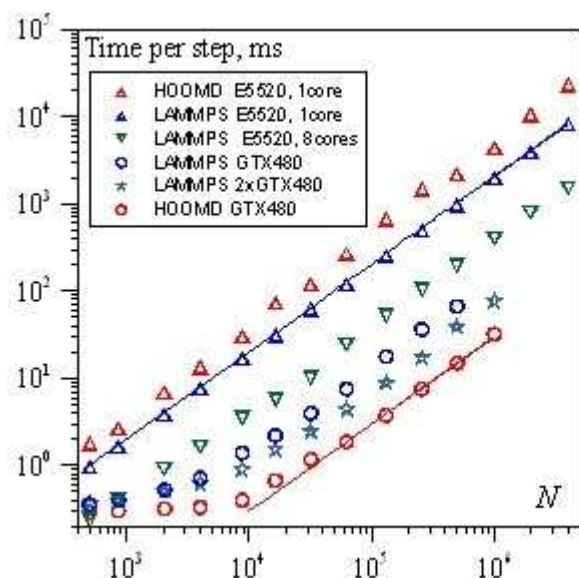


Рис. 1 Зависимость времени выполнения одного шага молекулярной динамики от количества частиц в системе для различных устройств и пакетом моделирования (показано на графике).

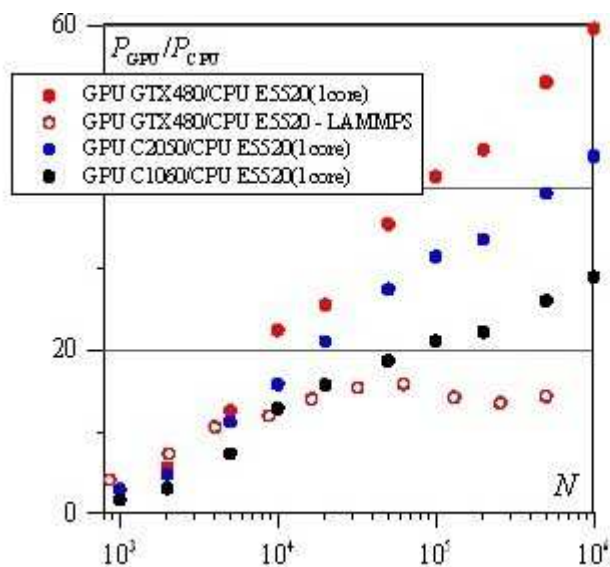


Рис. 2 Отношение производительности различных ГПУ и одного ядра ЦПУ Intel Xeon E5520 для различного числа частиц в МД ячейке (потенциал Леннарда-Джонса). Пустые точки – пакет LAMMPS, закрашенные – HOOMD.

3. Особенности реализации потенциал электрон-ионного взаимодействия на ГПУ

Потенциал электрон-ионного взаимодействия обладает рядом особенностей по сравнению с потенциалом Леннарда-Джонса, поэтому его адаптация для ГПУ потребовала существенной переработки функции расчета взаимодействий частиц, имевшейся в HOOMD.

Потенциал взаимодействия двух частиц зависит только от расстояния между ними. Сила определяется как минус градиент потенциала. Сила и потенциал включают сложные с вычислительной точки зрения функции $e^{-\alpha r}$ и \exp , поэтому для них были использованы интерполяционные таблицы. С учетом аппаратной поддержки линейной интерполяции в ГПУ, это позволило существенно ускорить вычисления.

Основной особенностью потенциала электрон-ионного взаимодействия является то, что он очень медленно убывает с ростом расстояния между частицами. Такие потенциалы называются дальнедействующими. Для систем с дальнедействующим потенциалом неприменим метод списка ближайших соседей, используемый традиционно для короткодействующих потенциалов. Это делает задачу вычислительно тяжелой даже для сравнительно небольшого числа частиц в системе. Кроме того,

необходимость знать координаты всех частиц системы для вычисления силы, действующей на одну частицу, делает неэффективным стандартный метод распараллеливания МД программ (декомпозиция по пространству), применяемый на системах с распределенной памятью. В этом случае ГПУ с общей памятью для всех потоков оказывается существенно более эффективным.

При выполнении программы на ГПУ выполняется декомпозиция по частицам, т.е. каждый поток вычисляет потенциальную энергию и силу для одной частицы. Таким образом распараллеливается неявный цикл по частицам. Особенно важным для быстродействия программы был выбор типа памяти, в которой хранятся интерполяционные таблицы. Было реализовано два варианта программы. Первый хранит массив частиц в текстурной памяти, а второй использует глобальную память. В версии с текстурной памятью использовалась аппаратная интерполяция, в версии с глобальной памятью интерполяция не использовалась. Сравнение их быстродействия будет рассмотрено ниже.

4. Производительность гибридных и традиционных систем для потенциала электрон-ионного взаимодействия

Для оценки производительности различных реализаций потенциала электрон-ионного взаимодействия было проведено тестовое моделирование неидеальной плазмы, полученной в результате ионизации кластеров натрия, содержащих 50-20000 атомов. При проведении тестирования использовалась та же вычислительная система, что и при проведении тестов для потенциала Леннарда-Джонса. Результаты для гибридной системы, показаны на рис. 3. Видно, что разница в производительности версии с глобальной памятью и версии с текстурной – незначительна. На рис. 3 показано, что отношение производительности ГПУ Nvidia GeForce 480GTX к 1 ядру ЦПУ Intel Xeon E5520 достигает 180 раз при максимальном числе частиц.

Следует отметить, что в данных тестах моделировалась только динамика электронов, а ионы считались неподвижными вследствие большого отношения масс. Такая постановка численного эксперимента оправдана, если речь идет об исследовании электронных колебаний, важных для расчета взаимодействия излучения с ионизованными кластерами.

Кроме указанной тестовой системы расчеты также проводились на кластере K100 ИПМ РАН, включающем 192 ГПУ Nvidia Tesla C2050. При этом на каждом ГПУ запускался расчет отдельной МД траектории с различным пространственным расположением ионов. Затем результаты усреднялись. В результате были получены новые данные о зависимости частоты и затухания основных мод колебаний электронов в ионизованном кластере в зависимости от размера кластера (результаты будут опубликованы отдельно). В отличие от расчетов на ЦПУ, где максимальное число частиц составило 1000, применение ГПУ позволило в 200 раз увеличить размер моделируемой системы и исследовать переход от системы с существенными размерными эффектами (кластер с 50 атомами) к практически однородной неидеальной плазме (крупные кластеры).

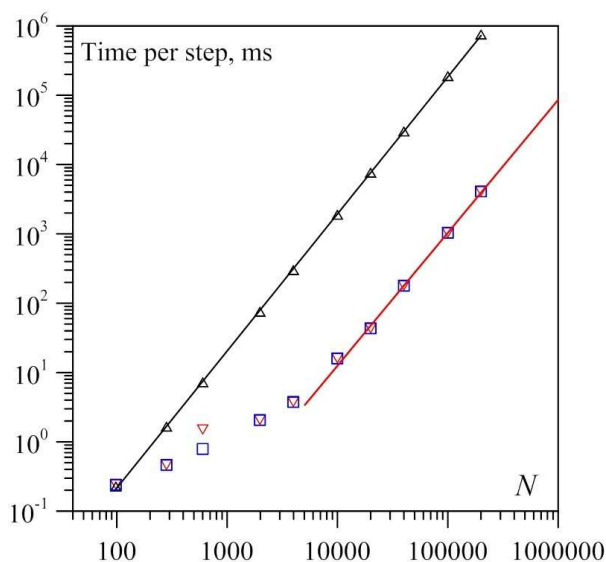


Рис.3 Зависимость времени выполнения одного шага молекулярной динамики от количества частиц в системе. Черные треугольники – расчет на одном ядре ЦПУ Intel Xeon E5520; красные треугольники – ГПУ Nvidia GeForce 480GTX, версия программы, использующая текстурную память; синие квадратики – ГПУ, версия программы, использующая глобальную память. Прямые – степенная интерполяция.

Работа поддержана по ПФИ ПРАН №13 и №14, а также по гранту Президента РФ МК-64941.2010.8 (Морозов И.В.).

ЛИТЕРАТУРА:

1. Морозов И.В., Норман Г.Э. Столкновения и плазменные волны в неидеальной плазме // ЖЭТФ. 2005. Т. 127. №2. С. 412–430.
2. T. Raitza, H. Reinholz, G. Röpke, I. Morozov, E. Suraud. Laser excited expanding small clusters: Single time distribution functions // Contributions to Plasma Physics. 2009. V. 49. P. 496-506.
3. Kuksin A.Yu., Morozov I.V., Norman G.E., Stegailov V.V., Valuev I.A. Standards for Molecular Dynamics Modeling and Simulation of Relaxation // Molecular Simulation. 2005. V. 31. № 14–15. P. 1005-1017.
4. Plimpton S.J. Fast Parallel Algorithms for Short-Range Molecular Dynamics // Journal of Computational Physics. 1995. V. 117. P. 1.
5. Stone J.E., Hardy D.J., Ufimtsev I.S., Schulten K. GPU-accelerated molecular modeling coming of age // J. Mol. Graph. Model. 2010. V. 29. P. 116.
6. Anderson J.A., Lorenz C.D., Travesset A. General purpose molecular dynamics simulations fully implemented on graphics processing units // Journal of Computational Physics. 2008. V. 227. P. 5342.
7. Янилкин А.В., Жиляев П.А., Куксин А.Ю., Норман Г.Э., Писарев В.В., Стегайлов В.В. Применение суперкомпьютеров для молекулярно-динамического моделирования процессов в конденсированных средах // Вычислительные методы и программирование. 2010. Т. 11. С. 111.
8. I.V. Morozov, A.M. Kazennov, R.G. Bystryi, G.E. Norman, V.V. Pisarev, V.V. Stegailov. Molecular dynamics simulations of the relaxation processes in the condensed matter on GPUs // Computer Physics Communications. 2011. V. 182. P. 1974–1978.