

Масштабируемые вычисления с применением гибридного кластера *

Б.М. Глинский, Д.А. Караваев, И.М. Куликов,
Н.В. Кучин, Н.В. Снытников

Федеральное бюджетное учреждение науки

Институт вычислительной математики и математической геофизики СО РАН

e-mail: gbm@opg.sccc.ru, kda@opg.sccc.ru, kulikov@ssd.sccc.ru,
kuchin@sccc.ru, nik@ssd.sccc.ru

Рассматриваются вычислительные технологии моделирования астрофизических процессов и сейсмических полей в неоднородных упругих средах на гибридном кластере Сибирского суперкомпьютерного центра. Такие задачи отличаются трудоемкостью численной и программной реализации в системах с высокой степенью параллелизма. Особенность реализации связана с архитектурой кластера, в состав которого входят гибридные вычислительные модули с CPU и GPU, а также с технологией совместного применения MPI и CUDA. Исследуются возможности масштабирования этих задач при их решении на гибридном кластере.

1. Введение

Проблема исследования масштабируемости параллельных алгоритмов при их реализации на современных многоядерных компьютерах и будущих суперЭВМ эксафлопсной производительности выходит за уровень технологических задач и требует научно-исследовательского подхода к ее решению. Вычислительные алгоритмы, как правило, являются более консервативными по сравнению с развитием средств вычислительной техники. В частности, с каждым днем требования к астрофизическим моделям всё возрастают и модели, которые несколько лет назад были актуальными, сейчас уже считаются устаревшими. Усложнение астрофизических моделей требует использования всё больших вычислительных ресурсов, а, следовательно, модификации и создания новых вычислительных методов и параллельных алгоритмов для решения таких задач. Большое количество программных реализаций и численных методов говорит об актуальности исследований в области разработки новых методов и их программных реализаций для решения задач астрофизики. Кроме того, несмотря на развитие астрофизических пакетов в сторону петафлопсных вычислений, таких как PetaGADGET [1], Enzo-P [2], PetaART [3] нужно отметить фундаментальные ограничения в масштабируемости AMR и SPH подходов, которые используются в основном числе программных пакетов для решения задач астрофизики [4,5]. Для моделирования некоторых астрофизических задач и задач физики плазмы необходимо решать систему уравнений, состоящую из уравнения Власова и уравнения Пуассона, на подробной сетке и на больших временных масштабах. Моделирование распространения упругих волн в 3D средах с

*При частичной финансовой поддержке грантов РФФИ № 12-01-31352 мол-а, 13-07-00589-а, 12-07-00065-а, 13-01-00231-а, 11-05-00937; МИП № 39 СО РАН, МИП № 130 СО РАН; гранта Президента РФ МК-4183.2013.9; Программы Президиума РАН Проект №15.9

целью построения трехмерных геометрий строения геофизических объектов является также очень трудоемкой вычислительной задачей.

2. Программный комплекс для моделирования взаимодействующих галактик GPUPEGAS

Задача сводится к численному решению газодинамических уравнений и уравнения Пуассона с помощью специально адаптированной для реализации на множестве графических ускорителей комбинации метода крупных частиц и метода Годунова. Будем рассматривать 3-х мерную модель динамики самогравитирующего газа в декартовых координатах, включающих в себя расширенную систему уравнений гравитационной газовой динамики в дивергентной форме, замкнутую уравнением состояния для идеального газа. За основу метода решения системы уравнений газовой динамики выбран метод крупных частиц [6], уже хорошо зарекомендовавший себя в ходе решения астрофизических задач [7]. В основе параллельной реализации решения гидродинамических уравнений лежит трехмерная декомпозиция расчетной области. По одной координате разрезание происходит средствами технологии MPI, по двум другим с использованием технологии CUDA (см. рис.1).

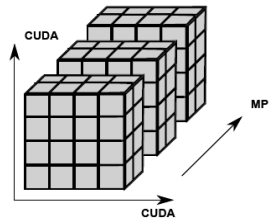


Рис. 1. Декомпозиция расчетной области для решения гидродинамических уравнений.

Это связано с топологией и архитектурой гибридного суперЭВМ NKS-30T+GPU Сибирского суперкомпьютерного центра ИВМиМГ СО РАН, который был использован для вычислительных экспериментов. Модификация численного метода решения гидродинамических уравнений позволяет на каждом этапе численного метода независимо вычислять значения потоков через каждую ячейку. Декомпозиция области на каждом этапе осуществляется с перекрытием одного слоя граничных точек соседних областей. Стоит отметить, что такая модификация метода не ограничивается только архитектурой CUDA, она также легко применима для технологии Intel MIC. Также может быть использована и другая сетевая топология. В случае использования гибридной реализации необходимо определить три понятия масштабируемости. SingleGPU performance (сильная масштабируемость в рамках одного GPU) – уменьшение времени счета одного шага одной и той же задачи при использовании большего числа графических ядер. MultiGPU performance (слабая масштабируемость при использовании многих GPU) – сохранения времени счета одного шага одного и того же объема задачи при одновременном увеличении количества GPU-ускорителей. FFTW performance (сильная масштабируемость при использовании библиотеки FFTW) – уменьшение времени счета одного шага одной и той же задачи при использовании большего числа процессоров или ядер. Результаты эффективности программной реализации приведены на рисунке 2.

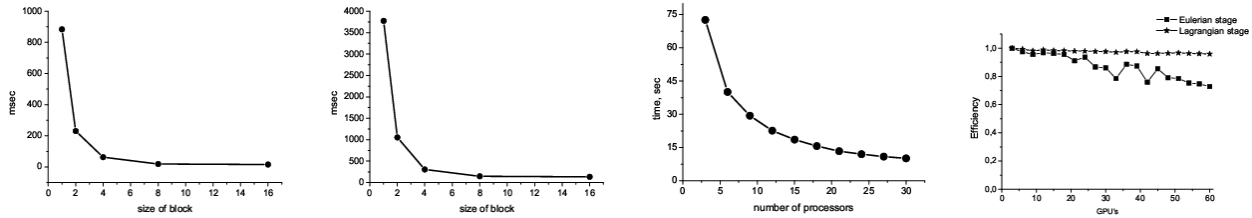


Рис. 2. Время счета эйлерова и лагранжевого этапов на одном графическом ускорителе. Время счета уравнения Пуассона в зависимости от использованных ядер. Эффективность параллельной реализации каждого из этапов в зависимости от использованных GPU.

Основным предназначением разработанного программного комплекса GPUPEGAS является моделирование столкновения галактик различных типов и под различными углами. В качестве модельной задачи рассмотрим задачу столкновения дисковых галактик под углом. Первое облако задается сферической областью, равномерно заполненной газом массы $M_{gas} = 16 \cdot 10^{41}$ кг, второе облако задается эллипсоидальной в отношении осей 1:2:1, наклоненный под углом 45 градусов к оси столкновения. В ходе встречного движения облаков с равными начальными скоростями $v_{cr} = 600$ км/с происходит их столкновение.

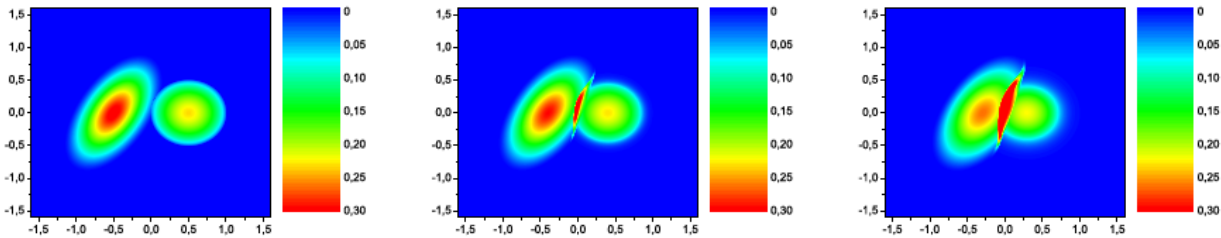


Рис. 3. Безразмерная плотность газовых облаков в моменты времени: начальная конфигурация (слева), $1 \cdot 10^{14}$ секунд (посередине), $2 \cdot 10^{14}$ секунд (справа).

3. Моделирование распространения упругих волн в трехмерных неоднородных средах с применением построителя 3D моделей

Численное моделирование распространения упругих волн в 3D изотропной неоднородной упругой среде [8] является трудоемкой вычислительной задачей. Одним из подходов решения является разностный метод [9]. Используемый параллельный алгоритм связан с применением сеточного метода. Построением трехмерных геометрий строения геофизических объектов и определения параметров упругой среды связано с решением набора прямых геофизических задач. Варьируя геофизические параметры и, соответственно, проводя расчеты по выбранным численным алгоритмам можно создать геометрическую модель изучаемых объектов. Для детальных исследований моделируемая упругая среда представляется наиболее подробно, что отражается на применяемой сетке. Поэтому эксперименты по численному моделированию проводятся на суперкомпьютерах, которые могут быть MPP архитектурой или гибридной. Из проведенных экспериментов применительно к задачам геофизики использование графических процессоров от Nvidia

показало их высокую эффективность. Для построения специфических 3D геометрий геофизических объектов разработана специализированная программа. Программа разработана для кластеров с использованием технологий параллельного программирования и позволяет задавать упругие параметры в каждой точке 3D сетки. Программа имеет достаточно гибкий способ задания различных сложных 3D геометрий, используя набор примитивов, объектов конструирования. Следует отметить, что эта программа используется в комплексе с программой численного моделирования. Таким образом, 3D геофизическая структура с определенными значениями упругих параметров создается в оперативной памяти параллельно и непосредственно на каждом из используемых вычислительных устройств. Основываясь на архитектуре гибридного кластера NKS-30T+GPU ССКЦ СО РАН, где вычислительные узлы состоят из нескольких CPU и GPU - графические процессоры, была выбрана реализация параллельного алгоритма, рис.1. Вычислительная область делится на подобласти, слои, вдоль оси Oz. Необходимые параметры для вычислений инициализированы на CPU и копируются на GPU. Использовалась одномерная топология, таким образом, одному MPI процессу соответствует один GPU, или один CPU. Вычисления по численному алгоритму для каждой из подобластей проводятся независимо на выделенном GPU с использованием CUDA в 2D. Поскольку у соседних GPU есть общие данные, расположенные в плоскости Oxy, то между ними реализованы обмены данными посредством CPU с использованием MPI. Разработана программа для численного моделирования 3D сейсмических полей с использованием комбинации MPI и CUDA. Времена счета в зависимости от количества доступных ядер для двух программных реализаций: используется только CPU и MPI; используется комбинация CPU и GPU, т.е. MPI и CUDA приведены ниже. На гибридном кластере NKS-30T+GPU проводились эксперименты по масштабированию алгоритма, когда область вычислений увеличивается пропорционально количеству вычислителей, и по изучению поведения программы при увеличении количества вычислительных ядер для зафиксированной области. Результаты представлены на рис.4, рис.5.

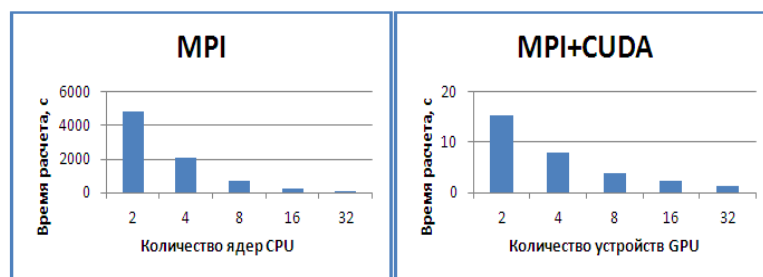


Рис. 4. Изменение времени расчета алгоритма численного моделирования в зависимости от числа вычислительных ядер CPU и количества устройств GPU.

4. Моделирование астрофизических задач и задач физики плазмы

Для моделирования некоторых астрофизических задач и задач физики плазмы необходимо решать систему уравнений, состоящую из уравнения Власова и уравнения Пуассона, на подробной сетке и на больших временных масштабах. Фактически, для проведения серийных численных экспериментов для нестационарной задачи с 1 миллиардом

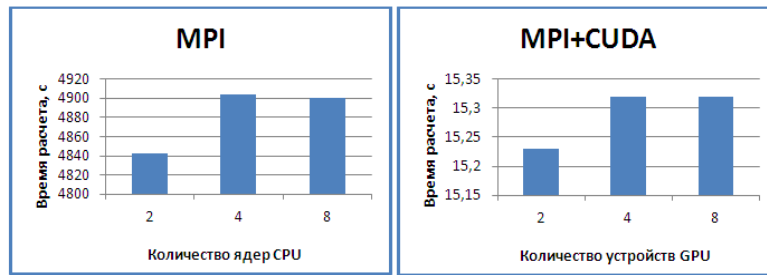


Рис. 5. Изменение времени расчета алгоритма численного моделирования при масштабировании.

сеточных узлов (сетка $1000 \times 1000 \times 1000$) и 100 миллиардов модельных частиц, необходимых для адекватного воспроизведения 6-мерной функции распределения вещества, один временной шаг должен обрабатываться за абсолютное время не больше 1 секунды. Учитывая, что для прослеживания траектории одной частицы требуется порядка 100 арифметических действий с плавающей точкой, то ясно, что эта задача требует использования компьютеров эксафлопсного класса. При этом создаваемые параллельные алгоритмы должны эффективно масштабироваться для использования на сотнях тысяч и миллионах вычислительных элементов (процессоров или ядер).

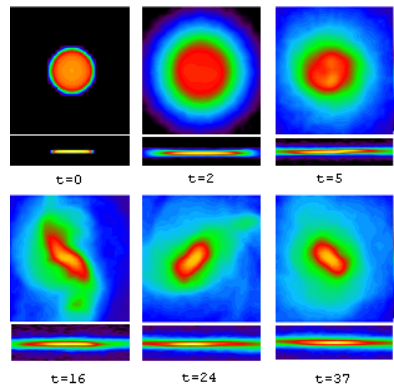


Рис. 6. Типичный вычислительный эксперимент: трехмерное моделирование звездной компоненты вращающейся дисковой галактики с образованием барообразной неустойчивости и изгибных деформаций. Выполнено 37 тыс. временных шагов.

Существующие параллельные алгоритмы используют подходы адаптивного измельчения сетки [10]; декомпозицию области, основанную на представлении гравитационного потенциала в виде суммы близкодействующей и далекодействующей частей [11]; алгоритмы решения СЛАУ, получаемых после аппроксимации уравнения Пуассона [12, 13]. Тем не менее, все они имеют определенные ограничения по масштабируемости для трехмерных задач, решаемых на тысячах процессоров или ядер. Мы предлагаем новый метод декомпозиции области для решения нестационарных задач, включающих в себя уравнение Пуассона и уравнение Власова. Метод основан на частичном предвычислении функции Грина для заданных сеточной функции потенциала и соответствующих краевых условий и методе сопряжения подобластей, предложенном в работе [14]. Общая схема метода для суперкомпьютеров с распределенной памятью описана в работе [15]. Далее приведем схему реализации этого алгоритма для суперкомпьютеров с ги-

бридной архитектурой, состоящих из CPU и GPU. Пример расчета приведен на рис.6. Поскольку каждой ячейке (узлу) сетки соответствует около 100 частиц, то основная вычислительная трудоемкость сосредоточена в интегрировании уравнений движения частиц, в то время как решение уравнения Пуассона занимает не более 1-2% от общей трудоемкости. В связи с этим оправдано хранить сеточные функции плотности и потенциала в памяти CPU, а хранение массивов с координатами и скоростями частиц в памяти GPU.

1. Выполняется параллельное решение уравнения Пуассона с помощью декомпозиции области.

1.1. Решение уравнения Пуассона с нулевыми граничными условиями. На этом шаге используется только CPU. Обмены данными между CPU отсутствуют.

1.2. Выполняется вычисление значений потенциала на границах подобластей и обмен значениями между соседними подобластями. Выполняется средствами MPI.

1.3. Решение уравнения Лапласа с вычисленными граничными условиями. Итоговое решение уравнения Пуассона равно сумме решений с шага 1.3 и 1.1.

2. Интегрирование уравнений движения частиц.

2.1. Передаем сеточные функции потенциала из памяти CPU в память GPU. Передаем массивы частиц, которые "переместились" из других подобластей, предназначенных для вычислений на других CPU. Выполняется средствами CUDA.

2.2. Вычисляем координаты частиц на новом шаге по времени (используя массивы координат частиц с предыдущего шага, хранящиеся в памяти GPU). Вычисляем сеточную функцию плотности. Вычисления выполняются параллельно, используется максимальное количество ядер GPU. Выполняется средствами CUDA.

2.3. Передаем сеточные функции плотности из памяти GPU в память CPU. Передаем массивы частиц, которые "перелетели" в соседние подобласти, предназначенные для вычислений на других CPU. Выполняется средствами CUDA.

2.4. Передаем массивы с "переместившимися" частицами в соседние подобласти (соответствующим CPU). Выполняется средствами MPI.

3. Переход на следующий временной шаг. Переход к этапу алгоритма 1.1.

5. Заключение

Для класса нестационарных задач гравитационной газовой динамики, описан новый вычислительный алгоритм на основе метода крупных частиц и метода Годунова, позволяющий проводить вычислительные эксперименты по изучению динамики самогравитирующего газа в трёхмерной постановке в широком диапазоне параметров. На основе предложенных алгоритмов создан программный комплекс GPUPEGAS для гибридных суперЭВМ, оснащенных графическими ускорителями. Проведенные эксперименты для двух программных реализаций алгоритма численного моделирования в геофизике показали, что алгоритм хорошо масштабируется и использование GPU дает значительный эффект. Несмотря на то, что при вычислениях приходится копировать данные между CPU и GPU использование гибридного кластера имеет свои преимущества и может быть использовано для трехмерных разностных методов. Экспериментальное исследование алгоритма для решения нестационарных задач, включающих в себя уравнение Пуассона и уравнение Власова, проведенное в Сибирском Суперкомпьютерном Центре, основанное на последовательности запусков с различным числом вычислительных ядер (вплоть до 50 тысяч) показало его линейную масштабируемость. Также было проведено

теоретическое исследование алгоритма и его масштабируемости вплоть до 6 миллионов вычислительных ядер (нитей).

Список литературы

- [1] FENG Y. ET AL. Terapixel Imaging of Cosmological Simulations // The Astrophysical Journal Supplement Series. 2011. V. 197, 18.
- [2] <http://mnggrid.ucsd.edu/projects/cello/>
- [3] <http://www.cs.iit.edu/~zlan/petaart.html>
- [4] FERRARI A. ET AL. A New Parallel SPH Method for 3D Free Surface Flows // High performance computing on vector systems 2009. 2010. Part 4. P. 179-188.
- [5] VAN STRAALLEN B. ET AL. Scalability challenges for massively parallel AMR applications // In Proceedings of the 2009 IEEE International Symposium on Parallel & Distributed Processing (IPDPS '09). 2009. IEEE Computer Society, Washington, DC, USA. P. 1-12.
- [6] VSHIVKOV V., LAZAREVA G., SNYTNIKOV A., KULIKOV I., TUTUKOV A. Hydrodynamical code for numerical simulation of the gas components of colliding galaxies // The Astrophysical Journal Supplement Series. 2011. V. 194, 47. 12 pp.
- [7] 7. ТУТУКОВ А.В., ЛАЗАРЕВА Г.Г., КУЛИКОВ И.М. Газодинамика центрального столкновения двух галактик: слияние, разрушение, пролет, образование новой галактики // *Астрономический журнал*. 2011. Т. 88, № 9. С. 837-851.
- [8] ГЛИНСКИЙ Б.М., КАРАВАЕВ Д.А., КОВАЛЕВСКИЙ В.В., МАРТЫНОВ В.Н. Численное моделирование и экспериментальные исследования грязевого вулкана "Гора Карабетова" вибросейсмическими методами. // *Вычислительные методы и программирование*. М.: Изд-во Моск. Гос. ун-та, 2010, Том 11, №1, С. 99-108.
- [9] VIRIEUX J. P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method // *Geophysics*, Volume 51, April 1986, Number 4, p. 889-901
- [10] HUANG J., GREENGARD L. A Fast Direct Solver for Elliptic Partial Differential Equations On Adaptively Refined Meshes // *SIAM J. Sci. Comput.* 2000. Vol.21. P.1551-1566.
- [11] BALLS G.T., COLELLA P. A Finite Difference Domain Decomposition Method Using Local Corrections for the Solution of Poisson's Equation // *J. Comp. Physics*. 2002. Vol.180. P.25-53.
- [12] ТЕРЕКHOV A.V. Parallel Dichotomy Algorithm for Solving Tridiagonal System of Linear Equations with Multiple Right-Hand Sides // *Parallel Computing*. 2010. Vol.36. P.423-438.
- [13] ЯНЕНКО Н.Н., КОНОВАЛОВ А.Н., БУГРОВ А.Н., ШУСТОВ Г.В. Об организации параллельных вычислений и "распараллеливании" прогонки // *Численные методы механики сплошной среды*. 1978. Т.9. С.139-146.
- [14] JAMES R.A. The Solution of Poisson's equation for Isolated Source Distributions // *J. Comp. Physics*. 1977. 25, pp.71-93.
- [15] N. SNYTNIKOV Scalable Parallel Algorithm for Solving the Collisionless Boltzmann-Poisson System of Equations // *Astronomical Society of the Pacific Conference Series*. 2012. V. 453. P. 393.