

# СО-ДИЗАЙН ПАРАЛЛЕЛЬНЫХ ЧИСЛЕННЫХ МЕТОДОВ В ЗАДАЧАХ ФИЗИКИ ПЛАЗМЫ И АСТРОФИЗИКИ

Б.М. Глинский<sup>1,2</sup>, И.М. Куликов<sup>1,2,3</sup>, А.В. Снытников<sup>1,2</sup>, И.Г. Черных<sup>1,2</sup>, В.А. Вшивков<sup>1,2</sup>

<sup>1</sup> *Институт вычислительной математики и математической геофизики СО РАН*

<sup>2</sup> *Новосибирский государственный университет*

<sup>3</sup> *Новосибирский государственный технический университет*

## Введение

Основной вопрос в моделировании плазмы заключается в том, чтобы правильно воспроизводить взаимодействие частиц с волнами, лежащее в основе плазменных неустойчивостей. Для адекватного моделирования плазменных неустойчивостей необходимо иметь не менее 16 узлов сетки на длине дебаевского радиуса. Число модельных частиц не может быть малым (менее 100 в каждой ячейке), иначе полученные результаты будут носить лишь приближенный характер в известных физических случаях, в новых же физических случаях вовсе ничего невозможно будет получить [1]. При этом задача является существенно трехмерной.

Конкретная задача заключается в моделировании возбуждения ленгмюровских волн релятивистским пучком заряженных частиц на установке ГОЛ-3 (ИЯФ СО РАН). Следует отметить, что исследование взаимодействия мощного пучка с плазмой актуально и во многих других задачах физики плазмы, в частности, в процессах «быстрого поджига» в установках инерциального синтеза и при изучении вспышек в солнечной короне.

Движение галактик в плотных скоплениях превращает столкновения между ними в важный эволюционный фактор, поскольку за хаббловское время рядовая галактика может испытать до десятка столкновений с другими галактиками своего скопления [2]. Математическое моделирование играет более чем важную роль в теоретическом исследовании таких процессов. Одной из важных вычислительных проблем в таких задачах является соотношение масштабов. Так линейный размер галактики составляет порядка  $10^4$  парсек, а линейный размер звезды составляет величину порядка  $10^{-7}$  парсек.

Поэтому для моделирования таких процессов в высоком разрешении необходимо использовать суперкомпьютеры из верхней части списка Top-500. Два из первых трёх (четыре из первых десяти) суперкомпьютера в ноябрьской версии 2013 года списка Top-500 оснащены графическими ускорителями и ускорителями Intel Xeon Phi. Наиболее вероятно первый суперкомпьютер эксафлопсной производительности будет построен на основе гибридного подхода. На сегодняшний день уже существуют коды для решения задач физики плазмы, оптимизированные для выполнения на суперЭВМ, оснащенных графическими ускорителями и ускорителями Intel Xeon Phi [3, 4].

Однако, разработка кодов для гибридных суперкомпьютеров не сугубо техническая задача, а отдельная сложная научная задача, требующая со-дизайна алгоритмов на всех стадиях решения задачи – от физической постановки до инструментов разработки. В контексте математического моделирования физических процессов под понятием со-дизайна мы понимаем построение физико-математической модели явления, численного метода, параллельного алгоритма и его программной реализации, эффективно использующей архитектуру гибридного суперкомпьютера.

Необходимость использования суперкомпьютеров из Top10 для моделирования в физике плазмы обусловлена существующей в настоящее время тенденцией к отказу от упрощенных моделей (в частности, гидродинамических) и переходом к более полному описанию на основе уравнению Власова. Для решения уравнения Власова используется метод частиц в ячейках [5]. В том случае, когда используются различные модификации метода частиц в ячейках, количественный результат может быть получен только при использовании очень большого числа модельных частиц в ячейке (1 -10 тыс. частиц). Таким образом проведение крупномасштабных расчетов в физике плазмы в настоящее время возможно только на суперЭВМ гибридной архитектуры, вследствие того, что суперкомпьютер классической архитектуры будет потреблять слишком много энергии [6].

В настоящее время проведены расчеты [7] взаимодействия релятивистского электронного пучка с плазмой, позволившие в квазиодномерном случае точно рассчитать инкремент двухпоточковой неустойчивости. Однако для этого пришлось значительно увеличить число модельных частиц (до 1000) в одной ячейке. Число узлов сетки  $120 \times 4 \times 4$ , размер области 1.2 (в безразмерных единицах). Длина области в дебаевских радиусах составляет 134.8. Таким образом, получаем следующую оценку размера сетки:  $2156 \times 2156 \times 2156$  при 1000 модельных частиц каждого типа в ячейке. Это означает объем памяти 1.4 Петабайт и вычислительную нагрузку порядка 1.5 PetaFLOPs (около 250 операций на каждую частицу) в течение одного временного шага. Исходя из этого можно сказать, что для решения рассматриваемой задачи во всей полноте, а именно точного расчета спектра плазменной турбулентности, необходимо применение машин с производительностью более 1

PetaFLOPs, то есть перспективных экзафлопс-компьютеров. Может показаться, что применения современных суперЭВМ петафлопсного класса будет достаточно, но во-первых, приведенные выше цифры оценочные, во-вторых, они показывают минимально необходимый объем вычислений (134 дебаевских радиуса, 16 ячеек на дебаевский радиус, 1000 частиц), в-третьих 7.5 PetaFLOPs - это на один временной шаг, которых требуются десятки тысяч.

В той ситуации, когда нет большого опыта расчета даже на петафлопсных суперЭВМ и, с другой стороны, известно, что при повышении размерности задачи всегда возникают ранее не проявившиеся трудности, очень сложно говорить о расчет на экзафлопс-суперкомпьютерах. Единственное, что можно в такой ситуации сделать -это постараться обеспечить максимальную масштабируемость, то есть предусмотреть возможность параллельного счета на всех этапах разработки алгоритма. Это означает необходимость применения со-дизайна для построения параллельных моделей в физике плазмы.

Постановка задачи динамики галактических объектов заключается в совместном решении уравнений гравитационной магнитной газовой динамики для описания газовой компоненты с учетом магнитного поля и самогравитации, а также решения задачи N-тел для описания динамики бесстолкновительной компоненты. Известно, что задача N-тел достаточно трудны для решения на суперЭВМ тем более с гибридной архитектурой и необходимо искать альтернативы для описания бесстолкновительной компоненты с помощью модели, численный метод для которой позволяет построить эффективный параллельный алгоритм. В работе [8] была построена такая модель, которая основывается на решении первых моментов уравнения Больцмана. Такой подход позволяет использовать единый численный алгоритм [9] для решения как уравнений магнитной газовой динамики, так и для решения уравнений для первых моментов уравнения Больцмана [10]. Такой подход был эффективно реализован на двух типах гибридных суперЭВМ: суперЭВМ, оснащенных графическими ускорителями [10], а также суперЭВМ, оснащенных ускорителями Intel Xeon Phi [4].

### **Со-дизайн методов решения задач физики плазмы**

Основой понятия со-дизайна является анализ эффективности параллельной реализации на всех этапах разработки алгоритма. В данном случае со-дизайн начинается еще на уровне физического анализа задачи. Из эксперимента известно, что модуляция плотности плазмы не превышает 300 %. Это означает, что количество модельных частиц в отдельной ячейке не может увеличиваться неограниченно. Таким образом, необходимость в динамической балансировке загрузки отсутствует, что повышает надежность и эффективность параллельной реализации.

На уровне математической модели предпочтение было отдано расчету поля на основе законов Фарадея и Ампера, без явного решения уравнения Пуассона. Использование этих уравнений и схемы Лэнгдона-Лазинского дает возможность получить метод расчета поля с почти неограниченным масштабированием при увеличении числа процессорных ядер.

Проведено сравнение отдельных показателей производительности четырех отечественных кластеров (СКИФ Cyberia, МВС-100К, СКИФ-МГУ "Чебышёв" и кластера НГУ), достигнутых на задаче о моделировании взаимодействия электронного пучка с плазмой. По результатам сравнения можно сделать вывод, что важнейшим показателем быстродействия кластера на реальных задачах является скорость доступа к оперативной памяти. Наибольшая скорость работы с помощью созданной программы достигнута на кластере СКИФ-МГУ "Чебышёв". Скорость обращения к памяти определялась по времени расчета движения модельных частиц в течение одного временного шага (при расчете движения каждой частицы происходит обращение к 9 трехмерным массивам, причем эти обращения являются неупорядоченными). Время счета для 2.5 млн. модельных частиц на 1 CPU составило 1.9 сек. (СКИФ Cyberia), 0.9 сек. (МВС-100К), 0.4 сек. (СКИФ-МГУ "Чебышёв") и 1.2 сек. (кластер НГУ). По скорости межпроцессорных обменов (скорость пересылки модельных частиц) лидируют СКИФ Cyberia и СКИФ-МГУ "Чебышёв". Следует отметить, что именно недостаточная в большинстве случаев скорость работы с оперативной памятью на большинстве кластеров предопределила переход на гибридные суперЭВМ на основе технологии CUDA, так как при использовании CUDA есть возможность серьезной оптимизации работы с памятью при использовании метода частиц в ячейках.

Со-дизайн при выборе инструментов параллельной реализации выглядит следующим образом. Для метода частиц с его очень большим количеством независимо обрабатываемых элементов (модельных частиц) использованием технологии CUDA является очень эффективным. Другие технологии параллельных вычислений, ориентированные на гибридные вычислительные устройства, такие как OpenCL, OpenMP, OpenACC также могут быть использованы, но именно CUDA дает возможность задействовать одновременно наибольшее количество параллельных процессов и в итоге получить максимальную эффективность.

Последний элемент со-дизайна - доработка параллельного вычислительного алгоритма с учетом архитектуры графических ускорителей. Традиционный вариант реализации метода частиц – все частицы в одном большом массиве - неэффективен при счете на GPU или, по крайней мере, не позволит воспользоваться преимуществами GPU. Поэтому частицы распределены по ячейкам, и расчет происходит в варианте "одна ячейка сетки – один потоковый блок GPU". Теоретически более эффективен вариант "группа соседних ячеек сетки – один потоковый блок GPU". Это вопрос дальнейшей оптимизации алгоритма.

### **Со-дизайн методов решения задач астрофизики**

При построении астрофизических моделей мы будем руководствоваться концепцией со-дизайна алгоритмов, в основе которой лежит разработка параллельных вычислительных технологий с учетом всех граней параллелизма. Так в работе [8] был предложен новый подход для описания бесстолкновительной компоненты, основанный на использовании первых моментов бесстолкновительного уравнения Больцмана. Такой подход имеет ограничения, когда необходимо рассматривать движения каждой отдельной частицы, а не кластера частиц (например, в задачах моделирования динамики планетообразования). Однако для моделирования взаимодействующих галактик такой подход был успешно апробирован [10]. Как уже было отмечено во введении постановка задачи динамики галактических объектов заключается в совместном решении уравнений гравитационной магнитной газовой динамики и уравнений для первых моментов бесстолкновительного уравнения Больцмана.

В последние два десятилетия из широкого диапазона газодинамических численных методов для решения нестационарных трехмерных астрофизических задач используются два основных подхода. Это лагранжев подход, в основном представленный SPH-методом и эйлеров подход с использованием адаптивных сеток или AMR. Основными проблемами SPH метода является плохое воспроизведение высоких градиентов решения и разрывов, подавление развития неустойчивостей, неоднозначный выбор ядра сглаживания и необходимость введения искусственной вязкости. Отдельно стоит проблема локального убывания энтропии в SPH методе. Основным недостатком сеточных методов — неинвариантность решения относительно поворота или галилеева неинвариантность при использовании расчетной сетки. Однако такая проблема может быть решена при использовании различных подходов к конструированию численных схем [2]. Один из общих недостатков обоих подходов – проблема масштабируемости. В последнее время для решения астрофизических задач активно стали использоваться лагранжево-эйлеровые методы. Такие методы объединяют достоинства и лагранжевого и эйлерового подходов. Авторским коллективом уже на протяжении нескольких лет развивается эйлерово-лагранжевый подход для решения астрофизических задач на основе комбинации Fluid-In-Cells метода и метода Годунова [2, 4, 9, 10]. В основе метода лежит решение уравнений газовой динамики в два этапа. На первом, эйлеровом, этапе система решается без адвективных членов, на втором, лагранжевом, этапе происходит учет адвективного переноса. Разделение на два этапа позволяет эффективно решить проблему галилеевой неинвариантности, а использование метода Годунова на эйлеровом этапе позволяет корректно воспроизводить разрывные решения.

Использование равномерных сеток позволяет выбрать произвольную декартовую декомпозицию расчетной области. На рисунке 2 приведены два способа декомпозиции расчетной области в случае использования суперЭВМ, оснащенных графическими ускорителями (слева) и ускорителями Intel Xeon Phi (справа).

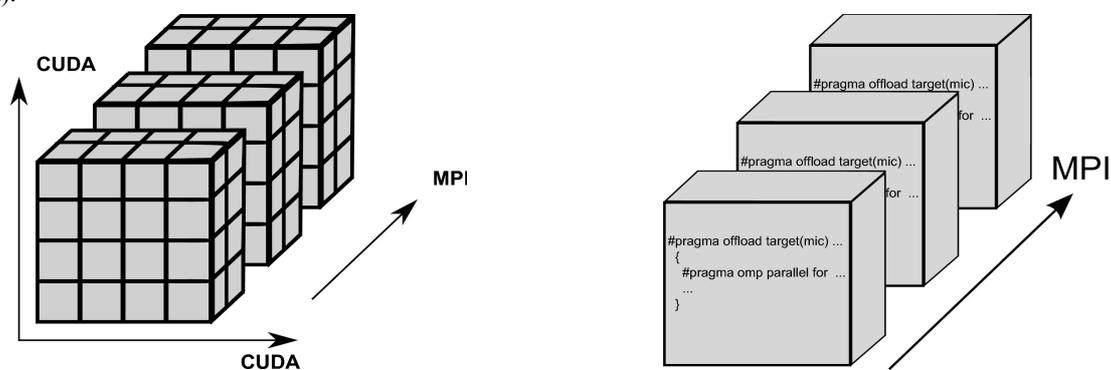


Рис. 2. Декомпозиция расчетной области

Использование такой декомпозиции расчетной области позволяет наиболее эффективно использовать возможности ускорителей. На рисунке 3 приведено ускорение в рамках одного графического ускорителя (слева) и в рамках одного ускорителя Intel Xeon Phi в различных режимах его использования (справа).

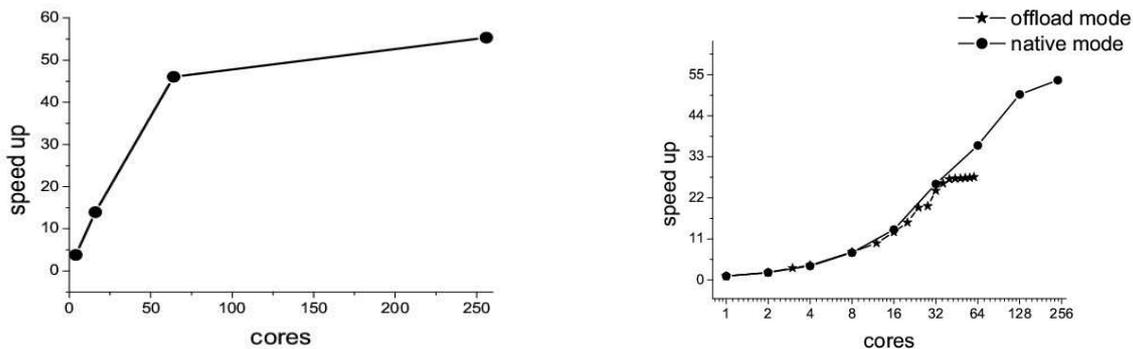


Рис. 3. Ускорение в рамках одного ускорителя

При использовании 60 графических и 32 ускорителей Intel Xeon Phi было достигнута эффективность на уровне выше 95%. Абсолютное время решения для кода AstroPhi выглядит следующим образом: 1 ядро Xeon 10,664 сек., 1 ядро Xeon Phi offload mode 81,708 сек. (столько же, сколько и native), 60 ядер Xeon Phi offload mode 2,960 сек. (ускорение в 28 раз), 60 ядер Xeon Phi native mode 1,547 сек. (ускорение в 53 раза). Для кода GPUPEGAS 1 ядро Xeon 10,664 сек., 1 ядро GPU 14,575 сек., 256 ядер GPU 0,265 сек. (ускорение 55 раз).

В заключение приведем результаты вычислительных экспериментов по моделированию столкновения двух галактик в описанной двух фазной модели.

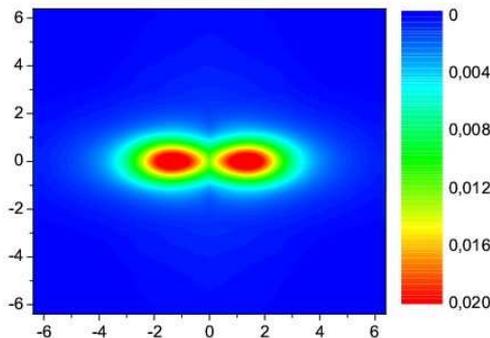


Рис. 4. Разлет галактик после их центрального столкновения

Подробности постановки задачи и результаты моделирования могут быть найдены в работе [10].

### Заключение

Разработка эффективных программных кодов для гибридных суперкомпьютеров отдельная сложная научная задача, требующая со-дизайна алгоритмов на всех стадиях решения задачи – от физической постановки до инструментов разработки. Понятие со-дизайна в контексте математического моделирования физических процессов понимается как построение физико-математической модели явления, численного метода, параллельного алгоритма и его программной реализации, эффективно использующей архитектуру гибридного суперкомпьютера.

Использование методики со-дизайна при моделировании плазменных процессов методом частиц в ячейках позволило достичь ускорения в 42 раза в расчетах на графическом ускорителе Nvidia Kepler. Ускорение в данном случае указано относительно 4 ядер процессора Intel Xeon. Масштабируемость программы, измеренная на MVC-100K составила 80% при использовании более 1000 процессорных ядер Intel Xeon. Таким образом благодаря использованию со-дизайна есть возможность провести трехмерные расчеты требуемой размерности при наличии необходимых ресурсов.

Благодаря со-дизайну параллельной вычислительной технологии решения задач астрофизики удалось достичь 55-кратного ускорения в рамках одного графического ускорителя и 96-процентной эффективности при использовании 60 GPU, а также 27-кратного в offload режиме и 53-кратного ускорения в native режиме в рамках одного ускорителя Intel Xeon Phi и 94-процентной эффективности при использовании 32 MIC. Такое эффективное использование ускорителей позволило производить расчеты на сетках 10243 за приемлемое время счета и смоделировать «тонкие» эффекты при столкновении галактик.

### ЛИТЕРАТУРА:

1. Ч.Бэдсел, А.Ленгдон, *Физика плазмы и численное моделирование*, М.: Энергоатомиздат 1989, -452с
2. V. Vshivkov, G. Lazareva, A. Snytnikov, I. Kulikov, A. Tutukov "Hydrodynamical code for numerical simulation of the gas components of colliding galaxies" // *The Astrophysical Journal Supplement Series*. 2011. V. 194, 47. 12 pp.
3. В.А.Вшивков, А.В.Снытников. «Особенности проведения экзафлопс-расчетов в физике плазмы» // *Вычислительные методы и программирование*, Том 13, 2012, стр. 44-48.
4. И.М. Куликов, И.Г. Черных, Б.М. Глинский "AstroPhi: программный комплекс для моделирования динамики астрофизических объектов на гибридных суперЭВМ, оснащенных ускорителями Intel Xeon Phi" // *Вестник ЮУрГУ, серия: Вычислительная математика и информатика*. 2013. Т. 2, № 4. С. 57-79.
5. Ю.Н.Григорьев, В.А.Вшивков. «Численное моделирование методами частиц-в-ячейках». Изд-во Сибирского отделения Российской академии наук, 2004
6. P.Muggli. *Proceedings of IPAC2013* (2013)
7. Е.А.Месяц, А.В.Снытников, К.В.Лотов. О выборе числа частиц в методе частиц-в-ячейках для моделирования задач физики плазмы // *Вычислительные технологии*, 2013. Т. 18, N 6, стр.
8. N. Mitchell, E. Vorobyov, G. Hensler "Collisionless Stellar Hydrodynamics as an Efficient Alternative to N-body Methods" // *Monthly Notices of the Royal Astronomical Society*. 2013. Vol. 428, I. 3. P. 2674-2687.
9. I. Kulikov "PEGAS: Hydrodynamical code for numerical simulation of the gas components of interacting galaxies" // *The Book Series of the Argentinian Association of Astronomy*. 2013. Vol. 4. P. 91-95.
10. I. Kulikov "GPUPEGAS: a new GPU-accelerated hydrodynamical code for numerical simulation of interacting galaxies" // *The Astrophysical Journal Supplement Series*. 2014 (in print). Arxiv: <http://arxiv.org/abs/1311.0861>