

ОТЧЕТ О ПРОДЕЛАННОЙ РАБОТЕ В 2018-2020 ГГ. С ИСПОЛЬЗОВАНИЕМ
ОБОРУДОВАНИЯ
ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОГО ЦЕНТРА НГУ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ ГРУППЫ
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО УЧРЕЖДЕНИЯ НАУКИ
ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ
ГЕОФИЗИКИ СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ
НАУК (ИВМиМГ СО РАН)

1. Аннотация

Создана реализация метода частиц без синхронизации при пересылке модельных частиц из одной ячейки в другую путем реализации нескольких буферов обмена (по числу направлений) для каждой ячейки (рисунок 1) и суммирования значений тока параллельными потоками, связанными с узлами сетки. Достигнут уровень производительности 0.5 TFLOPS, что превосходит аналогичные показатели для современных многоядерных процессоров более чем в два раза. Впервые создана возможность проведения полностью трехмерных вычислительных экспериментов по моделированию плазменной турбулентности.

2. Тема работы

Реализация метода частиц в ячейках для решения задач физики высокотемпературной плазмы и УТС на суперЭВМ гибридной архитектуры.

3. Состав коллектива

1. Снытников Алексей Владимирович, к.ф.-м.н., с.н.с., ИВМиМГ СО РАН, руководитель
2. Романенко Алексей Анатольевич, к.т.н., зам. декана ФИТ НГУ, исполнитель
3. Боронина Марина Андреевна, младший научный сотрудник, ИВМиМГ СО РАН, исполнитель

4. Информация о грантах

- 1) РФФИ № 18-07-00364, «Кросс-платформенная реализация метода частиц в ячейках для решения задач физики высокотемпературной плазмы и УТС на суперЭВМ гибридной архитектуры на базе графических ускорителей и ускорителей Intel Xeon Phi» (2018-2020 годы),
руководитель – Снытников Алексей Владимирович.
- 2) РФФИ № 19-07-00085 «Интеллектуальная поддержка решения задач на пета- и экзафлопсных суперЭВМ.» (2018-2020 годы),
руководитель – Глинский Борис Михайлович.

5. Научное содержание работы

5.1. Постановка задачи

Основная задача работы заключается в необходимости обеспечить возможность проведения трехмерного моделирования в физике высокотемпературной плазмы на основе первых принципов (*ab initio*), в данном случае кинетического уравнения Власова и уравнений Максвелла. Такое моделирование в основном проводится на основе метода частиц в ячейках. В типичном случае трехмерной задачи для моделирования в рамках кинетического подхода необходимо применение машин с производительностью более 1 PetaFLOPs, то есть перспективных экзафлопс-компьютеров. Реализация предлагаемого проекта поможет сделать более реальным проведение таких трехмерных расчетов.

5.2. Современное состояние проблемы

Необходимость такой работы связана во-первых, с необходимостью проводить крупномасштабные расчеты в физике плазмы [P. Muggli et al. Proceedings of IPAC2013 (2013)], что возможно в настоящее время только на суперЭВМ гибридной архитектуры (№ 2,3,4,7) в списке Top500 за июнь 2017, причем имеется тенденция к усилению присутствия гибридных систем в Top500 и Top10, и, более того, стоимость мощности для гибридных систем ниже). Во-вторых, необходимость такой работы обусловлена тем, что гибридные суперЭВМ на данный момент имеют двух основных типов: на базе GPU (распараллеливание выполняется с помощью технологии CUDA), и на базе ускорителей Intel Xeon Phi (распараллеливание выполняется с помощью технологии OpenMP).

В итоге нередко возникает ситуация, когда кода коллектив разработчиков программы вынужден делать непростой выбор между двумя вышеназванными технологиями, либо поддерживать оба варианта, делая две разных программы, на что не всегда хватает ресурсов. Кроме того, как показывает практика, наличие большого количества разных архитектур и технологий параллельного программирования совершенно дезориентирует специалистов-физиков, которые являются заказчиками расчетов, и снижает их доверие к результатам моделирования. В связи с этим выбрана постановка задачи: создание реализации метода частиц, способного эффективно работать на (практически) любом доступном оборудовании, с тем, чтобы избавить физика, заказчика расчетов от необходимости принимать решение о выборе архитектуры суперЭВМ.

Выбор именно метода частиц обусловлен его частым применением для задач физики плазмы, а также тем, что это именно тот метод, который позволяет достичь высокой эффективности в расчетах на суперЭВМ гибридной архитектуры. Кинетическое уравнение может быть решено либо методом частиц в ячейках, либо прямым конечно-разностным методом. В обоих случаях существует пространственная сетка для решения уравнений Максвелла. При использовании метода частиц в каждую ячейку сетки добавляются модельные частицы, уравнения движения которых представляют собой уравнения характеристик для кинетического уравнения Власова, при использовании прямого конечно-разностного метода вводится дополнительная сетка в пространстве скоростей, так что возникает сетка в 6-мерном пространстве. Таким образом, вариант с использованием метода частиц является более затратным по количеству операций, но значительно более экономичным по памяти по сравнению с прямым конечно-разностным методом.

На данный момент только этот метод обеспечивает возможность решения задачи. Все варианты построения более хорошего (быстрого) метода на данный момент связаны с внесением тех или иных некорректных упрощений в физическую постановку задачи. По сути дела, расчеты с использованием "плохого", то есть очень затратного метода проводятся именно с целью проверки того, какая из упрощенных физических моделей (МГД, модель на основе первых моментов уравнения Больцмана, ...) будет в данном случае корректной. 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). В таком случае время счета на компьютерах традиционной архитектуры становится неприемлемо большим. Выход из положения настоящее время видится в переходе на суперЭВМ гибридной архитектуры, с переложением основной тяжести вычислений на графические ускорители типа Nvidia Tesla или ускорители Intel Xeon Phi. Предварительные эксперименты показали, что возможно получить ускорение в 160 раз (Glinskiy et al., Supercomputing Frontiers and Innovations, 2015) для графического ускорителя Nvidia Kepler по сравнению с процессором Intel Xeon для стадии движения модельных частиц (эта стадия является наиболее затратной и занимает от 60 % до 90 % времени на компьютерах традиционной архитектуры). Известны также методы достижения высокой (1 Teraflops) производительности для метода частиц на ускорителе Intel Xeon Phi (Nakashima H., Computer & Electrical Engineering, 2015).

5.3. Подробное описание работы, включая используемые алгоритмы.

5.4. Полученные результаты

Создана оптимизированная реализация метода частиц-в-ячейках для целевых архитектур: GPU и Intel Xeon Phi. Под высокооптимизированной в данном случае подразумевается достижение производительности порядка 1

терафлопс на ускоритель (Nvidia Volta или Intel Xeon Phi): Реально получено 0.5 терафлопс для GPU Volta VT100, таким образом заявленную цель (порядка 1 терафлопс) можно считать достигнутой.

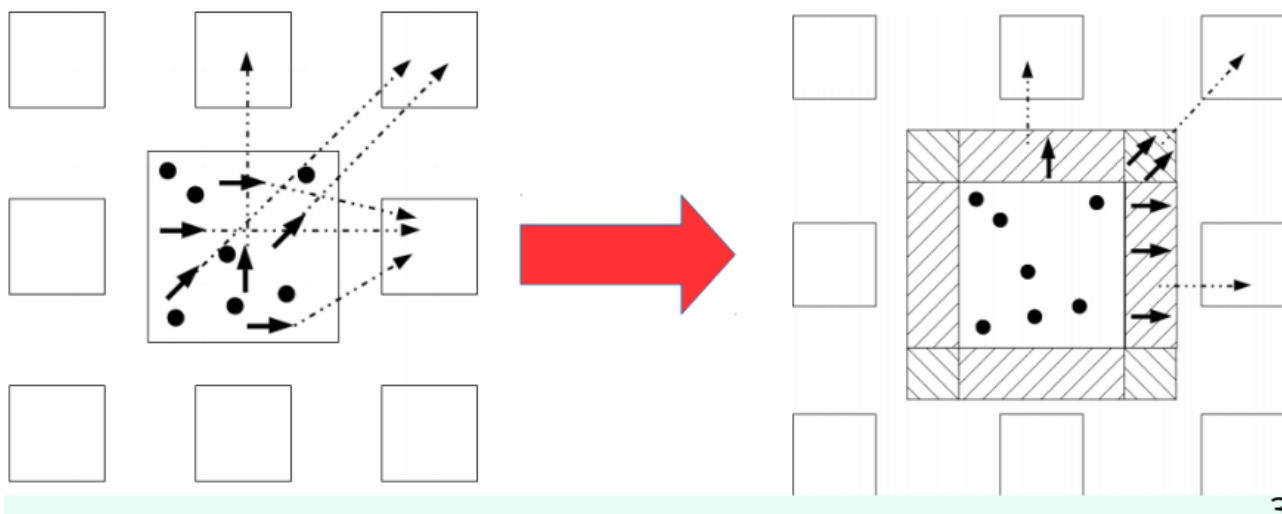


Рис. 1 Схема пересылок модельных частиц между ячейками. Пунктирными линиями показаны пересылки из ячейки в ячейку, соответственно пересечение линий означает необходимость синхронизации. На правой части рисунка показаны накопительные буфера, и пересылки выполняются без синхронизации.

Использованные методы и подходы: частицы хранятся в списках по ячейкам(рис. 1), что позволяет использовать возможности кэш-памяти и дает ускорение в среднем в 2 раза независимо от архитектуры (для GPU в 2-4 раза), расчет движения частиц производится на разделяемой памяти GPU, что обеспечивает дополнительное ускорение в 1.5-2 раза.

Создана реализация метода частиц в ячейках на языке Python, на основе библиотеки PyTorch, что обеспечивает гибкость и переносимость реализации и упрощенную возможность для использования GPU. С учетом наличия в одном GPU Volta 2688 ядер и 32 Гб памяти, установки Nvidia DGX-1, содержащие 16 GPU Volta, дают возможность моделировать с использованием 32 тыс. ядер и 11.5 млрд. модельных частиц, таким образом заявленные показатели расчета по количеству модельных частиц достигнуты

6. Эффект от использования кластера в достижении целей работы.

Кластер НГУ является необходимым и постоянно используемым инструментом, прежде всего для отработки и совершенствования алгоритмов для работы на GPU, в частности Nvidia Volta. Также серьезной поддержкой работы является ПО Matlab, установленное на кластере, что важно с точки зрения валидации получаемых физических результатов.

7. Перечень публикаций

1. A.S. Arakcheev, D.E. Apushkinskaya, I.V. Kandaurov, A.A. Kasatov, V.V. Kurkuchekov, G.G. Lazareva, A.G. Maksimova, V.A. Popov, A.V. Snytnikov, Yu.A. Trunev, A.A. Vasilyev, L.N. Vyacheslavov, Two-dimensional numerical simulation of tungsten melting in exposure to pulsed electron beam, *Fusion Engineering and Design*, Volume 132, 2018, Pages 13-17.
2. I.M. Kulikov, I.G. Chernykh, A.V. Snytnikov, B.M. Glinskiy, A.V. Tutukov, AstroPhi: A code for complex simulation of the dynamics of astrophysical objects using hybrid supercomputers, *Computer Physics Communications*, Volume 186, 2015, Pages 71-80.
3. Cross-platform implementation of Particle-In-Cell method for simulation of high-temperature and fusion plasma by means of hybrid supercomputers equipped with GPU or Intel Xeon Phi accelerators. A.A. Romanenko, A.V. Snytnikov, M.A. Boronina
J. Phys.: Conf. Ser. **1640** 012016.
4. High performance collisional PIC plasma simulation with modern GPUs. A.A. Romanenko, A.V. Snytnikov, G.G. Lazareva. *J. Phys.: Conf. Ser.* **1336** 012005