Проблемы распараллеливания метода частиц в ячейках для задачи взаимодействия электронного пучка с

В.А.Вшивков, А.В. Снытников
vsh,snytav@ssd.sscc.ru
Институт Вычислительной Математики и Математической Геофизики СО РАН
Новосибирск

Содержание

- Проблемы эффективного распараллеливания для большого числа процессоров
- Моделирование динамики плазмы методом частиц в ячейках
- Проведение больших численных расчетов на суперЭВМ
- О реализации метода частиц на GPU

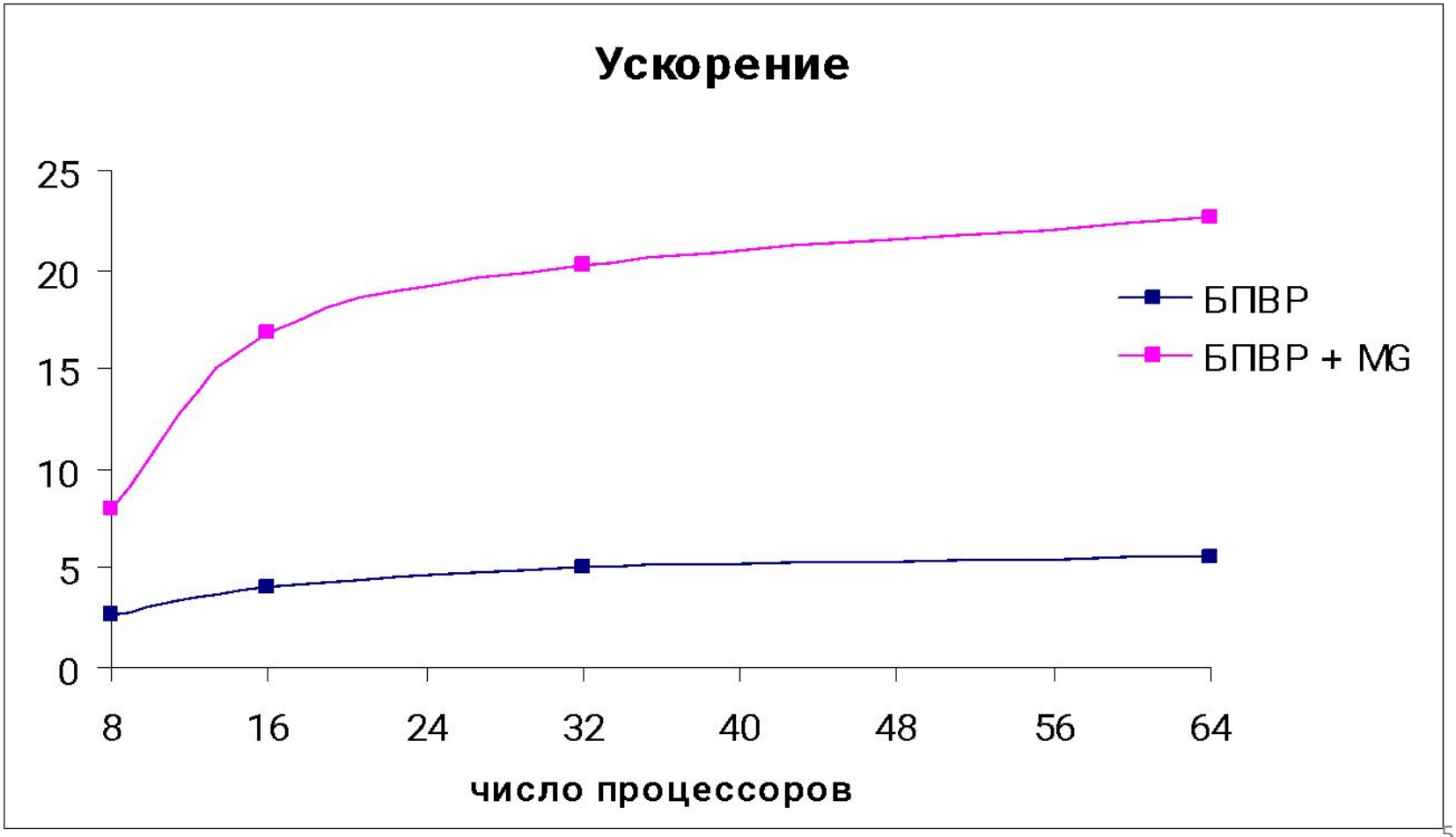
Проблемы эффективного распараллеливания для большого числа процессоров

- Решение уравнения Пуассона
- Параллельная прогонка
- Метод частиц в ячейках

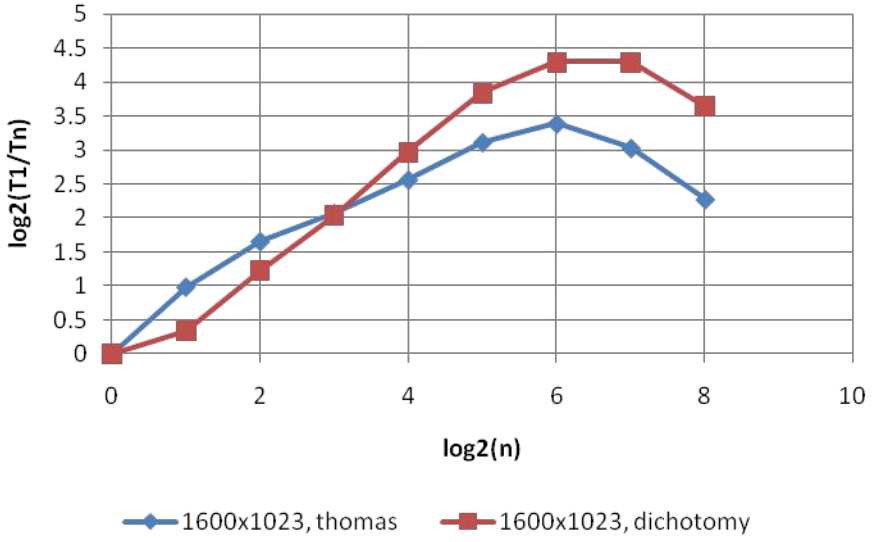
Rank	Site	Computer/Year Vendor	Cores	R_{max}	R_{peak}
1	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	548352	8162	8773.63
2	National Supercomputing Center in Tianjin China	NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C NUDT	186368	2566	4701
3	DOE/SC/Oak Ridge National Laboratory United States	Cray XT5-HE Opteron 6-core 2.6 GHz Cray Inc.	224162	1759	2331
4	National Supercomputing Centre in Shenzhen (NSCS) China	Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU Dawning	120640	1271	2984.3
5	GSIC Center, Tokyo Institute of Technology Japan	HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows NEC/HP	73278	1192	2287.63
6	DOE/NNSA/LANL/SNL United States	Cray XE6 8-core 2.4 GHz Cray Inc.	142272	1110	1365.81
7	NASA/Ames Research Center/NAS United States	SGI Altix ICE 8200EX/8400EX, Xeon HT QC 3.0/Xeon 5570/5670 2.93 Ghz, Infiniband SGI	111104	1088	1315.33
8	DOE/SC/LBNL/NERSC United States	Cray XE6 12-core 2.1 GHz Cray Inc.	153408	1054	1288.63
9	Commissariat a l'Energie Atomique (CEA) France	Bull bullx super-node S6010/S6030 Bull SA	138368	1050	1254.55
10	DOE/NNSA/LANL United States	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband IBM	122400	1042	1375.78

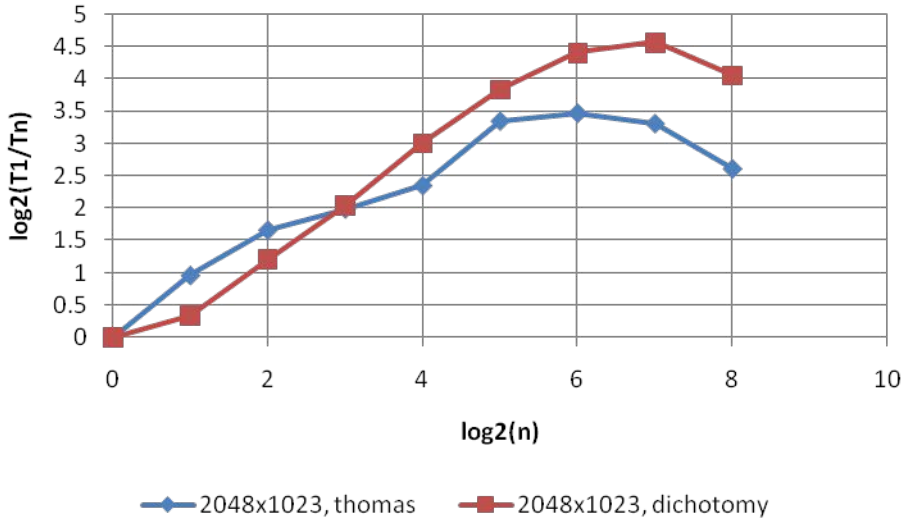
Моделирование динамики протопланетного диска:

Ускорение при использовании различных методов решения уравнения Пуассона



Зависимость логарифма ускорения от логарифма числа узлов





Tn – Время работы на N узлах

T1 – Время работы на 1 узле

n – количество узлов

Толстых М.А., Терехов А.В., Поливанов Н.С. МФТИ

Реализация массивно-параллельной глобальной модели атмосферы нового поколения

Международная суперкомпьютерная конференция «Научный сервис в сети Интернет: экзафлопсное будущее», Абрау-Дюрсо, 2011.

Всероссийская конференция **«Актуальные проблемы вычислительной математики и математического моделирования»**

13 - 15 июня 2012 года

Новосибирск, Россия

Тематика конференции:

Математическое моделирование и *Параллельные вычислительные методы*

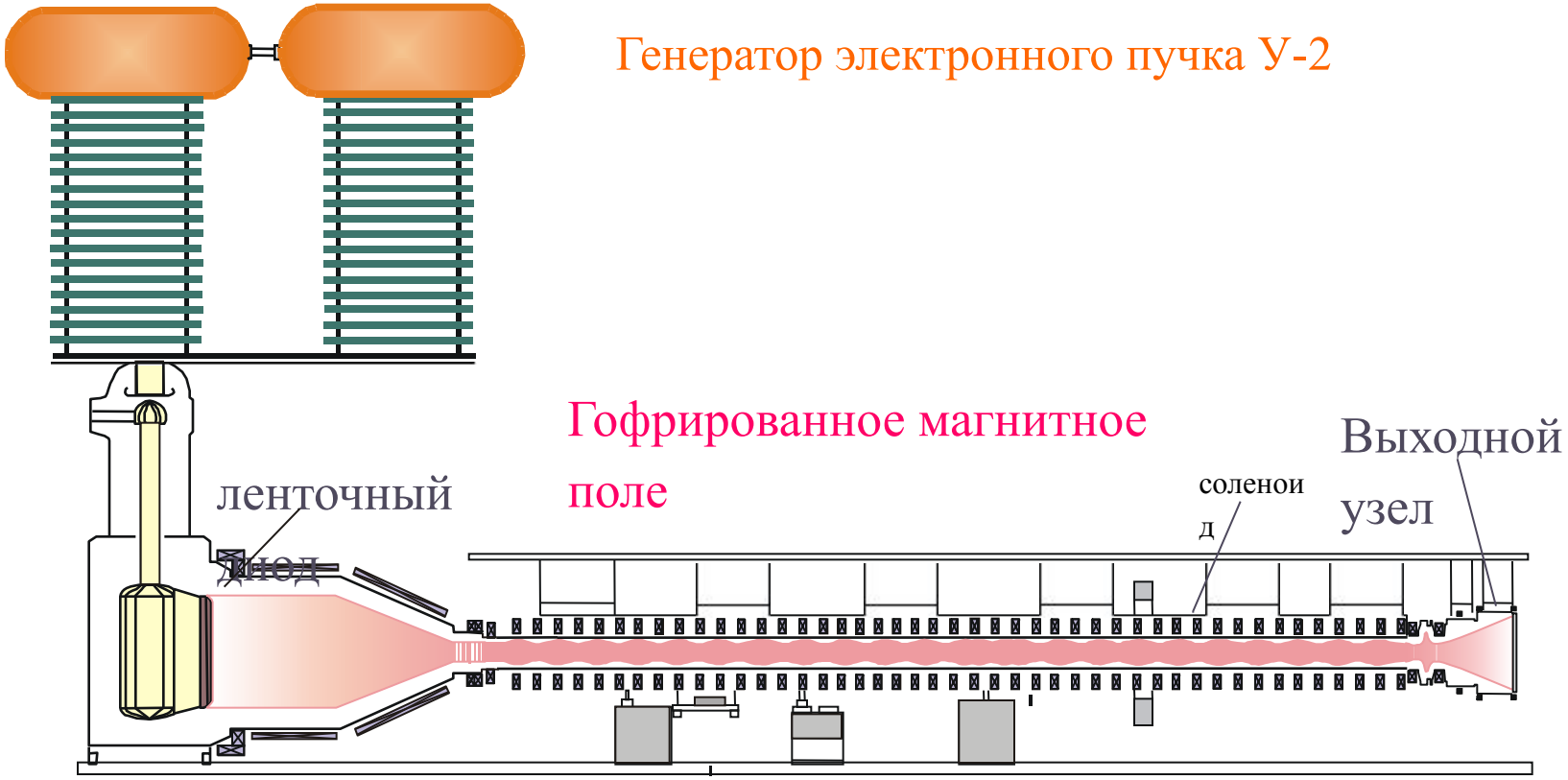
Председатели программного комитета:

академик Марчук Г.И.

академик Михайленко Б.Г.

Установка ГОЛ-3 (ИЯФ СО РАН)

- Установка ГОЛ-3 представляет собой многопробочную термоядерную ловушку открытого типа с плазмой высокой плотности, нагреваемой мощным релятивистским электронным пучком. Плазма установки ГОЛ-3 по своим параметрам является субтермоядерной.



Эффект аномальной теплопроводности

- В экспериментах на установке ГОЛ-3 (ИЯФ СО РАН) вследствие релаксации мощного электронного пучка наблюдается понижение электронной теплопроводности
- Коэффициент электронной теплопроводности уменьшается в 10^2-10^3 раз по сравнению с классическим значением для плазмы с такой плотностью и температурой
- Это позволяет лучше нагревать плазму и дольше удерживать ее в нагретом состоянии вследствие намного меньшего теплового потока на стенки установки

Система уравнений Власова-Максвелла

- Плазма описывается системой уравнений Власова-Максвелла:

$$\frac{\partial f_k}{\partial t} + (\vec{\upsilon}, \nabla) f_k + q_k \left(\vec{E} + \frac{1}{c}[\vec{\upsilon} \times \vec{H}] \right) \frac{\partial f_k}{\partial \vec{p}} = 0$$

$$rot\vec{H} = \frac{4\pi}{c}\vec{j} + \frac{1}{c}\frac{\partial \vec{E}}{\partial t} \qquad rot\vec{E} = -\frac{1}{c}\frac{\partial \vec{H}}{\partial t}$$

$$div\vec{E} = 4\pi\rho \qquad div\vec{H} = 0$$

$$\vec{j} = \sum_k q_k \int \vec{\upsilon} f_k(\vec{p}, \vec{r}, t) d\vec{p} \qquad \rho = \sum_k q_k \int f_k(\vec{p}, \vec{r}, t) d\vec{p}$$

где f_k - функция распределения частиц сорта k (электроны или ионы), c - скорость света, ρ - плотность электрического заряда, $\vec{j}$ - плотность электрического тока, q_k - заряд частицы сорта k.

Лагранжев этап

$$\frac{\partial f_k}{\partial t} + (\vec{\upsilon}, \nabla) f_k + q_k \left(\vec{E} + \frac{1}{c} [\vec{\upsilon} \times \vec{H}] \right) \frac{\partial f_k}{\partial \vec{p}} = 0$$

$$\frac{d\vec{p}_k}{dt} = q_k \left(\vec{E} + \frac{1}{c} [\vec{\upsilon}_k, \vec{H}] \right)$$

$$\frac{d\vec{r}_k}{dt} = \vec{\upsilon}_k$$

$$\vec{p}_k = \gamma_k m_k \vec{\upsilon}_k, \qquad \gamma_k = 1 / \sqrt{1 - \upsilon_k^2 / c^2}$$

Эйлеров этап

- Эйлеров этап:

$$\frac{\partial \vec{E}}{\partial t} = c \cdot rot\vec{H} - 4\pi \vec{j}$$

$$\frac{\partial \vec{H}}{\partial t} = -c \cdot rot\vec{E}$$

- Схема эйлерова этапа:

$$\frac{\vec{H}^{m+1/2} - \vec{H}^{m-1/2}}{\tau} = -c \cdot rot_h \vec{E}^{m},$$

$$\frac{\vec{E}^{m+1} - \vec{E}^{m}}{\tau} = c \cdot rot_h \vec{H}^{m+1/2} - 4\pi \vec{j}^{m+1/2}.$$

Восстановление плотности заряда по частицам

$$\rho_{i,l,k} = \sum_j q_j \bar{R}\left(\vec{r}_j - \vec{r}_{i,l,k}\right)$$

$$\bar{R}\left(\vec{r}_j - \vec{r}_{i,l,k}\right) = R\left(x_j - x_i\right) \cdot R\left(y_j - y_l\right) \cdot R\left(z_j - z_k\right)$$

- NGP:

$$R(x) = \begin{cases} \frac{1}{h}, & |x| \le \frac{h}{2} \\ 0, & |x| > \frac{h}{2} \end{cases}$$

- PIC:

$$R(x) = \begin{cases} \frac{1}{h}\left(1 - \frac{|x|}{h}\right), & |x| \le h \\ 0, & |x| > h \end{cases}$$

Схема вычисления токов

$$jx^{m+\frac{1}{2}}_{i,l-\frac{1}{2},k-\frac{1}{2}} = q\frac{\Delta x}{\tau}\left(\left(1-\delta_y\right)\left(1-\delta_z\right)+\frac{\Delta y\Delta z}{12h_yh_z}\right)$$

$$jx^{m+\frac{1}{2}}_{i,l-\frac{1}{2},k+\frac{1}{2}} = q\frac{\Delta x}{\tau}\left(\left(1-\delta_y\right)\delta_z-\frac{\Delta y\Delta z}{12h_yh_z}\right)$$

$$jx^{m+\frac{1}{2}}_{i,l+\frac{1}{2},k-\frac{1}{2}} = q\frac{\Delta x}{\tau}\left(\delta_y\left(1-\delta_z\right)-\frac{\Delta y\Delta z}{12h_yh_z}\right)$$

$$jx^{m+\frac{1}{2}}_{i,l+\frac{1}{2},k+\frac{1}{2}} = q\frac{\Delta x}{\tau}\left(\delta_y\delta_z+\frac{\Delta y\Delta z}{12h_yh_z}\right)$$

$$\delta_x = \frac{1}{h_x}\left(\frac{x_0+x_1}{2}-x_i\right)$$

$$\delta_y = \frac{1}{h_y}\left(\frac{y_0+y_1}{2}-y_l\right)$$

$$\delta_z = \frac{1}{h_z}\left(\frac{z_0+z_1}{2}-z_k\right).$$

$$\Delta x = x_1 - x_0$$

$$\Delta y = y_1 - y_0$$

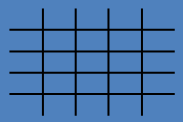
$$\Delta z = z_1 - z_0$$

$$\frac{\rho^{m+1}_{i-1/2,l-1/2,k-1/2}-\rho^{m}_{i-1/2,l-1/2,k-1/2}}{\tau} \equiv \frac{j^{m+1/2}_{x,i,l-1/2,k-1/2}-j^{m+1/2}_{x,i-1,l-1/2,k-1/2}}{h_x} +$$

$$+\frac{j^{m+1/2}_{y,i-1/2,l,k-1/2}-j^{m+1/2}_{y,i-1/2,l-1,k-1/2}}{h_y}+\frac{j^{m+1/2}_{z,i-1/2,l-1/2,k}-j^{m+1/2}_{z,i-1/2,l-1/2,k-1}}{h_z}$$

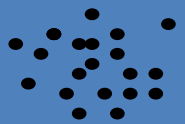
Алгоритм метода частиц-в-ячейках

Вычисление сеточных величин в узлах фиксированной сетки по частицам



Решение уравнений на фиксированной сетке (Эйлеров этап)

Интерполяция сил из узлов сетки в местоположение частиц



Движение частиц под действием сил (Лагранжев этап)

Модуль потока тепловой энергии электронов

$$\left|\vec{q}\right| = \left|\vec{v}_e\right| T_e$$

В соответствии с начальным предположением видно образование изолированных друг от друга областей с большим значением теплового потока

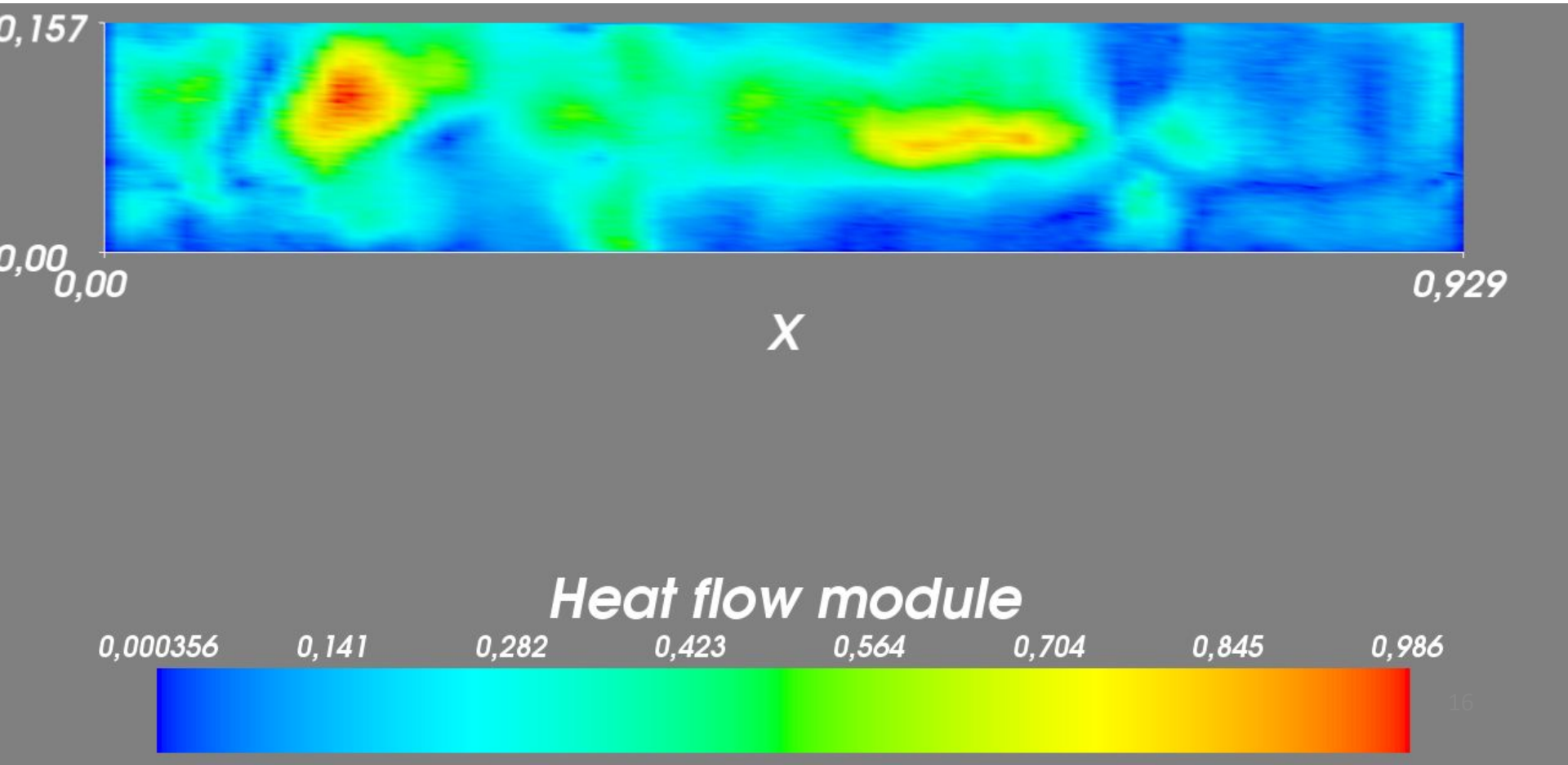
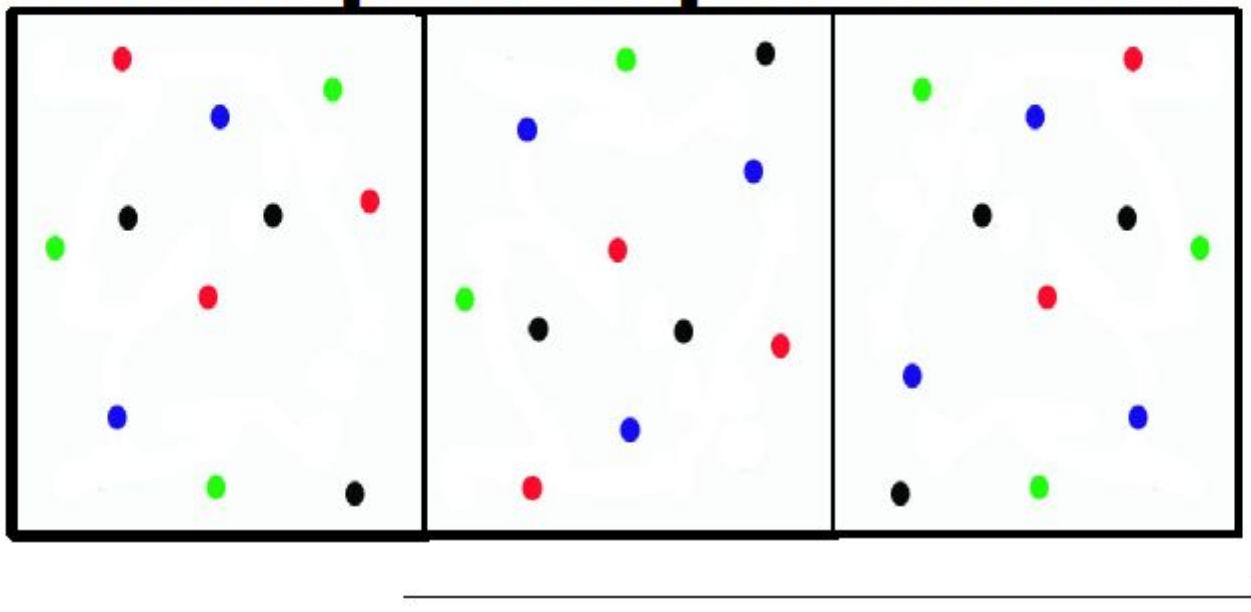


Схема распараллеливания



- N_{MAX} =
- M = 3

- Расчетная область делится на N_{MAX} подобластей вдоль координаты Y для решения уравнений Максвелла
- для каждой подобласти назначается группа из M процессорных элементов
- Далее частицы каждой подобласти разделяются дополнительно между M процессорными элементами

Проведение больших численных расчетов на суперЭВМ

- Оценка производительности суперЭВМ
- Повышение размерности задачи
- Компьютер — это не только процессоры
- Требования к системам хранения и передачи данных

Оценка производительности суперЭВМ

- Принятая единица — FlOpS (теоретические, или реально достигнутые, напр. LINPACK)
- Однако для реальных задач большее значение имеет быстродействие (и объем) оперативной памяти,
- а также жесткого диска

Время работы процедуры интегрирования на 1 шаге:
СКИФ-МГУ – 0.422 сек
МВС-100К – 0.896 сек

Значение объема жесткого диска
Пример конкретной задачи

- Релаксация мощного релятивистского пучка в высокотемпературной плазме, метод частиц-в-ячейках, сетка 512x64x64, 150 частиц в ячейке
- Изучается трехмерная динамика теплопроводности и фурье-образы основных величин (плотности, электрического поля — на данный момент 4 величины)
- Одна выдача занимает 160 Мб (архив 20)
- Необходимо от 100 до 400 моментов
- Требуется выдать все это на диск за ограниченное время работы программы — (один файл, СКИФ МГУ vs МВС-100К: 0.0134 сек. vs 0.0364 сек.)
- И не превысить дисковую квоту — возможно, это в большей степени вопрос администрирования — **но он существует**
- А потом еще передать по сети на локальный компьютер для обработки — по этой причине трехмерные выдачи делались пока только на НКС-30Т (ИВМиМГ СО РАН)

Повышение размерности задачи

- Существуют планы по поводу вычислений Exascale-масштабе.
- Тем не менее, лишь небольшое количество программ сейчас используют 1000 ядер (или больше), т.е. терафлопные мощности.
- Опыт проведения крупномасштабных расчетов свидетельствует, что при увеличении размерности на порядок появляются принципиально новые трудности в реализации алгоритма.
- Поэтому категорически нельзя сразу переходить от мелких, отладочных задач к крупномасштабным.
- А речь идет о повышении размерности на 6 порядков...

Компьютер — это не только процессоры

- Результат расчета в задачах физики плазмы (не только в рассмотренной выше) - это прежде всего, **трехмерные распределения плотности частиц, токов, распределения электромагнитного поля (сетка $2000^{3)}$.**
- Для сравнения численного результата с известными физическими закономерностями необходимо вычислить **фурье-образ рассматриваемой величины**
- Если они выдаются в двоичном формате, то размер одной такой выдачи составит 60 Гб. Но это один момент времени, в то время как требуется от 100 до 300 моментов времени с выдачей в течение одного расчета, то есть около **18 Петабайт**
- Более того, для решения какого-то отдельного вопроса в рамках задачи необходимо несколько (5-10) расчетов, то есть всего получается около **200**

Требования к системам хранения и передачи данных

- Объем диска - 200 Петабайт.
- Скорость диска - 270 Гбайт/сек (для обработки указанного массива данных в течение часа). Сейчас для SSD-дисков скорость чтения порядка 0.7 Гб/сек.
- Скорость сетевого соединения - 11 Гб/сек (для передачи этого массива данных по сети в течение суток), при том, что сейчас время передачи 1 Гб данных по внутренней сети ННЦ СО РАН занимает около получаса, т.е. 0.0005 Гб/сек.
- **Видно, что недостаток мощности систем хранения и передачи данных между текущим состоянием и перспективными экзафлопс-компьютерами ненамного меньше, чем по вычислительным мощностям (при том, что системам хранения данных традиционно уделяется меньше внимания).**

О реализации метода частиц на GPU

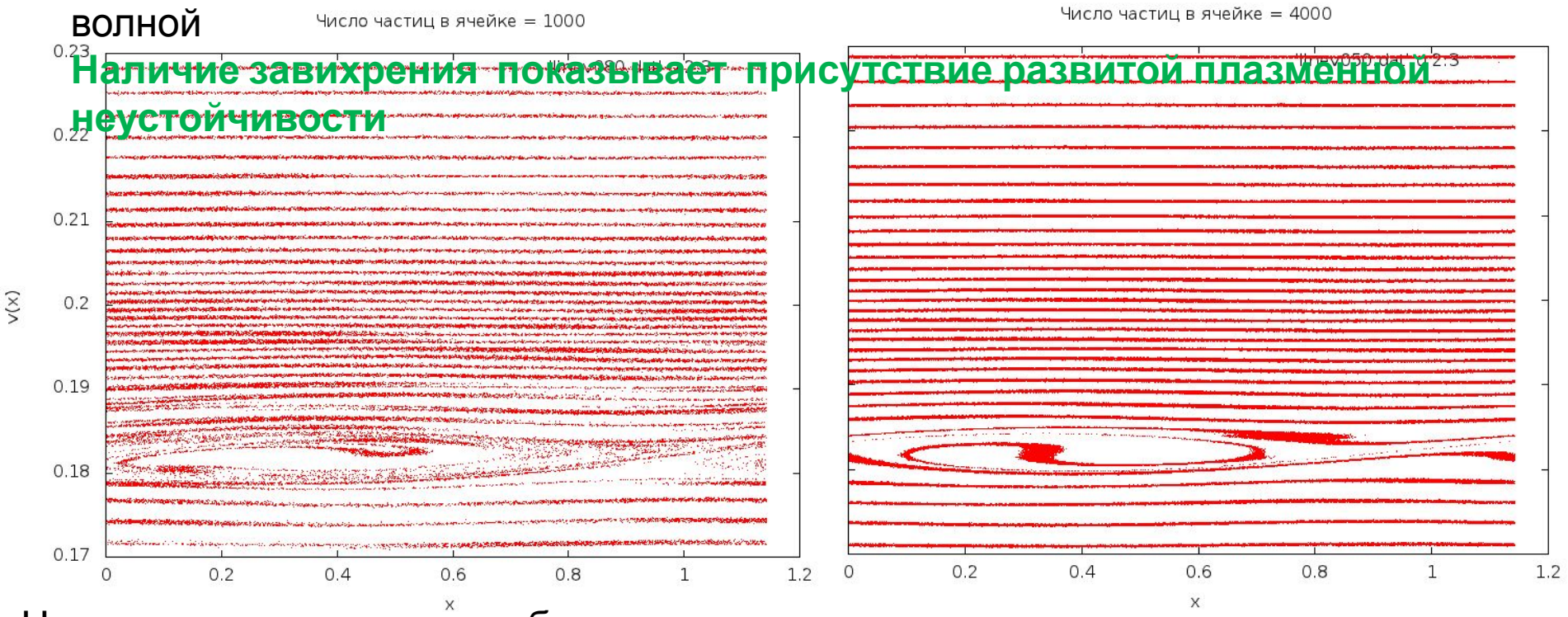
- Необходимость
- Методика
- Результаты

О необходимости использования большого числа частиц

На фазовых плоскостях показана скорость частиц пучка в зависимости от координаты

Частица смещается в том случае, когда она взаимодействует с плазменной волной

Наличие завихрения показывает присутствие развитой плазменной неустойчивости



На рисунках видно, что при большом количестве частиц процесс развития неустойчивости

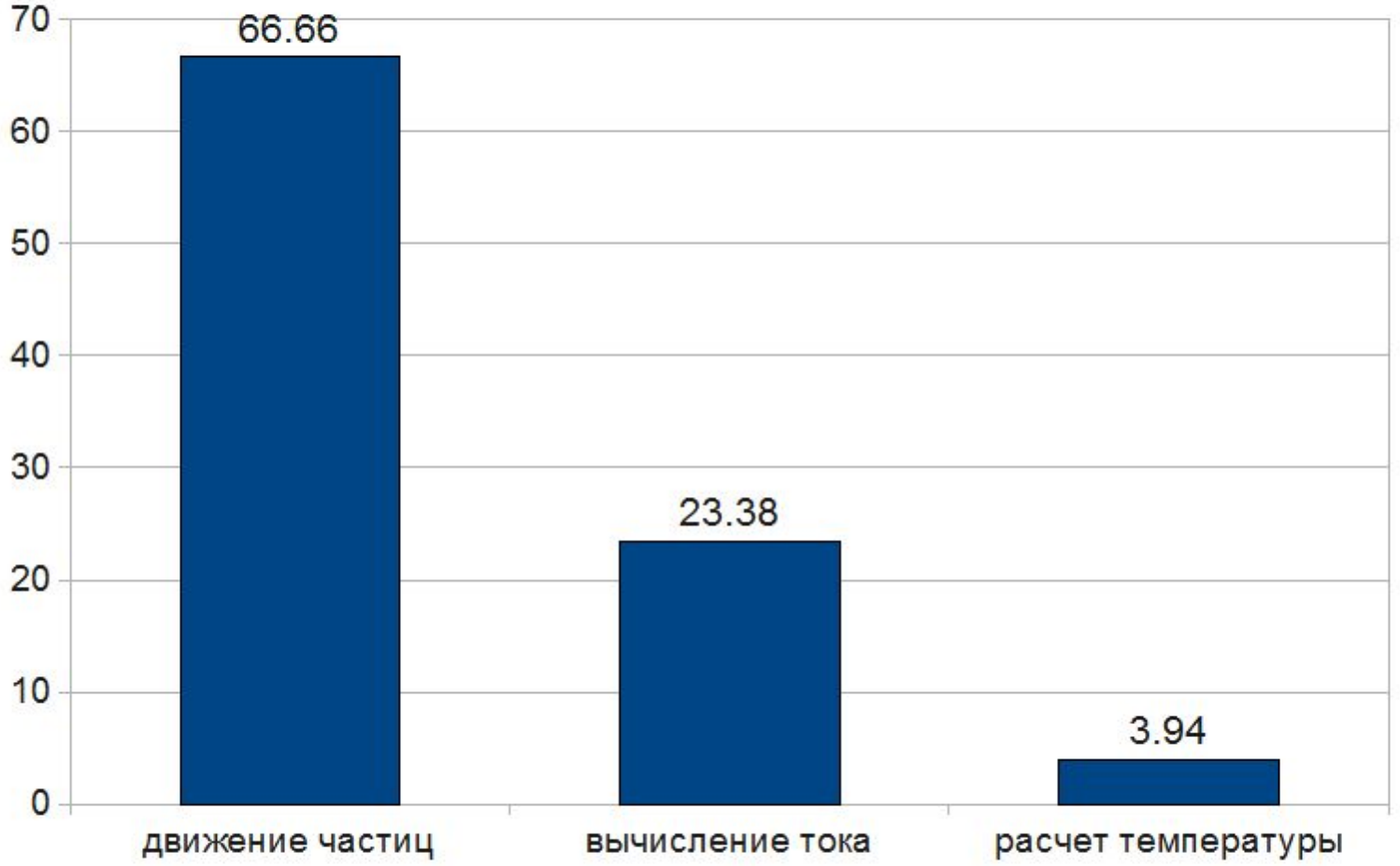
лучше соответствует теоретическим представлениям

Приципиально то, что в процессе образования неустойчивости (по физике) участвуе лишь

Оценка размера задачи

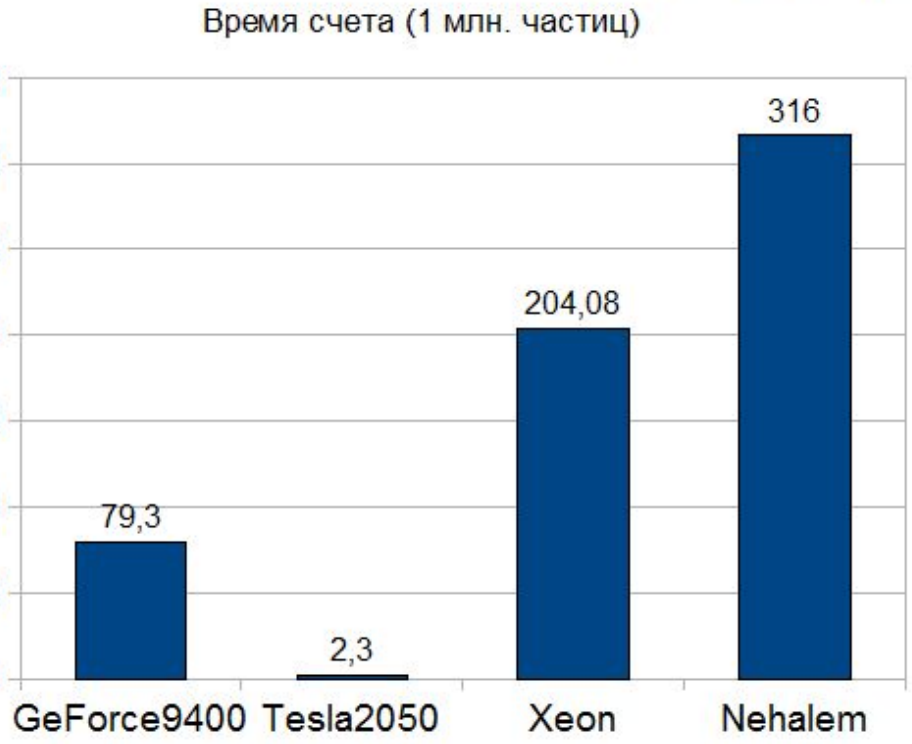
- В настоящее время проведены расчеты взаимодействия релятивистского электронного пучка с плазмой, позволившие в квазиодномерном случае точно рассчитать инкремент двухпотоковой неустойчивости
- Получено g = 0.081, точное значение g = 0.077 (К.В.Лотов и др., Физика плазмы, 2009).
- Однако для этого пришлось значительно увеличить число модельных частиц а именно до 1000 в одной ячейке.
- При этом величина дебаевского радиуса 8.9×10^{-3} в тех же единицах.
- Таким образом длина области в дебаевских радиусах составляет 134.8.
- Таким образом, получаем следующую оценку размера сетки: **2156x2156x2156 при 1000 модельных частиц каждого типа в ячейке.**
- **Это означает объем памяти 1.4 Петабайт и вычислительную нагрузку порядка 1.5 PetaFLOP (около 50 операций на каждую частицу)**

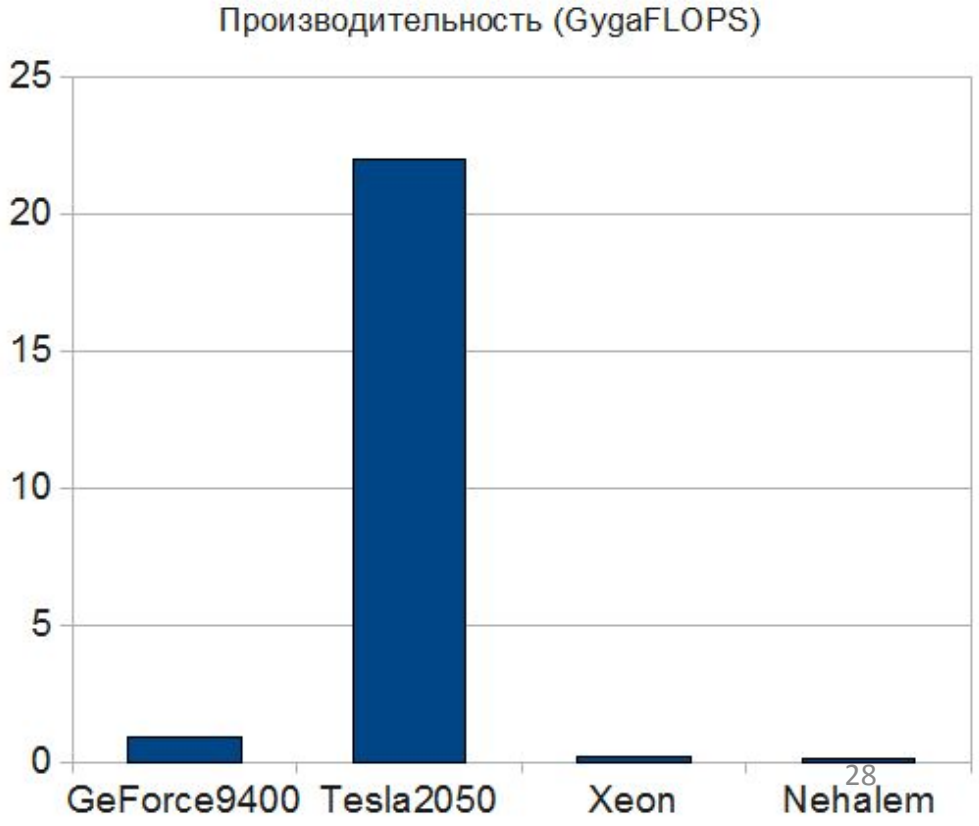
Распределение времени работы (в СКИФ-МГУ



Разница между движением частиц и вычислением тока в отсутствии обращений к большим одномерным массивам

Сравнение производительности CPU vs GPU





Использование текстур CUDA

- Что такое текстура: способ доступа к памяти
- Двух- или трехмерный массив с кэшированием, оптимизированным для двумерной адресации
- Двумерные координаты текстуры означают:
  - Номер ячейки, i
  - Номер частицы в ячейке, j
- Атрибуты частицы хранятся в 6 разных текстурах

Перспективы достижения экзафлопс-производительности для метода частиц-в-ячейках на GPU

- Используемая в настоящий момент одномерная декомпозиция области не может обеспечить достаточную масштабируемость
- Время счета одного временного шага составило 0.3 миллисекунды для одного миллиона частиц с двойной точностью (ГрафИТ!, НИВЦ МГУ).
- Так как для каждой частицы выполняется приблизительно 250 операций, то производительность одной карты Tesla может быть оценена как 833 ГигаФлопс (0.8 Терафлопс)

О перспективах достижения экзафлопс-производительности.Если...

- Взять за основу для рассуждений Tianhe-1A,
- Выделить для каждой подобласти один ускоритель Tesla и один универсальный процессор,
- Считать, что необходимое количество частиц помещается в оперативную память узла,
- Предположить, что время обмена данными между подобластями не превысит имеющегося сейчас,
- В таком случае компьютер Tianhe-1A дал бы для метода частиц в ячейках производительность порядка 5.6 PetaFLOPS.
- Такая же производительность могла бы быть достигнута при использовании порядка 250 тыс. 4-ядерных процессоров Xeon.

Заключение

1) В настоящее время параллельные методы и алгоритмы недостаточно разработаны, в связи с чем невозможно эффективно использовать существующие вычислительные мощности.

2) Для успешного создания эффективных параллельных алгоритмов и программ необходимо учитывать:

а) специфику задачи и метода;

б) архитектуру вычислительного комплекса.

ЦЕНТР КОЛЛЕКТИВНОГО ПОЛЬЗОВАНИЯ ССКЦ ПРИ ИВМиМГ СО РАН

Научный руководитель: академик Б.Г. Михайленко

Исполнительный директор: д.т.н. Б.М. Глинский

Зам. исполнительного директора: д.т.н. В.Э. Малышкин

Ученый секретать: к.ф.-м.н. И.Г. Черных

В состав ЦКП ССКЦ входят следующие лаборатории ИВМиМГ:

Лаб. Сибирский суперкомпьютерный центр

Лаб. Синтеза параллельных программ

Лаб. Вычислительной физики

Лаб. Параллельных алгоритмов решения больших задач

ОСНОВНЫЕ ЗАДАЧИ ЦКП ССКЦ

- Обеспечение работ институтов СО РАН и университетов Сибири по математическому моделированию в фундаментальных и прикладных исследованиях.
- Координация работ по развитию суперкомпьютерных центров Сибири, осуществляемая Советом по супервычислениям при Президиуме СО РАН.
- Организация обучения специалистов СО РАН и студентов университетов (ММФ и ФИТ НГУ, НГТУ) методам параллельных вычислений на суперкомпьютерах (поддержка ежегодных зимних и летних школ по параллельному программированию для студентов).
- Сотрудничество с INTEL, HP и промышленными организациями, тестирование новых процессоров.
- Сетевое взаимодействие с другими Суперкомпьютерными центрами СО РАН, Москвы и других городов России, а также зарубежных стран, совместная разработка технологий распределенных вычислений.

ВЫЧИСЛИТЕЛЬНЫЕ РЕСУРСЫ ЦКП ССКЦ

ССКЦ

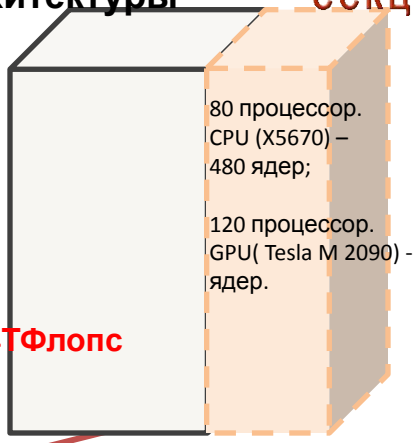
Кластер НКС-160
(hp rx1620)

168 процессор. Itanium 2, 1,6 ГГц; InfiniBand, Gigabit Ethernet (GE);
> 1 ТФлопс

Кластер НКС-30Т
(hp BL2X220c)

Общее число процессоров Intel Xeon E5450/E5540/X5670 576 (2688 ядер); InfiniBand, GE;
30 ТФлопс

Кластер гибридной архитектуры



80 процессор. CPU (X5670) – 480 ядер;

120 процессор. GPU(Tesla M 2090) - 61440 ядер.

85,4ТФлопс

GigabitEthernet
InfiniBand

СИСТЕМЫ ХРАНЕНИЯ ДАННЫХ (СХД)

Параллельная файловая система IBRIX для НКС-30Т 32 Тбайт

СХД для НКС-30Т 36 Тбайт (max-120 Тбайт)

СХД сервера с общей памятью 9 Тбайт (max-48 Тбайт)

СХД для НКС-160 3,2 Тбайт

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Системное
Общематематическое
Прикладное (ППП)

Сервер с общей памятью
(hp DL580 G5)

4 процессора (16 ядер) Intel Xeon Quad Core X7350, 2,93 ГГц;
256 Гбайт общая память;
187,5 ГФлопс

GE

GE

Сеть ИВМиМГ

Сеть Internet ННЦ

Спасибо за внимание!

Переход к безразмерным переменным

- скорость света **$c = 3 \times 10^{10}$ см/с**
- плотность плазмы **$n_0 = 10^{14}$ см$^{-3}$**
- плазменная электронная частота **$\omega_p = 1.6 \times 10^6$ сек$^{-1}$**

ГРАНТЫ, ПРИ ВЫПОЛНЕНИИ КОТОРЫХ ИСПОЛЬЗОВАЛИСЬ УСЛУГИ ЦКП ССКЦ В 2010 Г.

ССКЦ

Всего грантов, программ и проектов — 120

Из них Российских — 119, Международных — 1.

Грантов РФФИ – 41

Программ РАН – 24

Проектов СО РАН – 20

Программ Минобразнауки – 9

Другие – 26

_____

Всего публикаций – 142

Российских – 86

Зарубежных –56

Гранты по институтам:

ИВМиМГ – 30
ИВТ – 2
ИК – 11
ИКЗ (Тюмень) – 2
ИМ – 1
ИНГиГ – 4
ИНХ – 8
ИТ – 9
ИТПМ – 16
ИФП – 1
ИХБФМ – 1
ИХиХТ (Красноярск) – 3
ИХКиГ – 8
ИЦиГ – 11
ИЯФ – 4
НГТУ – 3
НГУ – 3
Унипро – 3

МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ЧИСЛЕННЫЕ МЕТОДЫ И ПАРАЛЛЕЛЬНЫЕ АЛГОРИТМЫ ДЛЯ РЕШЕНИЯ БОЛЬШИХ ЗАДАЧ СО РАН И ИХ РЕАЛИЗАЦИЯ НА МНОГОПРОЦЕССОРНЫХ СУПЕРЭВМ (МИП №26 СО РАН)

(Координатор, академик Б.Г. Михайленко)

- Проект объединяет 12 Институтов СО РАН: ИВМиМГ; ИНГГ; ИВТ; ИК; ИТПМ; ИЦиГ; ИВМ; ИФП; ИМ; ИХБиФМ; ИСЭ; ОФИМ.
- Основная цель проекта: эффективное решение больших задач СО РАН из разных научных областей на многопроцессорных суперЭВМ.
- Получены первые обобщающие результаты, как в распараллеливании алгоритмов, так и в параллельной реализации 3D моделей для различных областей науки.
- Основным инструментом при реализации проекта являются вычислительные средства ЦКП ССКЦ при МВМиМГ СО РАН.