



Комплексное тестирование высокопроизводительных вычислительных систем на основе программы для моделирования плазмы методом частиц в ячейках

А.В.Снытников,
Лаборатория Суперкомпьютерного Моделирования

Диссертация на соискание ученой степени доктора технических наук,
специальность 05.13.15
“Вычислительные машины, комплексы и компьютерные сети”

Актуальность

- необходимость создания именно **комплексного** теста производительности вычислительных систем,
- охватывающего все аспекты, влияющие на быстродействие:
 - процессорные ядра,
 - графические ускорители,
 - контроллеры памяти, межпроцессорные коммуникационные интерфейсы, контроллеры коммуникационной сети, сетевые коммутаторы,
 - средства доступа к энергонезависимой подсистеме хранения данных (дисковой подсистеме),
- позволяющего определить реальное время вычислений с учетом того, что данные находятся не в регистрах процессора и должны быть туда загружены
- таким образом выяснить реальное быстродействие системы на широком спектре задач математического моделирования.

Методология и методы исследования

- Построение комплексного теста выполняется на основе программы для моделирования методом частиц в ячейках. Данный метод хорошо подходит для тестирования ВС вследствие того, что он сочетает в себе несколько разных этапов, каждый из которых предъявляет высокие требования к какой-то определенной подсистеме тестируемой ВС. Это дает возможность именно комплексного тестирования ВС. Кроме того, данный метод является исключительно многовариантным, что дает дополнительные возможности для оценки производительности тестируемой ВС.
-
-

Основные положения, выносимые на защиту

- Пункт 1. Разработка научных основ создания вычислительных машин, комплексов и компьютерных сетей, исследования общих свойств и принципов функционирования вычислительных машин, комплексов и компьютерных сетей.
 - Методика комплексной оценки производительности ВС, основанная на программе для моделирования методом частиц в ячейках
- Пункт 2. Теоретический анализ и экспериментальное исследование функционирования вычислительных машин, комплексов и компьютерных сетей с целью улучшения их технико-экономических и эксплуатационных характеристик.
 - Методика определения для конкретной ВС абсолютной оценки, основанной на степени пригодности данной ВС для решения задач математического моделирования
- Пункт 4. Разработка научных методов и алгоритмов организации параллельной и распределенной обработки информации, многопроцессорных, многомашинных и специальных вычислительных систем.
 - Методика комплексного анализа производительности узлов мультиархитектурной ВС, оснащенной многоядерными процессорами и ускорителями вычислений

Глава 1. Описание реализации метода частиц в ячейках для высокопроизводительных ВС

$$\frac{\partial f_{i,e}}{\partial t} + \vec{v} \frac{\partial f_{i,e}}{\partial \vec{x}} + \vec{F} \frac{\partial f_{i,e}}{\partial \vec{v}} = 0$$

$$\nabla \times \vec{B} = 4\pi \vec{j} + \frac{1}{c} \frac{\partial \vec{E}}{\partial t}$$

$$\nabla \times \vec{E} = -\frac{1}{c} \frac{\partial \vec{B}}{\partial t}$$

$$\nabla \cdot \vec{E} = 4\pi \rho$$

$$\nabla \cdot \vec{B} = 0$$

• **Граничные условия:** периодические

$$\vec{p} = \gamma \vec{v}, \gamma^{-1} = \sqrt{1 - v^2}$$

$$\vec{F} = q_{i,e} \left(\vec{E} + \frac{1}{c} [\vec{v}, \vec{B}] \right)$$

$$\vec{j} = \sum_{i,e} q_{i,e} \int f_{i,e} \vec{v} d\vec{v}$$

$$\rho = \sum_{i,e} q_{i,e} \int f_{i,e} d\vec{v}$$

Начальные условия

$$\rightarrow \rho_e = 1000, \rho_b = 1$$

$$\rho = \rho_e + \rho_b$$

→ Импульсы электронов в плазме:
 $\rho_x, \rho_y, \rho_z \rightarrow$ максвелловское
 распределение, $\sigma = T_e = 1.0$

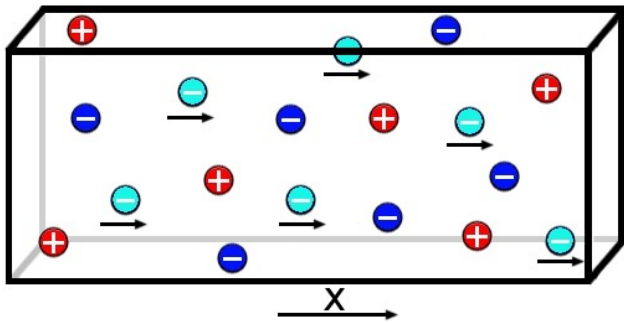
→ Импульсы ионов плазмы: 0

→ Импульс электронов пучка:
 $\rho_x = 50, \rho_y = \rho_z = 0$

Глава 1. Описание реализации метода частиц в ячейках для высокопроизводительных ВС



В методе частиц в ячейках среда разбивается на модельные частицы, траекториями движения которых являются характеристики кинетического уравнения Власова



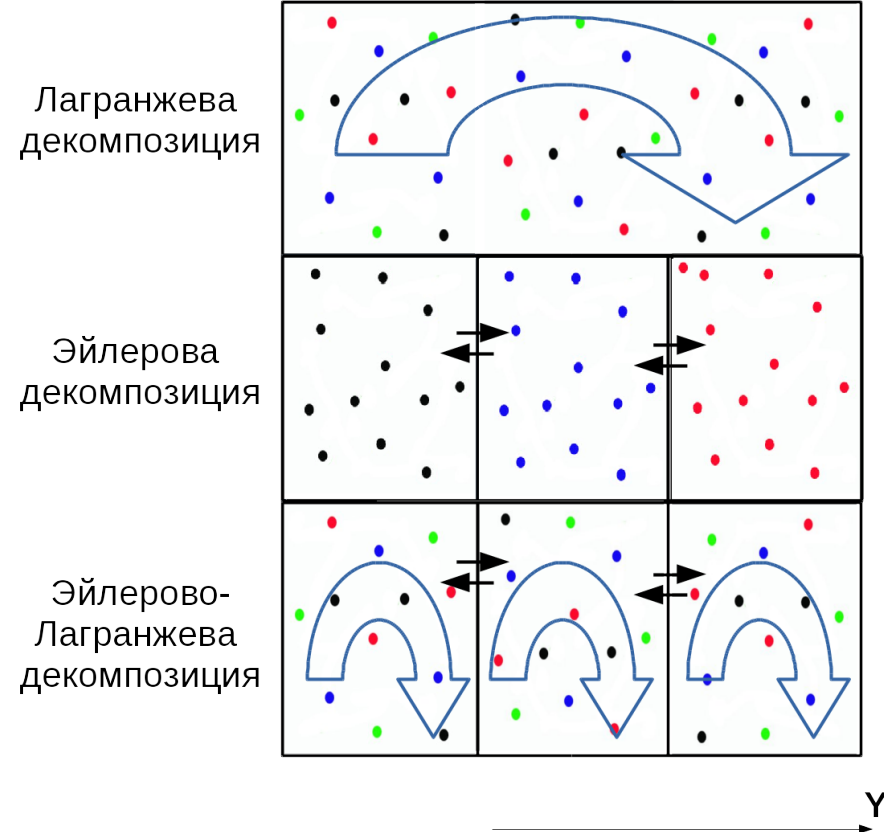
$$\frac{\partial p_{i,e}}{\partial t} = \kappa (E + [v, B]),$$
$$\frac{\partial r_{i,e}}{\partial t} = v_{i,e}.$$

$$p_{i,e} = \frac{v_{i,e}}{\sqrt{1 - v_{i,e}^2}}, \quad \kappa_e = -1, \quad \kappa_i = m_e / m_i.$$

- Уравнения движения модельных частиц решаются с помощью схемы с перешагиванием
- Уравнения Максвелла решаются методом FDTD (метод Yee)
- **Вычислительная сложность: 500 операций с плавающей точкой на частицу**

Реализация численных методов выполнена на основе работы: **Вшивков В.А.** и др.,⁶
Вычислительные технологии, Том 6, № 2, 2001.

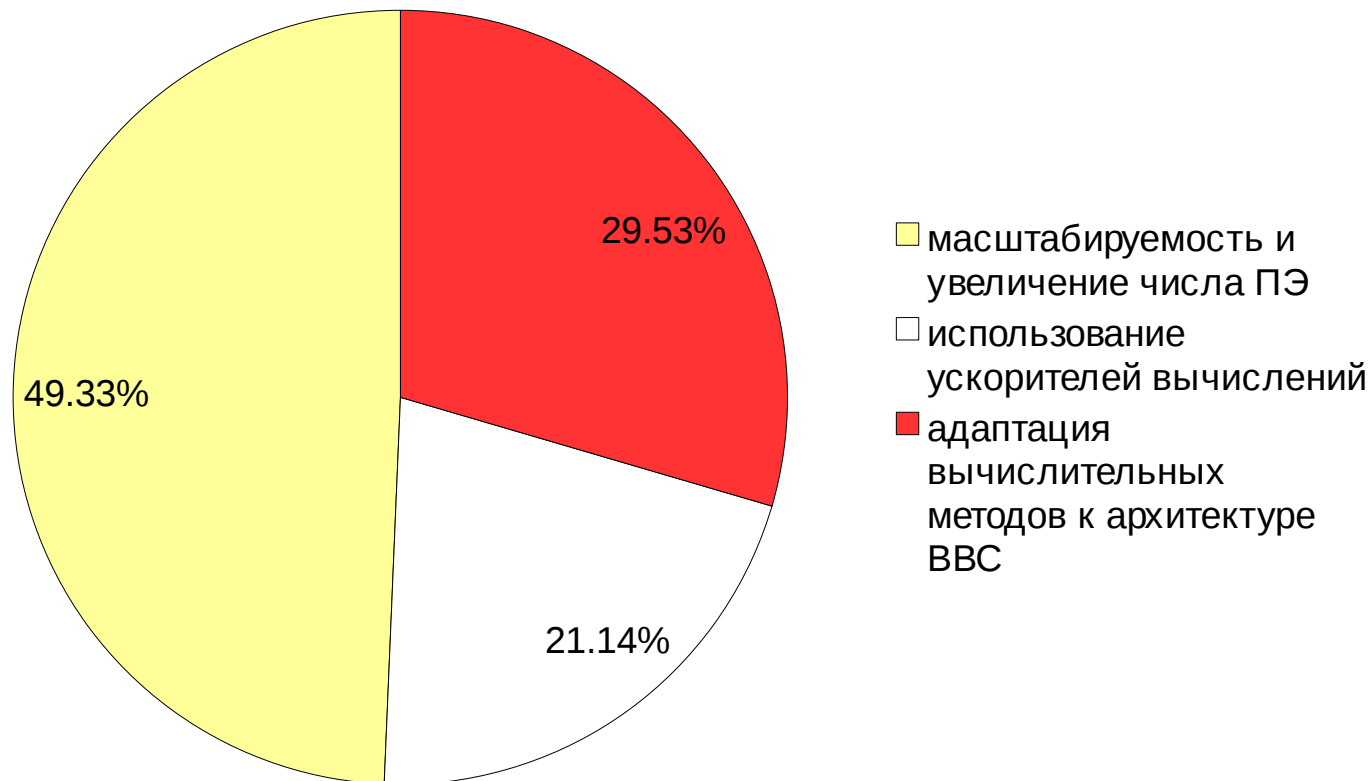
Глава 1. Описание реализации метода частиц в ячейках для высокопроизводительных ВС



- Реализовано три вида пространственной декомпозиции
- Реализация на GPU:
 - Частицы хранятся в виде массивов по ячейкам
 - Расчет движения частиц выполняется на разделяемой памяти
 - При пересылке частиц между ячейками исключена синхронизация
- Имеется методика переноса программы с GPU на Intel Xeon Phi (т.е. BBC как на основе многоядерных процессоров на основе ускорителей вычислений тестируются **на одной и той же программе**)

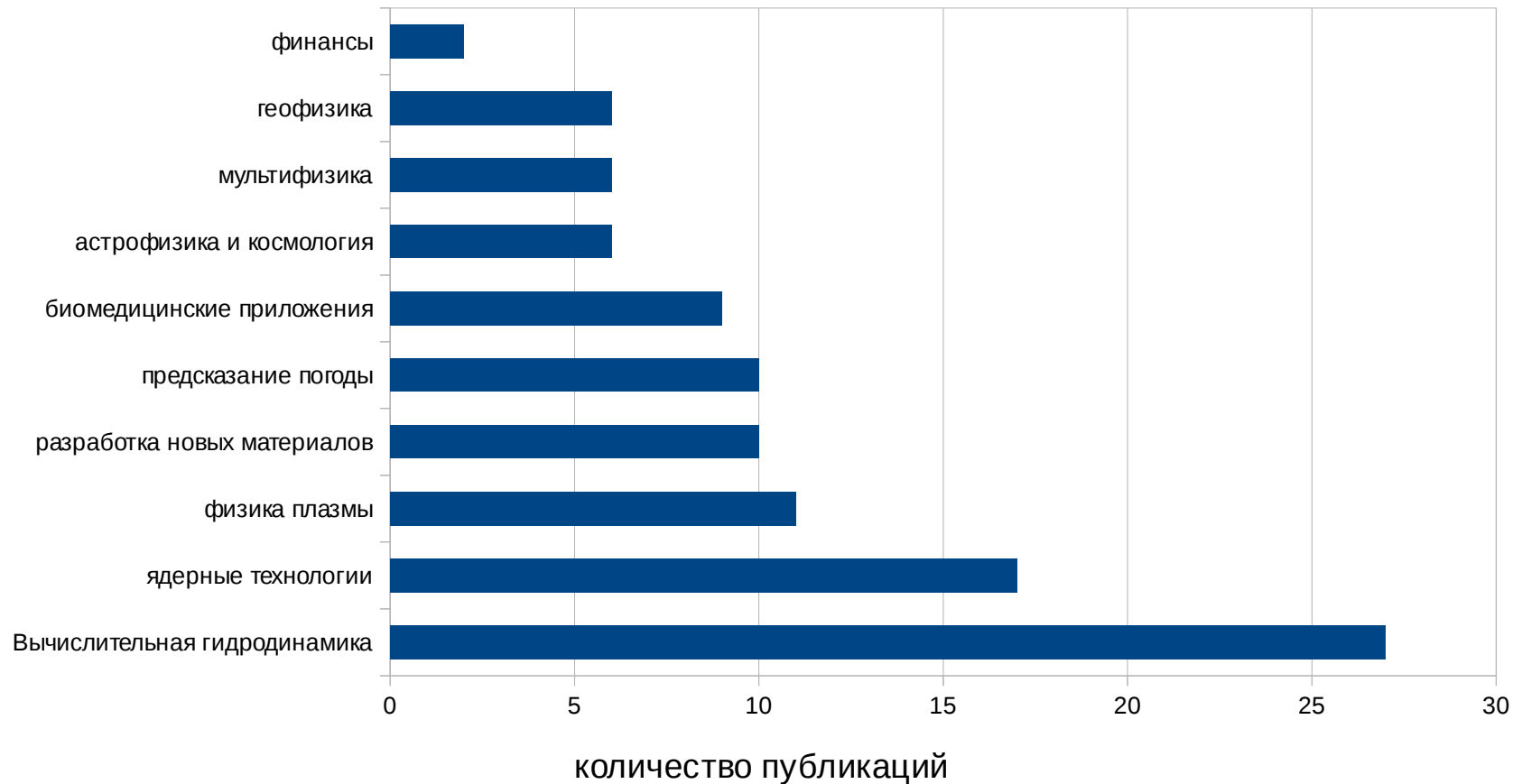
Глава 2. Физико-математические задачи и вычислительные методы в исследованиях, проводимых с использованием высокопроизводительных ВС

Обзор публикаций о перспективных ВВС



2010-2016 годы: Future Generation Computer Systems, Procedia Computers Science, Journal of Parallel and Distributed Computing, Parallel Computing, Journal of Computational Physics, Computer Physics Communications и др.

Глава 2. Физико-математические задачи и вычислительные методы в исследованиях, проводимых с использованием высокопроизводительных ВС



Глава 3. Комплексная оценка производительности ВС

- В настоящий момент списки Top50 и Top500 выстроены в порядке убывания пиковой производительности и производительности на тесте LinPack
- Это дает определенную информацию о сравнительной скорости работы представленных там машин.
- Но очень многие факторы, такие как
 - скорость работы и объем дисков,
 - Пропускная способность шины памяти и коммуникационной сети,
 - неоднородность оборудования и т.д. - остаются за пределами рассмотрения.
- А это именно те проблемы, с которыми придется столкнуться при попытке посчитать на кластере большую задачу.
- По этой причине тестирование проводится с помощью измерения времени, затрачиваемого на различные этапы программы, решающей реальную физическую задачу.

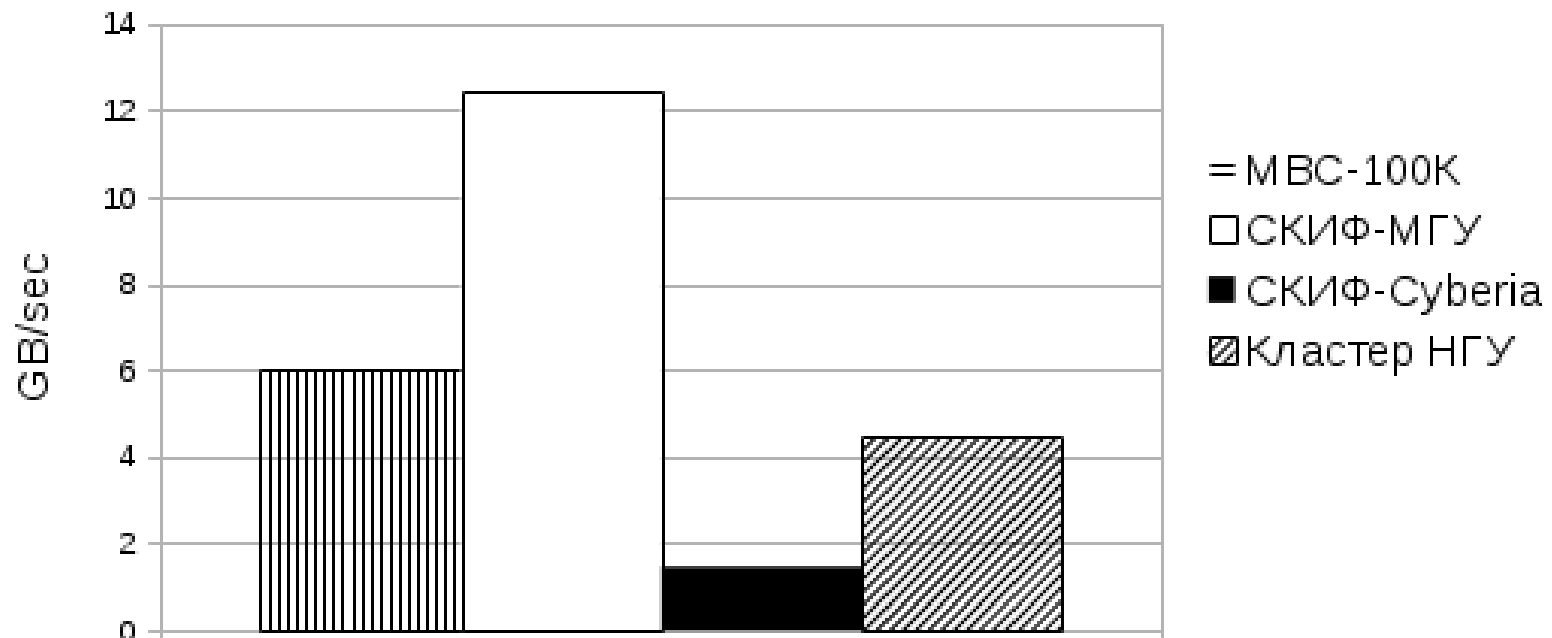
Глава 3. Комплексная оценка производительности ВС

Производительность системы памяти

$$W_{PIC,GB/sec} = \frac{W_P \times N_P P_{core}}{\Delta t}$$

- W_P - количество байт на одну модельную частицу, $W_P = 576$;
- N_P - количество модельных частиц на одно процессорное ядро (2.5×10^6);
- P_{core} - количество процессорных ядер;
- Δt - длительность временного шага, сек.

Пропускная способность памяти



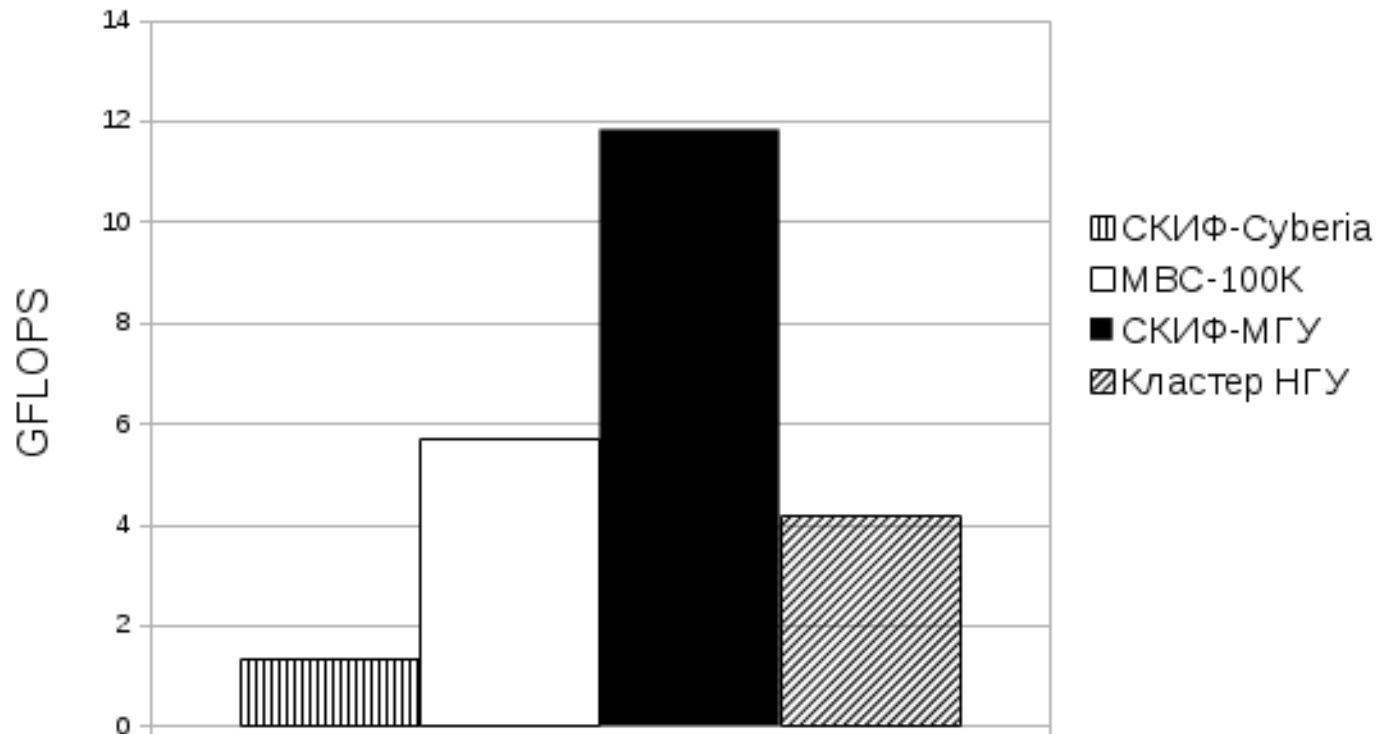
Глава 3. Комплексная оценка производительности ВС

Расчет производительности процессорных элементов на основе движения модельных частиц

$$N_{PIC,FLOPS} = \frac{F_P \times N_P \times P_{core}}{\Delta t}$$

- F_P - количество операций на одну модельную частицу, $F_P = 500$;
- N_P - количество модельных частиц на одно процессорное ядро (2.5×10^6);
- P_{core} - количество процессорных ядер;
- Δt - длительность временного шага. сек.

Производительность процессоров



Глава 3. Комплексная оценка производительности ВС

Расчет производительности процессорных элементов на основе движения модельных частиц

$$N_{PIC,FLOPS} = \frac{F_P \times N_P \times P_{core}}{\Delta t}$$

- F_P - количество операций на одну модельную частицу, $F_P = 500$;
- N_P - количество модельных частиц на одно процессорное ядро;
- P_{core} - количество процессорных ядер;
- Δt - длительность временного шага, сек.



Глава 3. Комплексная оценка производительности ВС

Производительность коммуникационной сети. Collective.

$$W_A = \frac{N_X \times N_Y / P_{SUB} \times N_Z \times 24}{T_A}$$

N_X, N_Y, N_Z - количество узлов сетки по X, Y и Z соответственно;

P_{SUB} - количество подобластей (если используется эйлерова декомпозиция)

T_A - длительность операции MPI_Allreduce (суммирование токов по всей области), сек.

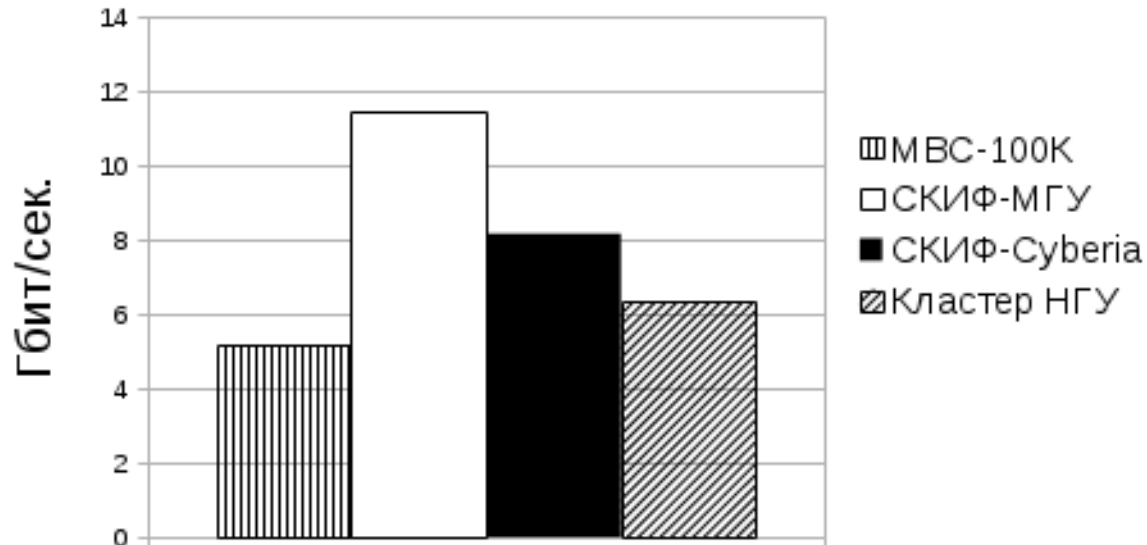
Глава 3. Комплексная оценка производительности ВС

Производительность коммуникационной сети. Peer-to-peer

$$W_S = \frac{U_P \times \nu N_P \times P_{core}}{T_{S,PIC}}$$

- U_P - количество байт на одну модельную частицу(48)
- N_P - количество модельных частиц на одно процессорное ядро (2.5×10^6);
- P_{core} - количество процессорных ядер;
- ν - доля пересылаемых частиц (nu 0.05);
- $T_{S,PIC}$ - время пересылки частиц, сек.

Производительность коммуникационной сети



Глава 3. Комплексная оценка производительности ВС

Формула для комплексной оценки ВС

$$\left. \begin{aligned} W_S &= \frac{U_P \times \nu N_P \times P_{core}}{T_{S,PIC}} \\ W_A &= \frac{N_X \times N_Y / P_{SUB} \times N_Z \times 24}{T_A} \end{aligned} \right\} \approx W_{PIC,GB/sec} = \frac{W_P \times N_P P_{core}}{\Delta t}$$

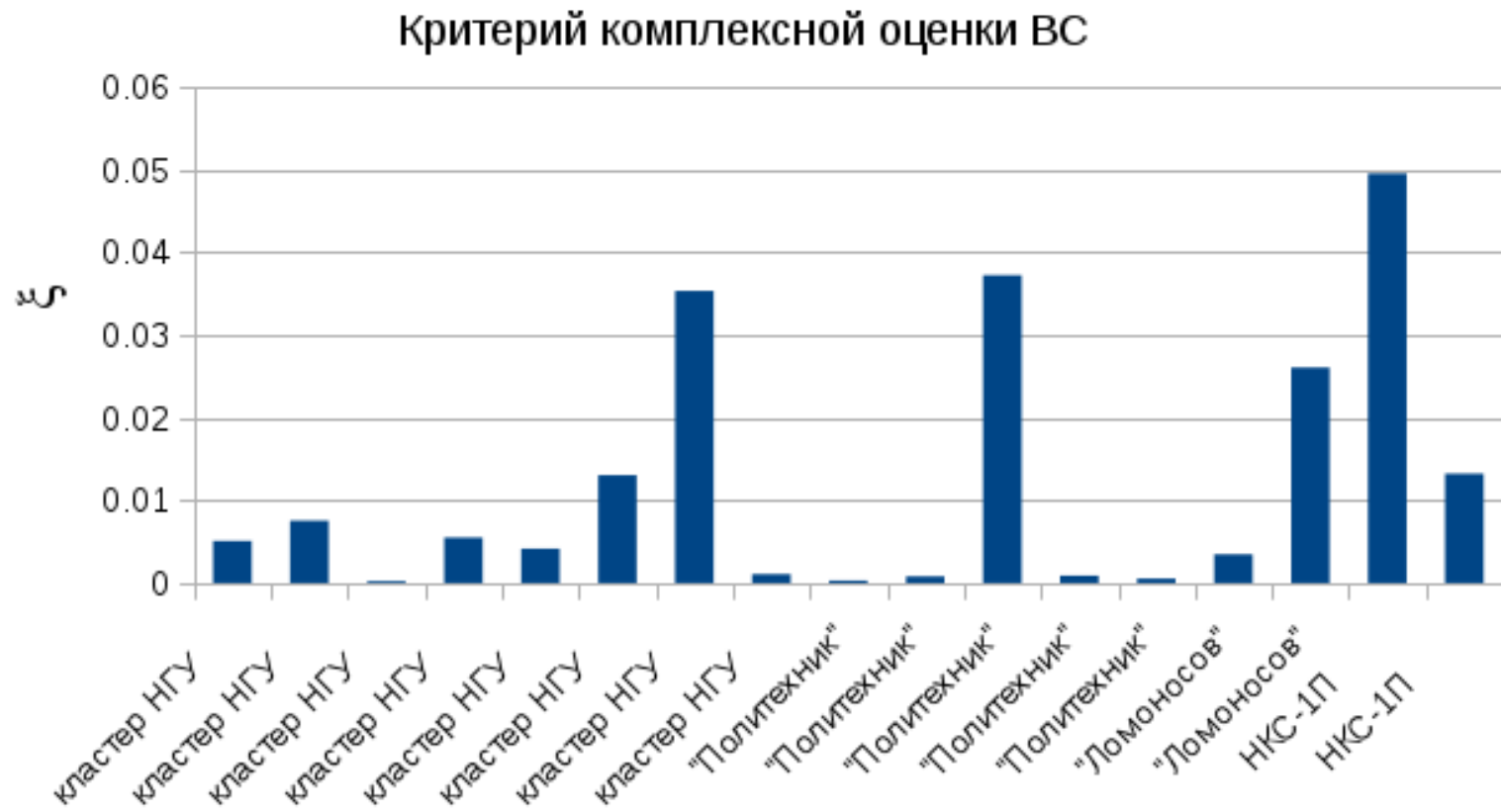
Для того, чтобы параллельная ВС могла быть признана адаптированной к задачам математического моделирования, она должна соответствовать следующим требованиям:

- Очень высокая производительность коммуникационной сети (W_S , и W_A), позволяющая пересылать все необходимые для расчета данные, не задерживая вычислений;
- Относительно высокая производительность оперативной памяти ($W_{PIC,GB/sec}$), позволяющая эффективно использовать ресурсы процессоров, т.е. **фактически совпадающая** с производительностью процессора ($N_{PIC, FLOPS}$).

Глава 3. Комплексная оценка производительности ВС

Формула для комплексной оценки ВС

$$\xi = \frac{W_{MPI}}{W_{PIC,GB/sec}},$$



Глава 4. Анализ масштабируемости, параллельной эффективности и ускорения параллельной ВС

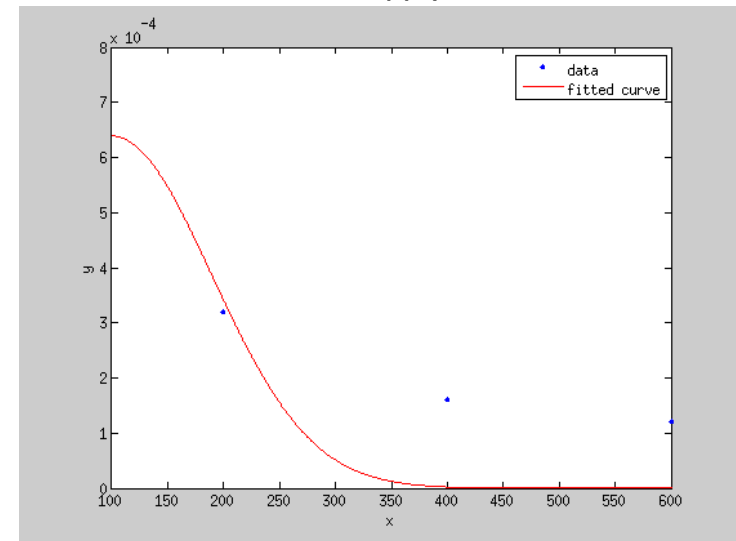
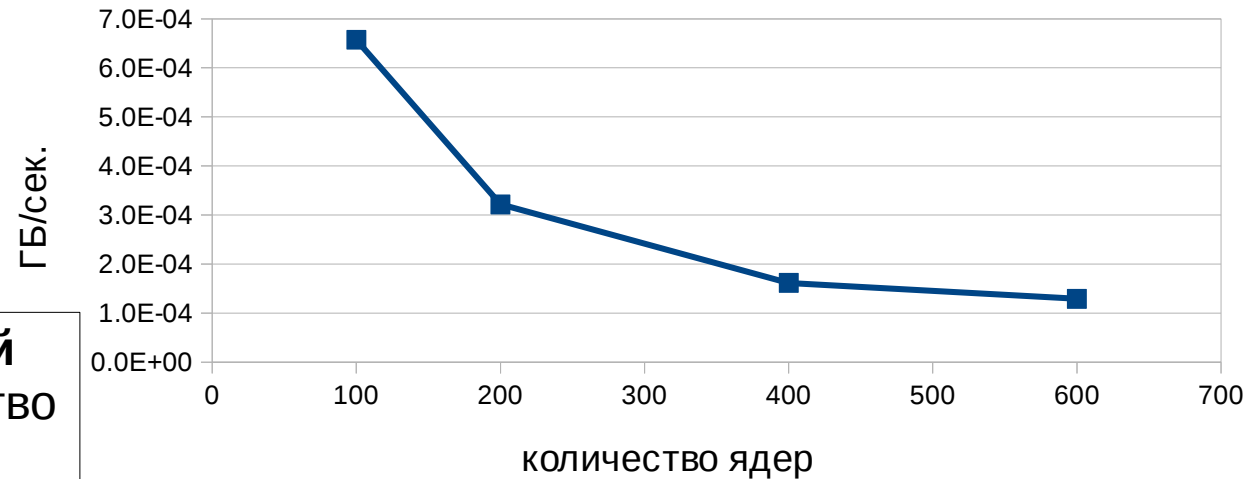
Вычисление характеристик коммуникационного оборудования ВС на основе измеренной масштабируемости метода частиц в ячейках

Производительность коммуникационной сети, W_A
кластер НГУ

$$W_A = \frac{N_X \times N_Y / P_{SUB} \times N_Z \times 24}{T_A}$$

Эффективный коммуникационный размер ВС - максимальное количество процессоров, которое может быть в рамках данной ВС эффективно использовано для решения одной задачи.

Количество процессоров, для которого производительность коммуникационной сети падает не более чем в **e** раз по сравнению с максимальной: **226**



Глава 4. Анализ масштабируемости, параллельной эффективности и ускорения параллельной ВС

Формулы для анализа данных о масштабируемости

$$\eta_N^{weak} = \frac{T^1(1)}{T^N(N)} \quad (4.3)$$

здесь $T^K(N)$ обозначает время счета задачи с характерной размерностью K при использовании N процессоров. это время состоит из двух основных частей:

- собственно времени счета $T_P^K(N)$;
- времени коммуникаций $T_C^K(N)$;

Таким образом,

$$\eta_N^{weak} = \frac{T^1(1)}{T_P^N(N) + T_C^N(N)} \quad (4.4)$$

при выполнении вычислений одновременно на однотипных процессорах можно считать, что $T^1(1) = T_C^N(N)$, таким образом формула 4.4 приводится к виду:

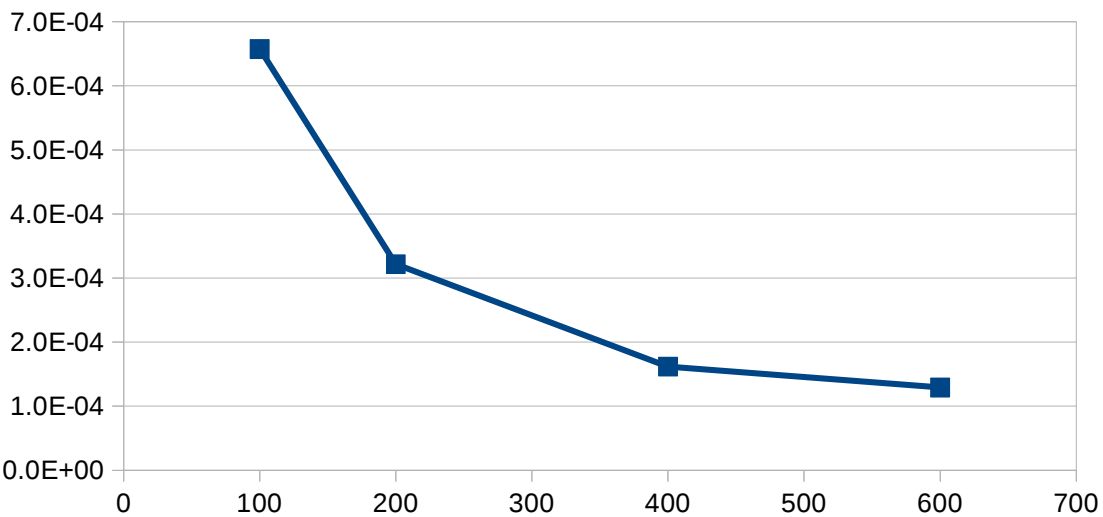
$$\eta_N^{weak} = \frac{1}{1 + \frac{T_C^N(N)}{T^1(1)}}. \quad (4.5)$$

Это означает, что при известном времени расчета задачи на одном процессоре можно восстановить время коммуникаций по эффективности в слабом смысле:

$$T_C^N(N) = T^1(1) \left(\frac{1}{\eta_N^{weak}} - 1 \right) \quad (4.6)$$

далее время коммуникаций может быть источником для вычисления производительности коммуникационной сети при известном объеме пересылок.

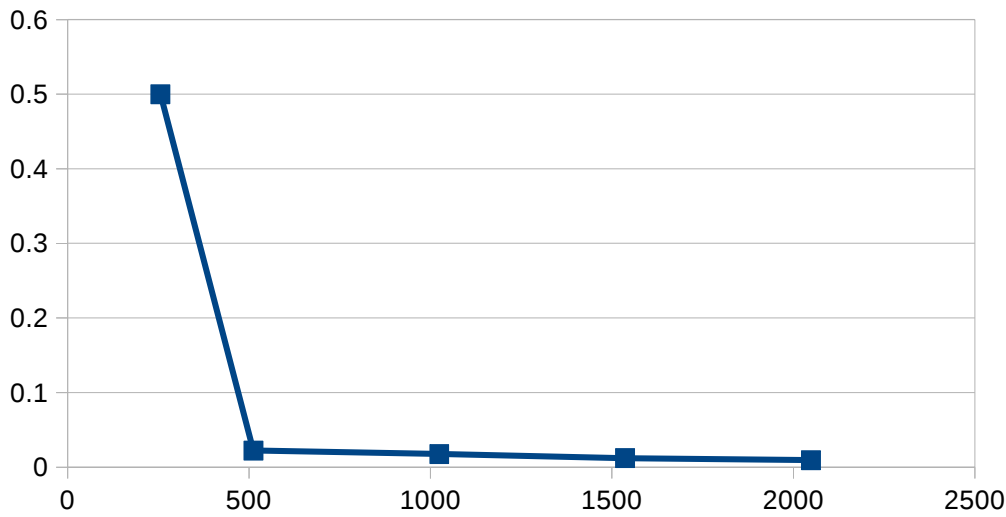
Кластер НГУ



**Эффективный
коммуникационный
размер**

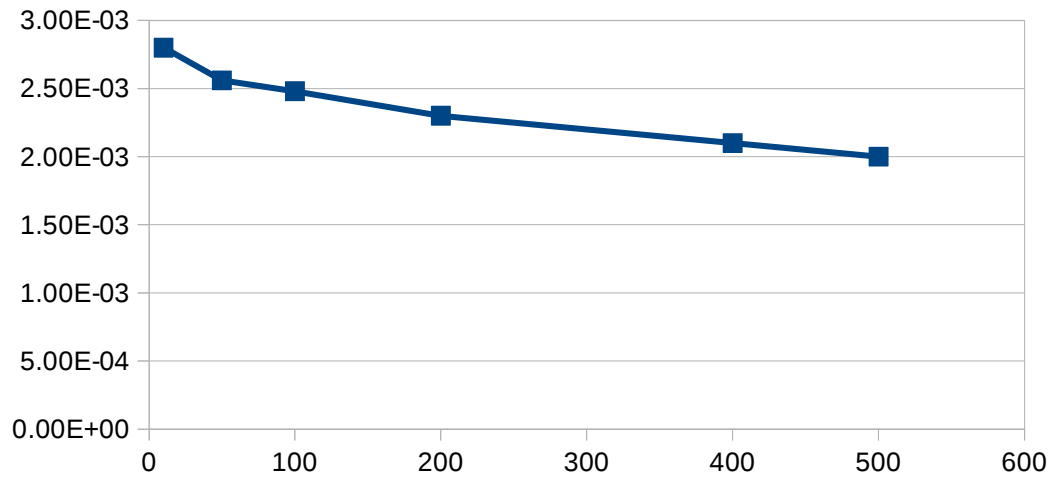
226

MBC-100K



425

Ломоносов
(раздел GPU)



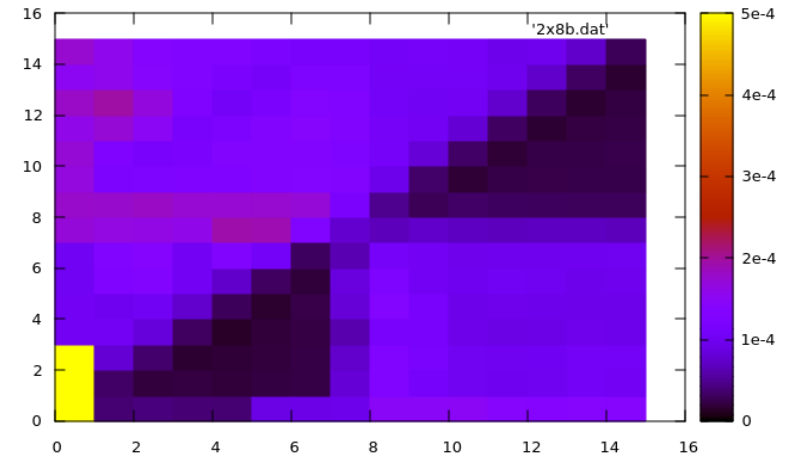
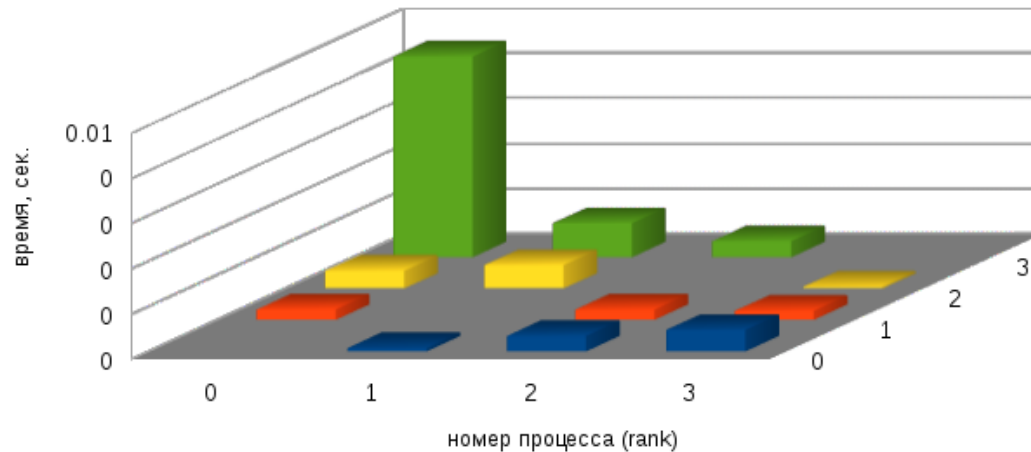
239

Глава 4. Анализ масштабируемости, параллельной эффективности и ускорения параллельной ВС

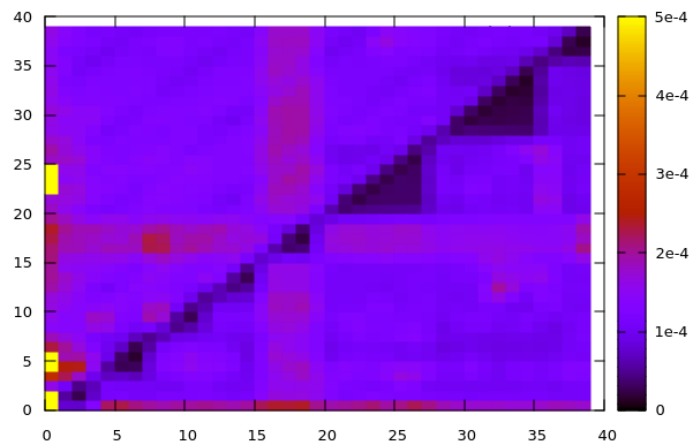
Определение коммуникационной структуры ВС

Длительность парной пересылки (MPI_Send)

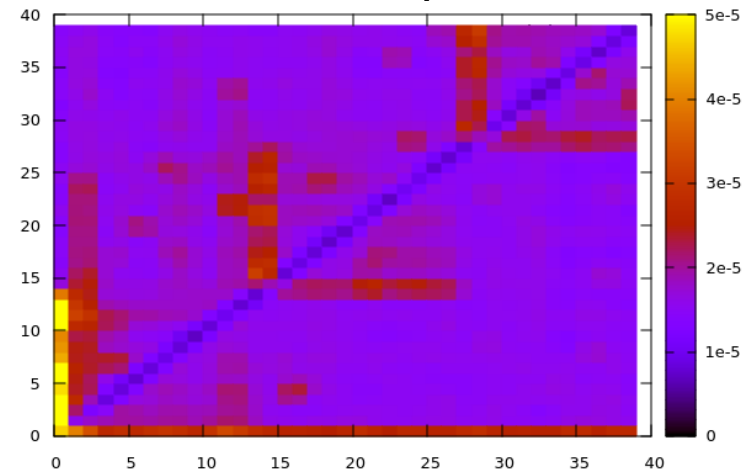
2 узла по 2 процесса



16 MPI-процессов, 2 узла по 8 процессов,
кластер НГУ



40 MPI-процессов, 10 узлов по 4 процесса,
кластер НГУ

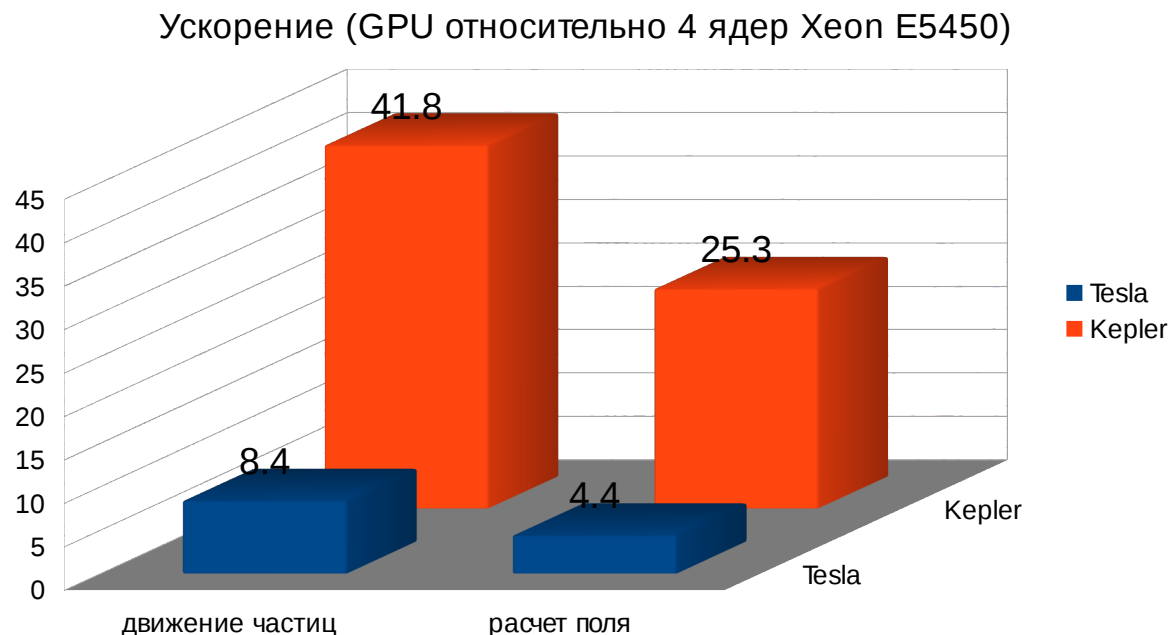


40 MPI-процессов, 10 узлов по 4 процесса,
кластер "Политехник"

Глава 5. Анализ производительности узлов мультиархитектурной ВС

Анализ производительности узлов с графическими ускорителями

- Движение модельных частиц является наиболее времяемкой частью расчета (до 90 %)
- В то же время именно эта часть алгоритма наиболее заметно ускоряется при переходе на GPU
- В перспективе это дает возможность провести трехмерное моделирование кинетического режима развития неустойчивости



Глава 5. Анализ производительности узлов мультиархитектурной ВС

Прогноз эффективности ВС на основе графических ускорителей

Вопрос, который наиболее часто задают специалисты по математическому моделированию применительно к мультиархитектурной ВС, оснащенной графическими ускорителями - это возможность **эффективной** реализации конкретного вычислительного алгоритма на данной мультиархитектурной ВС.

Для ответа на данный вопрос предлагается интерполяционная формула:

$$v_{pre} = v_{PIC} k + (1 - k) v_{B,E}$$

- v_{pre} - оценка скорости вычислений на GPU для рассматриваемого алгоритма,
- v_{PIC} - скорость вычислений с на этапе сдвига модельных частиц
- $v_{B,E}$ - на этапе расчета электромагнитного поля
- k - интерполяционный множитель

Глава 5. Анализ производительности узлов мультиархитектурной ВС

Прогноз эффективности ВС на основе графических ускорителей

большинство численных методов используемых в математическом моделировании находятся в промежуточном положении по отношению к используемым в методе частиц в ячейках алгоритму вычисления поля и алгоритму расчета движения частиц по следующим показателям:

- 1) **вычислительной интенсивности** (равномерное распределение вычислительно сложных фрагментов по тексту или отдельные высоконагруженные участки);
- 2) характеру **доступа к оперативной памяти** (регулярный или нерегулярный);
- 3) **объему используемых данных** (большой или маленький).

$$v_{pre} = v_{PIC}k + (1 - k)v_{B,E}$$

Вычислительный алгоритм	Интенсивность	Доступ к оперативной памяти	Объем данных	k
Расчет движения модельных частиц	низкая	нерегулярный	большой	1.0
Метод Монте-Карло	низкая	нерегулярный	средний	0.9
Метод SPH	низкая	нерегулярный	небольшой	0.6
Метод конечных элементов	высокая	нерегулярный	большой	0.5
Конечно-разностные схемы (явные)	высокая	регулярный	большой	0.2
Конечно-разностные схемы (явные)-2	высокая	регулярный	большой	0.1
Вычисление электромагнитного поля	высокая	регулярный	большой	0.0

Глава 5. Анализ производительности узлов мультиархитектурной ВС

Прогноз эффективности ВС на основе графических ускорителей

Вопрос, который наиболее часто задают специалисты по математическому моделированию применительно к мультиархитектурной ВС, оснащенной графическими ускорителями - это возможность **эффективной** реализации конкретного вычислительного алгоритма на данной мультиархитектурной ВС.

Для ответа на данный вопрос предлагается интерполяционная формула:

$$v_{pre} = v_{PIC} k + (1 - k) v_{B,E}$$

- v_{pre} - оценка скорости вычислений на GPU для рассматриваемого алгоритма,
- v_{PIC} - скорость вычислений с на этапе сдвига модельных частиц
- $v_{B,E}$ - на этапе расчета электромагнитного поля
- k - интерполяционный множитель

Глава 5. Анализ производительности узлов мультиархитектурной ВС

Прогноз эффективности ВС на основе графических ускорителей

$$v_{pre} = v_{PIC}k + (1 - k)v_{B,E}$$

Вычислительный алгоритм	Интенсивность	Доступ к оперативной памяти	Объем данных	k
Расчет движения модельных частиц	низкая	нерегулярный	большой	1.0
Метод Монте-Карло	низкая	нерегулярный	средний	0.9
Метод SPH	низкая	нерегулярный	небольшой	0.6
Метод конечных элементов	высокая	нерегулярный	большой	0.5
Конечно-разностные схемы (явные)	высокая	регулярный	большой	0.2
Конечно-разностные схемы (явные)-2	высокая	регулярный	большой	0.1
Вычисление электромагнитного поля	высокая	регулярный	большой	0.0

Заключение

- В работе предложена и реализована оригинальная методика комплексного тестирования мультиархитектурных параллельных вычислительных систем, основанная на используемой в реальных расчетах программе для математического моделирования. а именно на реализации метода частиц в ячейках.
- Особенности предложенной методики комплексного тестирования являются возможность определения для конкретной ВС абсолютной оценки, основанной на степени пригодности данной ВС для решения задач математического моделирования, метод измерения возрастания потока данных в коммуникационной сети ВС, а также оценка эффективности реализации вычислительных алгоритмов для мультиархитектурных ВС.

Рекомендации и перспективы дальнейшей разработки темы

- Основным направлением совершенствования разработанного теста является автоматическая выработка рекомендаций по оптимизации кода, охватывающих не только метод частиц в ячейках но и остальные наиболее часто используемые в математическом моделировании методы, под конкретную протестированную ВС.
- Одним из наиболее важных вариантов дальнейшего развития программы-теста, созданного в диссертационной работе является перенос на не охваченные в текущем варианте платформы, такие как Android, процессоры архитектуры Sunway, графические ускорители, использующие технологию OpenCL, в частности, AMD Firestream, а также на программируемые логические интегральные схемы (ПЛИС).
- Кроме того, необходимостью является апробация теста на крупных высокопроизводительных ВС мощностью более петафлопса, а также - возможно, после специальной адаптации схемы расчета движения частиц - на ВС векторной архитектуры.

Научные задачи

- Создать методику комплексной оценки производительности ВС в целом, основанную на программе для моделирования методом частиц в ячейках, с использованием крупноблочного распараллеливания на основе MPI.
- Разработать и протестировать формулы, обеспечивающие возможность экстраполяции результатов тестирования ускорения и эффективности распараллеливания по части узлов ВС на ВС в целом
- Создать методику проведения комплексного анализа производительности узлов мультиархитектурной ВС, оснащенной многоядерными процессорами и ускорителями вычислений на основе метода частиц в ячейках, с использованием мелкозернистого распараллеливания на основе CUDA и OpenMP.
- Разработать методику обобщения данных, полученных в ходе тестирования на основе метода частиц в ячейках для эффективной эксплуатации тестируемой ВС для широкого круга научных и народнохозяйственных задач, а также для улучшения быстродействия данной ВС.

Новизна

- Существенная научная новизна определяется тем, что метод частиц в ячейках является ресурсоемким методом, предъявляющим высокие требования к быстродействию не только процессора, но и оперативной памяти, и системы хранения данных. Высокая требовательность метода частиц к быстродействию ВС позволяет эффективно ценить быстродействие тестируемой ВС с той точки зрения, что если какая-либо из подсистем не оптимизирована и функционирует с низкой производительностью, то итоговая скорость счета по методу частиц будет очень мала. Далее известное разбиение метода частиц в ячейках на этапы различной сложности позволяет определить какая из подсистем тестируемой ВС вносит наибольший вклад в снижение производительности.
- Таким образом впервые появляется возможность точной диагностики ВС по результатам работы реализации вычислительного метода, используемой для получения актуальных физических результатов.
- Преимущества по сравнению с другими вариантами тестирования заключаются в
 - выявлении производительности ВС на реальном приложении, а не на специально сконструированном тесте - это важно с точки зрения динамики нагрузки на ВС (объем перемысливаемых частиц, например)
 - оценке интегральной производительности ВС, складывающейся из производительности процессора, ускорителей вычислений, кэш-памяти разных уровней, оперативной памяти и системы хранения данных
 - возможности выявления в ходе одного теста конкретной подсистемы ВС, наиболее заметно снижающих скорость счета реальных приложений.

Практическая значимость

- Созданный в рамках диссертационной работы комплексный тест позволяет проводить экспериментальное исследование функционирования вычислительных машин, комплексов и компьютерных сетей в широком диапазоне от больших высокопроизводительных ВС и метакомпьютеров, образованных соединением нескольких кластеров до ноутбуков.
- Это исследование дополняется проводимым в той же тестовой программе теоретическим анализом масштабируемости и параллельной эффективности. В ходе тестирования измеряются характеристики ВС - вычислительная производительность, скорость работы коммуникационной сети и системы ввода-вывода. Измеренные характеристики позволяют улучшить работы сетевой подсистемы ВС путем перепрограммирования сетевых адаптеров и изменения настроек системного ПО, а также дать рекомендации по улучшению технико-экономических и эксплуатационных характеристик ВС такого типа в дальнейшем.
- Реализация метода частиц в ячейках в представленной программе включает в себя также и возможность счета на графических ускорителях, и ускорители Intel Xeon Phi. Таким образом проведенные тесты покажут эффективность вычислений с использованием специализированных вычислителей в рамках анализируемой ВС, а также эффективность работы интерфейса, соединяющего графические и другие ускорители с основной частью ВС.
- В силу того, что скорость расчета по методу частиц в целом и эффективность распараллеливания в частности принципиально зависит от топологии внутренней сети высокопроизводительной ВС, а также от параметров коммуникационного оборудования и системного ПО, то результаты тестирования могут быть использованы для разработки создания структур и топологий компьютерных сетей, сетевых протоколов и служб передачи данных в компьютерных сетях.

•