Институт вычислительной математики и математической геофизики СО РАН

ОТОБРАЖЕНИЯ ПАРАЛЛЕЛЬНЫХ АЛГОРИТМОВ НА СУПЕРКОМПЬЮТЕРЫ ЭКЗАФЛОПСНОЙ ПРОИЗВОДИТЕЛЬНОСТИ НА ОСНОВЕ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Б.М. Глинский, А.С. Родионов, М.А. Марченко, Д.И. Подкорытов, Д.А. Караваев

http://www2.sscc.ru/

Благодарности

Работа выполнена в рамках выполнения Госконтракта Минобрнауки № 07.514.11.4016 (2011-2012гг) «Исследования и разработка методов имитационного моделирования функционирования гибридных экзафлопсных вычислительных систем»

Поддержана проектами РФФИ № 13-07-589, МИП 130 СО РАН, Программой РАН 4.9

ЦЕЛИ РАБОТЫ

1. Разработка новых эффективных методов имитационного моделирования параллельных программ, предназначенных для исполнения на суперкомпьютерах с пета и экзафлопсным уровнем производительности

2. Исследование возможности отображения параллельных алгоритмов на различные архитектуры суперЭВМ экзафлопсной производительности

ПРОБЛЕМЫ МАСШТАБИРОВАНИЯ АЛГОРИТМОВ

(тезисы)

- Исследование свойств масштабируемости параллельных алгоритмов является важной задачей при оценке эффективности их реализации, как для настоящих, так и будущих суперкомпьютеров пета- и экзафлопсного уровня.
- Данная проблема 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.
- Оценить поведение алгоритмов можно путем реализации их на имитационной модели, отображающей тысячи и миллионы вычислительных ядер.
- Имитационная модель позволяет выявить узкие места в алгоритмах, понять, как нужно модифицировать алгоритм, какие параметры необходимо настраивать при его масштабировании на большое количество ядер при заданной архитектуре вычислительной системы.

МОДЕЛЬ ЭКЗАФЛОПСНОЙ СУПЕРЭВМ

1. В настоящее время нет определенного мнения по архитектуре ЭВМ экзафлопсной производительности. В качестве одного из возможных решений предполагаем экстенсивное развитие инструментального вычислительного кластера НКС-30Т+GPU ЦКП ССКЦ СО РАН (многократное увеличение количества существующих ядер). Тем самым делается оценка производительности «снизу», поскольку естественно ожидать повышения характеристик ядер и интерконнекторов ЭВМ ЭП по сравнению с существующими.

2. Модель программы представляется взвешенным графом переходов между блоками программы с указанием параллельных ветвей. Временные задержки в блоках определяются на основе измерений, производимых в тестовых прогонах реальных программ на НКС-30Т+GPU. Прогоны реальных программ на конфигурациях с более чем 30 000 ядер позволяют надеяться на учёт в измеренных задержках эффектов от системной составляющей.

АГЕНТНО-ОРИЕНТИРОВАННАЯ СИСТЕМА ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ (AGNES)

AGNES (AGent NEtwork Simulator) – среда имитационного моделирования, создана на Java на основе JADE и состоит из двух типов агентов:

- управляющие агенты (УА), создающие среду моделирования;
- функциональные агенты (ФА), образующие модель, работающую в среде моделирования.

Достоинства пакета AGNES:

- отказоустойчивость;
- сбалансированное распределение нагрузки;
- наличие проблемно-ориентированных библиотек агентов;
- возможность динамического изменения модели в ходе эксперимента.

AGNES

Приложение AGNES – это распределенная мульти-агентная система, называемая платформой. Состоит из системы контейнеров, распределенных в сети. Обычно на каждом хосте находится по одному контейнеру (может быть несколько). Агенты существуют внутри контейнеров.

В качестве ***атомарной частицы*** в модели вычислений выбран ***вычислительный узел и исполняемый на нем код алгоритма***. Функциональный агент эмулирует поведение вычислительного узла кластера и программу вычислений на этом узле.

Вычисления - набор примитивных операций (вычисление на ядре; запись/чтение данных в память; парный обмен данными; синхронизация данных между вычислителями) и временных характеристик каждой операции.

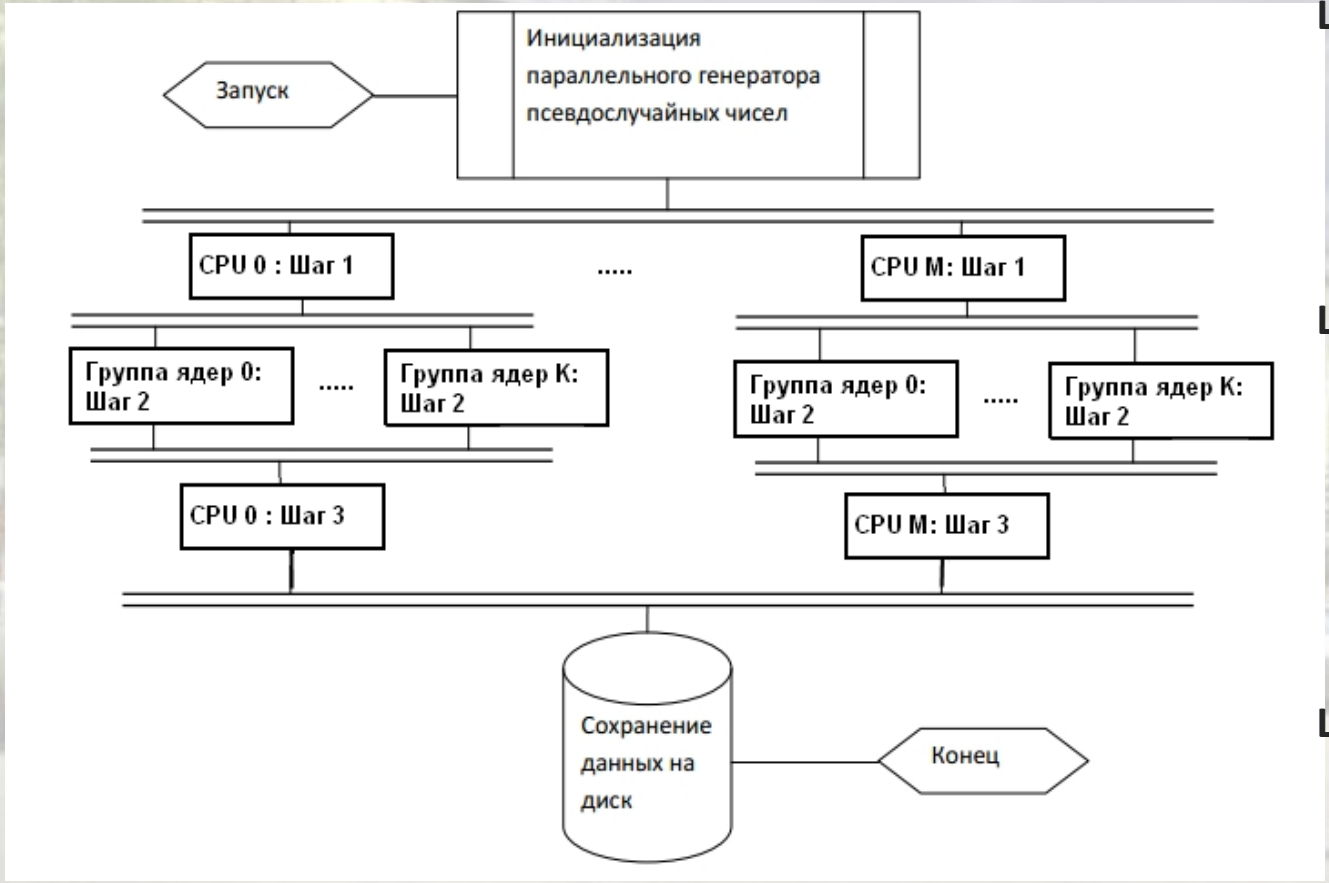
Система AGNES установлена в ЦКП ССКЦ ИВМиМГ СО РАН и доступна по ссылке

http://www2.sscc.ru/PPP/Mat-Libr/agnes.htm

Пример 1: Исследование масштабируемости распределённого статистического моделирования

- Задачи моделирования течений разреженного газа с учетом химических реакций, задачи переноса излучения и теории дисперсных систем и др.
- Архитектура: однородный MPP кластер
- **Масштабируемость** – точность вычислений методов МК зависит от количества независимых реализаций, поэтому увеличивая количество вычислителей (соответственно увеличивая количество реализаций в единицу времени), ожидаем пропорциональное уменьшение общего времени счета, при заданном уровне погрешности.

СХЕМА ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ МЕТОДОВ МОНТЕ-КАРЛО



Шаг 1: Подготовка к моделированию независимых реализаций на группах ядер

Шаг2: Моделирование реализаций, вычисление выборочных средних для группы

Шаг 3: сбор и осреднение данных

ИМИТАЦИЯ ВЫЧИСЛЕНИЙ МЕТОДОВ МОНТЕ-КАРЛО С ИСПОЛЬЗОВАНИЕМ AGNES

Используются два класса функциональных агентов:

- DataAgregator: ядро-«сборщик», собирает информацию об вычислениях, обрабатывает и агрегирует её.
- MonteCarlo: агент, имитирующий расчет методов Монте-Карло, ядро-«вычислитель». Каждый агент проводит независимые вычисления согласно схеме вычислений и взаимодействует только с соответствующим DataAgregator.

В результате работы модели собираются следующие отчеты:

- Набор времен, потраченных на каждую итерацию вычислений каждым агентом.
- Информация о количестве итераций вычислений, совершенных каждым агентом MonteCarlo.
- Информация об интенсивности получения данных агентами DataAgregator от вычислителей.

ИМИТАЦИЯ ИСПОЛНЕНИЯ МЕТОДА МОНТЕ-КАРЛО НА СУПЕРКОМПЬЮТЕРЕ
(MPP-архитектура)

Для ускорения расчётов в модели вводится двухуровневая система ядер-«сборщиков», что позволяет разгрузить центральное ядро-«сборщик».

Использовались следующие значения количества ядер-«сборщиков» на дополнительном уровне: *N=0* – случай только с одним ядром-«сборщиком»; *N=10; N=20; N=100*

Увеличение числа ядер-«сборщиков» привело к увеличению эффективности расчётов – большему ускорению от распараллеливания.

Исходные данные для имитационного моделирования получены с использованием библиотеки PARMONC:

http://www2.sscc.ru/SORAN-INTEL/paper/2011/parmonc.htm

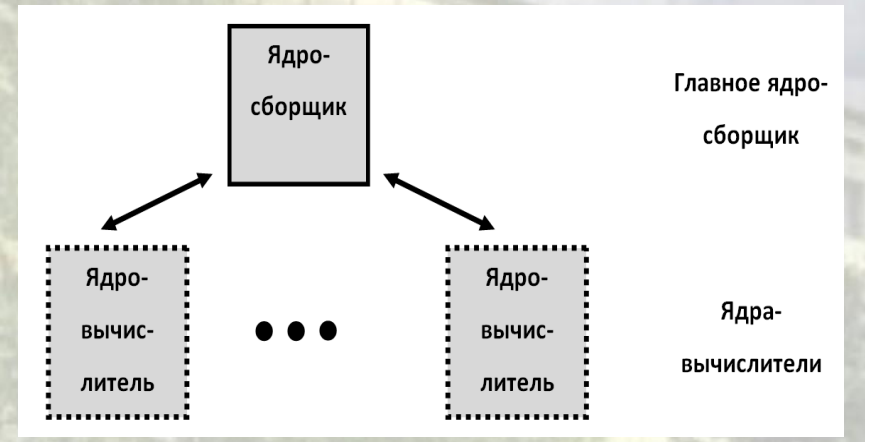
Ускорение от распараллеливания при расчётах на *M* ядрах определим так:

$$S_L(M) = \frac{T_L(M_{min})}{T_L(M)}$$

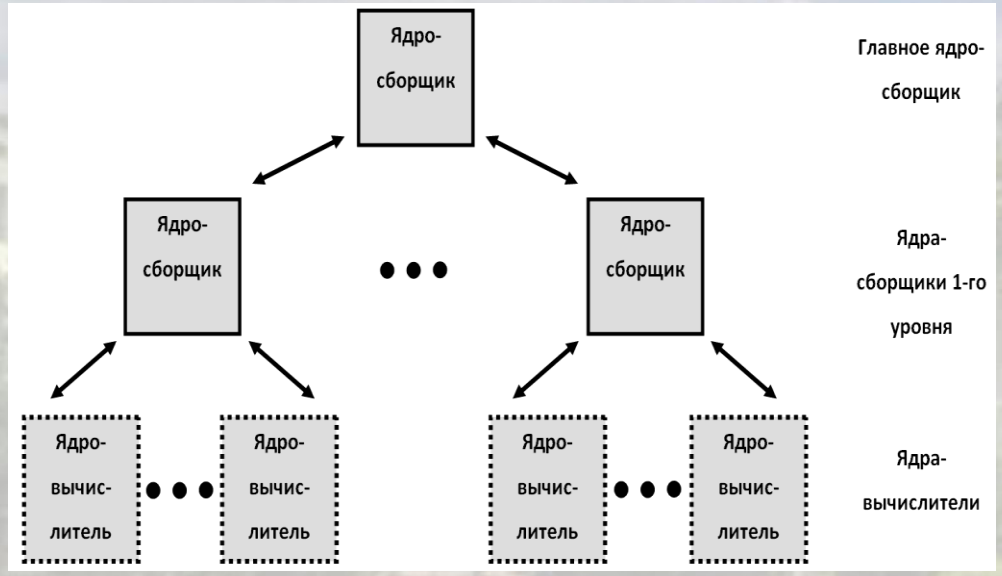
или, в предложении о пренебрежимо малом времени на обмен данными

$$S_L(M) = \frac{M}{M_{min}}$$

СХЕМА МОДЕЛИ МЕТОДА МОНТЕ-КАРЛО



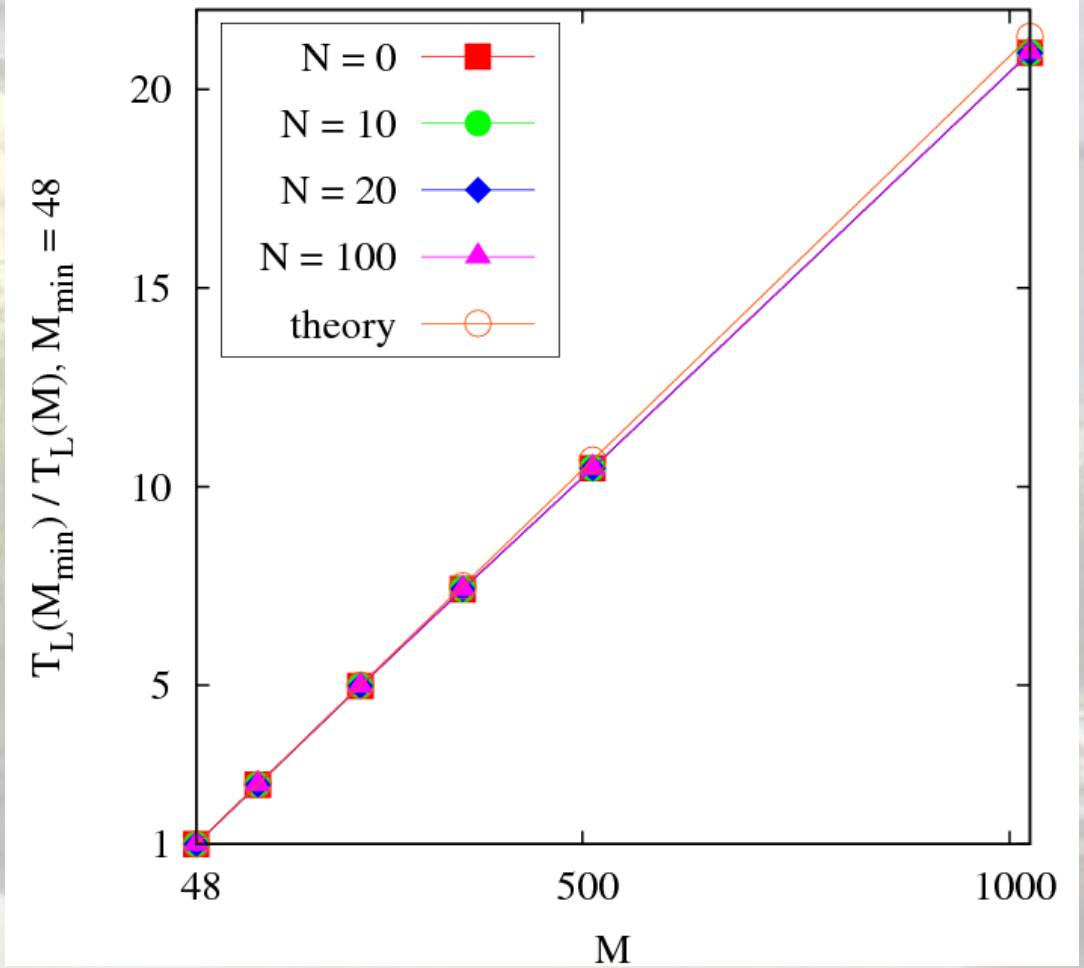
Одно главное ядро-сборщик



Добавлен уровень промежуточных ядер-сборщиков.

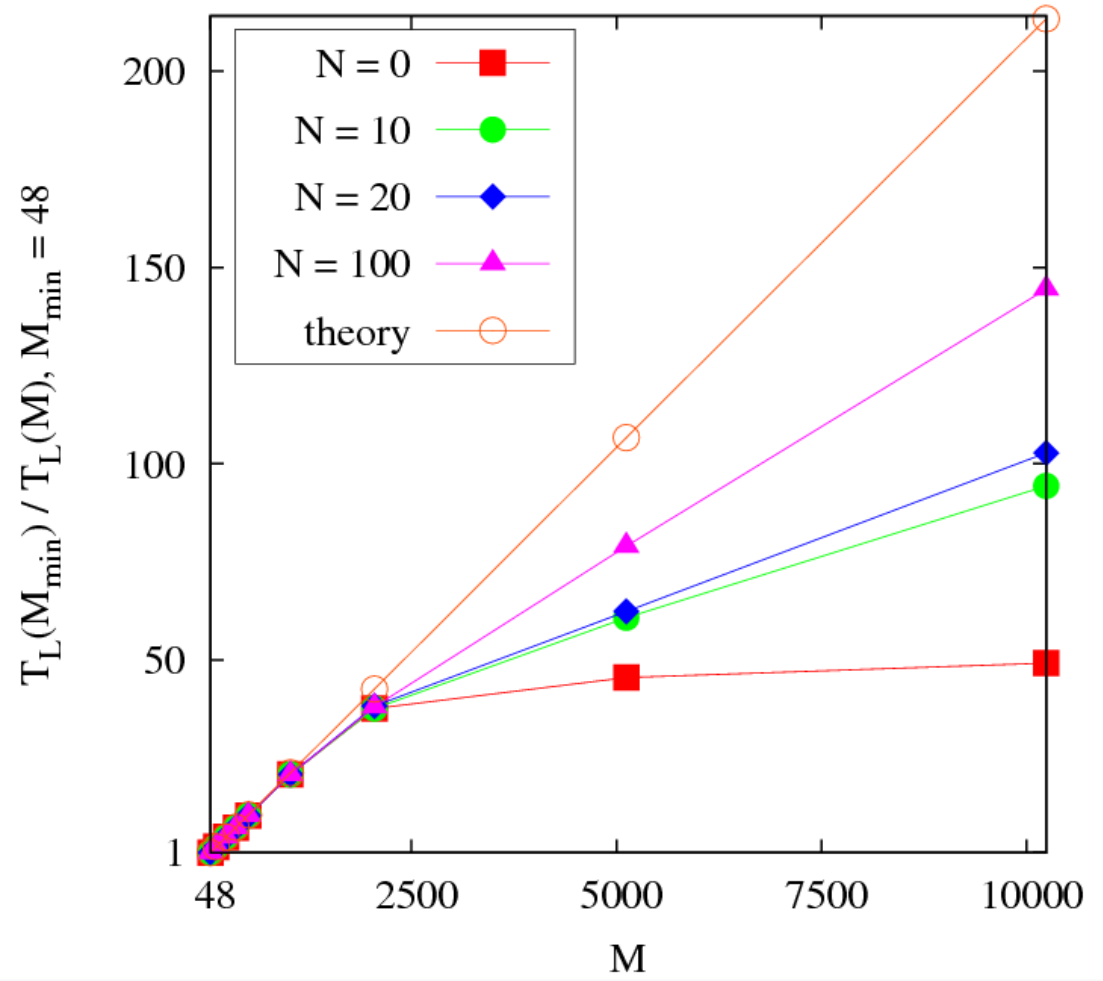
Организация передачи данных с использованием ядер-»сборщиков»

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ - 1



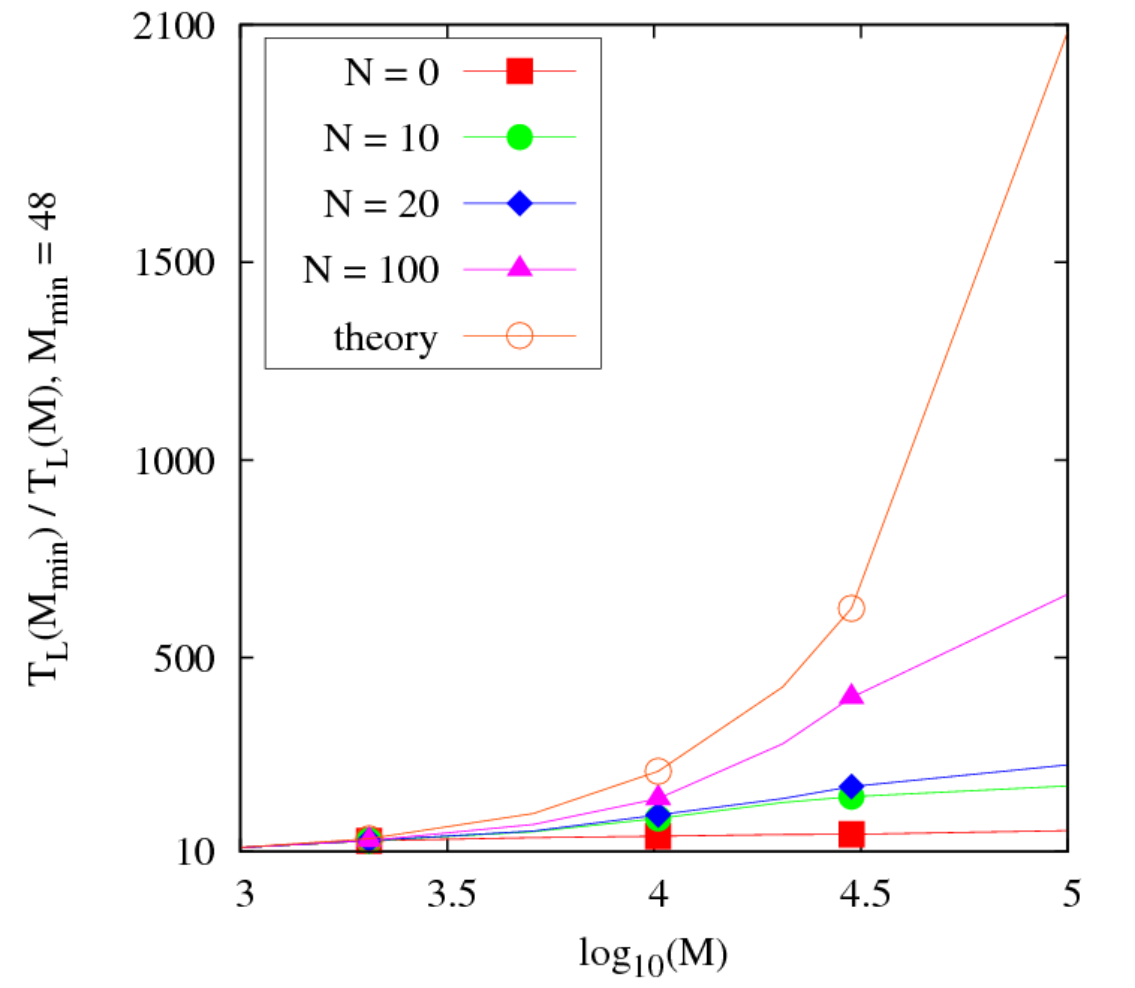
Сравнение ускорения до *M*=1000. Результаты ускорения для модели совпадают с ускорением при расчётах с использованием *PARMONC.*

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ - 2



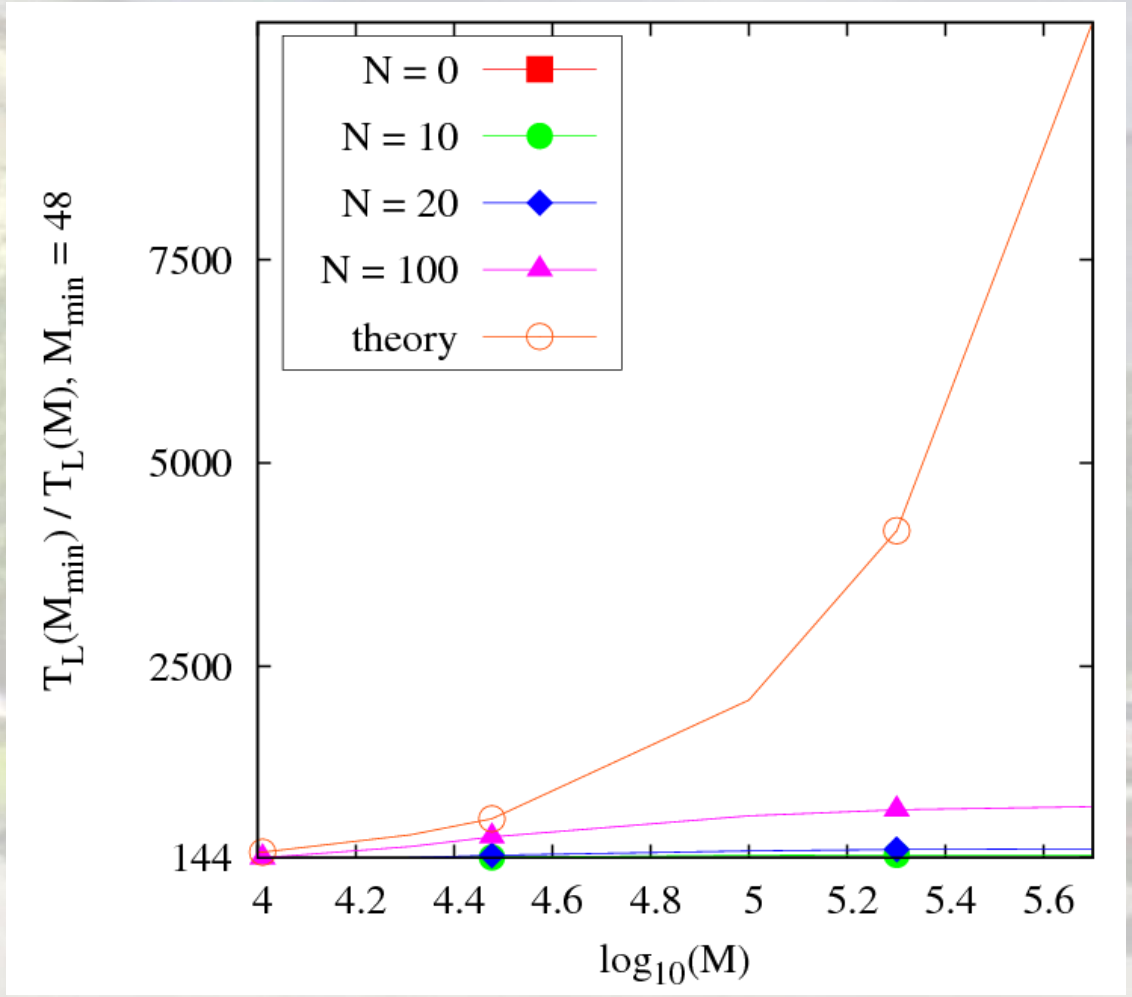
Сравнение ускорения до *M*=10 000

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ - 3



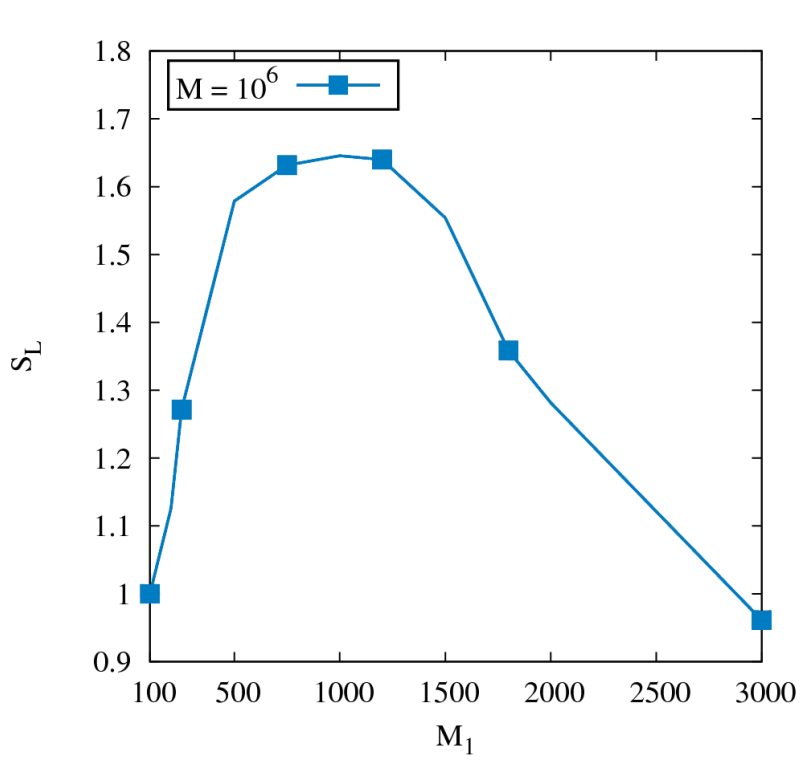
Сравнение ускорения до *M*=100 000. (горизонтальная ось – в логарифмическом масштабе).

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ - 4



Сравнение ускорения до *M*=500 000. (горизонтальная ось – в логарифмическом масштабе).

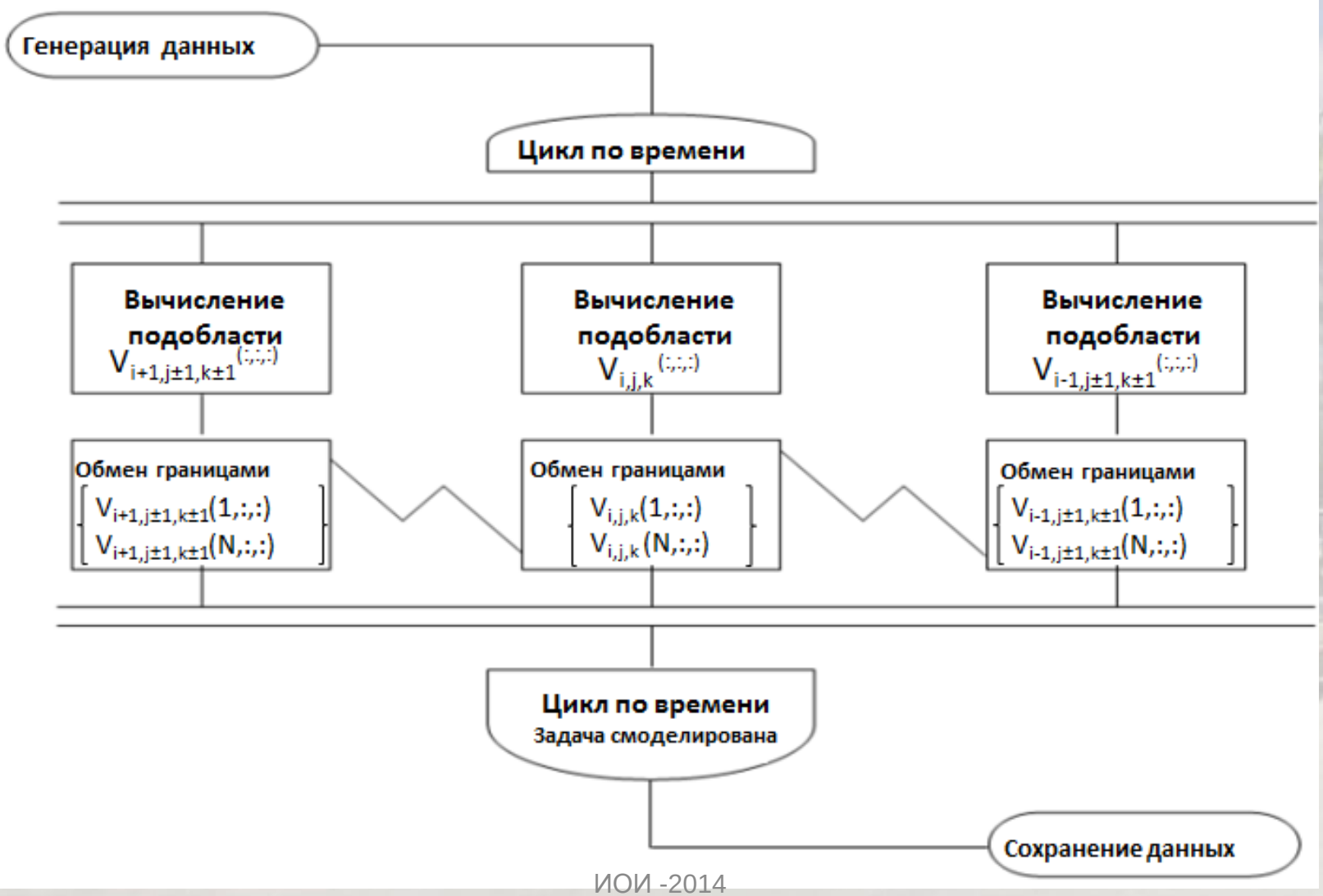
Зависимость относительного ускорения S_L от числа ядер-сборщиков M1 при общем числе моделируемых ядер M=10^6 .



Максимальная величина относительного ускорения достигается при M1, приближенно равном 1000. При меньшем значении M1 ядра-сборщики перегружены обработкой данных, поступающих от ядер-вычислителей, а при большем числе – перегружено главное ядро-сборщик, занятое обработкой поступающих данных от ядер-сборщиков.

СХЕМА ПАРАЛЛЕЛЬНОЙ РЕАЛИЗАЦИИ ВЫЧИСЛЕНИЙ В СЕТОЧНЫХ МЕТОДАХ И МЕТОДЕ ЧАСТИЦ

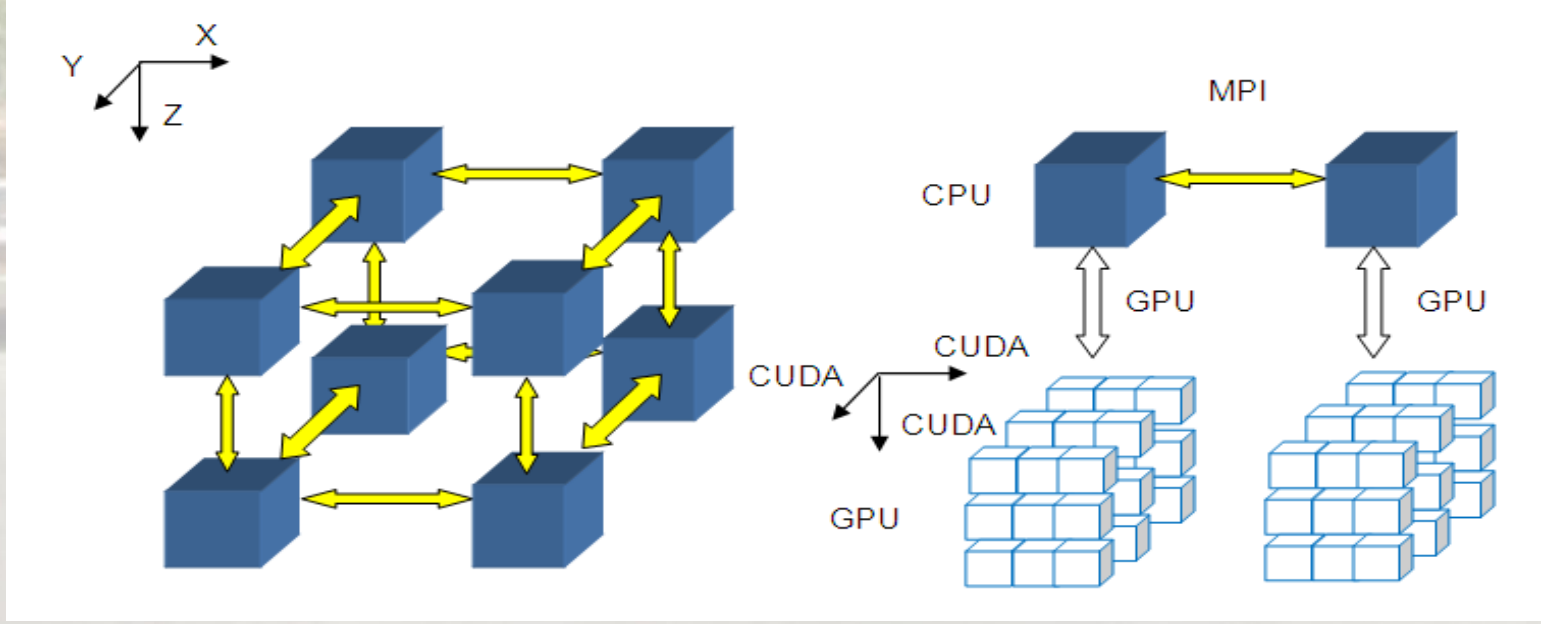
(Подход к решению задач механики сплошной среды)



ОРГАНИЗАЦИЯ ПАРАЛЛЕЛЬНОГО АЛГОРИТМА И ПРОГРАММЫ (гибридный кластер)

3D область моделирования разделяется на трехмерные подобласти по направлениям координатных осей; каждая из подобластей рассчитывается независимо на выделенном GPU, а обмены данными, между соседними GPU проводятся посредством CPU с использованием MPI. При этом вычисления для слоя производятся посредством CUDA в 3D.

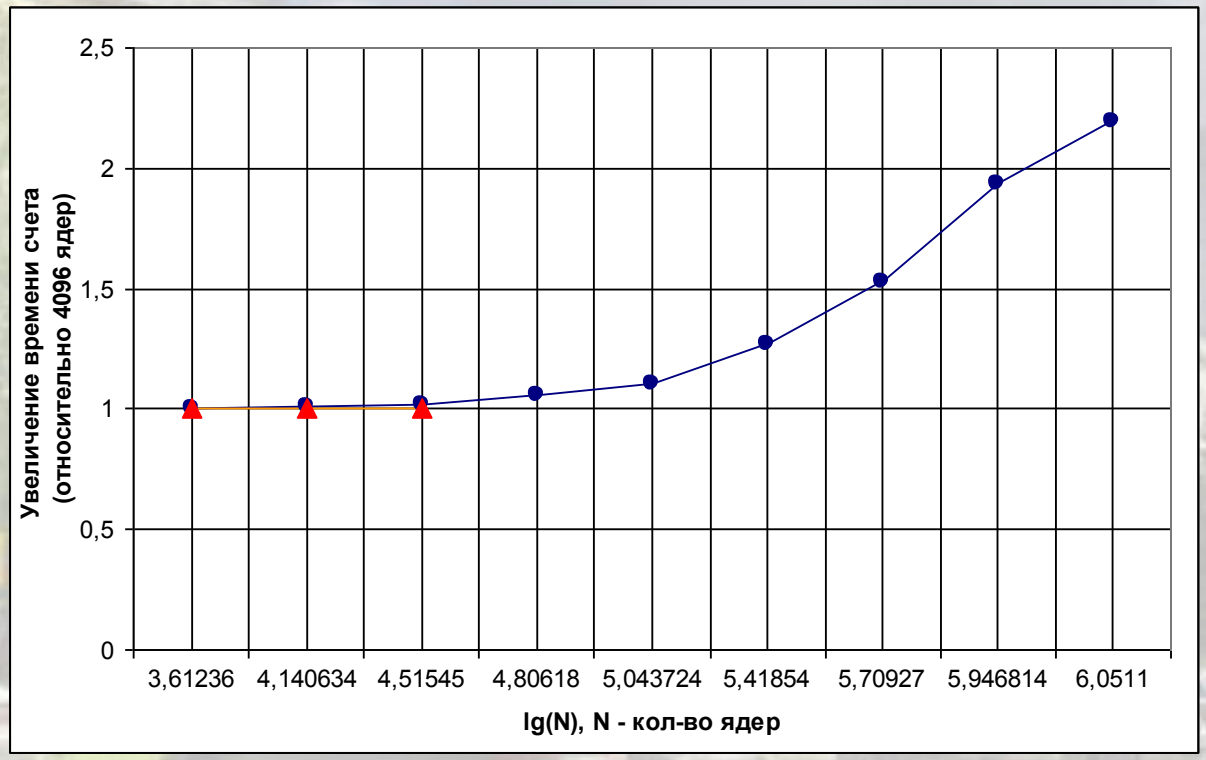
Схема декомпозиции расчетной области



ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СЕТОЧНОГО МЕТОДА С ИСПОЛЬЗОВАНИЕМ СИСТЕМЫ AGNES

- Область исследования делится вдоль осей на 3D подобласти, и полученные области загружаются на вычислители. Таким образом, получается, что у каждого вычислителя есть пересечение по данным максимум с 2-ми вычислителями по каждой из осей.
- Для имитации сеточных методов реализован класс функциональных агентов Grid — узел-вычислитель, имитирующий расчет сеточных методов на одном вычислителе.
- Масштабируемость – увеличивая количество вычислителей, увеличиваем размерность задачи, загружая на новые вычислители новые области такого же размера как и на предыдущих вычислителях. Ожидаем сохранение общего времени счета, при фиксированном количестве шагов алгоритма на одном вычислителе.

РЕЗУЛЬТАТЫ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ (гибридный кластер)



- Реальный расчет показан на начальном участке кривой (до 32768 ядер).
- Модельный расчет (до 1 124 864 ядер).
- На 500 000 ядер время выполнения алгоритма увеличилось в 1,5 раза, а для 1 000 000 ядер почти в 2,1 раза. Особенность данного алгоритма – увеличение числа обменов каждого узла с соседями на каждой итерации.

ПОРЯДОК ДЕЙСТВИЙ ПО ИМИТАЦИОННОМУ МОДЕЛИРОВАНИЮ С ПРИМЕНЕНИЕМ СИСТЕМЫ AGNES

1. Составить схему выполнения параллельной программы.
2. Прогнать параллельную программу на относительно небольшом количестве ядер.
3. Ввести в имитационную модель задержки, отображающие время счета на вычислительных ядрах и системные задержки, полученные из реальных расчетов.
4. Протестировать правильность расчета на имитационной модели путем сравнения на начальном участке реального и модельного расчетов.
5. Исследовать поведение алгоритма на модели при большом количестве ядер (от сотен тысяч до миллионов вычислительных ядер).
6. Провести, при необходимости, коррекцию вычислительной схемы, реализующей данный алгоритм.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

1. Разработана агентно-ориентированная система имитационного моделирования AGNES, позволяющая исследовать поведение вычислительного алгоритма при крупномасштабных вычислениях на заданной архитектуре суперЭВМ.
2. Проведена апробация разработанной системы при исследовании масштабируемости алгоритмов различных классов (алгоритмы статистического моделирования, сеточных методов, метода частиц), масштабируемых на вычислительные ресурсы MPP и гибридного кластеров.
3. Создан экспериментальный образец распределённой масштабируемой системы имитационного моделирования для крупномасштабных вычислений (СуперЭВМ – прикладная задача).

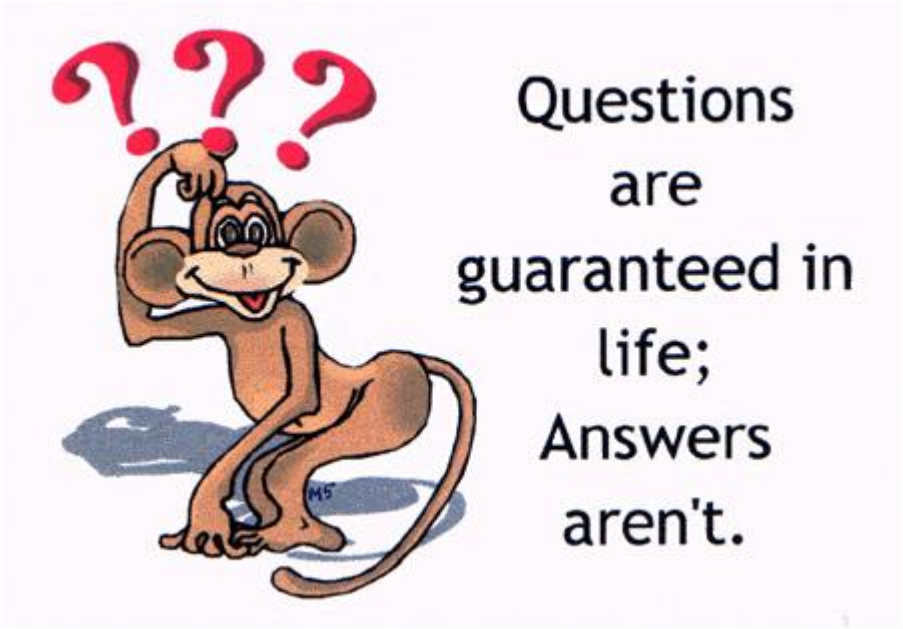
НАПРАВЛЕНИЕ ДАЛЬНЕЙШИХ ИССЛЕДОВАНИЙ

1. Продолжить экспериментальную эксплуатацию агентно-ориентированной системы имитационного моделирования AGNES.
2. Разработка средств представления перспективных архитектур в имитационных моделях суперЭВМ.
3. Прогонка на различных имитационных моделях экзафлопсных суперЭВМ разработанных масштабируемых параллельных алгоритмов: статистического моделирования, дискретной оптимизации на графах и сетях, численного моделирования 3D сейсмических полей, решения задач химической кинетики, астрофизики.
4. На основе анализа полученных результатов выбор оптимальной архитектуры экзафлопсной ЭВМ для решения различных классов задач.
5. Исследование других подходов к моделированию будущих экзафлопсных компьютеров различных архитектур.

ПУБЛИКАЦИИ

1. Б.М. Глинский, М.А. Марченко, Б. Г. Михайленко, А.С. Родионов, И.Г. Черных, Д.А. Караваев, Д.И. Подкорытов, Д. В. Винс. Отображения параллельных алгоритмов для суперкомпьютеров экзафлопсной производительности на основе имитационного моделирования // Информационные технологии и вычислительные системы, 2013, № 4, с. 3-14.
2. D. Podkorytov, A.S. Rodionov, O. Sokolova, A. Yurgenson, Using Agent-Oriented Simulation System AGNES for Evaluation of Sensor Networks // MACOM / Ed. by A. V. Vinel, B. Bellalta, C. Sacchi et al. Vol. 6235 of Lecture Notes in Computer Science. Springer, 2010. P. 247-250.
3. Б.М. Глинский, А.С. Родионов, М.А. Марченко, Д.И. Подкорытов, Д.В. Винс. Агентно-ориентированный подход к имитационному моделированию суперЭВМ экзафлопсной производительности в приложении к распределенному статисти-ческому моделированию // Вестник ЮУрГУ, 2012. № 18(277), Вып.12., с. 94-99.
4. Д.И. Подкорытов. Агентно-ориентированная среда моделирования сетевых систем AGNES // Ползуновский вестник, 2012. № 2/1, с. 93-106.
5. D. Podkorytov, A.S. Rodionov, H. Choo. Agent-based simulation system AGNES for networks modeling: review and researching // ICUIMC 2012 / Ed. by S.-H. Lee, L. Hanzo, R. Ismail et al. ACM, 2012. Paper 115, 4 pages.
6. B. Glinsky, A. Rodionov, M. Marchenko, D. Podkorytov, D. Weins. Scaling the Distributed Stochastic Simulation to Exaflop Supercomputers // Proceedings of 2012 IEEE 14th International Conference on High Performance Computing and Communications , p. 1131-1136.

Спасибо за внимание.
Ваши вопросы?



ЦКП СИБИРСКИЙ СУПЕРКОМПЬЮТЕРНЫЙ ЦЕНТР ИВМиМГ СО РАН

www2.sscc.ru

СОСТАВ ОСНОВНЫХ ТЕХНИЧЕСКИХ И ПРОГРАММНЫХ СРЕДСТВ ЦКП ССКЦ

КЛАСТЕР НКС-30Т+GPU
гибридной архитектуры

576 процессоров (2688 ядер) Intel Xeon E5450/E5540/X5670;

80 процессоров CPU (X5670) (480 ядер);

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

Общая пиковая производительность 2-х кластеров **115 Тфлопс**

СЕРВЕР
с общей памятью
(hp DL980 G7)

8 процессоров (80 ядер) Intel E7-4870; 768 Гфлопс
ОП - 1 Тбайт ;

СИСТЕМЫ ХРАНЕНИЯ ДАННЫХ (СХД)

Кластерная файловая система IBRIX для НКС-30Т
4 сервера, 32 Тбайта
Планируется расширение до 64 Тбайт

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Общесистемное:
RedHat 5.4 - Операционная система,
PBSPro 11.1 - Очередь заданий.

Средства разработки:
Intel Parallel Studio XE 2013
Intel Cluster Studio XE 2013

MPI 4.1 Intel, **Intel TraceAnalyzer/Collector, Intel C++, Intel Fortran**

NVIDIA CUDA 5,5
Portland Group PGI Accelerator 14.4.

Пакеты прикладных программ:
ANSYS CFD 14.5,
Gaussian 09

Пакеты, разработанные в ИВМиМГ СО РАН:
PARMONC,
AGNES.

СИБИРСКИЙ СУПЕРКОМПЬЮТЕРНЫЙ ЦЕНТР

Н.В. Кучин, Б.М. Глинский, Б.Г. Михайленко

ВЫЧИСЛИТЕЛЬНЫЕ РЕСУРСЫ

Кластер НКС-160
(hp rx1620)

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

!!! NEW
Кластер гибридной архитектуры НКС-30Т+GPU

576 (2688 ядер) процессоров Intel Xeon E5450/E5540/X5670

80 процессор. CPU (X5670) – 480 ядер;

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

Общая пиковая производ. 115 Тфлопс

GigabitEthernet
InfiniBand

Сеть ИВМиМГ — GE

Сеть Internet ННЦ — GE

Системное
Общематематическое
Прикладное (ППП)

СИСТЕМЫ ХРАНЕНИЯ ДАННЫХ (СХД)

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

Параллельная файловая система IBRIX для НКС-30Т 32 Тбайта

СХД для НКС-160
3,2 Тбайт

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

!!! NEW
Сервер с общей памятью
(hp DL980 G7)

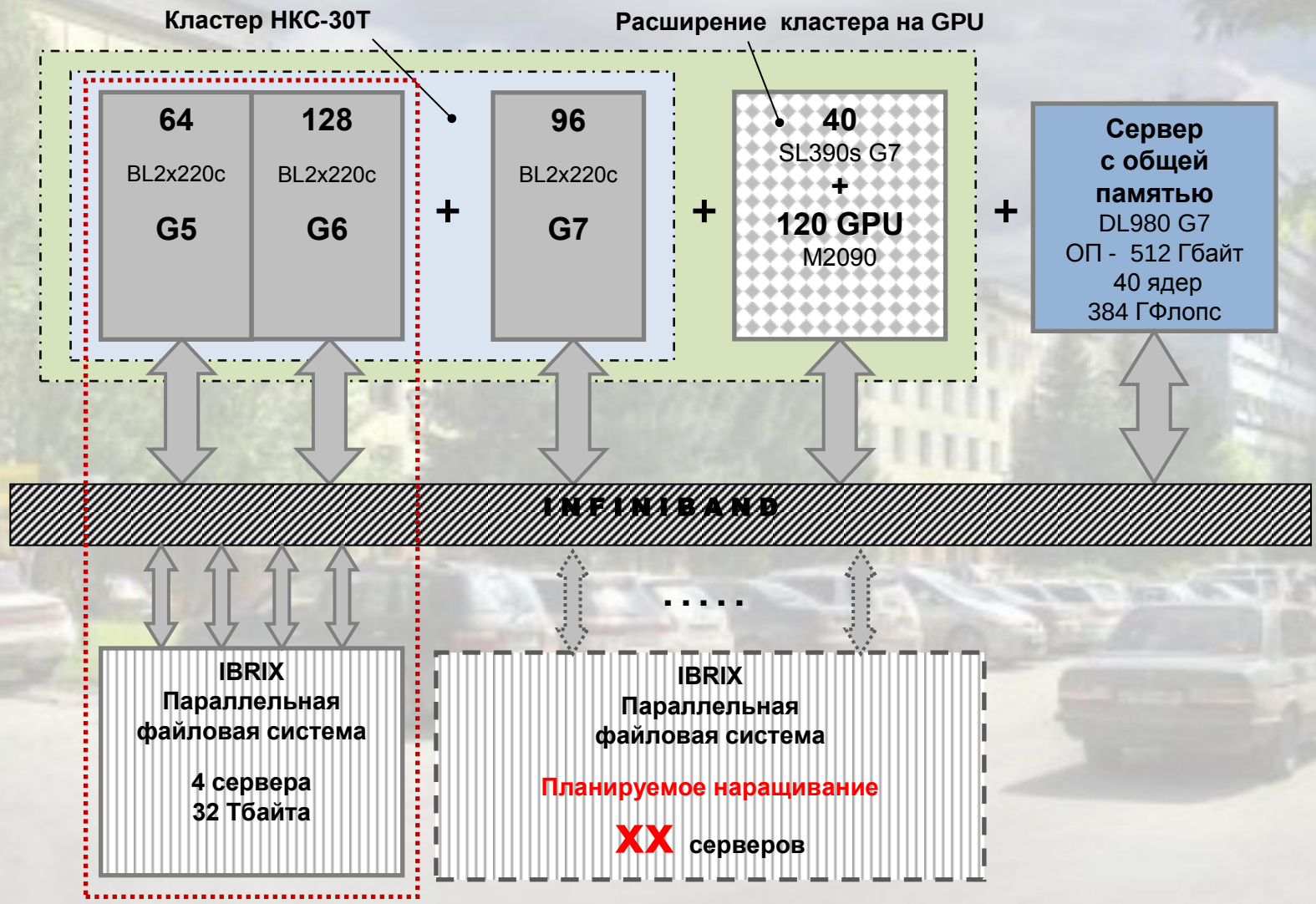
4 процессора (40 ядер)
Intel E7-4870;
ОП - 1024 Гбайт ; 768 Гфлопс.
Max: 8 процессоров (80 ядер),
2048 Гбайт, 768 Гфлопс.

Сервер с общей памятью
(hp DL580 G5)

4 процессора
(16 ядер)
Intel Xeon Quad
Core X7350, 2.93 Ггц;
ОП - 256 Гбайт;
187,5 Гфлопс.

СИБИРСКИЙ СУПЕРКОМПЬЮТЕРНЫЙ ЦЕНТР

ГИБРИДНЫЙ КЛАСТЕР (НКС-30Т+GPU)



СИБИРСКИЙ СУПЕРКОМПЬЮТЕРНЫЙ ЦЕНТР

Библиотека программ PARMONC для распределенных вычислений по методу Монте-Карло

Разработчик: М.А. Марченко, ИВМиМГ СО РАН, marchenko@sscc.ru

Библиотека PARMONC (сокращение от **PARallel MONte Carlo**) предназначена для распараллеливания трудоемких приложений метода Монте-Карло. При распараллеливании используется «естественная» крупноблочная фрагментированность алгоритмов метода Монте-Карло.

Для получения независимых параллельных потоков базовых случайных чисел используется тщательно протестированный, быстрый и надежный длиннопериодный генератор, разработанный в лаборатории методов Монте-Карло ИВМиМГ. Библиотечные подпрограммы вызываются из разных языков программирования без явного использования MPI. **Число используемых в PARMONC вычислительных ядер практически не ограничено и зависит только от используемой ЭВМ.**

Область применения: «большие» задачи статистического моделирования в естественных и гуманитарных науках (физика, химия, биология, медицина, экономика и финансы, социология и др.)

Пользователи библиотеки: отдел статистического моделирования в физике ИВМиМГ, Институт атомной энергетики, Обнинск; **заинтересованы**: ИСЭ СО РАН, Томск, Омский филиал ИМ СО РАН.

СИБИРСКИЙ СУПЕРКОМПЬЮТЕРНЫЙ ЦЕНТР

Система мультиагентного имитационного моделирования AGNES для моделирования больших систем с дискретными событиями

Разработчик: Д.И.Подкорытов, ИВМиМГ СО РАН, d.podkorytov@mail.ru

Система моделирования AGNES (**AGent NEtwork Simulator**) предназначена для имитационного моделирования больших систем, например для исследования свойств масштабируемости вычислительных алгоритмов при расчете их на суперкомпьютерах. AGNES функционирует на основе мультиагентной платформы **JADE** (Java Agent Development Framework), написанной на **JAVA**, поэтому AGNES может работать на любой системе, где установлена JVM (Java Virtual Machine). Система AGNES подходит для распределенного выполнения как на локальных сетях, так и на кластерах суперЭВМ.

В настоящее время AGNES использует только мощности CPU из выделенных на моделирование ресурсов, использование GPU пока не реализовано. **Число используемых в AGNES ядер CPU практически не ограничено и зависит только от используемой ЭВМ.**

Область применения: Моделирование сложных неоднородных систем с дискретными событиями, предпочтительно систем состоящих из большого числа схожих компонентов, легко поддающихся анализу и описанию.

Пользователи библиотеки: ИВМиМГ СО РАН, СибГУТИ.