

GPGPU

Цыпляев А.С., 6057/12

Введение

- GPGPU – *General-purpose graphics processing units* – GPU общего назначения
- Вычисления на GPU или GPGPU заключаются в использовании GPU (графического процессора) для универсальных вычислений в области науки и проектирования.

Введение

- **Графический процессор** – специализированный процессор в составе видеоадаптера (GPU + память) для обработки и отображения компьютерной графики
- Скорость обеспечивается возможностью распараллелить графические вычисления – обработку вершин, пикселей (фрагментов)...
- Графический процессор – мощный инструмент для параллельных вычислений с плавающей точкой

История

- История графических чипов началась с графических конвейеров с фиксированной функциональностью
- Они использовались, чтобы ускорить доступ к текстурам, рендеринг полигонов
- Далее - ускорение работы геометрией: трансформация вершин
- Постепенно их программируемость все возрастала
- Компания nVidia представила первый программируемый GPU
- Программы для GPU – вершинные и пиксельные шейдеры
- Большинство операций – с векторами и матрицами. Это привлекло внимание учёных

История

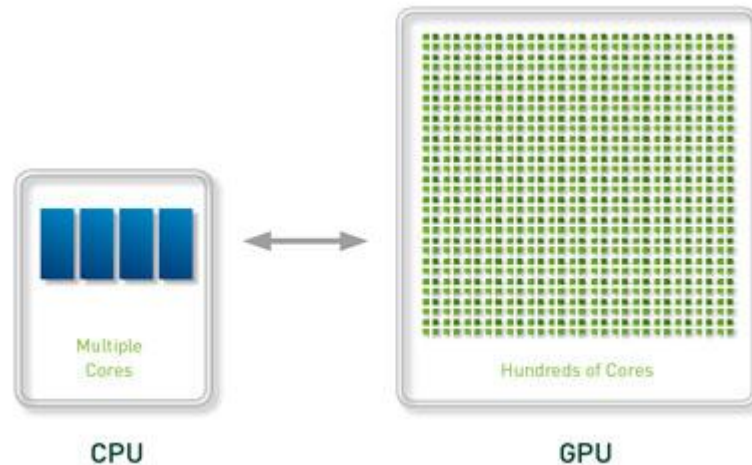
- 1999-2000 годах специалисты в компьютерной области и научные работники в таких сферах, как получение медицинских изображений и электромагнетизм, перешли на GPU для вычислительных приложений общего назначения
- Для GPGPU не существовало отдельного API и языка программирования. Разработчикам приходилось делать научные приложения, используя графический API и шейдеры. Это ограничивало доступность огромной производительности GPU для науки

История

- Первой реализацией функций GPGPU стала архитектура CUDA (*Compute Unified Device Architecture*) компании nVidia
- Реализации функций GPGPU на данный момент
 - *AMD FireStream*
 - *nVidia CUDA*
 - *DirectCompute от Microsoft*
 - *OpenCL – открытый отраслевой стандарт в области GPGPU*

Устройство

- GPU вычисления представлены совместным использованием CPU и GPU в гетерогенной модели вычислений
- Стандартная часть приложения выполняется на CPU, а более требовательная к вычислениям часть обрабатывается с GPU



Stream Processing

- Stream – поток элементов , требующих одинаковой обработки
- Kernel – функция обработки элементов
- GPU обрабатывает записи независимо друг от друга => нет разделяемых и статических данных
Для каждого элемента можно только прочесть его из входного потока, провести над ним вычисления, записать в выходной поток
- Возможно несколько входных и выходных потоков
- Пример: в CG – вершины, пиксели – элементы, шейдеры - Kernels

Концепции

- Вычислительные инструменты GPU:
 - Программируемый процессор (Programmable processor)
 - вершинные и фрагментные конвейеры выполняют Kernel функции
 - Растеризатор (Rasterizer)
 - Создаёт фрагменты и интерполирует константы, данные в вершинах, такие как цвет, текстурные координаты
 - Тектурный модуль (Texture Unit)
 - Read-only доступ к памяти
 - Фрэймбуфер (Framebuffer)
 - Write-only доступ к памяти

Концепции

- Текстура как поток (stream)
 - Представление данных в виде 2D решётки
 - Таким образом данные можно упаковать в текстуру
 - Некоторые операции над данными будут выполняться на GPU автоматически
- Kernels
 - Представим, что мы обходим данные в цикле данные представленные в решётке
 - Тогда Kernel – тело цикла
- Flow control
 - If-else блоки в Kernel функции (появились не так давно)

Техники

- Map
 - Простое применение kernel функции к потоку
 - Размер входа = размеру выхода
- Reduce
 - Уменьшение количества элементов в выходном потоке
 - Пример: уменьшение изображения
- Stream Filtering
 - Отбрасывание элементов потока по критерию
- Scatter
 - Изменение положения элемента в выходной решётке относительно входной
 - Пример: вершины

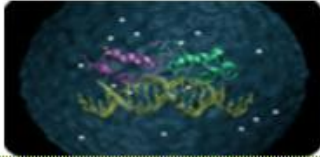
Техники

- Gather
 - Считывание нескольких элементов текстуры для выдачи одного
 - Пример: усреднение, размытие, blur
- Sort
 - Сортировка при помощи сортирующих цепей
- Search
 - Запуск множества параллельных функций поиска

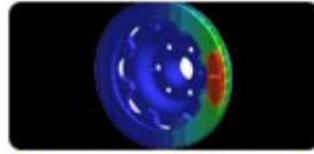
GPGPU в кластерах

- Кластер, в котором в узлах установлены GPU
- Классификация
 - Гомогенные
 - Все видеоадаптеры во всех узлах одного производителя
 - Гетерогенные
- nVidia имеет список Tesla Preferred Partners – список компаний способных собрать и установить кластер GPU nVidia
 - Cray, Dell, Hewlett-Packard, IBM, Silicon Graphics являются TPP
- AMD на данный момент не имеет подобного списка
- Примеры кластеров использующих GPGPU приводились в докладе TOP 500

Решаемые задачи



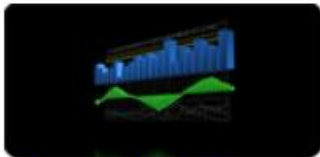
Bio-Informatics and Life Sciences



Computational Structural Mechanics



Computational Electromagnetics and Electrodynamics



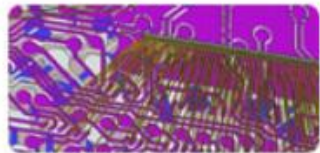
Computational Finance



Computational Fluid Dynamics



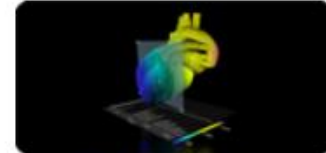
Data Mining, Analytics, and Databases



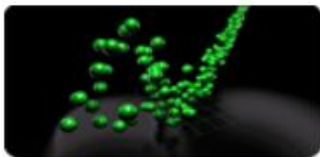
Electronic Design Automation



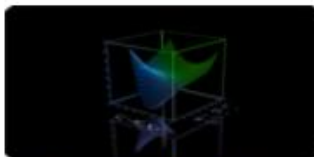
Imaging and Computer Vision



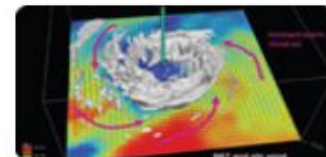
Medical Imaging



Molecular Dynamics



Numerical Analytics



Weather, Atmospheric, Ocean Modeling, and Space Sciences

Решаемые задачи

БИОНАУКИ



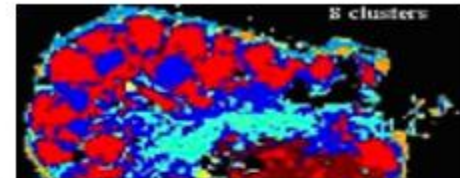
УНИВЕРСИТЕТ ШТАТА ИЛЛИНОЙС

Ускоренное моделирование молекул позволяет получать результаты исследований H1N1 еще быстрее.



ГАРВАРД

Более быстрое нахождение скрытых проблем сердца.



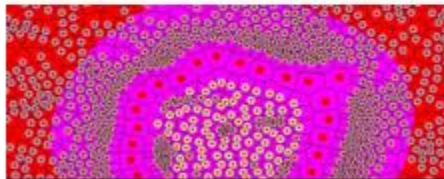
СЕВЕРО-ВОСТОЧНЫЙ УНИВЕРСИТЕТ

Достижения в работе с патологиями благодаря графическим процессорам.



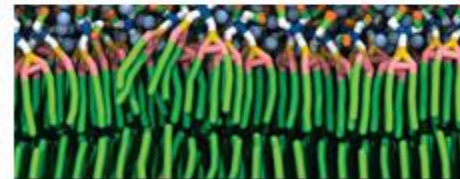
ИНСТИТУТ МАКСА ПЛАНКА

Институт Макса Планка.



GEORGIA TECH

Нахождение лекарств при помощи ускоренной программы MATLAB.

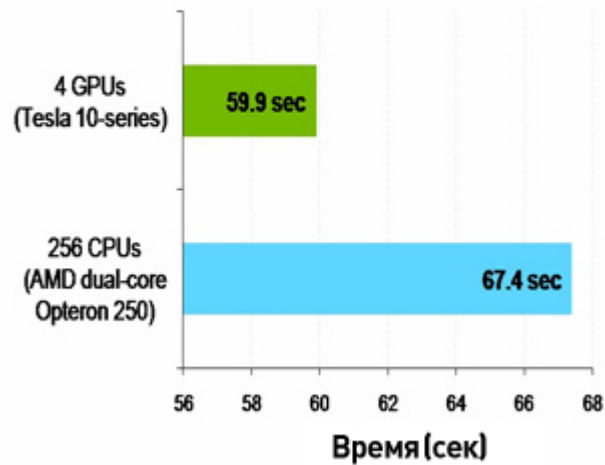


NCSA

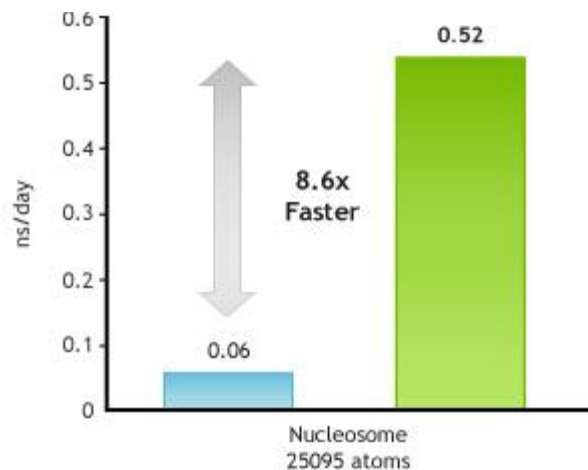
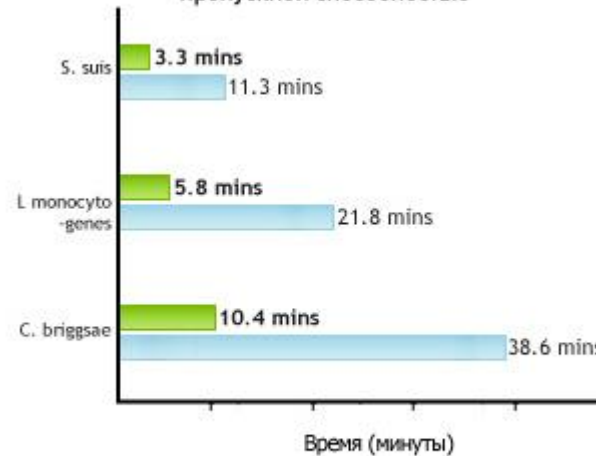
NCSA позволяет ускорять моделирование молекул.

Ускорение

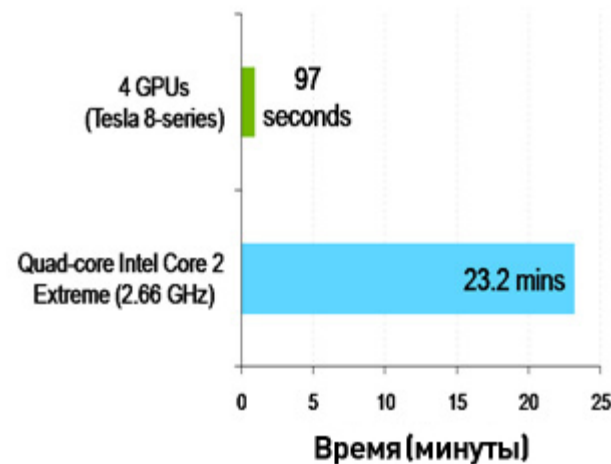
Компьютерная томография (СТ)



Построение последовательности ДНК с высокой пропускной способностью

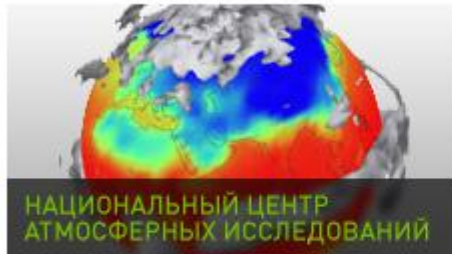


Время реконструкции ускоренного МРТ



Решаемые задачи

ВЫЧИСЛИТЕЛЬНАЯ ДИНАМИКА ЖИДКОСТЕЙ

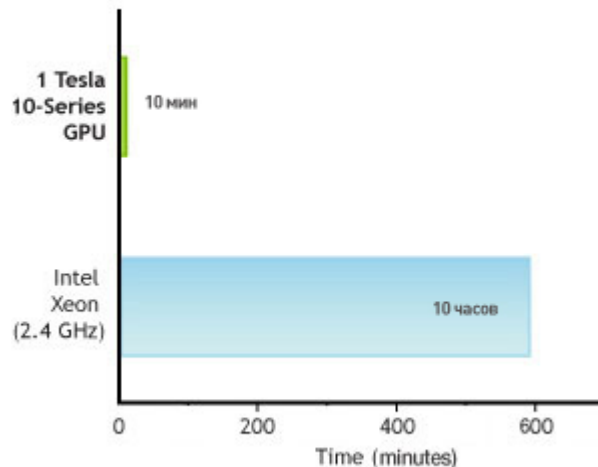


Национальный Центр Атмосферных исследований.

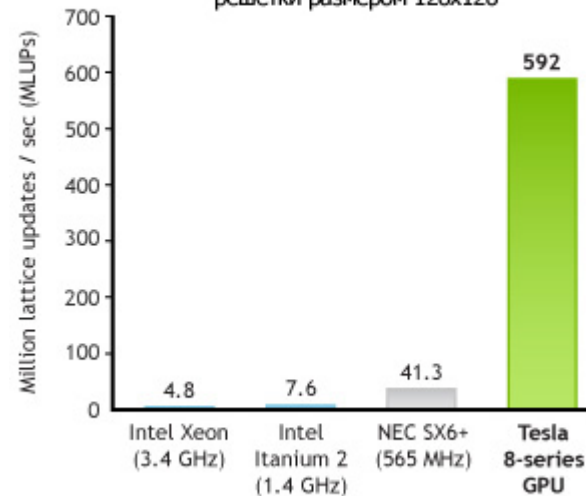


Ускорение вычислений методом решетки Больцмана при помощи MATLAB.

Симуляция цунами
3000km x 3000km (500m mesh)

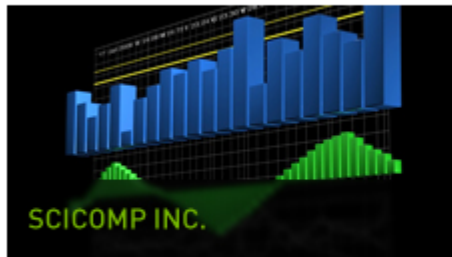


Метод решетки Больцмана для
решетки размером 128x128



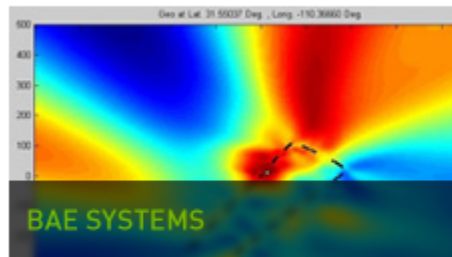
Решаемые задачи

ВЫЧИСЛИТЕЛЬНЫЕ ФИНАНСЫ



SciComp для моделирования производных.

ГЕОГРАФИЧЕСКИЕ ИНФОРМАЦИОННЫЕ УСЛУГИ



Ускорение геологоразведки при помощи MATLAB.

Решаемые задачи

ВЫВОД ИЗОБРАЖЕНИЙ

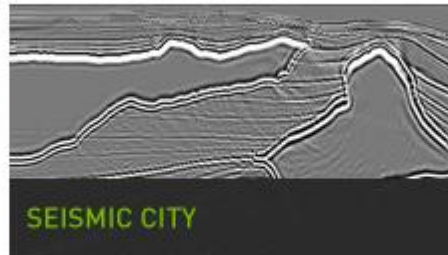


Возрастающее значение GPU в
Вооруженных Силах.

НЕФТЬ И ГАЗ



Градиент и CUDA. Эффективная
геологоразведка.

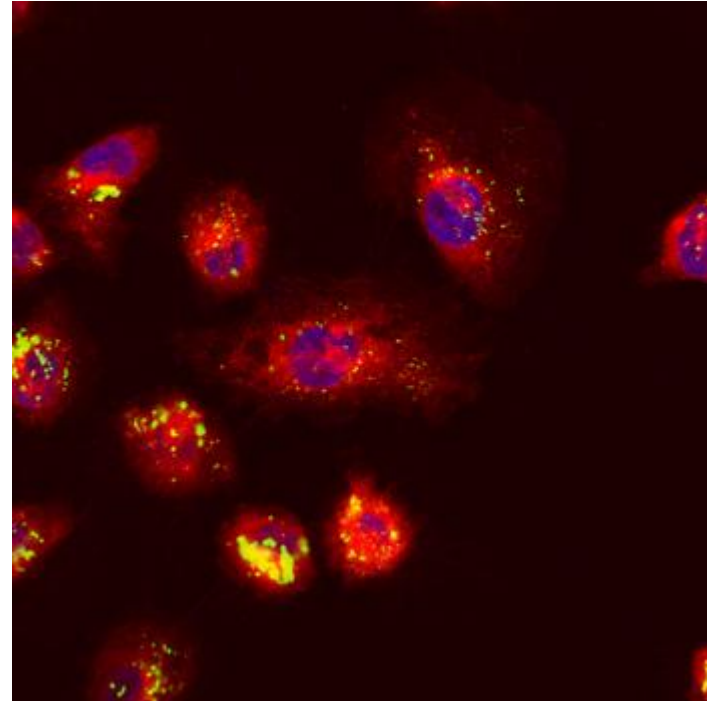
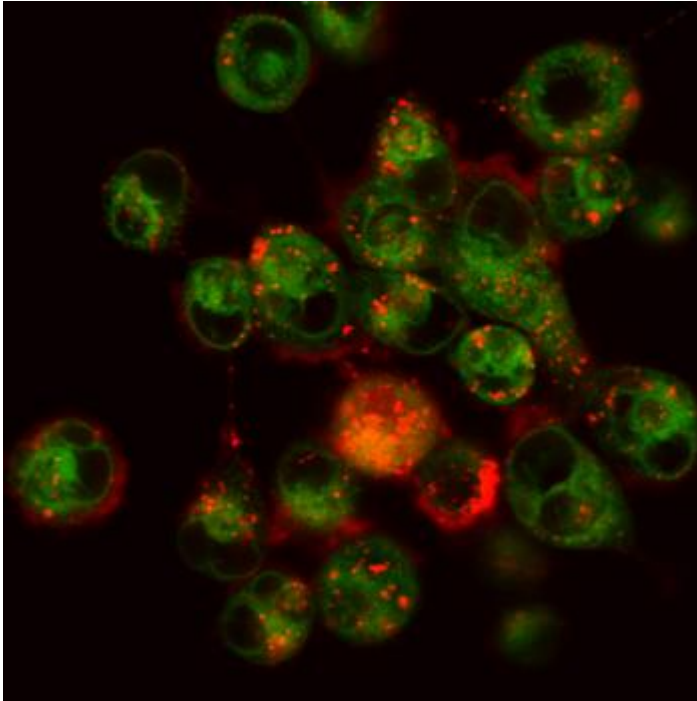


Технология для работы с
сейсмическими изображениями в
3D.

Диффузная флуоресцентная томография

- Метод оптической диагностики опухолей
- В организм вводятся специальные флуоресцентные маркеры, которые прикрепляются к злокачественным клеткам
- Подсветка тканей и ее регистрация позволяет определить место расположения опухоли
- Сложность метода ДФТ заключается в том, что свет, проходящий в биологических тканях, подвергается сильному рассеянию
- Специалисты **Института прикладной физики РАН** разработали специальные алгоритмы реконструкции трехмерного распределения флуорофоров в тканях, которые позволяют точно определить место расположения и геометрию опухоли
- **Алексей Катичев, младший научный сотрудник Института: “Перенос вычислений на архитектуру графических процессоров NVIDIA CUDA дал более чем стократный прирост производительности. Среднее время получения результата уменьшилось с двух с половиной часов до 1,5 минут”**

Диффузная флуоресцентная томография



Ссылки

- http://www.nvidia.ru/page/gpu_computing.html
- <http://www.nvidia.ru/object/tesla-case-studies-ru.html>
- http://www.nvidia.com/object/tesla_wtb.htm
- [http://www.3dnews.ru/video/what is faster gpu or cpu](http://www.3dnews.ru/video/what_is_faster_gpu_or_cpu)
- Wikipedia