

УДК: 004.942

Гибридные вычислительные системы на основе GPU для задач биоинформатики

А. Р. Джораев

NVIDIA,
119002, Москва, Арбат, д.10, 7 эт.

E-mail: adzhoraev@nvidia.com

Получено 31 мая 2010 г.

Статья посвящена преимуществам применения гибридных вычислительных систем на основе графических процессоров NVIDIA для решения задач моделирования молекулярной динамики, квантовой химии, секвенирования, приведены примеры приложений.

Ключевые слова: GPU, графический процессор, молекулярная динамика, квантовая химия

GPU-accelerated hybrid systems for high-performance computing in bio-informatics

A. R. Dzhoraev

NVIDIA Russia & CIS, 7th floor, Arbat 10, Moscow, 119002, Russia

Abstract. – Modern GPUs are massively-parallel processors, offering substantial amount of computational power in energy-efficient package. We discuss the benefits of utilizing this computing power for modeling problems in bio-informatics, such as molecular dynamics, quantum chemistry and sequence analysis.

Keywords: GPU, bioinformatics, molecular dynamics, quantum chemistry, sequence analysis

Citation: *Computer Research and Modeling*, 2010, vol. 2, no. 2, pp. 163–167 (Russian).

До последнего времени ключевым компонентом систем для высокопроизводительных вычислений, включая кластеры, был центральный процессор. Однако несколько лет назад у него появился серьезный конкурент – графический процессор (GPU).

Высокая производительность GPU объясняется особенностями его архитектуры.

В отличие от центрального процессора, который состоит из нескольких ядер, графический процессор изначально создавался как многоядерная структура, в которой количество компонентов измеряется сотнями. В графическом процессоре NVIDIA последнего поколения «Fermi» 512 вычислительных ядер. Также есть существенная разница и в принципах работы – архитектура CPU предполагает последовательную обработку информации, а GPU исторически предназначался для обработки компьютерной графики, поэтому рассчитан на массивно параллельные вычисления.

Каждая из этих двух архитектур имеет свои достоинства. Говорить об абсолютной замене CPU на GPU не имеет смысла – они не взаимозаменяют, а дополняют друг друга. CPU лучше работает с последовательными задачами, но при большом объеме обрабатываемой информации, с которой можно работать параллельно, очевидное преимущество имеет GPU.

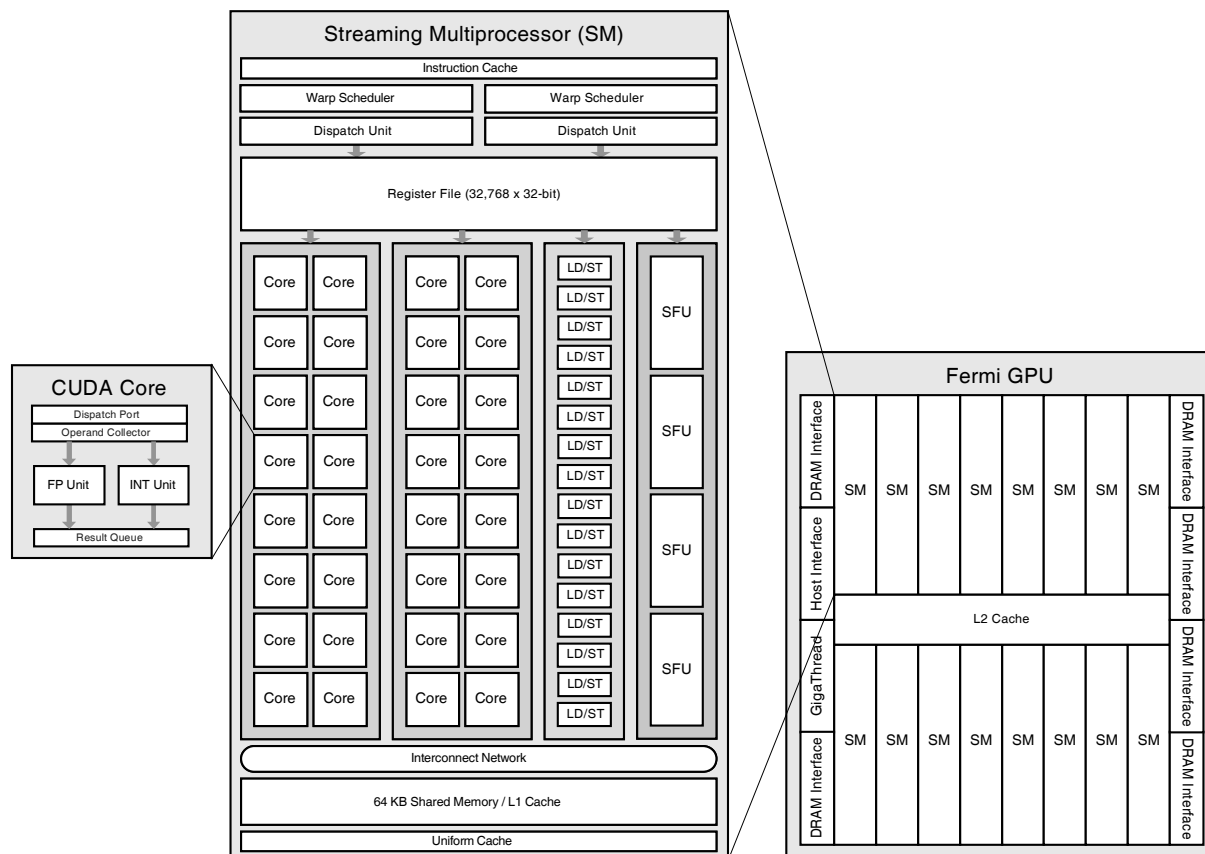


Рис. 1. Схематическое изображение GPU «Fermi» (справа) и одного SM (Streaming Multiprocessor) – модуля (в центре) и одного вычислительного ядра (слева) [NVIDIA's Next Generation CUDA™ Compute Architecture...]

Как показывают многочисленные испытания, пиковая производительность современного GPU в разы выше, чем производительность современного CPU (рис. 3). В скорости доступа к видеопамяти GPU также имеет значительное превосходство над CPU. Эффективная организация подсистемы памяти повышает общую эффективность графического процессора при работе с неграфическими задачами.

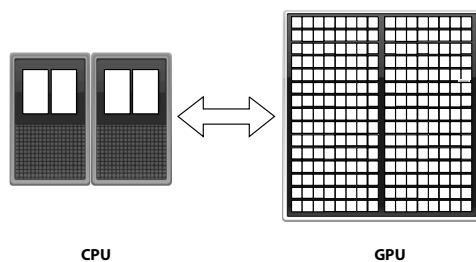


Рис. 2. В гибридной вычислительной системе совместно используются ресурсы CPU и GPU

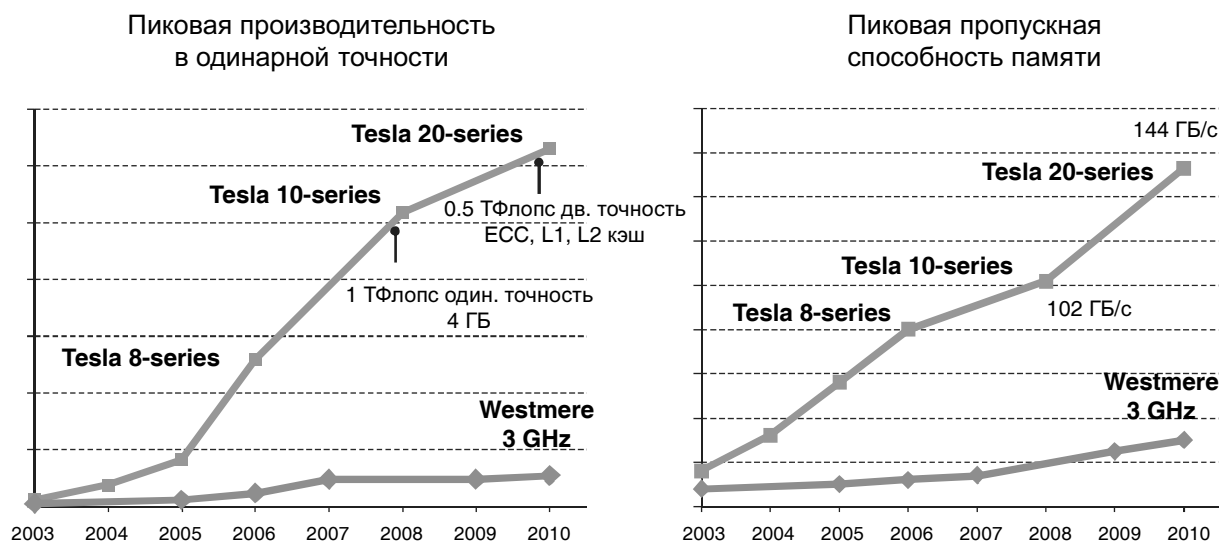


Рис. 3. Относительная производительность и скорость доступа к памяти процессоров GPU NVIDIA (■) и CPU Intel x86 (◆) в 2003–2010 годах

Высокая производительность массивно-параллельной архитектуры GPU обеспечивает гибридные системы колоссальной производительностью при решении самого широкого класса научных и прикладных вычислительных задач – от астрофизики и молекулярной динамики до гидродинамики и поиска залежей полезных ископаемых. Прирост в скорости решения таких задач достигает десятков и сотен раз.

До последнего времени основная проблема заключалась в том, как «научить» графический процессор выполнять несвойственные ему задачи. Пользователям-программистам приходилось строить гибридные системы из обычных видеокарт и программировать их, используя неудобные для этого графические API. Эту сложную задачу удалось успешно решить несколько лет назад, когда NVIDIA представила программно-аппаратную платформу CUDA, позволяющую запускать произвольный код на GPU.

CUDA предоставляет разработчику возможность создавать код, исполняющийся непосредственно на ядрах графического процессора, управлять его памятью, организовывать на нем сложные параллельные вычисления – проще говоря, использовать его как процессор общего назначения.

Сегодня программная модель CUDA поддерживает практически все наиболее популярные языки и интерфейсы программирования: C, C++, Fortran, Open CL, Direct Compute и др., предлагая все необходимые инструменты для разработки программного обеспечения (ПО).

Технология CUDA универсальна, и любой современный GPU NVIDIA может исполнять программный код общего назначения – будь то игровая карта GeForce, профессиональная Quadro или основа суперкомпьютерных систем – специальный вычислитель Tesla. Подобная масштабируемость решения – очевидное преимущество для пользователей. Код, разработанный

на видеокарте начального уровня, можно легко перенести на супервычислители Tesla, специально спроектированные для обработки больших массивов данных. Вычислители NVIDIA Tesla в разы быстрее обычных видеокарт при расчетах в двойной точности, надежнее (поддержка ECC) и оснащены большими объемами памяти.

Продукты серии «Tesla» предлагаются в двух вариантах исполнения – узел 1U для установки в стойку и PCI-E карта для установки в корпус обычного ПК (внешне напоминает видеокарту). Пиковая производительность системы последнего поколения Tesla S2050 достигает 2,06 ТФлопс в двойной точности.

Используя карты Tesla C2050, можно получить беспрецедентную производительность буквально на рабочем столе – компьютер с несколькими картами Tesla будет обладать мощностью, сравнимой с небольшим кластером, при этом система не требует создания серверной инфраструктуры, включается в обычную розетку и всегда доступна. Пиковая производительность карты Tesla C2050 достигает 515 ГФлопс в двойной точности.

Современные высокопроизводительные системы должны обеспечивать не только высокую производительность, но быть энергетически и экономически эффективными. В этом плане гибридные системы (использующие CPU и GPU одновременно) значительно опережают CPU-кластеры, демонстрируя существенно лучшие показатели производительности на Ватт потребляемой мощности. Производительность таких решений в расчете на стоимость владения также чрезвычайно высока.

Главным условием того, чтобы разработчики могли применять такие системы, является наличие ПО, которое их поддерживает. Благодаря активной работе NVIDIA по поддержке разработчиков появляется все больше и больше ПО, использующего мощь массивно-параллельной архитектуры GPU.

Табл. 1. Программные пакеты для осуществления расчетов в области молекулярной динамики, квантовой химии и биоинформатики, использующие технологию CUDA

| Приложение | Дополнительная информация |
|---|---|
| Молекулярная динамика | |
| AMBER | http://www.nvidia.com/object/amber_on_tesla.html |
| ACEMD | http://www.acellera.com/acemd/ |
| GROMACS | http://www.nvidia.com/object/gromacs_on_tesla.html |
| LAMMPS | http://www.nvidia.com/object/lammps_on_tesla.html |
| NAMD | http://www.nvidia.com/object/namd_on_tesla.html |
| Квантовая химия | |
| TeraChem | http://www.nvidia.com/object/terachem_on_tesla.html |
| Визуализация для молекулярной динамики и квантовой химии | |
| VMD | http://www.nvidia.com/object/vmd_on_tesla.html |
| Секвенирование | |
| CUDA-BLASTP | http://www.nvidia.com/object/blastp_on_tesla.html |
| CUDASW++ (SmithWaterman) | http://www.nvidia.com/object/swplusplus_on_tesla.html |
| MUMmerGPU | http://www.nvidia.com/object/mummer_on_tesla.html |
| GPU-HMMER | http://www.nvidia.com/object/hmmer_on_tesla.html |
| CUDA-MEME Motif Discovery | http://www.nvidia.com/object/meme_on_tesla.html |

Тенденция перехода на гибридные системы наблюдается во многих отраслях, в том числе и в биоинформатике, поскольку задачи молекулярной динамики, квантовой химии, генетики (и подобные) очень хорошо «ложатся» на архитектуру GPU. Специально для этой отрасли компанией NVIDIA в сотрудничестве с ведущими научными организациями мира был создан веб-

ресурс Bio Workbench [NVIDIA Tesla Bio Workbench], который является центром компетенции по применению GPU для решения задач из разных отраслей биологического знания. С этого ресурса можно скачать наиболее распространенное ПО для моделирования молекулярно-динамических процессов и квантовой химии, там есть информация по производительности ПО на разных платформах, рекомендуемые конфигурации, документация, а также есть возможность обсуждения и обмена опытом по данной тематике.

Ниже приводится список наиболее популярных приложений этой отрасли, которые перенесены на GPU. Разница в производительности этих приложений при работе на CPU и на GPU достигает десятков раз. Детальная информация по производительности доступна по ссылкам в таблице выше.

Данные приложения используются для создания новых материалов, оценки влияния различных веществ на белки, разработки медицинских препаратов, расшифровки генома. Использование вычислителей NVIDIA Tesla позволяет решать подобные задачи в десятки раз быстрее.

Дополнительная информация

NVIDIA's Next Generation CUDA™ Compute Architecture: Fermi (Whitepaper)
http://www.nvidia.com/content/PDF/fermi_white_papers/NVIDIA_Fermi_Compute_Architecture_Whitepaper.pdf

NVIDIA Tesla Bio Workbench
http://www.nvidia.ru/page/tesla_bio_workbench.html

NVIDIA CUDA
<http://www.nvidia.ru/cuda>

CUDA Programming Guide
<http://developer.nvidia.com>

Боресков А. В., Харламов А. А. Основы работы с технологией CUDA. – ДМК Пресс, 2010.