

## ПРИМЕНЕНИЕ GPU NVIDIA ДЛЯ НЕГРАФИЧЕСКИХ РАСЧЕТОВ

ГРЕБНОВ И.В., асп.

Рассматривается пакет прикладного программирования NVIDIA CUDA SDK, который позволяет производить неграфические расчеты на графических процессорах компании NVIDIA.

*Ключевые слова:* неграфические задачи, программируемый процессор, умножение матриц большой размерности.

### «GPU NVIDIA» APPLICATION FOR NON-GRAPHICAL CALCULATION

GREBNOV I.V., postgraduate

The article concerns NVIDIA CUDA SDK application package which enables to carry out non-graphic calculations using NVIDIA graphic processors (GPU).

*Key words:* non-graphical problems, programmed processor, high dimensionality matrices multiplication.

**Введение.** Еще в 2004 г. компания NVIDIA, мировой лидер в технологиях программируемых графических процессоров, поняла, что существует возможность использования GPU для решения неграфических задач. Но для этого нужно найти способ упрощения программирования и увеличения быстродействия расчетов. Для решения этой задачи была создана специальная группа по разработке нового аппаратного и программного обеспечения. Новое направление получило название CUDA (Compute Unified Device Architecture).

16 февраля 2007 г. компания NVIDIA выпустила первую публичную версию набора для разработчиков ПО NVIDIA CUDA SDK и компилятор C для расчетов на графических процессорах NVIDIA.

GPU – это мощный, программируемый процессор, который подходит для многих вычислительных приложений. С учетом недавнего представления технологии CUDA SDK вычислительные возможности GPU теперь доступны практически всем разработчикам через стандартный язык программирования C.

Новая технология доступна на новейших видеокартах класса G80 и выше. Эти графические процессоры с поддержкой CUDA SDK позволяют организовывать вычисления на более чем ста унифицированных процессорных ядрах с частотой свыше 1 ГГц, что в общей сложности дает производительность свыше 300 ГФлоп/с. Для сравнения, новейший Intel Core 2 eXtreme QX6800 имеет производительность всего порядка 100 ГФлоп/с. Сравнение производительности современных GPU и CPU представлено на рис. 1.

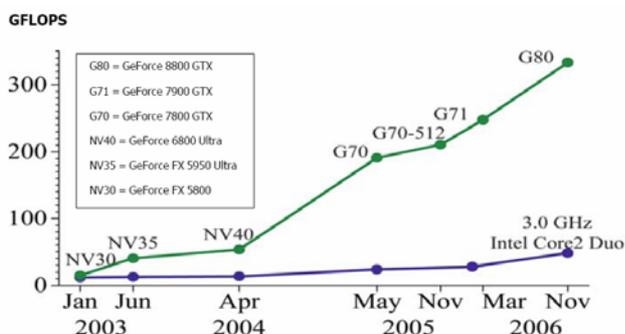


Рис. 1. Производительность GPU us CPU

**Архитектура системы.** GPU с CUDA работает как гибкий потоковый процессор, где тысячи вычислительных программ, называемых потоками, вместе решают сложные задачи. Каждый такой потоковый про-

цессор обладает небольшой локальной памятью, называемой регистрами. Все потоковые процессоры объединены в мультипроцессор. В свою очередь, каждый мультипроцессор обладает общей для процессоров кэш памятью и блоком управления (рис. 2).

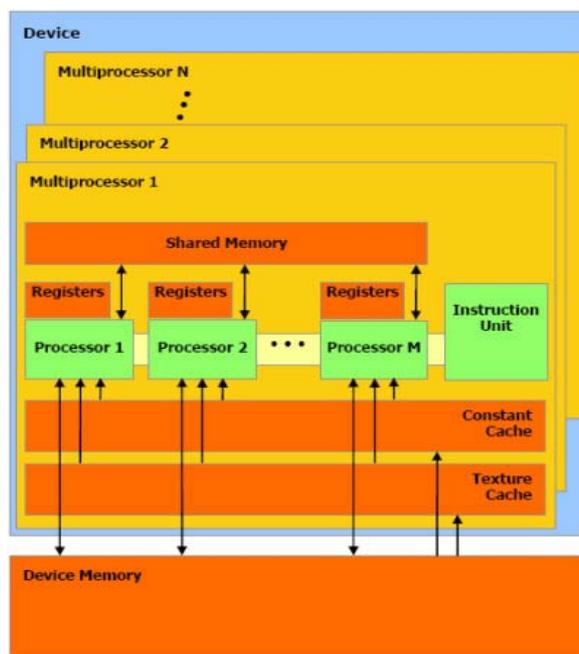


Рис. 2. Структурная схема G80

Между собой мультипроцессоры могут общаться и синхронизироваться только через память устройства.

Благодаря такой гибкой архитектуре компания NVIDIA может легко изменять количество мультипроцессоров в своих графических адаптерах.

CUDA может обращаться напрямую к OpenGL и DirectX. Кроме того, помимо унифицированных потоковых процессоров в графический процессор входят блоки растеризации и выборки текстур, и, в отличие от обычной памяти с функцией чтения/записи, они кэшируются, поэтому в некоторых случаях их тоже можно задействовать при решении специфических задач, что в общей сложности дает производительность свыше 500 ГФлоп/с.

**Вычислительная модель (рис. 3).** Каждый вызов функции для GPU NVIDIA представляет собой

сеть вычисления. В свою очередь, сеть разбивается на блоки, а блоки разбиваются на потоки. Идея заключается в том, что одиночный блок исполняется на одном мультимикропроцессоре, что позволяет потокам в пределах данного блока поддерживать связь и синхронизироваться друг с другом. Существует возможность задавать количество потоков на блок, а также количество блоков в сети. Следует отметить, что блоки исполняются параллельно на одном или нескольких мультимикропроцессорах и синхронизация на уровне блоков не поддерживается.

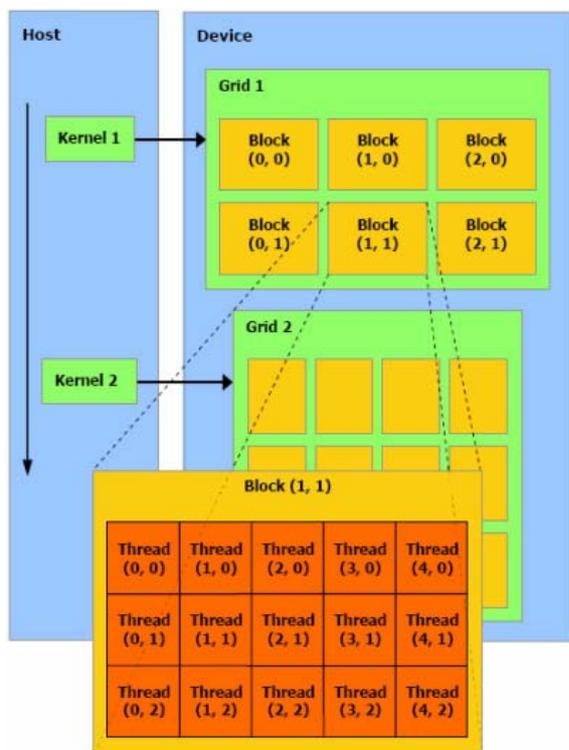


Рис. 3. Вычислительная модель

**Программная модель.** NVIDIA CUDA SDK состоит из двух основных компонентов:

- CUDA Display Driver;
- CUDA Toolkit.

CUDA Display Driver представляет собой модифицированную для CUDA версию драйвера. Драйвер выполняет функцию связующего звена, транслируя вызовы между программным обеспечением и оборудованием.

CUDA Toolkit представляет собой набор библиотек CUDA SDK, а также компилятор C для написания своих функций. В состав библиотеки CUDA SDK уже включены такие необходимые пакеты, как BLAS и FFT.

Компилятор языка C, который входит в состав CUDA Toolkit, на самом деле представляет собой надстройку над уже существующим компилятором системы. В этот компилятор введены новые ключевые слова `__device__` и `__host__`, с помощью которых можно указать, какие функции следует исполнять на GPU, а какие на CPU. Функции, предназначенные для исполнения на CPU, транслируются компилятором системы. Перенести функции на GPU нужно в том случае, если эти расчеты можно хорошо распараллелить и если они требуют небольшого количества локальной памяти. Следует также учесть ряд ограничений, накладываемых CUDA

Гребнов Илья Викторович,  
 ГОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,  
 аспирант,  
 телефон (4932) 26-98-29.

SDK, например, использование рекурсий и погрешностей арифметических расчетов.

**Результаты испытаний.** В качестве задачи для испытания NVIDIA CUDA SDK была выбрана задача умножения матриц большой размерности. Модель разбиения задачи на сеть, блоки и потоки представлена на рис. 4.

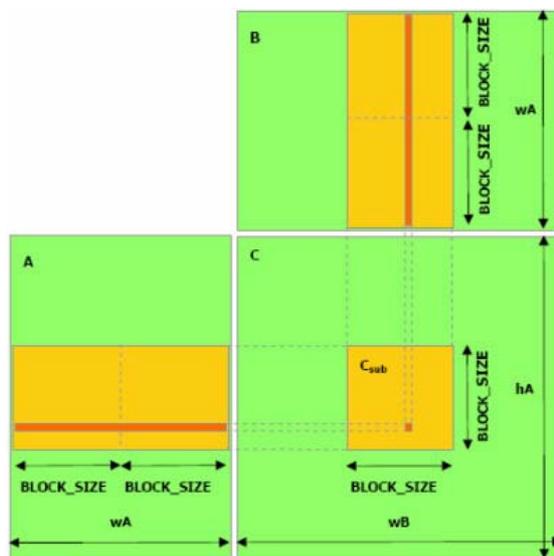


Рис. 4. Модель умножения

Умножение производилось по блокам, причем расчетом одного элемента конечной матрицы занимался отдельный поток. Также для увеличения быстродействия при расчете одного блока данные из матриц источников предварительно копировались в сверхбыструю локальную память мультимикропроцессора.

В результате испытания было установлено, что за одну секунду CUDA SDK делает в 2–3 раза больше вычислений, по сравнению с наиболее мощными современными CPU (сравнение производилось с двумя двухядерными процессорами Xeон и видеокартой NVIDIA Geforce 8800 GTS 320Mb), при экономии денежных средств в размере около 70%.

### Заключение

NVIDIA CUDA SDK и графические процессоры NVIDIA обладают большой вычислительной мощностью, однако необходимо отметить, что, по сравнению с традиционными процессорами, эти графические процессоры имеют и существенные недостатки, связанные с архитектурными особенностями. На текущий момент остается вопрос, насколько мощным является GPU с точки зрения применения для более сложных вычислений. Но в любом случае новая технология заслуживает внимания.

### Список литературы

1. <http://developer.nvidia.com/object/cuda.html>, «NVIDIA CUDA Programming Guide».
2. <http://forums.nvidia.com/>, «General CUDA GPU Computing Discussion».
3. <http://www.gpgpu.org/forums/>, «Advanced Development».