

Кульман Т.Н., Шлейхер В.С.
ИЗУЧЕНИЕ ВОЗМОЖНОСТЕЙ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИИ CUDA
Международный университет природы, общества и человека «Дубна», филиал «Протвино»
(кафедра информационных технологий)

Данная работа рассматривает технологию CUDA, позволяющую программисту использовать видеокарты в качестве мощных вычислительных устройств и о методике установки этой технологии на компьютере разработчика.

Технология NVIDIA CUDA™ (Compute Unified Device Architecture) - это среда разработки на C, которая позволяет программистам и разработчикам создавать программное обеспечение для сложных вычислительных задач, благодаря многоядерной вычислительной мощности графических процессоров. Другими словами, суть технологии заключается в возможности использования GPU-вычислений (Graphic Processor Unit) для обработки данных, которая обычно ложится «на плечи» центрального процессора.

GPGPU - универсальные вычисления на GPU

В 2003 году секция SIGGRAPH («Computations on GPUs») была выделена под новую область ИТ. Эта инициатива получила название *GPGPU* (General Programming on GPU, универсальные вычисления на GPU). Развитие GPGPU привело к созданию компанией NVIDIA технологии CUDA. Продукты с поддержкой NVIDIA CUDA ускоряют решение самых требовательных задач. CUDA присутствует в GeForce 8800, Quadro FX 5600/4600 и выше [1].

Особенности CUDA [2]:

- хорошо подходит для решения широкого круга задач с высоким параллелизмом;
- улучшает и упрощает модель программирования GPU, добавляет большое количество возможностей;
- предоставляет недорогой персональный суперкомпьютер без вложения дополнительных средств;
- подключается к приложениям с Microsoft Direct X и OpenGL;
- создана для Linux и Windows;
- является необходимой основой при построении высокоуровневых инструментов.

Характеристики [2]:

- интерфейс основан на языке программирования C;
- отдельный программно-аппаратный стек для вычислений;
- отсутствие необходимости в графических API;
- эффективная передача данных между системной и видеопамятью;
- доступ к разделяемой между потоками памяти размером в 16 Кб на мультипроцессор;
- аппаратная поддержка целочисленных и битовых операций;
- линейная адресация памяти с возможностью записи по произвольным адресам.

Модель программирования

Вычислительная архитектура CUDA основана на концепции одна команда на множество данных (Single Instruction Multiple Data, SIMD) и понятии мультипроцессора, устройства, хоста и ядра [3].

Блочно-сеточная организация [4] устроена следующим образом: потоки объединяются в блоки (thread block), блоки - в сетки (Grid). Потоки взаимодействуют между собой с помощью разделяемой памяти. Программа (ядро) исполняется над сеткой (рис.1). Одновременно исполняется одна сетка.

Для идентификации потоков используются двухмерные индексы. Блок выполняется на мультипроцессоре частями, или пулами, называемыми warp (в 32 потока). Задачи внутри пула исполняются в SIMD стиле.

Группировка блоков в сетки позволяет уйти от ограничений и применить ядро к большому числу потоков за один вызов.

Компоненты:

- новейший драйвер CUDA [5];
- полный набор инструментов CUDA [5];
- примеры кодов CUDA SDK [5] (например, замеры производительности с использованием таймеров, дискретное вейвлет-преобразование (одномерный вейвлет Хаара), свёртка изображений, демонстрация взаимодействия с OpenGL и Direct3D, вычисление цены опциона по методу Монте-Карло и др.).

Установка

Необходимы три файла:

- NVIDIADisplayWin2K(177_84)Int.exe;
- NVIDIA_CUDA_toolkit_2.0_win32.exe;
- NVIDIA_CUDA_SDK_2.02.0811.0240_win32.exe.

Перед установкой нужно убедиться в наличии периферийных устройств, обеспечивающих поддержку CUDA.

Первым устанавливается `NVIDIADisplayWin2K(177_84)Int.exe`. При данной установке происходит инсталлирование драйвера CUDA.

Вторым устанавливается `NVIDIA_CUDA_toolkit_2.0_win32.exe`.

Третьим - `NVIDIA_CUDA_SDK_2.02.0811.0240_win32.exe`.

После установки всех трёх файлов на компьютер необходимо совершить перезагрузку для корректной последующей работы программы.

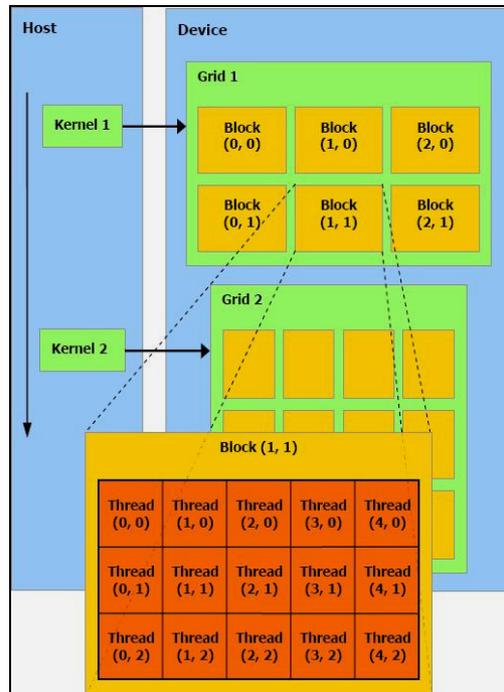


Рис. 1. Блочно-сеточная организация

Заключение

В будущем сложные вычисления предполагается выполнять с помощью параллельных алгоритмов. Технология CUDA является успешным примером применения этого подхода.

В данной работе была создана методика установки CUDA на ПК разработчика и изучены некоторые приемы программирования. В дальнейшем предполагается применение этой технологии в фирме ООО «Систел», занимающейся автоматизацией в энергетике, для решения расчетных задач и задач быстрого отображения графической информации.

Библиографический список

1. Технология NVIDIA CUDA. Документация CUDA. 1с.
2. Основы CUDA. www.ixbt.com.
3. CUDA, или Суперкомпьютеры – в массы. // Журнал для программистов C/C++. №6. С 3-4.
4. Lecture_22. // Лекции CUDA © NVIDIA Corporation 2006.
5. Программирование в CUDA. Документация CUDA. 3с.