

NVIDIA CUDA

ИСТОРИЯ

Комаров Сергей

GPU

GPU (device) это вычислительное устройство, которое:

- **Является сопроцессором к CPU (host)**
- **Имеет собственную память (DRAM)**
- **Выполняет одновременно очень много нитей**
- **CUDA появилась в 2006 году и представляет из себя программно-аппаратный комплекс производства компании NVIDIA**

CUDA 1.0

GeForce 8800 GTX выпущенная 8 ноября 2006.

Что получили разработчики:

- возможность напрямую, с помощью обычных указателей, адресовать память GPU
- 128 ядер
- 1350 MHz
- 768 Mb
- 681 млн транзисторов
- «сырость» CUDA-библиотек `cudaErrorNotYetImplemented`

НО! Эта архитектура во многом определила основные ограничения и возможности всей линейки графических адаптеров.

CUDA 1.1

GeForce 9800 GTX+

Что получили разработчики:

- 128 ядер
- 1836 MHz
- 512/1024 Мб
- 1.4 млрд транзисторов
- атомарные операции в “global memory”

CUDA 1.3

GeForce 285 GTX и Tesla C1060

Что получили разработчики:

- **240 ядер**
- **1476 MHz**
- **1/2 Gb**
- **1.4 млрд транзисторов**
- **числа с двойной точностью**
- **контроль четности памяти**

CUDA 2.0

GeForce 480 GTX и Tesla C2050

GeForce 580 GTX и Tesla M2090

Что получили разработчики:

- **480 ядер**
512 ядер
- **1401 MHz**
1544 MHz
- **1.5 Gb**
1.5/3 Gb
- **3 млрд транзисторов**
- **Кэш L1/L2**

CUDA 3.0

GeForce 680 GTX и Tesla K10

Что получили разработчики:

- **1536 ядер**
- **1006 MHz**
- **2 Gb**
- **3.5 млрд транзисторов**
- **Поддержка 64-битной адресации**
- **Унифицированное адресное пространство**
- **Tesla K10: добавлен еще один GPU**
Tesla K10 ~ 2 * *GeForce 680 GTX*

CUDA 3.5

Tesla K20X

Что получили разработчики:

- 2688 ядер
- 650 MHz
- 6 Gb
- 7.1 млрд транзисторов
- Повышена пиковая производительность чисел с одинарной и двойной точностью
- поток может породить другие
- поток может вызвать функцию из другой библиотеки

ТЕНДЕНЦИИ

- Maxwell – процессор, который будет поддерживать единую виртуальную память и предоставит CPU доступ к быстрой памяти, встроенной в GPU и наоборот
- Volta - процессор в котором модули памяти будут расположены прямо поверх ядер GPU
- SoC – System on Chip (система на чипе)