**ПРИМЕНЕНИЕ МУЛЬТИПРОЦЕССОРНОГО МОДУЛЯ CUDA ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА И СКОРОСТИ КОДИРОВАНИЯ ТВ-СИГНАЛОВ**

Аминев Д.А., Малкин Д.Е., Увайсов С.У.

*МИЭМ НИУ ВШЭ*

Рассматривается проблема повышения качества и скорости кодирования кодирования ТВ-сигналов. Приводится структурная схема архитектуры CUDA. Представлены результаты тестирования архитектуры относительно производительности.

**Usage of multiprocessor module CUDA for increasing of quality and speed of TV signals encoding. Aminev D.A., Malkin D.E., Uvaysov S.U.**

Issue of TV signals encoding speed increasing is studied. Structure chart of CUDA is shown. Testing results of encoders relate with performance is given.

В телекоммуникационных, телевизионных, графических системах и системах дистанционного зондирования возникает необходимость кодирования видеопотоков [, ]. В настоящее время наиболее распространенным стандарт является H.264 [].

В настоящее время кодирование ТВ-сигналов по стандарту H.264 реализуется программным путем, что требует значительных затрат процессорного времени. Однако технология вычислений на основе графических процессоров CUDA (Compute Unified Device Architecture) позволяет снять часть нагрузки с центрального процессора для высокопроизводительной обработки массивов данных больших объемов.

Технология CUDA реализуется на [графических](http://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D1%84%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D0%BF%D1%80%D0%BE%D1%86%D0%B5%D1%81%D1%81%D0%BE%D1%80) ускорителях начиная с [GeForce](http://ru.wikipedia.org/wiki/GeForce) восьмого поколения ([GeForce 8 Series](http://ru.wikipedia.org/wiki/GeForce_8_Series), [GeForce 9 Series](http://ru.wikipedia.org/wiki/GeForce_9_Series), [GeForce 200 Series](http://ru.wikipedia.org/wiki/GeForce_200_Series)), а также [Nvidia Quadro](http://ru.wikipedia.org/wiki/Nvidia_Quadro) и [Tesla](http://ru.wikipedia.org/wiki/NVidia_Tesla) компании [Nvidia](http://ru.wikipedia.org/wiki/Nvidia) [].

CUDA даёт возможность организовывать доступ к набору инструкций и памяти графического ускорителя, используя его для сложных [параллельных вычислени](http://ru.wikipedia.org/wiki/%D0%9F%D0%B0%D1%80%D0%B0%D0%BB%D0%BB%D0%B5%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5_%D0%B2%D1%8B%D1%87%D0%B8%D1%81%D0%BB%D0%B5%D0%BD%D0%B8%D1%8F)й. Графический ускоритель, поддержи­вающий технологию CUDA, имеет программируемую открытую архитектуру, по возможностям приближающуюся к современным [центральным процессорам](http://ru.wikipedia.org/wiki/%D0%A6%D0%B5%D0%BD%D1%82%D1%80%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%BF%D1%80%D0%BE%D1%86%D0%B5%D1%81%D1%81%D0%BE%D1%80).

В графическом ускорителе используется grid-модель памяти, кластерное моделирование потоков и [SIMD](http://ru.wikipedia.org/wiki/SIMD)-инструкции. Эта технология применяется при высокопроизводительных графических вычислениях и разработке графического API, совместимого с NVIDIA.

Общая структурная схема взаимодействия центрального двухъядерного процессора (Host) и графического процессора (Device) по технологии CUDA представлена на рис.1.

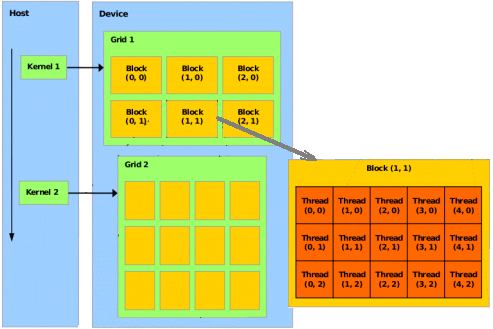


Рис.1. Cтруктурная схема взаимодействия центрального двухъядерного и графического процессоров по технологии CUDA.

Графический процессор (Device) является многоядерным (число ядер может быть до нескольких сотен). В показанном примере этот процессор состоит из двух гридов (Grid 1 и Grid 2), содержащих несколько блоков (Block), которые, в свою очередь, содержат определенное количество тредов (Thread). Здесь:

тред – наименьший элемент, выполняющий инструкцию;

блок – объединение нескольких тредов;

грид – объединение нескольких блоков;

ядро (Kernel) – приложение или программа, выполняемые графичес­ким процессором.

Устройство обладает общей памятью порядка 1 Гбайт и более, доступной всем блокам, а также разделяемой памятью от 16 Кбайт на каждый блок, за счет которой можно повысить быстродействие программы, поскольку использование разделяемой памяти уменьшает число конкурентных запросов к общей памяти.

Устройства с СUDA классифицируются по версиям «вычислительных возможностей» (compute capability). Сompute capability v 2.1 – максимальная из возможных [].

Самые последние аппаратные решения из серии Quadro и Tesla позволяют достигать максимального быстродействия с использованием CUDA.

Технология CUDA имеет некоторые недостатки. Во-первых, ее поддерживает и развивает только производитель NVidia. Во-вторых, все функции, выполняемые на устройстве, не поддерживают рекурсии и имеют некоторые другие ограничения.

Одной из задач обработки, которую можно решать с использованием платы CUDA, является реализация кодера сжатия по стандарту H.264 для телевидения высокой четкости (1920×1080 пикселов), когда скорость цифрового потока по интерфейсу 3Gb SDI от источника видеоинформации (камеры ТВЧ) достигает значения 3 Гбит/с.

Представляется, что ускорение обработки, достигаемое с помощью платы CUDA, позволит реализовать кодер H.264 для ТВЧ в реальном времени. Предлагается следующая структурная схема вычислительной системы на основе архитектуры CUDA (рис. 2).



Рис.2. Структурная схема архитектуры CUDA

Здесь входной поток видео от камеры ТВЧ поступает в оперативную память компьютера (ОЗУ) через модуль видеозахвата и контроллер USB 3.0. Затем этот поток кодируется в модуле GPU CUDA и записывается на RAID-систему []. При этом в процессе кодирования ресурсы центрального процессора не задействованы.

Кодирование проводилось с использованием кодека cudaH264enc в составе продукта Mediacoder, разработанного корейским ученым Huang Yi-chun Stanley []. Альтернативным кодеком на основе центрального процессора (ЦП) был выбран x264 [], распространяемый по лицензии GNU GPL [].

В проводимых тестах скорости потока несжатого видео составляли менее 3 Гбит/c ввиду отсутствия ТВЧ камеры. Однако это не повлияло на оценку сравнения характеристик кодеков.

При тестировании формат исходного видео – mpeg2 с параметрами 720х576p25; формат результирующего видео – H.264 с параметрами 720х576р25. Результаты тестирования двух кодеков приведены в таблице 1.

Таблица 1. Результаты тестирования.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Кодек | Параметры кодирования | Средняя скорость кодирования, fps | Загрузка ЦП, % | Метрика SSIM, среднее значение по всем кадрам | Размер закодирован-ного файла, Байт |
| x264 | Заданные по умолчанию | 645,0 | 47 | 0,50171 | 431 160 699 |
| Заданные по умолчанию с включенным no-deblock | 730,0 | 44 | 0,49848 | 431 429 438 |
| cudaH264enc | Заданные по умолчанию | 144,6 | 93 | 0,50252 | 453 559 160 |
| Заданные по умолчанию с включенным no-deblock | 145,4 | 93 | 0,49751 | 453 553 834 |

Из таблицы видно, что скорость кодирования с применением кодека cudaH264enc примерно в 5 раз больше скорости кодирования кодека X264, при этом загрузка центрального процессора составляет не более 50%. Метрика SSIM показывает незначительной разницу между качествами изображения для обоих кодеков. Размер закодированного файла показывает небольшую разницу в эффективности сжатия.

Тестирование проводилость на компьютере со следующей конфигурацией: процессор Intel Core i7-2600 CPU @ 3.4 GHz, видеокарта Nvidia GeForce GTX 560 Ti, операционная система Windows 7 x64.

Кодер х264 на основе центрального процессора также может обрабатывать видео с большей скоростью, однако при этом эффективность сжатия и качество изображения значительно снижаются.

Таким образом, реализация процесса кодирования ТВ-сигналов по стандарту H.264 на основе мультипроцессорного модуля CUDA дает значительный выигрышь в скорости кодирования при незначительной загрузке центрального процессора. Такую систему можно реализовать на основе рядового ноутбука с графическим процессором nvidia.

**Литература**

1. Серов А.В. Эфирное цифровое телевидение DVB-T/H. //Санкт-Петербург «БХВ-Петербург», 2010. – 464 с.
2. Аминев Д.А., Увайсов С.У. Анализ протоколов для передачи высокоскоростных потоков данных в межмодульных сетевых соединениях // VI международная научно-практическая конференция «Информационные и коммуникационные технологии в образовании, науке и производстве». - Протвино. -2-6 июля 2012. - С. 198–201.
3. ISO/IEC 14496-10 – cтандарт H.264 / AVC / MPEG-4 Part 10.
4. www.nvidia.ru – компания Nvidia – разработчик и производитель графических процессоров.
5. Аминев Д.А. Дисковые подсистемы: достижение максимальной скорости при наименьшем количестве дисков.// Цифровая обработка сигналов. - Москва. – 2008. - № 4, - С. 57–59.
6. mediacoderhq.com – универсальный медиа транскодер.
7. http://www.videolan.org/developers/x264.html - x264 - кодер H.264 / AVC
8. www.gnu.org/licenses/gpl.html – GNU GENERAL PUBLIC LICENSE.