

Расчеты на видеоплатах

Дрогалин С.В., ОрелГТУ

Вычислительная мощность графических процессоров позволяет намного быстрее обрабатывать видео, графику, ускорить исполнение различных программ.

Деятельность людей в большинстве случаев происходит последовательно. Это значит, что после выполнения одной задачи они приступают к другой, а мелкие незначительные проблемы решаются как бы между делом. Аналогично работает и главный процессор любого компьютера.

В большинстве случаев последовательное исполнение программы является необходимостью, так как результаты предыдущих вычислений используются для последующих операций. Но некоторые вычисления совершенно не зависят друг от друга и вполне могли бы производиться одновременно. Например, при обработке запросов к базам данных, построении научных моделей или кодировании видео. Для подобных задач центральный процессор имеет недостаточно оптимальную структуру.

Прекрасное дополнение к нему имеется практически в любом компьютере и называется видеоплатой. Они и разрабатывались исключительно для того, чтобы параллельно вычислять огромное количество точек изображения. Однако работать видеопроцессоры могут только через графические интерфейсы программирования DirectX и OpenGL.

Уже достаточно давно идут разговоры о возможностях использования видеокарт для вычислений, отличных от 3D-графики. Сложность и мощность видеочипов, степень их программируемости постоянно растут, а "специализация" вычислений становится все более "общей". Как результат уже существуют проекты ([BrookGPU](#), например), которые ставят своей целью вывести использование 3D-акселераторов из довольно узких игровых и профессиональных ниш.

Задача создания соответствующих инструментов для разработчиков является весьма длительной и может потребовать взаимодействия непосредственно

с производителями видеочипов. До сих пор сведений о реальной реализации каких-то вычислительных проектов не поступало, теперь на сайте [Beyond3D](#) появилась информация о работах по созданию версии клиента распределенных вычислений Folding@home, использующего для работы мощности чипа видеокарты. Кроме радости от "первого шага" в правильном направлении есть и "ложка дегтя" - скорее всего, работа клиента будет организована на видеокартах класса GeForce 6800 / Radeon X800 . Впрочем, к тому времени когда работы будут завершены, данные модели видеокарт уже могут получить широкое распространение.

Intel стремится выпустить свое дискретное графическое решение, — это не только желание выйти на рынок видеочипов, а скорее стремление не потерять положение на рынке процессоров. Ведь грядущий рынок GPGPU-решений может в корне изменить текущее положение вещей в индустрии. Ведь при таком развитии рынка, которое происходит сейчас, главным вычислительным блоком в ПК может стать видеокарта, которая возьмет на себя большинство самых ресурсоемких задач — это и игры, и кодирование/декодирование видео и аудио, архивирование данных, 3D моделирование, обработка изображений, физические расчеты и т. д. На все это уже сейчас способны графические решения.

Конечно, пока нет широкой поддержки этих возможностей видеочипов со стороны производителей ПО, но эта ситуация не будет длиться вечно. Например, Acrobat 8 и Acrobat Reader 8 используют мощности современных видеочипов для более плавного и приятного отображения документов, вспомните также NVIDIA Gelato 2.1 и технологию NVIDIA CUDA.

Современные средние видеочипы по совокупной вычислительной мощности на порядок превосходят 4-ядерные процессоры Intel, при несопоставимой цене.

Видеокарты не заменят процессоры (ведь видеочипы ориентированы на широкие математические параллельные вычисления). Однако, при покупке компьютера главным решением будет видеокарта. AMD разработала архитектуру Fusion, которая объединит CPU и GPU в один чип.

Грядущие изменения архитектуры ПК, смогут в разы увеличить вычислительные способности.

Устройства с поддержкой DirectX 8 стали первым камнем в фундаменте технологии универсальных графических процессоров. Как известно, для DirectX 8 требуется наличие свободно программируемых тендеров (своего рода вычислительных модулей) в GPU. Тем самым (по крайней мере, теоретически) можно исполнять почти любой программный код пиксельными тендерами.

Если сравнивать производительность центрального и графического процессоров при работе с числами с плавающей запятой, то в данном случае возможности GPU значительно выше. За последние четыре года производительность CPU выросла в четыре раза — с 5 Гфлоп/с (у Pentium 4) до 20 Гфлоп/с (у процессоров Core 2 Duo), то у графических процессоров за тот же период быстродействие увеличилось с 15 Гфлоп/с (ATI Radeon 9700) до невероятных 520 Гфлоп/с (серия NVIDIA GeForce 8800).

Преимущество в скорости вычислений объясняется возрастающим количеством шейдеров в графическом чипе. Если программа будет использовать одновременно каждый отдельный шейдер, то вычисления могут производиться сразу в 48 потоках (как, например, при использовании ATI Radeon X1950 XTX). Самые быстрые четырехъядерные процессоры работают в четыре потока, а общедоступные на сегодняшний день двухъядерные — в два потока.

Использование новой технологии тормозилось в первую очередь из-за проблем с созданием соответствующего ПО. До конца 2006 года не существовало стандартных библиотек SDK (Software development kit), с помощью которых, используя известные языки программирования Java, C или C++, можно было бы создавать программы для платформы GPUPU. Программирование происходило с использованием графических интерфейсов, DirectX и Open GL, и это требовало даже внесения изменений в структуру программ. Графические интерфейсы попросту не обладают целым рядом вычислительных операций, как, например, определение минимума и максимума в перечне данных. Такая ограниченность набора инструкций препятствует простому и быстрому переносу существующих

программ в среду GPGPU. Помимо этого функции графических процессоров у производителей отличаются так сильно, что программа, разработанная для одного чипа, как правило, не работоспособна при использовании чипа-конкурента, и это делает ее непригодной для массового рынка.

Тем не менее крупные производители видеочипов явно видят в технологии GPGPU большое будущее и огромный потенциал. Сегодня и AMD и NVIDIA уже создали собственные GPGPU – библиотеки для разработки программ. У AMD такая библиотека называется Close to the metal (CTM), у NVIDIA – Cuda. Оба набора библиотек работают с соответствующим компилятором для известного языка программирования C, который преобразует программу таким образом, что с ней может работать графический процессор. Кроме того, в сентябре 2006 года корпорация AMD представила видеоплату FireGL, задуманную только для работы с приложениями GPGPU. Первые серверы с графическим кластером уже работают.

Современные разработки по комбинированию аппаратной и программной составляющих ПК наконец-то открыли перед программистами доступ к потенциалу видеоплат, не требуя их коренной переделки. Самыми известными на сегодняшний день GPGPU-программами являются средства для расчета физических эффектов, а также проект Станфордского университета под названием Folding@home. Идущая в этом же направлении компания NVIDIA заявляет, что с помощью новой технологии ей удалось добиться увеличения скорости моделирования биржевых курсов почти в 200 раз.

В скором времени могут появиться антивирусные программы, которые не станут загружать CPU, а также веб-серверы, обрабатывающие запросы к базам данных с большой скоростью. По меньшей мере библиотеки CTM и Cuda стали основой, которая облегчает труд программистов.

Cg имеет общие с языком C структуры, типы данных, способы адресации элементов классов. С другой стороны, он имеет множество готовых конструкций для проведения вычислений на видеокарте.

Теперь представим часть программы на языке Cg.

```
struct C2E1v_Output {
    float4 position : POSITION;
    float4 color    : COLOR;
};

C2E1v_Output C2E1v_green(float2 position : POSITION)
{
    C2E1v_Output OUT;

    OUT.position = float4(position, 0, 1);
    OUT.color    = float4(0, 1, 0, 1); // RGBA green

    return OUT;
}
```

Список литературы:

1. Журнал CHIP [Текст]: информ. -аналит. журн. – 2007, июнь
М.: 2007- .- Ежемес. ISSN 1609 – 4212
2. Информатика и информационные технологии. Угринович Н. Д. – М.: Би-
ном. Лаборатория Занятий, 2002. – 512 с.: ил.