



Тайловый рендеринг

Тайловый ускоритель

В ближайшем будущем нас ожидают сразу два 3D-ускорителя для домашних ПК на базе тайлового рендеринга: это Kyro III и Columbia. Попробуем разобраться, что же принесет нам тайловая архитектура, попытаемся найти ее сильные и слабые стороны, а также проанализируем уже существующие чипы на ее основе.

Для начала давайте разберемся с основными принципами работы тайлового ускорителя и выясним, почему же нас не устраивает обычная, полигонная архитектура.

Что мы видим на экране монитора? Мы видим двухмерную проекцию трехмерного мира. В этом мире все объекты состоят из треугольников и многоугольников (полигонов). У каждого из полигонов есть ряд свойств: координаты (x, y, z) вершин, цвет вершин, глобальные свойства полигона (такие как текстура).

Очевидно, чтобы вы смогли взаимодействовать с виртуальным миром, компьютеру понадобится менять последовательно кад-

ры на экране монитора, на каждом из которых вы будете видеть результат того или иного действия, совершенного вами. Выстрелили вы из ружья — через определенное время увидите следы от пуль или, если повезет, сами пули, что в играх встречается нечасто.

В предыдущих публикациях мы уже касались такого понятия, как 3D-конвейер. Чтобы построить кадр, ускорителю необходимо выполнить ряд последовательных действий, которые и получили названия 3D-конвейера. Он состоит из нескольких ключевых этапов: трансформация и освещение, удаление невидимых поверхностей, текстурирование и затенение. Общее количество »

» этапов и подэтапов в различных ускорителях варьируется, но в последних моделях оно практически всегда больше десяти.

Ключевые этапы 3D-конвейера

Вкратце напомним, для чего служат этапы, обозначенные мной как ключевые. На этапе трансформации моделируются движение объекта и его перемещение, вращение, изменение размеров, формы и т. п. Это происходит путем изменения координат вершин и требует довольно интенсивного использования операций линейной алгебры и тригонометрических функций. На этапе расчета освещения (и затенения) объекта осуществляется вычисление освещенности каждого элементарного полигона. Удаление невидимых поверхностей (Hidden Surface Removal, HSR) — достаточно важный этап. На нем исключаются не видимые наблюдателем (виртуальной камерой) точки поверхности объекта. На этапе текстурирования на поверхность элементарных полигонов накладываются текстуры — специальные растровые картинки, создаваемые для имитации тех или иных материалов. Далее вычисляется цвет пикселя, исходя из цвета элементарных полигонов и их текстур, а также основываясь на данных, полученных на предыдущих этапах.

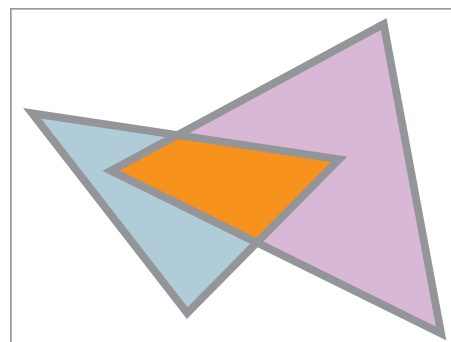
Здесь возникает вопрос: часть объектов может перекрываться другими объектами? Какие-то из них будут видны, какие-то нет. Допустим, у нас есть обычный треугольник. И так получилось, что его часть совпала с другим треугольником (рис. 1). Давайте разберемся, как поведут себя в этом случае традиционная и тайловая архитектуры.

Традиционный рендеринг

Что понимается под обычной (традиционной) архитектурой рендеринга? Обычная архитектура была изобретена еще в шестидесятых годах. Все полигоны сцены в этом случае рендерятся поочередно. Пока последний полигон не пройдет через 3D-конвейер, система не будет иметь ни малейшего представления о том, какая часть сцены завершена, а какая нет. Следовательно, через 3D-конвейер пройдут все полигоны, даже те, которых мы не увидим на экране, что приводит к бесполезной трате вычислительных ресурсов. После того как все полигоны «сошли» с 3D-конвейера, карта выясняет, какой именно пиксель выводить на

экран, а какой нет. Для решения этого вопроса в картах обычной архитектуры применяется Z-буфер — область в памяти ускорителя. В Z-буфере хранится значение удаленности каждого пикселя от виртуальной камеры. Каждый пиксель каждого полигона после прохождения предыдущих этапов сравнивается со значением, хранящимся в Z-буфере. Таким образом выясняется, какой из пикселей лежит ближе всего к виртуальной камере. Все пиксели, что лежат за ним на одной прямой, параллельной оси Z, прорисовываться не будут.

Таким образом, в случае, показанном на рис. 1, на 3D-конвейер сначала попадает первый треугольник, затем второй. Все это текстурируется, освещается, затеняется... Затем происходит Z-буферизация, и те пиксели нижнего треугольника (отмечены красным цветом), которые закрыты верхним треугольником, на экран не выводятся.



▲ Рис. 1. Пример Overdraw. Типичный случай нерационального использования ресурсов

Лишняя трата вычислительных ресурсов налицо.

Тайловый рендеринг

Теперь давайте разберемся, что такое тайловый рендеринг. Появился он уже достаточно давно, однако ускорители на базе

»



Преимущества тайлового рендеринга

Хорошие перспективы

Давайте рассмотрим, какие преимущества в теории нам сулит тайловая архитектура по сравнению с традиционной:

- Благодаря тому, что тайл достаточно маленький, большая часть операций над ним происходит во встроенных в чип буферах, что приводит к минимизации обращений к внешней памяти ускорителя. Как следствие, сохраняется пропускная способность шины обмена данными памяти с чипом — одного из самых слабых мест ускорителей с традиционной архитектурой (см. табл. 1). Учтите, что чем сложнее сцена и чем больше вы включаете различных технологий обработки графики, тем большая нагрузка приходится на шину.
- Операции, происходящие во встроенной в чип памяти, выполняются быстрее.
- Через процессы текстурирования и освещения (или затенения) проходят только пиксели, которые мы увидим на экране. Не происходит лишних обращений к текстурной памяти.

- Если еще несколько лет назад overdraw в играх было равно 1,2 или 1,5, то в настоящее время оно уже достигло значений 2 или 3. С выходом новых версий DirectX и OpenGL возникают все более реалистичные миры, число полигонов в сцене растет. Но в то же время растет и overdraw. В ближайшем будущем этот параметр будет равен уже 4 или 5. Тогда тайловый рендеринг проявит себя во всей красе.
- Хорошая масштабируемость. Если на одной карте будет два или четыре чипа на базе тайлового рендеринга, то такая карта будет работать практически в два или четыре раза быстрее одночипового варианта. Учитывая, что для тайловой архитектуры проблема с пропускной способностью шины не так болезненна и что часть памяти находится в самом чипе, улучшение технологии производства чипа или даже его частоты может дать более заметные результаты, нежели в случае с традиционной архитектурой.

Разрешение	Традиционная архитектура	Тайловая архитектура
640x480	2,104	1,104
800x600	3,287	1,406
1024x768	5,386	1,935
1280x1024	8,977	2,832
1600x1200	13,149	3,868

▲ Табл. 1. Необходимая пропускная способность шины обмена данными между памятью и чипом для традиционной и тайловой архитектур, Гбайт/с (данные PowerVR)



Гибрид тайлового и традиционного рендеринга

Все только начинается

Инженеры NVIDIA и ATI уже пробуют внедрить некоторые элементы тайлового рендеринга в традиционную архитектуру, ну или хотя бы снизить overdraw. Например, в ускорителе Radeon 8500 используется технология Hierarchical Z (она начала применяться еще и в Radeon). Благодаря этой технологии перед рендерингом экран так-

же разбивается на тайлы для определения видимых и невидимых объектов. Нечто подобное есть и в GeForce 3. Правда, на практике смесь тайловой и традиционной архитектуры значительно уступает чистой тайловой архитектуре. Все-таки до совершенства этим разработкам еще очень далеко.

» тайлового рендеринга до недавнего времени не пользовались особой популярностью. Тайловый рендеринг создавался с целью минимизации overdraw. Overdraw — параметр, характеризующий среднее число перекрытий в трехмерной сцене, — возникает тогда, когда один объект перекрывается другим, и нижний объект или его часть становятся уже не видимыми с позиции виртуальной камеры, как показано на рис. 1.

Специально для минимизации overdraw изображение разбивается на маленькие прямоугольники — тайлы. Их размер может быть различным: 32x32, 32x16, 16x16 пикселей. В начале 3D-конвейера каждый тайл проходит обработку в Z-буфере, расположенном в самом чипе, а не в памяти ускорителя, как это происходит в случае с традиционной архитектурой. Заметьте, что благодаря встроенному в чип буферу удается снизить число обращений к внешней памяти ускорителя и, соответственно, увеличить скорость работы. Вот, собственно, и

вся изюминка тайлового рендеринга. Все невидимые поверхности удаляются еще до текстурирования и расчета освещения, то есть на уровне чистых полигонных моделей. Кроме этого, благодаря тому, что Z-буфер встроен в чип, никаких существенных потерь производительности при 32-битном качестве Z-буфера по сравнению с 24-битным и 16-битным происходить не будет, поэтому в последних ускорителях на базе тайлового рендеринга во всех режимах используется 32-битный Z-буфер, что приводит к улучшению качества изображения.

Таким образом, в случае, показанном на рис. 1, все изображение сначала будет разбито на тайлы. Затем каждый тайл пройдет стадию отброса невидимых поверхностей, или Z-буферизацию. В результате этого дальнейшей обработке будет подвергаться только видимая часть нижнего треугольника, точнее, большая его часть, потому как небольшой overdraw все равно будет сохраняться.

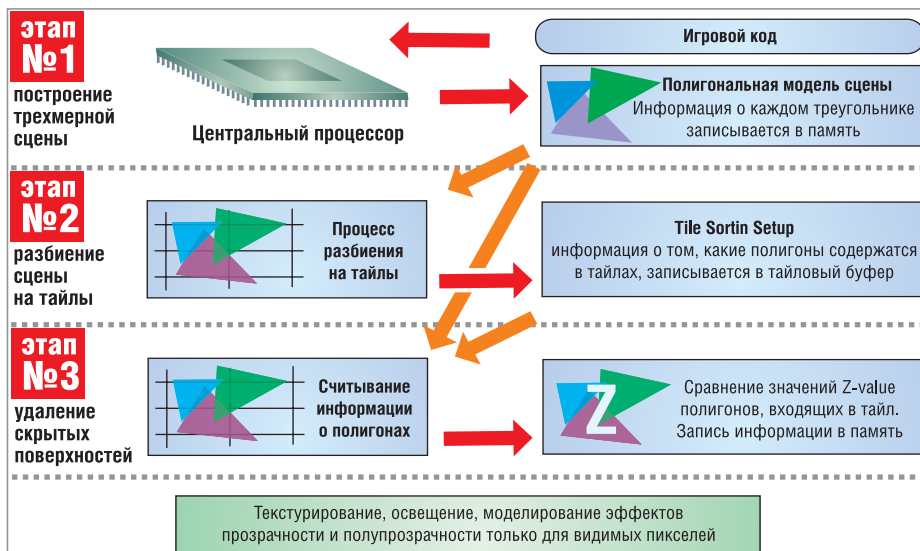
Построение изображения при помощи тайлового рендеринга

Для большей наглядности предлагаю рассмотреть полную сцену построения изображения чипом на базе тайлового рендеринга (рис. 2).

Что происходит, когда мы запускаем игру? Начинается построение трехмерной сцены, точнее, каркасной модели трехмерной сцены, состоящей из полигонов. Она строится полигон за полигоном, пока последний многоугольник не займет своего места. Вся работа здесь ложится на плечи CPU и блока T&L, если таковой на видеокарте имеется. Построение закончилось, вся сцена разместилась в кадровом буфере.

Далее все изображение разбивается на тайлы. Все полигоны считываются из кадрового буфера и проходят сортировку (Tile Sorting Setup), благодаря которой ускоритель будет знать, в каких именно тайлах находится данный полигон, ведь зачастую один полигон лежит в нескольких тайлах, как показано на рис. 3. Затем для каждого тайла выделяется определенная область во внутричиповой памяти — тайловый буфер, в котором хранится значение номера полигона (у каждого полигона в сцене есть свой уникальный номер), содержащегося в данном тайле. Номер полигона берется из кадрового буфера. Главное, чтобы алгоритм Tile Sorting Setup работал быстрее, чем на него поступают «свежие порции» полигонов. В принципе, во всех ускорителях тайлового рендеринга это предусмотрено. Этот алгоритм работает быстрее, чем центральный процессор, и довольно неплохо масштабируется в зависимости от скорости CPU.

Следующая стадия — удаление невидимых поверхностей (HSR). Этот алгоритм работает на тайловом уровне. На предыдущих этапах мы получили множество тайловых буферов, в которых располагаются номера полигонов. Один из полигонов в этом буфере должен определить цвет пикселей всего тайла. Очевидно, что этим полигоном должен быть видимый полигон. Но как же компьютер узнает, какой полигон видимый, а какой нет? Для этого разработан специальный алгоритм. У каждой компании, выпускающей продукты на базе тайлового рендеринга, он может быть своим. Однако в общих чертах его можно описать так: из тайлового буфера считывается номер первого полигона, по этому номеру из всех



▲ Рис. 2. Краткая схема тайлового рендеринга

» других буферов считывается информация о данном полигоне. На базе этой информации мы можем вычислить значение удаленности пикселей от виртуальной камеры (Z-value) для данного полигона в данном тайле. Затем из всех значений Z-value получается одно усредненное значение, которое записывается в память и помечается как Z-value первого полигона. Все то же самое повторяется и со вторым полигоном из этого же тайла, кроме операции записи значения Z-value, так как оно сразу же сравнивается со значением Z-value первого полигона. Если новое значение Z-value меньше того, что уже хранится в памяти (второй полигон ближе, чем первый), то старое значение затирается и на его место записывается полученное. Если же новое значение Z-value больше того, что уже хранится в памяти (второй полигон дальше, чем первый), то в памяти остается первое значение Z-value. Этот процесс будет повторяться до тех пор, пока все полигоны тайла не пройдут проверку и в нем не останется один, видимый глазу играющего и определяющий цвет пикселей тайла. Поскольку размер тайла довольно небольшой (например, PowerVR в серии чипов PCX2 использует в качестве тайла прямоугольник со сторонами $32 \times 16 = 512$ пикселей), все эти вычисления происходят довольно быстро.

Следующая стадия — стадия рендеринга. На этом этапе используется информация, полученная на предыдущих этапах.

Прорисовываются только видимые пиксели.

Наконец, на последней стадии тайл перемещается в кадровый буфер. Так, по тайлам, в нем собирается конечный кадр.

Тайловые архитектуры

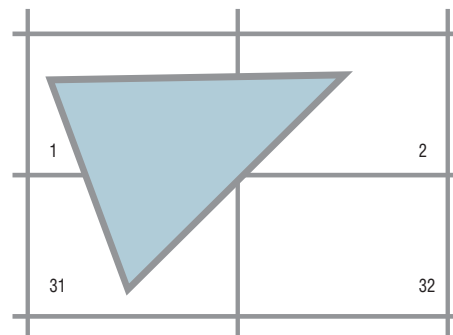
За время развития 3D-ускорителей были разработаны несколько архитектур тайлового рендеринга. Наиболее известными из них являются следующие.

Talisman

Talisman — разработка компании Microsoft. Она была начата еще в начале 90-х годов. Данная архитектура предполагала разбиение изображения на тайлы размером 32×32 пикселя. Также для ее реализации был разработан специальный процессор — Polygon Object Processor. Таким образом, требуемая пропускная способность шины памяти в разрешении 1024×768 должна была составить всего 220 Мбайт/с. К сожалению, судьба данной архитектуры нам не известна. В любом случае, если дальнейшая разработка уже не ведется, то на данный момент она устарела.

Giga3D

Giga3D — разработка компании GigaPixel, основанной в 1997 году. Как вы знаете, компанию GigaPixel купила 3Dfx. А 3Dfx была куплена компанией NVIDIA. Так что не удивляйтесь, если GeForce 5 будет базироваться на тайловом рендеринге. На базе



▲ Рис. 3. Полигон расположен в нескольких тайлах. Прямоугольники — это тайлы, цифры — номера тайлов

данной тайловой архитектуры было выпущено несколько чипов: GP-1, GP-2, GP-3 и GP-LP для ноутбуков. Да, должен был быть еще один чип — GP-4, но его разработка так и не была завершена. В качестве тайла используется квадрат размером 32×32 пикселя. Примечательной особенностью архитектуры Giga3D является практически «бесплатное» полноэкранное сглаживание.

Чип GP-1 обладает следующими характеристиками:

- ▶ частота ядра — 183 МГц;
- ▶ Pixel Fillrate — 732 млн пикселей/с, 4 конвейера рендеринга;
- ▶ Texel Fillrate — 6 млрд текселей/с, 8 блоков текстурирования на каждом конвейере;
- ▶ возможность наложения до 8 текстур за один проход;

»



Разработчики чипов и приложений

Мысли о тайлах

Imagination Technologies: «Тайловые архитектуры, безусловно, будут очень эффективны при работе со сценами, содержащими большое количество полигонов. В настоящий момент мы совершенствуем наши разработки в этой области и уже запатентовали достаточно большое число технологий».

NVIDIA: «В будущем, безусловно, число полигонов в сцене будет расти. Также увеличится и число вычислений. Поэтому будущее за тайловыми архитектурами. Хотя уже сейчас созданы довольно неплохие технологии и для традиционных архитектур, которые позволяют снизить overdraw. В будущем они смогут полностью его устранить».

ATI: «Тайловый рендеринг — один из путей снижения overdraw. В будущем он будет работать быстрее, чем традиционный. Вот почему мы уже сейчас начали вносить некоторые элементы тайлового рендеринга в наши продукты».

Matrox Graphics: «Тайловый рендеринг действительно представляет достаточно большой интерес как для разработчиков, так и для пользователей. Мы занимались «продвижением» на рынок карты Matrox m3D (на базе тайлового рендеринга), поэтому можно с уверенностью сказать, что тайловый рендеринг зачастую оказывается быстрее традиционного. Вдобавок к этому карты на базе тайловой архитектуры стоят не очень дорого».

Croteam (разработчики Serious Sam): «В настоящее время тайловый рендеринг является более предпочтительным, нежели традиционный рендеринг. Мы не можем сказать, что так будет и дальше. В любом случае нас ожидает один из следующих путей развития: либо разработчики полностью примут тайловый рендеринг и оптимизируют под него движки, либо мы останемся при традиционных способах рендеринга».

Epic Games (разработчики Unreal): «Мы не считаем, что тайловый рендеринг является будущим графических архитектур. Ведь на рынке тайловых ускорителей еще не было ни одного продукта, у которого бы не было определенного рода недоработок, недостатков и ошибок».



Полноэкранное сглаживание

Миф, старый как тайл

Миф о том, что тайловый рендеринг позволяет получать полноэкранное сглаживание (FSAA) без падения производительности, родился вместе с тайловой архитектурой. Для традиционного рендеринга FSAA — самая ресурсоемкая технология. При работе с высоким разрешением объем необходимых вычислительных операций увеличивается практически в геометрической прогрессии. Поэтому наблюдается значительное падение производительности. На тайловых системах при достаточном потенциале GPU полно-

экранное сглаживание окажется практически «бесплатным», так как оно будет применяться не ко всему кадру, а к каждому тайлу, благодаря чему обработка будет происходить несколько быстрее, чем в случае с традиционными системами. Однако пока все усилия создать чип, который сможет без падения производительности справляться с FSAA, не увенчались успехом. Кстати, именно эта проблема толкает разработчиков на создание новых методов FSAA, таких как Quincunx или Accuview от NVIDIA.

- » в данном чипе используются 48 запатентованных технологий;
- » число транзисторов — около 1 млн;
- » 128-битная шина обмена данными.

Довольно впечатляющие характеристики, если учесть, что он появился еще во времена GeForce 2 GTS. Особенно поражает число используемых транзисторов.

PowerVR

PowerVR — архитектура одноименной компании, получившая достаточно широкое распространение среди компьютеров класса desktop. В качестве тайла используется прямоугольник с размером 32x16 пикселей. Ничем особым среди других тайловых архитектур не выделяется. Наиболее известные чипы данной компании — Куро и Куро II. Вскоре появится и Куро III.

Другие

Компания Oak Technologies также разработала в свое время свою тайловую архитектуру. Наиболее известный чип на базе этой архитектуры — WARP 5. Однако 31 марта 1998 года эта компания прекратила разработку графических чипов.

Примечательно, что на графическом рынке вскоре родится новая тайловая архитектура разработки S3Graphics, которая будет использоваться в чипе Columbia.

Тайловая архитектура в Куро II

Рассмотрим тайловую архитектуру на примере видеокарты с чипом Куро II. Сравним его с GeForce 2 MX и GeForce 2 GTS. Как видно из табл. 2, Куро II уступает по техническим параметрам и GeForce 2 MX 400, и тем более Radeon 7500. Казалось бы, она дол-

жна неминуемо проиграть более именитым соперникам. Давайте обратимся к тестам.

Как показали результаты тестов, приведенные ниже, на практике действительно имеет место превосходство тайлового рендеринга над традиционным. Особенно это проявляется в тяжелых условиях — при включенном FSAA в высоких разрешениях. Недаром в тестовом пакете на базе движка Unreal 2 данная карта держится середнячком среди таких именитых соперников, как серия GeForce 2, серия GeForce 3, Radeon 8500 и 7500, зачастую обгоняя даже GeForce 2 Pro. Заметьте, что у Куро II достаточно низкое падение производительности в зависимости от разрешения и глубины цвета по сравнению со многими картами для DirectX 7. Однако не все так гладко: у Куро II достаточно большое падение производительности при активации трилинейной фильтрации. В некоторых играх тайловый рендеринг заметно сдает позиции, взять хотя бы тот же Giants, где Куро II проигрывает GeForce 2 MX 400 почти в два раза.

Однако в целом картина довольно радужная. У тайлового рендеринга, безусловно, есть будущее. В играх следующего поколения с возрастающим overdraw — вот где себя сможет проявить тайловый рендеринг. Конечно, разработчикам придется несладко — ведь им предстоит подружить блок T&L с тайлами, добиться максимальной оптимизации своих архитектур. Но Unreal 2 уже не за горами, и картам традиционной архитектуры (Radeon 8500, GeForce 3 Ti 500, GeForce 4 Ti 4600) в этих играх придется несладко.

■ ■ ■ Алексей Мирошниченко



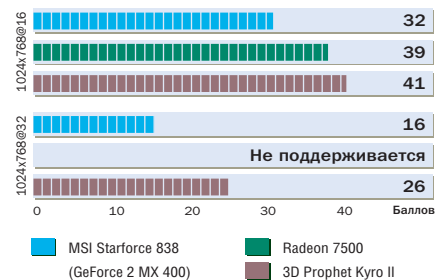
	MSI Starforce 838	Radeon 7500	3D Prophet Kyro II
Чип	GeForce 2 MX 400	Radeon 7500	Kyro II
Частота ядра, МГц	200	270	175
Частота памяти в пересчете на SDR (DDR=SDRx2), МГц	200	230	175
Тип памяти	DDR SDRAM	DDR SDRAM	SDR SDRAM
Число конвейеров рендеринга	2	2	2
Число блоков текстурирования на каждом	2	3	1
Аппаратный блок T&L	есть	есть	отсутствует
Число транзисторов, млн	25	30	15

▲ Табл. 2. Технические данные исследуемых видеокарт

Quake 3 Arena, High Quality, demo1



Quake 3 Arena, HQ, demo1, FSAA 4x



▲ Тестирование в Quake 3 Arena