

УДК 004.921

АНАЛІЗ АРХІТЕКТУР ГРАФІЧНИХ ВІДЕОКАРТ

Романюк О.Н., Обідник Д.Т., Поліщук О.В., Величко П.О.

Вінницький національний технічний університет

Розглянуто найпоширеніші архітектури відеокарт.

В останні роки істотно розширилося застосування засобів комп'ютерної графіки в найрізноманітніших галузях діяльності людини. У сучасному світі, де роль інформації стає все більш важливою, цілком природно, що просторові зображення, що мають високий ступінь інформативності й реалістичності, перетворюються в її основний носій. При формуванні таких зображень необхідно відображати графічні сцени великого обсягу та детальності, що обумовлює необхідність використання апаратних засобів для генерації зображень, зокрема, відеокарт.

Під відеокартою (GPU – Graphics processing unit) розуміють програмно-апаратну систему генерації та виведення на екран графічних зображень, яка містить графічний процесор або серію графічних процесорів [1, 2, 3, 5]. Відеокарта відповідає за геометричні перетворення тривимірної сцени, растеризацію та передачу оброблених даних у вигляді зображення на екран.

Серед видів відеокарт можна виділити окремі відеокарти (Dedicated graphics cards), вбудовані відеокарти (IGP – Integrated graphics solutions), гібридні процесори (APU – Accelerated Processing Unit), графічні процесори загального призначення (GPGPU – General Purpose GPUs).

Сучасні окремі відеокарти використовуються для розрахунку та виведення графічної інформації у графічних станціях, персональних комп'ютерах, ноутбуках тощо. Вони підключаються до материнської плати за допомогою шин PCI Express або Accelerated Graphics Port і, як правило, мають свою власну відеопам'ять. Основні вимоги до їх продуктивності визначаються потребами у сфері конструювання та ігровій індустрії. Висока швидкодія відеокарт забезпечується, в основному, за рахунок розпаралелення обчислень. Сучасні професійні відеокарти містять 2048–3072 потокових процесорів, кількість яких щороку зростає [4, 3, 5]. Вони забезпечують високу гнучкість програмування певних етапів графічного конвеєра.

Вбудовані відеокарти, гібридні процесори та графічні процесори загального призначення поступаються за швидкістю окремим відеокартам [4]. Вони застосовуються у цифрових пристроях

невеликих розмірів, наприклад, ноутбуках, нетбуках, мобільних телефонах та планшетах тощо.

Сучасні відеокарти розвиваються у напрямку підвищення продуктивності за рахунок впровадження нових архітектурних рішень, які забезпечують все вищий рівень розпаралелення обчислень [6, 7]. Тільки професійні та окремі ігрові відеокарти здатні обробляти обсяги інформації, необхідні для динамічного відтворення фотореалістичних зображень високополігональних тривимірних сцен. Разом з тим, нові архітектури відеокарт дозволяють застосовувати нові підходи та методи побудови зображень, направлені на обробку саме високодеталізованої комп'ютерної графіки.

Головним блоком кожної відеокарти є графічний процесор або серія таких процесорів. Основними компаніями-виробниками графічних кристалів є AMD (Advanced Micro Devices) і NVIDIA Corporation, які є основними конкурентами на ринку графічних процесорів для відеокарт. Конкуренцію цим виробникам на ринку професійної комп'ютерної графіки створює компанія Intel Corporation, яка розробила новий центральний процесор, що має графічне ядро (22-нанометрові гібридні процесори Ivy Bridge Core i3, i5 та i7, загальну структуру яких наведено на рисунку 1) Процесор Intel Ivy Bridge Core складається з 2–8 обчислювальних ядер, графічного ядра та контролера, який забезпечує з'єднання між процесором та інтерфейсом материнської плати. Ці складові з'єднані між собою за допомогою кільцевої шини. Така структура процесора як графічної системи забезпечує швидку передачу даних від обчислювальних ядер до графічного. Проте, на відміну від відеокарт, графічне ядро не має своєї власної відеопам'яті та використовує оперативну пам'ять комп'ютера. Такі процесори економні з фінансової точки зору, проте вони не можуть забезпечити формування графічних зображень високодеталізованих тривимірних сцен (рис 1).

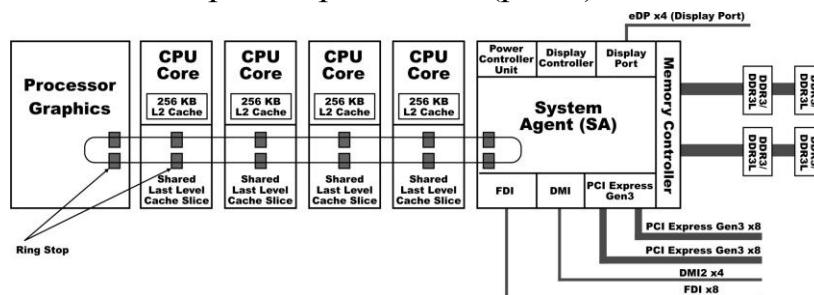


Рис. 1. Структура гібридних процесорів Intel Ivy Bridge Core

Типову структуру передових відеокарт можна розглянути на прикладі графічного процесора NVIDIA Kepler GK110, реалізованого

на 28 нм технології виробництва. Повна структура процесора Kepler GK110 містить 15 мультипроцесорних блоків (SMX) і шість 64-бітових контролерів пам'яті (Memory Controller) (рис 2). Мультипроцесорні блоки SMX з'єднані з кешом другого рівня (L2 Cache), що забезпечує швидкий обмін інформацією з контролерами, які розподіляють інформацію для розпаралелення обчислень.

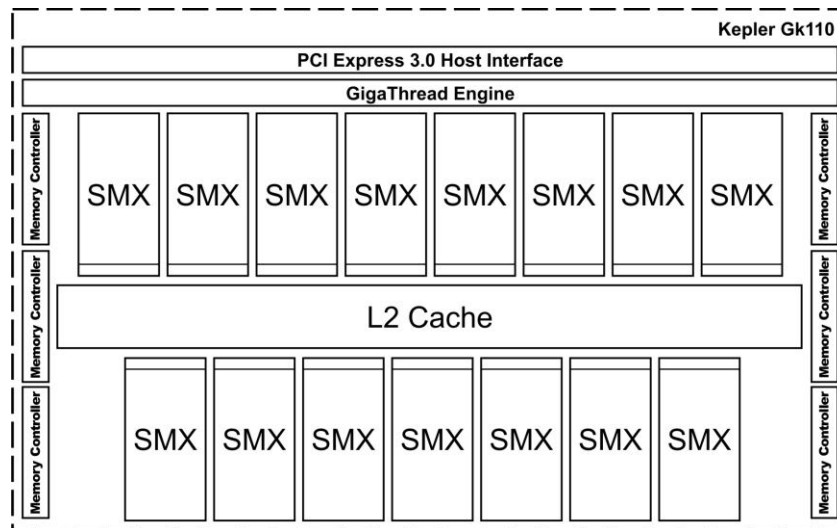


Рис. 2. Повна структурна схема мікропроцесора Kepler GK110

Структуру мультипроцесорного блока SMX наведено на рисунку 3. Кожен мультипроцесорний блок SMX містить 192 програмованих ядра CUDA (Core), 64 блоки подвійної точності (DP Unit), 32 спеціальних функціональних блоки (SFU), 32 блоки введення-виведення (LT/ST), а також 16 текстурних блоків (Tex). Кожне ядро CUDA (Core) може виконувати арифметичні операції як над цілими, так і над числами з плаваючою точкою, забезпечуючи при цьому можливість розрахунку кожного з етапів графічного конвеєра в одному ядрі. Така структура графічного процесора Kepler GK110 орієнтована на формування графічних зображень виокремлених тривимірних сцен, оскільки забезпечує можливість паралельної обробки 2880 примітивів одночасно. Таким чином графічний процесор Kepler GK110 (рис 3) дає можливість програмування не тільки шейдерних етапів графічного конвеєра, а й етапів тесселяції (додаткової тріангуляції) та текстурування, забезпечуючи при цьому високий рівень розпаралелення обчислень.

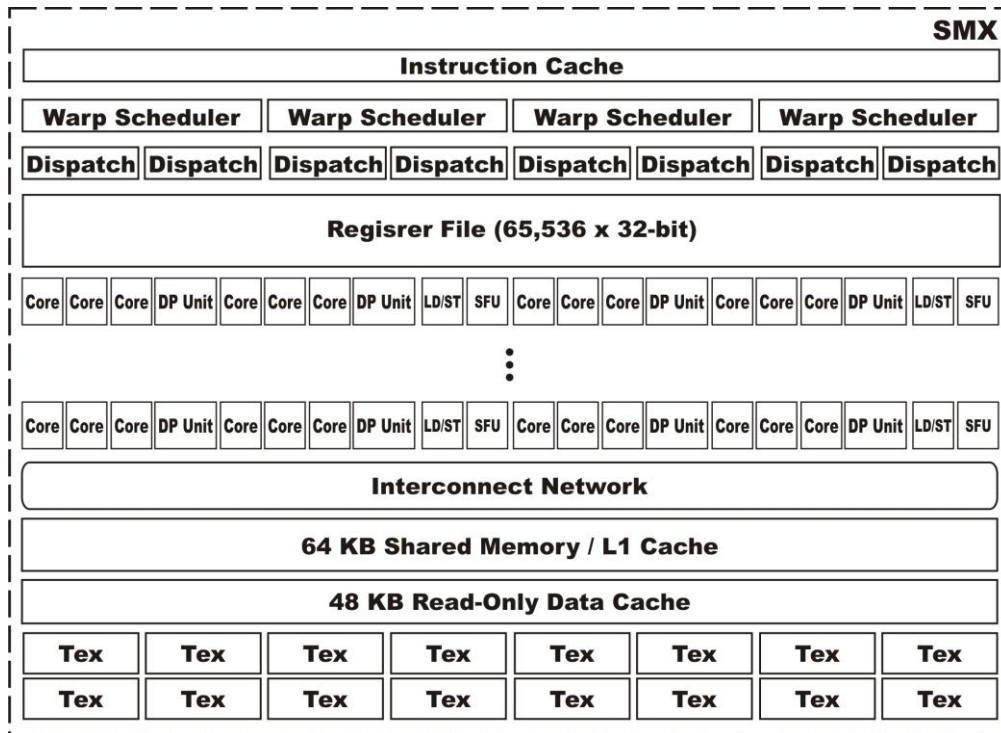


Рис. 3. Структура мультипроцесорного блока SMX

Архітектура Kepler GK110 підтримує динамічний паралелізм, який забезпечує паралельний доступ центрального процесора до кожного з мультипроцесорних блоків SMX. У кожному мультипроцесорному блоці SMX наявні спеціальні 4 блоки Warp Scheduler, які формують 32 паралельних потоки. При цьому кожне з ядер CUDA (Core) може бути паралельно завантажено розрахунками завдяки системі паралельного доступу Hyper-Q, а також має доступ до відеопам'яті через спеціальну дворівневу систему кешу.

Для покращення продуктивності графічного прискорювача Kepler GK110 підтримує команди перемішування (Shuffle Instructions), що виконуються на рівні регістрів і не потребують завантаження процесорів. Також Kepler GK110 забезпечує виконання атомарних операцій (Atomic Operations) типу додавання, порівняння та вибору. Ці операції виконуються без переривання іншими потоками. Вони використовуються для паралельного сортування, зниження кількості обчислень і побудови паралельних структур без блокування.

Графічний процесор Kepler GK110 містить спеціальні апаратні блоки текстурування, що підтримують спеціальні процедури MIP-mapping і фільтрації. При цьому обмеження у кількості текстур сцени немає і вони можуть динамічно завантажуватись у відеопам'ять.

Для обробки значного обсягу даних Kepler GK110 підтримує технологію NVIDIA GPUDirect, яка дозволяє об'єднувати декілька

графічних прискорювачів не тільки через шину PCI Express, а й через мережне з'єднання, забезпечуючи при цьому прямий доступ до відеопам'яті та ефективні процедури обміну даними між графічними прискорювачами.

Структура сучасних графічних процесорів, проаналізована на прикладі Kepler GK110, орієнтована на формування графічних зображень високодеталізованих тривимірних сцен. Вона забезпечує можливість програмування шейдерних етапів графічного конвеєра, етапів тесселяції та текстурування, що дає можливість впроваджувати власні методи геометричних перетворень полігональних моделей, методів інтерполяції та зафарбовування тощо.

На базі архітектури Kepler побудована GPU NVIDIA GK104 (рис 4), яка, у свою чергу, є основним блоком відеокарти NVIDIA GTX 680.

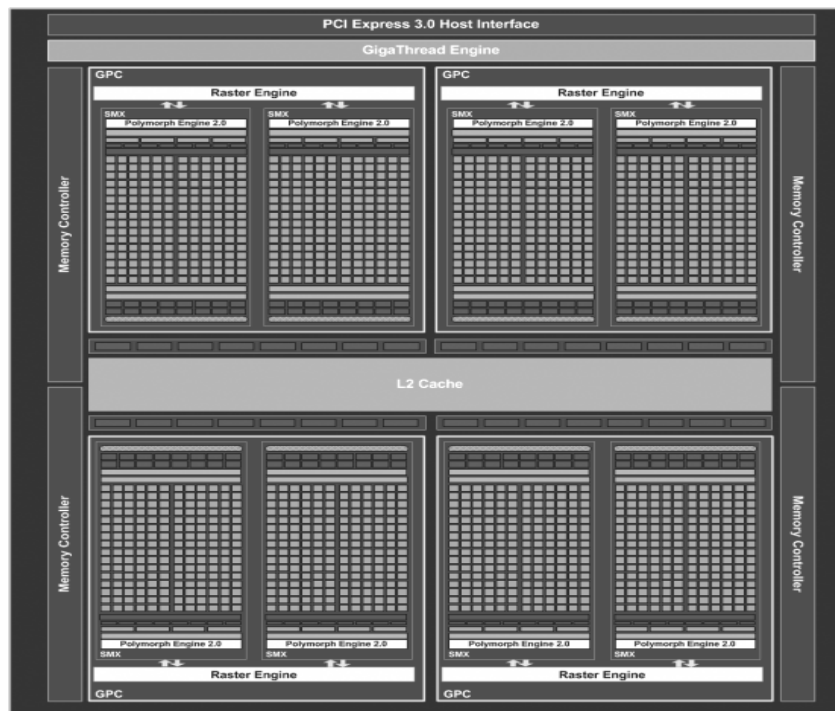


Рис. 4. Структурна схема мікропроцесора Kepler

Kepler є розвитком архітектури Fermi. Головна відмінність полягає в зменшеному с 40 до 28 нм техпроцесі. Між ними дійсно багато спільного. Kepler має чотири кластери обробки графіки (Graphics Processing Clusters), в кожному з яких по два SMX (Next Generation Streaming Multiprocessors), вони ж є потоковими процесорами. У загальній складності виходить вісім SMX.

Кожен поточковий мультипроцесор має 192 CUDA-ядра. Нагадаємо, що в поколінні Fermi один мультипроцесор вмщував в себе тільки 32 Cuda-ядра. Також, кожен SMX несе в собі текстурні

блоки і двигун PolyMorph (PolyMorph Engine) версії 2.0. Пропускна здатність зросла до 8 GT / s (Гігатранзакцій в секунду). Площа кристала GTX 580 знизилася з 520 до 294 мм. Кількість транзисторів зросла з 3000 до 3500 млн.. Кількість блоків растеризації (ROPs) зменшилося з 48 до 32 щодо GTX 580. Загальний обсяг відеопам'яті становить 2048 Мбайт. На сьогоднішній день NVIDIA GTX 680 - єдина відеокарта, пам'ять якої по дефолту працює на таких частотах. Частота відеопам'яті разом з шиною пам'яті дають загальну пропускну здатність 192,3 Гб / с. Тактова частота самого GPU становить 1006 МГц.

Технологія Turbo Boost дозволяє збільшувати тактову частоту роботи ядер вище номінальної, якщо при цьому не перевищуються встановлені норми енергоспоживання і TDP. Це дуже зручно, так як процесор сам починає працювати на підвищених частотах без втручання користувача. Змінилася система визначення тактових частот графічного процесора, тому NVIDIA відмовилася від подальшого використання концепції Hotclocks.

Архітектура Fermi передбачає (рис 5), що обробка комп'ютерної графіки більше не є єдиним завданням графічних процесорів, хоча і залишається одним з пріоритетних напрямів.

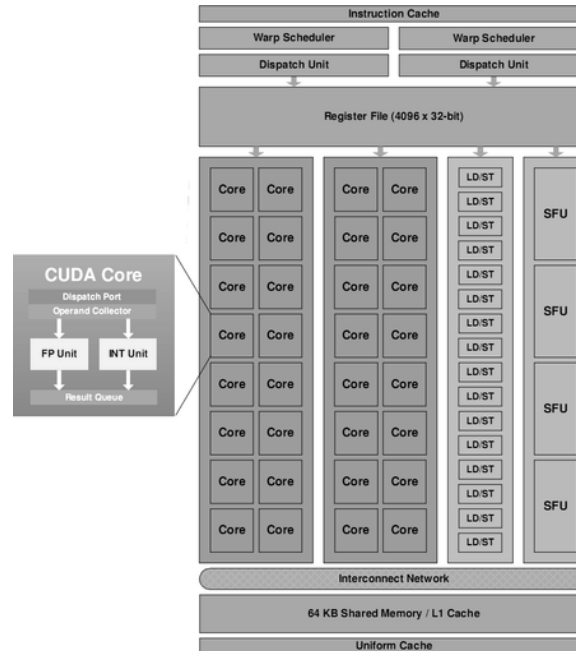


Рис. 5. Структурна схема Fermi

NVIDIA позиціонує нову архітектуру переважно на ринок суперкомп'ютерів та інших високопродуктивних розрахункових рішень, що передбачає як високу швидкість розрахункових операцій,

так і високу надійність разом з високою зручністю програмування. Для цього ринку ключовою вимогою є підтримка речових обчислень подвійної точності (з плаваючою комою) і механізмів знаходження та корекції помилок (ECC) в оперативній пам'яті та підсистемах кеш-пам'яті для підвищеної відмовостійкості.

Fermi складається з 512 шейдерних конвеєрів, які згруповані в 16 кластерів. Кожен кластер складається з 2 субкластерів по 16 конвеєрів кожен, а також 2 локальних планувальників завдань, 2 блоків вибірки, 4 блоків спецфункцій, 16 блоків навантаження. На кожен кластер доводиться 64Кб вбудованої пам'яті, яка повинна бути поділена між локальною пам'яттю і кеш-пам'яттю 1-го рівня. Передбачається виділення 16 Кб під локальну пам'ять і 48 Кб під кеш-пам'ять або навпаки. Отже, корпорація NVIDIA створивши архітектури Fermi і Kepler зробила революцію у світі графічних відеоадаптерів.

Література

1. Башков Е. А. Современное алгоритмическое и аппаратное обеспечение виртуальных систем трехмерного моделирования окружающей обстановки / Е. А. Башков, С. А. Зори, С. В. Ковальский // Наукові праці ДонНТУ. Серія «Інформатика, кібернетика та обчислювальна техніка». – Донецьк : ДонНТУ, 2009. – Вип. 10 (153). – С.81–91. – ISSN 1996-1588.
2. Горбань П. NVIDIA выпускает графическую демонстрацию Kepler [Электронный ресурс] / П. Горбань – 2012. – Режим доступа : <http://nv-world.ru/news/nvidia-release-new-dawn-demo/>.
3. AMD Radeon HD 7770 и HD 7750: недорогие видеокарты на архитектуре Graphics Core Next [Электронный ресурс] /Tom's Hardware Guide, 2012. – Режим доступа : http://www.thg.ru/graphic/obzor_amd_hd7770_i_hd7750/print.html.
4. Мурки Т. Устройство процессоров Intel Ivy Bridge [Электронный ресурс] / Т. Мурки – 2012. – Режим доступа : <http://www.ixbt.com/cpu/ivy-bridge-architecture-1.shtml>
5. Lefohn Aaron. GPGPU : General-Purpose Computation on Graphics Processors / Aaron Lefohn, I. Buck, P. McCormick, J. Owens, P. Tim, R. Strzodka // Tutorial in IEEE Visualization 2005. – 535 p.
6. Шиллинг А. Архитектуры высокопроизводительных графических систем / А. Шиллинг, В. Штрассер, Г. Книттель // Тюбингенский университет, «Открытые системы», Германия. – электронное издание № 5. – Режим доступа : <http://www.osp.ru/os/1995/05/178728/>.
7. Bao H., Hua W. Real-Time Graphics Rendering Engine / Hujun Bao, Wei Hua // Springer, 2011. – 300 p. – ISSN 1995-6819.

Надійшла до редколегії 11.06.2013