

УДК 004. 925.9

АНАЛІЗ ГРАФІЧНОЇ АХІТЕКТУРИ NVIDIA PASCAL

Романюк О. Н, Романюк О. В., Войт Б. Л.
Вінницький національний технічний університет, Україна

Проаналізовано особливості архітектури NVIDIA PASCAL. Розглянуто організацію інтерфейсу, багатошарової пам'яті, технологію багатобуферного формування зображень, технологію мультипроекування зображень.

Графічні процесори забезпечують формування реалістичних графічних сцен з високою продуктивністю [1].

Графічні процесори NVIDIA нового покоління відомі під кодовим ім'ям Pascal [2-4] і оприлюднені у 2016 році. Відеокarti на базі архітектури Pascal забезпечують високу продуктивність і енергоефективність. Вони створені з використанням надшвидкісної технології FinFET [5] і підтримують DirectX™ 12 [6] для швидкого, плавного та енергоефективного обчислювального процесу.

Володіючи 16 млрд. транзисторів, побудованих на базі 16-нм техпроцесу FinFET, графічний процесор Pascal є найбільшим в світі FinFET чіпом. Він створений для забезпечення високої продуктивності та енергоефективності для задач з практично необмеженими обчислювальними вимогами.

Однією з особливостей Pascal є потрійний буфер для виведення кадрів, що забезпечує одночасно наднизькі затримки в відображенні та забезпеченні вертикальної синхронізації. В одному буфері зберігається виведене зображення, в іншому - останній відформований кадр, у третьому - малюється поточний.

Чіпи на базі архітектури Pascal підтримують багатошарову пам'ять з високою пропускнуою здатністю (stacked DRAM за класифікацією NVIDIA, конкретний тип - SK Hynix high-bandwidth memory (HBM) [7] або Micron hybrid memory cube, уніфікований адресний простір пам'яті для CPU і GPU, шину NVLink для зв'язку з CPU і GPU у системах для суперкомп'ютерних обчислень, а також цілий ряд нових графічних, обчислювальних і мультимедійних функцій, які будуть підтримуватися DirectX 12, OpenGL 5 [8] і іншими майбутніми програмними інтерфейсами.

NVIDIA реалізувала об'єднання процесора і GPU, Вони зможуть користуватися ресурсами один одного. Це дозволить краще розподіляти навантаження в DX12- середовищі.

Pascal - це перша архітектура, оснащена високошвидкісним двонаправленим зв'язком між GPU і CPU.

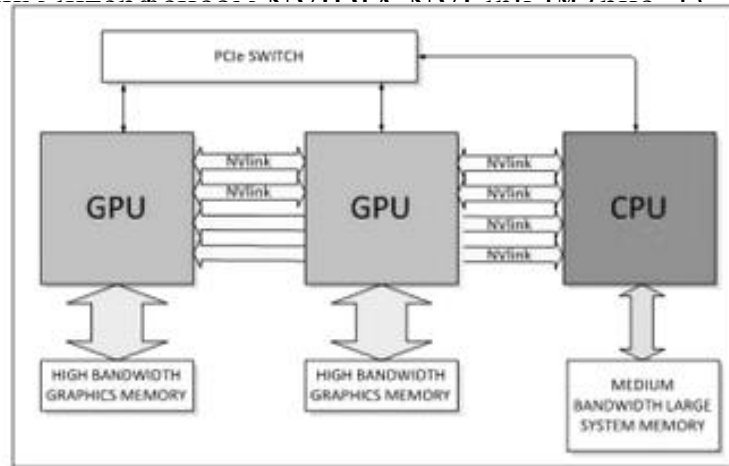


Рисунок 1 - Інтерфейс NVIDIA

Ця технологія дозволяє масштабувати додатки на кількох GPU, забезпечуючи п'ятикратний приріст продуктивності порівняно з кращими на сьогодні рішеннями в своєму класі.

Технологія 3D Memory або Stacked Memory усуває недоліки зв'язку між GPU і відеопам'яттю. Інтерфейс NVLink покликаний збільшити ефективність з'єднання між GPU і CPU, а також між декількома GPU.

У даному відношенні на першому місці знаходиться пропускна здатність, яка сьогодні забезпечується поширеним інтерфейсом PCI Express. 16 ліній PCI Express 3.0 забезпечують пропускну здатність 15,75 Гбайт/с або 128 GT/s. NVIDIA, наприклад, з архітектурою "Maxwell" додано технологію стиснення пам'яті, яка покликана задовольнити зростаючі вимоги до пропускну здатності пам'яті.

NVIDIA використовує для NVLink з'єднання «точка-точка». Одне з'єднання NVLink охоплює вісім ліній. У Pascal буде використовуватися чотири з'єднання NVLink. За інформацією NVIDIA, кількість з'єднань може змінюватися залежно від цільового ринку. З'єднання NVLink можна гнучко комбінувати, якщо того вимагають сценарії використання.

Основа для NVLink - швидкодіючий, 8-смуговий, подвійний двонаправлений зв'язок. Pascal GPUs підтримує багато цих зв'язків, забезпечуючи гнучкість конфігурації. Відеокарти можуть сполучатися разом (або відеокарта і процесор), щоб сформувати єдиний зв'язок

GPU↔CPU або використовуватися індивідуально, щоб створити мережу GPU↔CPU або зв'язок GPU↔GPU, що дозволяє використовувати обчислювальні елементи разом, для спільного вирішення різного роду завдань. Якщо центральний процесор не підтримує функцію NVLink, то можна просто об'єднувати GPU.

NVLink більш ніж в два рази ефективніше PCIe 3.0 [9].

Архітектура Pascal задовольняє вимоги дисплеїв нового покоління, включаючи VR-дисплеї і дисплеї ультрависокої роздільної

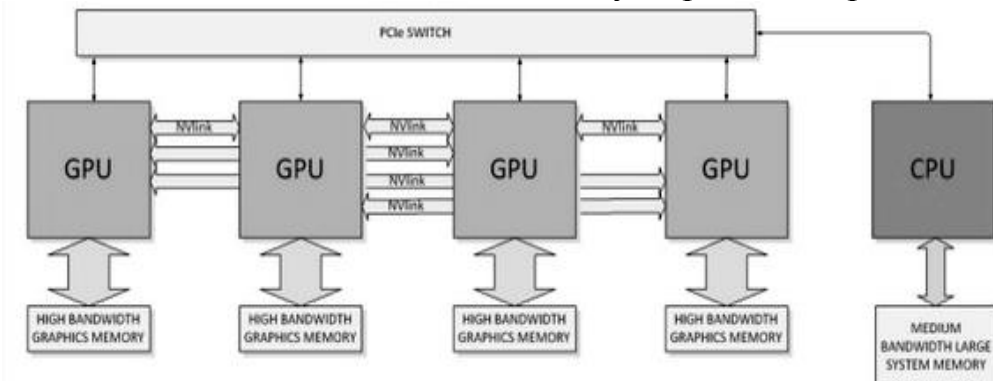


Рисунок 2 – Шина NVLink

здатності, підтримуючи підключення декількох моніторів. Вона оснащена технологіями NVIDIA GameWorks™, які забезпечують плавний і кінематографічний ігровий процес. Крім того, вона підтримує революційну технологію для запису зображення з охопленням в 360 градусів.

У Pascal модифікували ту частину, яка відповідає за геометричну складову картинки. Це необхідно для мультимоніторних конфігурацій і роботи з VR-шоломами: при належній підтримці з боку ігрового двигуна – відеокарта може один раз порахувати геометрію і отримати кілька проєкцій геометрії для кожного з екранів. Це істотно знижує навантаження в VR не тільки при роботі з трикутниками (тут досягається дворазовий приріст), але і в роботі з піксельною складовою.

У Pascal з'явилася апаратна підтримка технології мультипроєктування, яка дозволяє працювати одночасно з різними проєкціями зображення (рис. 3).

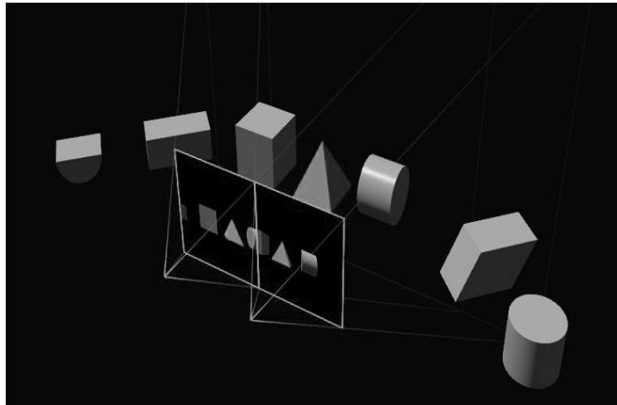


Рисунок 3. – Приклад проекції зображень з використанням технології мультипроектування

Спеціальний блок Simultaneous Multi-Projection всередині PolyMorph Engine відповідає за формування різних проекцій при обробці одного потоку геометрії. Цей блок обробляє геометрію одночасно для 16 проекцій з одним або двома центрами перспективи. Це не вимагає повторної обробки геометрії і дозволяє реплікувати дані до 32 разів (16 проекцій на дві точки).

Рішення, закладені в Kepler і продовжені в поколінні Maxwell, залишилися і в Pascal: перші споживчі відеокарти випущені на базі чіпа GP104, який складається з чотирьох кластерів обробки графіки, повнорозмірний, шестикластерний GP100 під маркою TITAN X [10].

З моменту появи GDDR5 [11] компанія NVIDIA працювала над наступним поколінням швидкісної пам'яті. Результатом взаємодії з розробниками пам'яті стала поява GDDR5X зі швидкістю передачі даних 10 Гбіт/с. Робота з настільки швидкою пам'яттю висуває нові вимоги до розведення електричних ланцюгів. Тому були перероблені лінії передачі даних між GPU і мікросхемами пам'яті, змінено структуру самого чіпа. Все це дозволяє ефективно працювати з надшвидкісним відеобуфером. Серед переваг GDDR5X-більш низька робоча напруга на рівні 1,35 В.

GeForce підтримують спеціальні алгоритми стиснення даних у пам'яті, що дозволяє більш ефективно використовувати кеш і передавати більше даних при тій же пропускній здатності. Підтримується кілька методів, залежно від типу даних вибирається свій алгоритм стиснення. Важливу роль відіграє алгоритм стиснення кольору delta color compression. Завдяки йому кодується не колір

кожного окремого пікселя, а різниця між пікселями при послідовній передачі даних. Обчислюється усереднений колір тайла і дані про усунення кольору для кожного пікселя цього тайла.

ЛІТЕРАТУРА

1. Романюк О.Н. Класифікація графічних відеоадаптерів / О. Н. Романюк, Р. Ю. Довгальок, С. В. Олійник // Наукові праці Донецького національного технічного університету .Серія Інформатика, кібернетика та обчислювальна техніка”. Випуск 14(188). 2011. С. — 211-215.
2. Nvidia Pascal Architecture Overview [Електронний ресурс] Режим доступу: <https://www.notebookcheck.net/Nvidia-Pascal-Architecture-Overview.165493.0.html>.
3. Архитектура NVIDIA PASCAL [Електронний ресурс] Режим доступу: <http://www.nvidia.com.ua/object/gpu-architecture-ru.html>.
4. Nvidia Pascal: обзор архитектуры [Електронний ресурс] Режим доступу: <https://occlub.ru/testodrom-mobile/2342-nvidia-pascal-obzor-arkhitektury?showall=1>.
5. FinFET [Електронний ресурс] Режим доступу: https://en.wikipedia.org/wiki/Multigate_device.
6. DirectX 12 [Електронний ресурс] Режим доступу: https://ru.wikipedia.org/wiki/Direct3D_12.
7. High Bandwidth Memory [Електронний ресурс] Режим доступу: https://ru.wikipedia.org/wiki/High_Bandwidth_Memory.
8. Open Graphics Library [Електронний ресурс] Режим доступу: <https://en.wikipedia.org/wiki/OpenGL>.
9. PCI Express [Електронний ресурс] Режим доступу: https://ru.wikipedia.org/wiki/PCI_Express.
10. NVIDIA TITAN X: большой Pascal [Електронний ресурс] Режим доступу: <https://3dnews.ru/939245>.
11. Graphics Double Data Rate/https: [Електронний ресурс] Режим доступу: <https://3dnews.ru/939245//ru.wikipedia.org/wiki/GDDR5>.

Отримано 29.06.2017