

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**ОДЕСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ ХАРЧОВИХ ТЕХНОЛОГІЙ**

**ХІ МІЖНАРОДНА  
НАУКОВО-ПРАКТИЧНА  
КОНФЕРЕНЦІЯ**

**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І  
АВТОМАТИЗАЦІЯ – 2018**

**Збірник доповідей**

**Частина II**

Одеса,  
4-5 жовтня 2018

## ЗМІСТ

|  |    |
|--|----|
| <i>МОРОЗ А. Н.</i>   | 3  |
| <i>НОЖКО Т.Г.</i>  | 4  |
| <i>УЕНОРОВ В.В., РОНЛЕВИНА Н.О.</i>                                | 6  |
| <i>РОМАНЮК О.Н., ЛИСЕНКО Є.С., ВОЙТ Б.Л.</i>                       | 7  |
| <i>РОМАНЮК С. О., НЕЧИПОРУК М. Л.</i>                              | 10 |
| <i>РОМАНЮК О. Н., ПАНФІЛОВА Ю. О., ЧАН А. Л. В.</i>                | 13 |
| <i>РИБАЛКО І. І., БОГДАНОВА Л. М., АНОСОВ В. Л.</i>                | 16 |
| <i>СКАКОВСЬКИЙ Ю.М., БАБКОВ А.В.</i>                               | 17 |
| <i>СТАНОВЬКА Т.П., СПРОМЛЯ С.Г., БОЛТАЧ С.В.</i>                   | 20 |
| <i>СУЛІМА Ю.Ю., СУЛІМА Ю.Є.</i>                                    | 22 |
| <i>ТРАЧ Н.Р., ВОЛКОВ В.Э.</i>                                      | 24 |
| <i>ЮРЧЕНКО В. В., БОГДАНОВА Л. М., АНОСОВ В. Л.</i>                | 25 |
| <i>УАНАКОВ В.Р.</i>  | 27 |
| <i>ГНАТЕНКО В.Ю., СТУПЕНЬ П.В.</i>                                 | 29 |
| <i>ЛЕОНТЬЄВА І.О., ХОБІН В.А.</i>                                  | 31 |
| <i>КОРНІЄНКО Ю.К., БОЙЦОВА О.С., ШАМРАЙ О.А.</i>                   | 33 |
| <i>КОРНІЄНКО Ю.К., КОТЛИК С.В., БОЙЦОВА О.С., ШАМРАЙ О.А.</i>      | 35 |
| <i>ІВАНОВА А.Г., ОЛЬШЕВСЬКА О.В.</i>                               | 38 |
| <i>ШЕРШУН О.О., ОЛЬШЕВСЬКА О.В.</i>                                | 40 |
| <i>ВОЛКОВА А.Ю., ПРУС В.В., ОЛЬШЕВСЬКА О.В.</i>                    | 42 |
| <i>ХАРАШ К.М., ОЛЬШЕВСЬКА О.В.</i>                                 | 43 |
| <i>БОГДАНОВ А.С., КОРНІЄНКО Ю.К.</i>                               | 45 |
| <i>СКАЛІЙ Д.О., ОЛЬШЕВСЬКА О.В.</i>                                | 47 |
| <i>ДЖИДЖУЛА М.В., КОРНІЄНКО Ю.К.</i>                               | 48 |
| <i>ЄПІФАНОВА А.О., КОРЖАН В.С., ОЛЬШЕВСЬКА О.В., ЛОМОВЦЕВ П.Б.</i> | 49 |

## АНАЛІЗ АРХІТЕКТУРИ VOLTA ВІДЕОКАРТ

*Проаналізовано архітектуру VOLTA відеокарт. Визначено особливості архітектури. Розглянуто побудову тензорних ядер, комбінованої чіпової пам'яті, управління потоками, оптимізованого набору інструкцій. Наведено технічні характеристики графічної архітектури NVIDIA Volta.*

Для побудови реалістичних зображень широко використовуються відеокарти. Сучасні відеокарти не обмежуються лише звичайним виведенням зображень, вони мають вбудовані процесори, які може здійснювати додаткову обробку, звільняючи від цих задач центральний процесор. Компанія NVIDIA розробила нову архітектуру Volta для побудови нового покоління відео карт високої потужності. Оскільки споживачам відео карт важливо знати особливості архітектури нового покоління, то в статті запропоновано аналіз нововведень

NVIDIA Volta — це одна з найпотужніших у світі архітектура графічних процесорів, побудований на основі досягнень у галузі штучного інтелекту та високопродуктивних розрахунків. Перший процесор на базі Volta став GPU для дата-центрів Tesla V100, що забезпечує високу швидкість і масштабування навчання та інферса глибинних нейронних мереж, а також прискорення високопродуктивних та графічних розрахунків.

Сьоме покоління графічних архітектур NVIDIA Volta використовує 21 млрд транзисторів, що забезпечують продуктивність глибокого навчання, що є рівноцінним для 100 центральних процесорів. Пікова продуктивність Volta в 5 раз вища потужності архітектури Pascal (поточна графічна архітектура NVIDIA), і в 15 разів вище Maxwell. Завдяки об'єднанню ядер CUDA і нової ядра Volta Tensor в уніфікованій архітектурі, один сервер на базі GPU Tesla V100 зможе замінити сотні центральних процесорів у високопродуктивних розрахунках.

NVLink піднімає високошвидкісний інтерфейс між графічними процесорами і між графічними та центральними процесорами на новий рівень, що вдвічі збільшує пропускну спроможність порівняно з попереднім поколінням NVLink. Пам'ять HBM2 DRAM зі швидкістю передачі даних 900 ГБ / с, збільшує пропускну здатність на 50% порівняно з попереднім поколінням. Оптимізоване під Volta програмне забезпечення, включаючи CUDA, cuDNN і TensorRT, доступне для сторонніх розробників.

Volta — одна із мікроархітектур графічного процесора для формування високореалістичних зображень. Це архітектура відеокарт, що використовує техніку трасування променів у реальному часі. Дана техніка являє собою одну з технік рендерингу зображень. Технологія побудови зображення тривимірних моделей за допомогою метода трасування променів в комп'ютерних програмах відстежує зворотню траєкторію поширення променів від камери до екрану. Як тільки промені світла перетинають якийсь об'єкт, відбувається перевірка властивостей об'єкта для того, щоб розрахувати точний колір кожного пікселя для його виведення на екран. Таким чином забезпечується більш висока реалістичність створюваних сцен. [1]

До основних особливостей графічного процесора Volta належать:

- наявність тензорних ядер зі змішаною точністю обчислень, призначених для матричних обчислень у задачах глибокого навчання (які спеціалізуються на операціях матричного множення великих матриць вхідних даних, за рахунок чого зростає ефективність [2]);
- кеш-пам'ять зі зниженими затримками доступу;
- оптимізований набір інструкцій для спрощення декодування та скорочення затримок в процесі виконання інструкцій;
- оптимізації з метою досягнення високої тактової частоти і кращої енергоефективності.

Щоб зростання продуктивності відео карт відповідала потребам ринку, компанія Nvidia впровадила в свій обчислювальний процесор Volta новий тип ядер — тензорні ядра (Tensor Core). Ці ядра — найважливіша особливість нової архітектури Volta, яка і допоможе отримати багаторазовий ріст продуктивності в задачах навчання і інференсу великих нейромереж.

Вісім тензорних ядер в кожному мультипроцесорі SM виконують в загальному тисячі двадцять чотири операції з плаваючою комою за такт, що у вісім разів швидше, ніж може забезпечити мультипроцесор архітектури Pascal з використанням стандартних FP32-операцій. В цілому, якщо порівнювати GV100 з обчислювальним процесором GP100, це призводить до збільшення продуктивності в задачах глибокого навчання в 12 разів.

Процесор включає 64 FP32-ядра та 32 FP64-ядра на кожен SM, які використовують схему розподілу ресурсів. Детальна схема SM зображена на рис. 1. Об'єднання кеша L1 для даних і роздільної пам'яті в один блок забезпечує обидва типи доступу до пам'яті. Сумарна ємність цих типів пам'яті становить 128 КБ на кожен процесор. Кеш першого рівня в архітектурі Volta працює як канал потокової передачі даних з високою пропускною здатністю, одночасно забезпечуючи доступ до часто використовуваних даних з низькою затримкою, поєднуючи ці дві характеристики.

У графічних процесорах Volta чіпи пам'яті розташовані в одному корпусі з кристалом GPU (багато кристальна компоновка). Дане розташування пам'яті і GPU дозволяє збільшити пропускну здатність з'єднання між ними до 1 ТБ / с. [3]

В кожному відділі є кеш інструкцій нульового рівня. Графічні процесори даної архітектури мають можливість конфігурувати обсяги пам'яті до 96 КБ на кожен SM. Роздільні FP32 і INT32 ядра дозволяють одночасно виконувати FP32- і INT32-інструкції. Незалежне управління (планування) потоками забезпечує більш точну синхронізацію і кращу взаємодію між паралельними потоками. Однією з головних цілей при розробці нової архітектури було скорочення часу розробки, необхідного для ефективної роботи програм на GPU, а також для забезпечення більшої гнучкості у взаємодії між потоками.

Одне з найважливіших змін у нових мультипроцесорів Volta полягає в тому, що в них входять розрізнені ядра FP32 і INT32, що дозволяють одночасно виконувати FP32- та INT32-інструкції на повній швидкості, що підвищує утилізацію блоків і загальну продуктивність графічного процесора. Декодер та планувальники запускають по одній інструкції і по одній варпі за кожним такт, і все виконується вдвічі більше вказівок і варпів на кожний мультипроцесор за такт, в порівнянні з Pascal.

Попередні сімейства GPU, включаючи Pascal, не можуть виконувати FP32 і INT32 інструкції одночасно, а тільки по черзі, що негативно позначається на продуктивності в деяких завданнях. Нова архітектура мультипроцесорів дозволяє запускати FP32-інструкції в повному темпі і використовувати половину, що залишилася слотів видач, і для виконання інших типів інструкцій: INT32, FP64, завантаження / збереження, розгалуження, спеціальних функцій SFU і т. Д., Підвищуючи ефективність використання обчислювальних можливостей GPU.

Моделі виконання SIMT (single instruction, multiple thread — одна інструкція і багато потоків) в попередніх GPU компанії і Volta дещо відрізняються. Модель SIMT в Volta дозволяє використовувати рівний паралелізм між всіма потоками, незалежно від варпа, підтримуючи стан виконання для кожного потоку, включаючи лічильник програми і стек викликів.

Обчислювальний процесор GV100 став першим GPU, який підтримує незалежне управління (планування) потоками, що забезпечує більш точну синхронізацію і кращу взаємодію між паралельними потоками. Однією з головних цілей при розробці нової архітектури було скорочення часу розробки, необхідного для ефективної роботи програм на GPU, а також для забезпечення більшої гнучкості у взаємодії між потоками.

Для того, аби підвищити ефективність паралельного виконання, в Volta є оптимізатор планування (Schedule optimizer), який групує активні потоки з одного і того ж варпа в блоки SIMT. Це зберігає високий темп виконання SIMT з великою гнучкістю: в Volta потоки можуть розходитися і сходитися в варпа, а GPU буде групувати потоки, які виконують один і той же код, і запускати їх паралельно.

Декодер і планувальники запускають по одній інструкції та по одному варпові на кожен процесор за такт. Водночас архітектура графічних процесорів дозволяє запускати FP32-інструкції в повному темпі та використовувати решту слотів видачі для інших типів інструкцій: FP64, INT32, завантаження та збереження, розгалуження, спеціальних функцій SFU тощо. [4]

В архітектурі Volta застосовується нова комбінована швидка чіпова пам'ять, що поєднує кеш-пам'ять першого рівня для даних і пам'ять, що розділяється. Така реалізація покращує продуктивність, спрощує складність програмування для деяких завдань і знижує необхідність в ручній оптимізації, необхідної для досягнення продуктивності, близькою до пікової.



Рисунок 1. — Схема SM

Проаналізовано архітектуру VOLTA відеокарт. Визначено особливості архітектури. Розглянуто побудову тензорних ядер, комбінованої націпової пам'яті, управління потоками, оптимізованого набору інструкцій. Наведено технічні характеристики графічної архітектури NVIDIA Volta.

#### СПИСОК ЛІТЕРАТУРИ

1. А. Берилло, “Nvidia Volta: новая вычислительная архитектура” *ИХВТ Новости, технологии, обзоры гаджетов, смартфонов, бытовой техники и автомобилей* [Електронний ресурс]. Режим доступу: <https://www.ixbt.com/video4/nvidia-volta.shtml#sm>.
2. А. Берилло, “Nvidia Volta: новая вычислительная архитектура. Тензорные ядра” *ИХВТ Новости, технологии, обзоры гаджетов, смартфонов, бытовой техники и автомобилей* [Електронний ресурс]. Режим доступу: <https://www.ixbt.com/video4/nvidia-volta.shtml#tensor>.
3. О. Н. Романюк, О. В. Даньковська, та С. І. Вяткін, “АНАЛІЗ АРХІТЕКТУР ВІДЕОКАРТ КОМПАНІЇ NVIDIA” *Збірник матеріалів Міжнародної науково-практичної Інтернет-конференції, Вінниця, грудень 2014 р.*, [Електронний ресурс]. Режим доступу: <https://is.gd/rZUfwb>.
4. В. Капрусь, “Видеокарты NVIDIA на базе архитектуры Volta получают поддержку технологии трассировки лучей в реальном времени.” *ИТСуа, 21 березня 2018 р.*, [Електронний ресурс]. Режим доступу: <https://itc.ua/news/videokartyi-nvidia-na-baze-arhitekturyi-volta-poluchat-podderzhku-tehnologii-trassirovki-luchey-v-realnom-vremeni/>.
5. “Стали известны спецификации игровой видеокарты NVIDIA GeForce GTX 1180 (Volta)” *ТЕХНОТ, 8 травня 2018 р.*, [Електронний ресурс]. Режим доступу: <https://tehnot.com/stali-izvestny-spetsifikatsii-igrovoj-videokarty-nvidia-geforce-gtx-1180-volta/amp/>.

**XI МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ**

**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І АВТОМАТИЗАЦІЯ – 2018**

*ОДЕСА*  
*4 – 5 ЖОВТНЯ, 2018*

Збірник включає доповіді учасників XI Міжнародної науково-практичної конференції «Інформаційні технології і автоматизація – 2018»

**Редакційна колегія:** Котлик С.В., Хобін В.А.

**Комп'ютерний набір і верстка:** Шамрай О.А.

**Відповідальний за випуск:** Котлик С.В.

НТТБ ОНАХТ

