

ВИКИПЕДИЯ

Volta (микроархитектура)

Материал из Википедии — свободной энциклопедии

Volta — микроархитектура профессиональных графических процессоров класса Server/Datacenter представленная в мае 2017 года, и разработанная корпорацией NVIDIA Corporation в качестве преемника микроархитектуры Pascal. Она названа в честь Алессандро Джузеппе Антонио Анастасио Вольта — итальянского физика, химика и физиолога, одного из основоположников учения об электричестве. О планах разработки этой микроархитектуры было объявлено ещё в марте 2013 года^[1]. Однако о первом продукте было объявлено только в мае 2017 года^[2]. Это первый чип NVIDIA, включающий специально разработанные тензорные ядра для машинного обучения систем искусственного интеллекта, в которых улучшены процессы глубокого обучения по сравнению с обычными ядрами CUDA.

Nvidia Volta	
Кодовое имя	GV100
Дата выпуска	Май 2017
Производители	<u>TSMC (12 нм FinFET)</u>
Тип памяти	<u>HBM2</u>
← <u>Pascal</u>	<u>Turing (consumer), Ampere (consumer, → professional)</u>

Первой видеокартой, использовавшей GPU с этой микроархитектурой, стала карта уровня дата-центра [Tesla V100](#), вошедшая в систему [Nvidia DGX-1](#). Она также используется в [Nvidia Quadro GV100](#) и [Nvidia Titan V](#). Не существует массовых видеокарт [десктопного уровня](#) серии [GeForce](#) на базе микроархитектуры Volta. Для видеокарт десктопного уровня серии [GeForce 20](#) была разработана упрощённая микроархитектура [Turing](#), на основе которой эти карты были анонсированы в августе 2018 года, которая также пришла на смену микроархитектуры Pascal.

Содержание

[Технические подробности](#)

[Продукты](#)

[Примечания](#)

[Ссылки](#)

Технические подробности

Архитектурные усовершенствования микроархитектуры Volta включают следующее:

- [CUDA Compute Capability 7.0](#)
- Память с высокой пропускной способностью 2-го поколения ([HBM2](#)).
- [NVLink 2.0](#): шина с высокой пропускной способностью между [центральной процессором](#) и графическим процессором, а также между несколькими графическими процессорами. Обеспечивает гораздо более высокие скорости передачи, чем те, которые достижимы при использовании [PCI Express](#); обеспечивает скорость 25 Гбит/с на полосу и до 200 Гбайт/с (4 × 50 Гбайт/с) на один GPU.
- 672 Tensor-ядра: Тензорное ядро — это объект, который умножает две матрицы FP16 4 × 4, а затем добавляет к результату третью матрицу FP16 или FP32 с помощью операций умножения примесей и получает результат FP32, который при необходимости можно понизить до результатов FP16. Тензорные ядра предназначены для ускорения обучения [нейронных сетей](#).
- [PureVideo Feature Set I](#) аппаратное [декодирование видео](#).

Продукты

Volta анонсируется как микроархитектура графического процессора в поколении SoC [Tegra Xavier](#), ориентированная на автономные автомобили.

На ежегодной конференции Nvidia GPU Technology Conference 10 мая 2017 года Nvidia официально анонсировала микроархитектуру Volta наряду с GPGPU-ускорителем Tesla V100. Графический процессор Volta GV100 построен по 12-нанометровому техпроцессу с использованием HBM2-памяти с пропускной способностью 900 ГБ/с.

Nvidia официально анонсировала NVIDIA TITAN V 7 декабря 2017 года.

Nvidia официально анонсирует Quadro GV100 27 марта 2018 года.

- Многопоточный мультипроцессор охватывает *64 ядра CUDA* и *4 TMU*.
- Кластер обработки графики охватывает четырнадцать многопоточных мультипроцессоров.
- CUDA-ядра: Единицы отображения текстур: Единицы вывода.
- Ядро Tensor — это FPU смешанной точности, специально разработанный для матричной арифметики.

Примечания

1. Geoff Gasiор (2013-03-19). "Nvidia's Volta GPU to feature on-chip DRAM" (<https://techreport.com/news/24529/nvidia-volta-gpu-to-feature-on-chip-dram>). The Tech Report. Архивировано (<https://web.archive.org/web/20190501155929/https://techreport.com/news/24529/nvidia-volta-gpu-to-feature-on-chip-dram>) 1 мая 2019. Дата обращения: 14 марта 2017.
2. Ryan Smith (2017-05-10). "The NVIDIA GPU Tech Conference 2017 Keynote Live Blog" (<https://www.anandtech.com/show/11360/the-nvidia-gpu-tech-conference-2017-keynote-live-blog>). Архивировано (<https://web.archive.org/web/20190325121842/https://www.anandtech.com/show/11360/the-nvidia-gpu-tech-conference-2017-keynote-live-blog>) 25 марта 2019. Дата обращения: 3 ноября 2018.

Ссылки

- Архитектура NVIDIA Volta (<https://www.nvidia.com/ru-ru/data-center/volta-gpu-architecture/>) (рус.). Официальный сайт NVIDIA Corporation. Дата обращения: 21 сентября 2023.
- GTC17: все подробности архитектуры NVIDIA Volta (<https://www.hardwareluxx.ru/index.php/artikel/hardware/grafikkarten/41864-gtc17-volta-nvidia.html>) (рус.). Hardwareluxx.ru (12 мая 2017). Дата обращения: 21 сентября 2023.
- Nvidia Volta: новая вычислительная архитектура (<https://www.ixbt.com/video4/nvidia-volta.shtml>) (рус.). Hardwareluxx.ru (6 июня 2017). Дата обращения: 21 сентября 2023.

Источник — [https://ru.wikipedia.org/w/index.php?title=Volta_\(микроархитектура\)&oldid=140198650](https://ru.wikipedia.org/w/index.php?title=Volta_(микроархитектура)&oldid=140198650)

Эта страница в последний раз была отредактирована 14 сентября 2024 в 18:56.

Текст доступен по лицензии Creative Commons «С указанием авторства — С сохранением условий» (CC BY-SA); в отдельных случаях могут действовать дополнительные условия.

Wikipedia® — зарегистрированный товарный знак некоммерческой организации «Фонд Викимедиа» (Wikimedia Foundation, Inc.)