从GPU到GPGPU (Lite)

报告人: 蒲镜羽 2023.12.15

目前归属PLCT实验室图形栈小队

Github: KiritakeKumi

Mail:root@niconiconi.us

Twitter: @nowkiritakekumi

Website:niconiconi.us

• GPU: 图形处理单元(Graphic Processing Unit),是一种专门设计用于处理图形和图像计算任务的电子芯片。被开发用于加速图形渲染在视频游戏和三维图形中应用。

• GPGPU: (General-Purpose Computing on Graphics Processing Units) 利用图形处理单元 (GPU) 进行通用计算的技术。最初设计用于处理图形数据的GPU具有强大的并行处理能力,而GPGPU技术使这些能力可以应用于非图形的计算任务。

一切的开始 NVIDIA

- 2006年, NVIDIA发布了革命性的GeForce 8800 GTX, 属于GeForce 8系列家族
- 支持单精度 (32位) 浮点计算 ,并给出了名为Compute Unified Device Architecture的 通用计算方案,即CUDA



AMD

• AMD在其Radeon HD 4000系列显卡上支持了OpenCL

AMD: "性能有限"



CUDA? OpenCL?

- ・共同点
- **1.并行计算**: OpenCL和CUDA都是为了利用多核心处理器(如GPU)的并行计算能力而设计的。
- 2.高性能计算:两者都被广泛应用于高性能计算任务,如科学模拟、数据分析、机器学习等。
- 3.编程模型: OpenCL和CUDA都提供了一套编程模型和API, 允许开发者编写能在并行硬件上运行的程序。

CUDA? OpenCL?

・区别

1.厂商和平台支持:

- 1. CUDA是NVIDIA开发的,专门针对其GPU。它只在NVIDIA的GPU上运行。
- 2. OpenCL是一个开放标准,由Khronos Group维护,旨在提供跨平台的并行编程。它支持多种厂商的多种处理器,包括NVIDIA和AMD的GPU,以及CPU。

2.性能优化:

- 1. CUDA通常在NVIDIA的GPU上提供更好的性能和更深入的硬件优化。
- 2. OpenCL提供的是更广泛的硬件兼容性。

3.编程语言和生态系统:

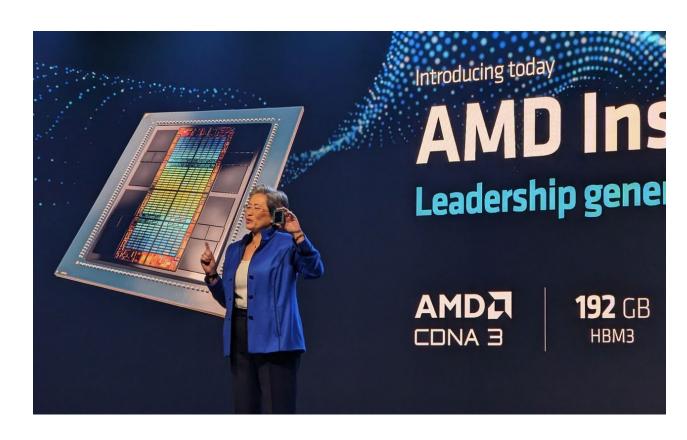
- 1. CUDA主要使用CUDA C/C++,这是一种类似于C/C++的语言,但包含了NVIDIA的扩展。
- 2. OpenCL使用一种类似于C99的语言,有自己的语法和功能集,同时也支持其他语言绑定。

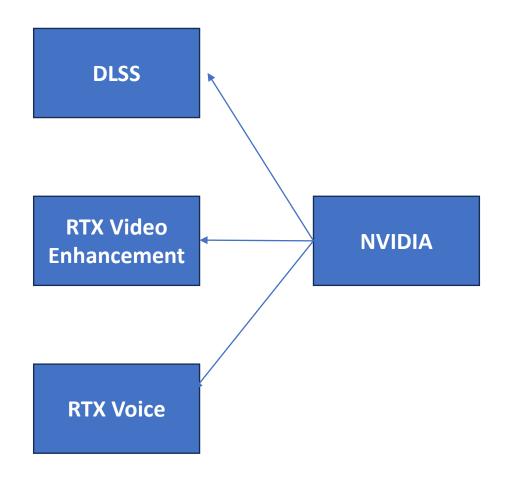
ROCm!

- AMD在2016年推出Radeon Open Compute Platform (ROCm)
- ROCm是一个开源平台,旨在提供用于高性能计算(HPC)和机器学习的软件栈,特别是针对AMD的GPU,用来和NVIDIA的CUDA竞争。

- 但是这东西现在能用了嘛?
- 能用,好不好用另说.jpg









那么此时此刻国内的GPU厂商都在干什么呢?

中国人民的老朋友 Imagination

- 被Apple挖空后,Imagination 也被中国注资了
- •特点
 - ・文档很少
 - · 官方驱动曾经很糟糕 (现在也很糟糕)
 - Imagination 曾经给苹果供 IP (PowerVR) , 相对强

战未来.jpg 摩尔线程

- 核基础设计可能来自于某厂
- 也提供了一个"兼容CUDA"的计算框架 同时这里的计算单元叫"MUSA"
- 纸面参数为14.4TFLOPS 的单精度浮点算力
- 某厂本身对于D3D这边的驱动就很**一般**,目前性能属于是发布后稳步更新驱动爬坡。
- 现在已经能畅玩?原神了!
- 虽然纸面性能很强,但是有很多部分在拖后腿。





穷的变形的 AMD

• 2016年:

推土机的大规模失败, Zen 设计还未完全定型投产

- 濒临破产、靠定制 (Xbox, PS) 甚至变卖资产 (顶级企业卖大楼.jpg) 维生
- 授权费三亿美金,某司获得 Zen 1 除 FPU 与安全子系统外的绝大多数设计

图形栈小组的公开课程

https://space.bilibili.com/296494084/channel/collectiondetail?sid=16
35641

• 虽然暂时停更了,但是很快会重新更新。

