GPU架构和概念

图示

AI 生成的内容可能不正确。

因此，GPU 的设计侧重于并行处理大块数据（三角形或多边形），这是图形应用程序的一项要求。与每个时钟周期只能处理几十个并行线程或进程的 CPU 相比，GPU 能够同时处理数千个并行线程。

为了让 GPU 执行任务，在某个时刻，数据必须从 CPU 传输到 GPU。当任务完成且结果要写入文件时，GPU 必须将数据发送回 CPU。GPU 要执行的指令同样是从 CPU 发送到 GPU 的。每一次这样的传输操作都由 PCI 总线进行协调。虽然在本章中我们不会讨论如何实现这些操作，但我们会探讨 PCI 总线的硬件性能限制。

由于这些限制，设计不佳的 GPU 应用程序的性能可能比仅使用 CPU 的代码还要差。我们还将讨论 GPU 的内部架构以及 GPU 在内存和浮点运算方面的性能。

文本, 信件

AI 生成的内容可能不正确。

每个计算设备都有单独的GPU、GPU内存和多个计算单元（CU）。

文本

AI 生成的内容可能不正确。

文本, 信件

AI 生成的内容可能不正确。

GPU 编程模型

文本

AI 生成的内容可能不正确。

文本

AI 生成的内容可能不正确。

文本, 信件

AI 生成的内容可能不正确。

工作组是由多个工作项（Work Item）组成的，工作项是 GPU 并行计算中的最小执行单元，多个工作项组合成工作组，共同协作完成一个相对独立的子任务。

计算单元是 GPU 硬件的物理组成部分，负责执行具体的计算操作，它可以同时处理多个工作组中的工作项。

文本, 信件

AI 生成的内容可能不正确。

文本, 信件

AI 生成的内容可能不正确。

文本, 信件

AI 生成的内容可能不正确。

基于指令的GPU编程

文本

AI 生成的内容可能不正确。