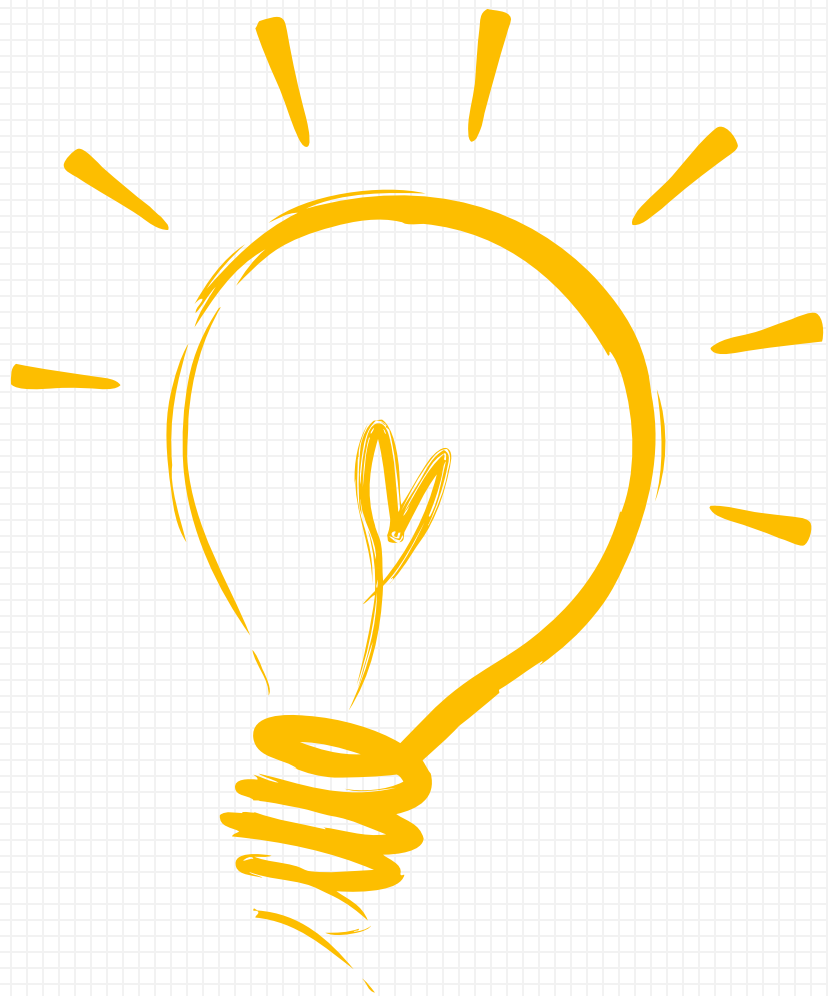


CUDA简介

CUDA并行编程系列课程
主讲：权双

CONTENTS



01 GPU硬件平台

02 CPU+GPU异构架构

03 CUDA介绍

什么是GPU

- ★ GPU 意为图形处理器。GPU 也常被称为显卡。与它对应的一个概念是CPU。
- ★ GPU: 数据运算
- ★ CPU:逻辑运算



什么是GPU

★ GPU性能指标:

- 1、核心数
- 2、GPU显存容量
- 3、GPU计算峰值
- 4、显存带宽



CPU+GPU异构架构

GPU不能单独计算，CPU+GPU组成异构计算架构；

CPU起到控制作用，一般称为主机（Host）；

GPU可以看作CPU的协处理器，一般称为设备（Device）；

主机和设备之间内存访问一般通过PCIe总线链接。

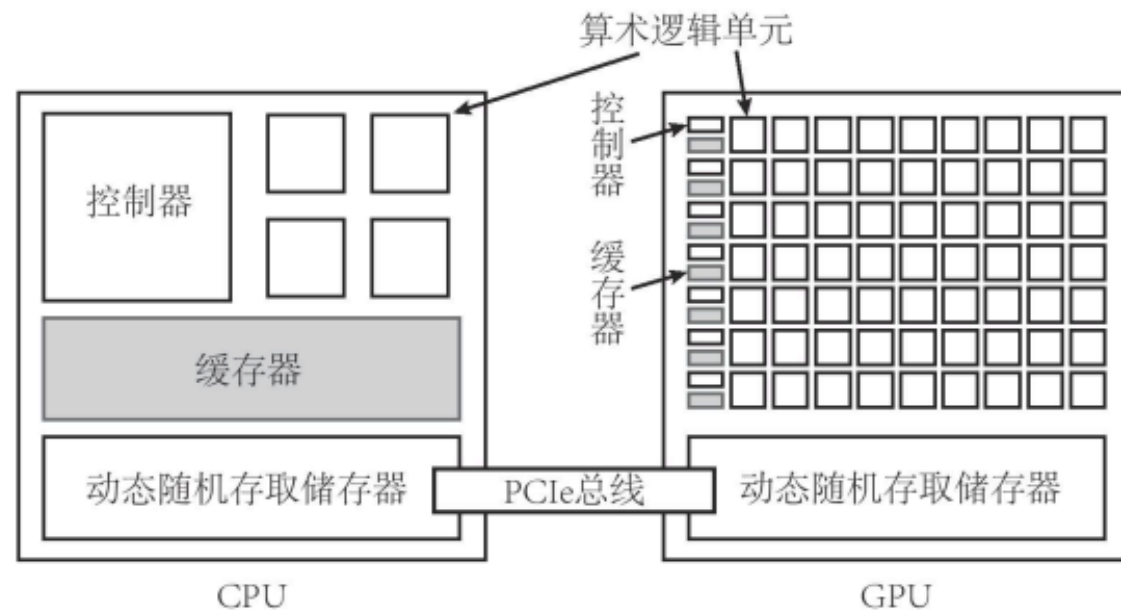


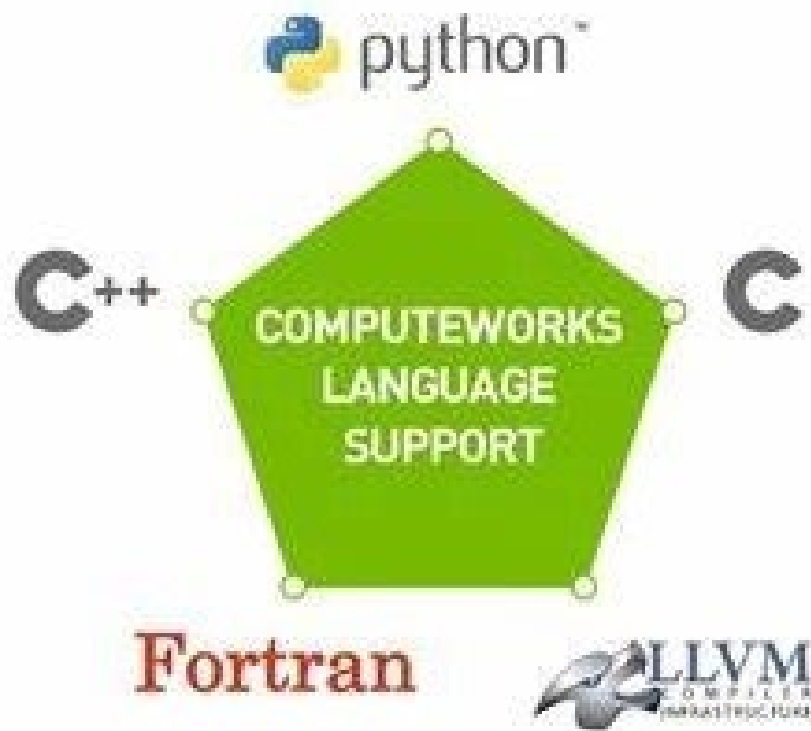
图 1.1: CPU（左）和非集成 GPU（右）的硬件架构示意图。

什么是CUDA



- ★ 2006年, NVIDIA公司发布CUDA;
- ★ CUDA建立在NVIDIA的GPU上的一个通用并行计算平台和编程模型
- ★ 基于GPU的并行训练已经是目前大火的深度学的标配
- ★ 目前最新版本CUDA12.2

CUDA编程语言



- ★ CUDA旨在支持各种语言和应用程序编程接口
- ★ 最初基于C语言，目前越来越多支持C++
- ★ CUDA还支持Python编写
- ★ 本课程主要基于C++的CUDA编程

CUDA运行时API

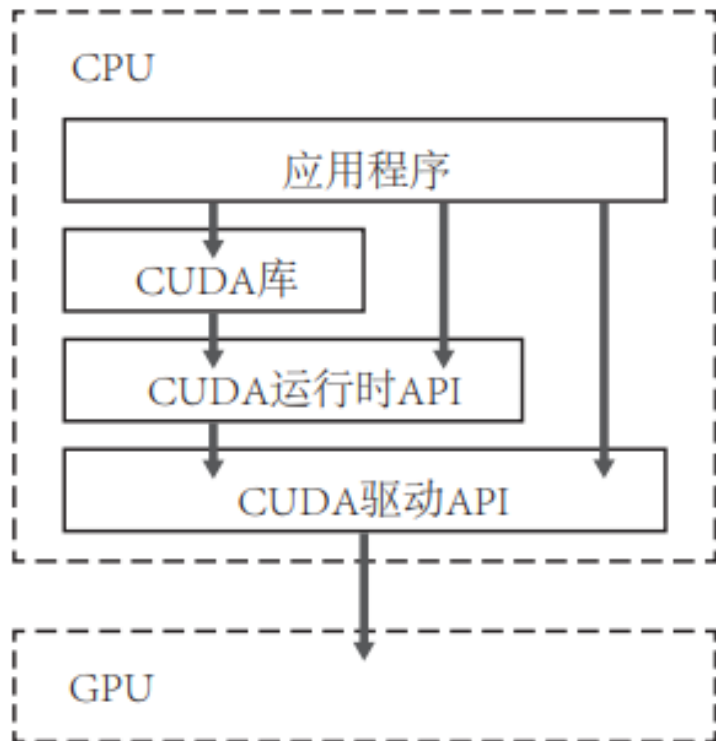


图 1.2: CUDA 编程开发环境概览。

- ★ CUDA 提供两层 API 接口，CUDA 驱动 (driver) API 和 CUDA 运行时 (runtime) API
- ★ 两种 API 调用性能几乎无差异，课程使用操作对用户更加友好 Runtime API
- ★ 本课程主要基于 C++ 的 CUDA 编程

THANKS

谢谢聆听

