

GPU并行计算和CUDA程序开发

周 斌 @NVIDIA 2014年4月



授课老师

▶ 周 斌

- ▶ NVIDIA CUDA Fellow、中科大CUDA课程讲师
- ▶ 曾任NVIDIA 高性能计算开发技术资深工程师
- ▶ 科研:信号和信息处理、视频图像处理
- ▶ 兴趣:数据分析, 密码学和密码分析, 无人机系统
- ▶ 涉及领域:
- ▶ 气象数值预报系统、视频图像处理、高速信号处理、生物信息学、网络搜索、加密系统等
- ▶ 关注GPU系统架构、高性能应用、移动互联网等

对象和目标

- ▶ 有志于从事CUDA/GPU系统开发的开发者
 - ▶ 使用CUDA/GPU并行计算系统的科研工作者
 - ▶ 计算机，电子，自动化，生医等相关专业的硕士研究生或高年级本科生

 - ▶ 了解和掌握GPU/CUDA并行计算系统的分析，设计，开发，调试和优化方法
 - ▶ GPU并行计算系统的分析能力，编程能力，开发经验

 - ▶ 风格：中英夹杂
-

预修课程和教材

► 基础知识

- 1) 计算机体系结构基础 2) C语言程序设计
- 3) 计算机算法基础 4) 线性代数

► 课程内容参考

- 1. CUDA C Programming Guide, NVIDIA Corp.
- 2. CUDA Best Practice Guide, NVIDIA Corp.
- 3. Programming Massively Parallel Processors, 2010, David Kirk and Wen-mei Hwu

► 声明：参考和引用了如下的工作

- 1. Patrick Cozzi, CIS 565, University of Pennsylvania
- 2.
- 3.
- 4. Udacity CS 344 Intro to Parallel Computing



内容

- ▶ 1) CPU体系架构概述
 - ▶ 2) 并行程序设计概述
 - ▶ 3) CUDA开发环境搭建和工具配置
 - ▶ 4) GPU体系架构概述
 - ▶ 5) GPU编程模型
 - ▶ 6) CUDA编程(1)
 - ▶ 7) CUDA编程 (2)
 - ▶ 8) CUDA编程 (3)
 - ▶ 9) CUDA程序分析和调试工具
 - ▶ 10) CUDA程序基本优化
 - ▶ 11) CUDA程序深入优化
 - ▶ 12) 最新NVIDIA GPU 和 CUDA特性
-

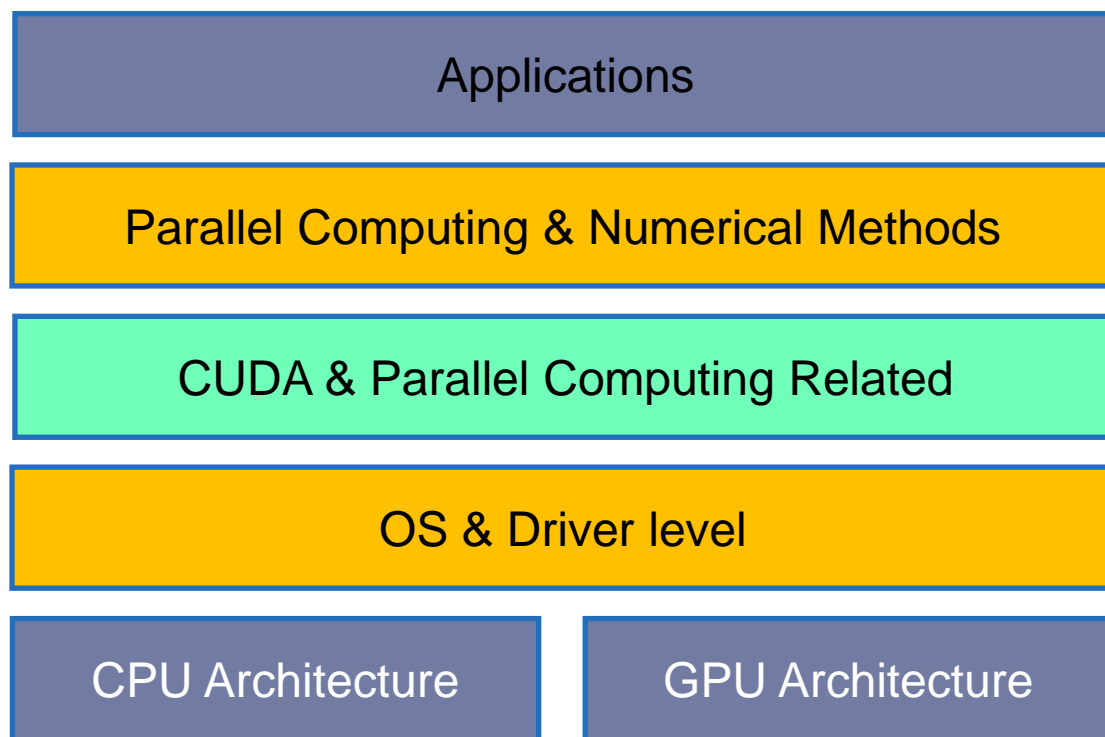


简单测试

- ▶ Programming Model?
- ▶ Memory Model?
- ▶ WARP?
- ▶ Occupancy?
- ▶ Optimization
 - ▶ Compute Bound or memory Bound?
 - ▶ Little's law?
- ▶ Others
 - ▶ CUDA-GDB
 - ▶ Parallel Nsight?



生态环境



课程资源

- ▶ NVIDIA CUDA Zone

- ▶ <http://developer.nvidia.com/category/zone/cuda-zone>

- ▶ QQ群

- ▶ GPU深度开发



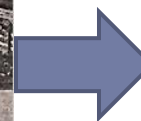
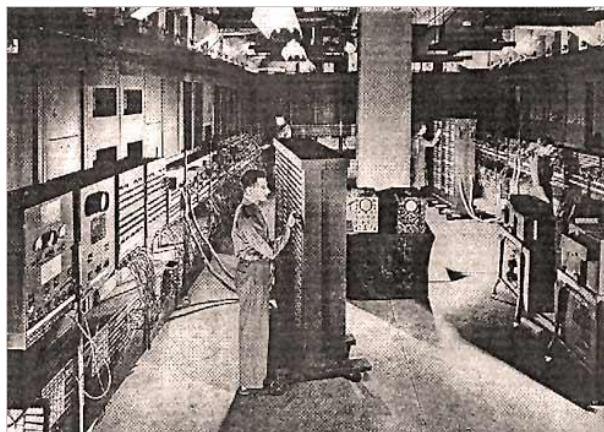
开始课程之前的问题...

- ▶ 1) 为什么我们要使用计算机?

为了更好地解决计算问题

- ▶ 2) 你需要什么样的计算机? 畅想...

速度无穷快(1Phz?)
无穷多内存(1EB?)
智能化的接口
(人工智能)



可惜世间总是太多无奈…

- ▶ 现在的CPU系统已经遇到各种瓶颈
- ▶ 只能向多核及并行系统发展
- ▶ 顺势而生的 *GPU* – Graphics *P*rocessing *U*nit

