GPU并行计算和CUDA程序开发及优化

周　斌　　2014年秋

课程概况

- 时间
  - 7~16周 周六 上午 9:45~12:10
  - 实验课程 9~16周 周六下午 14:00 – 17: 00； 晚18:00-21:00;
- 地点
  - 课程:西区3教副楼3C121
  - 实验： 电四楼209
- 人物
  - 老师+助教+同学
- 事件
  - 课程+实验+....

计划增加一次课程

- 时间
  - 周日 上午 9:50~11:20 及 下午 2:00~3:30
- 地点
  - 课程:西区3教副楼3C121 或待定
- 人物
  - 老师+助教+同学
- 事件
  - 课程.

授课老师

- 周　斌
  - 山东省科学院海仪所海洋遥感遥测研究室主任
  - NVIDIA CUDA Fellow、中科大客座研究员
  - 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

授课老师

- 邮件：
  - synosy@gmail.com
- 答疑：
  - 课上课下
  - QQ: GPU深度开发
  - 邮件
  - iqiyi 搜索 CUDA 教程
- 风格：中英夹杂

名人堂

李索恒

陈凯

曹旭磊

？？？？

2012年秋季学期：李索恒
2013年春季学期：陈　凯
2013年秋季学期：曹旭磊

成绩优秀的同学获得了NVIDIA公司提供的奖励
本学期奖励继续么？

课程情况总结

- 2012年秋
  - 90人选课，平均成绩83分，除1人退课外，全部及格并基本符合正态分布
  - 12个Project
    - 内容: MD5&SHA-1; FDTD; Ray-tracing; Page-Rank;Sparse-Solver; Smith-Waterman….
  - 3人获得NVIDIA公司的奖励
- 2013年春
  - 23人选课，平均成绩85分，全部及格
  - 1人获得NVIDIA公司的奖励

课程情况总结

- 2013年秋
  - 168人选课，平均成绩85分，12人退课，全部及格并基本符合正态分布；20个Project
  - 3人获得NVIDIA公司的奖励

The Students

- Gain the abilities of:
  - Construction of GPU hardware systems
  - CUDA/GPU Programming for their research work
  - Deploying CUDA-accelerated applications

- Highly Evaluted

	benefit(5.0)	Total Evaluation(5.0)	Teaching Overall
Univ. Average	4.29	4.31	38.83
School. Average	4.29	4.31	38.83
This Course	4.50	4.44	40.61

- Comments from students:
  - Of High quality, learned a lot, very useful. Hope continuing.
  - Very practical.
- Some Other Universities Students went to USTC for it.

The Projects

Ray - Tracing

Sparse Solver

FDTD Simulation

WPA2-PSK Attack

Page Rank, Search Engine

...

CUDA + Hadoop

3D Reconstruction

Image&Video Processing

Speech Recognization

Machine Learning

Monte Carlo method

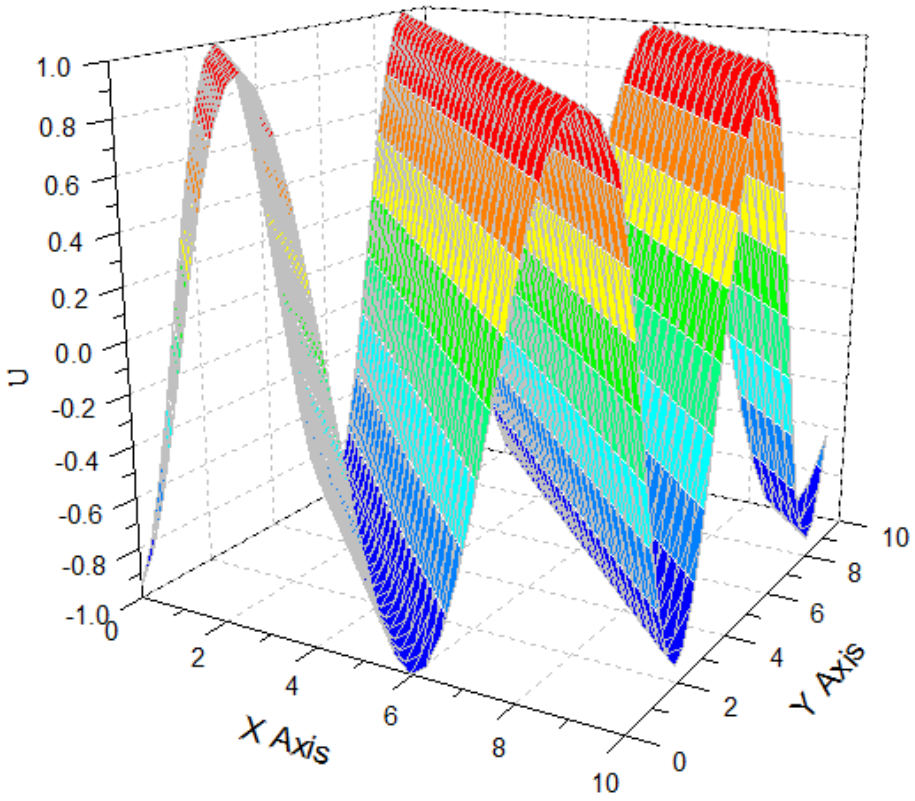
...

3D wave equation: >100x Speedup (dubious)

3D实现环境（服务器）

CPU	**E5-2620**
GPU	**Tesla K20c**
CUDA	**5.0**
3-D Grid	Support

z方向固定，k=128时， x, y, u的三维图

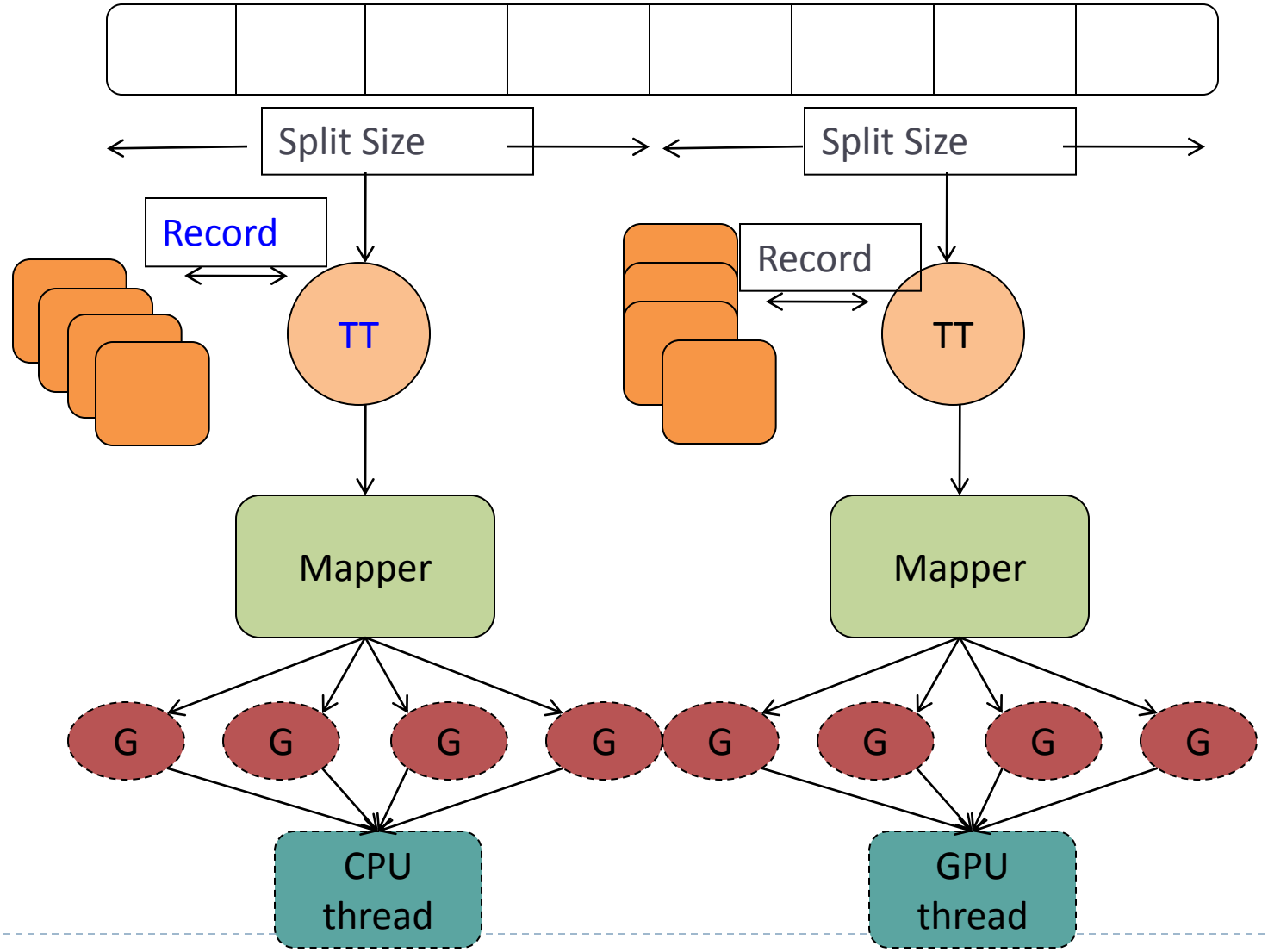


origin软件的分辨问题，在x,y方向上只取了64个点，进行作图，显得图形不圆滑，若是取所有点绘图，图形会非常圆滑

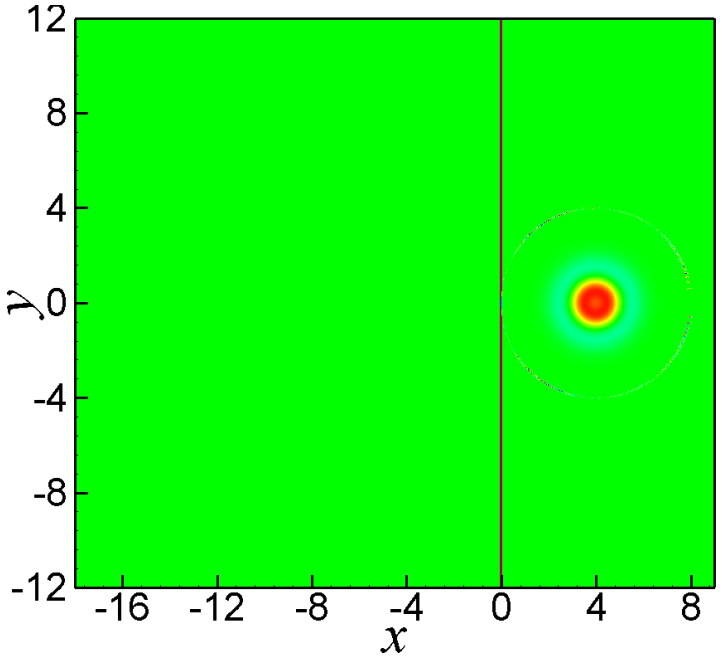
三种方式的结果对比

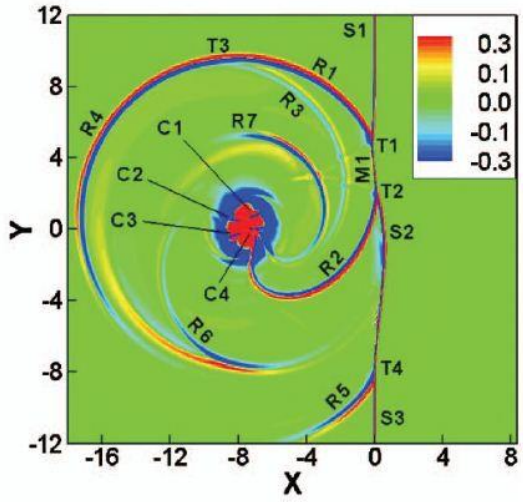
	执行时间(ms)
CPU	3170000
GPU3D	94.29084
GPU3D_SHARED	92.726463

CUDA && Hadoop: Basic Function



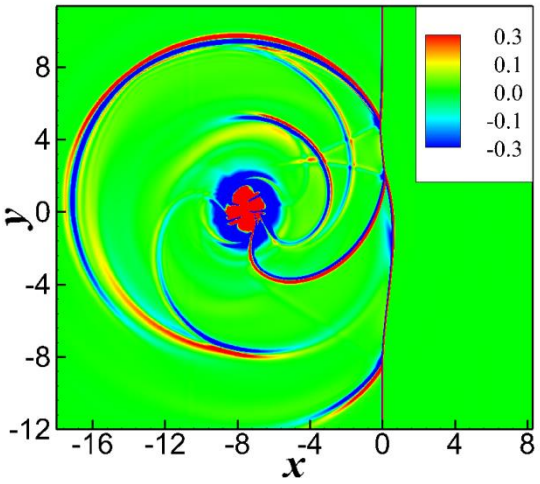
FDTD Simulation: 10x Speedup



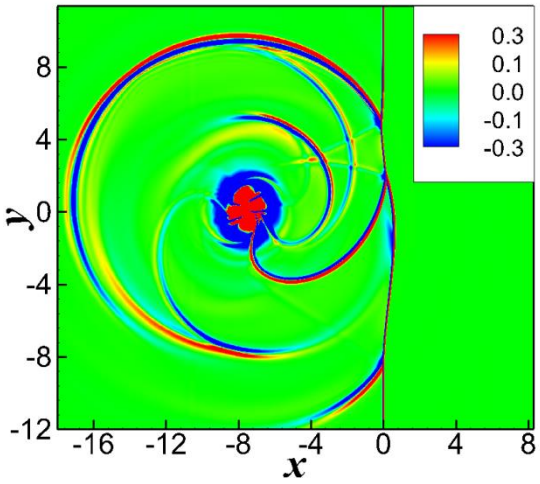


(h) $t = 12.0$

t = 12.0, reference



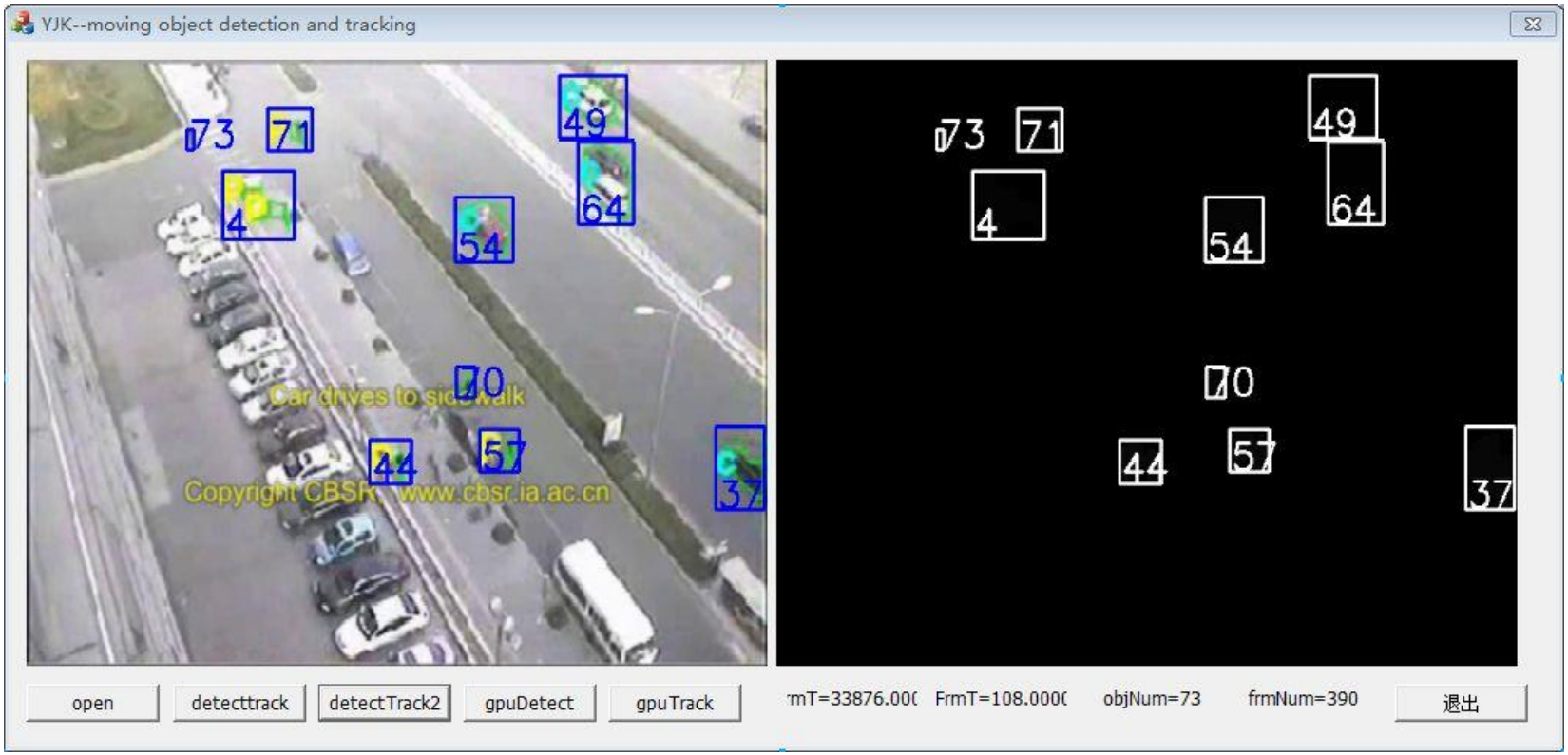
t = 12.0, single GPU



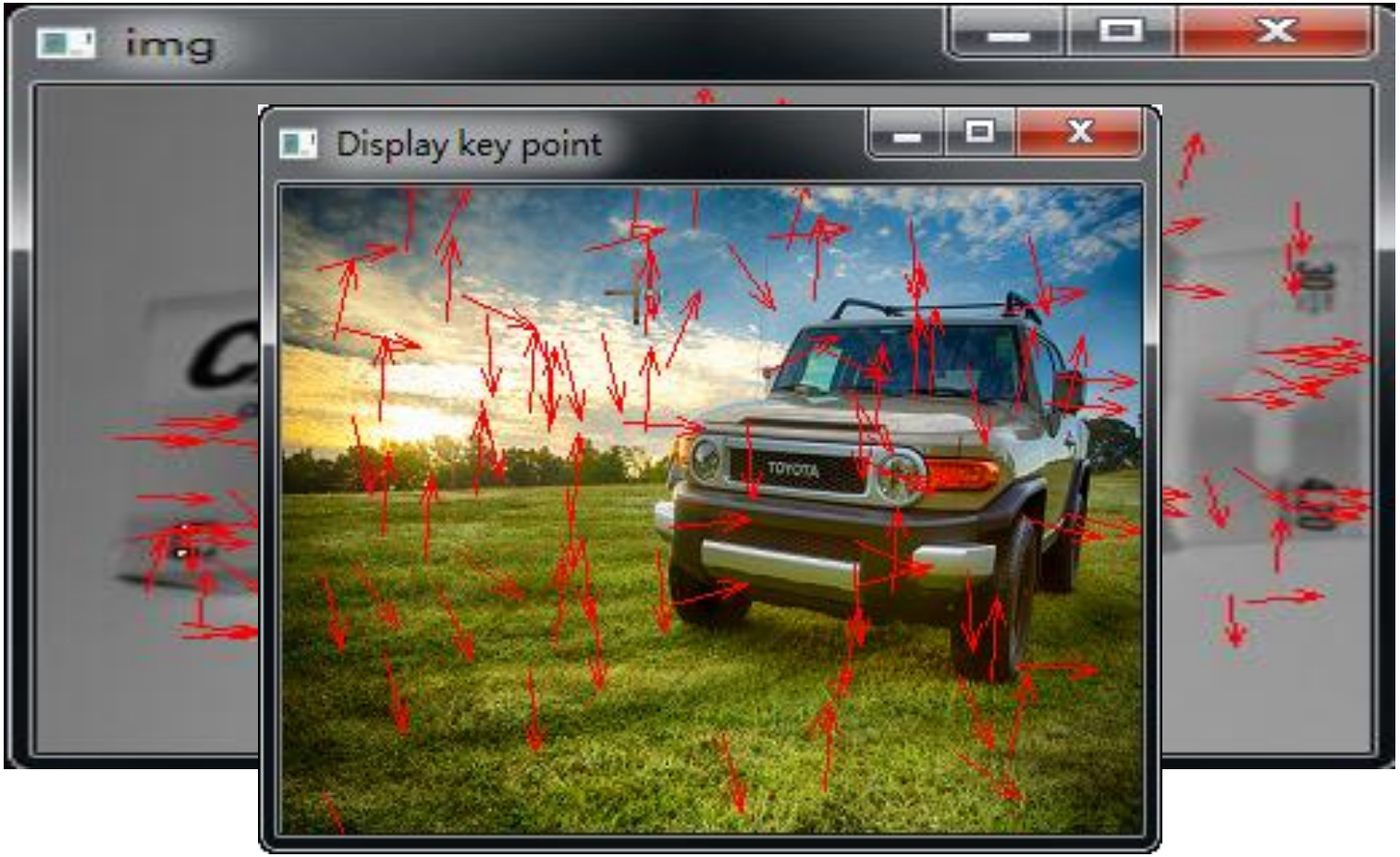
t = 12.0, dual GPUs

Video Multi-Target Tracking: Speedup 5x

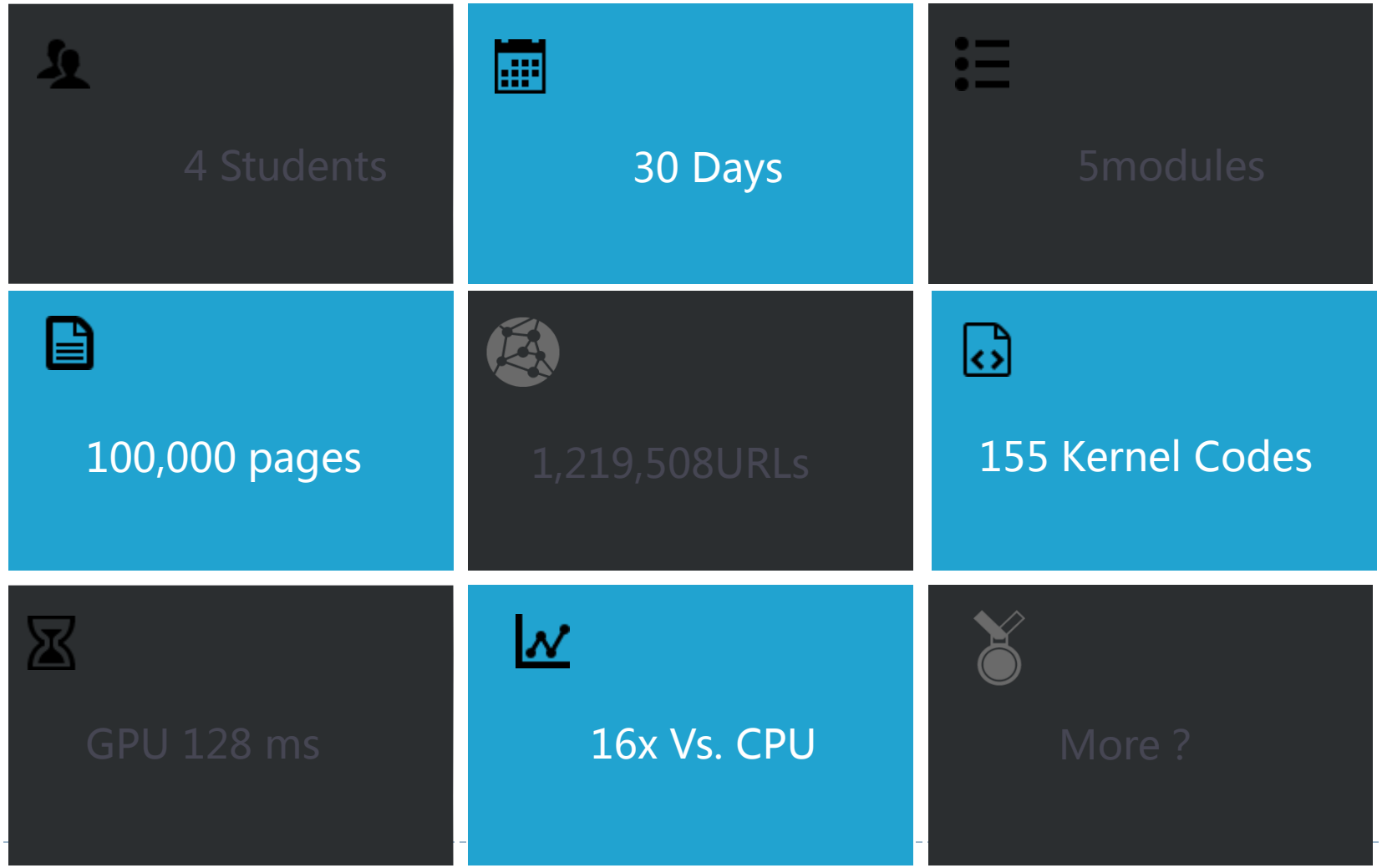
Improvement needed



SIFT on GPU: 3x speedup



GPU-based Search Engine: 16x speedup



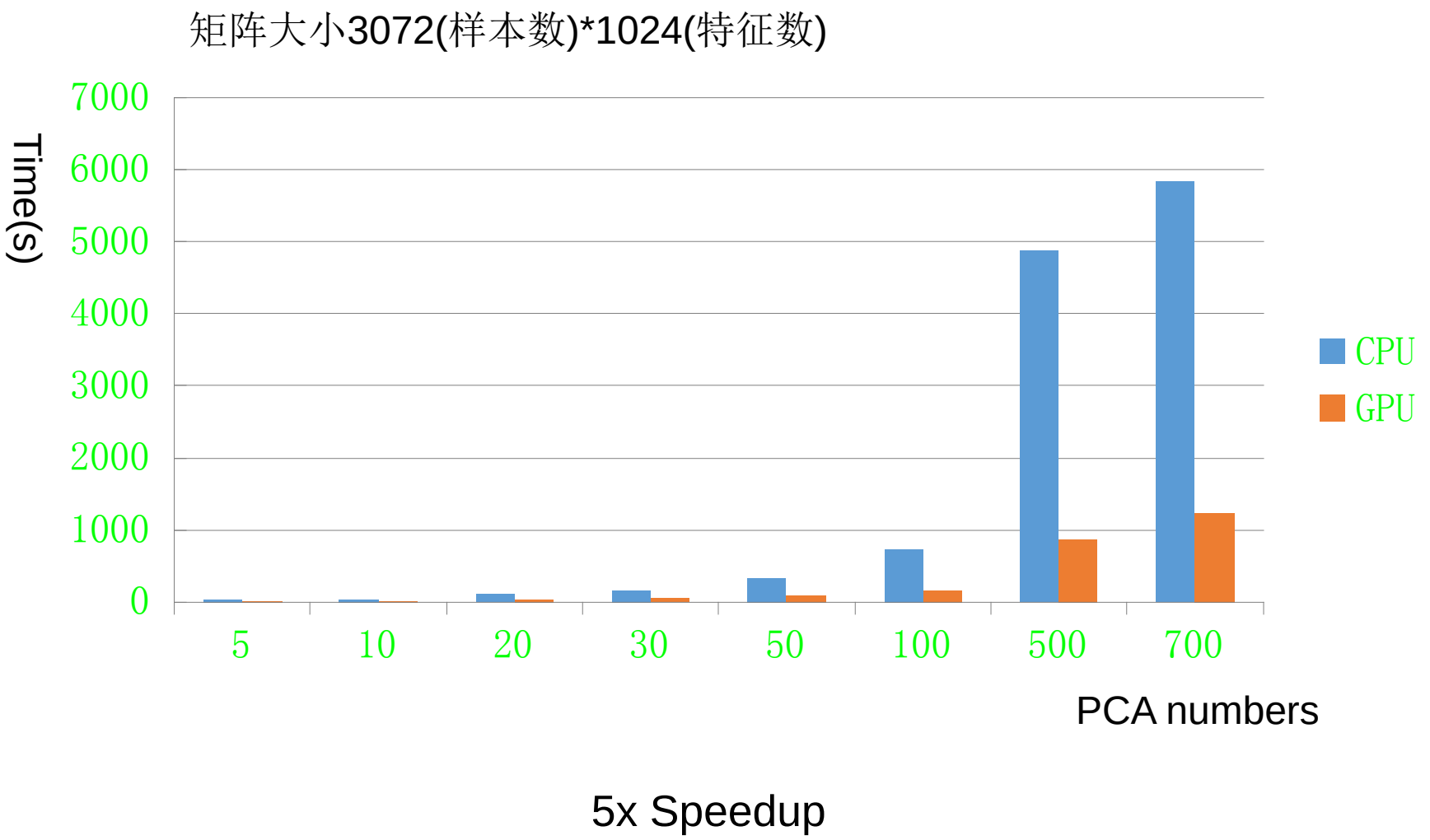
搜索首页


192.168.1.168:8080/webForSearch/index.jsp
SA12011
baidu
bing
翻译
豆瓣FM
有道
学术搜索
百度网盘
科研
项目
食谱
购物
Online
娱乐
IT
四轴
充电
其他书签
CLCL
BETA
中科大
Search

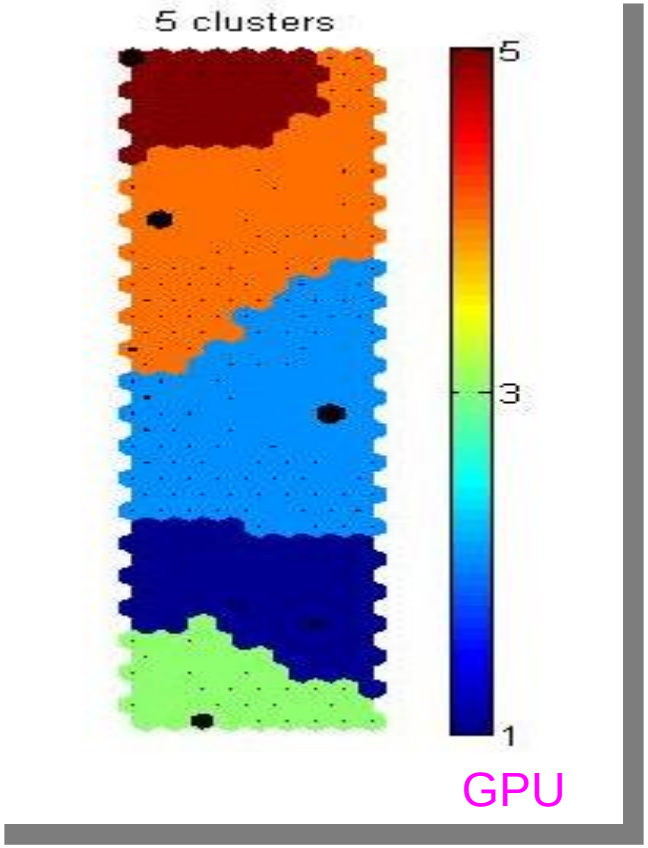
搜索结果

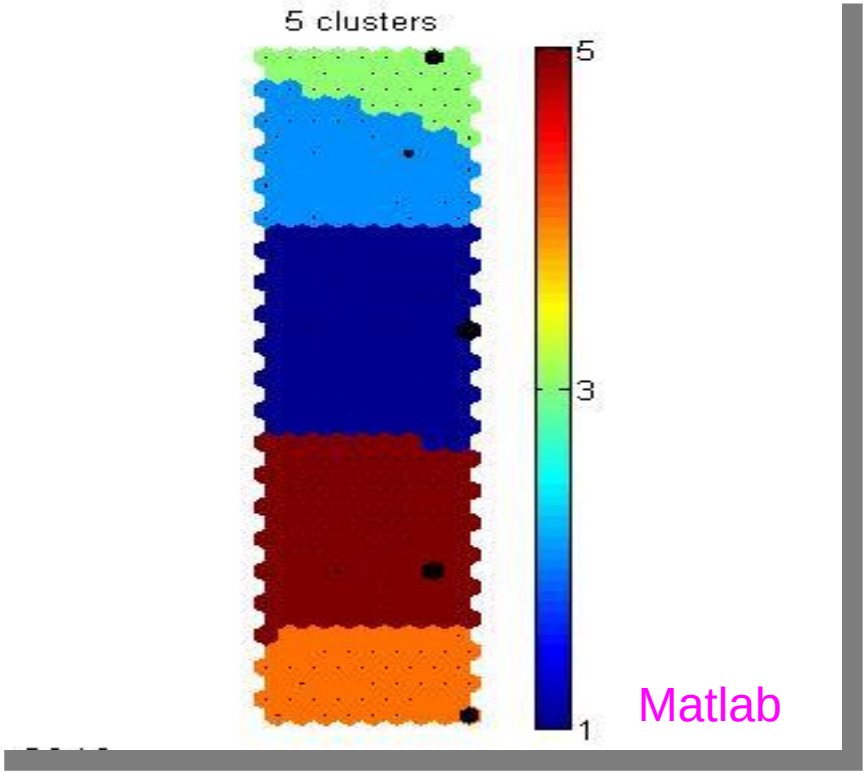

查询结果
192.168.1.168:8080/webForSearch/MyJsp.jsp
SA12011 baidu bing 翻译 豆瓣FM 有道 学术搜索 百度网盘 科研 项目 食谱 购物 Online 娱乐 IT 四轴 充电 其他书签
Search For 中科大 Find Total: 246total time for serching: 2ms
中科大“学术猫”受捧 -新华财经-新华网
PageRank=0.00003324
中科大“学术猫”受捧 -新华财经-新华网 中科大“学术猫”受捧 2013年11月15日 05:30:58 | 责任编辑：李天真 | 来源： 光明网 11月12日，中国科学院院士、中科大教授more
如果在中科大遇见你 中国科学技术大学新闻网
PageRank=0.00002995
如果在中科大遇见你 中国科学技术大学新闻网 如果在中科大遇见你 2013-11-01 逍遥派小掌门 如果在中科大遇见你，在初冬的早晨，要相约去吃一次食堂的早餐，慢慢的习惯，习惯阳光和你都在身旁，想要more
三千人腌250吨泡菜 -新华财经-新华网
PageRank=0.00003324
中国新闻网 当地时间2013年11月13日，韩国首尔，韩国举行爱心泡菜分享活动，超过三千名参加者现场组成爱心形状，集体腌制250吨的泡菜。 1 1 支持键盘翻页←左右→ 分享到： 更多图片 中科大more
稀世粉钻8300万美元 -新华财经-新华网
PageRank=0.00003324
地时间11月13日，“粉红之星”粉钻在瑞士日内瓦拍卖，这枚罕见的宝石最终以8300万美元成交，写下史上宝石拍卖新高。图为“粉红之星”粉钻。（资料图） 1 1 支持键盘翻页←左右→ 分享到： 更多图片 中科大more
[科技日报]接续断裂的创新链条— 中科大先研院为成果转化探路 中国科学技术大学新闻网
PageRank=0.0000243
[科技日报]接续断裂的创新链条— 中科大先研院为成果转化探路 中国科学技术大学新闻网 [科技日报]接续断裂的创新链条— 中科大先研院为成果转化探路 2013-12-06 本报记者 吴长锋 11月11日more
黄山雨后现“佛光” -新华财经-新华网
PageRank=0.00002719
黄山风景区雨后放晴，云海飘荡。黄山风景区始信峰出现今年入冬以来最为持久的“佛光”奇观，令人叹为观止。中新社发方也广德 摄 图片来源：CNSPHOTO 1 1 支持键盘翻页←左右→ 分享到： 更多图片 中科大more

PCA on GPU: 5x Speedup

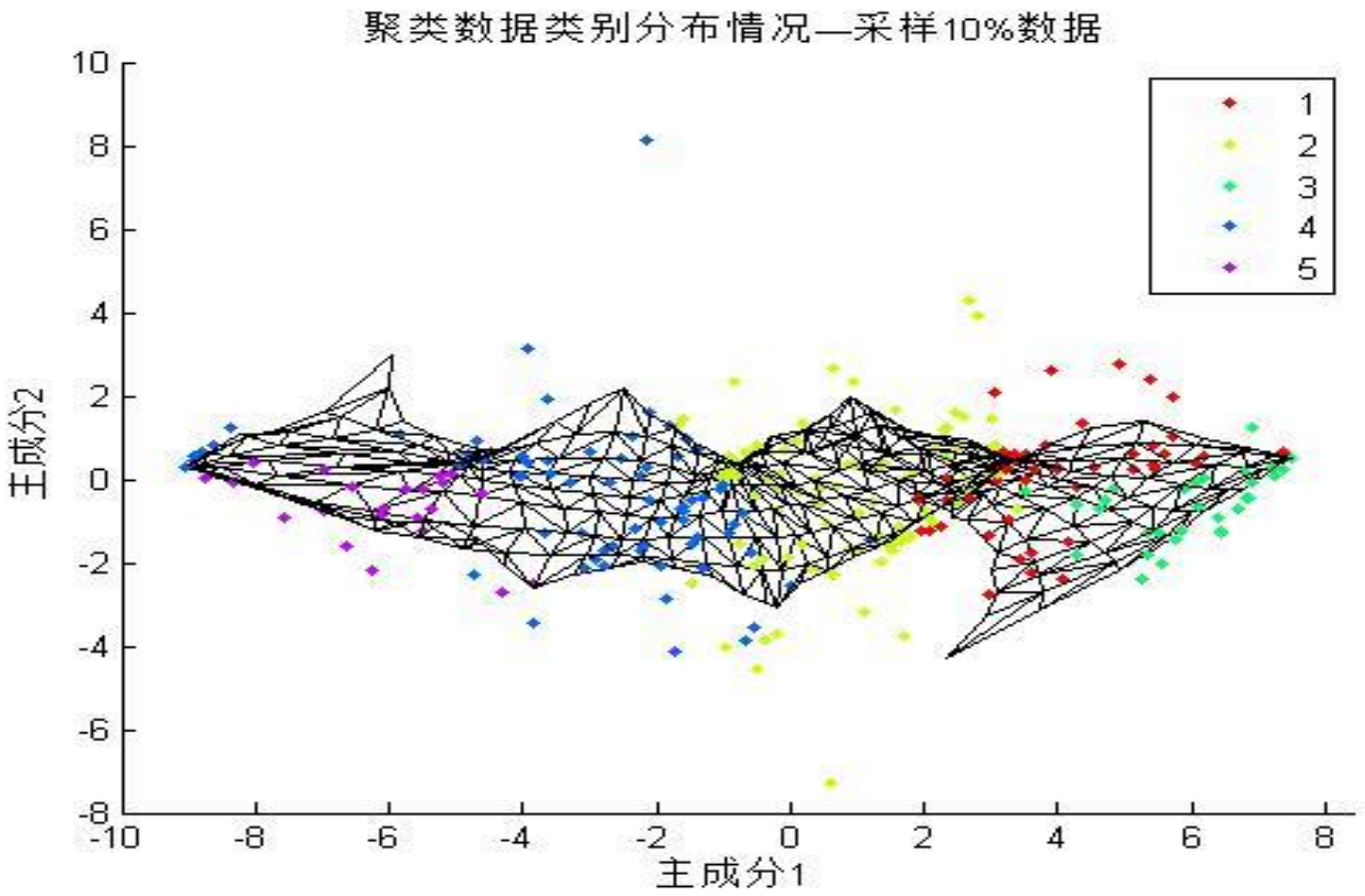


GPU-SOM(Self Organizing Map): 50x





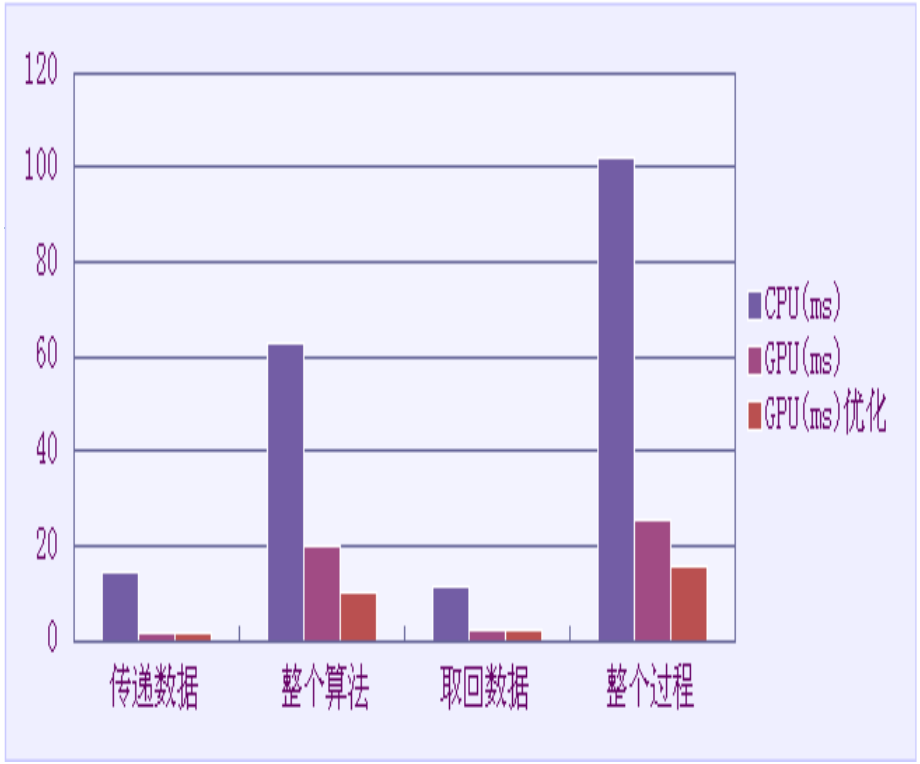
Data Clustering



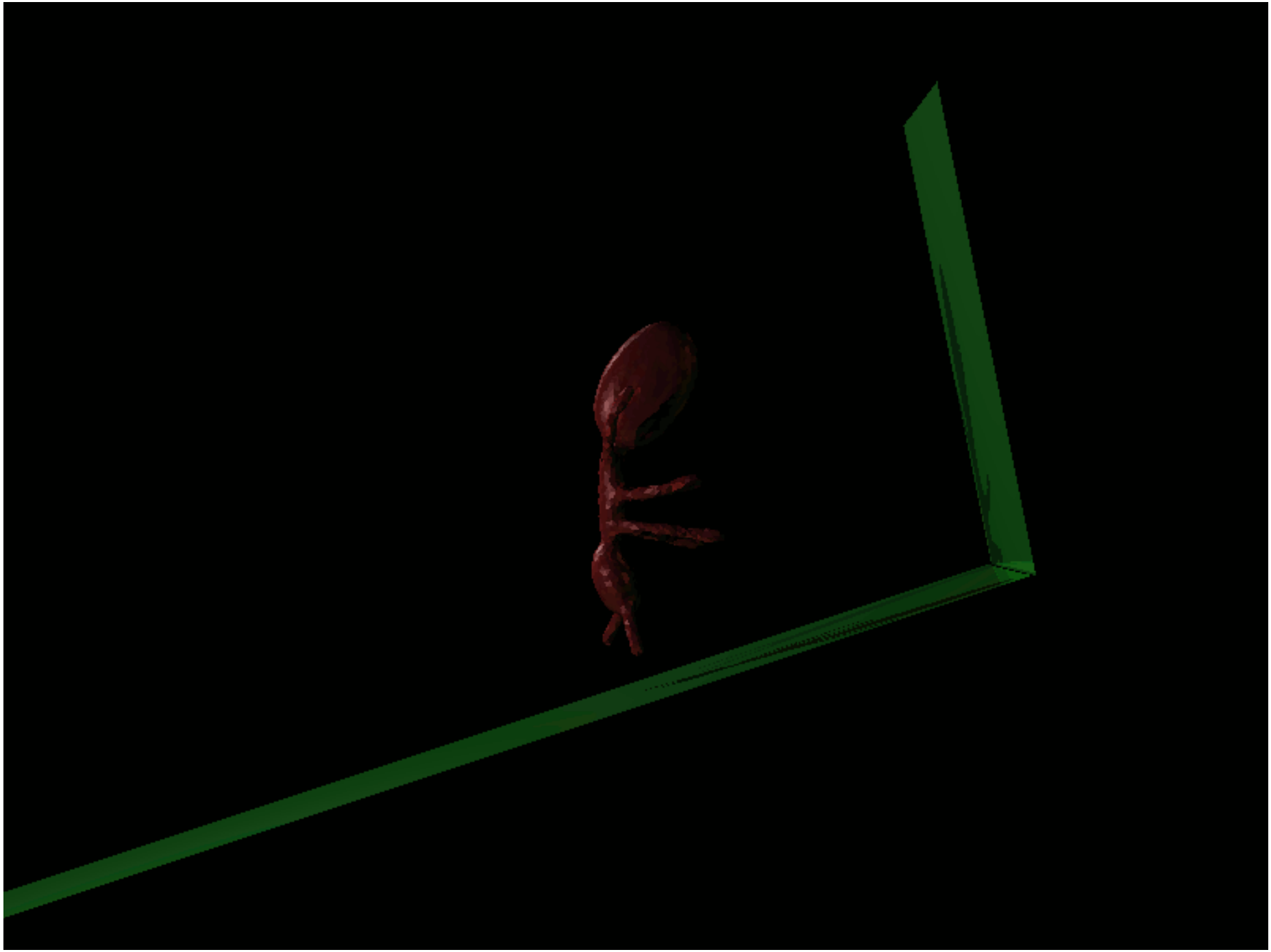
DeFrog: RealTime, 20x Speedup

Image Enhancement: 6x

1，硬件平台：
CPU------Pentium 2.2GHz
GPU------G210M,计算能力1.2
2，软件平台：
操作系统为 windows XP
运行环境为 VS2008，cuda toolkit 5.5，opencv
3处理的图片的大小1200x800



	Data Transfer	Compute	Data Retrieve	Totally
CPU(ms)	14.211776	63.003777	11.579264	102.012772
GPU(ms)	1.878592	20.22376	2.439744	25.761248
GPU(ms)-Opt	1.878592	10.21354	2.439744	15.761248



One of the world's Best SHA-1 on NVIDIA GPU

Performance Comparison

- **GTX 460**, 7 multiprocessors, 336 sp @ 1557MHz
  - '00000000' – '00999999', totally 10^6 messages
  - Our competitors are set into brute-force mode

	SHA-1	MD5
Our implementation (Hash/s)	**400.3M**	671.5M
oclHashcat-lite (Hash/s)	361.0M	**1329.3M**
IGHASHGPU (Hash/s)	371.5M	675.1M

课程资源

- NVIDIA CUDA Zone
  - http://developer.nvidia.com/category/zone/cuda-zone
- 网络教材和课程
  - http://www.bb.ustc.edu.cn/
- 相关素材
  - 将通过网络和电子邮件发布
- Others
  - 请各位多利用网络搜索引擎

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

Know the backgrounds and computer level of the students

- Many Students don’ t have solid programming backgrounds
- Adding Some contents about C programming will be helpful
- **Add Some Basic Computer Architecture Contents**
- **Add Some Basic numerical methods**
- **Add some Parallel Computing contents**

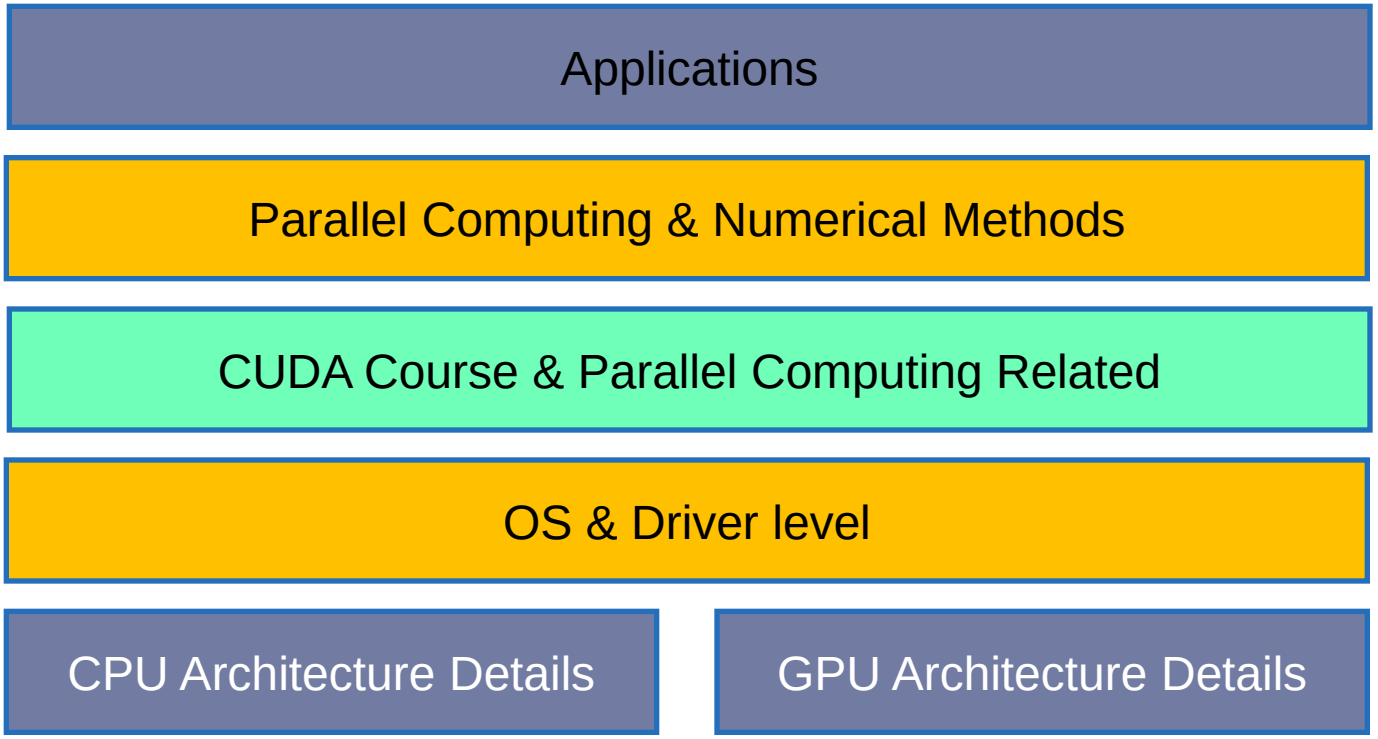
Prerequisite

- C Programming Language
- Basic Computer Architecture
- Basic Algebra and Numerical Methods
- Domain Knowledge
- Linux OS

Prerequisite

- Many students have high scores but not really know how and why
- Don’ t take Prerequisite seriously when they claim they are ok. They are really not!
- However, give them basic trainings and review on the prerequisites

课程定位


Applications
Parallel Computing & Numerical Methods
CUDA Course & Parallel Computing Related
OS & Driver level
CPU Architecture Details
GPU Architecture Details

课程目标

- 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并行计算系统的分析，设计，开发，调试和优化方法
- 系统分析能力，编程能力，开发经验

目标说明

- Not theoretical Algebra or Numerical methods
  - But needs to add some basic algebra
- Not Computer Architecture Course
  - But needs to add some basic contents
- Not Parallel Computing Course
  - But needs to add some contents

课程模式

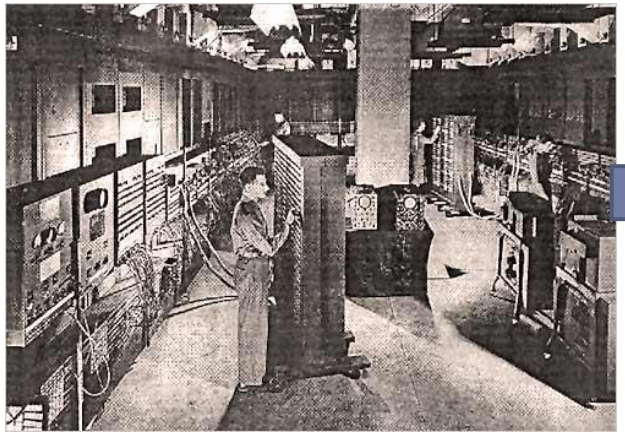
- 强调面向应用和实践
- 最新的业界动态和技术
- 课堂讲解+上机实验
- 评分方式
  - 1) 平时作业和实验报告 40%
  - 2) 课程考试 25%
  - 3) 大作业 35%
- 学生小组制

作业提交

- 每周六，周日 讲完课程布置作业
- 在下一个周四24:00之前提交
- 提交方式
  - 网络系统：以服务器时间为准
  - 电子邮件：以我或助教的电脑时间为准
  - 邮寄方式：以当地邮局邮戳为准
  - 当面提交：以我或助教的手机手表时间为准
- 晚交一天扣本次作业10%的分，直至扣完
- 严禁抄袭，凡抄袭和被抄袭者都是0分

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…
  - 速度更快
  - 内存更大
  - 无穷多的外存
  - 智能化的接口



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* – *G*raphics *P*rocessing *U*nit

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

课程内容

- a) what to cover & what not to?
- b) GPU Architecture & CPU Architecture ?
- c) CUDA Programming
- d) CUDA Optimizations
- e) advance Topics : MPI + CUDA, Multi-GPU, Clusters, Numerical Algorithms with GPU,OpenCL, OpenGL+CUDA, Cloud Computing

What to cover

- Basic C Programming & Linux OS
- Basic Computer Architecture
- Basic Parallel Computing
- Basic CPU Architecture
- Basic GPU Architecture
- CUDA Programming Language
- CUDA Optimization
- MPI+OpenMP
- Compiling & Debugging

What to cover

- Tools + Third Party Software
  - Compiling
  - Debugging Tools
  - Nsight, CUDA GDB
  - Visual Profiler
- Needed Linux OS

What NOT to cover?

- GPU architecture Details
- Pipeline
- Programmable Vertex and Pixel Processors
- Texture Details
- Low-level drivers

GPU Architecture & CPU Architecture ?

- Basic Information will be enough
- ALU
- Register
- Data Path
- Cache
- Instruction Execution
- SIMD+SIMT
- SM,SMX Info

CUDA Programming

- C Programming Improvements
- Basic Programming Guide
- Cover as much as possible
- No low-level programming
  - Such as PTX, Assemble…
- 10 hours of class

CUDA Programming cont’d

- More Practice based
- Every class is followed by lab work
  - Lab work including whole development process
- Every class is followed by a small programming project
  - Vector Add
  - Parallel reduction
  - Matrix Multiplication
  - And so on…
- Debugging & Profiling are to be added
- Tools
- More TA, more basic programming guidance.

CUDA Optimizations

- Know how and why
- Focusing on bottleneck analysis
- More bandwidth-limited applications
- Coalesced Memory Access
- Using Visual Profiler to find problems
- Emphasize Algorithm Design
- Not focusing on Program Optimization

Advance Topics

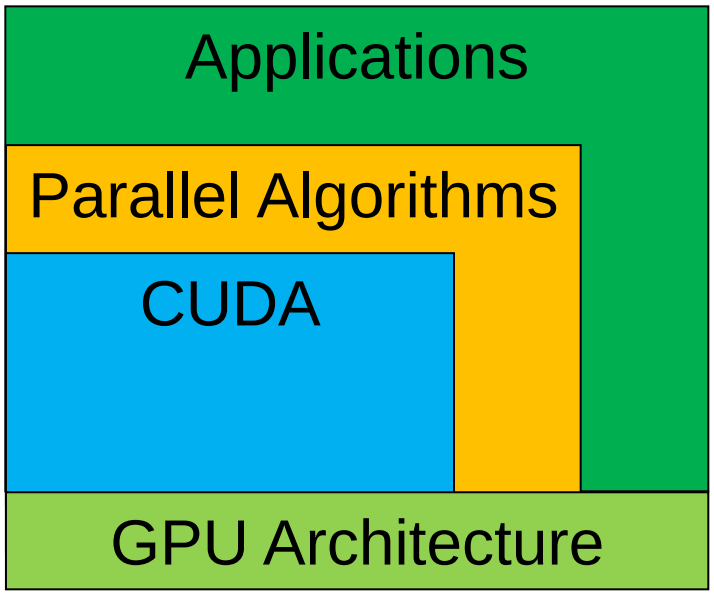
- MPI + CUDA,
- MultiGPU
- Remote GPU Clusters
  - To promote after-class study
- Numerical Algorithms with GPU
  - Numerical Errors
- OpenCL
- OpenGL+CUDA
- Cloud Computing & Grid

Research Fields Related Contents

- Asking Students to participate in teaching
- Students work within 3~5 Groups
- Every group raise a specific topic
- Students give talks
- More research activities involvements

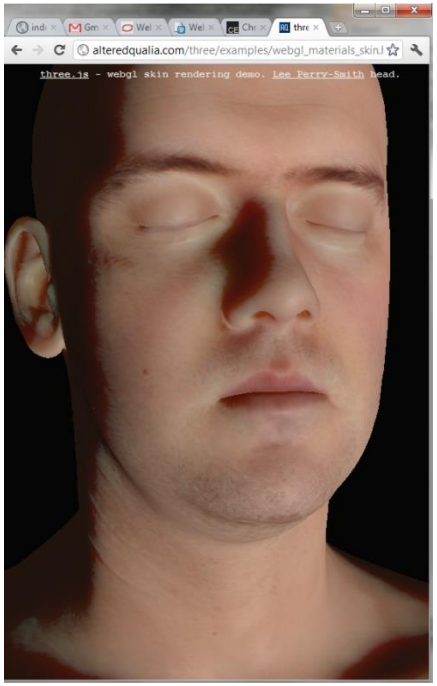
课程内容的关系

- 应用是最关键的
  - Application Driven!!!
- 并行处理:了解架构
- 编程
- 优化



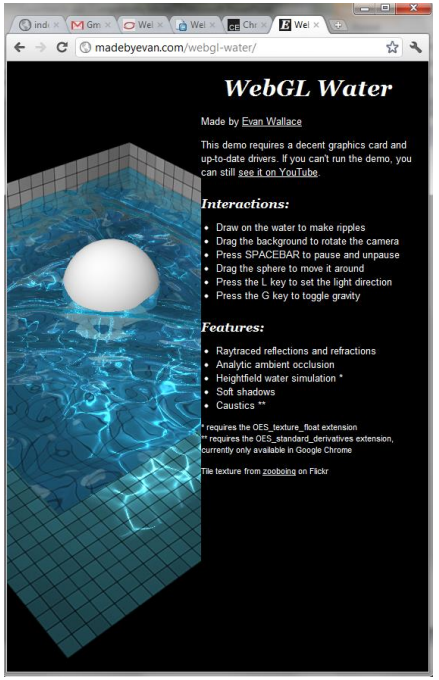
课程内容

- WebGL 展示



WebGL Skin

http://alteredqualia.com/three/examples/webgl_materials_skin.html



WebGL Water

http://madebyevan.com/webgl-water/

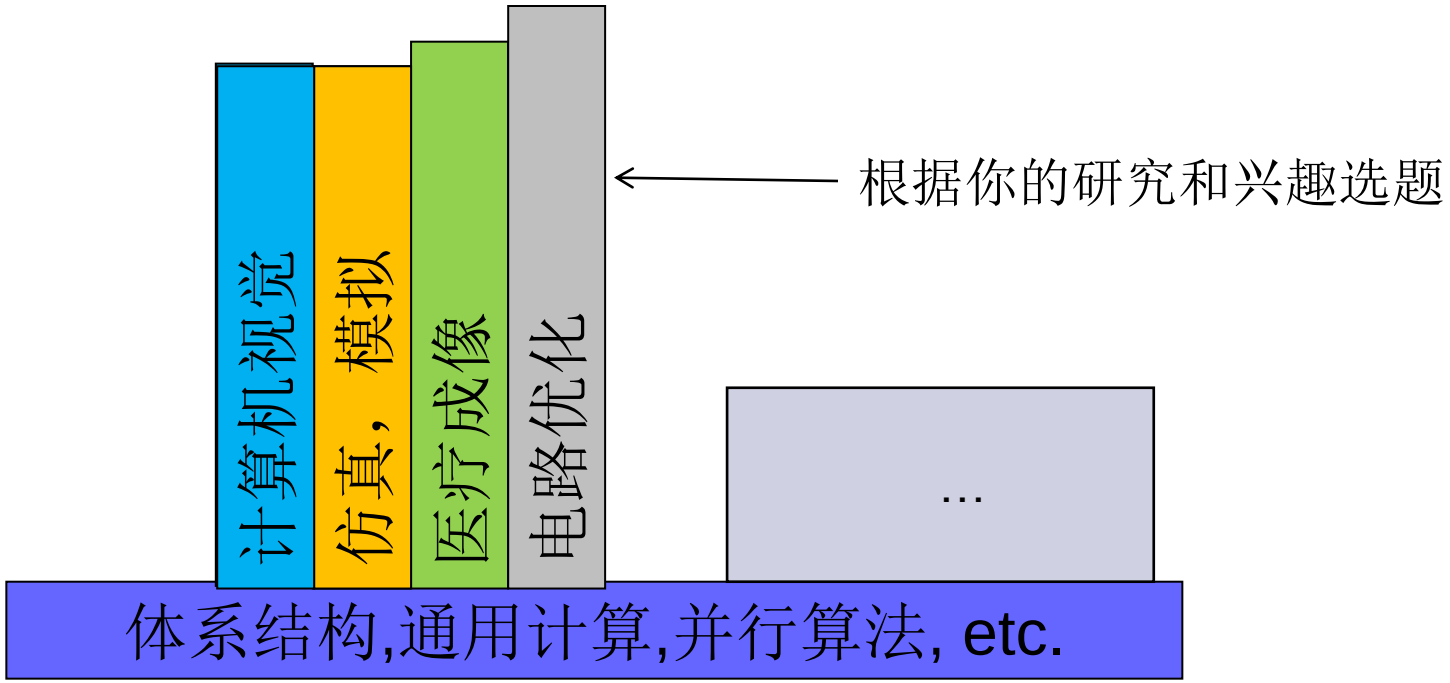
课程内容

- GPU Compute + Real-Time Rendering

http://www.nvidia.com/object/GTX_400_games_demos.html

课程内容

- 大作业:



硬件需求

- 上机实验需要具有 *NVIDIA Fermi* 以上的GPU
- 作业将同样需要,可以到实验室使用
- 更新驱动
  - http://www.nvidia.com/Download/index.aspx
- 如何看GPU型号?

软件和后续学习

- 将学习使用Windows和 Linux 系统
- 喜欢Linux系统
  - 相信科大同学的能力
- 联系助教获得其他实验室资源
- 联系教师获得其他资源
- 实验地点：电四楼 209 实验室