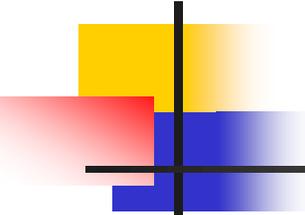


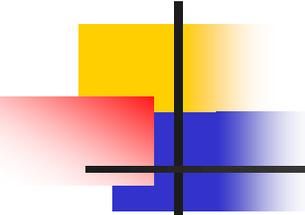
《并行程序设计》课程改革 —基于oneAPI的多架构编程

南开大学 王刚
2023.11



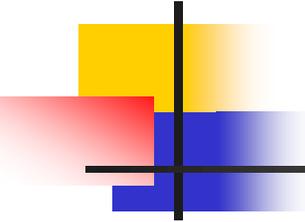
提纲

- 并行程序设计课程教学面临的挑战
- 基于oneAPI的多架构并行编程教学改革
- 南开并行计算教学改革进展



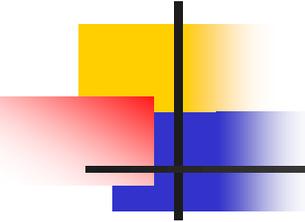
提纲

- 并行程序设计课程教学面临的挑战
- 基于oneAPI的多架构并行编程教学改革
- 南开并行计算教学改革进展



“两个能力培养”

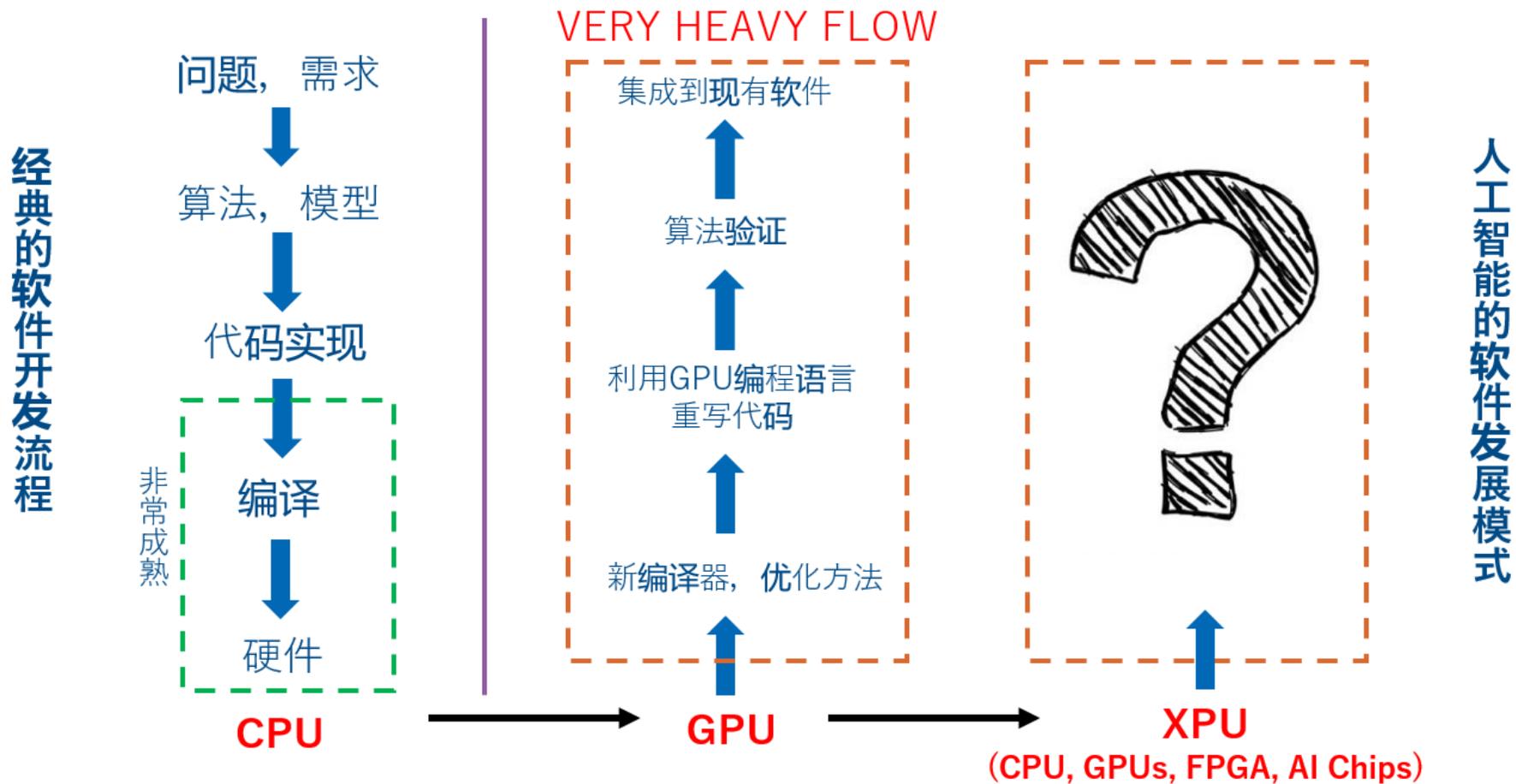
- 系统能力培养：培养构建计算机底层硬件系统和基础软件系统的能力
 - 卡脖子领域
 - 涉及课程：计算机体系结构、计算机组成原理、操作系统、编译系统原理、...
- 问题求解能力培养：培养对实际问题建模、设计算法、分析等完整的求解能力
 - 涉及课程：程序设计、离散数学、数据结构、算法设计与分析、...

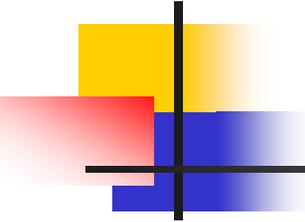


并行程序设计课程的核心价值

- 并行程序设计课兼具两个能力培养任务
 - 并行程序=数据结构+算法+并行体系结构
- 并行程序设计的任务
 - 在理解问题的基础上，设计相应的高效求解问题的（适合特定并行体系结构）数据结构和算法，采用某种适合特定（并行）计算平台的高级语言完成解题方案的编码
追求正确、高效（在并行计算下有更丰富的内涵）和简明

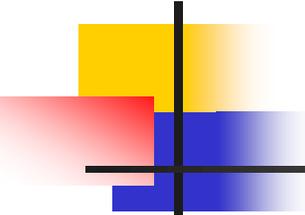
(并行) 软件开发面临的挑战





并行程序设计课程面临的挑战

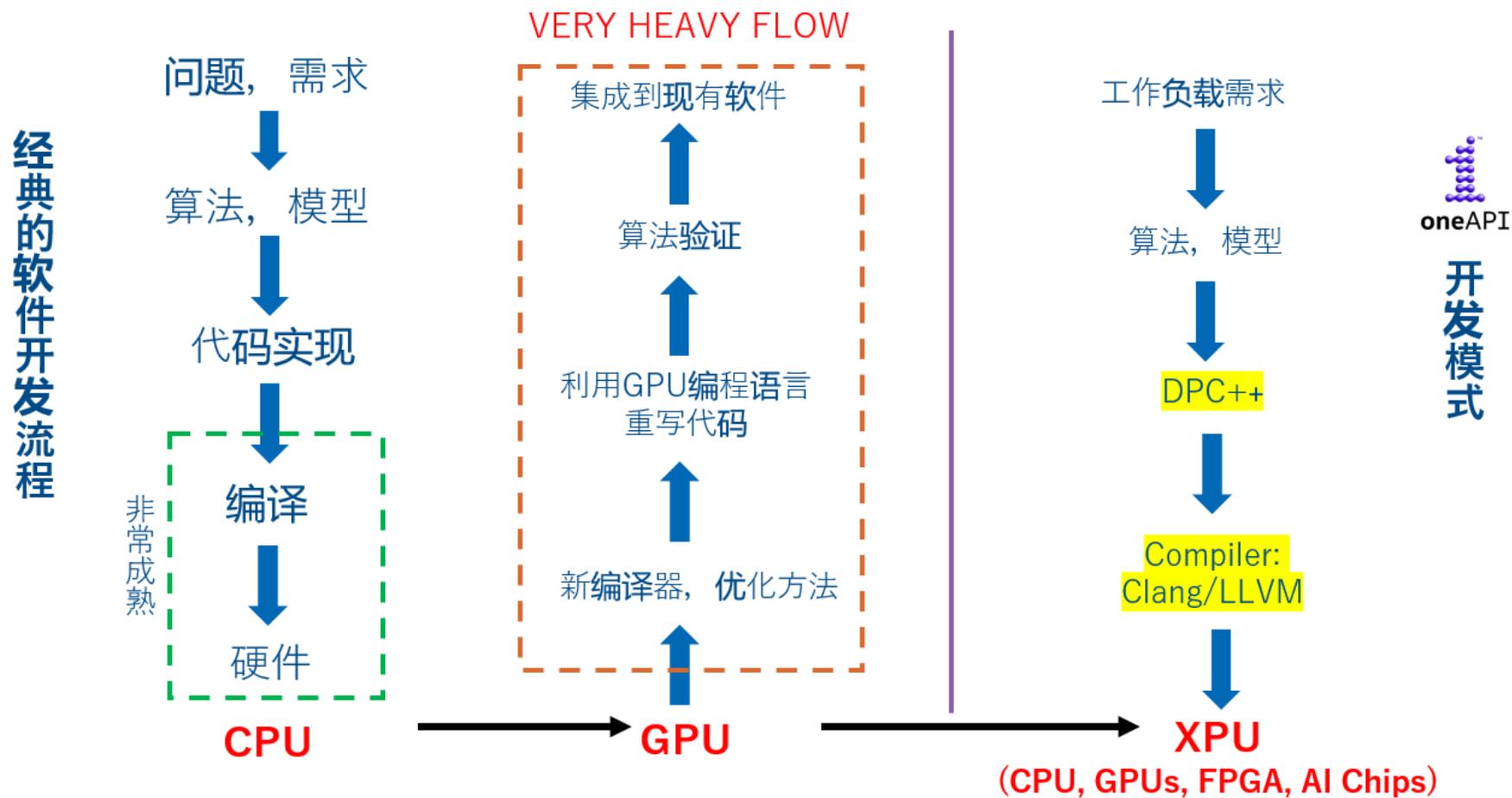
- 大学计算机教育中，还多是基于传统并行编程模型和开发工具构建来并行程序设计课程
- 需要给学生讲授不同架构上的并行编程方法才能适应当前百花齐放的并行架构
 - SIMD编程（X86上的SSE/AVX、ARM上的Neon）、多线程编程（Pthread、OpenMP）、MPI编程、GPU编程（CUDA）、...

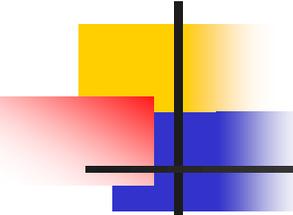


提纲

- 并行程序设计课程教学面临的挑战
- 基于oneAPI的多架构并行编程教学改革
- 南开并行计算教学改革进展

oneAPI提供了一种新的模式





oneAPI提供了新的教改思路

- 并行程序设计课程细化为三方面能力的培养
 - 并行计算系统能力的培养
 - 了解现代并行计算机的组织——并行执行控制机制和互连网络拓扑结构、及其对并行程序的影响
 - 并行程序的正确性、性能与并行体系结构紧密相关
 - 问题并行求解思维能力的培养
 - 提前抛出问题，引导学生思考求解方案——建模、设计编程工具无关/并行平台（部分）无关的抽象数据结构（ADT）和算法流程（操纵ADT）、并行算法复杂性的全面分析
 - 并行编程全流程能力的培养
 - 针对问题并行求解方案，引出具体并行平台及其编程工具特性，讲解ADT和算法正确、高效实现方法
 - 程序调试/调优能力的培养，与算法复杂性理论分析的呼应
 - 开发环境部署能力的培养，实际问题求解中很重要的能力
 - ✓ 基于oneAPI的跨架构并行编程能力！

引入SYCL/DPC++编程教学

DPC++代码解剖

```
void dpcpp_code(int* a, int* b, int* c) {  
    // 设置一个DPC++设备队列  
    queue q;  
    // 为输入和输出向量设置缓冲  
    buffer buf_a(a, range<1>(N));  
    buffer buf_b(b, range<1>(N));  
    buffer buf_c(c, range<1>(N));  
    // 将命令组函数对象提交给队列  
    q.submit([&](handler &h){  
        // 创建设备访问器，以访问在全局内存中分配的缓冲  
        accessor A(buf_a, h, read_only);  
        accessor B(buf_b, h, read_only);  
        accessor C(buf_c, h, write_only);  
        // 指定设备kernel体为一个Lambda函数  
        h.parallel_for(range<1>(N), [=](auto i){  
            C[i] = A[i] + B[i];  
        });  
    });  
}
```

Kernel调用是并行执行的

对范围中的每个元素调用kernel

Kernel调用可以访问调用id

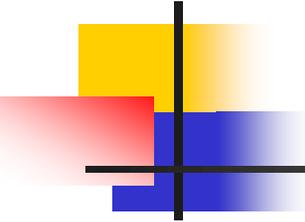
Step 1: 创建一个设备队列
(可通过设备选择器或使用默认选择器来指定一个设备类型)

Step 2: 创建缓冲 (既表示主机端内存、也表示设备端内存)

Step 4: 创建缓冲访问器，以访问设备上的缓冲数据

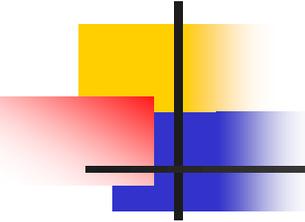
Step 5: 发送一个kernel (lambda) 去执行

Step 6: 编写一个kernel



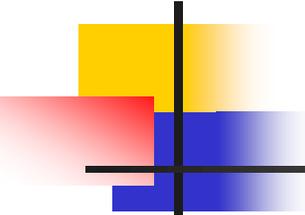
多架构编程对比教学

- 不同架构下SYCL/DPC++程序优化，以及oneAPI进阶编程知识
 - SIMD: AVX-512、OpenMP: 任务卸载、...
- 不同架构下，传统编程模式与oneAPI多架构统一编程模型的对比
- 使学生了解不同架构下并行算法/并行程序设计与优化技术的同时，掌握高效的多架构统一编程方法



实验教学改革

- 安装部署oneAPI开发环境（以及借助英特尔的DevCloud云端教学平台）
- 期末研究报告选题范例：基于oneAPI
 - 不同架构下传统编程模式与oneAPI编程模型的对比：编程工作量、目标程序性能、...
 - oneAPI开发生态的研究：代码编写辅助、辅助库、分析调试工具、...
 - 基于oneAPI的并行优化研究题目



提纲

- 并行程序设计课程教学面临的挑战
- 基于oneAPI的多架构并行编程教学改革
- 南开并行计算教学改革进展

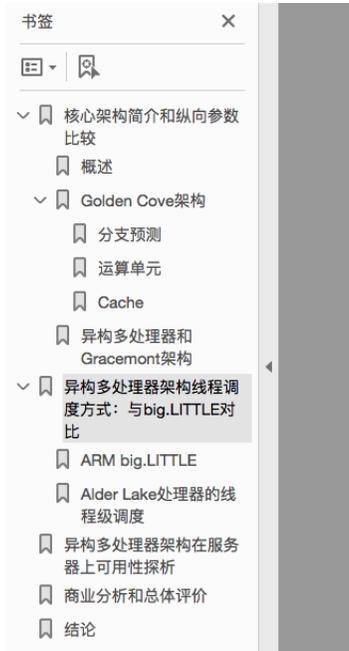
南开并行计算教学发展

- 90年代就开设研究生《并行计算》课程
- 2005年左右就引入GPGPU教学内容
 - 学研结合：生物信息学、搜索引擎、存储系统、密码学中的算法并行优化
 - 第一届CUDA校园大赛程序设计大赛第一名
- 2016年开设本科生《并行程序设计》课程
- 2020年获天津市工程专业学位优秀课程



与英特尔教改合作初见成效

- 2021年、2022年、2003参加oneAPI技术讲座、完成DPC++线上课程学生超过300人
- 在DevCloud上采用oneAPI完成编程作业和期末研究报告千余人次



计算机学院
并行程序设计作业报告

Intel Core 12 代 (Alder Lake):
针对消费级市场的异构多处理器优化

姓名：冯朝芑
学号：2012039
专业：计算机科学与技术

作业/期末研究报告选题示例

○ SIMD编程：AVX-512

书签

- 普通高斯消元
 - 实验简介
 - 实验设计
- arm平台--neon
- X86平台--sse、avx、avx512**
- 实验总结与其他优化尝试
- 基于Gröbner基计算的特殊高斯消元
- 总结

一、普通高斯消元

并行程序设计实验报告

n	串行耗时/s	sse 耗时/s	avx 耗时/s	avx512 耗时/s	串行/sse	串行/avx	串行/avx512
64	0.000403	0.000192	0.000151	0.000156	2.099	2.667	2.579
128	0.00327	0.00144	0.000983	0.000799	2.279	3.327	4.094
256	0.026	0.011	0.00713	0.00476	2.354	3.643	5.459
512	0.206	0.0854	0.0541	0.032	2.416	3.815	6.435
1024	1.676	0.705	0.429	0.237	2.378	3.901	7.064
2048	13.373	5.563	3.364	1.926	2.404	3.976	6.942

表 3: X86 平台非对齐性能测试结果

利用表中的数据做曲线图，如图4所示，串行（serial）用 S 来表示。分析图与表中的数据可以发现，并行化的效果为 $avx512 > avx > sse$ 。其中，sse 算法的效能问题规模 n 较小时随规模增大略有提高，之后趋于稳定，与串行算法的耗时比约在 2.3、2.4 左右。avx 算法的效能随问题规模的增大要强于 sse 算法，同样是先增大后稳定，最后耗时比可达到 4。avx512 算法受问题规模影响最大，在 n 较小时性能的变化最为显著，在 n 较大时可达到的稳定耗时比也最高，为 7 左右。图中两条红色竖线分别代表 L1 cache、L2 cache 对应的问题规模，可以发现 L1 cache 填满前，曲线大幅上扬，优化效果明显提升；在填满 L1 cache、未填满 L2 cache 时，曲线缓慢

作业/期末研究报告选题示例

○ Pthread编程：火焰图分析

书签

- 概述
- 单层卷积并行化
 - 数据划分
 - 任务划分：划分卷积过滤器（特征划分）
 - x86与arm平台程序性能对比与总结
 - 单层卷积火焰图分析**
 - x86平台对比
 - 不同平台下mutex代价测量实验
 - 其他卷积层参数作自变量的性能测量
- 多层卷积层并行
 - 池化
 - 感想与总结

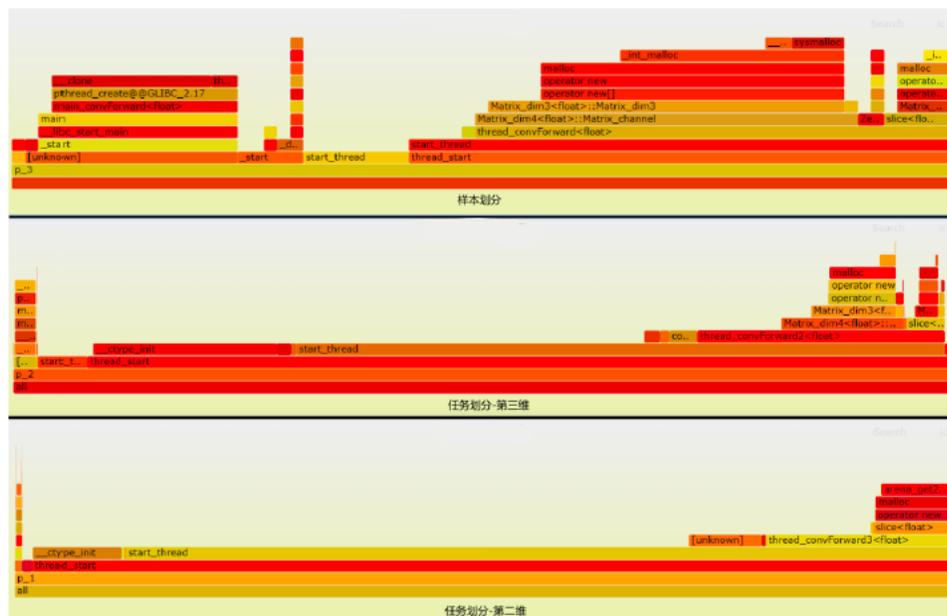


图 2.9: 循环划分火焰图

火焰图纵轴表示调用栈，调用栈越深，火焰就越高，函数下方为其父函数；横轴表示抽样数，如果一个函数在 x 轴占据的宽度越宽，则表示其执行时间长。一般而言，如果火焰图中，有“平顶”（plateaus）现象，就表示呈现平顶的函数可能存在性能问题。事实上，程序实际性能也是平顶越不明显，性能越

作业/期末研究报告选题示例

○ Pthread编程：x86 vs. arm mutex性能分析

书签



概述

单层卷积并行化

数据划分

任务划分：划分卷积过滤器（特征划分）

x86与arm平台程序性能对比与总结

单层卷积火焰图分析

x86平台对比

不同平台下mutex代价测量实验

其他卷积层参数作自变量的性能测量

多层卷积层并行

池化

感想与总结

如图 2.14 所示，结果发现两个平台的互斥量核心实现逻辑大抵是相似的，而互斥量在上锁时也不只是记录锁的状态，同时也通过 mutex 记录了持有锁的线程；利用 Code:Blocks 进行 Debug，可以在 CPU register 中查询到对应寄存器数值：x86 平台在 lock 时将 r8 中保存线程标识（起始地址）赋值给 mutex，在解锁时通过比对 mutex 的数值与试图解锁的线程标识是否相符，相符则开锁利用 xchg 完成原子交换，arm 平台则对应用 mrs, tpidr_el0 完成线程切换，这也符合互斥量只有持锁者才可开锁的设置。

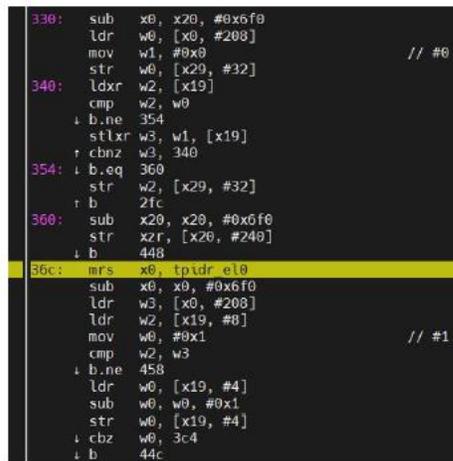
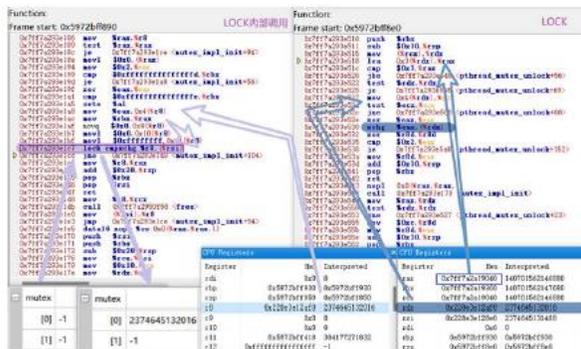


图 2.15: x86 平台与 arm 平台 Mutex 汇编语句对比

作业/期末研究报告选题示例

○ 期末研究报告：DPC++加速高斯消去



(四) 实验结果

将 CUDA 移植后的 DPC++ 代码在本地（笔记本电脑）进行测试，结果如表5。
(Device:) Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz 集显

测试样例	7	8	9
串行算法一/ms	23.54	297.85	807.32
串行算法二/ms	41.41	484.92	1741.74
基于算法二的 CUDA 算法/ms	14.19	140.56	398.83
由 CUDA 算法移植的 DPC++ 算法/ms	26.86	383.82	663.09

表 5: OneAPI 实验结果

从实验结果发现，在各测试样例下，由 CUDA 移植后的 DPC++ 程序都要比原 CUDA 用时长，并且 DPC++ 程序要比 CUDA 程序慢一倍。

分析其原因，我们发现，DPC++ 程序使用的设备是 Intel 集显，而 CUDA 所用的是笔记本电脑上的 NVIDIA 独立显卡。因此其性能会有一定差异。但由于实验设备有限（只有笔记本电脑），我们暂时无法在其他设备上程序测试。

作业/期末研究报告选题示例

○ 期末研究报告：DPC++加速一些算法



在调用规约 API 时，首先要创建计算结果的存储位置，并创建规约对象。

```
1  std::pair<int, int> operator_identity = {  
2      std::numeric_limits<int>::max(),  
3      std::numeric_limits<int>::min()  
4  };  
5  buffer<decltype(operator_identity)> result_buf(1);  
6
```

31 of 36

7 使用 ONEAPI 设计其他并行算法

```
7  auto reduction_object = reduction(  
8      result_buf,  
9      h,
```

英特尔oneAPI创新大使征文

- 组织学生参加2022年《英特尔oneAPI创新大使征文活动》，2人获得一等奖，2人获得二等奖

2022/5/26 13:18

英特尔oneAPI—高斯消元算法并行化_zouxianyu的博客-CSDN博客



CSDN

博客

下载·课程

学习 PK

社区

认证

开源

云服务 新

oneapi

搜索

邹先予

英特尔oneAPI—高斯消元算法并行化

原创

zouxianyu



于 2022-05-15 22:14:21 发布



57



收藏

版权

文章标签：

oneapi

孟笑朵

英特尔 oneAPI —— 实现COO格式稀疏矩阵乘积openMP并行化

原创

neverstop_xd



于 2022-05-17 23:53:36 发布



100



收藏

文章标签：

oneapi

英特尔oneAPI创新大使征文

- 组织学生参加2022年《英特尔oneAPI创新大使征文活动》，2人获得一等奖，2人获得二等奖

熊宇轩

英特尔oneAPI - 使用parallel_for完成高斯消去的并行

化

原创 suhipek 已于 2022-05-17 21:34:23 修改 58 收藏 1 版权
文章标签: oneapi 矩阵 算法

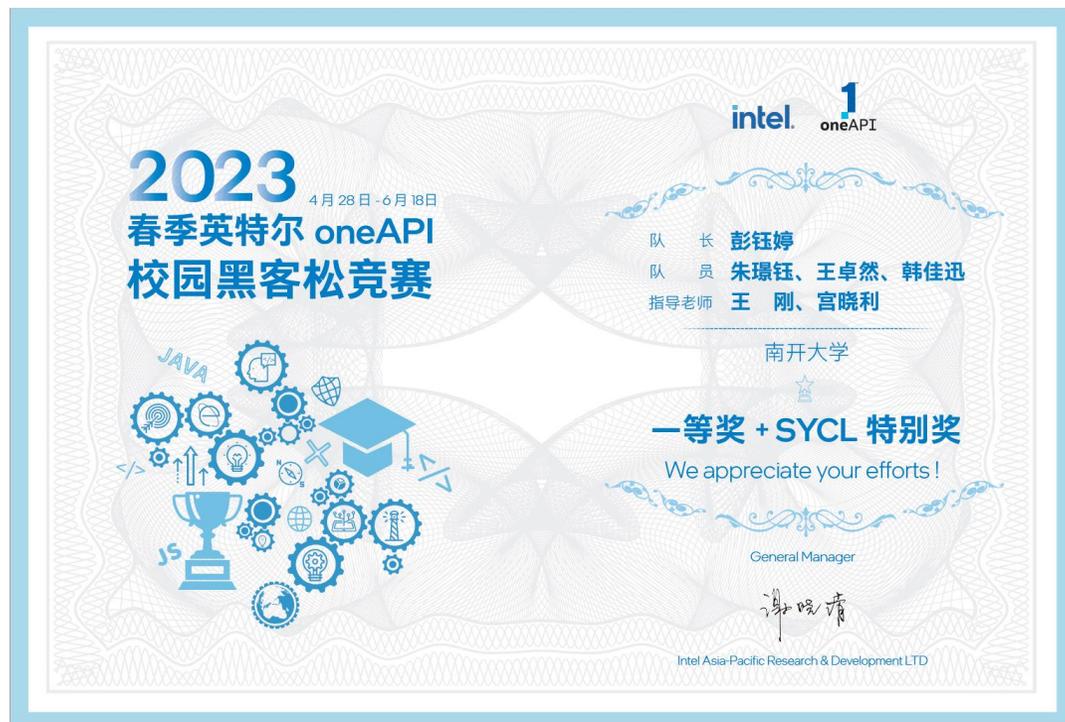
英特尔 oneAPI

戴廷钧

用 AI Analytics Toolkit 实现强化学习

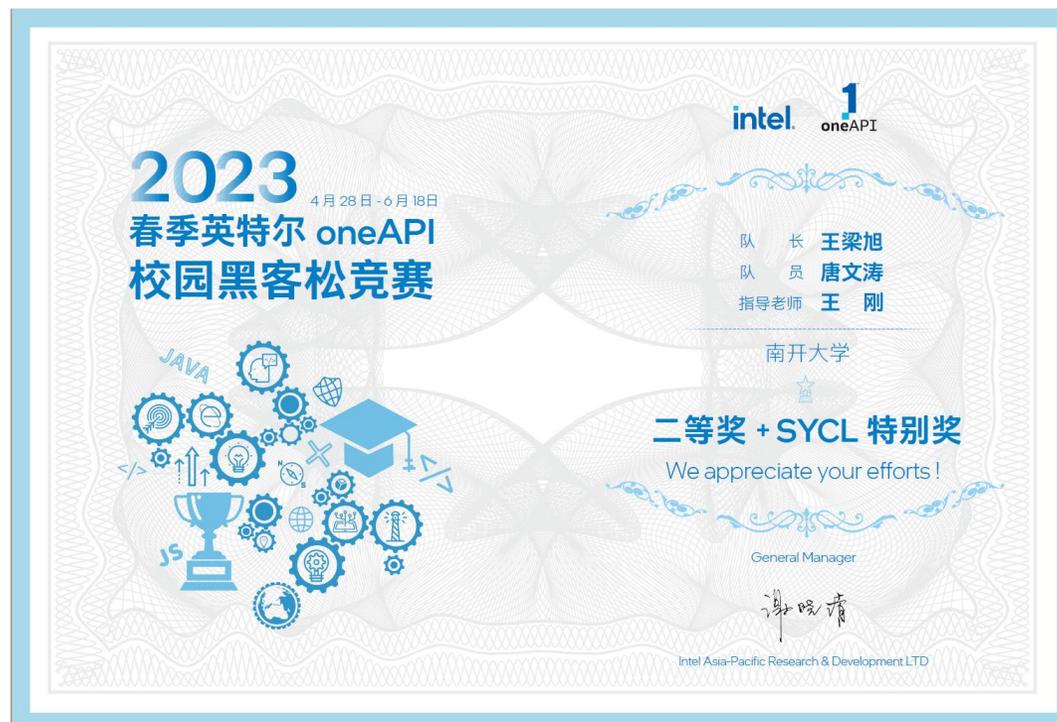
英特尔oneAPI校园黑客松

- 组织学生参加2023年《英特尔oneAPI校园黑客松》，获得1个一等奖、1个二等奖、1个优秀分享奖以及2个SYCL特别奖



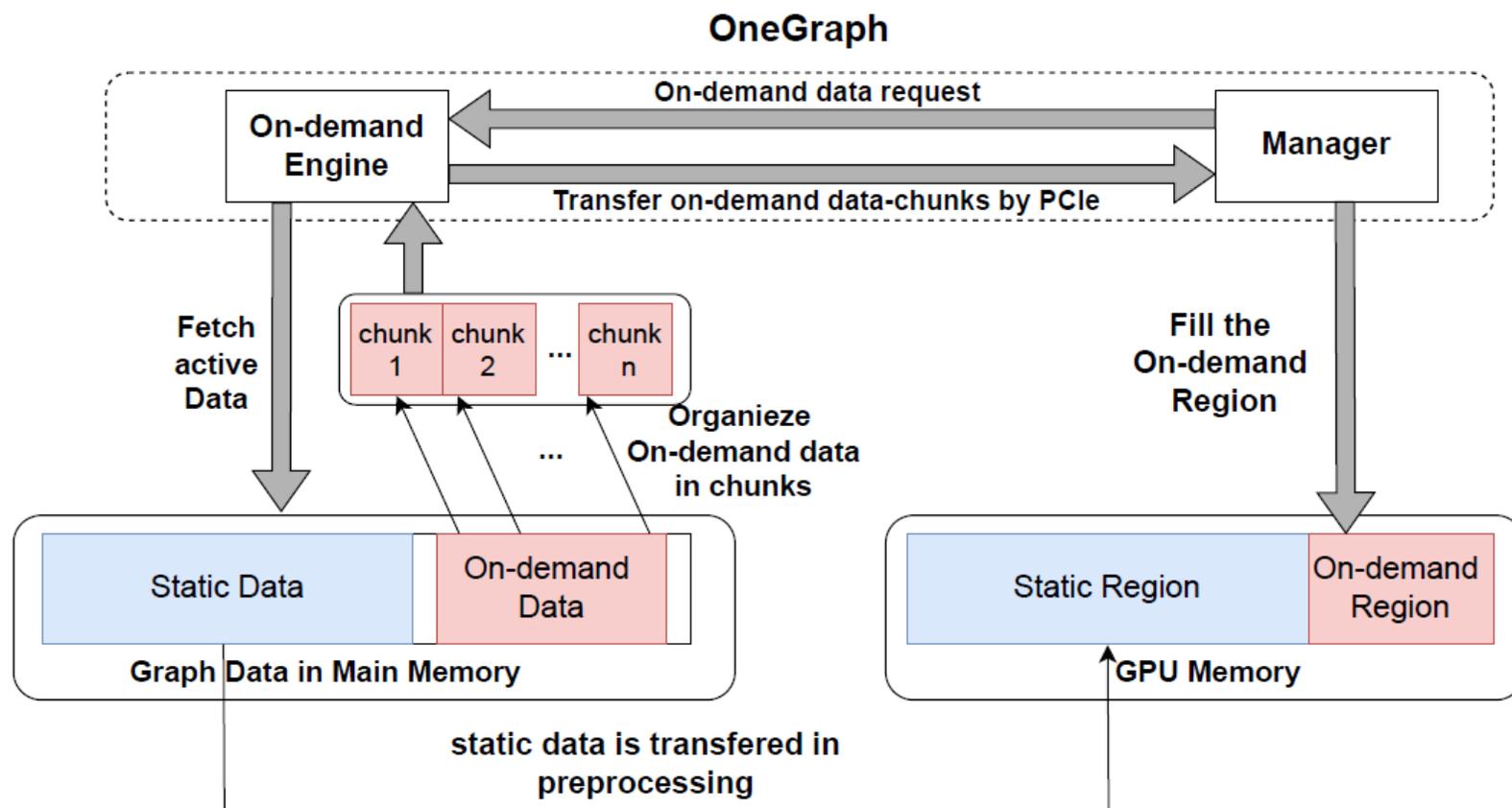
英特尔oneAPI校园黑客松

- 组织学生参加2023年《英特尔oneAPI校园黑客松》，获得1个一等奖、1个二等奖、1个优秀分享奖以及2个SYCL特别奖



基于oneAPI的大规模图计算异构加速框架设计

- 已被《CCF Transactions on High Performance Computing》录用



助教团队的传帮带作用

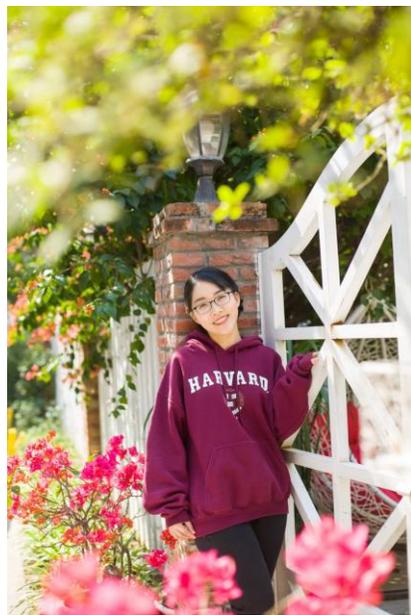
- 完成实验教学指导书修订、实验平台搭建、辅导上机、批改作业等任务



周浩



安祺



麦隽韵



李世阳



曾泉胜

助教团队的传帮带作用

- 完成实验教学指导书修订、实验平台搭建、辅导上机、批改作业等任务



徐一帆



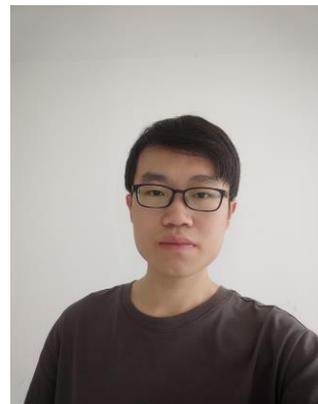
陈静怡



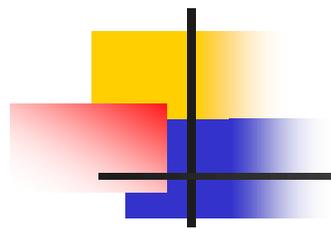
贺祎昕



孙辉



李君龙



THANKS!