



北京大学
PEKING UNIVERSITY

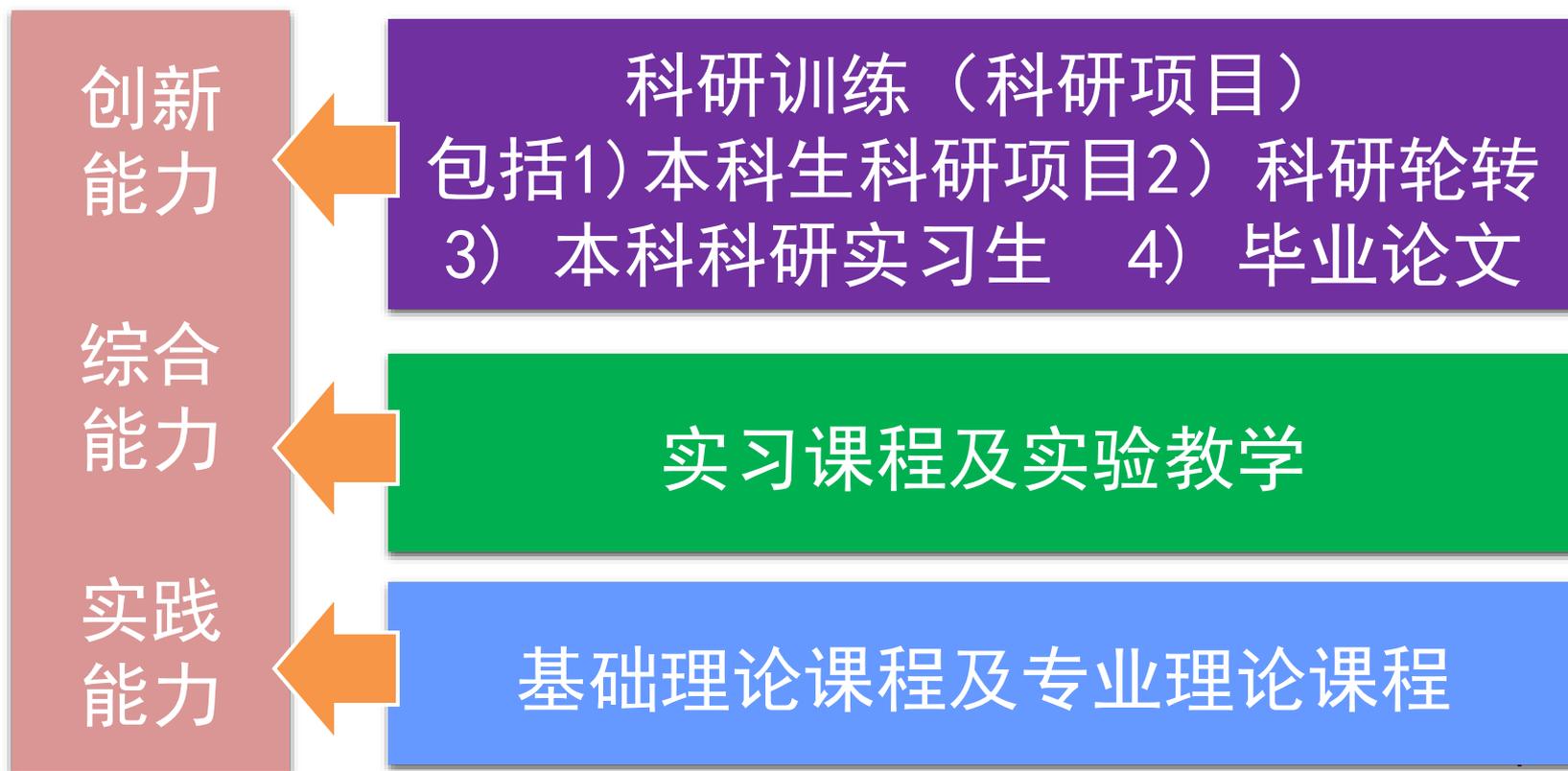
关于编译原理等系统类课程 教学的若干调整性思考

北京大学计算机科学技术系 刘先华

2020年11月20日 · 重庆

北大计算机系学生实践与能力训练

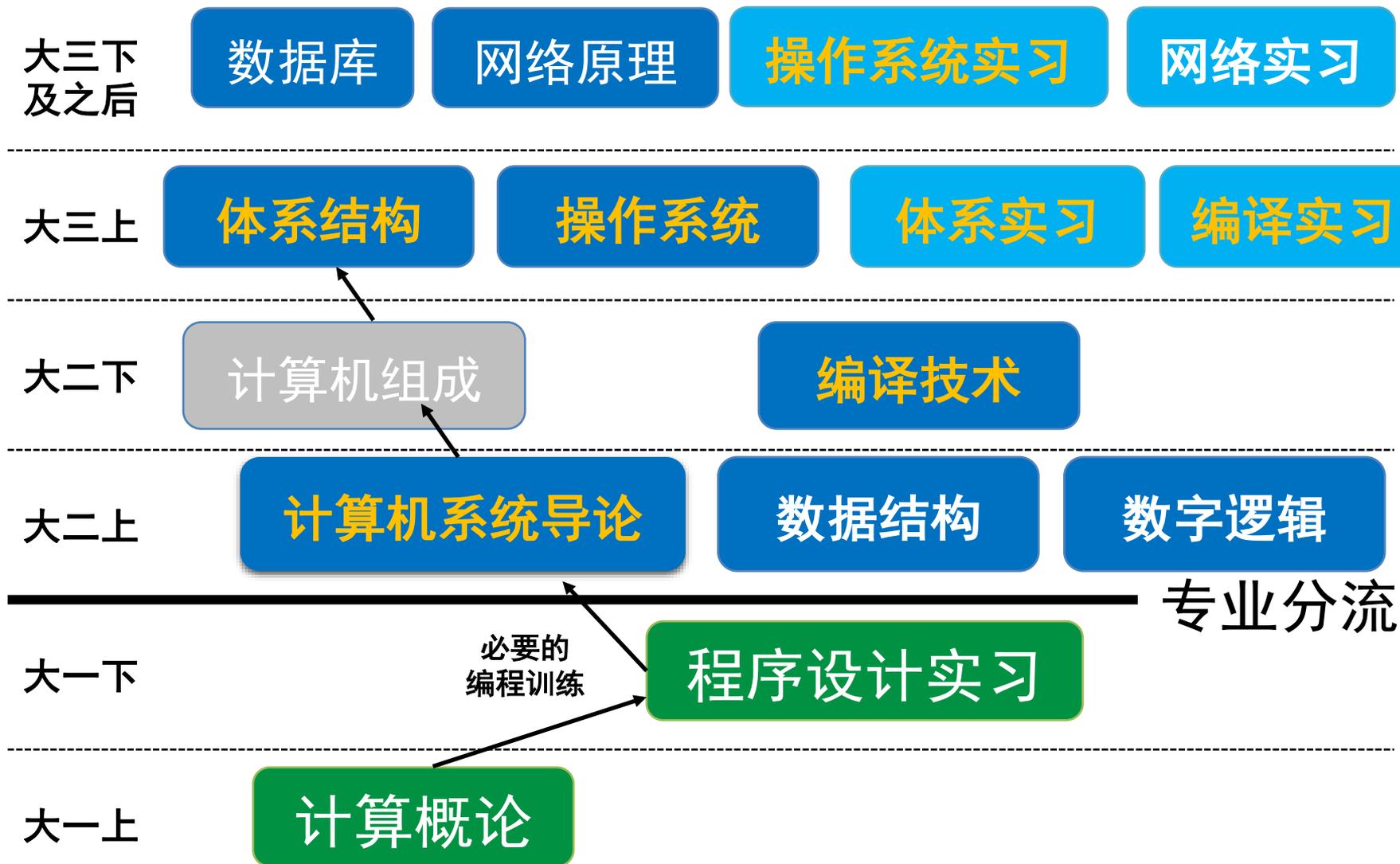
北大计算机系建立了一整套实践与能力训练的制度，将实践与能力训练与本科生课程训练并列本科生培养的两个组成部分。实践课程是其中的一个重要组成部分。



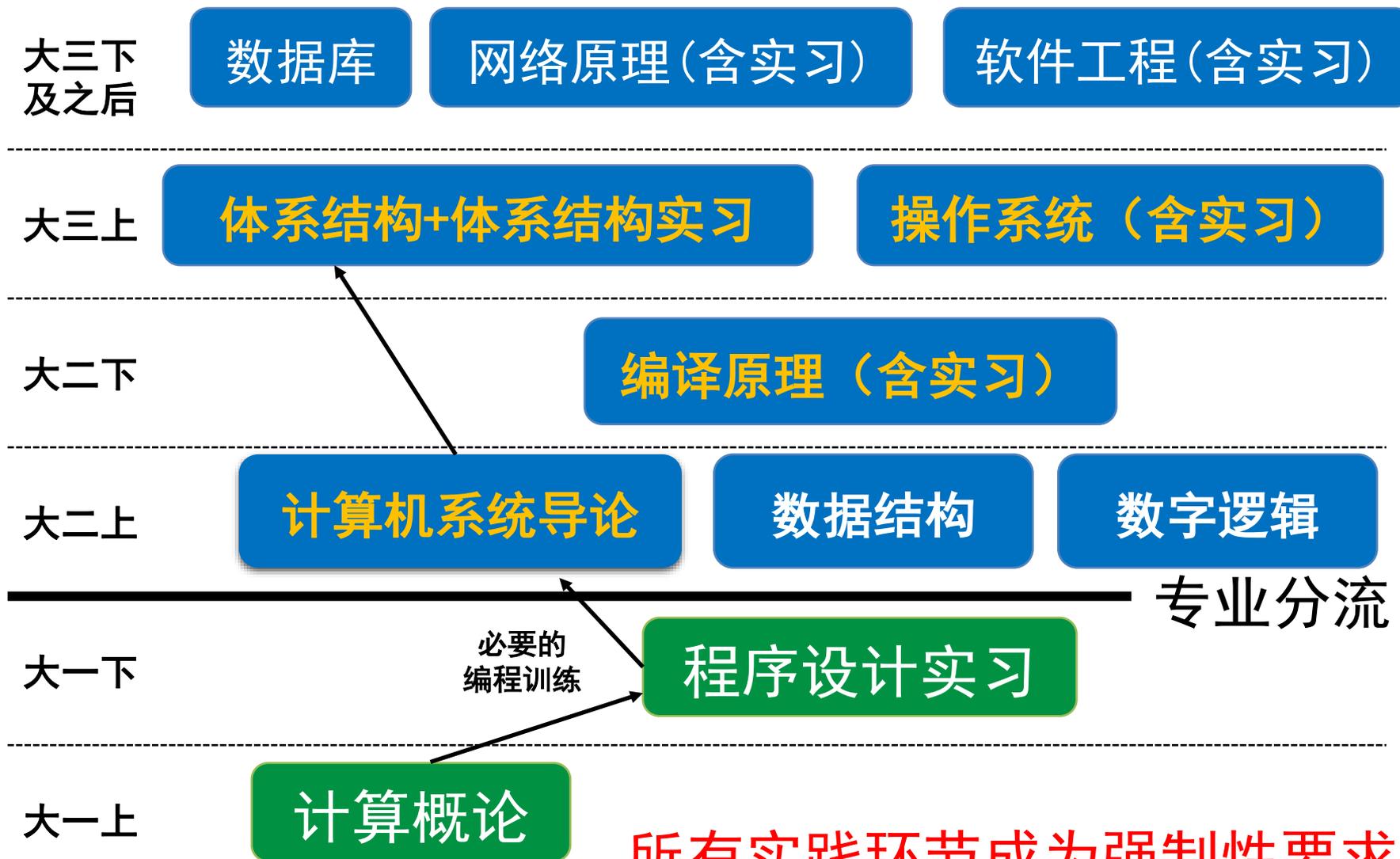
系统类课程中实践环节的重要性

- “欲诚其意者，先致其知；**致知在格物**。物格而后知至，知至而后意诚”——《大学》
- “不闻不若闻之，闻之不若见之，见之不若知之，**知之不若行之。学至於行之而止矣。行之，明也。**”——《荀子·儒效》
- 古人学问无遗力，少壮工夫老始成。**纸上得来终觉浅，绝知此事要躬行。**
——陆游《冬夜读书示子聿》

北大计算机系课程要求（部分）

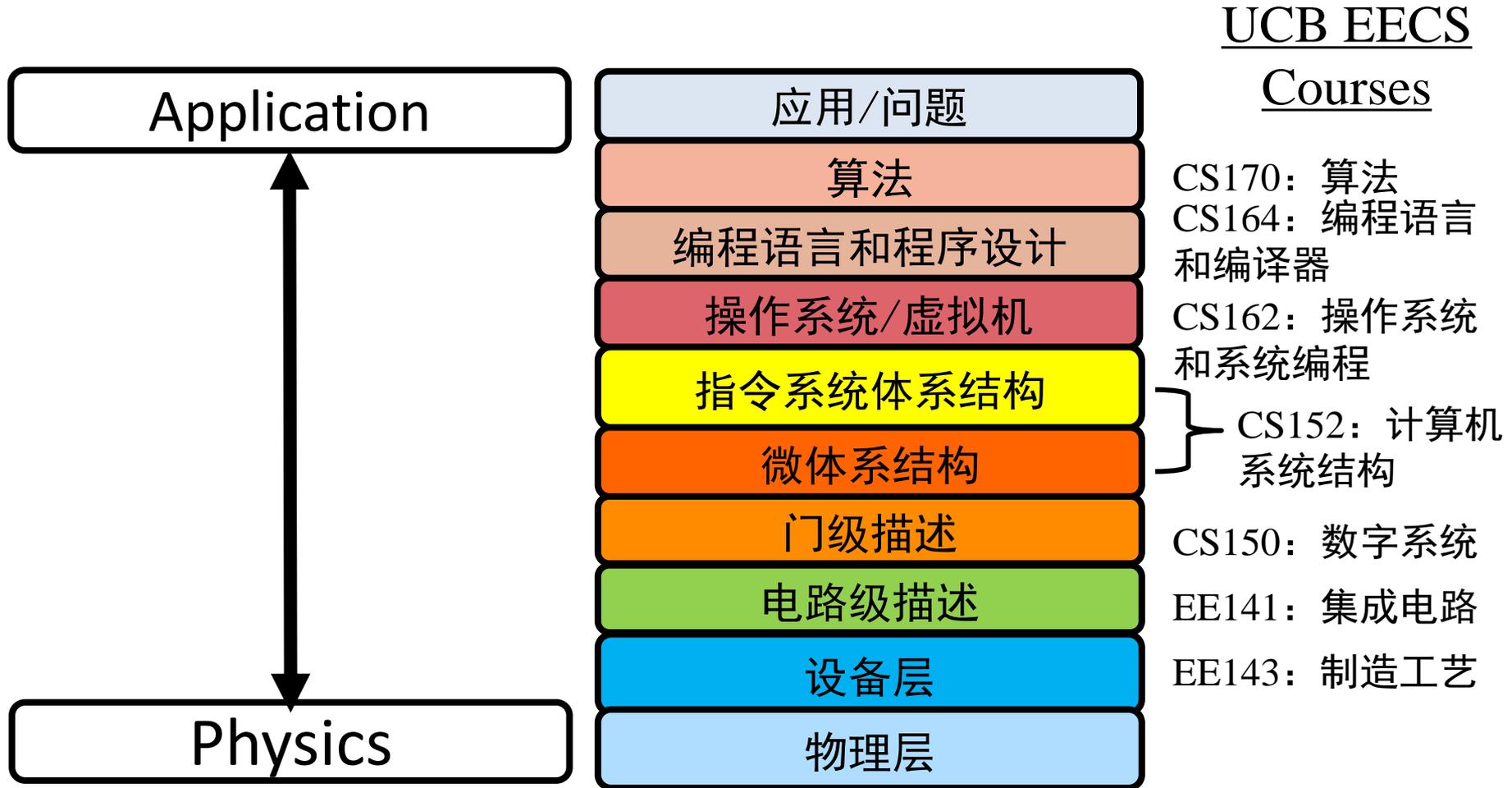


北大计算机系课程要求（部分）



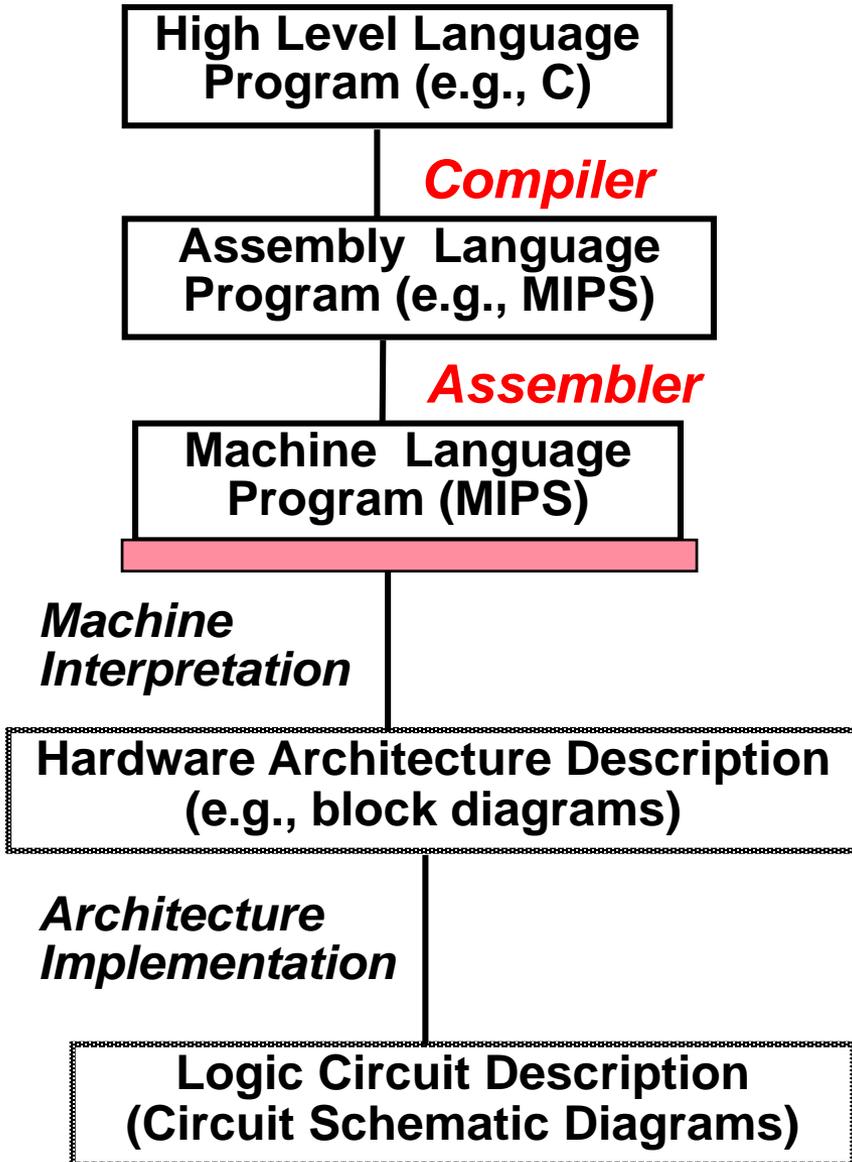
所有实践环节成为强制性要求

计算机系统设计中的层次观



Gap too large to bridge in one step

计算机系统导论

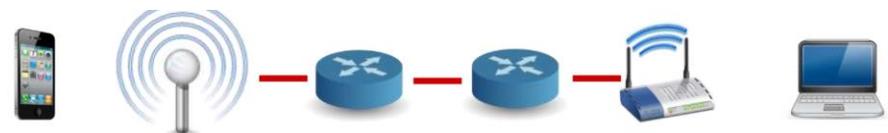
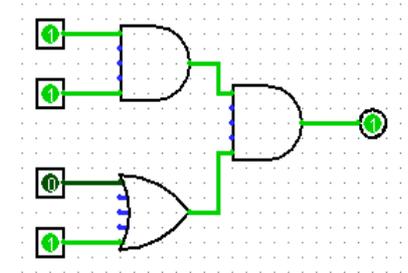
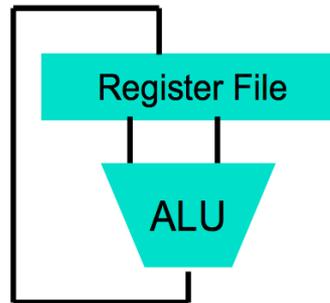


```
temp = v[k];
v[k] = v[k+1];
v[k+1] = temp;
```

```
lw $t0, 0($2)
lw $t1, 4($2)
sw $t1, 0($2)
sw $t0, 4($2)
```

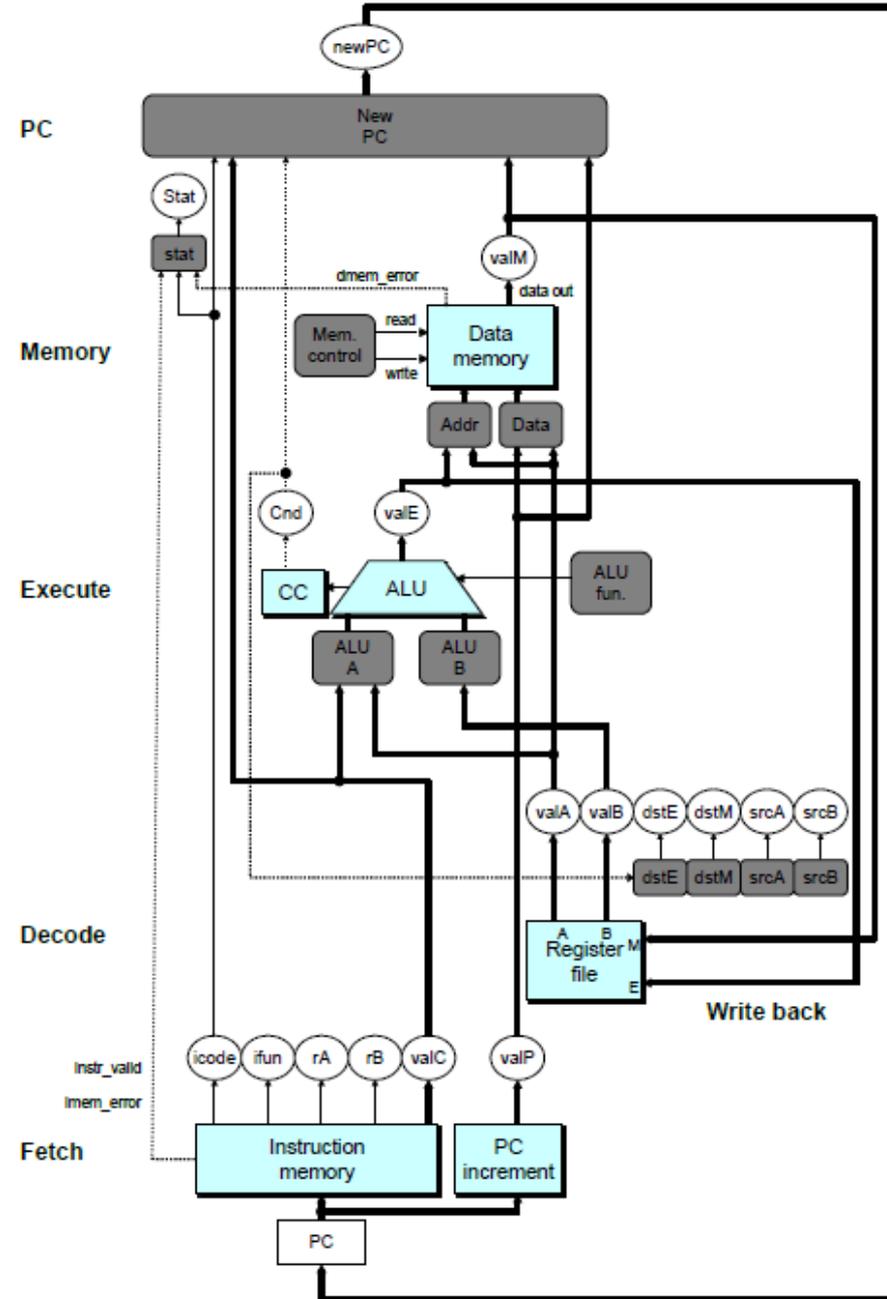
Anything can be represented
as a *number*,
i.e., data or instructions

```
0000 1001 1100 0110 1010 1111 0101 1000
1010 1111 0101 1000 0000 1001 1100 0110
1100 0110 1010 1111 0101 1000 0000 1001
0101 1000 0000 1001 1100 0110 1010 1111
```



ICS中的HCL

		OPq rA, rB
Fetch	icode,ifun	icode:ifun $\leftarrow M_1[PC]$
	rA,rB	rA:rB $\leftarrow M_1[PC+1]$
	valC	
	valP	valP $\leftarrow PC+2$
Decode	valA, srcA	valA $\leftarrow R[rA]$
	valB, srcB	valB $\leftarrow R[rB]$
Execute	valE	valE $\leftarrow valB \text{ OP } valA$
	Cond code	Set CC
Memory	valM	
Write back	dstE	R[rB] $\leftarrow valE$
	dstM	
PC update	PC	PC $\leftarrow valP$



ICS中的HCL2C和HCL2V

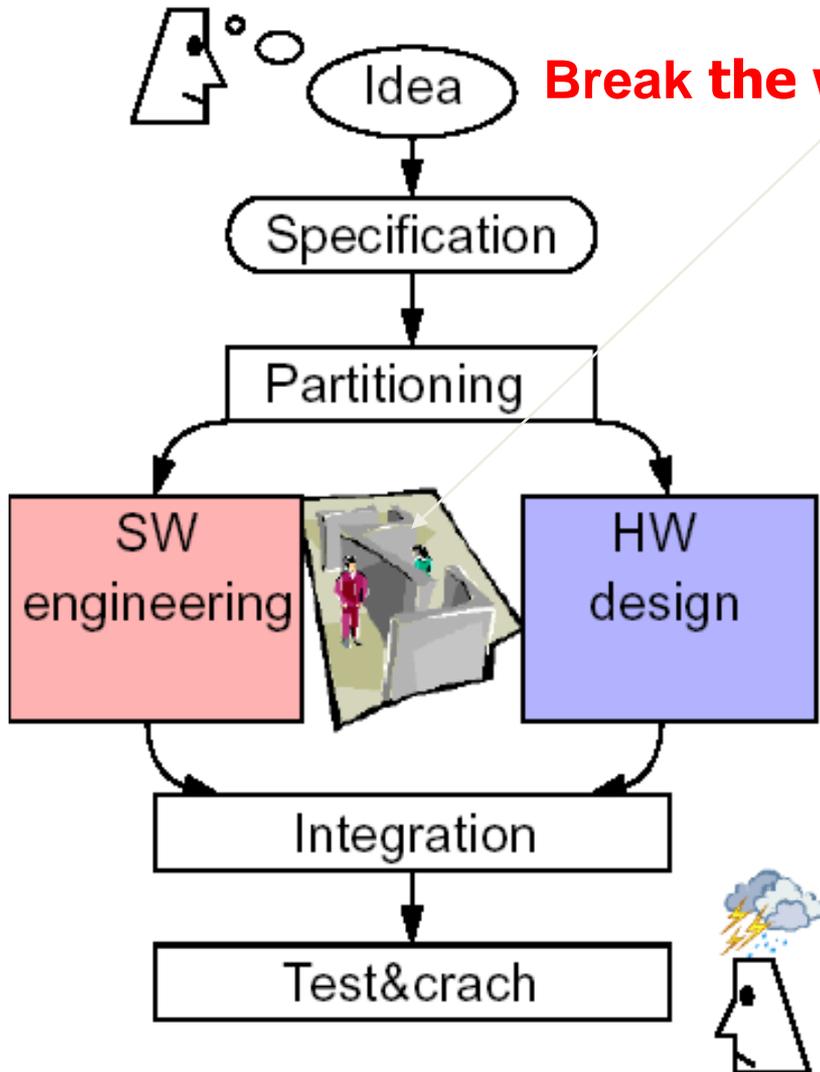
```
statements: /* empty */
  | statements statement
  ;
```

```
statement:
  QUOTE QSTRING { insert_code($2); }
  | BOOLARG VAR QSTRING { add_arg($2, $3, 1); }
  | WORDARG VAR QSTRING { add_arg($2, $3, 0); }
  | BOOL VAR ASSIGN expr SEMI { gen_func($2, $4, 1); }
  | WORD VAR ASSIGN expr SEMI { gen_func($2, $4, 0); }
  ;
```

```
expr:
  VAR { $$=$1; }
  | NUM { $$=$1; }
  | LPAREN expr RPAREN { $$=$2; }
  | NOT expr { $$=make_not($2); }
  | expr AND expr { $$=make_and($1, $3); }
  | expr OR expr { $$=make_or($1, $3); }
  | expr COMP expr { $$=make_comp($2,$1,$3); }
  | expr IN LBRACE exprlist RBRACE { $$=make_ele($1, $4); }
  | LBRACK caselist RBRACK { $$=$2; }
  .
```

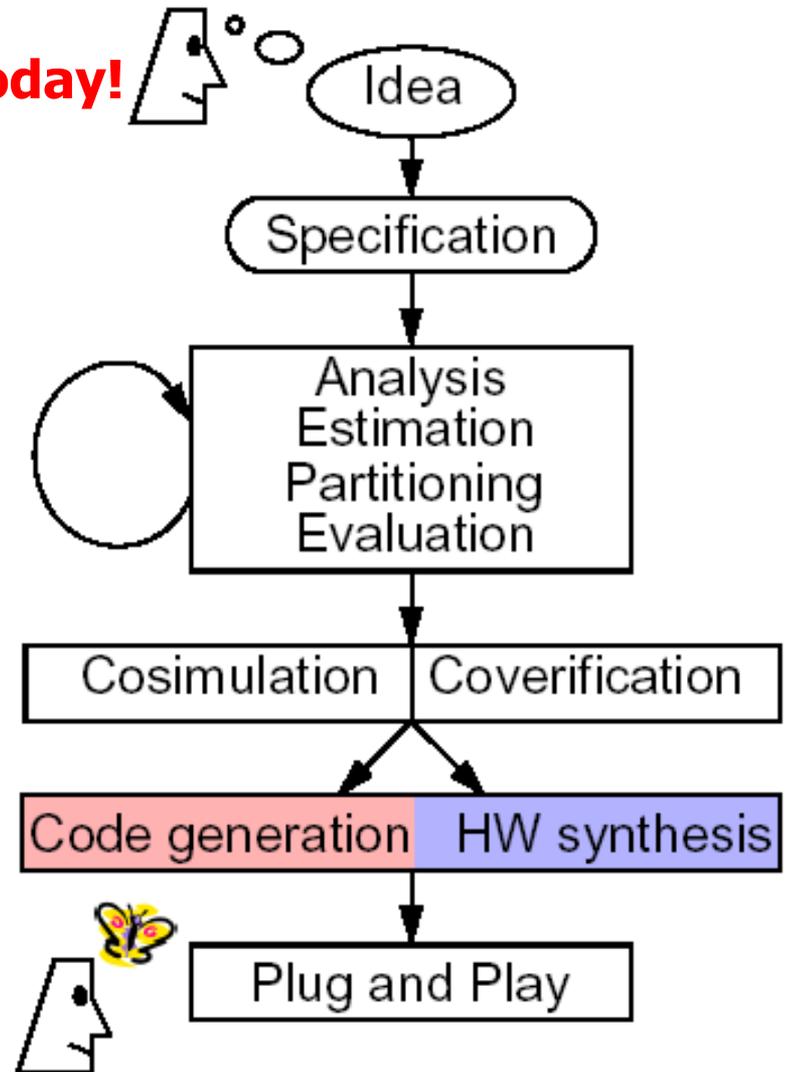
设计流程的变化

Good old way today



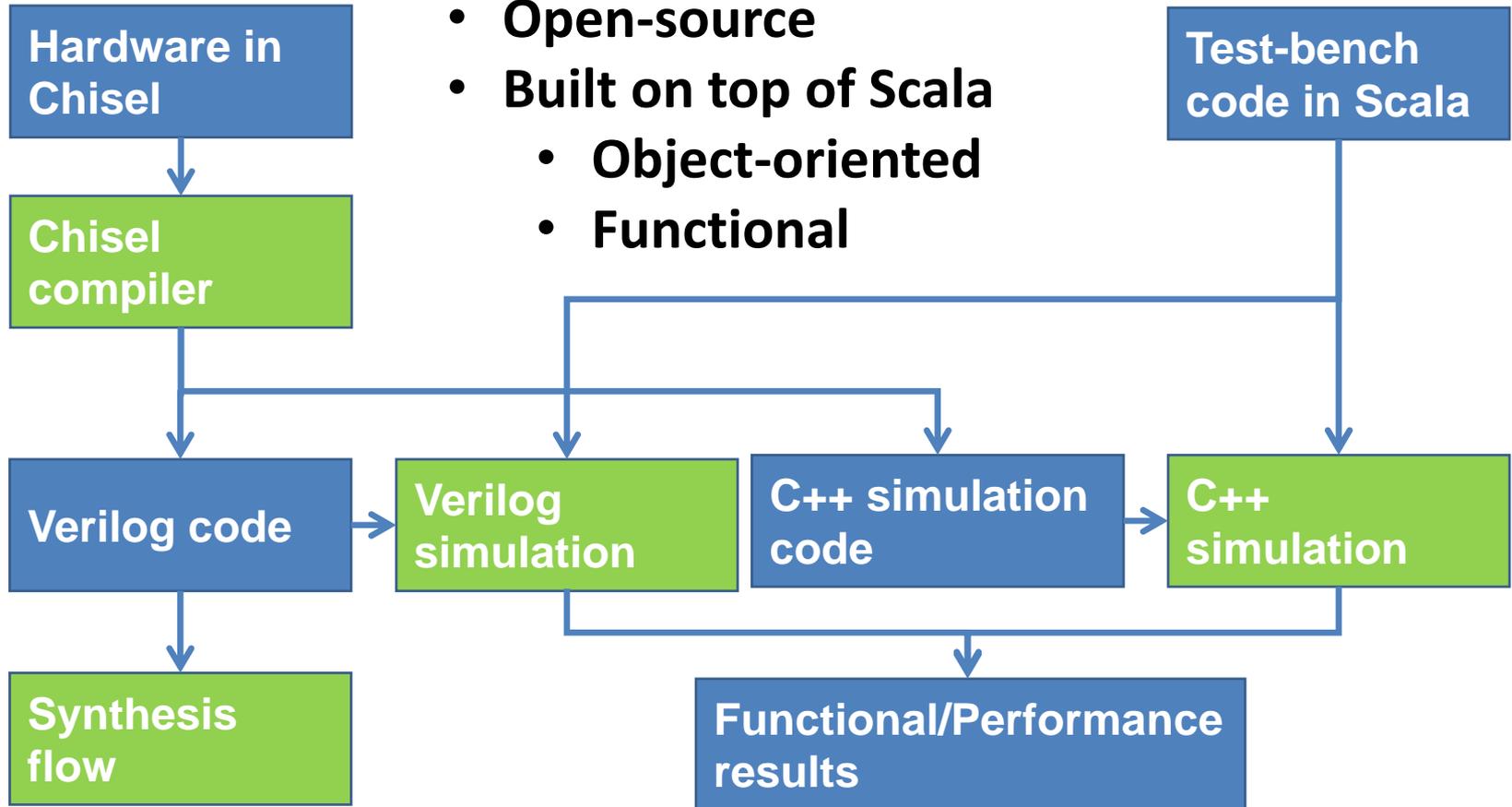
Break the wall today!

Contemporary way



Chisel @ UC Berkeley

- Developed @ UC Berkeley
- Open-source
- Built on top of Scala
 - Object-oriented
 - Functional



 Tool

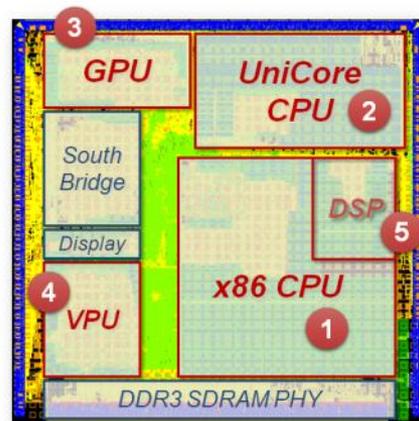
 Input/output

“不是备课的” 备课

参与研制北大自主CPU及配套计算机系统，从事过的工作包括：

- 编译器（lcc/gcc/llvm）、汇编器等二进制工具、调试器、库函数及动态链接器的设计、移植和优化
- 功能级、性能级模拟器的设计研发
- 二进制翻译工具及静态链接优化工具的设计和优化
- 虚拟机的移植和优化
- 浏览器中Javascript解释引擎的分析和优化
- 处理器性能评测、分析及优化设计

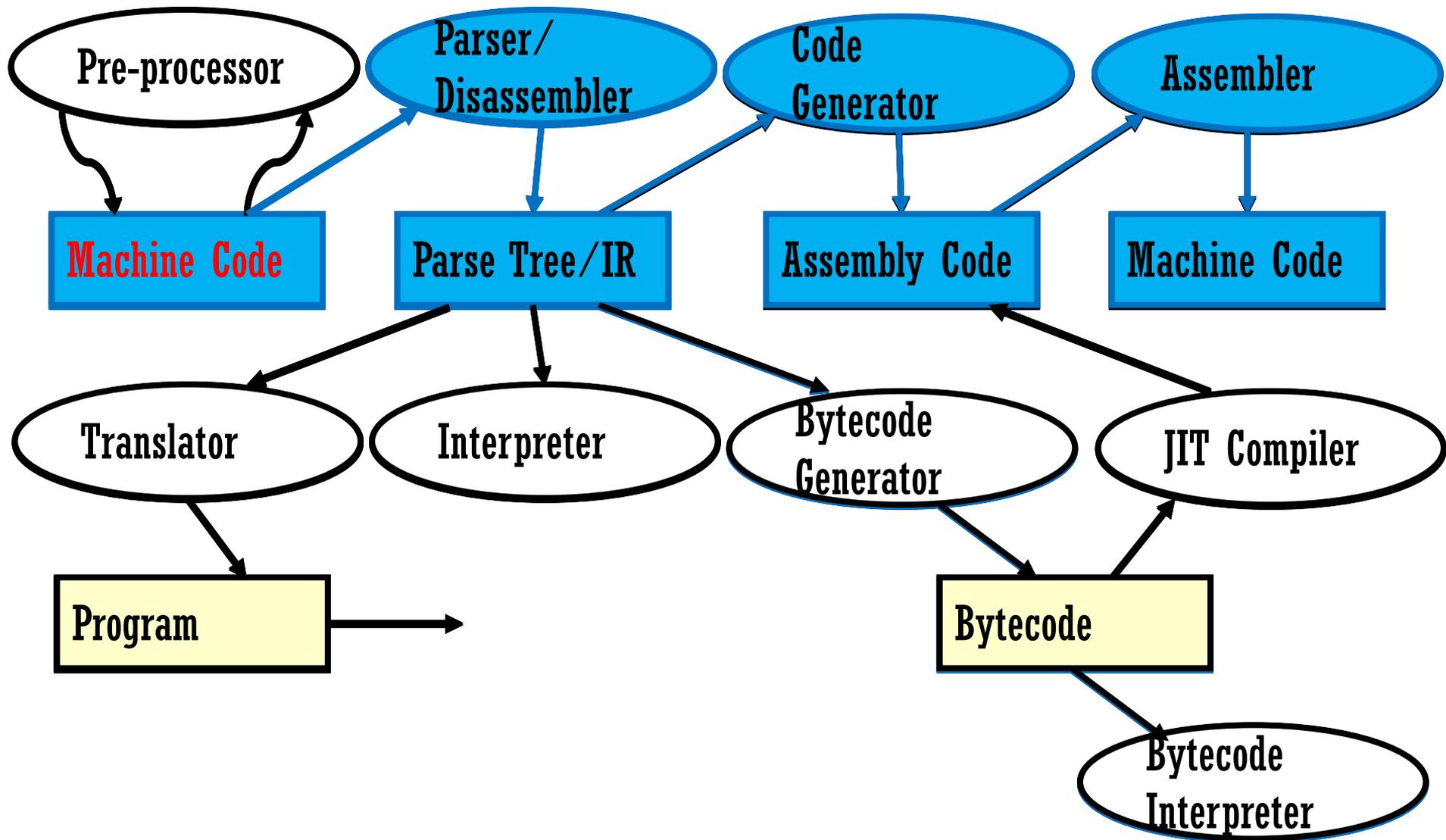
结合实际科研实践经验，对“计算机组成与系统结构实习”和“编译实习”进行教学改革和课程讲授，获得学生好评。



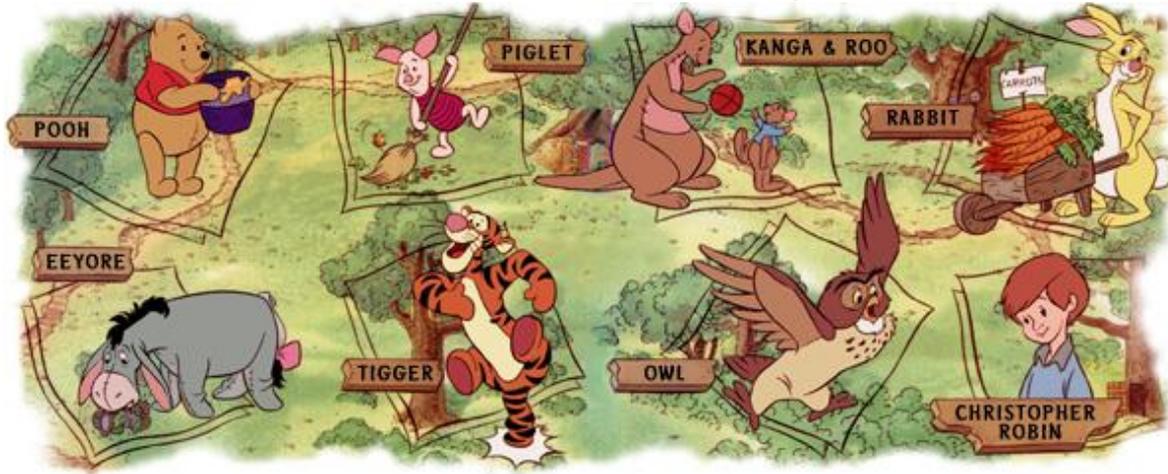
“盲人摸象”的启示



编译器、解释器、虚拟机及其它



例：“编译实习”课程具体任务



明确任务 规划题目	9.20	课程介绍, 题目介绍(MiniC+MiniJava)	
	9.27	Lex/Flex + YACC/Bison	
	10.11	文法和参考代码说明, 环境安装	
词法分析 语法分析 类型检查	10.18	交流、答疑 (文法讨论、词语法分析)	
	10.25	交流、答疑 (表达式和控制流翻译)	
	11.01	点评, MiniC2EEyore	
			检查点A
	11.08	点评, MiniC2EEyore	
中间代码生成 (代码优化)	11.15	交流、答疑 (语法树构建、活性分析)	
	11.22	交流、答疑 (寄存器分配、代码输出)	
	11.29	点评, EEyore2Tigger	
	12.06	交流、答疑 (I/O处理, 汇编生成)	
目标代码生成	12.13	交流、答疑 (优化)	
	12.20	点评, Tigger2RISCV	
			检查点C
实习报告	12.27	交流、答疑 (实习报告)	
	01.03	总结、PRE	
			12.31 23:59

对编译课程的一些调整

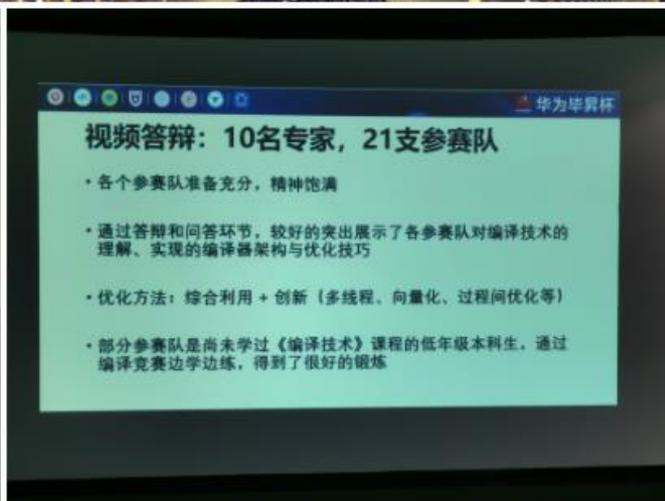
内容上的调整

- 阶段性审查
- 中间表示的强化
- 后端机器结构的影响
- 授课进度上的一些变化（优化的适当增强）

考核方式上的调整

- 单人一组
- 机器评测+答辩
- 允许合理使用工具（lex+yacc, JavaCC, ANTLR）
- 拥抱开源，但需要更严格的审查

计算机编译系统设计赛



CG Online Judge

Secure | https://course.educg.net/admin/index.jsp



CG Online Judge 我的课程 ▾ GitLab

帮助

编译原理 (北京大学)

- 公告与问卷
- 课程信息
- 学生
- 作业
- 考试
- 在线实验
- 在线课堂
- 答疑
- 汇总成绩
- 助教
- 设置
- 进入学生端

作业题库

- 题目概览
- 录入题目
 - 选择题
 - 填空题
 - 判断题
 - 简答题
 - 拍照上传题
 - 编程题
 - 程序片段编程题
 - 接口编程题
 - SQL评测题
 - 算法可视化
 - 文件上传题
 - 项目题
 - 通用评测题

回收站

作业管理

- 作业概览
- 布置作业
- 回收站

比赛与闯关^{beta}

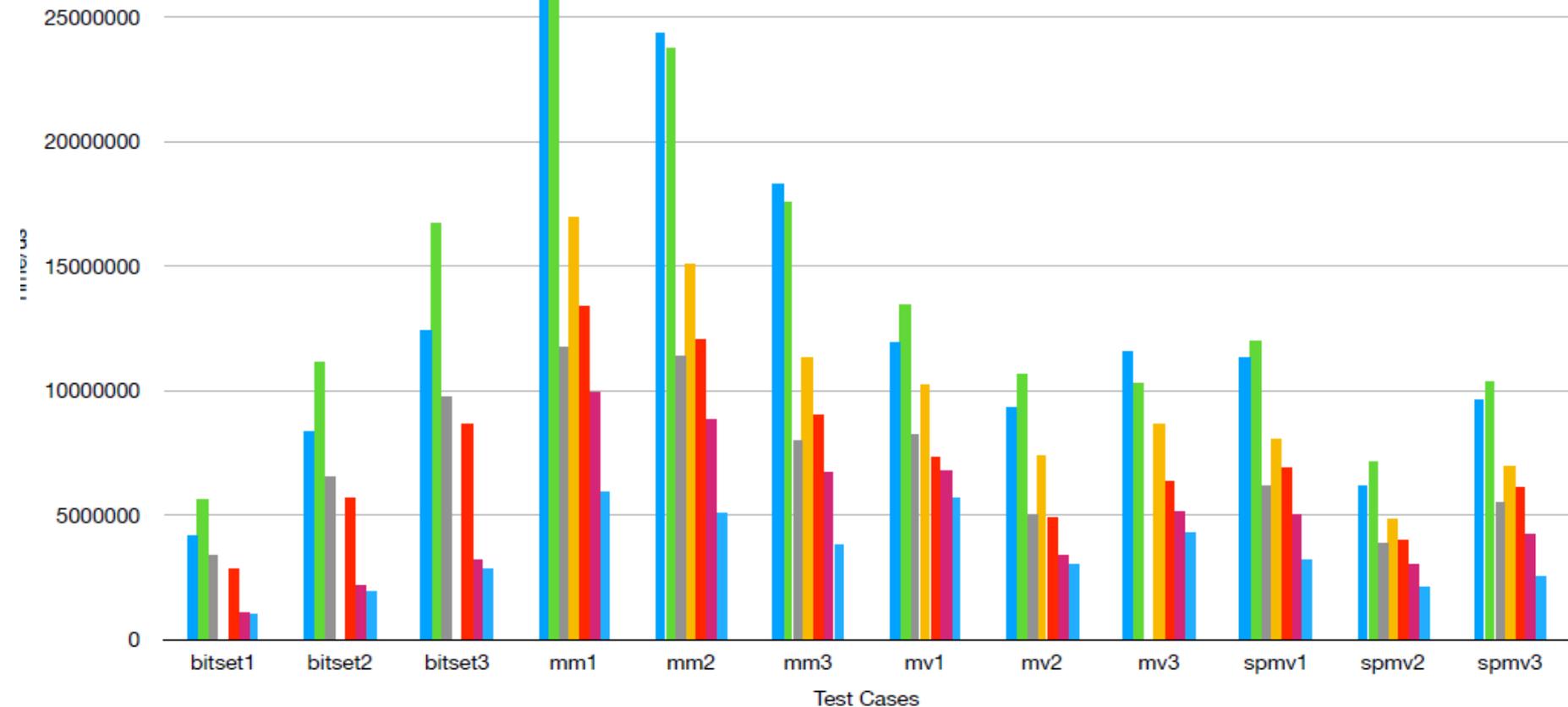
- 比赛概览
- 组织比赛
- 回收站

数据统计

- 作业活动数据
- 答题明细数据
- 常见问题

题目	类型	难度	章节	度量	操作	修改 / 录入时间
<input type="checkbox"/> 306278. 第二阶段测试: 生成 RISC-V 汇编	通用评测题			代码行: - 正确率: - 完成时间: -	编辑 删除 </> 样例代码 🔖 标签 📄 复制到试题库	邢其正 2020-11-16 18:19:18 2020-11-11 20:17:13
<input type="checkbox"/> 306277. 第二阶段测试: 生成 Tigger	通用评测题			代码行: - 正确率: - 完成时间: -	编辑 删除 </> 样例代码 🔖 标签 📄 复制到试题库	邢其正 2020-11-16 18:19:07 2020-11-11 20:11:03
<input type="checkbox"/> 306269. 第一阶段测试: 生成 Eeyore	通用评测题			代码行: - 正确率: 0.17 完成时间: 338分 钟	编辑 删除 </> 样例代码 🔖 标签 📄 复制到试题库	邢其正 2020-11-16 18:18:45 2020-11-11 18:45:15
<input type="checkbox"/> 305803. 样例题目	通用评测题	2		代码行: - 正确率: - 完成时间: -	编辑 删除 🔴 样例代码 🔖 标签 📄 复制到试题库	2020-11-06 10:25:34 2020-11-06 10:25:33
<input type="checkbox"/> 300835. 初赛	通用评测题	5		代码行: - 正确率: - 完成时间: -	编辑 删除 </> 样例代码 🔖 标签 📄 复制到试题库	2020-10-20 13:26:35 章胜奇 2020-10-20 13:26:35
<input type="checkbox"/> 300834. 赛后性能测试:shuffle	通用评测题			代码行: - 正确率: - 完成时间: -	编辑 删除 🔴 样例代码 🔖 标签 📄 复制到试题库	2020-10-20 13:26:35 章胜奇 2020-10-20 13:26:35
<input type="checkbox"/> 300833. 赛后性能测试:median	通用评测题			代码行: - 正确率: - 完成时间: -	编辑 删除 🔴 样例代码 🔖 标签 📄 复制到试题库	2020-10-20 13:26:35 章胜奇 2020-10-20 13:26:35
<input type="checkbox"/> 300832. 赛后性能测试:ft	通用评测题			代码行: - 正确率: - 完成时间: -	编辑 删除 🔴 样例代码 🔖 标签 📄 复制到试题库	2020-10-20 13:26:35 章胜奇 2020-10-20 13:26:35
<input type="checkbox"/> 300831. 赛后性能测试:conv	通用评测题			代码行: - 正确率: - 完成时间: -	编辑 删除 🔴 样例代码 🔖 标签 📄 复制到试题库	2020-10-20 13:26:35 章胜奇 2020-10-20 13:26:35
<input type="checkbox"/> 300830. 赛后性能测试:transpose	通用评测题			代码行: - 正确率: - 完成时间: -	编辑 删除 🔴 样例代码 🔖 标签 📄 复制到试题库	2020-10-20 13:26:35 章胜奇 2020-10-20 13:26:35
<input type="checkbox"/> 300829. 决赛赛后测试	通用评测题	5		代码行: - 正确率: - 完成时间: -	编辑 删除 </> 样例代码 🔖 标签 📄 复制到试题库	2020-10-20 13:26:34 章胜奇 2020-10-20 13:26:34

横向性能对比



扩展和自选任务题目示例

- 一种3D场景的标记语言TDML的设计实现
- 一种音乐处理语言MPL的设计实现
- 一种可以处理维基百科页面的语言Wikify
- 一种将指令式语言到函数式语言的转换器
- 面向SQL数据库查询语言的扩展及优化
- 基于python的OpenMP扩展实现
- Scheme语言记忆化功能的编译支持
- 一种元编译器的设计实现
- 面向应用程序热点的RISC-V指令系统扩展

扩展和自选任务题目示例

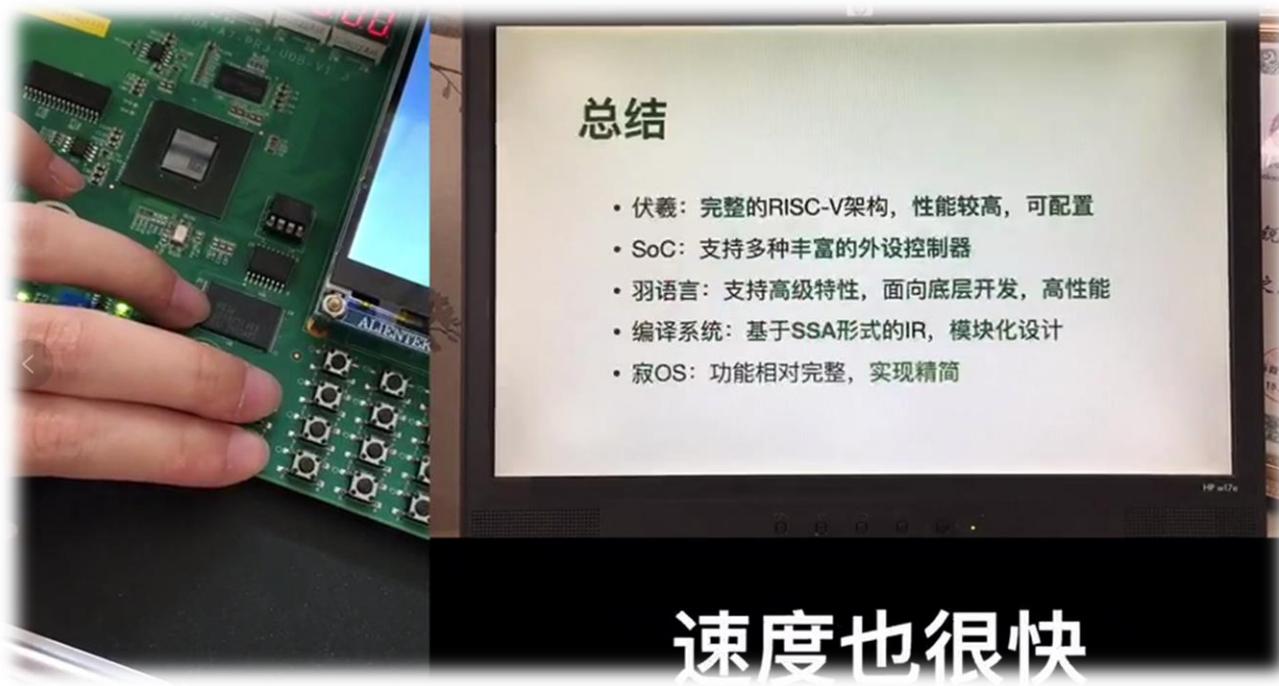
- GDFa在GCC后端的实现
- 基于OpenMP的循环并行调度优化
- 基于LLVM的圈复杂度计算工具的设计与实现
- Static Race Detection on LLVM IR
- Android应用静态污点分析和其中的回调库函数
- 基于控制流检测的程序安全执行检测技术
- Memory leak fixing in C language
- Comparing Pointer Analysis Effectiveness
- 针对Scratch Pad Memory的数据重用分析

授课场景



计算机系统实践环节的设计思考

注重基础课程，同时促进各学科之间的相互融合和相互促进。将计算机组织与系统结构实验、编译实习、操作系统实习等课程进行整合，以锻炼学生的综合能力。

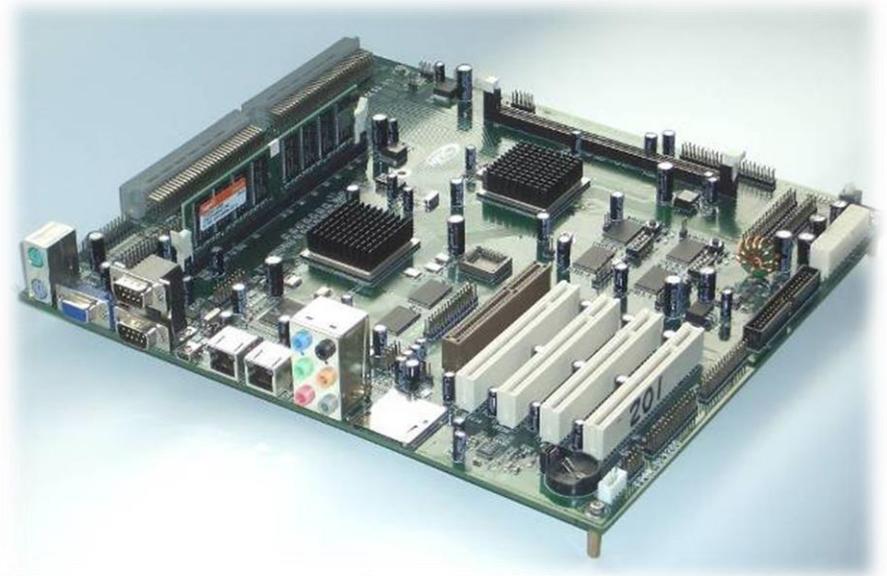
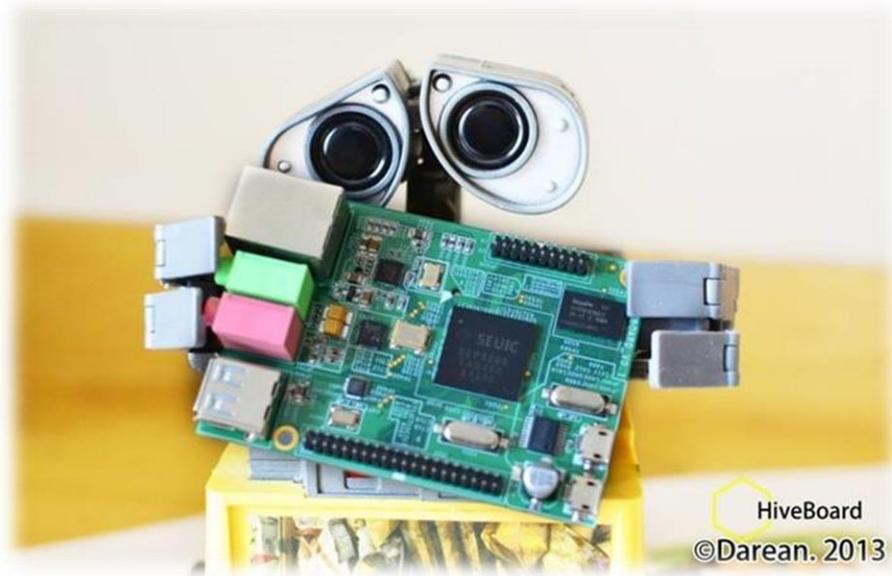


尝试性联合实践内容及要求

- 设计并实现一个编译器，可以将用给定的C语言的子集编写的源程序编译成指定指令系统的汇编程序（RISC-V指令系统架构、北京大学自定义的UniCore指令系统体系结构）
 - 完成词法分析和语法分析
 - 完成寄存器分配、代码生成和优化，生成汇编代码
 - 用汇编器生成可执行文件并在真机或模拟器上运行
- 为给定指令系统设计一个模拟器，使得编译器生成的代码可以在该模拟器上正确运行
 - 使用模拟器进行数据统计并开展性能分析
 - 调整处理器结构、存储层次并进行分析 and 评估

采用的平台

- Linux + GCC + ARM/UniCore/RISC-V (Real System)
- Xilinx FPGA



A25-18110



A25-30031



A25-19110



A25-61101



A25-6110X



A25-61100

学生的一些反馈

- “这一次实习不仅是完成课程任务，更让我们的思维方式得到了极大的锻炼。”
- “这次编译和体系联合实验是我们在大学两年半做过的最完整的实习，两门紧密结合在一起的实习课程让我们连贯的体会了编译和体系密切的关系。”
- “在最终我们编译器编译出来的程序能够正确快速的在我们实现的模拟器上运行的时候，那种满足感也是无以言表的。这两门实习，虽然是我们大学生活中最难的、也是最耗时的两门实习，但是确是对我们来说意义最为重要的实习之一，相信我们学到的能够对得起我们所付出的。”

学生的一点反馈

我们最喜欢的事是什么？



写编译器！



写编译器！



还是写编译器！



实验内容及要求（后继思考的改进）

- 完成一个简单操作系统的设计，系统可被该编译器编译成功并在真机或模拟器上运行
 - 支持中断控制、存储管理、进程切换等功能
 - 增加必要的管理和通信模块
 - 开展性能分析和评估
- 使用VHDL完成简单处理器的设计和验证，该处理器应可运行简单的程序或控制程序（OS）
 - 处理器采用多周期通路或流水线设计
 - 处理器应支持中断和必要的存储管理功能
 - 处理器可支持前述编译器生成的代码的运行
 - 处理器可支持前述简单操作系统的运行（扩展）

一些可能的贯通

- 仿真器/模拟器的设计技术路线选择
- 虚拟机的理解
- ICS等课程的若干工具的改进（助教角色）
 - HCL Tool @ 计算机系统导论
 - IR Interpreter @ 编译原理
 - 代码查重工具
 -

一点小分享

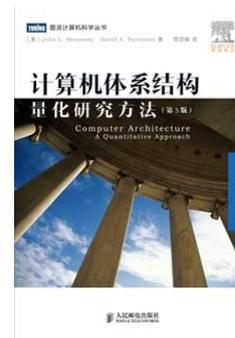
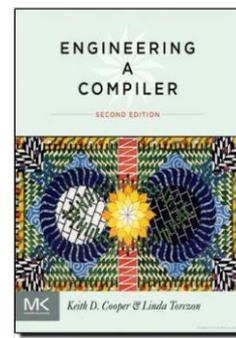
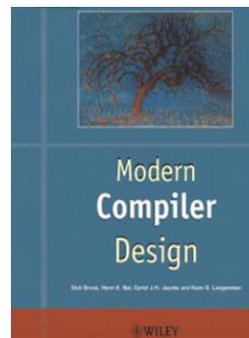
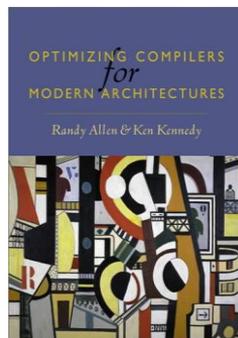
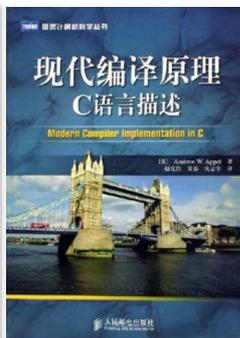
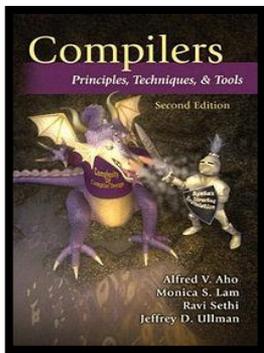
□ 注重**基本词汇、概念的源与流**。

compiler, interpreter, arch-i-itecture, chem-is-try

inter-rupt/trap/system call

MIPS=Microprocessor without Interlocked Pipeline Stages

□ 尝试**设计思想/哲学的传递**，试着让学生有**对美的感受**。



□ 注重设计实现中的**趣味性**。

授课体会心得

□实践环节十分重要

□纸上得来终觉浅，绝知此事要躬行

□允许多样性，鼓励多样性

□横看成岭侧成峰，远近高低各不同

□鼓励同学展开创新性尝试

□桐花万里丹山路，雏凤清于老凤声

一些思考

- 有必要提供**多层次、模块化**的课程体系，开设不同深度要求的课程，以适应不同兴趣和特长的学生。实验班的课程可以比普通班课程有更高的要求。
- **寻找更真实的目标环境和评测场景**，采用互动式、研讨式和案例式等方法，加大课程设计和综合实践的深度。
- **横向评测，以赛代练，以赛带练。**



北京大學
PEKING UNIVERSITY

衷心感谢
敬请指正

设置综合性的训练



计算机硬件基础课程

计算机软件基础课程