



中国科学技术大学
University of Science and Technology of China

编译教学与编译器研发能力培养

张昱

0551-63603804, yuzhang@ustc.edu.cn

中国科学技术大学
计算机科学与技术学院



编译课程的特点与现状

是计算机科学与技术（CST）专业的“**经典**”核心课

□ 其涉及的原理和技术：十分普遍 [龙书]

□ 有助于培养学生的

■ 系统能力（认知、分析、开发与应用）

■ 形式化描述能力

问题：高等教育的大众化(1998-)、CST专业细分

→ 各校**取消或减少**编译课的课时，部分高校**缺实践**

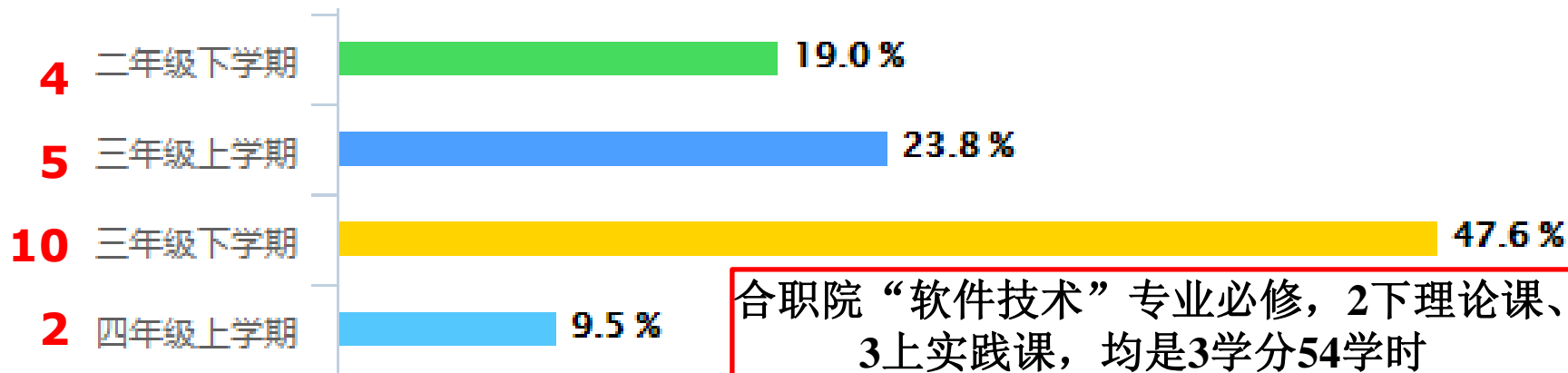
中科大：**1999级** 80学时理论课+独立的40学时实践课

2000级及之后：60+40理论课



问卷统计数据 (20所高校, 2017.7-8)

□ 编译开课学期



合职院“软件技术”专业必修，2下理论课、3上实践课，均是3学分54学时

	理论课时	实验课时	学分	各校每年上编译学生数
平均值	46.2	17.38	3.3	147
最大值	64	54	6	400
最小值	32	0	2	10
				合计：3087



编译教学的困难与问题

- 课程难、缺师资、无助教或助教质量不高
- 普遍反映在实验方面
 - 有2所学校无课程实验，导致学习更为抽象
 - 没有好的实验教材，缺少案例分析
 - 编译原理技术与实际应用如何结合
 - 多数学校：即便实验部分简单，但完成度也不高
- 教材
 - 偏重理论，缺乏对具体实现的介绍
如何理论问题实践化，抽象问题通俗化？
 - 内容陈旧



大环境的变化

□ 开源软件盛行

GitHub (2007-) : 65+百万项目、23+百万用户

□ 编程语言百花齐放

■ 众多: GitHub -- 开源项目涉及316种编程语言

■ 演变: C - C90, C99, C11; C++ - 98, 03, 06, 11, 14, 17

■ 流行的变化

GitHub(2016): 1JavaScript, 2Java, 3Python, 4Ruby, 5PHP, 6C

Tiobe <https://www.tiobe.com/tiobe-index/>

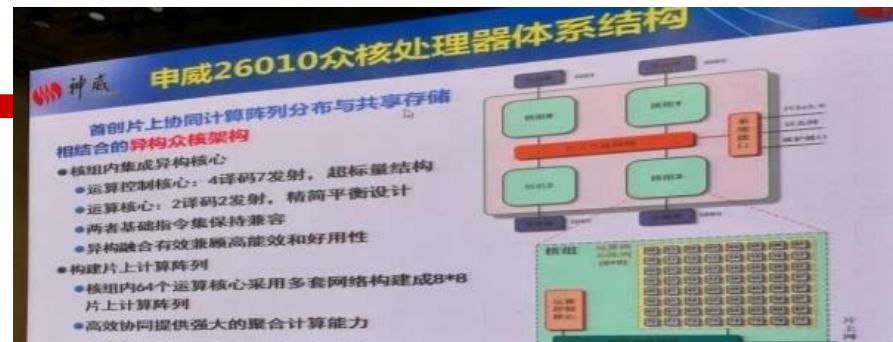
□ 2017.8: 1Java, 2C, 3C++, 4C#, 5Python, 6VB .NET

□ 2010年: 1Java, 2C, 3PHP, 4C++, 5VB, 6C#, 7Python

□ 1970年: 1Fortran, 2Lisp, 3Cobol, 4Algol60, 5APL, 6Snobol4



大环境的变化



- 开源软件盛行
- 编程语言百花齐放
- 目标平台/语言多样、异构

■ 另一种编程语言

■ CISCs (复杂指令集) : [x86](#)、[IA64](#)、...

■ RISCs (精简指令集) : [MIPS](#)、[ARM](#)、...

■ 多核/众核

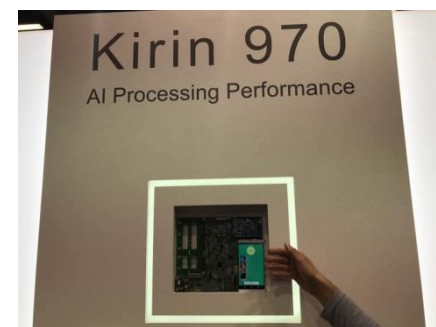
■ 加速器

□ GPU_s ([CUDA](#)、[OpenCL](#))、FPGA_s

□ TPU, NPU

■ 量子计算机

张昱：编译教学与编译器研发能力培养





大环境的变化

□ 开源软件盛行

□ 编程语言百花齐放、目标平台/语言多样

□ 编译器的形式

解释、交叉编译、即时编译、增量编译、预先编译等

□ 复杂系统：规模日趋庞大+多语言混合+异构

■ 应用的多样性

■ 提高软件开发生产力(productivity)

■ 改善软件的安全性、可靠性和可维护性

■ 支持并行(parallelism)与并发(concurrency)、新架构等

■ 移动和分发，模块化，多范型

讲授基础语言原理

[Thomas Ball, Benjamin Zorn;

CACM2015.5]

工业界期待迎来更多学习了程序设计语言原理的毕业生。

编

程人员日益成为紧缺人才，这一问题已得到广泛关注（请查看 www.code.org 网站上的相关内容，或观看 2013 年和 2014 年 12 月计算机科学周期间举办的“编码一小时”（Hour of Code）课程）。尽管会编程在支持人们发挥创造力方面拥有巨大潜力，但我们应记住，编程语言仍将继续发展，以解决层出不穷的新问题，而语言原理总是万变不离其宗。如本观点文章所论述，语言基础在本行业所使用的复杂软件系统的设计与实现中发挥日益重要的作用。工业界亟需更多语言原理方面的人才，为客户提供可靠而高效的软件解决方案。

在过去，为应对复杂系统在设计¹与实现方面的难题，很多重要语言原理应运而生。John McCarthy 在 1959 年前后为 Lisp 语言提出了垃圾回收机制，而该机制目前广泛应用于 Java 和 C# 等现代编程语言以及

语言基础在本行业所使用的复杂软件系统的设计与实现中发挥日益重要的作用。工业界亟需更多语言原理方面的人才，为客户提供可靠而高效的软件解决方案。

被广泛使用的编程语言奉为圭臬。² 类型系统按照程序表达式计算出的值来对其进行分类，³ 让编译程序可以证明某类错误不存在，更高效地优化编码。Hoare 的断言方法为如何确定程序的正确性提供了框架。⁴

我们希望构建更加复杂的系统，要想以可预见的方式实现可靠的系统，能在更高层次表达编程人员意图的全新机制必不可少。随着新类型的系统不断问世，为了让更多人员能对此类系统编程，就必须设计全新的编程语言。表达编程人

企业和社会团体抽象的弱类型系统（此处为 C 语言）就会付出惨重代价。编程的可靠系统，家公司都已开发编程语言。例如，（<http://golang.org>）rust 语言（<http://rust-lang.org>）以及 Microsoft 的 Sing# 语言（<http://singularity.codeplex.com/>）。这些语言通过可为程序执行提供更多安全保障的全新类型系统提高了编程水平。

领域专用语言（DSL）进一步提高了编程水平，通过限制表达能力实现了较通用语言更高层次的行为保证。SQL 数据库查询语言是一个经典例子，该语言基于关系代数，⁵ 支持复杂的查询优化。

DSL 仍将在行业中派上用场。Google 的 Map/Reduce 数据并行执行模型⁶ 推动了基于 SQL 的 DSL 的大量出现，其中包括 Yahoo 的 Pig



课程目标与培养的能力

- 理解编程语言的设计和实现,了解编程语言的理论
- **编程能力**(理解底层的机制和可能的影响要素)
快速分析和解决实际编程中的问题,解释产生的现象
- **形式化能力和语言设计能力**
形式描述语言的语法和语义,能设计领域专用语言**DSL**
- **工程能力**
操控上规模的软件、过程管理、个人/团队、沟通、文档
- **创新思维**
了解现代编译系统以至现代计算机系统、创新、业界动态

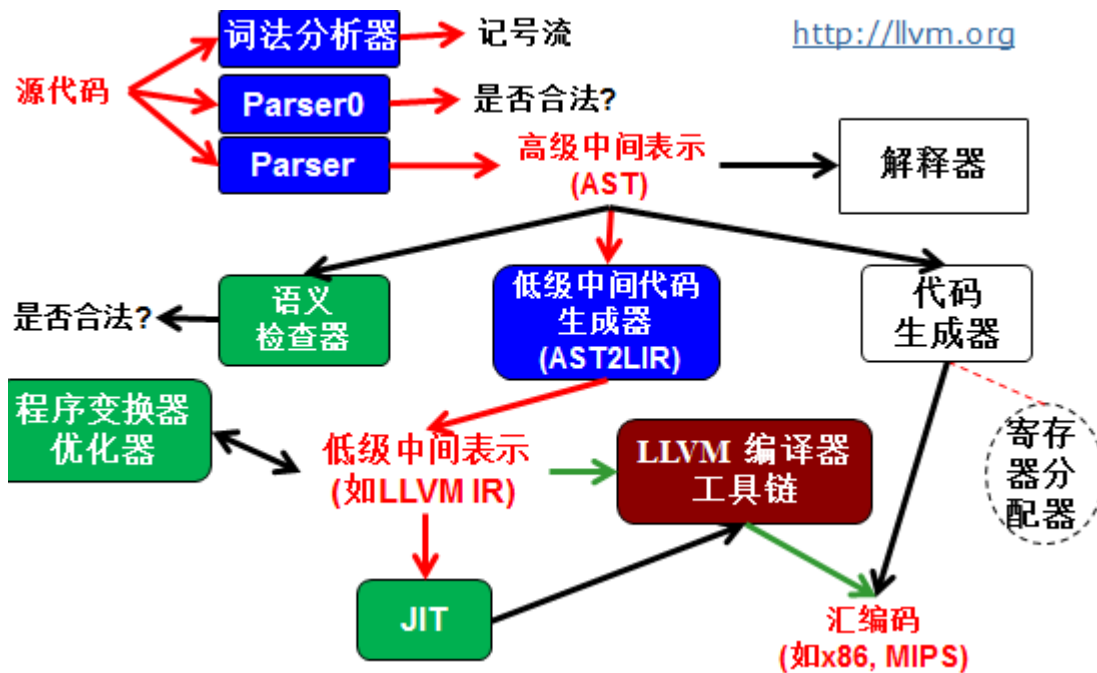


实施总体思路和方法

□ 强化课程实践（实验占总评 50%）

- git版本管理（2012-）
- 基础实验（做小型编译器）
- 阅读理解（开源软件如LLVM）
- 可选的综合实践

<http://staff.ustc.edu.cn/~yuzhang/compiler>





教学内容和方法、支撑

- 结合编程实际问题来讲解：分析的角度和因素
- 知识点：基础、优化（仅介绍基础结构和框架）
- ✓ 增加对不同编译器形式体系的介绍

✓ 分析技术

注重**编译器效率** => 让更多的人能**快速设计** 语言

lex+Yacc(LR) → **ANTLR 4 (ALL(*)[OOPSLA2014]2017-)**

ANTLR: 支持Java、C++、C#、Python、Go、JavaScript、Swift

✓ LLVM(源于UIUC, 为解决OpenGL函数到不同GPU的翻译)

- 1) 基础实验要求能生成**LLVM IR** (结合教程) (2015-)
- 2) 特定模块的**导读和问题回答**: 解析、CSA (2015-)、JIT (2017-)
- 3) 编译LLVM (2016-) (体验大工程的编译、时空问题及解释)



□ 支撑：优秀本科生作助教(2014-)

参与实践案例的构造
已形成良性循环



研讨：编译教学和编译器研发能力

- 内容选择：理论、实验
- 与业界的实际需求如何结合？
- 如何应对各种变化：源、目标、开源编译器等
- 师资、助教
- 资源库：案例、业界问题等