



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

# 编译原理课程与系统能力培养

张 昱

中国科学技术大学 计算机科学与技术学院

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

安徽省高校计算机教育研究会年会[11/28/2015,合肥]

# 内容概要

- ① 中科大的程序语言系列课程简介
- ② 多层次的编译原理课程
- ③ 编译原理课程实践案例
- ④ 教学上采取的策略
- ⑤ 总结

# 内容概要

- ① 中科大的程序语言系列课程简介
- ② 多层次的编译原理课程
- ③ 编译原理课程实践案例
- ④ 教学上采取的策略
- ⑤ 总结



# 中科大程序语言系列课程简介

## ➤ 程序设计语言课程

- 程序设计I（C语言） 40/30学时，2.5学分，一上，通修
- 程序设计II 40/40，3，一下，通修
- Java 软件开发基础 2学分，选修

## ➤ 编译原理和技术 60/40学时，4学分，三上，专业核心课

## ➤ 程序设计语言基础 60/20，3.5学分，三下，专业方向课

- JavaScript中的闭包、动态绑定；函数式语言；并发.....

## ➤ 并行计算 2.5学分

## ➤ 程序设计语言理论：研究生课程

# 内容概要

- 01 中科大的程序语言系列课程简介
- 02 多层次的编译原理课程
- 03 编译原理课程实践案例
- 04 教学上采取的策略
- 05 总结



# 多层次的编译原理课程

## ➤ 计算机专业

分两个级别，同时间段授课，学生可以选择

### ➤ 英才班（30+）

加大实验强度，基础实验+扩展实验，占总分**50%**

### ➤ 普通班

基础实验，实验占总分**20~30%**

学生均要开发出小型实验型语言的编译器

## ➤ 计算机双学位

➤ 周六上课，4节/次，弱化实验（2个），内容适当删减



# 普通高校本科/双学位的教学目标

通过编程语言**实现技术**的学习，提高：

- 学习、理解编程语言的能力
- 在程序开发中应用编程语言的能力
  - ✓ 提高程序排错的能力  
即快速理解、定位和解决在程序开发与程序运行中碰到的问题的能力
  - ✓ 提高编写高质量代码的能力



# 重点高校本科的教学目标

通过对和编程语言相关的**理论和技术**的学习，提高

- 学习、理解编程语言的能力
  - 在程序开发中应用编程语言的能力
  - 在软件工程中应用这些理论和技术的能力
- 
- ✓ 形式语言和自动机理论
  - ✓ 语法制导的翻译技术
  - ✓ 类型论和类型系统
  - ✓ 数据流分析的理论基础



# 内容概要

- ① 中科大的程序语言系列课程简介
- ② 多层次的编译原理课程
- ③ 编译原理课程实践案例**
- ④ 教学上采取的策略
- ⑤ 总结



# 编译原理课程实践案例

本学期的英才班课程实践安排

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

<https://www.zybuluo.com/sangyy/note/166600>

- **7个基础实验**：和讲授内容配套，循序渐进
- **最后1个扩展实验**：自由选题，1月份逐个答辩
- **实验环境**：**Linux+命令行编译运行（Makefile、shell脚本）**
  - 学生平时在个人电脑上实验
  - 助教上讨论课、QQ交流等
  - 长期在线的服务器：**git版本管理**，学生专有的库
- **助教**：以往的研究生 → 近年的**大四学生**（上届学本课程）



# 编译原理课程实践案例(续)

## ➤ P1 预备阶段

- 熟悉实验环境: **Linux**、**LLVM 3.6.0**、**GCC**, **Makefile**
- 熟悉**C1**语言的特征: 编写**C1**程序, 用**GCC/Clang**编译
- 初步了解编译的过程
- 提交的目录结构要求
  - ✓ **README**
  - ✓ **Makefile**
  - ✓ **src/**
  - ✓ **doc/**
  - ✓ **bin/** : **shell脚本**
  - ✓ **test/**



# 编译原理课程实践案例(续)

## ➤ P2 词法分析

- 学习LLVM提供的教程Kaleidoscope
  - ✓ 理解其中的词法分析过程
  - ✓ 对上述词法分析进行扩展：多行注释、八进制数
  - ✓ 修改扩展后的词法分析部分，使之输出记号流
- 学习使用Flex，用Flex生成C1的词法分析器
- 阅读Clang源代码中的词法分析文件，并回答问题
- 提交的目录结构要求



# 编译原理课程实践案例

## ➤ P3 Kaleidoscope语法分析

- 理解Kaleidoscope的语法分析和抽象语法树
- 扩展while循环

## ➤ P4 C1语言的语法分析与错误恢复

- 学习Bison
- 用Flex+bison为C1语言构造能识别正确 C1程序的分析器
- 提交的目录结构要求



# 编译原理课程实践案例

## ➤ P5 生成C1的抽象语法树(AST)

- 理解所提供的`asgn2ast`样例
- 为C1生成AST

## ➤ P6 Clang源码阅读：理解语法分析和静态检查

- 用`ninja`编译`clang`，了解其中由`td`描述文件自动生成代码的技术
- 阅读`clang`源码中的部分源程序文件，来理解`clang`的语法分析和静态语义检查的实现机制
- 了解`clang`编译工具的构造（可选的库+驱动程序）

## ➤ P7 代码生成：为C1程序生成LLVM IR代码

安徽省高校计算机教育研究会年会[11/28/2015,合肥]

# 内容概要

- 01 中科大的程序语言系列课程简介
- 02 多层次的编译原理课程
- 03 编译原理课程实践案例
- 04 教学上采取的策略**
- 05 总结



# 教学上采取的策略

- 介绍许多从实际编译和运行时碰到的问题
  - 多以C语言作为实例语言
    - ✓ 学过、非类型安全的语言、可生成汇编便于阅读
  - 引导用所学知识去分析、解决、理解实际编程中的问题和发生的现象 → 让学生出题
  - 不同的编码方法、不同的实现技术产生的不同影响
- 介绍现代常用语言的编译运行系统的结构
  - 编译、连接生成本地码：C → 汇编码
  - 即时编译器+垃圾收集：Java虚拟机、JavaScript引擎
  - 安卓代码的编译、运行：Delvik → ART

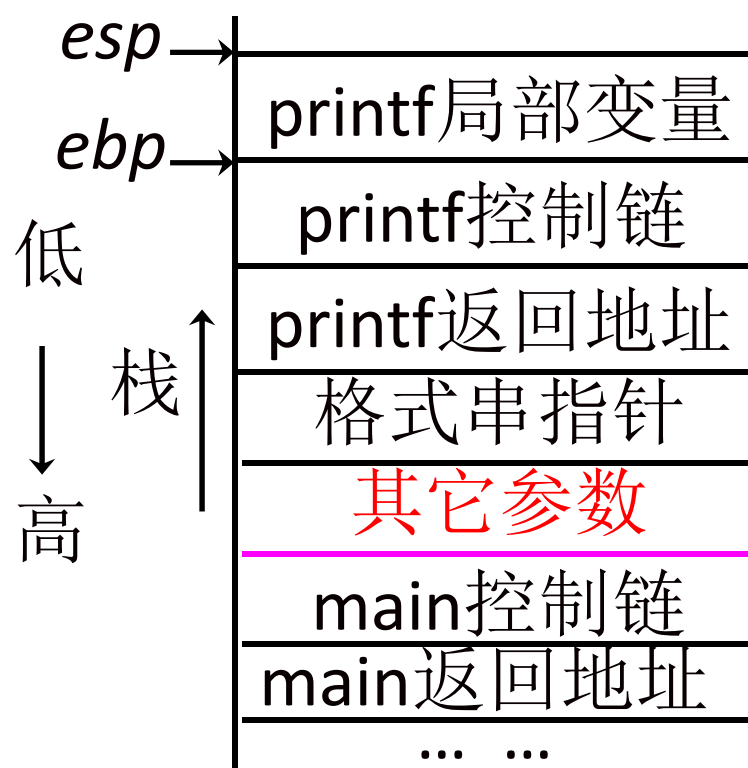




# 例题1 全局存储分配

下面的程序运行时输出3个整数。试从运行时存储空间的组织和printf的实现来分析，为什么此程序会有3个整数输出？

```
main() {  
    printf("%d, %d, %d\n");  
}
```





## 例题2 编译和连接

**cc**是UNIX系统上C语言编译命令，**-l**是连接库函数的选择项。某程序员自己编写了两个函数库 **libuser1.a** 和 **libuser2.a**，当用命令

**cc test.c -luser1.a -luser2.a**

编译时，报告有未定义的符号，而改用命令

**cc test.c -luser2.a -luser1.a**

时，能得到可执行程序。试分析原因

(备注：库名中的**lib**在命令中省略。该命令和命令**cc test.c libuser1.a libuser2.a**的效果一致)



## 例题2 编译和连接

**cc test.c -luser1.a -luser2.a**

**cc test.c -luser1.a -luser2.a**

解答

**test.c**

引用a

**libuser1.a**

定义b

**libuser2.a**

定义a

引用b

**test.c**

引用a

**libuser2.a**

定义a

引用b

**libuser1.a**

定义b



## 例题3 代码生成

下面的程序在X86/Linux机器上编译后的运行结果是7，而在SPARC/SUNOS机器上的运行结果是6。试分析运行结果不同的原因

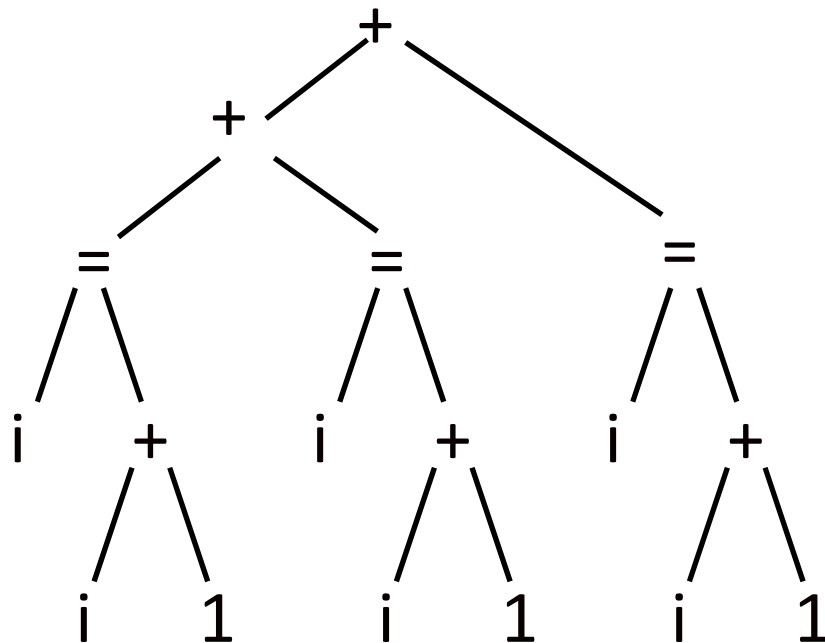
```
main()
{
    long i;

    i = 0;
    printf("%ld\n", (++i)+(++i)+(++i) );
}
```



## 例题3 代码生成

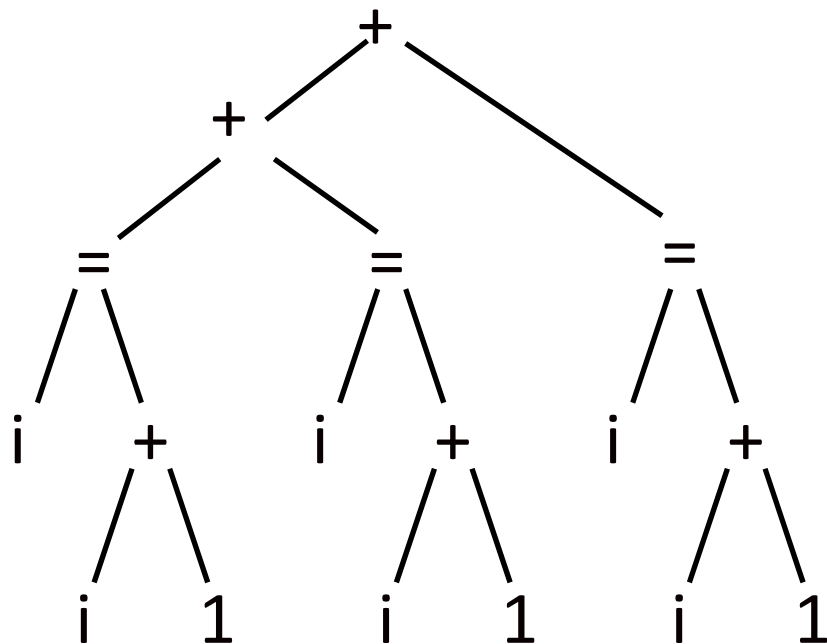
按一般的代码生成， $i = i + 1$ 的计算结果保留在寄存器中，因此这三个 $i = i + 1$ 的计算次序不会影响最终的结果。结果应该是6






按一般的代码生成， $i = i + 1$ 的计算结果保留在寄存器中，因此这三个 $i = i + 1$ 的计算次序不会影响最终的结果。结果应该是6

结果是7的话，一定是某个 $i = i + 1$ 的结果未保留在寄存器中。上层计算对它的引用落在计算另一个 $i = i + 1$ 的后面



# 内容概要

- ① 中科大的程序语言系列课程简介
- ② 多层次的编译原理课程
- ③ 编译原理课程实践案例
- ④ 教学上采取的策略
- ⑤ 总结



# 总结

- 教学与实践内容选择的建议
  - 第1门程序语言：细讲、GUI→命令行
  - 以后的语言课学习：相比以前语言的特点、实现养成查语言规范、编译器版本的说明来了解语言特征
  - 程序设计方法学及一定规模的程序的程序设计
  - 编译原理和技术（针对一般的高校）
    - 淡化对分析器的自动生成技术的叙述
    - 增强对实际编译、运行问题的分析
    - 引入编译器源码的阅读





# 资源

➤ 教师的课程主页

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

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

➤ 中科大省级精品课程资源

[http://www.bb.ustc.edu.cn/jpkc/sheng\\_ji.html](http://www.bb.ustc.edu.cn/jpkc/sheng_ji.html)

➤ **2013年**《编译原理和技术》省级精品资源共享课

➤ **2009年**《编译原理》省级精品课程

➤ 两年**1**次的龙星计划课程（**2014、2016**）

<http://dragonstar.ict.ac.cn/dragonstar/curriculum.asp>

*欢迎省内同仁反馈、交流探讨！*

安徽省高校计算机教育研究会年会[11/28/2015,合肥]



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

Q&A

**谢谢！**

安徽省高校计算机教育研究会年会[11/28/2015,合肥]