

从硬件工程师的角度理解软件；从软件工程师的角度理解硬件

INTRODUCTION TO COMPUTER ARCHITECTURE

计算机体系结构基础

胡伟武◎等著

龙芯中科技术有限公司



机械工业出版社
China Machine Press

计算机类专业
系统能力培养
系列教材

INTRODUCTION TO COMPUTER ARCHITECTURE

计算机体系结构基础

胡伟武 苏孟豪 王焕东 汪文祥 章隆兵
肖俊华 刘 苏 陈新科 吴瑞阳 李晓钰 著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

计算机体系结构基础 / 胡伟武等著. —北京: 机械工业出版社, 2017.9
(计算机类专业系统能力培养系列教材)

ISBN 978-7-111-57978-6

I. 计… II. 胡… III. 计算机体系结构—高等学校—教材 IV. TP303

中国版本图书馆 CIP 数据核字 (2017) 第 218615 号

本书由国内从事微处理器设计的一线科研人员编写而成, 为培养“造计算机”的人才探索广受欢迎的本科教育方法。作者从微处理器设计的角度出发, 充分考虑计算机体系结构的学科完整性, 强调体系结构、基础软件、电路和器件的融会贯通。全书包括指令系统结构、计算机硬件结构、CPU 微结构、并行处理结构、计算机性能分析等主要内容, 重点放在作为软硬件界面的指令系统结构, 以及包含 CPU、GPU、南北桥协同的计算机硬件结构上。

本书可作为高等学校“计算机体系结构”课程的本科生教材, 同时也适合相关专业研究生或计算机技术人员参考阅读。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 温莉芳

责任校对: 李秋荣

印刷: 三河市宏图印务有限公司

版次: 2017 年 9 月第 1 版第 1 次印刷

开本: 185mm×260mm 1/16

印张: 19.5

书号: ISBN 978-7-111-57978-6

定价: 49.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

丛书序言

——计算机专业学生系统能力培养和系统课程设置的研究

未来的5~10年是中国实现工业化与信息化融合,利用信息技术与装备提高资源利用率、改造传统产业、优化经济结构、提高技术创新能力与现代管理水平的关键时期,而实现这一目标,对于高效利用计算系统的其他传统专业的专业人员需要了解和掌握计算思维,对于负责研发多种计算系统的计算机专业的专业人员则需要具备系统级的设计、实现和应用能力。

1. 计算技术发展特点分析

进入21世纪以来,计算技术正在发生重要发展和变化,在20世纪个人机普及和Internet快速发展基础上,计算技术从初期的科学计算与信息处理进入了以移动互联、物物相联、云计算与大数据计算为主要特征的新型网络时代,在这一发展过程中,计算技术也呈现出以下新的系统形态和技术特征。

(1) 四类新型计算系统

1) **嵌入式计算系统** 在移动互联网、物联网、智能家电、三网融合等行业技术与产业发展中,嵌入式计算系统有着举足轻重的作用。例如,移动互联网中的移动智能终端、物联网中的汇聚节点、三网融合后的电视机顶盒等是复杂而新型的嵌入式计算系统;除此之外,新一代武器装备,工业化与信息化融合战略实施所推动的工业智能装备,其核心也是嵌入式计算系统。因此,嵌入式计算将成为新型计算系统的主要形态之一。在当今网络时代,嵌入式计算系统也日益呈现网络化的开放特点。

2) **移动计算系统** 在移动互联网、物联网、智能家电以及新型装备中,均以移动通信网络为基础,在此基础上,移动计算成为关键技术。移动计算技术将使计算机或其他信息智能终端设备在无线环境下实现数据传输及资源共享,其核心技术涉及支持高性能、低功耗、无线连接和轻松移动的移动处理机及其软件技术。

3) **并行计算系统** 随着半导体工艺技术的飞速进步和体系结构的不断发展,多核/众核处理机硬件日趋普及,使得昔日高端的并行计算呈现出普适化的发展趋势。多核技术就是在处理器上拥有两个或更多一样功能的处理器核心,即将数个物理处理器核心整合在一个内核中,数个处理器核心在共享芯片组存储界面的同时,可以完全独立地完成各自操作,从而能在平衡功耗的基础上极大地提高CPU性能。并行计算对计算系统微体系结构、系统软件与编程环境均有很大影响,同时,云计算也是建立在廉价服务器组成的大规模集群并行计算基础之上。因

此，并行计算将成为各类计算系统的基础技术。

4) **基于服务的计算系统** 无论是云计算还是其他现代网络化应用软件系统，均以服务计算为核心技术。服务计算是指面向服务的体系结构（SOA）和面向服务的计算（SOC）技术，它是标识分布式系统和软件集成领域技术进步的一个里程碑。服务作为一种自治、开放以及与平台无关的网络化构件可使分布式应用具有更好的复用性、灵活性和可增长性。基于服务组织计算资源所具有的松耦合特征使得遵从 SOA 的企业 IT 架构不仅可以有效保护企业投资、促进遗留系统的复用，而且可以支持企业按需应变的敏捷性和先进的软件外包管理模式。Web 服务技术是当前 SOA 的主流实现方式，其已经形成了规范的服务定义、服务组合以及服务访问。

(2) “四化”主要特征

1) **网络化** 在当今网络时代，各类计算系统无不呈现出网络化发展趋势，除了云计算系统、企业服务计算系统、移动计算系统之外，嵌入式计算系统也在物联时代通过网络化成为开放式系统。即，当今的计算系统必然与网络相关，尽管各种有线网络、无线网络所具有的通信方式、通信能力与通信品质有较大区别，但均使得与其相联的计算系统能力得以充分延伸，更能满足应用需求。网络化对计算系统的开放适应能力、协同工作能力等也提出了更高的要求。

2) **多媒体化** 无论是传统 Internet 应用服务，还是新兴的移动互联网服务业务，多媒体化是其面向人类、实现服务的主要形态特征之一。多媒体技术是利用计算机对文本、图形、图像、声音、动画、视频等多种信息进行综合处理、建立逻辑关系和人机交互作用的新技术。多媒体技术使计算机可以处理人类生活中最直接、最普遍的信息，从而使得计算机应用领域及功能得到了极大的扩展，使计算机系统的人机交互界面和手段更加友好和方便。多媒体具有计算机综合处理多种媒体信息的集成性、实时性与交互性特点。

3) **大数据化** 随着物联网、移动互联网、社会化网络的快速发展，半结构化及非结构化的数据呈几何倍增长。数据来源的渠道也逐渐增多，不仅包括了本地的文档、音视频，还包括网络内容和社交媒体；不仅包括 Internet 数据，更包括感知物理世界的的数据。从各种类型的数据中快速获得有价值信息的能力，称为大数据技术。大数据具有体量巨大、类型繁多、价值密度低、处理速度快等特点。大数据时代的来临，给各行各业的数据处理与业务发展带来重要变革，也对计算系统的新型计算模型、大规模并行处理、分布式数据存储、高效的数据处理机制等提出了新的挑战。

4) **智能化** 无论是计算系统的结构动态重构，还是软件系统的能力动态演化；无论是传统 Internet 的搜索服务，还是新兴移动互联的位置服务；无论是智能交通应用，还是智能电网应用，无不显现出鲜明的智能化特征。智能化将影响计算系统的体系结构、软件形态、处理算法以及应用界面等。例如，相对于功能手机的智能手机是一种安装了开放式操作系统的手机，可以随意安装和卸载应用软件，具备无线接入互联网、多任务和复制粘贴以及良好用户体验等能力；相对于传统搜索引擎的智能搜索引擎是结合了人工智能技术的新一代搜索引擎，不仅具

有传统的快速检索、相关度排序等功能,更具有用户角色登记、用户兴趣自动识别、内容的语义理解、智能信息化过滤和推送等功能,其追求的目标是根据用户的请求从可以获得的网络资源中检索出对用户最有价值的信息。

2. 系统能力的主要内涵及培养需求

(1) 主要内涵

计算机专业学生的系统能力的核心是掌握计算系统内部各软件/硬件部分的关联关系与逻辑层次;了解计算系统呈现的外部特性以及与人和物理世界的交互模式;在掌握基本系统原理的基础上,进一步掌握设计、实现计算机硬件、系统软件以及应用系统的综合能力。

(2) 培养需求

要适应“四类计算系统,四化主要特征”的计算技术发展特点,计算机专业人才培养必须“与时俱进”,体现计算技术与信息产业发展对学生系统能力培养的需求。在教育思想上要突现系统观教育理念,在教学内容中体现新型计算系统原理,在实践环节上展现计算系统平台技术。

要深刻理解系统化专业教育思想对计算机专业高等教育过程所带来的影响。系统化教育和系统能力培养要采取系统科学的方法,将计算对象看成一个整体,追求系统的整体优化;要夯实系统理论基础,使学生能够构建出准确描述真实系统的模型,进而能够用于预测系统行为;要强化系统实践,培养学生能够有效地构造正确系统的能力。

从系统观出发,计算机专业的教学应该注意教学生怎样从系统的层面上思考(设计过程、工具、用户和物理环境的交互),讲透原理(基本原则、架构、协议、编译以及仿真等),强化系统性的实践教学培养过程和内容,激发学生的辩证思考能力,帮助他们理解和掌控数字世界。

3. 计算机专业系统能力培养课程体系设置总体思路

为了更好地培养适应新技术发展的、具有系统设计和系统应用能力的计算机专门人才,我们需要建立新的计算机专业本科教学课程体系,特别是设立系统级综合性课程,并重新规划计算机系统核心课程的内容,使这些核心课程之间的内容联系更紧密、衔接更顺畅。

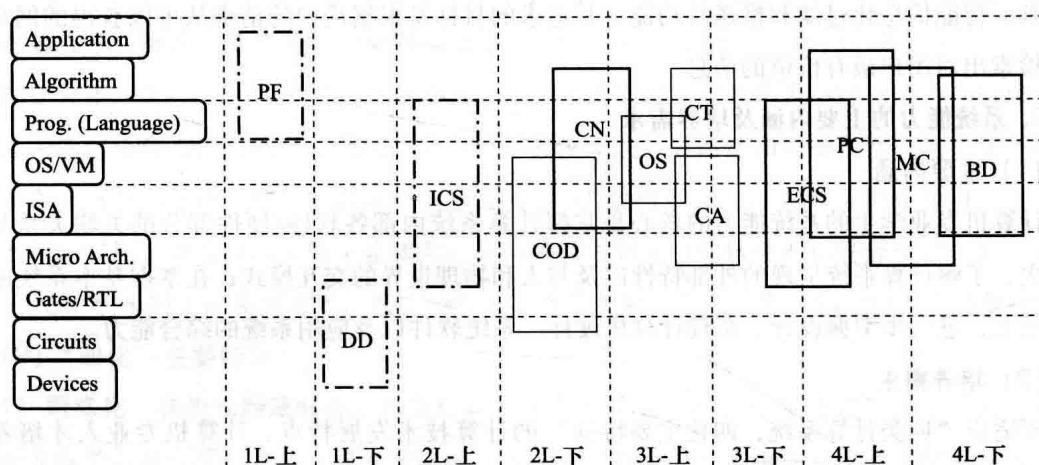
我们建议把课程分成三个层次:计算机系统基础课程、重组内容的核心课程、侧重不同计算系统的若干相关平台应用课程。

第一层次核心课程包括“程序设计基础(PF)”“数字逻辑电路(DD)”和“计算机系统基础(ICS)”。

第二层次核心课程包括“计算机组成与设计(COD)”“操作系统(OS)”“编译技术(CT)”和“计算机系统结构(CA)”。

第三层次核心课程包括“嵌入式计算系统(ECS)”“计算机网络(CN)”“移动计算(MC)”“并行计算(PC)”和“大数据并行处理技术(BD)”。

基于这三个层次的课程体系，相关课程设置方案如下图所示。



图中左边部分是计算机系统的各个抽象层，右边的矩形表示课程，其上下两条边的位置标示了课程内容在系统抽象层中的涵盖范围，矩形的左右两条边的位置标示了课程大约在哪个年级开设。点画线、细实线和粗实线分别表示第一、第二和第三层次核心课程。

从图中可以看出，该课程体系的基本思路是：先讲顶层比较抽象的编程方面的内容；再讲底层有关系统的具体实现基础内容；然后再从两头到中间，把顶层程序设计的内容和底层电路的内容按照程序员视角全部串起来；在此基础上，再按序分别介绍计算机系统硬件、操作系统和编译器的实现细节。至此的所有课程内容主要介绍单处理器系统的相关内容，而计算机体系结构主要介绍各不同并行粒度的体系结构及其相关的操作系统实现技术和编译器实现技术。第三层次的课程没有先后顺序，而且都可以是选修课，课程内容应体现第一和第二层次课程内容的螺旋式上升趋势，也即第三层次课程内容涉及的系统抽象层与第一和第二层次课程涉及的系统抽象层是重叠的，但内容并不是简单重复，应该讲授在特定计算系统中的相应教学内容。例如，对于“嵌入式计算系统（ECS）”课程，虽然它所涉及的系统抽象层与“计算机系统基础（ICS）”课程涉及的系统抽象层完全一样，但是，这两门课程的教学内容基本上不重叠。前者着重介绍与嵌入式计算系统相关的指令集体系结构设计、操作系统实现和底层硬件设计等内容，而后者着重介绍如何从程序员的角度来理解计算机系统设计与实现中涉及的基础内容。

与传统课程体系设置相比，新的课程体系中最大的不同在于有一门涉及计算机系统各个抽象层面的能够贯穿整个计算机系统设计和实现的基础课程：“计算机系统基础（ICS）”。该课程讲解如何从程序员角度来理解计算机系统，可以使程序员进一步明确程序设计语言中的语句、数据和程序是如何在计算机系统中实现和运行的，让程序员了解不同的程序设计方法为什么会有不同的性能等。

此外，新的课程体系中，强调课程之间的衔接和连贯，主要体现在以下几个方面。

1) “计算机系统基础”课程可以把“程序设计基础”和“数字逻辑电路”之间存在于计

算机系统抽象层中的“中间间隔”填补完整并很好地衔接起来，这样，到2L上结束的时候，学生就可以通过这三门课程清晰地建立单处理器计算机系统的整机概念，构造出完整的计算机系统的基本框架，而具体的计算机系统各个部分的实现细节再通过后续相关课程来细化充实。

2) “数字逻辑电路”“计算机组成与设计”“嵌入式计算系统”中的实验内容之间能够很好地衔接，可以规划一套承上启下的基于FPGA开发板的综合实验平台，让学生在一个统一的实验平台上从门电路开始设计基本功能部件，然后再以功能部件为基础设计CPU、存储器和外围接口，最终将CPU、存储器和I/O接口通过总线互连为一个完整的计算机硬件系统。

3) “计算机系统基础”“计算机组成与设计”“操作系统”和“编译技术”之间能够很好地衔接。新课程体系中“计算机系统基础”和“计算机组成与设计”两门课程对原来的“计算机系统概论”和“计算机组成原理”的内容进行了重新调整和统筹规划，这两门课程的内容是密切关联的。对于“计算机系统基础”与“操作系统”“编译技术”的关系，因为“计算机系统基础”以Intel x86为模型机进行讲解，所以它为“操作系统”（特别是Linux内核分析）提供了很好的体系结构基础。同时，在“计算机系统基础”课程中为了清楚地解释程序中的文件访问和设备访问等问题，会从程序员角度简单引入一些操作系统中的相关基础知识。此外，在“计算机系统基础”课程中，会讲解高级语言程序如何进行转换、链接以生成可执行代码的问题；“计算机组成与设计”中的流水线处理等也与编译优化相关，而且“计算机组成与设计”以MIPS为模型机进行讲解，而MIPS模拟器可以为“编译技术”的实验提供可验证实验环境，因而“计算机系统基础”和“计算机组成与设计”两门课程都与“编译技术”有密切的关联。“计算机系统基础”“计算机组成与设计”“操作系统”和“编译技术”这四门课程构成了一组计算机系统能力培养最基本的核心课程。

从“计算机系统基础”课程的内容和教学目标以及开设时间来看，位于较高抽象层的先行课（如程序设计基础和数据结构等课程）可以按照原来的内容和方式开设和教学，而作为新的“计算机系统基础”和“计算机组成与设计”先导课的“数字逻辑电路”，则需要对传统的教学内容，特别是实验内容和实验手段方面进行修改和完善。

有了“计算机系统基础”和“计算机组成与设计”课程的基础，作为后续课程的操作系统、编译原理等将更容易被学生从计算机系统整体的角度理解，课程内容方面不需要大的改动，但是操作系统和编译器的实验要以先行课程实现的计算机硬件系统为基础，这样才能形成一致的、完整的计算机系统整体概念。

本研究还对12门课程的规划思路、主要教学内容及实验内容进行了研究和阐述，具体内容详见公开发表的研究报告。

4. 关于本研究项目及本系列教材

机械工业出版社华章公司在较早的时间就引进出版了MIT、UC-Berkeley、CMU等国际知名院校有关计算机系统课程的多种教材，并推动和组织了计算机系统能力培养相关的研究，对

国内计算机系统能力培养起到了积极的促进作用。

本项研究是教育部 2013 ~ 2017 年计算机类专业教学指导委员会“计算机类专业系统能力培养研究”项目之一，研究组成员由国防科技大学王志英、北京航空航天大学马殿富、西北工业大学周兴社、南开大学吴功宜、武汉大学何炎祥、南京大学袁春风、北京大学陈向群、中国科技大学安虹、天津大学张钢、机械工业出版社华章公司温莉芳等组成，研究报告分别发表于中国计算机学会《中国计算机科学技术发展报告》及《计算机教育》杂志。

本系列教材编委会在上述研究的基础上对本套教材的出版工作经过了精心策划，选择了対系统观教育和系统能力培养有研究和实践的教师作为作者，以系统观为核心编写了本系列教材。我们相信本系列教材的出版和使用，将对提高国内高校计算机类专业学生的系统能力和整体水平起到积极的促进作用。

“计算机类专业系统能力培养系列教材”编委会组成如下：

主任 王志英

副主任 马殿富

委员 周兴社 吴功宜 何炎祥 袁春风 陈向群 安虹 温莉芳

秘书 姚蕾

“计算机类专业系统能力培养系列教材”编委会

推荐序

“计算机体系结构”(Computer Architecture)也称为“计算机系统结构”,是计算机科学与技术一级学科下最重要的二级学科。“计算机体系结构”是研究怎么造计算机而不是怎么用计算机的学科。我国学者在如何用计算机的某些领域的研究已走到世界前列,例如最近很红火的机器学习领域,中国学者发表的论文数和引用数都已超过美国,位居世界第一。但在如何造计算机的领域,参与研究的科研人员较少,科研水平与国际上还有较大差距。2016年国家自然科学基金会计算机学科的面上项目共有4863项申请,但申报“计算机体系结构”(F0203)方向的项目只有22项,占总申报项目的0.45%,而计算机图像与视频处理方向的申报项目有439项。

做计算机体系结构方向研究的科研人员较少与大学及研究生的课程教育直接相关。计算机体系结构是工程性很强的学科,而我国的大学老师大多没有机会实际参与设计CPU和操作系统,对计算机的软硬件工作过程不能融会贯通,教学时只能照本宣科,学生只学到一些似懂非懂的名词概念,难以培养“造计算机”的兴趣。目前全国许多高校使用从国外翻译的体系结构教材,John L. Hennessy和David A. Patterson合著的《计算机体系结构:量化研究方法》已经不断改版至第5版,被认为是计算机体系结构的经典教材,但此书有近千页之厚,本科生未必都能接受。国内也出版了不少体系结构(系统结构)方面的教材,但多数兼顾了研究生和参考书的需求。因此,迫切需要一本为本科生量身定制的计算机体系结构精品教材。

摆在读者面前的这本《计算机体系结构基础》就是为满足本科教育而编著的精品教材。过去出版的体系结构教材大多是“眼睛向上”编写的,作者既考虑了做本科教材的需求,又考虑了参考书的需求,为了体现参考书的技术前瞻性,往往会包含一些未经受考验的新技术。而本书是作者在2011年已经出版的硕士生教材《计算机体系结构》的基础上,“眼睛向下”编著的本科生教材,多年的研究生授课使作者十分明确本科生应学习哪些体系结构的基础知识。凡写进这本教科书的内容都是本科生应该掌握的知识,不会为追求时髦而增加额外的内容。

与过去出版的计算机体系结构教科书相比,本书有以下几个特点:


第一个特点是特别重视知识的基础性。计算机发明至今已经70余年,曾经用来造计算机的技术多如牛毛,计算机期刊与会议上发表的文章数以万计,但是许多技术如过往烟云,已经被丢进历史的垃圾堆。我在美国读博士时,一位很有权威的教授讲了一个学期计算机体系结构课,基本上都是讲并行计算机的互连结构(Interconnection),如蝶形互连(Butterfly)、超立方体互连(Hypercube)、胖树互连(Fat Tree)等,现在这些内容已不是计算机界普遍关心的问题。20世纪90年代,计算机体系结构国际会议(ISCA)几乎成了专门讨论缓存(Cache)技术的会议,但没有几篇文章提出真正可用的缓存技术,以至于计算机界的权威John L. Hennessy

教授 1997 年说出这样的话：“把 1990 年以来计算系统结构方面所有的论文都烧掉，对计算机系统结构没有任何损失。”本书作者在“自序”中写道：“计算机体系结构千变万化，但几十年发展沉淀下来的原理性的东西不多，希望从体系结构快速发展的很多现象中找出一些内在的、本质的东西。”毛泽东在《实践论》中归纳总结了十六个字：“去粗取精，去伪存真，由此及彼，由表及里。”本书作者遵循这十六个字的精神，对几十年的计算机体系结构技术做了认真的鉴别、选择和对比、分析，写进教科书的内容是经得起历史考验的基础知识。

第二个特点是强调“一以贯之”的系统性。“计算机系统结构”的关键词是“系统”而不是“结构”，国外做计算机系统结构研究的学者介绍自己时往往是说：“我是做系统（System）研究的。”计算机专业的学生应具有系统层面的理解能力，能站在系统的高度解决应用问题。对计算机系统是否有全面深入的了解是区别计算机专业人才和非专业人才的重要标志。长期以来我们采用“解剖学”的思路进行计算机教学，按照硬件、软件、应用等分类横切成几门相对独立的课程，使得计算机系毕业的学生对整个计算机系统缺乏完整的理解。如果问已经学完全部计算机课程的学生，在键盘上敲一个空格键到屏幕上的 PPT 翻一页，在这一瞬间计算机中哪些硬件和软件在运转，如何运转，可能绝大多数学生都讲不清楚。本书有若干章节专门讲述计算机的软硬件协同、计算机系统启动过程等，着力培养学生的全局思维能力。为了使学生一开始就对计算机有全局的框架性认识，此教材的第 1 章对全书内容做了尽可能通俗易懂的描述，这是追求系统性教学的刻意安排。本书作者强调：“一个计算机体系结构设计人员就像一个带兵打仗的将领，要学会排兵布阵。要上知天文、下知地理，否则就不会排兵布阵，或者只会纸上谈兵地排兵布阵，只能贻误军国大事。”这里讲的“天文”是指应用程序、编译程序和操作系统；“地理”是指逻辑、电路和工艺。只有上下贯通，才能真正掌握计算机体系结构。

第三个特点是强调能在硅上实现的实践性。由于 CMOS 电路集成度的指数性提高，一块 CPU 芯片已可以集成几十亿晶体管。计算机体系结构的许多知识现在都体现在 CPU 中，因此从某种意义上讲，不懂 CPU 设计就不能真正明白计算机体系结构的奥妙。CPU 的结构通常称为微体系结构，主要在硕士课程中讲授，但本科生的体系结构课程也应学习在硅上能实现的技术。陆游诗云：“纸上得来终觉浅，绝知此事要躬行。”只会 P2P 的学习（从 Paper 到 Paper 的学习）往往学不到真本事，只有最后能“躬行”到硅上的知识才是过硬的知识。本书作者有十几年从事 CPU 设计的经验，能正确区分哪些是纸上谈兵的知识，哪些是能落实到硅上的知识，这是他们独特的优势。在中国科学院大学的本科教学中，计算机体系结构课程还辅以高强度的实验课，实践证明这对学生真正理解课堂学到的知识大有好处。

本书内容用于课堂教学只有两年时间，还需要不断听取学生的反馈意见和同行的批评建议，希望经过几年的完善修改，本书能真正成为受到众多大学普遍欢迎的精品教材。



2017 年 8 月

F O R E W O R D

自 序

计算机专业有几门“当家”的核心课程是关于“如何造计算机”的，硬件方面以计算机组成原理和计算机体系结构为主，软件方面以操作系统和编译原理为主。其他如离散数学、编程语言、数据结构、数字逻辑等作为计算机专业的学科基础课也很重要，除了计算机专业，其他使用计算机的专业如自动化专业、电子专业也在学。

我从2001年就开始从事龙芯处理器的研发，并从2005年起在中国科学院大学教授计算机体系结构课程，其间接触了很多从各高校计算机专业毕业的学生，发现他们在大学时主要练就了诸如编程等“怎么用计算机”的本领，对操作系统和体系结构这种“如何造计算机”的课程，或者没有系统学习，或者只学到一些概念。比如对于“从打开电源到计算机启动再到登录界面”或者“从按一下空格键到翻一页PPT”这样的过程，如果问及计算机系统内部包括CPU、南北桥、GPU在内的硬件以及包括操作系统和应用程序在内的软件是如何协同工作的，计算机专业毕业的学生几乎没有人说得明白。

我1986年到中国科学技术大学计算机系学习的时候，教授我计算机体系结构课程的老师都是亲自造过计算机的，他们能够讲明白计算机软硬件工作的原理性过程。改革开放以来，我国主要使用国外的CPU和操作系统“攒”计算机，学术界也几乎不从事CPU和操作系统这种核心技术的研究工作，全国两千多个计算机专业主要使用国外教材或者翻译的国外教材教授学生“如何造计算机”。由于计算机体系结构和操作系统都是工程性很强的学科，而任课老师却没有机会参与设计CPU和操作系统，因此教学生的时候难免照本宣科，使学生只学到一些概念，难以对计算机的软硬件工作过程融会贯通。

发展以CPU和操作系统为代表的自主基础软硬件，是国家的战略需求，而人才培养是满足该战略需求的必要条件。因此，自2005年开始，我便结合龙芯CPU的实践在中国科学院研究生院开设计算机体系结构课程，并于2011年依托清华大学出版社出版了《计算机体系结构》教材。2014年，中国科学院大学设立并开始招收本科生，要求我也给本科生讲授计算机体系结构课程。刚开始觉得难度很大，因为计算机体系结构非常复杂，给研究生讲清楚都不容易，给本科生讲清楚就更难。

经过反复思考，觉得可以利用这个机会，建设包括本科生、硕士生、博士生在内的计算机体系结构课程体系，由浅入深地培养“造计算机”的人才。为此，我们计划编写一套分别面向本科生、硕士生、博士生的“计算机体系结构”课程教材。

面向本科生的教材为《计算机体系结构基础》。主要内容包括：作为软硬件界面的指令系统结构，包含 CPU、GPU、南北桥协同的计算机硬件结构，CPU 的微结构，并行处理结构，计算机性能分析等。上述面面俱到的课程安排主要是考虑到体系结构学科的完整性，但重点是软硬件界面及计算机硬件结构，微结构则是硕士课程的主要内容。

面向硕士生的教材为《计算机体系结构》。主要介绍 CPU 的微结构，包括指令系统结构、二进制和逻辑电路、静态流水线、动态流水线、多发射流水线、运算部件、转移猜测、高速缓存、TLB、多核对流水线的影响等内容。

面向博士生的教材为《高级计算机体系结构》。中科院计算所的“高级计算机体系结构”课程是博士生精品课程的一部分，主要强调实践性，使学生通过设计真实的（而不是简化的）CPU，运行真实的（而不是简化的）操作系统，对结构设计、物理设计、操作系统软件做到融会贯通。

在此基础上，还将推出计算机体系结构实验平台和实验教材。

这套教材的编写突出以下特点：一是系统性，体系是“系统的系统”，很难脱离软硬件环境纯粹就体系结构本身讲解计算机体系结构，需要对体系结构、基础软件、电路和器件融会贯通；二是基础性，计算机体系结构千变万化，但几十年发展沉淀下来的原理性的东西不多，希望从体系结构快速发展的很多现象中找出一些内在的、本质的东西；三是实践性，计算机体系结构是实践性很强的学科，要设计在“硅”上运行而不是在“纸”上运行的体系结构。

胡伟武

前 言

计算机体系结构是一门比较抽象的学科，很有可能经过一个学期的学习只学到一些概念。本课程教学希望达到三个目的。一是建立学生的系统观。计算机系统的复杂性体现在计算机中各部分之间的关系非常复杂。如苹果 iPhone 的 CPU 性能不如 Intel 的 X86 CPU，但用户体验明显好于桌面计算机，这就是系统优化的结果。希望学生学完这门课程后能够从系统的角度看待计算机，不再简单地以主频论性能，或者简单地把用户体验归结于 CPU 的单项性能。二是掌握计算机体系结构的若干概念。计算机体系结构中的概念很多，虽然抽象，但是必须掌握。比如计算机体系结构的四大设计原则，指令系统结构，处理器流水线，等等。三是掌握一些重点知识并具备一些重点能力。主要包括：计算机的 ABI 接口，存储管理中的虚实地址转换过程，通过 IO 地址空间扫描进行 IO 设备初始化，计算机系统的启动过程，重要总线如 AXI 总线、内存总线、PCIe 总线的信号及其时序，用 Verilog 编写 RTL 代码的能力，先行进位加法器的逻辑，两位一乘补码乘法器逻辑，用 Perf 或 Oprofile 进行性能分析的能力，等等。

本书第一部分为引言，介绍体系结构研究内容、主要性能指标、发展趋势以及设计原则。计算机体系结构 (Computer Architecture) 是描述计算机各组成部分及其相互关系的一组规则和方法，是程序员所看到的计算机属性。计算机体系结构主要研究内容包括指令系统结构 (Instruction Set Architecture, 简称 ISA) 和计算机组织结构 (Computer Organization)。微体系结构 (Micro-architecture) 是微处理器的组织结构，并行体系结构是并行计算机的组织结构。冯·诺依曼结构的存储程序和指令驱动执行原理是现代计算机体系结构的基础。

本书第二部分介绍以指令系统结构为核心的软硬件界面，包括指令系统总体介绍、指令集结构、异常与中断、存储管理、软硬件协同等内容。贯穿该部分内容的核心思想是建立高级语言 (如 C 语言) 与指令系统结构的关系。例如，C 语言的语句与指令系统的关系，算术语句可直接映射为相关运算指令，for 循环映射为条件跳转，switch 语句映射为跳转索引和跳转表等；操作系统中地址空间的组织与指令访问内存的关系，静态全局变量映射到地址空间的静态数据区、局部变量映射到堆栈区、动态分配的数据则映射到进程空间的堆中；操作系统中进程和线程的表示及切换在指令和地址映射方面的具体体现；敲击键盘和移动鼠标等事件如何通过指令系统的外部中断传递到 CPU，以及指令系统对操作系统处理外部中断的必要支持；等等。

本书第三部分介绍计算机硬件结构。该部分的核心思想是搞清楚计算机内部包括 CPU、GPU、内存、IO 之间是如何协同完成软件规定的各种操作的。例如，在计算机开机过程中，

BIOS 完成硬件初始化后把操作系统从硬盘拷贝到内存执行的过程中南北桥与 CPU 是如何配合的；CPU 和 GPU 是如何协同操作完成计算机屏幕显示的，在显示过程中哪些活是 CPU 干的，哪些活是 GPU 干的；包括以太网、USB 等在内的各种 IO 接口的驱动在硬件上的具体体现是什么；等等。

本书第四部分介绍微结构。该部分的核心思想是建立指令系统和晶体管之间的“桥梁”。微结构是决定 CPU 性能的关键因素。由于微结构是“计算机体系结构”硕士课程的主要内容，因此本科课程的微结构内容在追求系统地介绍有关概念的基础上，重点把先行进位加法器和五级静态流水线讲透，希望学生通过对先行进位加法器和五级静态流水线的深入了解，举一反三地了解微结构的实现方式。微结构中动态流水线、乱序执行和多发射等内容只做概念性的介绍。

本书第五部分介绍并行处理结构。应用程序的并行行为是并行处理的基础，现代计算机通过多层次的并行性开发来提高性能。并行处理编程模型包括消息传递模型（如 MPI）和共享存储模型（如 OpenMP）等。多核处理器的设计需要考虑存储一致性模型、高速缓存一致性协议、片上互连、多核同步等核心问题。

本书第六部分介绍计算机的性能分析方法。性能不是由一两个具体指标（如主频）决定的，而是若干因素综合平衡的结果；性能评测也没有绝对合理公平的办法，不同的计算机对不同的应用适应性不一样，对某类应用甲计算机比乙计算机性能高，对另外一类应用可能反之。巨大的设计空间和工作负载的多样性，导致计算机系统的性能分析和评价成为一个非常艰巨的任务。计算机性能分析的主要方法包括理论建模，用模拟器进行性能模拟，以及对实际系统进行性能评测等。

上述面面俱到的课程安排主要是考虑到体系结构学科的完整性，但本科课程重点是软硬件界面及计算机硬件结构。对于一般高校，并行处理结构和计算机系统性能分析可以不讲。

在选修本课程前，学生应对 C 语言程序设计、数字逻辑电路有一定的基础。本课程试图说明一个完整的计算系统的工作原理，其中涉及部分操作系统的知识。为了有更好的理解，还可以同时选修操作系统课程。课程中的实例和原理介绍以 MIPS 体系结构为主。与传统课程中讲授的 X86 体系结构相比，MIPS 结构相对简单明晰而又不失全面。学生可以通过配套的实验课程，自底而上构建自己的计算机系统，包括硬件、操作系统以及应用软件，从而对“如何造计算机”有更深刻的认识。

由于时间仓促，本书还存在不少问题，主要包括不少内容的详略失衡，文字严谨性不够，举例不够导致很多内容过于抽象。好在总体结构基本完整了，在下一版改版时可以重点解决上述问题。下一版改版时争取精选 100 个例子覆盖本书重要内容，使计算机体系结构进一步从抽象走向具体化。

目 录

丛书序言

推荐序

自序

前言

第一部分 引言

第1章 引言	2
1.1 计算机体系结构的研究内容	2
1.1.1 一以贯之	2
1.1.2 什么是计算机	5
1.1.3 计算机的基本组成	6
1.2 衡量计算机的指标	8
1.2.1 计算机的性能	8
1.2.2 计算机的价格	10
1.2.3 计算机的功耗	11
1.3 计算机体系结构的发展	12
1.3.1 摩尔定律和工艺的发展	13
1.3.2 计算机应用和体系结构	17
1.3.3 计算机体系结构发展	18
1.4 体系结构设计的基本原则	20
1.4.1 平衡性	20
1.4.2 局部性	21
1.4.3 并行性	22
1.4.4 虚拟化	23
1.5 本章小结	24
习题	24

第二部分 指令系统结构

第2章 指令系统	26
2.1 指令系统简介	26
2.2 指令系统设计原则	26
2.3 指令系统发展历程	28
2.3.1 指令集的演变	28
2.3.2 存储管理的演变	29
2.3.3 运行级别的演变	31
2.4 本章小结	32
习题	32
第3章 指令集结构	33
3.1 地址空间	33
3.2 操作数	36
3.2.1 数据类型	36
3.2.2 访存地址	36
3.2.3 寻址方式	36
3.3 指令操作和编码	37
3.4 RISC 指令集比较	38
3.5 C 语言的机器表示	43
3.6 本章小结	46
习题	47
第4章 异常与中断	48
4.1 异常分类	48
4.2 异常处理	48
4.3 中断	50
4.3.1 中断的优先级和原子性	50
4.3.2 向量化中断	51
4.3.3 中断传递机制	52

4.4 本章小结	52	7.3.2 CPU - 北桥 - 南桥三片 结构	81
习题	52	7.3.3 CPU - 弱北桥 - 南桥 三片结构	81
第5章 存储管理	53	7.3.4 CPU - 南桥两片结构	82
5.1 存储管理的原理	53	7.3.5 SoC 单片结构	82
5.2 TLB 的结构和使用	55	7.4 处理器和 IO 设备间的通信	82
5.3 TLB 异常的处理	57	7.4.1 IO 寄存器寻址	83
5.4 本章小结	60	7.4.2 处理器和 IO 设备之间的 同步	83
习题	61	7.4.3 存储器和 IO 设备之间的 数据传送	84
第6章 软硬件协同	62	7.4.4 IO 中断控制器	85
6.1 函数调用规范	62	7.5 本章小结	86
6.1.1 MIPS ABI 整数寄存器 约定	62	习题	86
6.1.2 MIPS ABI 函数调用 约定	63	第8章 计算机总线接口技术	87
6.1.3 MIPS 堆栈布局	64	8.1 总线概述	87
6.2 中断的生命周期	66	8.2 总线分类	88
6.3 系统调用过程	67	8.3 片上总线	88
6.4 同步与通信	68	8.4 内存总线	93
6.4.1 基于互斥的同步机制	68	8.5 系统总线	99
6.4.2 非阻塞的同步机制	69	8.5.1 HyperTransport 总线	99
6.5 本章小结	70	8.5.2 HyperTransport 包格式	102
习题	70	8.6 IO 总线	103
第三部分 计算机硬件结构		8.6.1 PCIE 总线	103
第7章 计算机组成原理和结构	72	8.6.2 PCIE 包格式	104
7.1 冯·诺依曼结构	72	8.7 本章小结	105
7.2 计算机的组成部件	73	习题	105
7.2.1 运算器	73	第9章 计算机系统启动过程 分析	106
7.2.2 控制器	73	9.1 处理器核初始化	106
7.2.3 存储器	74	9.1.1 处理器复位	107
7.2.4 输入设备	77	9.1.2 调试接口初始化	107
7.2.5 输出设备	77	9.1.3 TLB 初始化	108
7.3 计算机系统硬件结构发展	79	9.1.4 Cache 初始化	109
7.3.1 CPU - GPU - 北桥 - 南桥四片结构	80	9.2 总线接口初始化	111