



普通高等教育“十五”国家级规划教材

# 计

# 算机体系结构

## 第二版

张晨曦 王志英 张春元 戴葵 肖晓强



高等教育出版社

普通高等教育“十五”国家级规划教材

# 计算机体系结构

第二版

张晨曦 王志英 张春元 戴 葵 肖晓强

高等教育出版社

## 内容提要

本书是普通高等教育“十五”国家级规划教材,第一版被列为“面向 21 世纪课程教材”,并于 2002 年获全国普通高等学校优秀教材二等奖。为了跟踪和反映计算机技术的发展,新版教材在第一版的基础上进行了较大的修改,补充了新内容,进一步提高了可读性和系统性。

本书除了着重论述体系结构的基本概念、基本原理、基本结构和基本分析方法以外,还强调采用量化的分析方法,使读者能更具体、实际地分析和设计计算机体系结构。

全书共 7 章:计算机体系结构的基本概念,计算机指令集结构设计,流水线技术,指令级并行,存储层次,输入/输出系统,多处理机。主要内容有:计算机体系结构的概念以及并行性概念的发展,DLX,流水线的基本概念和性能分析,流水线的相关问题,向量处理机,指令动态调度,超标量和超长指令字技术,分支处理技术,Cache 的基本知识及提高性能的方法,廉价磁盘冗余阵列 RAID,I/O 系统性能分析,多处理机的存储器体系结构、互连网络、同步机制以及同时多线程技术等。全书比较全面地介绍了当今计算机体系结构的发展前沿,并配有大量的实例分析。

本书可作为高等学校计算机及相关专业本科生或研究生计算机体系结构课程的教材,也可供从事计算机事业的工程技术人员参考。

## 图书在版编目(CIP)数据

计算机体系结构 / 张晨曦等. — 2 版. — 北京:高等教育出版社,2005.6(2006 重印)

ISBN 7-04-016647-X

I. 计... II. 张... 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(2005)第 036718 号

出版发行 高等教育出版社  
社 址 北京市西城区德外大街 4 号  
邮政编码 100011  
总 机 010-58581000  
经 销 蓝色畅想图书发行有限公司  
印 刷 天津新华二印刷有限公司  
开 本 787×1092 1/16  
印 张 21.5  
字 数 470 000

购书热线 010-58581118  
免费咨询 800-810-0598  
网 址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
网上订购 <http://www.landaco.com>  
<http://www.landaco.com.cn>  
畅想教育 <http://www.widedu.com>  
版 次 2000 年 6 月第 1 版  
2005 年 6 月第 2 版  
印 次 2006 年 4 月第 3 次印刷  
定 价 26.90 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 16647-00

## 第一作者简介

张晨曦,男,1960年9月生,汉族,福建龙岩人,国防科技大学计算机学院教授,博士生导师,国家级“中青年有突出贡献专家”。主要成就有:先后主持了四项国家自然科学基金项目,曾获“国家杰出青年基金”资助。发表论文八十多篇,其中在国外发表二十多篇。有18篇被国际著名八大检索工具收录。撰写专著两部(第二作者),教材一本(第一作者)。其中专著《新一代计算机》由荷兰 North-Holland 出版社出版,另一部1992年获“国家教委优秀专著特等奖”,1993年获“全国优秀科技图书一等奖”。教材《计算机体系结构》2002年获“全国普通高等学校优秀教材二等奖”。获部委级科技进步一等奖两项(排名第二),二等奖一项(排名第一),获部委级教学成果二、三等奖各一项。负责研制的新型 Internet 动画制作/演播系统“网动王”于1999年3月通过了专家鉴定,达到了国际先进水平,并已在远程教育和 CAI 中获得了较广泛的应用。1991年被国家教委授予“做出突出贡献的中国博士”光荣称号,被评为湖南省科技青年“十佳”之一,1993年被评为“全军优秀教师”,1993年和1995年两次获“霍英东青年教师奖”;1995年获“中国青年科技奖”。从15岁起当中学教师,对教学方法和现代教育技术有深入的研究,提出了面向远程教育和 CAI 的教学方法——多媒体图形解析教学法。业余爱好:摄影。

通讯地址:湖南长沙 国防科技大学 计算机学院

邮编:410073

网址:www.FotoSky.com, www.GotoSchool.net

E-mail: architecture21@sohu.com

## 第二版前言

本书是高等学校计算机专业本科生及研究生计算机体系结构课程的通用教材。在内容的选择上,本书不打算覆盖体系结构的各个方面,不想成为大而全的参考手册,而是重点论述现代大多数计算机都采用的比较成熟的思想、结构和方法等,并且力求做到深入浅出、通俗易懂。

“计算机体系结构”是计算机领域的一门重要学科。它强调从总体结构、系统分析这一角度来研究计算机系统。通过学习这本书,读者能把在“计算机组成原理”、“数据结构”、“操作系统”、“汇编语言程序设计”等课程中所学的软、硬件知识有机地结合起来,从而建立起计算机系统的完整概念。

本书第一版于2000年由高等教育出版社出版,并被列为“面向21世纪课程教材”。该书已被许多学校选为教材,并于2002年获得全国普通高等学校优秀教材二等奖。为了跟踪和反映计算机技术的发展,我们在第一版的基础上进行了较大的修改,并补充了新内容,进一步提高了可读性和系统性。

本书除了着重论述体系结构的基本概念、基本原理、基本结构和基本分析方法以外,还强调采用量化的分析方法。这种方法使我们能更具体、实际地分析和设计计算机体系结构。书中用了大量的例题说明如何进行量化分析。

全书共包括7章。第1章论述计算机体系结构的概念以及体系结构和并行性概念的发展,讨论影响计算机系统设计的成本和价格因素,并介绍定量分析技术基础。第2章论述计算机指令集结构设计中的一些问题,包括寻址技术、指令集的功能设计、操作数的类型和大小、指令格式等,并且介绍一种指令集结构的实例——DLX。第3章为“流水线技术”,论述流水线的概念和性能分析、DLX流水线、流水线中的相关问题,并对向量处理机的结构及特点、关键技术以及性能评价进行讨论。第4章为“指令级并行”,论述利用硬、软件技术开发程序中存在的指令间并行性的技术和方法,包括指令动态调度、超标量技术、分支处理技术和超长指令字技术等。第5章为“存储层次”,论述Cache的基本知识、降低Cache失效率的方法、减少Cache失效开销的方法以及减少命中时间的方法,并对虚拟存储器进行讨论。第6章为“输入/输出系统”,论述外部存储设备、I/O设备与CPU和存储器的连接、廉价磁盘冗余阵列RAID,并讨论I/O系统性能分析。第7章为“多处理机”,论述多处理机的存储器体系结构、互连网络、同步机制以及同时多线程技术,并讨论一个多处理机实例。

作为教材,使用者可以根据自己的需求,选取相应的内容进行教学。有3种建议方案:

方案1:

全部7章的内容,根据具体要求,可以安排60~80个学时。

方案2:

第 1 章 → 第 2 章 → 第 3 章 → 第 5 章 → 第 6 章

如果要建立相对完整的单机系统体系结构的概念,可以采用本方案。建议安排 50~60 个学时。

方案 3:

第 1 章 → 第 2 章 → 第 3 章 → 第 5 章

对于仅仅需要建立基本体系结构概念的使用者,可以采用本方案。本方案在对第 3 章、第 5 章内容略作删减的情况下,只需约 40 个学时。如果读者已经了解指令集的有关概念,还可以跳过第 2 章。在这种情况下,可以安排约 32 学时的短课程。

本书由国防科技大学计算机学院的张晨曦教授编写第 5 章,王志英教授编写第 7 章,张春元教授编写第 1 章和第 4 章,戴葵副教授编写第 2 章和第 3 章,肖晓强副教授编写第 6 章。

由于作者水平有限,书中难免有错误和不妥之处,敬请读者批评指正。

作 者

2004 年 12 月于长沙

# 第一版前言

本书是高等学校计算机专业本科生及研究生计算机体系结构课程的通用教材。为了适应面向 21 世纪计算机类专业教学内容和课程体系改革的需要,国家教育部高等学校计算机科学与技术教学指导委员会统一组织编写了计算机科学与技术专业“九五”规划教材。其中,《计算机体系结构》、《计算机组成原理》和《微型计算机技术》是重点组织的系列教材。本书是该系列教材之一。

在内容的选择上,本书不打算覆盖计算机体系结构的各个方面,不想成为大而全的参考手册,而是重点论述现代大多数计算机都采用的比较成熟的思想、结构和方法等,并且力求做到深入浅出,通俗易懂。

本书除了着重论述计算机体系结构的基本概念、基本原理、基本结构和基本分析方法以外,还强调采用量化的分析方法。这种方法使我们能更具体、实际地分析和设计计算机体系结构。书中用了大量的例题说明如何进行量化分析。

本书共七章。第一章论述计算机体系结构的概念以及体系结构和并行性概念的发展,并简单地讨论影响计算机系统设计的成本和价格因素。第二章论述计算机指令集结构设计中的一些问题,包括寻址技术、指令集的功能设计、操作数的类型和大小、指令格式等,并且介绍一种指令集结构的实例——DLX。第三章为流水线技术,论述流水线的基本概念、分类以及性能计算方法,并对流水线中的相关问题以及向量计算机进行讨论。第四章为指令级并行,论述利用硬、软件技术开发程序中存在的指令间并行性的技术和方法,包括指令调度、超标量技术、分支处理技术和超长指令字技术。第五章为存储层次,论述 Cache 的基本知识、降低 Cache 失效率的方法、减少 Cache 失效开销的方法以及减少命中时间的方法,并对主存和虚拟存储器进行讨论。第六章为输入输出系统,论述存储设备、总线和通道,并讨论 I/O 与操作系统的关系以及 I/O 系统设计。第七章为多处理机,论述多处理机的存储器体系结构、互连网络以及同步与通信,并对并行化技术和多处理机实例进行讨论。

本书由国防科技大学计算机学院的张晨曦教授编写第五章,王志英教授编写第七章,张春元博士编写第一章和第四章,戴葵博士编写第二章和第三章,朱海滨博士编写第六章。

本书由复旦大学的朱传琪教授主审,提出了宝贵的修改意见。在本书编写过程中,得到了中国科学技术大学陈国良教授的大力支持,并得到了国防科学技术大学计算机学院领导和有关师生的多方面帮助。在此一并表示衷心的感谢。

由于作者水平有限,书中难免有错误和不妥之处,敬请读者批评指正。

本书有配套的 CAI 课件和 PowerPoint 讲稿,详见:

<http://www.GotoSchool.net>

作者

1999年3月于长沙

# 郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

**反盗版举报电话：**(010) 58581897/58581896/58581879

**传 真：**(010) 82086060

**E - mail：**dd@hep.com.cn

**通信地址：**北京市西城区德外大街4号

高等教育出版社打击盗版办公室

**邮 编：**100011

**购书请拨打电话：**(010)58581118

策划编辑	倪文慧
责任编辑	张海波
封面设计	于文燕
责任绘图	朱 静
版式设计	范晓红
责任校对	张 颖
责任印制	韩 刚



# 目 录

<b>第 1 章 计算机体系结构的基本概念</b> .....	1	<b>第 2 章 计算机指令集结构设计</b> .....	43
1.1 引言 .....	1	2.1 指令集结构的分类 .....	43
1.2 计算机体系结构的概念 .....	2	2.1.1 指令集结构的分类 .....	43
1.2.1 计算机系统层次概念 .....	3	2.1.2 通用寄存器型指令集结构分类 .....	44
1.2.2 计算机体系结构 .....	5	2.2 寻址技术 .....	46
1.2.3 计算机组织和计算机实现技术 .....	6	2.3 指令集结构的功能设计 .....	49
1.3 计算机体系结构的发展 .....	9	2.3.1 CISC 指令集结构的功能设计 .....	50
1.3.1 存储程序计算机体系结构 及其发展 .....	9	2.3.2 RISC 指令集结构的功能设计 .....	52
1.3.2 计算机的分代和分型 .....	13	2.3.3 控制指令 .....	53
1.3.3 应用需求的发展 .....	14	2.4 操作数的类型、表示和大小 .....	55
1.3.4 计算机实现技术的发展 .....	15	2.5 指令集格式的设计 .....	57
1.3.5 技术的挑战和体系结构的 生命周期 .....	17	2.5.1 寻址方式的表示方法 .....	57
1.4 影响计算机体系结构的成本和 价格因素 .....	20	2.5.2 指令集格式的选择 .....	58
1.4.1 计算机系统的成本和价格 .....	20	2.6 DLX 指令集结构 .....	59
1.4.2 时间因素 .....	23	2.6.1 DLX 指令集结构 .....	59
1.4.3 集成电路的成本 .....	24	2.6.2 DLX 指令集结构效能分析 .....	64
1.5 定量分析技术基础 .....	27	2.7 小结 .....	65
1.5.1 计算机性能的评测 .....	27	习题 2 .....	66
1.5.2 测试程序 .....	28	<b>第 3 章 流水线技术</b> .....	68
1.5.3 性能设计和评测的基本原则 .....	32	3.1 流水线的概念 .....	68
1.5.4 CPU 的性能 .....	34	3.1.1 流水线的概念 .....	68
1.6 计算机体系结构中并行性的 发展 .....	36	3.1.2 流水线的分类 .....	70
1.6.1 并行性概念 .....	36	3.2 DLX 的基本流水线 .....	73
1.6.2 提高并行性的技术途径 .....	37	3.2.1 DLX 的一种简单实现 .....	73
1.6.3 并行计算的应用需求 .....	40	3.2.2 基本的 DLX 流水线 .....	77
1.7 小结 .....	40	3.2.3 流水线性能分析 .....	80
习题 1 .....	41	3.3 流水线中的相关 .....	86
		3.3.1 流水线的结构相关 .....	86
		3.3.2 流水线的数​​据相关 .....	89
		3.3.3 流水线的控制相关 .....	97
		3.4 MIPS R4000 流水线计算机	

简介 .....	104	5.2 Cache 基本知识 .....	177
3.4.1 MIPS R4000 整型流水线 .....	104	5.2.1 映象规则 .....	177
3.4.2 MIPS R4000 浮点流水线 .....	107	5.2.2 查找方法 .....	180
3.5 向量处理机 .....	108	5.2.3 替换算法 .....	181
3.5.1 向量处理方式和向量处理机 .....	108	5.2.4 写策略 .....	182
3.5.2 提高向量处理机性能的主要 技术 .....	111	5.2.5 Cache 的结构 .....	184
3.5.3 向量处理机的性能评价 .....	115	5.2.6 Cache 性能分析 .....	186
3.6 小结 .....	119	5.2.7 改进 Cache 性能 .....	190
习题 3 .....	120	5.3 降低 Cache 失效率的方法 .....	191
<b>第 4 章 指令级并行</b> .....	123	5.3.1 增加 Cache 块大小 .....	193
4.1 指令级并行的概念 .....	123	5.3.2 提高相联度 .....	195
4.1.1 循环展开调度的基本方法 .....	124	5.3.3 Victim Cache .....	197
4.1.2 相关性 .....	128	5.3.4 伪相联 Cache .....	197
4.2 指令的动态调度 .....	133	5.3.5 硬件预取技术 .....	199
4.2.1 动态调度的原理 .....	133	5.3.6 由编译器控制的预取 .....	200
4.2.2 动态调度算法之一:记分牌 .....	134	5.3.7 编译器优化 .....	202
4.2.3 动态调度算法之二:Tomasulo 算法 .....	142	5.4 减少 Cache 失效开销 .....	206
4.3 控制相关的动态解决技术 .....	152	5.4.1 让读失效优先于写 .....	207
4.3.1 分支预测缓冲 .....	152	5.4.2 子块放置技术 .....	207
4.3.2 分支目标缓冲 .....	155	5.4.3 请求字处理技术 .....	208
4.3.3 基于硬件的前瞻执行 .....	157	5.4.4 非阻塞 Cache 技术 .....	209
4.4 多指令流出技术 .....	161	5.4.5 采用两级 Cache .....	210
4.4.1 静态超标量技术 .....	162	5.5 减少命中时间 .....	213
4.4.2 动态多指令流出技术 .....	164	5.5.1 容量小、结构简单的 Cache .....	214
4.4.3 超长指令字技术 .....	167	5.5.2 虚拟 Cache .....	214
4.4.4 多流出处理器受到的限制 .....	168	5.5.3 写操作流水化 .....	216
4.5 小结 .....	170	5.5.4 Cache 优化技术总结 .....	217
习题 4 .....	170	5.6 主存 .....	218
<b>第 5 章 存储层次</b> .....	173	5.7 虚拟存储器 .....	225
5.1 存储器的层次结构 .....	173	5.7.1 虚拟存储器基本原理 .....	225
5.1.1 从单级存储器到多级存储器 .....	173	5.7.2 快表 .....	226
5.1.2 存储层次的性能参数 .....	174	5.8 进程保护和虚存实例 .....	228
5.1.3 “Cache-主存”和“主存-辅存” 层次 .....	175	5.8.1 进程保护 .....	229
5.1.4 存储层次的 4 个问题 .....	177	5.8.2 页式虚存举例:Alpha AXP 的存储 管理和 21064 的 TLB .....	230
		5.9 Alpha AXP 21064 存储层次 .....	232
		5.10 小结 .....	236

习题 5 .....	237	6.7.4 块服务器和文件服务器 .....	280
<b>第 6 章 输入/输出系统</b> .....	239	6.8 小结 .....	280
6.1 引言 .....	239	习题 6 .....	281
6.1.1 I/O 系统性能与 CPU 性能 .....	239	<b>第 7 章 多处理机</b> .....	283
6.1.2 I/O 系统的可靠性 .....	240	7.1 引言 .....	283
6.2 外部存储设备 .....	241	7.1.1 并行计算机体系结构的分类 .....	283
6.2.1 磁盘设备 .....	241	7.1.2 通信模型和存储器的结构模型 .....	285
6.2.2 Flash 存储器 .....	244	7.1.3 并行处理面临的挑战 .....	287
6.2.3 磁带设备 .....	245	7.2 对称式共享存储器体系结构 .....	289
6.2.4 光盘设备 .....	246	7.2.1 多处理机 Cache 一致性 .....	289
6.3 可靠性、可用性和可信性 .....	247	7.2.2 实现一致性的基本方案 .....	290
6.4 廉价磁盘冗余阵列 RAID .....	249	7.2.3 监听协议及其实现 .....	292
6.4.1 RAID0 .....	251	7.3 分布式共享存储器体系结构 .....	293
6.4.2 RAID1 .....	252	7.3.1 基于目录的 Cache 一致性 .....	294
6.4.3 RAID2 .....	254	7.3.2 目录协议及其实现 .....	296
6.4.4 RAID3 .....	255	7.4 互连网络 .....	298
6.4.5 RAID4 .....	256	7.4.1 互连网络的性能参数 .....	298
6.4.6 RAID5 .....	257	7.4.2 静态连接网络 .....	301
6.4.7 RAID6 .....	258	7.4.3 动态连接网络 .....	305
6.4.8 RAID7 .....	259	7.5 同步 .....	309
6.4.9 RAID 的实现与发展 .....	259	7.5.1 基本硬件原语 .....	310
6.5 I/O 设备与 CPU 和存储器的 连接 .....	260	7.5.2 用一致性实现锁 .....	311
6.5.1 总线 .....	260	7.5.3 同步性能问题 .....	313
6.5.2 设备的连接 .....	263	7.5.4 大规模机器的同步 .....	316
6.5.3 通道 .....	265	7.6 同时多线程 .....	319
6.6 I/O 系统性能分析 .....	271	7.6.1 将线程级并行转换为指令级 并行 .....	320
6.6.1 I/O 性能与系统响应时间 .....	271	7.6.2 同时多线程处理器的设计 .....	321
6.6.2 Little 定律 .....	273	7.6.3 同时多线程的性能 .....	322
6.6.3 M/M/1 排队系统 .....	274	7.7 多处理机实例 .....	323
6.6.4 M/M/m 排队系统 .....	276	7.8 小结 .....	327
6.7 I/O 与操作系统 .....	277	习题 7 .....	327
6.7.1 DMA 和虚拟存储器 .....	277	<b>主要参考文献</b> .....	329
6.7.2 I/O 和 Cache 数据一致性 .....	278		
6.7.3 异步 I/O .....	279		

# 第1章 计算机体系结构的基本概念

## 1.1 引言

人们目前使用的数字电子计算机(通常称之为计算机)诞生于1945年。近六十年来,计算机技术一直处于发展和变革之中。今天,用1万元人民币购置的使用 Pentium 4 微处理器的桌面型个人计算机,其性能已经超过了在20世纪80年代初期用1千万美元购置的 Cray-1 巨型计算机。20世纪50年代,人们认为在银行里用计算机来完成现金存取业务的想法是荒唐可笑的,因为当时最便宜的计算机也需要50万美元;20世纪60年代是大型主机的时代,价值几百万美元的计算机主要用于大型金融系统和科学计算,人们认为用计算机控制汽车行驶是天方夜谭,因为当时最小巧的计算机的体积就有轿车那么大;20世纪70年代,小型计算机的身影出现在学校、实验室,人们在多个终端上分时共享一台主机;20世纪80年代,出现了放在桌面上的个人计算机(Personal Computer, PC)和工作站(workstation),它们以微处理器为核心,成本低廉,因此计算机迅速普及到个人;20世纪90年代,以微软 Windows 的成熟为代表,人性化的计算机图形操作界面开始普及,Internet 迅速成长,WWW 成为人们了解世界的新窗口,掌上计算机第一次成为现实,基于微处理器的数字电子消费品,如 VCD、电子游戏机、移动电话(手机)等大量涌现在人们身边;进入21世纪,人们看到了更加丰富多彩的计算机应用,如高清晰度电视、数字化网络视频服务、智能移动电话、GPS/PDA/移动电话等复合多功能设备、MP3 播放器、智能卡、无线网络、网络游戏、Wi-Fi 笔记本、高品质数码相机……;2003年11月18日,中国体育总局正式宣布:电子竞技成为中国体育总局承认的第99个正式体育项目!微处理器、计算机、数字化等对人们的生活、工作、学习等各方面均产生了重大及深远的影响。

计算机的发展始终受到制造技术和计算机体系结构技术的双重影响。在计算机诞生的头25年中,计算机性能增长相对缓慢。在这个过程中,制造技术和体系结构同时发挥着作用,其间充满了尝试和创新,目前广泛使用的存储程序计算机的完整概念就是这个时期产生的,人们通常称之为冯·诺依曼(Von Neumann)计算机结构。20世纪70年代以后,由于集成电路的出现,计算机性能出现了极大的飞跃,产生了一系列著名的计算机系统,包括 IBM 360/370 系列、DEC PDP 系列、CDC 6600/7600 系列等,整个20世纪70年代计算机性能的增长速度达到每年25%~30%,这种增长主要归功于以集成电路为代表的制造技术的发展。1971年,在 Intel 公司,出现了计算机历史上的第一片微处理器 4004,这是一次对计算机设计和制造技术影响深远的革命。从20世纪70年代到80年代中期,采用微处理器的计算机性能增长达到每年35%;与当时广泛使用的大、中、小型计算机相比,微处理器的性能增长更多地依赖于集成电路技术的发展。进入20世纪80年代以后,计算机体系结构产生了一次重大变革,出现了现在称之为精简

指令集计算机(Reduced Instruction Set Computer, RISC)的处理器设计技术。此后,计算机体系结构变革不断,集成电路技术为计算机设计提供的技术空间得到了充分的发挥,计算机系统的性能增长超过了每年 50%。20 世纪 90 年代中期以来,尤其是 Intel 公司的 Pentium Pro 的出现,基于 RISC 体系结构核心的微处理器大批量上市,极大地推动了计算机产业的发展,计算机的系统性能出现了全面提升。目前,计算机性能增长达到每年 50% 以上,其中包括器件技术在内的计算机制造技术增长了约 8%,其余约 42% 的部分主要依靠计算机体系结构发展的支持。1995 年时,实际的微处理器的最高性能与仅仅依靠制造技术获得的性能增长相比,差距几乎为 5 倍,而到 2001 年,这两者的差距已经达到几乎 15 倍。

微处理器出现以后,计算机系统设计、计算机市场和计算机应用都出现了较大的变化。首先,计算机用户是最直接的受益者。第二,对于市场而言,微处理器的大批量生产促成了计算机产品的批量化、标准化和市场化,这种变化也促进了计算机设计、生产和应用的良性发展。微处理器的使用,使得绝大多数计算机系统制造商无需再进行中央处理器的设计和制造了,计算机系统设计的复杂性和风险也大大下降。第三,大量兼容的微处理器、标准化的接口、高度兼容的计算机系统的出现,避免了系统程序和应用程序的重复开发。高级语言,特别是 C/C++ 语言,成为计算机系统的必备语言,减少了对目标代码兼容性的依赖,系统和应用软件开发的风险都大大减少了;操作系统和计算机语言的标准化,促成了一大批独立、标准操作系统和高级语言的出现,例如包括 Linux 在内的 UNIX 操作系统体系的产生和发展,极大地降低了采用新型体系结构的费用和 risk。目前,计算机市场基本上可化分为微处理器、计算机系统、操作系统、应用开发平台和应用系统等多层结构,世界上绝大部分公司都把自己的精力集中于其中一个或几个层面上,而不是所有层面都兼顾。

计算机系统的这种变化对现实世界的影响是巨大的。20 世纪 80 年代末,当人们第一次接触 80386 微处理器的 32 位(b)PC 系统时,主流操作系统为 DOS,进行字处理的软件主要采用字符界面,Windows 操作系统、多媒体和所见即所得的概念刚刚诞生,只有在特殊情况下才会使用局域网,功能也十分有限,而一台具有基本配置的 80386 计算机系统价格在 2 万元人民币以上,极少有个人购置计算机。32 位的 80386 微处理器,其性能相当于一台百万次计算机。人们目前广泛使用的基于 Pentium 4 微处理器的计算机系统,其性能相当于一台数亿次的 32 位计算机,Windows XP 甚至 Linux 都具有华丽、友好的基于图形的用户界面,网络连接、多媒体功能已经成为基本配置,字符界面难免踪影,家庭拥有多台计算机已经司空见惯,计算机已经可以为人们提供大量非常有用的功能,同时这些功能的使用也相对容易了,这些变化已经使人们的生活出现了更多的所谓数字化的成分。

目前,从嵌入式控制器到巨型计算机,都采用微处理器进行系统设计。我们可以预计,在计算机系统设计和计算机应用领域中,基于微处理器的系统将在相当长的时间内占支配和统治地位。这种现状对计算机体系结构的研究影响巨大,本书中所研究的很多概念都是基于这个出发点的。

## 1.2 计算机体系结构的概念

设计一种新型计算机系统首先必须面临的问题是什么呢?我们会列举出很多因素,其中最

主要的有新型计算机的主要特点和性能,具体包括指令集设计、功能组织、逻辑设计、实现技术等。实现技术包括集成电路设计、制造和封装技术、系统制造和组装技术、供电、冷却技术等。另外,设计时人们往往追求在限定的造价范围内,使新型计算机具有最高的性能,这个过程和选购或者组装 PC 时对系统的选配是非常相似的。实际上组装一台满足要求的 PC,就可以看成是一个简单化了的计算机系统设计 and 制造过程。如何采用先进的计算机体系结构和生产技术,制造出具有高性价比(performance cost rate)的计算机系统,是所有通用计算机设计的共同目标。

### 1.2.1 计算机系统中的层次概念

现代计算机系统是由软件和硬/固件组成的十分复杂的系统。为了对这个系统进行描述、分析、设计和使用,人们从不同的角度提出了观察计算机的观点和方法。本节从计算机语言的角度,把计算机系统按功能划分成多级层次结构。

这里的计算机语言主要是指计算机程序设计语言,其基本构成元素是语句,核心是描述计算机系统可以理解并能够按规定执行的操作。随着计算机系统和应用领域的发展,人们已设计出一系列的计算机语言,从面向机器的语言(如机器语言、汇编语言),到各种高级程序设计语言(如 Java、C/C++、FORTRAN、Pascal),还有各种面向问题的语言或者叫应用语言(如面向数据库查询的 SQL 语言、面向数字系统设计的 VHDL 语言、面向人工智能的 PROLOG 语言),人们可以根据应用需求,选择或者设计符合自己要求的计算机语言。总体上看,计算机语言是由低级向高级发展的,高一级的语言的语句相对于低级语言而言,其功能更强,更便于应用,但又都是以低级语言为基础的。

计算机语言可分成一系列的层次(level)或级,最底层语言的功能最简单,最高层语言的功能最强。对于用某一层语言编写程序的程序员来说,他一般是不管其程序在机器中是如何执行的,只要程序正确,他就应该能得到预期的结果。这样,对这层语言的程序员来说,他似乎有了一种新的机器,这层语言就是这种机器的机器语言,该机器能执行相应层语言写的全部程序。因此计算机系统就可以按语言的功能划分成多级层次结构,每一层以一种不同的语言为特征。现代计算机系统的层次结构如图 1.1 所示,图中第 4 级以上完全由软件实现。人们将由软件实现的机器称为虚拟机器(virtual machine),以区别于由硬件或固件实现的实际机器。

第 1 级是微程序机器级,其机器语言是微指令集,工作于该级的程序员实际上是计算机系统的设计人员,他们使用的是为程序描述的、实际上是计算机指令集中每一条指令的功能。用微指令编写的微程序一般是直接由硬件解释实现的。

第 2 级是传统机器级。这一级的机器语言是该机的指令集,程序员用机器指令集编写的程序可以由微程序进行解释。这个解释程序运行在第 1 级上。由微程序解释的指令集又称作仿真(emulation)。实际上,在第 1 级上可以有一个或数个能够在它上面运行的解释程序,每一个解释程序都定义了一种指令集。因此,可以通过仿真在一台机器上实现多种指令集。

计算机系统中也可以没有微程序机器级。在这些计算机系统中是用硬件来直接实现传统机器的指令集的,而不必由任何解释程序进行干预。目前广泛使用的 RISC 技术就是采用这样的设计思想,处理器的指令集用硬件直接实现,目的是加快指令执行的速度。

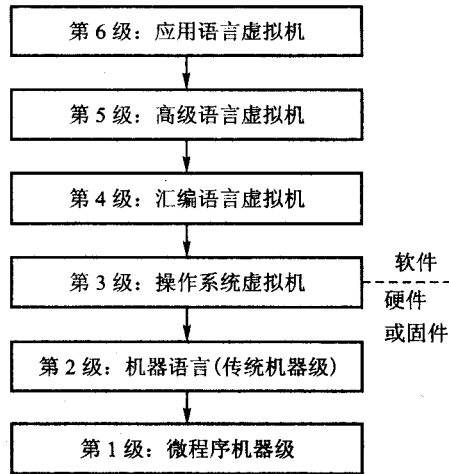


图 1.1 计算机系统的多级层次结构

第 3 级是操作系统虚拟机。从操作系统的基本功能来看,一方面它要直接管理传统机器中的软、硬件资源,另一方面它又是传统机器的引申。它提供了传统机器所没有的某些基本操作和数据结构,如文件结构与文件管理的基本操作、存储体系和多道程序以及多重处理所用的某些操作、设备管理等,这些基本操作和数据结构往往和指令集一起,以整体形式提供给更高层次的虚拟机使用,称之为系统功能调用和系统参数。

第 4 级是汇编语言虚拟机。这一级的机器语言是汇编语言,用汇编语言编写的程序,首先要翻译成第 3 级和第 2 级语言,然后再由相应的机器执行。完成汇编语言翻译的程序就叫做汇编程序。

第 4 级上出现了一个重要变化。通常,第 1、2 和 3 级是用解释(interpretation)方法实现的,而第 4 级或更高级则经常是用翻译(translation)方法实现。

翻译和解释是语言实现的两种基本技术。它们都是以执行一串  $N$  级指令来实现  $N+1$  级指令,但二者仍存在着差别。翻译技术是先把  $N+1$  级程序全部变换成  $N$  级程序后,再去执行新产生的  $N$  级程序,在执行过程中  $N+1$  级程序不再被访问。而解释技术是每当一条  $N+1$  级指令被译码后,就直接去执行一串等效的  $N$  级指令,然后再去取下一条  $N+1$  级的指令,依此重复进行。在这个过程中不产生翻译出来的程序,因此解释过程是边变换、边执行的过程。在实现新的虚拟机器时,这两种技术都被广泛使用。一般来说,解释执行比翻译花的时间多,但占用存储空间较少。

第 5 级是高级语言虚拟机。这一级的机器语言就是各种高级语言,目前高级语言已达数百种。高级语言的翻译过程就是编译(compile),完成翻译的程序就是编译器(compiler)或编译程序。用高级语言,如 C/C++、Pascal、FORTRAN 等,所编写的程序一般是由编译器翻译到第 4 级或第 3 级上。个别的高级语言也有用解释方法实现的,如绝大多数 BASIC 语言系统。

第 6 级是应用语言虚拟机。这一级是为使计算机满足某种特殊用途而专门设计的,因此这一级语言就是各种面向问题的应用语言。可以设计专门用于人工智能、教育、行政管理、计算机

设计等方面的虚拟机,这些虚拟机也是当代计算机应用领域的重要研究课题。应用语言编写的程序一般是由应用程序包翻译到第5级上,也有越级翻译到更低一些级别的机器上的。

### 1.2.2 计算机体系结构

计算机体系结构(computer architecture)这个词目前已被广泛使用。Architecture本来用在建筑方面,译为“建筑学、建筑术、建筑样式、构造、结构”等。这个词被引入计算机领域后,最初的译法也各不相同,以后趋向译为“体系结构”或“系统结构”,但关于它的定义的细节仍有多种不同的说法。

经典的关于“计算机体系结构”的定义是1964年C. M. Amdahl在介绍IBM 360系统时提出的,其具体描述为“计算机体系结构是程序员所看到的计算机的属性,即概念性结构与功能特性”。

按照计算机系统的多级层次结构,不同级程序员所看到的计算机具有不同的属性。例如,机器(汇编)语言程序员所看到计算机的主要属性是该机指令集的功能特性;而高级语言虚拟机程序员所看到的计算机的主要属性是该机所配置的高级语言所具有的功能特性。显然,不同的计算机系统,从机器语言程序员或汇编语言程序员的角度来看,是具有不同的属性的。但是,从高级语言(如Pascal、Java)程序员看,它们就几乎没有什么差别,都具有相同的属性。或者说,这些传统机器级所存在的差别是高级语言程序员所“看不见”的,或者说是他们不知道的。在计算机技术中,对这种本来是存在的事物或属性,但从某种角度看又好像不存在的概念称为透明性(transparency)。通常,在一个计算机系统中,低层机器的属性往往对高层机器的程序员是透明的,如传统机器级的概念性结构和功能特性,对高级语言程序员来说是透明的。由此看出,在层次结构的各个级上都有它的体系结构。Amdahl提出的体系结构是指机器语言级程序员所看见的计算机属性。

以上这些属性是机器语言程序设计者(或者编译程序生成系统)为使其所设计(或生成)的程序能在机器上正确运行所必须遵循的计算机属性,它包含概念性结构和功能特性两个方面。目前,对于通用寄存器型机器,这些属性主要是指:

- ① 数据表示(硬件能直接辨认和处理的数据类型)。
- ② 寻址规则(包括最小寻址单元、寻址方式及其表示)。
- ③ 寄存器定义(包括各种寄存器的定义、数量和使用方式)。
- ④ 指令集(包括机器指令的操作类型和格式、指令间的排序和控制机构等)。
- ⑤ 中断系统(中断的类型和中断响应硬件的功能等)。
- ⑥ 机器工作状态的定义和切换(如管态和目态等)。
- ⑦ 存储系统(主存容量、程序员可用的最大存储容量等)。
- ⑧ 信息保护(包括信息保护方式和硬件对信息保护的支持)。
- ⑨ 输入/输出(I/O)结构(包括I/O连接方式、处理机和存储器与I/O设备间数据传送的方式和格式以及I/O操作的状态等)等。

这些属性是计算机系统中由硬件或固件完成的功能,程序员在了解这些属性后才能编出在传统机器上正确运行的程序。因此,经典计算机体系结构概念的实质是计算机系统中软、硬件界面的



确定,也就是指令集的设计,该界面之上由软件的功能实现,界面之下由硬件和固件的功能来实现。

以上比较全面地讨论了经典的计算机体系结构概念。随着计算机技术的发展,计算机体系结构所包含的内容是不断变化和发展的。这里使用的是广义的计算机体系结构的概念,它除了包括经典的计算机体系结构中关于指令集设计的概念范畴外,还包括了对计算机组织和计算机实现技术的研究。

### 1.2.3 计算机组织和计算机实现技术

计算机组织(computer organization)是计算机系统的逻辑实现,而计算机实现(computer implementation)是计算机系统的物理实现。它们各自包含不同的内容,但又有紧密的关系。

下面以系列机(family machine)为例说明这些概念之间的关系。首先由 IBM 公司在 IBM S360 系统中提出的系列机的概念,被认为是计算机发展史上一个重要里程碑。直到现在,各计算机厂家仍按系列机的思想发展自己的产品。现代计算机不但系统系列化,其构成部件和软件也系列化,如 CPU、硬盘、操作系统、高级语言等。而至今计算机领域中产量最大的系列计算机莫过于 IBM PC 及其兼容系列 PC 机和 Intel 的 x86 系列微处理器。

所谓系列机就是指具有相同的体系结构,但具有不同组织和实现的一系列不同型号的机器,如 IBM S370 系列有 370/115、125、135、145、158、168 等一系列从低速到高速的各种型号。它们各有不同的性能和价格,采用不同的组织和实现技术,但从程序设计者所看到的机器属性却是相同的。在中央处理机中,它们都执行相同的指令集,但在低档机上可以采用指令的分析和指令的执行顺序进行的方式,而在高档机上则采用重叠、流水和其他并行处理方式等。

PC 系列计算机从 1981 年开始生产至今,其历史悠久,生产厂家众多,产量庞大,但发展的技术路线各异,品种复杂,因此对 PC 系列机进行分类的方法有很多,其中最基本的分类是根据所采用处理器的类型来进行的。下面从处理器、处理器字宽、主要 I/O 总线、存储空间、主要操作系统和计算机结构等方面看一看 PC 系列机的一些典型特性,表 1.1 给出了处理器、处理器字宽、I/O 总线、存储空间和主要操作系统的比较,图 1.2 为 PC 系列机典型结构的比较。其中 ISA (Industry Standard Architecture)总线是 AT 总线的标准化以后的名称。

表 1.1 PC 系列机典型特性比较

计算机	时间	处理器	字宽	主要 I/O 总线	存储空间	主要操作系统
PC 和 PC XT	1981	8088	16 b	PC 总线	20 b	DOS
PC AT	1982	80286	16 b	AT(ISA)	24 b	DOS、XENIX
80386 PC	1985	80386	32 b	ISA/EISA	32 b	DOS、Windows 3.0
80486 PC	1989	80486	32 b	ISA + VL	32 b	DOS、Windows 3.1
Pentium PC	1993	Pentium	32 b	ISA + PCI	32 b	DOS、Windows 3.1
Pentium II PC	1997	Pentium II	32 b	ISA + PCI + AGP	32 b	Windows 95
Pentium III PC	1999	Pentium III	32 b	PCI + AGP + USB	32 b	Windows 98、2000
Pentium 4 PC	2000	Pentium 4	32 b	PCI - E + AGP + USB	32 b	Windows Me、XP