

高等学校计算机专业规划教材

计算机组成与操作系统

李东 朱东杰 陈源龙 编著

*Computer Organization
and Operating system*



机械工业出版社
China Machine Press

高等学校计算机专业规划教材



计算机组成与操作系统

李东 朱东杰 陈源龙 编著



*Computer Organization
and Operating system*



机械工业出版社
China Machine Press

图书在版编目(CIP)数据

计算机组成与操作系统 / 李东, 朱东杰, 陈源龙编著. —北京: 机械工业出版社, 2015.10
(高等学校计算机专业规划教材)

ISBN 978-7-111-51686-6

I. 计… II. ①李… ②朱… ③陈… III. ①计算机组成原理—高等学校—教材 ②操作系统—高等学校—教材 IV. TP3

中国版本图书馆CIP数据核字(2015)第234541号

本书突出“内容全面、习题和例题丰富”的特色, 适应国际计算机教育发展的新要求, 强调培养计算机相关专业大学生对计算机系统的全面认知与设计开发能力。

本书全面系统地介绍了计算机组成与操作系统的基本原理和基本概念, 主要包括计算机的发展历程、计算机的基本组成与工作原理、计算机的分类与性能评价、多核处理器、信息在计算机中的表示与运算方法、计算机指令与指令集、CISC与RISC、处理器的基本组成、控制器的组成与工作原理、存储器与存储系统(含Cache)、总线技术、8086/8088汇编语言程序设计、I/O接口、中断系统、外部设备(如键盘、鼠标、显示器、硬盘、光盘)、操作系统的基本概念与功能、文件系统的基本概念与实现、进程管理与处理器调度、内存管理与虚拟存储器。此外, 附录中还介绍了图灵机模型、数制及其转换、相联存储器。

本书行文流畅, 深入浅出。例题涵盖了2009~2014年的计算机专业硕士研究生入学全国统一考试试题的详尽解答, 实用性强。

本书可作为计算机科学与技术、软件工程、物联网工程、信息安全、信息管理与信息系统、信息与计算科学、生物信息技术等专业的“计算机组成原理”“操作系统”“计算机系统”“微机原理”“计算机组成与操作系统”课程的教材, 也可供准备参加计算机专业硕士研究生入学全国统一考试的考生、计算机行业的工程技术人员以及其他自学者学习参考。

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: 张梦玲

责任校对: 董纪丽

印刷: 北京诚信伟业印刷有限公司

版次: 2015年10月第1版第1次印刷

开本: 185mm×260mm 1/16

印张: 25.75

书号: ISBN 978-7-111-51686-6

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: hzsj@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

前 言

目前计算机的发展已经进入“后 PC”时代，移动互联网、物联网、智能手机的快速发展，不仅改变了计算机的形态，而且提高了对计算机相关专业学生的知识结构和能力水平的要求。培养计算机相关专业学生对计算机系统的全面认知与设计开发能力已经成为国际计算机教育领域的普遍共识。

通过本书的学习，读者不仅可以了解计算机硬件以及核心软件操作系统的工作原理和协同技术，而且可以掌握汇编语言程序设计的基本方法，对计算机系统有全面的认知，初步具备计算机系统的设计开发能力和解决问题的能力，建立起计算思维和创新意识。

本书内容新颖、全面，实用性强。它的特点是：

①内容全面，深度适宜，与技术发展保持同步。

本书涵盖了计算机组成原理、操作系统和汇编语言程序设计三门课程的全部内容，介绍了 Unicode 字符集、多核处理器和铁电存储器等前沿技术。事实上，本书还覆盖了截至 2014 年 10 月的计算机技术的最新成果。

②通过全面认识计算机系统，培养学生的计算思维和创新意识，提高他们解决问题的能力。

本书对计算机的基本硬件与核心软件进行了深入的论述，使学生全面认识计算机系统，领会到计算思维的核心就是抽象和自动化以及计算思维就在我们身边，帮助学生打破计算机的神秘感，培养学生的计算思维和创新意识，提高他们解决问题的能力。

③例题和习题丰富、形式多样、实用性强。

本书的例题涵盖了 2009 ~ 2014 年的计算机专业硕士研究生入学全国统一考试的试题。读者不仅可以从本书得到对这些试题的详尽解答，而且可以根据不同章节试题类例题的多少了解考试的重点和难点。每一章都配有丰富的习题，题型有简答题、填空题、选择题、设计题，甚至提供了学习型习题。读者可以通过阅读习题掌握新的知识，并应用这些知识独立解决问题。

本书既可作为计算机科学与技术、软件工程、物联网工程、信息安全、信息管理与信息系统、信息与计算科学、生物信息技术等专业的“计算机组成原理”“操作系统”“计算机系统”“微机原理”“计算机组成与操作系统”课程的教材，也可供准备参加计算机专业硕士研究生入学全国统一考试的考生、计算机行业的工程技术人员以及其他自学者学习参考。

本书由哈尔滨工业大学的李东教授、陈源龙博士以及哈尔滨工业大学（威海分校）的朱东杰博士共同编写。其中，第 1 ~ 6 章由李东编写，第 7 ~ 9 章由陈源龙编写，第 10 ~ 14 章由朱东杰编写，且李东教授负责全书统稿。哈尔滨工业大学计算机学院的史先俊、舒燕君、徐冰、黄庆成、刘松波、王伟、吴锐、张宇、孙春奇、程丹松、张英涛，以及哈尔滨工业大学（威海分校）计算机学院的柏军、张策、何辉、石代锋参与了书中例题和习题的编撰

工作。吉帅、聂建宏、王文邦、郑钰、范晟铭、潘辉等协助完成书稿的校对工作。

机械工业出版社华章分社的朱劼编辑也为本书的出版提供了大量的帮助，并提出指导性建议，谨此表示衷心的感谢。

由于编者水平有限，书中难免会有疏漏和错误，恳请读者批评指正。读者若有反馈意见或同行教师若有教学心得，欢迎联系我们：Lee@hit.edu.cn。

本书相关习题，可从 www.hzbook.com/Books/8759.html 上获得，也可通过扫描下列二维码获取。



编 者

2015 年 7 月

目 录

前言

第 1 章 计算机系统引论 1

1.1 计算机系统概论..... 1

1.1.1 计算机系统硬件和软件 1

1.1.2 冯·诺依曼计算机模型 3

1.1.3 计算机硬件的基本组成 5

1.1.4 计算机的体系结构、组成与实现..... 6

1.1.5 操作系统与计算机系统的层次结构 9

1.2 计算机的发展历程..... 12

1.3 计算机系统的分类..... 17

1.3.1 概述 17

1.3.2 弗林分类法 18

1.4 计算机的性能评价指标..... 19

1.5 微处理器与微型计算机..... 25

1.5.1 微处理器与微型计算机的产生与发展 25

1.5.2 多核处理器 28

1.6 中国计算机事业的发展历程..... 30

1.7 计算机的应用及发展趋势..... 32

1.7.1 计算机的应用 32

1.7.2 计算机发展的动力 36

1.7.3 计算机的发展趋势 37

第 2 章 计算机中信息的表示与运算 40

2.1 数据的表示..... 40

2.1.1 定点数的表示 40

2.1.2 浮点数的表示 44

2.2 定点数的运算..... 50

2.2.1 逻辑运算 50

2.2.2 移位运算 51

2.2.3 加法与减法运算 53

2.2.4 乘法运算 57

2.2.5 除法运算 63

2.2.6 算术逻辑单元 67

2.3 浮点数的运算..... 71

2.3.1 浮点数加、减运算 71

2.3.2 浮点数乘、除运算 74

2.4 面向错误检测与纠错的数据编码..... 74

2.5 字符与字符串..... 79

2.6 面向存储与传输的数据编码..... 83

第 3 章 处理器 87

3.1 处理器的指令集..... 87

3.1.1 概述 87

3.1.2 指令的操作码与操作数 88

3.1.3 寻址方式 89

3.1.4 指令的基本功能与指令集设计 93

3.1.5 指令的格式 98

3.1.6 面向多媒体处理的增强指令 101

3.2 处理器的组成与工作过程..... 103

3.2.1 处理器的基本功能和基本组成 103

3.2.2 计算机的工作过程 107

3.2.3 采用流水线技术的处理器 108

3.3 从 CISC 到 RISC..... 114

3.3.1 RISC 产生的背景 114

3.3.2 RISC 的定义 116

3.3.3 指令级并行技术 117

3.4 Intel 80x86 系列微处理器 125

3.4.1 Intel 8086/8088 微处理器 125

3.4.2 Intel 80286、80386 和 80486 微处理器 133

3.4.3 Intel Pentium 系列微处理器 136

3.5 ARM 系列微处理器	139	6.4.3 Pentium 微处理器的主存储器组成	209
3.5.1 概述	139	6.4.4 存储芯片的发展	210
3.5.2 ARM 微处理器的模式、工作状态和寄存器组织	140	6.5 高速缓冲存储器	213
3.5.3 ARM 微处理器的存储器组织和寻址方式	143	6.5.1 Cache 的工作原理	213
3.5.4 ARM 微处理器的指令集	146	6.5.2 地址映像与变换	216
第 4 章 控制器	153	6.5.3 替换算法	221
4.1 概述	153	6.5.4 写入策略	223
4.2 硬布线控制器	158	6.5.5 两级 Cache 与分裂型 Cache	224
4.3 微程序控制器	161	6.5.6 Cache 的性能评价	224
第 5 章 总线技术	167	第 7 章 8086/8088 汇编语言程序设计	228
5.1 概述	167	7.1 引言	228
5.2 总线的设计与实现	170	7.1.1 汇编语言的基本概念与学习汇编语言的重要性	228
5.3 总线控制	173	7.1.2 8086/8088 汇编语言的基本语法	229
5.3.1 总线仲裁	173	7.2 顺序程序设计	233
5.3.2 总线通信控制	177	7.2.1 概述	233
5.4 总线的性能指标	180	7.2.2 寻址方式的表示	234
5.5 总线标准	183	7.2.3 表达式的使用	236
5.5.1 微型计算机的系统总线标准	183	7.2.4 属性取代符	238
5.5.2 微型计算机的局部总线标准	186	7.2.5 属性分离符 / 数值回送操作符	239
5.5.3 I/O 总线标准举例	187	7.2.6 运算符的优先级	241
第 6 章 存储系统	190	7.2.7 乘除法运算的实现	241
6.1 存储器的分类与性能评价	190	7.2.8 算术移位和逻辑移位	243
6.1.1 存储器的分类	190	7.3 分支结构程序设计	244
6.1.2 存储器的性能评价	190	7.4 循环结构程序设计	247
6.2 存储器访问的局部性原理与层次结构的存储系统	191	7.5 字符串操作程序设计	254
6.2.1 存储器访问的局部性原理	191	7.6 宏、条件汇编与重复汇编	257
6.2.2 层次结构的存储系统	191	7.7 子程序设计	263
6.3 半导体存储器	193	7.8 8086/8088 微处理器的其他指令与应用	270
6.3.1 随机访问半导体存储器	193	第 8 章 计算机的外部设备	274
6.3.2 只读存储器	200	8.1 输入设备	274
6.4 主存储器	203	8.1.1 键盘	274
6.4.1 主存储器的组成	203	8.1.2 鼠标	276
6.4.2 提高主存储器访问带宽的方法	206	8.2 输出设备	277

8.2.1 阴极射线管显示器	277	11.2 I/O 核心子系统	324
8.2.2 平板显示器	282	11.2.1 I/O 调度	324
8.2.3 打印机	283	11.2.2 高速缓存与缓冲区	325
8.3 辅存设备	285	11.2.3 设备分配与回收	325
8.3.1 硬盘	285	11.2.4 假脱机技术	326
8.3.2 光盘	288	11.2.5 I/O 出错处理	327
8.3.3 U 盘和固态硬盘	289	第 12 章 文件管理	328
第 9 章 I/O 接口	291	12.1 文件系统基础	328
9.1 I/O 技术的发展	291	12.1.1 文件概念	328
9.2 I/O 接口的组成与工作原理	293	12.1.2 文件的逻辑结构	329
9.3 中断系统	297	12.1.3 目录结构	330
9.3.1 中断的处理过程	297	12.1.4 文件共享	332
9.3.2 中断屏蔽	300	12.1.5 文件的访问控制	333
9.3.3 中断控制器 8259A	302	12.2 文件系统的实现	335
9.3.4 8086/8088 微处理器的中断 系统	304	12.2.1 文件系统的层次结构	335
9.4 DMA 技术	306	12.2.2 目录的实现	336
9.5 通道技术	309	12.2.3 文件的实现	336
第 10 章 操作系统概述	311	12.3 磁盘组织与管理	340
10.1 操作系统的概念、特征、功能和 提供的服务	311	12.3.1 磁盘的组织	340
10.1.1 操作系统的概念	311	12.3.2 磁盘调度算法	341
10.1.2 操作系统的特征	312	12.3.3 磁盘中物理块的优化分布	342
10.1.3 操作系统的功能	313	12.3.4 磁盘的错误处理	343
10.1.4 操作系统提供的服务	313	第 13 章 进程管理与处理器调度	345
10.2 操作系统的发展与分类	314	13.1 进程的状态与转换	345
10.2.1 操作系统的发展	314	13.2 进程的组织与控制	347
10.2.2 操作系统的分类	317	13.2.1 进程组织	347
10.3 操作系统的运行环境	318	13.2.2 进程控制	348
10.3.1 内核态与用户态	318	13.3 进程的同步与互斥	351
10.3.2 中断与异常	318	13.4 进程间通信	362
10.3.3 系统调用	318	13.5 死锁及其处理	363
10.3.4 操作系统的启动过程	319	13.6 处理器调度	366
10.4 现代操作系统的新发展	320	13.7 线程的概念与多线程模型	370
第 11 章 I/O 管理	321	第 14 章 内存管理与虚拟存储器	373
11.1 概述	321	14.1 内存管理基础	373
		14.2 连续分配的内存管理	375
		14.3 非连续分配的内存管理	378

14.3.1 段式存储管理	378	附录	394
14.3.2 页式存储管理	379	附录 A 图灵机模型	394
14.3.3 段页式存储管理	381	附录 B 图灵奖历年获得者	394
14.4 虚拟存储器	382	附录 C 数制及其转换	397
14.5 页面替换与页面分配	386	附录 D EBCDIC 码	400
14.5.1 页面替换	386	附录 E 8086/8088 指令格式	401
14.5.2 页面分配	387	附录 F 相联存储器	402
14.6 Cache 与虚拟存储器的协同	390	参考文献	404

第1章 计算机系统引论

计算思维 (Computational Thinking) 是指运用计算机科学的基础概念进行问题求解、系统设计及人类行为理解等涵盖计算机科学之广度的一系列思维活动。计算思维的核心是抽象 (Abstraction) 和自动化 (Automation)。其中“抽象”是指用符号 (例如数字) 来表示客观事物, “自动化”是指让计算机解决问题而无须人的干预。

周以真 (Jeannette M. Wing) 教授, 美国卡内基·梅隆大学

1.1 计算机系统概论

1.1.1 计算机系统硬件和软件

1. 硬件

计算机系统由硬件 (Hardware) 和软件 (Software) 构成。其中, 硬件指的是计算机系统的物理实体, 例如键盘、显示器和主机 (箱) 及连接它们的线缆。

键盘属于输入设备 (Input Device)。典型的输入设备还有鼠标、摄像头、扫描仪等。

显示器属于输出设备 (Output Device)。常见的输出设备还有打印机、绘图仪、音箱等。

近年来, 随着触摸屏技术的发展, 显示器既可以用于输出, 也可以用于输入。

主机箱中放置有主板 (Main Board) 和硬盘驱动器、软盘驱动器、光盘驱动器及联系它们的线缆。

主板是一个印制线路板 (Printed Circuit Board, PCB), 用于集成并连接处理器、存储器、输入输出控制器等芯片, 图 1-1 就是一个典型的微型计算机的主板。由于计算机中所有的操作和命令都有一定的时间顺序, 所以主板上需要设置定时部件, 如时钟 CLK (Clock) 和时序信号发生器 (Timing Signal Generator, TSG)。

主板上的存储器是可以被处理器直接访问的, 称为“主存储器 (Main Memory), 简称主存”或“内存器 (Internal Memory), 简称内存”。

相应地, 硬盘 (Hard Disk, HD)、软盘 (Floppy Disk, FD)、光盘 (Optical Disk Memory, ODM)、磁带等称为“辅助存储器 (Secondary Memory), 简称辅存”或“外存储器 (External Memory), 简称外存”。

辅存的容量比主存的容量要大得多, 每位的平均价格也要低得多, 但是它的访问速度却明显慢于主存。因此, 为了保证速度很快的处理器有较高的工作效率, 计算机的设计者规定处理器只与速度较快的主存交换信息, 而不直接访问辅存。辅存中的信息要装入到主存后才能供处理器使用。

由于典型的主存都是“易失性存储器 (Volatile Memory)”, 即掉电或断电后, 存储在其中的信息将消失, 所以辅存是计算机系统中不可或缺的存储体。需要长久保存的程序和数据必须存储在“非易失的存储器 (Non-volatile Memory)”——辅存里。

辅存既为处理器提供程序和数据, 又接受 / 存储处理器发回的计算结果, 所以它属于输入 / 输出设备 (Input/Output Device)。

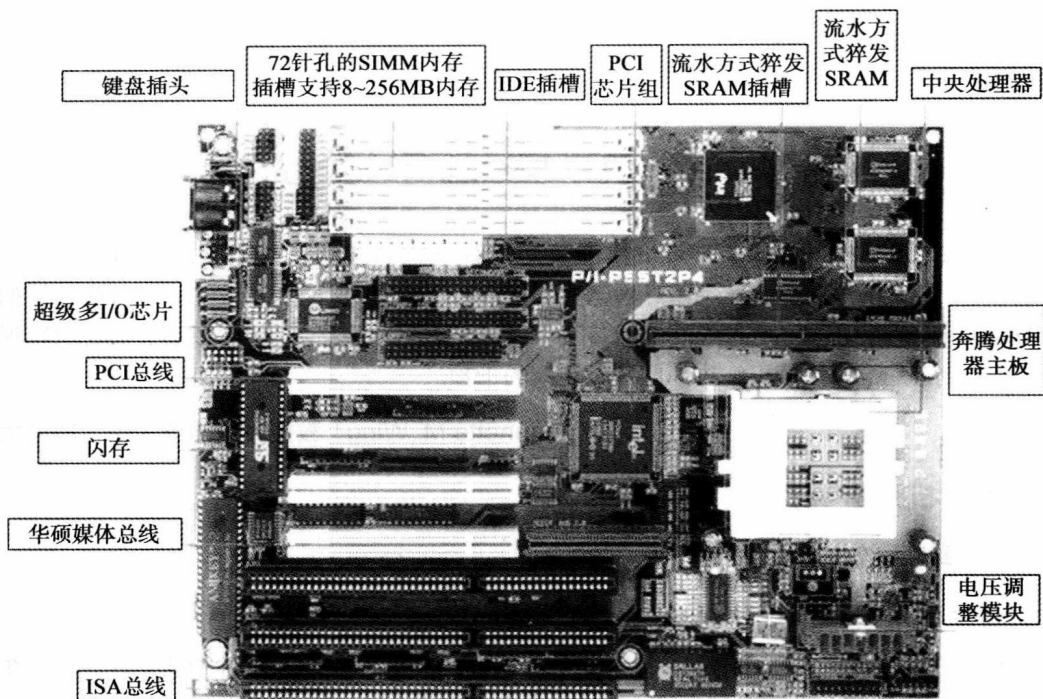


图 1-1 一个典型的微型计算机的主板

在有些场合，除主板及其上芯片之外的硬件器件称为外部设备或外围设备（Peripheral Equipment），简称外设。外设包括：输入设备、输出设备、辅存。

2. 软件与软件 / 硬件的等价性原理

客观地说，计算机的“神通广大”是由其中的软件所赋予的。离开软件，计算机就毫无用处。软件指的是人们事先编制好的、能解决具体问题的程序（Program）及相应数据（Data）。

由于软件存储在主存和辅存中，对人而言具有“看不见、摸不着”的特点，但它绝对是计算机的“灵魂”，计算机的功能完全是通过运行程序来实现的。

通常，软件分为系统软件和应用软件，也可以更详细地分为：系统软件、应用软件、支持软件、测试与维护软件。

系统软件是指构成一个计算机系统所必需的基本软件，与具体用户无关。常见的系统软件有操作系统（Operating System, OS）和数据库管理系统（Database Management System, DBMS）。

应用软件是由用户根据各自的应用需要而安装的、能解决专用领域特殊问题的软件，如 AutoCAD、3D MAX、Protel、Matlab 等。

支持软件是指用于帮助和支持软件开发的软件，如汇编程序、高级语言编译器、文本编辑器、设计工具软件、项目管理软件、配置管理软件等。

测试与维护软件是指用于软件故障诊断、错误隔离、系统调试及检测系统可靠性的软件。

除去软件和硬件外，还有一个概念就是“固件（Firmware）”。对于那些不再需要改动而且经常被调用的软件，为了使其有更快的执行速度，可以将其存储在访问速度较快的、具有非易失性的只读存储器（Read Only Memory, ROM）芯片中。这相当于将软件“固化”在硬件

(ROM 芯片)中。如果需要修改或升级软件,只需要更换一块 ROM 芯片即可。

这种吸收软件、硬件各自优点,性能介于软件和硬件之间(执行速度快于软件,灵活性优于硬件),以硬件形式出现的软件,被称为“固件”。

在器件成本不断下降的今天,固件越来越多地被用来提高计算机的性能。

事实上,计算机是面向算法的机器,而程序/软件是某个算法的实现。一种算法可以由硬件或固件来实现,也可以由软件来实现。如向量或数组运算,在低档的计算机上由软件来实现,而在高档的计算机(如大型计算机或超级计算机)上则由硬件来实现。所以说,硬件和软件在逻辑功能上是完全等价的,即软件的功能在原理上可以由硬件来实现,硬件的功能在原理上也可以由软件来实现,只存在性能与实现成本上的差别,这称为“软件/硬件的等价性原理(The Principle of Equivalence of Hardware and Software)”。

软件/硬件的等价性原理以及软件、硬件、固件三个概念的提出,无疑验证了老子的伟大论断:道生一,一生二,二生三,三生万物。

随着软件开发技术的发展,人们提出了“软件工程”的概念。所谓“软件工程”,是指将系统化的、规范化的、可量化的方法应用于软件的开发、运行和维护中,即将工程化应用于软件的整个生命周期中。

在软件工程的意义上,软件是“计算机程序、过程、规则和与这些程序、过程、规则有关的文档,以及从属于计算机系统运行的数据。”

1.1.2 冯·诺依曼计算机模型

在第一台电子计算机(Electronic Numerical Integrator And Computer, ENIAC)设计与研制的过程中,他们的设计者曾向美国普林斯顿大学高等研究院(Institute of Advanced Study, IAS)的科学家冯·诺依曼(von Neumann)进行过咨询。ENIAC 投入使用后,冯·诺依曼曾到现场参观,对 ENIAC 表现出强烈的兴趣。在对 ENIAC 存在的不足(ENIAC 是专用计算机,它的功能是由电路连线来决定的,改变功能时,需要人为地改变电路连线)进行深入思考的基础上,冯·诺依曼等人于 1946 年 6 月发表了一篇旨在构建一台通用计算机的技术报告《关于电子计算装置逻辑结构初探》。这份报告提出了基于“存储程序”控制的“离散变量自动电子计算机(Electronic Discrete Variable Automatic Computer, EDVAC)”的设计方案。

“存储程序”的思想是:计算机的用途和硬件完全分离。硬件采用固定性逻辑,提供某些固定不变的功能。通过编制不同的程序来满足不同用户对计算机的应用需求。

依照这个思想,在计算机上求解一个问题,需要将求解该问题的过程分解成一系列简单、有序的计算步骤,一个步骤由计算机提供的一条计算机指令(Instruction)完成。然后将这些有序的计算步骤——对应成计算机能够识别并可执行的指令——汇总在一起,就形成所谓的程序,并存储在计算机中。计算机通过逐条、顺序执行程序中的指令来完成问题的求解。“存储程序”思想体现了计算思维的核心概念——自动化。

根据《关于电子计算装置逻辑结构初探》设计的“冯·诺依曼计算机”的特点如下。

1) 具备五大功能:数据存储、操作判断与控制、数据处理、数据输入和数据输出。对应五个功能部件:存储器(Memory)、控制单元(Control Unit, CU)、算术逻辑运算单元(Arithmetic Logic Unit, ALU)、输入单元(Input Unit)和输出单元(Output Unit),如图 1-2 所示。图 1-2 中,有两股信息在流动。一股是控制流(即操作命令),它从 CU 发出,分散流向各部件;另一股是数据流(包括指令和地址),它在 CU 的控制下,从一个部件流到另一个部件。

2) 存储器由一组一维排列、线性编址的存储单元组成, 每个存储单元的位数是相等且固定的, 存储单元按地址访问。这是最简单、最易于实现的信息存储与查找方案, 就如同一排平房, 房间大小都是一样的, 按照房间号查找房间, 房间号逐间递增。

3) “程序”由一条一条的指令有序排列而成, 而指令由操作码和地址码两部分组成。操作码规定了该指令的操作类型(即功能), 地址码指示存储操作数和运算结果的存储单元地址。操作数的数据类型由操作码来规定, 操作数可能是定点数、浮点数、双精度浮点数、十进制数、逻辑数、字符或字符串等。

4) 指令和数据均采用二进制表示, 并以二进制形式进行运算。二进制的计算规则是最简单的, 加法仅有 4 种: $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, $1 + 1 = 10$ 。

把纷繁复杂的信息抽象成 0/1, 这是计算机的根, 是计算机的哲学, 是计算思维核心概念“抽象”的具体体现——信息符号化, 符号数字化, 数字 0/1 化。

5) 为了简化计算机的控制与组成, 程序(指令)与数据同等地、不加区分地存储在同一个存储器中, 但可以从时间和空间上, 将它们区分开。在取指周期中, 从存储器流向控制器的是指令; 在执行周期中, 存储器与运算器交换的是数据。

6) 为了“逐条、顺序执行程序中的指令”, 冯·诺依曼提出了一个极易实现的解决方案: 设置一个“程序计数器(Program Counter, PC)”来指示下一条将要执行的指令的地址。在一般情况下, 每执行完一条指令, PC 就会自动加 1, 以指向下一条指令的存储单元。

当然, 为了赋予计算机更多的“灵性”, PC 的值也可以通过执行特殊的指令来修改, 从而达到改变指令执行顺序的目的。这样, 执行指令的顺序就不受限于指令存储的顺序了。

不过, 尽管冯·诺依曼非常聪明, 冯·诺依曼计算机的结构还是存在一些问题, 后来人们对它进行了改进和发展——大学, 止于至善!

例如, 由于以 ALU 为中心, 输入/输出单元与存储器之间的数据传送都要经过 ALU, 这使得 ALU 无法专注于运算, 低速的输入/输出和高速的运算不得不相互等待, 串行工作。因此, “冯·诺依曼计算机”很快就被改进成以存储器为中心(如图 1-3 所示)。这样, 输入/输出设备就可以与运算器并行工作, 输入设备也可与输出设备并行工作, 提高了设备的效率和利用率。同时这使得计算机五个功能单元的互联更加简单。

时至今日, 冯·诺依曼计算机结构经历近七十年, 依然占据统治地位。这是为什么呢?

说到底, 冯·诺依曼的最大贡献就是将计算机系统分为五个部件并确定了这五个部件的互联结构。在考虑到 ALU 既需要从存储器中读取数据, 又需要将运算结果写回到存储器, 请读者自行设计一个计算机五个部件的互联结构, 看看能否比图 1-3 更简单。

抽象起来, “冯·诺依曼计算机结构”就是“以一个部件为中心, 实现五个部件互联的星

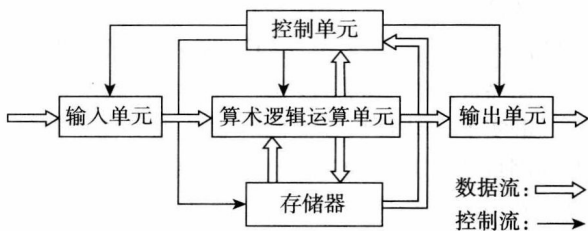


图 1-2 冯·诺依曼计算机结构

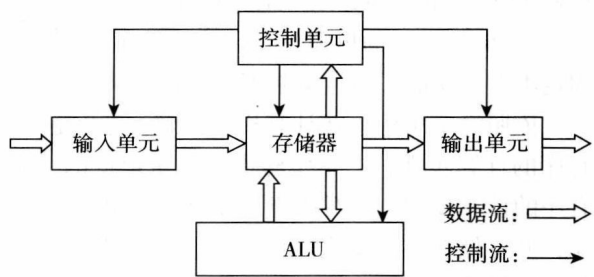


图 1-3 现代冯·诺依曼计算机结构

形结构”。这样实现的“五个部件互联”是最简单的、连接链路也是最少的。由于外围的四个部件只能与中心部件进行数据通信，而且有些数据通路还是单向的。进而冯·诺依曼计算机所需要的数据传送功能也是最少的、最简单的。

综上所述，“冯·诺依曼计算机”可以概括为：用最简单、最易于实现的思想（二进制及其运算规则）来表示数据并实现运算，用最简单的互联结构来组成一台计算机。

这不仅体现了计算思维，还很好地体现了“简单就是美”的工程哲学理念！

1.1.3 计算机硬件的基本组成

1. 概述

从组成的角度看，计算机由 CU、ALU、存储器、输入单元和输出单元组成。在具体实现时，通常将 CU 和 ALU 集成在一起，构成“处理单元（Processing Unit, PU）”。处理单元也被称为“处理器（Processor）”。

一台计算机通常只拥有一个 PU，而这个 PU 又是计算机的核心部件，所以这样的处理单元又称为“中央处理单元或中央处理器（Central Processing Unit, CPU）”。事实上，一台计算机可以拥有多个 PU，这样的计算机叫作“并行计算机（Parallel Computer）”或“高性能计算机（High Performance Computer）”。追求更高的计算性能，是科学研究的永恒主题。

2. PU 的组成

PU 是计算机系统的核心，由 ALU、CU 以及一些暂存单元——寄存器（Register）组成。

ALU 能完成“加”“减”“乘”“除”等算术运算和“与”“或”“非”“异或”等逻辑运算。但它并不是只靠一套电路来“智慧地”完成各种运算，而是用不同的电路来完成不同的运算，即由加法器来完成算术运算，逻辑运算器来完成逻辑运算，移位器（Shifter）来完成移位运算，求补器（Complementer）来完成取反 / 求补码运算。

CU 是 PU 的指挥机构，由程序计数器 PC、存放当前指令的指令寄存器（Instruction Register, IR）、解释指令的指令译码器（Instruction Decoder, ID）、发出各种命令信号的控制信号发生器（Control Signal Generator, CSG）及相应的控制逻辑组成。CU 依据指令译码器产生的一系列操作命令 / 信号来指挥、协调 PU 乃至计算机系统中各个部件的工作。

PU 内部的寄存器有：若干个存放数据的数据寄存器、若干个存放操作数地址的地址寄存器和一个存放各种“标志（Flag）”的标志寄存器（Flag Register, FR）。同样是一串“0101”，如果存储在数据寄存器中，计算机则将其理解为一个数；如果存储在地址寄存器中，计算机则将其理解为一个操作数的地址。

最常用的数据寄存器是用于存放加法的一个操作数及运算结果的累加寄存器 ACC（Accumulator）。

有些文献将这些寄存器称为“寄存器文件（Register File）”或“寄存器堆”。有的计算机中，寄存器既可以存数据，也可以存地址。这样的寄存器称为通用寄存器（General Purpose Register, GPR）

引入“标志”来表示 / 区分计算过程的各种状态是计算思维的一个具体体现。例如，当加法运算的结果为零时，“零标志（Zero Flag, ZF）”被置为 1，否则为 0。运算结果的符号被复制到符号标志（Sign Flag, SF）中。SF 为 1，表示运算结果为负数，否则为正数。当运算结果中“1”的个数为零或偶数时，“奇偶标志（Parity Flag, PF）”被置为 1，否则为 0。类似的还有

“进位标志 (Carry Flag, CF)” 和 “溢出标志 (Overflow Flag, OF)” 等。这些标志分别占据标志寄存器中的不同位置, 后继的指令就可以根据 ZF、SF、PF 的值来选择不同的操作。

访问主存是 PU 经常执行的操作。为了实现这个操作, PU 内部设置了专门存放访存地址的寄存器 MAR (Memory Access Register)、专门存放与主存交换数据的寄存器 MDR (Memory Data Register)。有的文献称 MDR 为 MBR (Memory Buffer Register)。

在用户看来, 只要把主存地址送入 MAR, 启动读命令, 在一个访存周期内, 目标数据就会从主存被读入到 MDR 中; 或者只要把主存地址送入 MAR, 并把目标数据送入 MDR, 启动写命令, 在一个访存周期内, 目标数据就会从 MDR 被写到主存中。

图 1-4 显示了计算机系统硬件的基本组成。

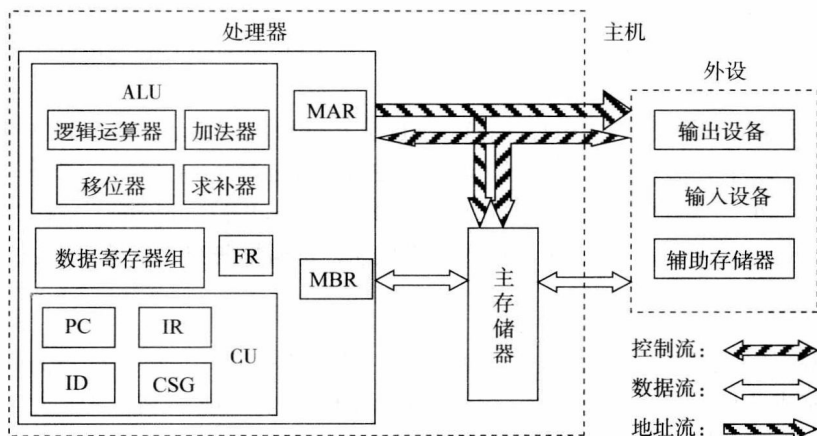


图 1-4 计算机系统硬件的基本组成

1.1.4 计算机的体系结构、组成与实现

1. 计算机的体系结构与系列机

“计算机体系结构 (Computer Architecture)” 是计算机学科中最重要的概念之一。但这一概念并不是同计算机或计算机科学与技术学科一起诞生的, 它是在第一台电子计算机诞生近 20 年后, 由安达尔 (G M. Amdahl) 等人于 1964 年在设计 IBM System 360 时提出的。

在 20 世纪 50 年代, 计算机系统的设计大多是逐个进行的。由于设计者的不同, 计算机一般具有各自与众不同的指令系统, 程序员按照其特定的汇编符号编写汇编程序并在特定的机器上运行。当时用户拥有的计算机数目极少, 计算机的昂贵也使得用户很少更换计算机。用户只关心自己计算机的功能与性能, 而不提其他需求。

从 20 世纪 50 年代末期开始, 随着计算机的日益普及, 需要编写的程序越来越多, 程序也越来越大, 程序的开发成本越来越高、开发周期越来越长。这时, 人们就希望能够在自己部门的机器上运行别的部门开发的程序, 针对旧机器开发的程序能够不加修改地在新买的机器上运行。因此, 为了保证优秀的软件能够长期使用, 保护用户的投资, 降低软件开发人员的重复工作量, 使用户或软件开发人员能够把更多的资金或精力投入到新的软件上去, 程序可移植性 (Portability) 的概念被提了出来。所谓可移植性是指针对在一台计算机上能够运行的程序, 不加修改或只需少许加工就可以在另外一台计算机上正确运行, 并给出相同的结果。不具备可移植性的程序, 就不能在其他计算机上运行, 被其他用户共享。

这时,计算机的发展进入到第二代阶段,计算机的应用领域也从传统的科学计算领域,逐渐扩大到数据处理、过程控制和事务管理等领域,计算机生产厂商不断增多,计算机的型号五花八门,既有面向科学计算的超级计算机,又有面向大规模数据处理和事务管理的大型计算机,还有低端用户的中、小型机。

因此,产品多样性和软件难于移植的矛盾就摆在了计算机设计师的面前。此时,IBM 公司投资 50 亿美元计划开发一个大型计算机系统,总设计师安达尔雄心勃勃,决心一举解决这个矛盾。

首先,安达尔计划让新的计算机系统的功能齐全、无所不能。用户只要购买了一台新机器,就可以满足他对计算机的所有需求。为此,这个新系统被命名为 System 360 (S/360),寓意它能够满足全方位(360 度)的应用。

其次,安达尔希望 S/360 长盛不衰,十年、二十年,甚至上百年之后仍然受到用户的喜爱。用户一旦购买了 S/360,就没必要购买别的厂商的机器。由于用户始终使用 S/360,也就不为程序难于移植到其他机器上而发愁了。

只要肯投入,付得起高价钱,实现第一个计划并不是一件难事。但由于新技术的不断提出、新材料的不断涌现、计算机厂商之间竞争日益激烈,计算机系统更新换代的速度越来越快,计算机系统的生命周期也越来越短。在这种情况下,如何实现安达尔的第二个计划呢?

通过对用户和市场进行深入的分析,安达尔定下了解决问题的哲学:一切为用户着想,而计算机系统最直接的用户就是程序员。

对于程序员而言,他们最大的期望就是一劳永逸,即一旦掌握某一种机器的属性及其编程方法,就永远只对这一种机器编程,不愿意再学习其他机器的属性及其编程方法。但同时他们又想拥有性能不断提升的新机器来充分施展软件的功能。

这样,不变的机器属性及其编程方法与对不断提升的机器性能的期望构成了一对矛盾。

为此,安达尔提出了一个新的概念——计算机体系结构来解决这一矛盾,并将其定义为程序员所看到的机器的属性,即机器的概念性结构和功能特性。在当时,程序员主要指汇编程序员。

安达尔认为:计算机的设计/制造者可以利用新技术、新材料来设计制造性能更高的新机器,但只要新机器保持原机器的体系结构,程序员就无须学习新的机器属性及机器语言,原有的软件自然不用做任何修改就能在新机器上运行。

安达尔还考虑到:用户对性能的需求是多方面的,但是对软件可移植性的要求是一致的,所以可以通过生产具有相同体系结构,但不同档次的机器来同时满足用户这两方面的要求。

可见,安达尔真是全心全意地为用户着想。

安达尔将同一厂商生产的具有相同体系结构的机器定义为系列计算机(简称系列机)。系列机可以是不同年代生产的机器,也可以是相同年代生产但档次不同的机器。安达尔设计的 IBM S/360 就是计算机历史上的第一个系列机,如图 1-5 所示。

若将软件兼容(Software Compatibility)定义为同一个程序可以不加修改地在具有相同体系结构的各个档次的机器上正确运行,唯一区别仅仅在于运行时间长短不同,则安达尔就是以软件兼容(统一机器语言)的方式来解决程序的可移植性问题。

通过提出“计算机体系结构”概念,安达尔解决了 IBM S/360 的软件可移植性问题。

计算机体系结构和系列机这两个概念的提出,一举打破了 20 世纪 50 年代计算机“手工作坊式”的设计生产方式,使得已有的软件资源得到了充分的利用,有效地减少了开发软件的投资与工作量,很好地解决了软件开发环境要求相对稳定,而新技术、新材料又不断涌现之间的

矛盾,使得程序员能够把更多的时间和精力投入到完善软件功能、提高软件质量的工作中,并为积累和重用软件、提高软件生产率奠定了基础。同时,计算机厂商又可以不断地采用新技术、新材料,为市场提供性能更高、价格更低的新机器,来更好地满足用户不断增长的应用需求。计算机产业从此走向了供需两旺、蓬勃发展的新时期。

其后,IBM公司在20世纪70年代后期和90年代又先后推出了与System/360兼容的System/370系列和System/390系列。

系列机的概念一经提出,各个著名计算机厂商纷纷效仿,先后分别提出了各自不同的系列机产品。

DEC公司在PDP-11系列之后,又推出了VAX-11(Virtual Address eXtension of PDP11)系列计算机。CDC公司的6600、7600、CYBER系列超级计算机也曾一度非常显赫。CRAY计算机公司一直研制并生产着CRAY系列超级计算机。英特尔公司从1980年开始推出的Intel 80x86系列微处理器更是创下了产量最大的系列机的纪录。

因此,计算机体系结构和系列机概念的提出是计算机发展史上的一个重要里程碑。“计算机体系结构”这一概念至今仍在计算机科学与技术领域中处于重要的地位,按“系列机”的思想来设计计算机仍是所有计算机厂商所必须遵循的原则。

同时,这一概念也影响了软件开发,导致软件工程领域出现了一个新的概念——软件体系结构。欲了解这一概念及相关技术,请学习“软件体系结构”课程或相关书籍。

2. 计算机组成(Computer Organization)与计算机实现(Computer Implementation)

伴随计算机体系结构概念的提出,两个与之相关联的概念也相继出现。它们是:计算机组成和计算机实现。

计算机组成是计算机体系结构的逻辑实现。一种计算机体系结构可以有多种不同的计算机组成,例如实现加法,不同的计算机可以选择不同的逻辑电路。计算机组成设计的任务是按照所要求的性能价格比,最佳、最合理地用各种器件和部件组成计算机,以实现设计规定的体系结构。

计算机实现是计算机组成的物理实现。一种计算机组成也可以有多种不同的计算机实现。计算机实现的设计内容包括:中央处理器、主存储器等部件的物理结构、封装技术和制造工艺;信号传输的频率、电气属性;电源、冷却、微组装技术和整机装配技术等。

例如,常见的微处理器散热方式有风冷、热管和液(水)冷三种。

风冷散热器(Air-Cooled Radiator)通常由一个散热风扇和一个散热片组成,如图1-6a所示。其散热原理是散热片通过和芯片表面接触,吸收芯片产生的热量,然后再通过风扇将散热片的热量吹散到空气中。风冷散热器具有结构简单、价格便宜、易于安装等优点,不足之处是工作时噪声较大。此外,风冷散热器对环境依赖较高,当气温升高时,其散热效果会明显降低。

热管散热器(Heat Pipe Radiator)是一种具有极高导热性能的传热元件,它通过在全封闭真空管内的液体(通常是水)的蒸发与凝结来传递热量,原理类似于电冰箱的压缩机制冷,散热效率较高。该类散热器通常由一个散热风扇和一组散热管组成,如图1-6b所示。

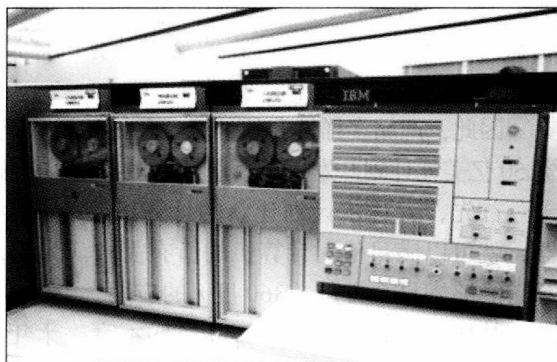


图 1-5 IBM S/360 的外观照片