

王道考研系列

2013年 计算机组成原理 联考复习指导

2013NIAN
JISUANJI ZUCHENG YUANLI
LIANKAO FUXI ZHIDAO

● 王道论坛 组编

更多计算机
考研和学习交流
尽在www.cskaoan.com

王道论坛由清华、北大、浙大、上交、哈工大、中科院等名校计算机研究生共同创办，致力于给报考计算机专业研究生的考生提供帮助和指导。王道论坛是唯一一家专注于名校计算机考研的论坛。

王道论坛组编的一系列计算机考研辅导书，融入了论坛众多名校高分选手的智慧，以及论坛交流精华，我们希望将其打造成“书本+在线”的学习方式。对于书中的疑难点，欢迎大家在论坛交流和讨论。

目前已有越来越多的名校采用上机的形式考查考生的动手编程能力，为方便大家练习，我们搭建了九度OJ(ac.jobdu.com)，收录了全国各大高校的复试上机题，希望能给考生复习上机考试提供强有力的支持。



中南大学出版社
www.csupress.com.cn

2013 年计算机组成原理 联考复习指导

王道论坛 组 编

主要作者介绍

李福龙：

本科毕业于哈尔滨工程大学，2010 年考入哈尔滨工业大学计算机科学与技术学院，初试状元，初试总分 410 分。

夏俭磊：

本科毕业于华北电力大学，2010 年考入哈尔滨工业大学计算机科学与技术学院，初试总分 387 分。

李昊迪：

本科毕业于哈尔滨工业大学，2011 年考入哈尔滨工业大学计算机科学与技术学院，初试总分 372 分，专业成绩 128 分。

宋景凯：

本科毕业于长春大学，2011 年考入哈尔滨工业大学计算机科学与技术学院，初试总分 366 分，专业成绩 131 分。



中南大学出版社
www.csupress.com.cn

内 容 简 介

.....

本书严格按照最新计算机考研大纲的计算机组成原理部分，对大纲所涉及的知识点进行集中梳理，力求内容精炼、重点突出、深入浅出。本书精选了名校历年考研真题，并给出详细的解题思路，力求达到讲练结合、灵活掌握、举一反三的功效。创新的“书本+在线”的学习方式，网上答疑，通过本书可大大提高考生的复习效果，达到事半功倍的复习效率。

本书可作为考生参加计算机专业研究生入学考试的备考复习用书，也可作为计算机专业的学生学习计算机组成原理课程的辅导用书。

王道论坛

www.cskaoyan.com

序

.....

当前,随着我国经济和科技高速发展,特别是计算机科学突飞猛进的发展,对计算机相关人才,尤其是中高端人才的需求也将不断增长。硕士研究生入学考试可视为人生的第二次大考试,它是改变命运、实现自我理想的又一次机会,而计算机专业一直是高校考研的热门专业之一。

自计算机专业研究生入学考试实行统一命题以来,初试科目包含了最重要的四门基础课程(数据结构、计算机组成原理、操作系统、计算机网络),很多学生普遍反映找不到方向,复习也无从下手。倘若有一套能够指导考生如何复习的好书,必将对考生的帮助匪浅。我的学生风华他们策划和编写了这一系列的计算机专业考研辅导书,重点突出,层次分明。他们结合了自身的复习经验、理解深度以及对大纲把握程度的体会,对考生而言是很有启发和指导意义的。

计算机这门学科,任何机械式的死记硬背都是收效甚微的。在全面深入复习之后,首先对诸多知识点分清主次,并结合做题,灵活运用所掌握的知识点,再选择一些高质量的模拟试题来检测自己理解和掌握的程度,查漏补缺。这符合我执教40余年来一直坚持“教材-习题集-试题库”的教学体系。

从风华他们策划并组建编写团队到初稿成型,直至最后定稿,我能体会到风华和他的团队确实倾注了大量的精力。这本书的出版一定会受到广大考研学生的欢迎,它会使你在考研的路上得到强有力的帮助。

序

前 言

.....

2011 年，由王道论坛组织名校高分选手，编写了 4 本单科辅导书。单科书是基于王道之前作品的二代作品，不论是编排方式，还是内容质量上都较前一版本的王道书有了较大的提升。这套书也参考了同类优秀的教材和辅导书，更是结合了高分选手们自己的复习经验。无论是对考点的讲解，还是习题的选择和解析，都结合了他们对专业课复习的独特见解。“王道考研系列”单科书，一共 4 本：

《2013 年数据结构联考复习指导》

《2013 年计算机组成原理联考复习指导》

《2013 年操作系统联考复习指导》

《2013 年计算机网络联考复习指导》

2011 年的单科书由于是第一年出版，时间较为仓促，小错误相对较多，给读者的复习带来了一些不便。今年，我们不仅修正了去年发现的全部错误，还对考点讲解做出了尽可能的优化，也重新筛选了部分习题，尤其是对习题的解析做出了更好的改进。

综观最近 4 年的考研试题，难度越来越高、题型越来越灵活，因此考取高分也越来越难。对于报考名校的考生，尤其是跨专业的考生来说，普遍会认为计算机专业课范围广、难度大、考题灵活。而对于一个想继续在计算机专业领域深造的考生来说，认真学习和扎实掌握这 4 门计算机专业中最基础的专业课，是最基本的前提。

当然，掌握专业课的知识点没有捷径可言，考生也不应怀有任何侥幸心理，扎扎实实地打好基础、踏踏实实地做题巩固，最后灵活致用才是高分的保障。我们只希望这套书能够指导大家复习考研，但学习还是得靠自己，高分不是建立在任何空中楼阁之上的。

“王道考研系列”的特色是“书本 + 在线”，你在复习中遇到的任何困难，都可以在王道论坛上发帖，论坛的热心道友，以及辅导员都会积极参与并与你交流。你的参与就是我们最大的鼓舞，任何一个建议，我们都会认真考虑，也会针对大家的意见对本书进行修订。

我们虽然尽最大努力来保证本书质量，但由于编写的时间仓促，以及编者的水平有限，书中如有错误或任何不当之处，望广大读者指正，我们将及时改正。

予人玫瑰，手有余香，王道论坛伴你一路同行！

目 录

第 1 章 计算机系统概述	(1)
1.1 计算机发展历程	(1)
1.1.1 计算机硬件的发展	(1)
1.1.2 计算机软件的发展	(2)
1.1.3 计算机的分类与发展方向	(2)
1.1.4 本节习题精选	(3)
1.1.5 答案与解析	(3)
1.2 计算机系统层次结构	(4)
1.2.1 计算机系统的组成	(4)
1.2.2 计算机硬件的基本组成	(4)
1.2.3 计算机软件的分	(7)
1.2.4 计算机的工作过程	(7)
1.2.5 计算机系统的多级层次结构	(8)
1.2.6 本节习题精选	(9)
1.2.7 答案与解析	(11)
1.3 计算机的性能指标	(13)
1.3.1 计算机的主要性能指标	(13)
1.3.2 几个专业术语的概念	(14)
1.3.3 本节习题精选	(14)
1.3.4 答案与解析	(16)
1.4 常见问题和易混淆知识点	(19)
第 2 章 数据的表示和运算	(21)
2.1 数制与编码	(22)
2.1.1 进位计数制及其相互转换	(22)
2.1.2 真值和机器数	(24)
2.1.3 BCD 码	(24)
2.1.4 字符与字符串	(25)
2.1.5 校验码	(26)
2.1.6 本节习题精选	(29)
2.1.7 答案与解析	(31)
2.2 定点数的表示与运算	(33)

2.2.1	定点数的表示	(33)
2.2.2	定点数的运算	(37)
2.2.3	本节习题精选	(44)
2.2.4	答案与解析	(50)
2.3	浮点数的表示与运算	(56)
2.3.1	浮点数的表示	(56)
2.3.2	浮点数的加减运算	(60)
2.3.3	本节习题精选	(61)
2.3.4	答案与解析	(65)
2.4	算术逻辑单元 ALU	(71)
2.4.1	串行加法器和并行加法器	(71)
2.4.2	算术逻辑单元的功能和结构	(74)
2.4.3	本节习题精选	(76)
2.4.4	答案与解析	(77)
2.5	常见问题和易混淆知识点	(78)
第3章	存储系统的层次结构	(81)
3.1	存储器的分类	(82)
3.1.1	存储器的分类	(82)
3.1.2	存储器的性能指标	(83)
3.1.3	本节习题精选	(83)
3.1.4	答案与解析	(84)
3.2	存储器的层次化结构	(85)
3.2.1	多级存储系统	(85)
3.2.2	本节习题精选	(86)
3.2.3	答案与解析	(87)
3.3	半导体随机存储器	(87)
3.3.1	半导体存储芯片	(87)
3.3.2	SRAM 存储器和 DRAM 存储器	(88)
3.3.3	只读存储器	(91)
3.3.4	本节习题精选	(91)
3.3.5	答案与解析	(93)
3.4	主存储器与 CPU 的连接	(95)
3.4.1	连接原理	(95)
3.4.2	主存容量的扩展	(95)
3.4.3	存储芯片的地址分配和片选	(97)
3.4.4	存储器与 CPU 的连接	(98)
3.4.5	本节习题精选	(99)
3.4.6	答案与解析	(101)

3.5	双口 RAM 和多模块存储器	(106)
3.5.1	双端口 RAM	(106)
3.5.2	多模块存储器	(107)
3.5.3	本节习题精选	(109)
3.5.4	答案与解析	(110)
3.6	高速缓冲存储器	(111)
3.6.1	程序访问的局部性原理	(111)
3.6.2	Cache 的基本工作原理	(111)
3.6.3	Cache 和主存的映射方式	(112)
3.6.4	Cache 中主存块的替换算法	(115)
3.6.5	Cache 写策略	(115)
3.6.6	本节习题精选	(116)
3.6.7	答案与解析	(118)
3.7	虚拟存储器	(123)
3.7.1	虚拟存储器的基本概念	(123)
3.7.2	页式虚拟存储器	(124)
3.7.3	段式虚拟存储器	(125)
3.7.4	段页式虚拟存储器	(126)
3.7.5	快表 TLB	(126)
3.7.6	虚拟存储器与 Cache 的比较	(126)
3.7.7	本节习题精选	(127)
3.7.7	答案与解析	(130)
3.8	常见问题和易混淆知识点	(133)
第 4 章	指令系统	(135)
4.1	指令格式	(135)
4.1.1	指令的基本格式	(135)
4.1.2	定长操作码指令格式	(137)
4.1.3	扩展操作码指令格式	(137)
4.1.4	本节习题精选	(138)
4.1.5	答案与解析	(140)
4.2	指令寻址方式	(143)
4.2.1	指令寻址和数据寻址	(143)
4.2.2	常见的数据寻址方式	(144)
4.2.3	本节习题精选	(148)
4.2.4	答案与解析	(152)
4.3	CISC 和 RISC 的基本概念	(155)
4.3.1	复杂指令系统计算机 CISC	(156)
4.3.2	精简指令系统计算机 RISC	(156)

4.3.3	CISC 和 RISC 的比较	(156)
4.3.4	本节习题精选	(157)
4.3.5	答案与解析	(158)
4.4	常见问题和易混淆知识点	(159)
第5章	中央处理器	(161)
5.1	CPU 的功能和基本结构	(161)
5.1.1	CPU 的功能	(161)
5.1.2	CPU 的基本结构	(162)
5.1.3	本节习题精选	(163)
5.1.4	答案与解析	(165)
5.2	指令执行过程	(167)
5.2.1	指令周期	(167)
5.2.2	指令周期的数据流	(168)
5.2.3	指令执行方案	(170)
5.2.4	本节习题精选	(171)
5.2.5	答案与解析	(173)
5.3	数据通路的功能和基本结构	(175)
5.3.1	数据通路的功能	(175)
5.3.2	数据通路的基本结构	(175)
5.3.3	本节习题精选	(176)
5.3.4	答案与解析	(181)
5.4	控制器的功能和工作原理	(187)
5.4.1	控制器的结构和功能	(187)
5.4.2	硬布线控制器	(187)
5.4.3	微程序控制器	(192)
5.4.4	本节习题精选	(192)
5.4.5	答案与解析	(201)
5.5	指令流水线	(205)
5.5.1	指令流水线的基本概念	(205)
5.5.2	流水线的分类	(207)
5.5.3	影响流水线的因素	(208)
5.5.4	流水线的性能指标	(209)
5.5.5	超标量流水线的概念	(210)
5.5.6	本节习题精选	(211)
5.5.7	答案与解析	(213)
5.6	多核处理器的基本概念	(217)
5.6.1	多核处理器的发展简述	(217)
5.6.2	多核处理器的基本概念	(217)

5.6.3 多核处理器的主要技术和挑战	(217)
5.6.4 本节习题精选	(218)
5.6.5 答案与解析	(218)
5.7 常见问题和易混淆知识点	(218)

第6章 总线

6.1 总线概述	(220)
6.1.1 总线基本概念	(220)
6.1.2 总线的分类	(221)
6.1.3 系统总线的结构	(221)
6.1.4 总线的性能指标	(223)
6.1.5 本节习题精选	(223)
6.1.6 答案与解析	(225)
6.2 总线仲裁	(227)
6.2.1 集中仲裁方式	(228)
6.2.2 分布仲裁方式	(229)
6.2.3 本节习题精选	(230)
6.2.4 答案与解析	(230)
6.3 总线操作和定时	(231)
6.3.1 总线传输的4个阶段	(231)
6.3.2 同步定时方式	(232)
6.3.3 异步定时方式	(232)
6.3.4 本节习题精选	(233)
6.3.5 答案与解析	(234)
6.4 总线标准	(235)
6.4.1 常见的总线标准	(235)
6.4.2 本节习题精选	(236)
6.4.3 答案与解析	(236)
6.5 常见问题和易混淆知识点	(237)

第7章 输入/输出系统

7.1 I/O系统基本概念	(238)
7.1.1 输入/输出系统	(238)
7.1.2 I/O控制方式	(239)
7.1.3 本节习题精选	(239)
7.1.4 答案与解析	(240)
7.2 外部设备	(240)
7.2.1 输入设备	(240)
7.2.2 输出设备	(240)

7.2.3	外存储器	(242)
7.2.4	本节习题精选	(245)
7.2.5	答案与解析	(246)
7.3	I/O 接口	(247)
7.3.1	I/O 接口的功能	(248)
7.3.2	I/O 接口的基本结构	(248)
7.3.3	I/O 接口的类型	(249)
7.3.4	I/O 端口及其编址	(249)
7.3.5	本节习题精选	(249)
7.3.6	习题精选	(250)
7.4	I/O 方式	(251)
7.4.1	程序查询方式	(251)
7.4.2	程序中断方式	(252)
7.4.3	DMA 方式	(257)
7.4.4	通道方式	(260)
7.4.5	本节习题精选	(262)
7.4.6	答案与解析	(268)
7.5	常见问题和易混淆知识点	(275)
参考文献	(277)

王道论坛

www.cskaoyan.com

第 1 章 计算机系统概述

【考纲内容】

- (一)计算机发展历程
- (二)计算机系统层次结构
1. 计算机硬件的基本组成
2. 计算机软件的分类
3. 计算机的工作过程
- (三)计算机性能指标
- 吞吐量、响应时间，CPU 时钟周期、主频、CPI、CPU 执行时间，MIPS、MFLOPS。

【考题分布】

年份	单选题/分	综合题/分	考查内容
2009 年	1 题 ×2	0	冯·诺依曼计算机基本特点与指令执行过程
2010 年	1 题 ×2	√ ^①	计算机性能指标；MAR 与地址空间、MDR 与字长的关系
2011 年	1 题 ×2	0	计算机的性能指标的定义
2012 年	1 题 ×2	√	CPU 执行时间的相关计算；MIPS 的计算

①打“√”表示有综合应用题部分涉及本章的知识点。

本章是组成原理的概述，易对有关概念或性能指标出选择题，也可能综合后续章节的内容出有关性能分析的综合题。掌握本章的基本概念，是学好后续章节的基础。部分知识点在初学时理解不甚深刻也无需担忧，相信随着后续章节的学习一定会有更为深入的理解。

1.1 计算机发展历程

1.1.1 计算机硬件的发展

1. 计算机的四代变化

从 1946 年世界上第一台电子数字计算机 ENIAC (Electronic Numerical Integrator And Computer) 问世以来，计算机的发展已经经历了四代。

(1)第一代计算机(1946 ~ 1957 年)——电子管时代

特点：逻辑元件采用电子管；使用机器语言进行编程；主存用延迟线或磁鼓存储信息，容量极小；体积庞大，成本高；运算速度较低，一般只有每秒几千到几万次。

(2) 第二代计算机(1958 ~ 1964 年)——晶体管时代

特点: 逻辑元件采用晶体管; 运算速度提高到每秒几万到几十万次; 主存使用磁芯存储器; 软件开始使用高级语言, 如 FORTRAN, 有了操作系统的雏形。

(3) 第三代计算机(1965 ~ 1971 年)——中小规模集成电路时代

特点: 逻辑元件采用中小规模集成电路; 半导体存储器开始取代磁芯存储器; 高级语言发展迅速, 操作系统也进一步发展, 开始有了分时操作系统。

(4) 第四代计算机(1972 年 ~ 现在)——超大规模集成电路时代

特点: 逻辑元件采用大规模集成电路和超大规模集成电路, 并产生了微处理器; 诸如并行、流水线、高速缓存和虚拟存储器等概念用在了此代计算机中。

2. 计算机元件的更新换代

1) 摩尔定律。当价格不变时, 集成电路上可容纳的晶体管数目, 约每隔 18 个月便会增加一倍, 性能也将提升一倍。也就是说我们现在和 18 个月后花同样的钱买到 CPU, 后者的性能是前者的两倍。这一定律揭示了信息技术进步的速度。

2) 半导体存储器的发展。1970 年, 仙童公司生产出第一个较大容量的半导体存储器, 至今, 半导体存储器经历了 11 代: 单芯片 1 KB、4 KB、16 KB、64 KB、256 KB、1 MB、4 MB、16 MB、64 MB、256 MB 和现在的 1 GB。

3) 微处理器的发展。自 1971 年 Intel 公司开发出第一个微处理器 Intel 4004 至今, 微处理器经历了 Intel 8008(8 位)、Intel 8080(8 位)、Intel 8086(16 位)、Intel 8088(16 位)、Intel 80286(16 位)、Intel 80386(32 位)、Intel 80486(32 位)、Pentium(32 位)、Pentium pro(64 位)、Pentium II(64 位)、Pentium III(64 位)、Pentium 4(64 位)等。

1.1.2 计算机软件的发展

计算机软件技术的蓬勃发展, 也为计算机系统的发展做出了很大的贡献。

计算机语言的发展经历了面向机器的机器语言和汇编语言、面向问题的高级语言。其中高级语言的发展真正促进了软件的发展, 它经历了从科学计算和工程计算的 FORTRAN、结构化程序设计 PASCAL 到面向对象的 C++ 和适应网络环境的 Java。

与此同时, 直接影响计算机系统性能提升的各种系统软件也有了长足的发展, 特别是微机的操作系统, 从 DOS 发展到目前的视窗与网络操作系统(代表分别为 Windows 与 UNIX)。

1.1.3 计算机的分类与发展方向

电子计算机可分为电子模拟计算机和电子数字计算机。

数字计算机又可按用途可分为专用计算机和通用计算机。这是根据计算机的效率、速度、价格以及运行的经济性和适应性来划分的。

通用计算机又分为巨型机、大型机、中型机、小型机、微型机和单片机 6 类, 它们的体积、功耗、性能、数据存储量、指令系统的复杂程度和价格依次递减。

此外, 计算机按指令和数据流还可分为:

1) 单指令流和单数据流系统(SISD), 也即传统冯·诺依曼体系结构。

2) 单指令流和多数据流系统(SIMD), 包括阵列处理器和向量处理器系统。

3)多指令流和单数据流系统(MISD),这种计算机实际上不存在。

4)多指令流和多数据流系统(MIMD),包括多处理器和多计算机系统。

计算机的发展趋势正向着“两极”分化。一极是微型计算机向更微型化、网络化、高性能、多用途方向发展;另一极则是巨型机向更巨型化、超高速、并行处理、智能化方向发展。

1.1.4 本节习题精选

一、单项选择题

1. 电子计算机的发展已经经历了4代,这4代计算机的主要元件分别是()。

- A. 电子管、晶体管、中小规模集成电路、激光器件
- B. 晶体管、中小规模集成电路、激光器件、光介质
- C. 电子管、晶体管、中小规模集成电路、大规模集成电路
- D. 电子管、数码管、中小规模集成电路、激光器件

2. 微型计算机的发展以()技术为标志。

- A. 操作系统
- B. 微处理器
- C. 磁盘
- D. 软件

3. 可以在计算机中直接执行的语言和用助记符编写的语言分别是()。

- I. 机器语言
- II. 汇编语言
- III. 高级语言
- IV. 操作系统原语
- V. 正则语言

- A. II、III
- B. II、IV
- C. I、II
- D. I、V

4. 只有当程序执行时才将源程序翻译成机器语言,并且一次只能翻译一行语句,边翻译边执行的是()程序,把汇编语言源程序转变为机器语言程序的过程是()。

- I. 编译
- II. 目标
- III. 汇编
- IV. 解释
- A. I、II
- B. IV、II
- C. IV、I
- D. IV、III

5. 到目前为止,计算机中所有的信息仍以二进制方式表示的理由是()。

- A. 节约元件
- B. 运算速度快
- C. 由物理器件的性能决定
- D. 信息处理方便

1.1.5 答案与解析

一、单项选择题

1. C

此题也可以根据元件的先进程度的升序得出答案。

2. B

微型计算机的发展是以微处理器的技术为标志的。

3. C

机器语言是计算机唯一可以直接执行的语言,汇编语言用助记符编写,以便记忆。

4. D

解释程序的特点是翻译一句执行一句,边翻译边执行;由高级语言转化为汇编语言的过程叫做编译,把汇编语言源程序翻译成机器语言程序的过程称为汇编。

5. C

二进制只有 1 和 0 两个数字，刚好和逻辑电路里的高、低电平对应，实现起来比较方便且简单可靠，故由物理器件的性能决定。

1.2 计算机系统层次结构

1.2.1 计算机系统的组成

硬件系统和软件系统共同构成了一个完整的计算机系统。硬件是指有形的物理设备，是计算机系统中实际物理装置的总称。软件则是指在硬件上运行的程序和相关的文档及文档。

一个计算机系统性能的好坏，很大程度上是由软件的效率和作用来表征的，而软件性能的发挥又离不开硬件的支持。对于某一功能来说，其既可以用软件实现，也可以用硬件实现，则称为**软硬件在逻辑上是等效的**。在设计计算机系统时，要进行软硬件的功能分配。通常来说，一个功能若使用较为频繁而且用硬件实现的成本理想的话，使用硬件解决可以提高效率。而用软件实现可以提高灵活性，但是效率往往不如硬件实现高。

1.2.2 计算机硬件的基本组成

1. 早期的冯·诺依曼机

冯·诺依曼在研究 EDVAC 机时提出了“存储程序”的概念，“存储程序”的思想奠定了现代计算机的基本结构，以此概念为基础的各类计算机通称为冯·诺依曼机，其特点如下。

- 1) 计算机硬件系统由运算器、存储器、控制器、输入设备和输出设备 5 大部件组成。
- 2) 指令和数据以同等地位存于存储器内，并可按地址寻访。
- 3) 指令和数据均用二进制代码表示。
- 4) 指令由操作码和地址码组成，操作码用来表示操作的性质，地址码用来表示操作数在存储器中的位置。
- 5) 指令在存储器内按顺序存放。通常，指令是顺序执行的，在特定条件下，可根据运算结果或根据设定的条件改变执行顺序。
- 6) 早期的冯·诺依曼机以运算器为中心，输入/输出设备通过运算器与存储器传送数据。

典型的冯·诺依曼计算机结构如图 1-1 所示。

注意：“存储程序”的概念是指将指令以代码的形式事先输入到计算机主存储器中，然后按其在存储器中的首地址执行程序的第一条指令，以后就按照该程序的规定顺序执行其他指令，直至程序执行结束。

2. 现代计算机的组织结构

在微处理器问世之前，运算器和控制器分离，而且存储器的容量很小，故而设计成以运算器为中心，其他部件都通过运算器完成信息的传递，如图 1-1 所示。

而随着微电子技术的进步，同时计算机需要处理、加工的信息量也与日俱增，大量 I/O

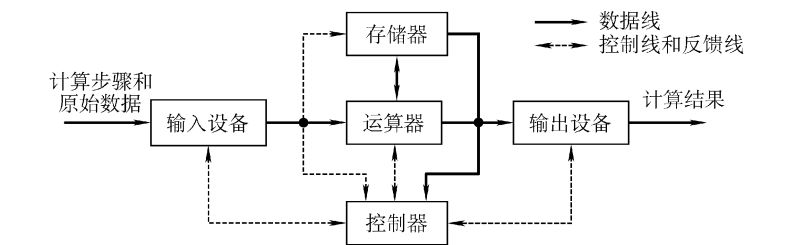


图 1-1 典型的冯·诺依曼计算机结构

设备的速度和 CPU 的速度差距悬殊，故而以运算器为中心的结构不能够满足计算机发展的要求。现代计算机已经发展为以存储器为中心，使 I/O 操作尽可能地绕过 CPU，直接在 I/O 设备和存储器之间完成，以提高系统的整体运行效率，其结构如图 1-2 所示。

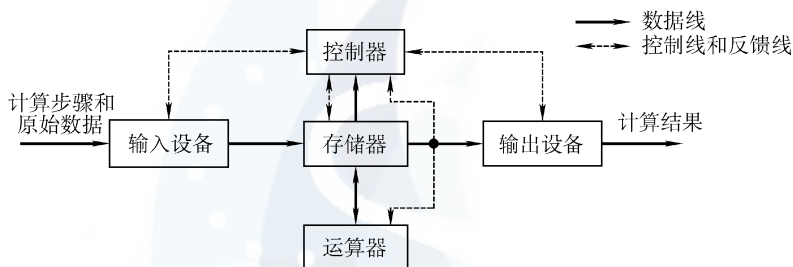


图 1-2 以存储器为中心的计算机结构

目前绝大多数现代计算机仍遵循冯·诺依曼的存储程序的设计思想。

3. 计算机的功能部件

传统冯·诺依曼计算机和现代计算机的结构虽然有所不同，但功能部件是一致的，它们的功能部件包括：

(1) 输入设备

输入设备的主要功能是将程序和数据以机器所能识别和接受的信息形式输入到计算机。

最常用也是最基本的输入设备是键盘，此外还有鼠标、扫描仪、摄像机等。

(2) 输出设备

输出设备的任务是将计算机处理的结果以人们所能接受的形式或其他系统所要求的信息形式输出。

最常用、最基本的输出设备是显示器、打印机。计算机的输入、输出设备(简称 I/O 设备)是计算机与外界联系的桥梁，是计算机中不可缺少的一个重要组成部分。

(3) 存储器

存储器是计算机的存储部件，用来存放程序和数据。

存储器分为主存储器(简称主存，也称内存储器)和辅助存储器(简称辅存，也称外存

储器)。CPU 能够直接访问的存储器是主存储器。辅助存储器用于帮助主存储器记忆更多的信息,辅助存储器中的信息必须调入主存后,才能为 CPU 所访问。

主存储器由许多**存储单元**组成,每个存储单元包含若干个**存储元件**,每个元件存储一位二进制代码“0”或“1”。故而存储单元可存储一串二进制代码,称这串代码为**存储字**,这串代码的位数称为**存储字长**,存储字长可以是一个字节(8 bit)或者是字节的偶数倍。

主存储器的工作方式是按**存储单元的地址**进行存取的,这种存取方式称为按地址存取方式(相联存储器是按内容访问的)。

主存储器的最基本组成如图 1-3 所示。存储体存放二进制信息,**地址寄存器(MAR)**存放访问地址,经过地址译码后找到所选的存储单元。**数据寄存器(MDR)**是主存和其他部件的中介机构,用于暂存要从存储器中读或者写的信息,时序控制逻辑用于产生存储器操作所需的各种时序信号。

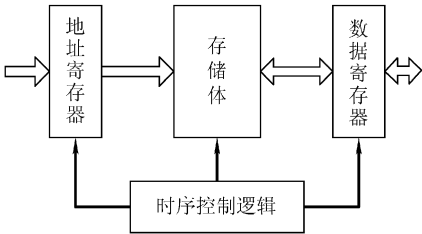


图 1-3 主存储器逻辑图

MAR 的位数对应存储单元的个数,如 MAR 为 10 位则有 $2^{10} = 1024$ 个存储单元,记为 1 KB。

MDR 的位数和存储字长相等。

(4) 运算器

运算器是计算机的执行部件,用于对数据进行加工处理,完成算术运算和逻辑运算。算术运算是按照算术运算规则如加、减、乘、除进行的运算;逻辑运算则是如与、或、非、异或、比较、移位等原则进行的运算。

运算器的核心是算术逻辑单元 ALU(Arithmetic and Logical Unit)。运算器包含若干通用寄存器,用于暂存操作数和中间结果,如累加器(ACC)、乘商寄存器(MQ)、操作数寄存器(X)、变址寄存器(IX)、基址寄存器(BR)等,其中前 3 个寄存器是必须有的。

运算器内还有程序状态寄存器(PSW),保留各类运算指令或测试指令的结果的各类状态信息,以表征系统运行状态。

(5) 控制器

控制器是计算机的指挥中心,由其“指挥”各部件自动协调地进行工作。控制器由程序计数器(PC)、指令寄存器(IR)、控制单元(CU)组成。

PC 用来存放当前欲执行指令的地址,可以自动 +1 以形成下一条指令的地址,它与主存的 MAR 之间有一条直接通路。

IR 用来存放当前的指令,其内容来自主存的 MDR。指令中的操作码 OP(IR)送至 CU,用以分析指令并发出各种微操作命令序列,而地址码 Ad(IR)送往 MAR 来取操作数。

如图 1-4 所示为一个更细化的计算机组成框图。现代计算机一般是将运算器和控制器集成到同一个芯片上,合称为中央处理器,简称 CPU。CPU 和主存储器共同构成主机,而计算机中除去主机的其他硬件装置(如 I/O)统称为外部设备(简称外设)。也就是说,外设主要包括外存和 I/O 设备。

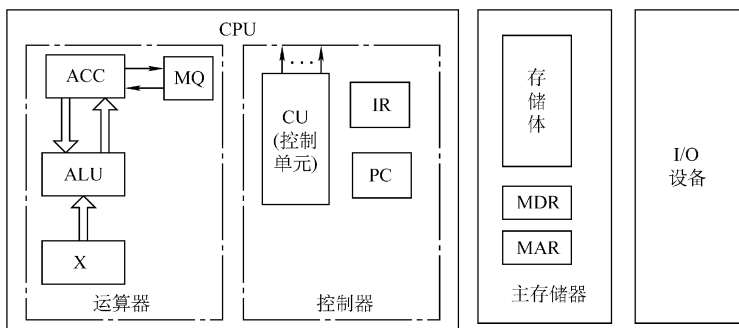


图 1-4 细化的计算机组成框图

1.2.3 计算机软件的分类

1. 系统软件和应用软件

软件按其功能分类可以分为系统软件和应用软件。

系统软件是一组保证计算机系统高效、正确运行的基础软件，通常作为系统资源提供给用户使用。主要有操作系统(OS)、数据库管理系统(DBMS)、语言处理程序、分布式软件系统、网络软件系统、标准库程序、服务性程序等。

应用软件是指用户为解决某个应用领域中的各类问题而编制的程序，如各种科学计算类程序、工程设计类程序、数据统计与处理程序等。

注意：数据库管理系统(DBMS)和数据库系统(DBS)是有区别的。DBMS是位于用户和操作系统之间的一层数据管理软件，是系统软件；而DBS是指计算机系统中引入数据库后的系统，一般由数据库、数据库管理系统、数据库管理员(DBA)和应用系统构成。

2. 三个级别的语言

1) 机器语言：又称为二进制代码语言，需要编程人员记忆每一条指令的二进制编码。机器语言是计算机唯一可以直接识别和执行的语言。

2) 汇编语言：汇编语言用英文单词或其缩写代替二进制的指令代码，更容易为人们记忆和理解。汇编语言的程序必须经过一个称为汇编程序的系统软件的翻译，将其转换为计算机的机器语言后，才能在计算机的硬件系统上执行。

3) 高级语言：高级语言(如C、C++、Java等)更多地是为了方便程序设计人员写出解决问题的处理方案和解题过程的程序。通常高级语言需要经过编译程序编译成汇编语言程序，然后经过汇编操作得到机器语言程序，或者直接由高级语言程序翻译成机器语言程序。

1.2.4 计算机的工作过程

计算机的工作过程分为以下几个步骤。

- 1) 把程序和数据装入到主存储器中。
- 2) 从程序的起始地址运行程序。

3) 用程序的首地址从存储器中取出第一条指令，经过译码、执行步骤等控制计算机各功能部件协同运行，完成这条指令功能，并计算下一条指令的地址。

4) 用新得到的指令地址继续读出第二条指令并执行，直到程序结束为止；每一条指令都是在取指、译码和执行的循环过程中完成的。

下面以取数指令(即将指令地址码指示的存储单元中的操作数取出后送至运算器的ACC中)为例，其信息流程如下：

取指令： $PC \rightarrow MAR \rightarrow M \rightarrow MDR \rightarrow IR$

分析指令： $OP(IR) \rightarrow CU$

执行指令： $Ad(IR) \rightarrow MAR \rightarrow M \rightarrow MDR \rightarrow ACC$

此外，每取完一条指令，还必须为取下条指令作准备，形成下一条指令的地址，即 $(PC) + 1 \rightarrow PC$ 。

注意： (PC) 指程序计数器PC中存放的内容。 $PC \rightarrow MAR$ 应理解为 $(PC) \rightarrow MAR$ ，即程序计数器中的值经数据通路送到MAR，也表示数据通路时括号可省略(因为只是表示数据流经的途径，而不强调数据本身的流动)。但是运算时括号不能省略，即 $(PC) + 1 \rightarrow PC$ 不能写为 $PC + 1 \rightarrow PC$ 。当题目中 $(PC) \rightarrow MAR$ 的括号没有省略时，考生最好也不要省略。

1.2.5 计算机系统的多级层次结构

现代计算机是一个硬件与软件组成的综合体。由于面对的应用范围越来越广，所以必须有复杂的系统软件和硬件的支持。由于软件、硬件的设计者和使用者都从不同的角度，以及各种不同的语言对待同一个计算机系统，因此，他们各自看到的计算机系统的属性及对计算机系统提出的要求也就不一样。

计算机系统的多级层次结构，就是针对上述情况，根据从各种角度所看到的机器之间的有机联系，分清彼此之间的界面，明确各自的功能，以便构成合理、高效的计算机系统。

关于计算机系统层次结构的分层方式，目前尚无统一的标准，这里采用如图1-5所示的层次结构。

第1级是微程序机器层，这是一个实在的硬件层，它由机器硬件直接执行微指令。

第2级是传统机器语言层，它也是一个实际的机器层，由微程序解释机器指令系统。

第3级是操作系统层，它由操作系统程序实现。操作系统程序是由机器指令和广义指令组成的，这些广义指令是为了扩展机器功能而设置的，它是由操作系统定义和解释的软件指令，所以这一层也称为混合层。

第4级是汇编语言层，它为用户提供一种符号化的语言，借此可编写汇编语言源程序。

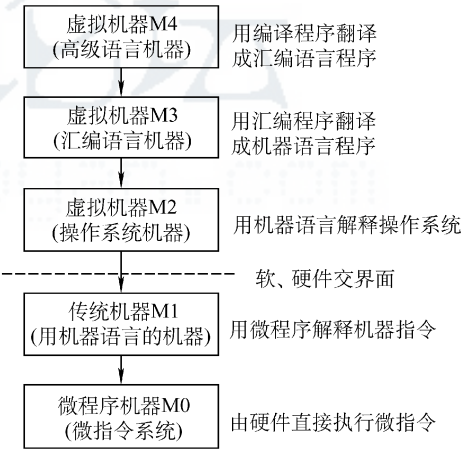


图 1-5 计算机系统的多级层次结构

这一层由汇编程序支持和执行。

第5级是高级语言层,它是面向用户的,为方便用户编写应用程序而设置的。该层由各种高级语言编译程序支持和执行。

在高级语言层之上,还可以有应用层,由解决实际问题 and 应用问题的处理程序组成,如文字处理软件、数据库软件、多媒体处理软件和办公自动化软件等。

通常把没有配备软件的纯硬件系统称为“裸机”。第3层~第5层称为虚拟机,简单来说,就是软件实现的机器。虚拟机只对该层的观察者存在,这里的分层和计算机网络的分层类似,对于某层的观察者来说,只能通过该层次的语言来了解和使用计算机,至于下层是如何工作就不必关心了。

层次之间的关系紧密,下层是上层的基础,上层是下层的扩展。随着超大规模集成电路技术的不断发展,部分软件功能将由硬件来实现,因而软硬件交界面的划分也不是绝对的。

本课程主要讨论传统机器 M1 和微程序机器 M0 的组成原理及设计思想。

1.2.6 本节习题精选

一、单项选择题

- 完整的计算机系统应包括()。
 - 运算器、存储器、控制器
 - 外部设备和主机
 - 主机和应用程序
 - 配套的硬件设备和软件系统
- 冯·诺依曼机的基本工作方式是()。
 - 控制流驱动方式
 - 多指令多数据流方式
 - 微程序控制方式
 - 数据流驱动方式
- 下列()是冯·诺依曼机工作方式的基本特点。
 - 多指令流单数据流
 - 按地址访问并顺序执行指令
 - 堆栈操作
 - 存储器按内容选择地址
- 【2009年计算机联考真题】冯·诺依曼计算机中指令和数据均以二进制形式存放在存储器中,CPU区分它们的依据是()。
 - 指令操作码的译码结果
 - 指令和数据的寻址方式
 - 指令周期的不同阶段
 - 指令和数据所在的存储单元
- 以下说法错误的是()。
 - 硬盘是外部设备
 - 软件的功能与硬件的功能在逻辑上是等效的
 - 硬件实现的功能一般比软件实现具有更高的执行速度
 - 软件的功能不能用硬件取代
- 存放欲执行指令的寄存器是()。
 - MAR
 - PC
 - MDR
 - IR
- 在CPU中,跟踪下一条要执行的指令的地址的寄存器是()。
 - PC
 - MAR
 - MDR
 - IR
- CPU不包括()。
 - 地址寄存器
 - 指令寄存器(IR)

C. 地址译码器

D. 通用寄存器

9. 在运算器中, 不包含()。

A. 状态寄存器

B. 数据总线

C. ALU

D. 地址寄存器

10. 下列关于 CPU 存取速度的比较中, 正确的是()。

A. Cache > 内存 > 寄存器

B. Cache > 寄存器 > 内存

C. 寄存器 > Cache > 内存

D. 寄存器 > 内存 > Cache

11. 一个 8 位的计算机系统以 16 位来表示地址, 则该计算机系统有()个地址空间。

A. 256

B. 65 535

C. 65 536

D. 131 072

12. ()是程序运行时的存储位置, 包括所需的数据。

A. 数据通路

B. 主存

C. 硬盘

D. 操作系统

13. 下列()属于应用软件。

A. 操作系统

B. 编译程序

C. 连接程序

D. 文本处理

14. 关于编译程序和解释程序, 下面说法错误的是()。

A. 编译程序和解释程序的作用都是将高级语言程序转换成机器语言程序

B. 编译程序编译时间较长, 运行速度较快

C. 解释程序方法较简单, 运行速度也较快

D. 解释程序将源程序翻译成机器语言, 并且翻译一条以后, 立即执行这条语句

15. 下列叙述中正确的是()。

A. 寄存器的设置对汇编语言是透明的

B. 实际应用程序的测试结果能够全面代表计算机的性能

C. 系列机的基本特性是指令系统向后兼容

D. 软件和硬件在逻辑功能上是等价的

16. 指令流通常是()。

A. 从主存流向控制器

B. 从控制器流向主存

C. 从控制器流向控制器

D. 从主存流向主存

17. 在 CPU 的组成中, 不包括()。

A. 运算器

B. 存储器

C. 控制器

D. 寄存器

18. 下列()不属于系统程序。

A. 数据库系统

B. 操作系统

C. 编译程序

D. 以上 3 种都属于系统程序

19. 关于相联存储器, 下列说法正确的是()。

A. 只可以按地址寻址

B. 只可以按内容寻址

C. 既可以按地址寻址又可以按内容寻址

D. 以上说法均不完善

20. 计算机系统的层次结构可以分为 6 层, 其层次之间的依存关系是()。

A. 上下层之间相互无关

B. 上层实现对下层的功能扩展, 而下层是实现上层的基础

- C. 上层实现对下层的扩展作用, 而下层对上层有限制作用
- D. 上层和下层的关系是相互依存、不可分割的

二、综合应用题

1. 什么是存储程序原理? 按此原理, 计算机应具有哪几大功能?

1.2.7 答案与解析

一、单项选择题

1. D

A 是计算机主机的组成部分, 而 B、C 只涉及了计算机系统的部分内容, 都不完整。

2. A

早期的冯·诺依曼机以运算器为中心, 且是单处理机, B 是多处理机。冯·诺依曼机最根本特征是采用“存储程序”原理, 基本工作方式是控制流驱动方式。

3. B

A 是不存在的机器, B 是对“存储程序”的阐述, 故正确。C 是与题干无关的选项。D 是相连存储器的特点。

4. C

通常完成一条指令可分为取指阶段和执行阶段。在取指阶段通过访问存储器可将指令取出; 在执行阶段通过访问存储器可将操作数取出。这样, 虽然指令和数据都是以 0、1 代码形式存在存储器中, 但 CPU 可以根据指令周期的不同阶段, 判断出在取指阶段访问存储器取出的 0、1 代码是指令, 在执行阶段访存取出的 0、1 代码是数据。

5. D

软件和硬件具有逻辑上的等效性, 硬件实现具有更高的执行速度, 软件实现具有更好的灵活性, 通常对执行频繁、硬件实现代价不是很高的功能由硬件实现。故选 D。

6. D

IR 存放当前欲执行的指令, PC 存放下一条指令的地址, 不要将它们混淆。此外, MAR 用来存放欲访问的存储单元地址, MDR 存放从存储单元取来的数据。

7. A

在 CPU 中, 程序计数器 PC 用来跟踪下一条要执行的指令在主存储器中的地址。

8. C

地址译码器是主存的构成部分, 不属于 CPU。地址寄存器虽然一般属于主存, 但是现代计算机中绝大多数 CPU 内集成了地址寄存器。

9. D

运算器的核心部分是算术逻辑运算单元(ALU)。地址寄存器位于 CPU 内, 但并没有集成在运算器与控制器中。地址寄存器用来保存当前 CPU 所访问的内存单元的地址。由于内存和 CPU 之间存在着操作速度上的差别, 所以必须使用地址寄存器来保持地址信息, 直到内存的读/写操作完成为止。

10. C

寄存器在 CPU 内部, 速度最快。Cache 采用高速的 SRAM 制作, 而内存常用 DRAM 制

作，其速度较 Cache 慢。本题也可根据存储器层次结构的速度关系得出答案。

11. C

8 位计算机表明计算机字长为 8 位，即一次可以处理 8 位的数据，而 16 位表示地址码的长度，故而该机器有 $2^{16} = 65\,536$ 个地址空间。

12. B

计算机只能从主存中取指令与操作数，不能直接与外存交换数据。

13. D

操作系统属于大型系统软件；编译程序属于语言处理程序；连接程序属于服务性程序，故选 D。

14. C

编译程序是先完整编译后运行的程序，如 C、C++ 等；解释程序是一句一句翻译且边翻译边执行的程序，如 JavaScript、Python 等。由于解释程序要边翻译成机器语言边执行，故而一般速度较编译程序慢。

15. C

寄存器的设置对汇编语言不透明，汇编程序员要对寄存器进行直接操作。全面代表计算机性能的是实际软件的运行情况。软件和硬件在逻辑上是等效的，但不是等价的。向后兼容指的是时间上向后兼容，即新机器兼容使用以前机器的指令系统。

16. A

指令是存放在主存中的，在主存中取出指令后送入控制器进行分析并发出各种操作序列。因此指令流是从主存流向控制器的。

17. B

CPU 由运算器和控制器两个部件组成，而运算器和控制器中都含有寄存器。而存储器是一个独立的部件。

18. A

数据库系统是指在计算机系统中引入数据库后的系统，一般由数据库、数据库管理系统、应用系统、数据库管理员构成，其中数据库管理系统是系统程序。

19. C

相联存储器既可以按地址寻址又可以按内容（通常是某些字段）寻址，为与传统存储器区别，又称为按内容寻址的存储器。

20. B

在计算机多层次结构中，上下层是可以分割的，且上层是下层的功能实现。此外，上层在下层的基础上实现了更加丰富的功能，仅有下层而没有上层也是可以的。

二、综合应用题

1. 解答：

存储程序是指将指令以代码的形式事先输入到计算机主存储器中，然后按其在存储器中的首地址执行程序的第一条指令，以后就按照该程序的规定顺序执行其他指令，直至程序执行结束。

计算机按照此原理应该具有 5 大功能：数据传送功能、数据存储功能、数据处理功能、操作控制功能、操作判断功能。

1.3 计算机的性能指标

1.3.1 计算机的主要性能指标

1. 机器字长

机器字长是指计算机进行一次整数运算(即定点整数运算)所能处理的二进制数据的位数,通常与 CPU 的寄存器位数、加法器有关。机器字长一般等于内部寄存器的大小,字长越长,数的表示范围越大,计算精度就越高。计算机字长通常都选定为字节(Byte, 8 位)的整数倍,通常是 2、4、8 倍。不同的计算机,字长可以不相同。

注意: 机器字长、指令字长和存储字长的关系(见本章疑难点 4)。

2. 数据通路带宽

数据通路带宽是指数据总线一次所能并行传送信息的位数。这里所说的数据通路宽度是指外部数据总线的宽度,它与 CPU 内部的数据总线宽度(内部寄存器的大小)有可能不同。

注意: 各个子系统通过数据总线连接形成的数据传送路径称为数据通路。

3. 主存容量

主存容量是指主存储器所能存储信息的最大容量,通常以字节来衡量,也可以用字数 \times 字长(如 512 K \times 16 位)来表示存储容量。其中, MAR 的位数反映了存储单元的个数, MAR 的位数反映了可寻址范围的最大值(而不一定是实际存储器的存储容量)。

例如, MAR 为 16 位,表示 $2^{16} = 65\ 536$,即此存储体内有 65 536 个存储单元(可称作 64 K 内存,1K=1 024),若 MDR 为 32 位,表示存储容量为 64 K \times 32 位。

4. 运算速度

(1) 吞吐量和响应时间

1) **吞吐量:** 指系统在单位时间内处理请求的数量。它取决于信息能多快地输入内存, CPU 能多快地读取指令,数据能多快地从内存取出或存入,以及所得结果能多快地从内存送给一台外部设备。这些步骤中的每一步都关系到主存,因此,系统吞吐量主要取决于主存的存取周期。

2) **响应时间:** 指从用户向计算机发送一个请求,到系统对该请求做出响应并获得它所需要的结果的等待时间。通常包括 CPU 时间(运行一个程序所花费的时间)与等待时间(用于磁盘访问、存储器访问、I/O 操作、操作系统开销等时间)。

(2) 主频和 CPU 时钟周期

1) **CPU 时钟周期:** 通常为节拍脉冲或 T 周期,即主频的倒数,它是 CPU 中最小的时间单位,每个动作至少需要一个时钟周期。

2) **主频(CPU 时钟频率):** 机器内部主时钟的频率,它是衡量机器速度的重要参数。主频的倒数是 CPU 时钟周期。对于同一个型号的计算机,其主频越高,完成指令的一个执行步骤所用的时间越短,执行指令的速度越快。

注意: CPU 时钟周期 = 1/主频,主频通常以 MHz(兆赫兹)为单位,1Hz 表示每秒 1 次。

(3)CPI(Clock cycle Per Instruction),即执行一条指令所需的时钟周期数。

(4)CPU 执行时间,指运行一个程序所花费的时间。

$$\text{CPU 执行时间} = \text{CPU 时钟周期数} / \text{主频} = (\text{指令条数} \times \text{CPI}) / \text{主频}$$

上式表明,CPU 的性能(CPU 执行时间)取决于三个要素:①主频(时钟频率);②每条指令执行所用的时钟周期数(CPI);③指令条数。

(5)MIPS 和 MFLOPS

1) **MIPS**(Million Instructions Per Second),即每秒执行多少百万条指令。

$$\text{MIPS} = \text{指令条数} / (\text{执行时间} \times 10^6) = \text{主频} / \text{CPI}$$

2) **MFLOPS**(Million Floating-point Operations Per Second),即每秒执行多少百万次浮点运算。 $\text{MFLOPS} = \text{浮点操作次数} / (\text{执行时间} \times 10^6)$ 。

1.3.2 几个专业术语的概念

1) **系列机**:具有基本相同的体系结构,使用相同的基本指令系统的多个不同型号的计算机组成的一个产品系列。

2) **兼容**:指计算机软件或硬件的通用性,即使用或运行在某个型号的计算机系统硬件、软件也能应用于另外一个型号的计算机系统时,则称这两台计算机在硬件或软件上存在兼容性。

3) **软件可移植性**:指把使用在某个系列计算机中的软件直接或进行很少修改就能运行在另外一个系列计算机中的可能性。

4) **固件**:将程序固定在 ROM 中组成的部件称为固件。固件是一种具有软件特性的硬件,固件的性能指标介于硬件与软件之间,吸收了软、硬件各自的优点,其执行速度快于软件,灵活性优于硬件,是软、硬件结合的产物。例如,目前操作系统已实现了部分固化(把软件永恒地存储于只读存储器中)。

1.3.3 本节习题精选

一、单项选择题

1.【2010 年计算机联考真题】下列选项中,能缩短程序执行时间的措施是()。

- I. 提高 CPU 时钟频率 II. 优化数据通路结构 III. 对程序进行编译优化
A. 仅 I 和 II B. 仅 I 和 III C. 仅 II 和 III D. I、II、III

2.【2011 年计算机联考真题】下列选项中,描述浮点数操作速度指标的是()。

- A. MIPS B. CPI C. IPC D. MFLOPS

3.关于 CPU 主频、CPI、MIPS、MFLOPS,说法正确的是()。

- A. CPU 主频是指 CPU 系统执行指令的频率,CPI 是执行一条指令平均使用的频率
B. CPI 是执行一条指令平均使用 CPU 时钟的个数,MIPS 描述一条 CPU 指令平均使用 CPU 时钟数

C. MIPS 是描述 CPU 执行指令的频率,MFLOPS 是计算机系统的浮点数指令

D. CPU 主频指 CPU 使用的时钟脉冲频率,CPI 是执行一条指令平均使用 CPU 时钟数

4.存储字长是指()。

- A. 存放在一个存储单元中的二进制代码组合

- B. 存放在一个存储单元中的二进制代码位数
 C. 存储单元的个数
 D. 机器指令的位数
5. 以下说法错误的是()。
- A. 计算机的机器字长是指数据存储与运算的基本单位
 B. 寄存器由触发器构成
 C. 计算机中一个字的长度都是 32 位
 D. 磁盘可以永久性存放数据和程序
6. 下列关于机器字长、指令字长和存储字长的说法中, 正确的是()。
- I. 三者在数值上总是相等的 II. 三者在数值上可能不等
 III. 存储字长是存放在一个存储单元中的二进制代码位数
 IV. 数据字长就是 MDR 的位数
- A. I、III B. I、IV C. II、III D. II、IV
7. 32 位微机是指该计算机所用 CPU()。
- A. 具有 32 位寄存器 B. 能同时处理 32 位的二进制数
 C. 具有 32 个寄存器 D. 能处理 32 个字符
8. 用于科学计算的计算机中, 标志系统性能的主要参数是()。
- A. 主时钟频率 B. 主存容量 C. MFLOPS D. MIPS
9. 若一台计算机的机器字长为 4 字节, 则表明该机器()。
- A. 能处理的数值最大为 4 位十进制数
 B. 能处理的数值最多为 4 位二进制数
 C. 在 CPU 中能够作为一个整体处理 32 位的二进制代码
 D. 在 CPU 中运算的结果最大为 2^{32}
10. 在 CPU 的寄存器中, () 对用户是完全透明的。
- A. 程序计数器 B. 指令寄存器 C. 状态寄存器 D. 通用寄存器
11. 计算机操作的最小单位时间是()。
- A. 时钟周期 B. 指令周期 C. CPU 周期 D. 中断周期
12. CPU 的 CPI 与下列哪个因素无关? ()
- A. 时钟频率 B. 系统结构 C. 指令集 D. 计算机组织
13. 从用户观点看, 评价计算机系统性能的综合参数是()。
- A. 指令系统 B. 吞吐率 C. 主存容量 D. 主频率
14. 当前设计高性能计算机的重要技术途径是()。
- A. 提高 CPU 主频 B. 扩大主存容量
 C. 采用非冯·诺依曼 D. 采用并行处理技术
15. 下列关于“兼容”的叙述, 正确的是()。
- A. 指计算机软件与硬件之间的通用性, 通常在同一系列不同型号的计算机间存在
 B. 指计算机软件或硬件的通用性, 即它们在任何计算机间可以通用
 C. 指计算机软件或硬件的通用性, 通常在同一系列不同型号的计算机间通用
 D. 指软件在不同系列计算机中可以通用, 而硬件不能通用

16. 下列说法正确的是()。

- I. 在微型计算机的广泛应用中, 会计电算化属于科学计算方面的应用
 - II. 决定计算机计算精度的主要技术是计算机的字长
 - III. 计算机“运算速度”指标的含义是每秒钟能执行多少条操作系统的命令
 - IV. 利用大规模集成电路技术把计算机的运算部件和控制部件做在一块集成电路芯片上, 这样的一块芯片叫单片机
- A. I、III B. II、IV C. II D. I、III、IV

二、综合应用题

1. 设主存储器容量为 64 K×32 位, 并且指令字长、存储字长、机器字长三者相等。写出如图 1-4 所示各寄存器的位数, 并指出哪些寄存器之间有信息通路。
2. 用一台 40 MHz 的处理器执行标准测试程序, 它所包含的混合指令数和响应所需的时钟周期见表 1-1。求有效的 CPI、MIPS 速率和程序的执行时间(I 为程序的指令条数)。
3. 微机 A 和 B 是采用不同主频的 CPU 芯片, 片内逻辑电路完全相同。

表 1-1 测试程序包含的混合指令数及响应所需的时钟周期

指令类型	CPI	指令混合比
算术和逻辑	1	60%
高速缓存命中的访存	2	18%
转移	4	12%
高速缓存失效的访存	8	10%

- 1) 若 A 机的 CPU 主频为 8MHz, B 机为 12MHz, 则 A 机的 CPU 时钟周期为多少?
- 2) 若 A 机的平均指令执行速度为 0.4MIPS, 那么 A 机的平均指令周期为多少?
- 3) B 机的平均指令执行速度为多少?

表 1-2 每条指令所占的比例及 CPI 数

指令类型	指令所占比例	CPI
算术逻辑指令	43%	1
Load 指令	21%	2
Store 指令	12%	2
转移指令	24%	2

4. 某台计算机只有 Load/Store 指令能对存储器进行读/写操作, 其他指令只对寄存器进行操作。根据程序跟踪试验结果, 已知每条指令所占的比例及 CPI 数, 见表 1-2。
- 求上述情况的平均 CPI。

假设程序由 M 条指令组成。算术逻辑运算中 25% 的指令的两个操作数中的一个已在寄存器中, 另一个必须在算术逻辑指令执行前用 Load 指令从存储器中取到寄存器中。因此有人建议增加另一种算术逻辑指令, 其特点是一个操作数取自寄存器, 另一个操作数取自存储器, 即寄存器—存储器类型, 假设这种指令的 CPI 等于 2。同时, 转移指令的 CPI 变为 3。求新指令系统的平均 CPI。

1.3.4 答案与解析

一、单项选择题

1. D

CPU 的时钟频率, 也即 CPU 主频, 一般来说, 一个时钟周期内完成的指令数是固定

的,所以主频越高,CPU的速度也就快,程序的执行时间就越短,I正确。数据在功能部件之间传送的路径称为数据通路,数据通路的功能是实现CPU内部的运算器和寄存器以及寄存器之间的数据交换。优化数据通路结构,可以有效提高计算机系统的吞吐量,从而加快程序的执行速度,II正确。计算机程序需要先转化成机器指令序列才能最终得到执行,通过对程序进行编译优化可以得到更优的指令序列,从而使得程序的执行时间也越短,III正确。

2. D

MFLOPS 即每秒执行多少百万条浮点数运算,用于描述浮点数操作速度指标。

3. D

4. B

存储体由许多存储单元组成,每个存储单元又包含若干个存储元件,每个存储元件能寄存一位二进制代码“0”或“1”。可见,一个存储单元可存储一串二进制代码,称这串二进制代码为一个**存储字**,这串二进制代码的位数称为**存储字长**。

5. C

计算机中一个字的长度可以是16、32、64位等,不一定是32位的。

6. C

机器字长、指令字长和存储字长,三者数值上可以相等也可以不相等,视不同机器而定。一个存储单元中的二进制代码的位数称为存储字长。存储字长等于MDR的位数,而数据字长是一次存取数据的长度,它可以不等于MDR的位数。

7. B

计算机的位数,即机器字长,也就是计算机一次能处理的二进制数的长度。此外应注意,操作系统的位数是操作系统可寻址的位数,与机器字长是不一样的。

8. C

MFLOPS 是指每秒执行多少百万次浮点运算,该参数用来描述计算机的浮点运算性能,而用于科学计算的计算机主要就是评估浮点运算的性能。

9. C

机器字长是计算机内部一次可以处理的二进制数的位数,故该计算机一次可处理 $4 \times 8 = 32$ 位的二进制代码。

10. B

汇编程序员可以通过指定待执行指令的地址来设置PC的值,状态寄存器、通用寄存器只有为汇编程序员可见,才能实现编程,而IR、MAR、MDR是CPU的内部工作寄存器,对程序员均不可见。

11. A

时钟周期即CPU频率的倒数,是最基本的时间单位,其余选项均大于时钟周期。另外,CPU周期又称机器周期,由多个时钟周期组成。

12. A

CPI是执行一条指令所需的时钟周期数,系统结构、指令集、计算机组织都会影响CPI,而时钟频率并不会影响到CPI,但可以加快指令的执行速度。如执行一条指令需要10个时钟周期,则一台主频为1GHz的CPU,执行这条指令要比一台主频为100MHz的

CPU 快。

13. B

主频、主存容量和指令系统(间接影响 CPI)并不是综合性能的体现。吞吐率指系统在单位时间内处理请求的数量,是评价计算机系统性能的综合参数。

14. D

提高 CPU 主频、扩大主存容量对性能的提升是有限度的。采用并行技术是实现高性能计算的重要途径,现今超级计算机均采用多处理器来增强并行处理能力。

15. C

兼容指计算机软件或硬件的通用性,故 A、D 错。B 中,它们在任何计算机间可以通用,错误。C 中,兼容通常在同一系列不同型号的计算机,正确。

16. C

会计电算化属于计算机数据处理方面的应用, I 错误。II 显然正确。计算机“运算速度”指标的含义是每秒钟能执行多少条指令, III 错误。这样集成的芯片称为 CPU, IV 错误。

二、综合应用题

1. 解答:

由主存容量为 $64\text{ K} \times 32\text{ 位}$, 因 $2^{16} = 64\text{ K}$, 则地址总线宽度为 16 位, 32 位表示数据总线宽度, 故 MAR 为 16 位, PC 为 16 位, MDR 为 32 位。

因指令字长 = 存储字长 = 机器字长

则 IR、ACC、MQ、X 均为 32 位

寄存器之间的信息通路有:

PC → MAR

Ad(IR) → MAR

MDR → IR

取数: MDR → ACC, 存数: ACC → MDR

MDR → X

2. 解答:

CPI 即执行一条指令所需的时钟周期数。本标准测试程序共包含 4 种指令, 那么 CPI 就是这 4 种指令的数学期望。即

$$\text{CPI} = 1 \times 60\% + 2 \times 18\% + 4 \times 12\% + 8 \times 10\% = 2.24$$

MIPS 即每秒执行百万条指令数。已知处理器时钟频率为 40 MHz, 即每秒包含 40 M 个时钟周期, 故

$$\text{MIPS} = 40 / \text{CPI} = 40 / 2.24 = 17.9$$

程序的执行时间 $T = \text{CPI} \times T_{\text{IC}} \times I$, 其中 T_{IC} 是一个 CPU 时钟的时间长度, 是 CPU 时钟频率 f 的倒数。故

$$T = \text{CPI} \times T_{\text{IC}} \times I = \text{CPI} \times (1/f) \times I = 5.6 \times 10^{-8} \times I \text{ 秒}$$

3. 解答:

1) A 机的 CPU 主频为 8 MHz, 所以 A 机的 CPU 时钟周期 = $1/8\text{ MHz} = 0.125\text{ }\mu\text{s}$ 。

2) A 机的平均指令周期 = $1/0.4\text{ MIPS} = 2.5\text{ }\mu\text{s}$ 。

3) A 机平均每条指令的时钟周期数 = $2.5\text{ }\mu\text{s} / 0.125\text{ }\mu\text{s} = 20$ 。

因微机 A 和 B 片内逻辑电路完全相同，所以 B 机平均每条指令的时钟周期数也为 20。由于 B 机的 CPU 主频为 12 MHz，所以 B 机的 CPU 时钟周期 = 1/12 MHz = 1/12 μs。B 机的平均指令周期 = 20 × (1/12) = 5/3 μs。
B 机的平均指令执行速度 = 1/(5/3) μs = 0.6 MIPS。
另解：B 机的平均指令执行速度 = A 机的平均指令执行速度 × (12/8) = 0.4 MIPS × (12/8) = 0.6 MIPS。

4. 解答：

①本处理机共包含 4 种指令，那么 CPI 就是这 4 种指令的数学期望。即：

$$CPI = 1 \times 43\% + 2 \times 21\% + 2 \times 12\% + 2 \times 24\% = 1.57$$

②设原指令总数为 M，则新指令总数为

$$M - 0.43M \times 0.25 = 0.8925M$$

增加另一种算术逻辑指令后，每种指令所占的比例及 CPI 数，见下表：

指令类型	指令所占比例	CPI
算术逻辑指令	$(0.43M - 0.43M \times 0.25) / 0.8925M = 0.3613$	1
算术逻辑指令(新)	$(0.43M \times 0.25) / 0.8925M = 0.1204$	2
Load 指令	$(0.21M - 0.43M \times 0.25) / 0.8925M = 0.1148$	2
Store 指令	$0.12M / 0.8925M = 0.1348$	2
转移指令	$0.24M / 0.8925M = 0.2689$	3

所以：CPI' = 1 × 0.3613 + 2 × 0.1204 + 2 × 0.1148 + 2 × 0.1348 + 3 × 0.2689 = 1.9076。

1.4 常见问题和易混淆知识点

1. 同一个功能既可以由软件实现也可以由硬件实现吗？

软件和硬件是两种完全不同的形态，硬件是实体，是物质基础；软件是一种信息，看不见、摸不到。但是在逻辑功能上，软件和硬件是等效的。因此，在计算机系统中，许多功能既可以由硬件直接实现，也可以在硬件的配合下由软件实现。

例如，乘法运算既可以用专门的乘法器(主要由加法器和移位器组成)实现，也可以用乘法子程序(主要由加法指令和移位指令等组成)来实现。

2. 翻译程序、汇编程序、编译程序、解释程序的区别和联系？

翻译程序是指把高级语言源程序翻译成机器语言程序(目标代码)的软件。

翻译程序有两种：一种是**编译程序**，它将高级语言源程序一次全部翻译成目标程序，每次执行程序时，只要执行目标程序，因此，只要源程序不变，就无须重新翻译；另一种是**解释程序**，它将源程序的一条语句翻译成对应的机器目标代码，并立即执行，然后翻译下一条源程序语句并执行，直至所有源程序语句全部被翻译并执行完。所以解释程序的执行过程是翻译一句执行一句，并且不会生成目标程序。

汇编程序也是一种语言翻译程序，它把汇编语言源程序翻译为机器语言程序。汇编语

言是一种面向机器的低级语言，是机器语言的符号表示，与机器语言一一对应。

编译程序与汇编程序的区别：如果源语言是诸如 C、C++、Java 等“高级语言”，而目标语言是诸如汇编语言或机器语言之类的“低级语言”，这样的—个翻译程序称为编译程序。如果源语言是汇编语言，而目标语言是机器语言，这样的—个翻译程序称为汇编程序。

3. 什么叫透明性？透明是指什么都能看见吗？

在计算机领域中，站在某一类用户的角度，如果感觉不到某个事物或属性的存在，即“看”不到某个事物或属性，则称为“对该用户而言，某个事物或属性是透明的”。这与日常生活中的“透明”概念（公开、看得见）正好相反。

例如，对于高级语言程序员来说，浮点数格式、乘法指令等这些指令的格式、数据如何在运算器中运算等都是透明的；而对于机器语言或汇编语言程序员来说，指令的格式、机器结构、数据格式等则不是透明的。

在 CPU 中，IR、MAR 和 MDR 对各类程序员都是透明的。

4. 机器字长、指令字长、存储字长的区别和联系？

机器字长：计算机能直接处理的二进制数据的位数，机器字长一般等于内部寄存器的大小，它决定了计算机的运算精度。

指令字长：一个指令字中包含二进制代码的位数。

存储字长：一个存储单元存储二进制代码的长度。它们都必须是字节的整数倍。

指令字长一般都取存储字长的整数倍，如果指令字长等于存储字长的 2 倍，就需要 2 次访存取出一条指令，因此，取指周期为机器周期的 2 倍，如果指令字长等于存储字长，则取指周期等于机器周期。

早期的计算机存储字长一般和机器的指令字长与数据字长相等，故访问一次主存便可以取出一条指令或一个数据。随着计算机的发展，指令字长可变，数据字长也可变，但它们必须都是字节的整数倍。

5. 计算机体系结构和计算机组成的区别和联系？

计算机体系结构是指机器语言或汇编语言程序员所看得到的传统机器的属性，包括指令集、数据类型、存储器寻址技术等，大都属于抽象的属性。

计算机组成是指如何实现计算机体系结构所体现的属性，它包含许多对程序员来说是透明的硬件细节。例如，指令系统是属于结构的问题，但指令的实现，即如何取指令、分析指令、取操作数、如何运算等都属于组成的问题。因此，当两台机器指令系统相同时，只能认为它们具有相同的结构，至于这两台机器如何实现其指令，完全可以不同，即可以认为它们的组成方式是不同的。例如，一台机器是否具备乘法指令，是一个结构的问题，但实现乘法指令采用什么方式，则是一个组成的问题。

许多计算机厂商提供一系列体系结构相同的计算机，而它们的组成却有相当大的差别，即使是同一系列的不同型号机器，其性能和价格差异很大。例如，IBM System/370 结构就包含了多种价位和性能的机型。

第 2 章 数据的表示和运算

【考纲内容】

(一)数制与编码

进位计数制及其相互转换；真值和机器数

BCD 码；字符与字符串；校验码

(二)定点数的表示和运算

1. 定点数的表示

无符号数的表示；有符号数的表示

2. 定点数的运算

定点数的移位运算；原码定点数的加/减运算；补码定点数的加/减运算

定点数的乘/除运算；溢出概念和判别方法

(三)浮点数的表示和运算

1. 浮点数的表示

IEEE754 标准

2. 浮点数的加/减运算

(四)算术逻辑单元 ALU

1. 串行加法器和并行加法器

2. 算术逻辑单元 ALU 的功能和结构

【考题分布】

年份	单选题/分	综合题/分	考查内容
2009 年	2 题 × 2	0	C 语言中隐式类型转换；浮点加法运算
2010 年	2 题 × 2	0	定点数的运算及溢出判断；不同类型数的机内表示法及其强制转换
2011 年	1 题 × 2	1 题 × 11	IEEE754 浮点数的表示；C 语言中 unsigned int 和 int 的表示、及其类型转换原理、补码加法运算的实现原理及溢出判断
2012 年	3 题 × 2	0	C 语言中 unsigned int 和 unsigned short 的转换；IEEE 754 单精度浮点数的范围；字符串的小端存储、C 语言中 int、char 和 short 的表示

本章内容较为繁杂，由于计算机中数的表示和运算方法与人们日常生活中的表示和运算是不同的，因此理解也较为困难。综观最近几年的真题，不难发现，unsigned、short、int、long、float、double 等在 C 语言中的表示、运算、溢出判断、隐式类型转换、强制类型转换，IEEE754 浮点数的表示，以及浮点数的运算都是联考考查的重点，需要牢固掌握。

2.1 数制与编码

2.1.1 进位计数制及其相互转换

1. 进位计数法

进位计数法是一种计数的方法。常用的进位计数法有十进制数、二进制数、十六进制数、八进制数等。十进制数是人们在日常生活中的最常用数，而在计算机中通常用二进制数、八进制数和十六进制数。

在进位计数法中，每个数位所用到的不同数码的个数称为**基数**。十进制的基数为10(0~9)，每个数位计满10就向高位进位，即“逢十进一”。

十进制数101，其个位的1显然与百位的1所表示的数值是不同的。每个数码所表示的数值等于该数码本身乘以一个与它所在数位有关的常数，这个常数称为**位权**。一个进位数的数值大小就是它的各位数码按权相加。

一个r进制数($K_n K_{n-1} \cdots K_0 K_{-1} \cdots K_{-m}$)的数值可表示为

$$K_n r^n + K_{n-1} r^{n-1} + \cdots + K_0 r^0 + K_{-1} r^{-1} + \cdots + K_{-m} r^{-m} = \sum_{i=n}^{-m} K_i r^i$$

式中，r是基数； r^i 是第i位的位权(整数位最低位规定为第0位)； K_i 的取值可以是0, 1, ..., r-1共r个数码中的任意一个。

(1) 二进制数

计算机中用得最多的是基数为2的计数制，即二进制。二进制只有0和1两种数字符号，计数“逢二进一”。它的任意数位的权为 2^i ，i为所在位数。

(2) 八进制数

八进制作为二进制的一种书写形式，其基数为8，有0~7共8个不同的数字符号，计数“逢八进一”。因为 $r=8=2^3$ ，所以只要把二进制中的3位数码编为一组就是一位八进制数码，两者之间的转换极为方便。

(3) 十六进制数

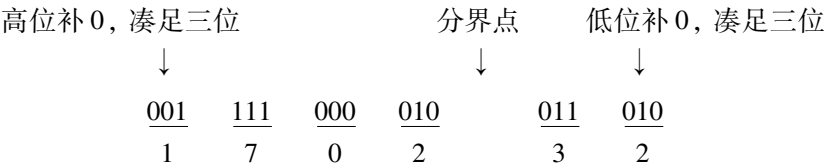
十六进制也是二进制的一种常用的书写形式，其基数为16，“逢十六进一”。每个数位可取0~9、A、B、C、D、E、F中的任意一个，其中A、B、C、D、E、F分别表示10~15。因为 $r=16=2^4$ ，故4位二进制数码与1位十六进制数码相对应。

2. 不同进制数之间的相互转换

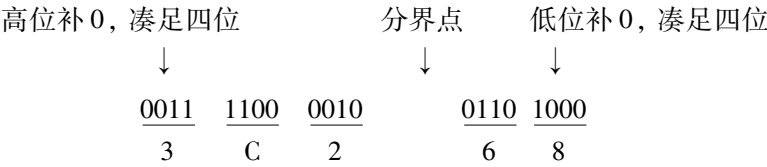
(1) 二进制转换为八进制和十六进制

对于一个二进制混合数(既包含整数部分，又包含小数部分)，在转换时应以小数点为界。其整数部分，从小数点开始往左数，将一串二进制数分为3位(八进制)一组或4位(十六进制)一组，在数的最左边可根据需要加“0”补齐；对于小数部分，从小数点开始往右数，也将一串二进制数分为3位一组或4位一组，在数的最右边也可根据需要加“0”补齐。最终使总的位数成为3或4的整数倍，然后分别用对应的八进制或十六进制数取代。

【例2-1】 将二进制数1111000010.01101分别转换为八进制数和十六进制数。



所以，对应的八进制数为 $(1702.32)_8 = (1111000010.01101)_2$



所以，对应的十六进制数为 $(3C2.68)_{16} = (1111000010.01101)_2$

同样，由八进制或十六进制转换成二进制，只需将每一位改为3位或者4位二进制数即可(必要时去掉整数最高位或者小数最低位的0)。八进制和十六进制之间的转换也能方便地实现，十六进制转换为八进制(或八进制转换为十六进制)时，先将十六进制(八进制)转换为二进制，然后由二进制转换为八进制(十六进制)较为方便。

(2)任意进制转换为十进制

将任意进制的数各位数码与它们的权值相乘，再把乘积相加，就得到了一个十进制数。这种方法称为按权展开相加法。例如，

$$(11011.1)_2 = 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} = 27.5$$

(3)十进制转换为任意进制

一个十进制数转换为任意进制数，常采用基数乘法。这种转换方法对十进制数的整数部分和小数部分将分别进行处理，对于整数部分用除基取余法；对于小数部分用乘基取整法，最后将整数部分与小数部分的转换结果拼接起来。

除基取余法(整数部分的转换)：整数部分除基取余，最先取得的余数为数的最低位，最后取得的余数为数的最高位(即除基取余，先余为低，后余为高)，商为0时结束。

【例 2-2】 将十进制数 123.6875 转换成二进制数。

整数部分：



故整数部分 $123 = (1111011)_2$

乘基取整法(小数部分的转换)：小数部分乘基取整，最先取得的整数为数的最高位，最后取得的整数为数的最低位(即乘基取整，先整为高，后整为低)，乘积为0(或满足精度要求)时结束。

小数部分：

乘基取整		
0.6875		
×	2	
1.3750		
	0.3750	1 最高位
×	2	
0.7500		
	0	0
×	2	
1.5000		
	0.5000	1
×	2	
1.0000		
	1	最低位

故小数部分 $0.6875 = (0.1011)_2$ 。
所以， $123.6875 = (1111011.1011)_2$ 。

注意：关于十进制转换为任意进制为何采用除基取余法和乘基取整法，以及取的数放置位置的原理。请结合第 1 节中 r 进制数的数值表示的公式思考，而不应是死记硬背。

2.1.2 真值和机器数

在日常生活中，通常用正、负号来分别表示正数（正号可省略）、负数，如 +15、-8 等。这种带“+”或“-”符号的数称为真值。真值是机器数所代表的实际值。

在计算机中，通常采用数的符号和数值一起编码的方法来表示数据。常用的有原码、补码和反码表示法。这几种表示法都将数据的符号数字化，通常用“0”表示“正”，用“1”表示“负”。如 0, 101（这里的逗号“，”实际上并不存在，仅为区分符号位与数值位）表示 +5。这种把符号“数字化”的数称为机器数。

2.1.3 BCD 码

二进制编码的十进制数(Binary-Coded Decimal, BCD)通常采用 4 位二进制数来表示一位十进制数中的 0~9 这 10 个数码。这种编码方法使二进制和十进制之间的转换得以快速进行。但 4 位二进制数可以组合出 16 种代码，故必有 6 种状态为冗余状态。

下面列举几种常用的 BCD 码。

1)8421 码(最常用)：它是一种有权码，设其各位的数值为 $b_3、b_2、b_1、b_0$ ，则权值从高到低依次为 8、4、2、1，则它表示的十进制数为 $D = 8b_3 + 4b_2 + 2b_1 + 1b_0$ 。如 $8 \rightarrow 1000; 9 \rightarrow 1001$ 。

如果两个 8421 码相加之和小于或等于 $(1001)_2$ ，即 $(9)_{10}$ ，则不需要修正；如果相加之和大于或等于 $(1010)_2$ ，即 $(10)_{10}$ ，则要加 6 修正(从 1010 到 1111 这 6 个为无效码，当运算结果落在这个区间时，需要将运算结果加上 6)，并向高位进位，进位可以在首次相加或修正时产生。

1+8=9	4+9=13	9+7=16
0001	0100	1001
+ 1000	+ 1001	+ 0111
1001	1101	10000
不需要修正	+ 0110 修正进位	+ 0110 修正进位
	10011	10110

2)余 3 码：它是一种无权码，是在 8421 码的基础上加上 $(0011)_2$ 形成的，因每个数都多余“3”，故称为余 3 码。如 $8 \rightarrow 1011; 9 \rightarrow 1100$ 。

3)2421 码：它也是一种有权码，权值由高到低分别为 2、4、2、1，特点是大于等于 5 的 4 位二进制数中最高位为 1，小于 5 的最高位为 0。如 5→1011 而不是 0101。

2.1.4 字符与字符串

由于计算机内部只能识别和处理二进制代码，所以字符都必须按照一定的规则用一组二进制编码来表示。

1. 字符编码 ASCII 码

目前，国际上普遍采用的一种字符系统是 7 位二进制编码的 ASCII 码（读音“阿斯科”），可表示 10 个十进制数码、52 个英文大写字母和小写字母（A ~ Z，a ~ z）和一定数量的专用符号（如 \$、%、+、= 等），共 128 个字符。

在 ASCII 码中，编码值 0 ~ 31 为控制字符，用于通信控制或设备的功能控制；编码值 127 是 DEL 码；编码值 32 是空格 SP；编码值 32 ~ 126 共 95 个字符称为可印刷字符。

提示：0 ~ 9 的 ASCII 码值为 48(011 0000) ~ 57(011 1001)，即去掉高 3 位，只保留低 4 位，则正好是二进制形式的 0 ~ 9。

2. 汉字的表示和编码

在 1981 年的国家标准 GB 2312 - 80 中，每个编码用两个字节表示，收录了一级 3 755 个、二级 3 008 个汉字，各种符号 682 个，共计 7 445 个。

目前最新的汉字编码是 2000 年公布的国家标准 GB 18030，收录了 27 484 个汉字。编码标准采用 1B、2B 和 4B。

汉字的编码包括汉字的输入编码、汉字内码、汉字字形码三种，它们是计算机中用于输入、内部处理和输出三种用途的编码。区位码是国家标准局于 1981 年颁布的标准，用两个字节表示一个汉字，每个字节用七位码，它将汉字和图形符号排列在一个 94 行 94 列的二位代码表中。区位码是 4 位十进制数，前 2 位是区码，后 2 位是位码，所以称为区位码。

而国标码则是将十进制的区位码转换为十六进制数后，再在每个字节上加上 20H。国标码两个字节的最高位都是 0，ASCII 码的最高位也是 0。为了方便计算机区分中文字符和英文字符，将国标码两个字节的最高位都改为“1”，这就是汉字内码了。

区位码和国标码都是输入码，其和汉字内码的关系（十六进制）：

$$\text{国标码} = (\text{区位码})_{16} + 2020\text{H}$$

$$\text{汉字内码} = (\text{国标码})_{16} + 8080\text{H}$$

3. 字符串的存放

字符串是指连续的一串字符，通常方式下，它们占用主存中连续的多个字节，每个字节存储一个字符。主存字由 2 或 4 个字节组成时，在同一个主存字中，既可按先存储低位字节、后存储高位字节的顺序（即从低位字节向高位字节顺序）存放字符串的内容（又称小端模式），也可按从先存储高位字节、后存储低位字节的顺序（即从高位字节向低位字节顺序）存放字符串的内容（又称大端模式）。

这两种存放方式都是常用方式，不同计算机可以选用其中任何一种（甚至是同时采

I	F	空	A
>	B	空	T
H	E	N	空
R	E	A	D
(C)	空

图 2 - 1 字符串的存放（大端模式）

用)。例如，字符串：IF_A > B_THEN_READ(C)_，其从高位字节到低位字节依次存放在主存中，如图 2-1 所示。

其中主存单元长度(字长度)由 4 个字节组成。每个字节中存放相应字符的 ASCII 值，注意空格“_”也占一个字节的位置。因此，每个字节分别存放十进制的 73、70、32、65、62、66、32、84、72、69、78、32、82、69、65、68、40、67、41、32。

2.1.5 校验码

校验码是指能够发现或能够自动纠正错误的数据编码，也称为检错纠错编码。校验码的原理是通过增加一些冗余码，来检验或纠错编码。

通常某种编码都由许多码字构成，任意两个合法码字之间最少变化的二进制位数，称为数据校验码的码距。对于码距不小于 2 的数据校验码，开始具有检错的能力。码距越大，检、纠错能力就越强，而且检错能力总是大于或等于纠错能力。

下面介绍 3 种常用的校验码。

1. 奇偶校验码

在原编码上加一个校验位，它的码距等于 2，可以检测出一位错误(或奇数位错误)，但不能确定出错的位置，也不能够检测出偶数位错误，增加的冗余位称为奇偶校验位，如图 2-2 所示。

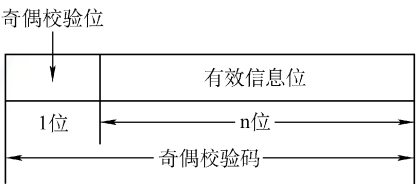


图 2-2 奇偶校验码的格式

奇偶校验实现的方法：由若干位有效信息(如一个字节)，再加上一个二进制位(校验位)组成校验码，如图 2-2 所示。校验位的取值(0 或 1)将使整个校验码中“1”的个数为奇数或偶数，所以有两种可供选择的校验规律。

奇校验码：整个校验码(有效信息位和校验位)中“1”的个数为奇数。

偶校验码：整个校验码(有效信息位和校验位)中“1”的个数为偶数。

【例 2-3】 给出两个编码 1001101 和 1010111 的奇校验码和偶校验码。

设最高位为校验位，余 7 位是信息位，则对应的奇偶校验码为：

1001101 11001101(奇校验) 01001101(偶校验)

1010111 01010111(奇校验) 11010111(偶校验)

缺点：具有局限性，奇偶校验只能发现数据代码中奇数位出错情况，但不能纠正错误，常用于对存储器数据的检查或者传输数据的检查。

2. 海明(汉明)校验码

海明码是广泛采用的一种有效的校验码，它实际上是一种多重奇偶校验码。其实现原理是在有效信息位中加入几个校验位形成海明码，并把海明码的每一个二进制位分配到几个奇偶校验组中。当某一位出错后，就会引起有关的几个校验位的值发生变化，这不但可以发现错位，还能指出错位的位置，为自动纠错提供了依据。

根据纠错理论得

$$L - 1 = D + C \quad \text{且 } D \geq C$$

即编码最小码距 L 越大，则其检测错误的位数 D 越大，纠正错误的位数 C 也越大，且纠错能力恒小于或等于检错能力。海明码就是根据这一理论提出的具有纠错能力的一种

编码。

求海明码的步骤如下。

【例 2-4】 在 $n=4$ 、 $k=3$ 时，求 1010 的海明码？

(1) 确定海明码的位数

设 n 为有效信息的位数， k 为校验位的位数，则信息位 n 和校验位 k 应满足：

$$n + k \leq 2^k - 1 \quad (\text{若要检测两位错，则需再增加 1 位校验位，即 } k + 1 \text{ 位})$$

海明码位数为 $n + k = 7 \leq 2^3 - 1$ 成立，则 n 、 k 有效。设信息位 $D_4D_3D_2D_1$ (1010)，共 4 位，校验位 $P_3P_2P_1$ ，共 3 位，对应的海明码为 $H_7H_6H_5H_4H_3H_2H_1$ 。

(2) 确定校验位的分布

规定校验位 P_i 在海明位号为 2^{i-1} 的位置上，其余各位为信息位，因此：

P_1 的海明位号为 $2^{1-1} = 2^0 = 1$ ，即 H_1 为 P_1 。

P_2 的海明位号为 $2^{2-1} = 2^1 = 2$ ，即 H_2 为 P_2 。

P_3 的海明位号为 $2^{3-1} = 2^2 = 4$ ，即 H_4 为 P_3 。

将信息位按原来的顺序插入，则海明码各位的分布如下。

$$\begin{array}{ccccccc} H_7 & H_6 & H_5 & H_4 & H_3 & H_2 & H_1 \\ D_4 & D_3 & D_2 & P_3 & D_1 & P_2 & P_1 \end{array}$$

(3) 分组，以形成校验关系

每个数据位用多个校验位进行校验，但要满足：被校验数据位的海明位号等于校验该数据位的各校验位海明位号之和。另外，校验位不需要再被校验。

	$P_1(H_1)$	$P_2(H_2)$	$P_3(H_4)$
D ₁ 放在 H ₃ 上, 由 P ₂ P ₁ 校验:	3= 1	2	
D ₂ 放在 H ₅ 上, 由 P ₃ P ₁ 校验:	5= 1		4
D ₃ 放在 H ₆ 上, 由 P ₃ P ₂ 校验:	6=	2	4
D ₄ 放在 H ₇ 上, 由 P ₃ P ₂ P ₁ 校验:	7= 1	2	4
	第 1 组	第 2 组	第 3 组

(4) 校验位取值

校验位 P_i 的值为第 i 组 (由该校验位校验的数据位) 所有位求异或。

根据 3) 中的分组有：

$$P_1 = D_1 \oplus D_2 \oplus D_4 = 0 \oplus 1 \oplus 1 = 0$$

$$P_2 = D_1 \oplus D_3 \oplus D_4 = 0 \oplus 0 \oplus 1 = 1$$

$$P_3 = D_2 \oplus D_3 \oplus D_4 = 1 \oplus 0 \oplus 1 = 0$$

所以，1010 对应的海明码为 1010010 (下画线为校验位，其他为信息位)

(5) 海明码的校验原理

每个校验组分别利用校验位和参与形成该校验位的信息位进行奇偶校验检查，就构成了 k 个校验方程：

$$S_1 = P_1 \oplus D_1 \oplus D_2 \oplus D_4$$

$$S_2 = P_2 \oplus D_1 \oplus D_3 \oplus D_4$$

$$S_3 = P_3 \oplus D_2 \oplus D_3 \oplus D_4$$

若 $S_3S_2S_1$ 的值为“000”，则说明无错；否则说明出错，而且这个数就是错误位的位号，如 $S_3S_2S_1 = 001$ ，说明第 1 位出错，即 H_1 出错，直接将该位取反就达到了纠错的目的。

3. 循环冗余校验(CRC)码

CRC 的基本思想是：在 K 位信息码后再拼接 R 位的校验码，整个编码的长度为 N 位，因此，这种编码又称为 (N, K) 码，如图 2-3 所示。



图 2-3 循环冗余校验码的格式

CRC 码基于线性编码理论，在发送端，将要传送的 K 位二进制信息码左移 R 位，将它与生成多项式 $G(x)$ 做模 2 除法，生成一个 R 位校验码，并附在信息码后，构成一个新的二进制码(CRC 码)，共 $(K + R)$ 位。在接收端，则利用生成多项式对接收到的编码做模 2 除法，以检测和确定出错的位置，如无错则整除，其中生成多项式是接收端和发送端的一个约定。

任意一个二进制数码都可以用一个系数仅为“0”或“1”的多项式与其对应。生成多项式 $G(x)$ 的最高幂次为 R ，转换成对应的二进制数有 $R + 1$ 位。例如，生成多项式 $x^3 + x^2 + 1$ 对应的二进制数为 1101，而二进制数 1011 对应的多项式为 $x^3 + x^1 + 1$ 。下面用一个例子来介绍 CRC 的编码和检测过程。

【例 2-5】 设生成多项式为 $G(x) = x^3 + x^2 + 1$ ，信息码为 101001，求对应的 CRC 码。
 $R =$ 生成多项式最高幂次 $= 3$ ， $K =$ 信息码长度 $= 6$ ， $N = K + R = 9$ 。

生成多项式 $G(x)$ 对应的二进制码为 1101。

(1) 移位

将原信息码左移 R 位，低位补 0。
得到 101001000。

(2) 相除

对移位后的信息码，用生成多项式进行模 2 除法，产生余数。

模 2 除法：模 2 加法和减法的结果相同，都是做异或运算，模 2 除法和算术除法类似，但每一位除(减)的结果不影响其他位，也就是不借位，步骤如下。过程如图 2-4 所示。

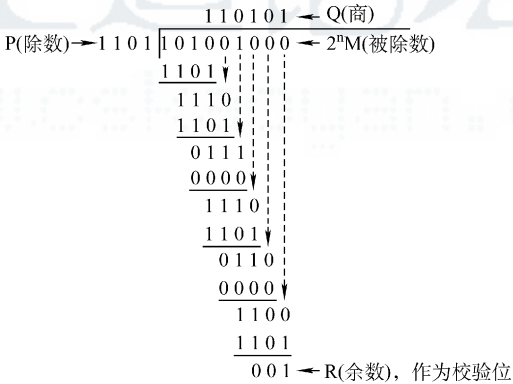


图 2-4 CRC 码的生成过程(模 2 取余)

①用除数对被除数最高几位做模 2 减(异或)，不借位。

②除数右移一位，若余数最高位为1，商为1，并对余数做模2减。若余数最高位为0，商为0，除数继续右移一位。

③循环直到余数位数小于除数时，该余数为最终余数。

模2除法过程如图2-4所示，得到的余数为001，则报文101001编码后的报文(即CRC码)为101001001(下划线为校验位)。

(3) 检错和纠错

接收端收到的CRC码，用生成多项式 $G(x)$ 做模2除法，若余数为0，则码字无错。

若接收端收的CRC码为 $C_9C_8C_7C_6C_5C_4C_3C_2C_1 = 101001011$ ，将这个数据与1101进行模2除法，得到的余数为010，则说明 C_2 出错，将 C_2 取反即可。

2.1.6 本节习题精选

一、单项选择题

- 下列各种数制的数中，最小的数是()。
A. $(101001)_2$ B. $(101001)_{BCD}$ C. $(52)_8$ D. $(233)_{16}$
- 两个数7E5H和4D3H相加，得()。
A. BD8H B. CD8H C. CB8H D. CC8H
- 若十进制数为137.5，则其八进制数为()。
A. 89.8 B. 211.4 C. 211.5 D. 1011111.101
- 四位机器内的数值代码，则它所表示的十进制真值可能为()。
I. 16 II. -1 III. -8 IV. 8
A. I、II、III B. II、IV C. II、III、IV D. 只有IV
- 一个16位无符号二进制数的表示范围是()。
A. 0 ~ 5 536 B. 0 ~ 66 535
C. -32 768 ~ 32 767 D. -32 768 ~ 32 768
- 下列说法有误的是()。
A. 任何二进制整数都可以用十进制表示
B. 任何二进制小数都可以用十进制表示
C. 任何十进制整数都可以用二进制表示
D. 任何十进制小数都可以用二进制表示
- 下列编码中，()不是合法的8421码。
A. 0111 1001 B. 0000 0001 C. 1010 0101 D. 0001 1001
- 已知计算机中用8421码表示十进制数，A和B的编码表示分别为0011 1000和0010 0011，则A+B的结果为()。
A. 0101 1011 B. 0110 0001 C. 0110 1011 D. 0101 0001
- 下列关于ASCII编码，正确的描述是()。
A. 使用8位二进制代码，最右边一位为1
B. 使用8位二进制代码，最左边一位为0
C. 使用8位二进制代码，最右边一位是0
D. 使用8位二进制代码，最左边一位是1

10. 在一个按字节编址的计算机中,若数据在存储器中以小端方案存放。假定 int 行变量 i 的地址为 08000000H, i 的机器数为 01234567H, 地址 08000000H 单元的内容是()。

- A. 01H B. 23H C. 45H D. 67H

11. 以下关于校验码的叙述中,正确的是()。

- I. 校验码的码距必须大于 2
II. 校验码的码距越大,检、纠错能力越强
III. 增加奇偶校验位的位数,可以提高奇偶校验的正确性
IV. 采用奇偶校验可检测出一位数据错误的位置并加以纠正
V. 采用海明校验可检测出一位数据错误的位置并加以纠正
VI. CRC 码是通过除法运算来建立数据和校验位之间的约定关系的
A. I、III、V B. II、IV、VI C. I、V、VI D. II、V、VI

12. 设在网络中传送采用偶校验的 ASCII 码,当收到的数据位为 10101001 时,可以断定()。

- A. 未出错 B. 出现偶数位错
C. 未出错或出现偶数位错 D. 出现奇数位错

13. 下列校验码中,奇校验正确的有()。

- A. 110100111 B. 001000111 C. 010110011 D. 110100111

14. 用 1 位奇偶校验能检测出 1 位主存错误的百分比为()。

- A. 0% B. 100% C. 50% D. 无法计算

15. 已知大写英文字母“A”的 ASCII 码值为 41H, 现字母“F”被存放在某个存储单元中,若采用偶校验(假设最高位作为校验位),则该存储单元中存放的十六进制数是()。

- A. 46H B. C6H C. 47H D. C7H

16. 用海明码来发现并纠正 1 位错,信息位为 8 位,则检验位的位数为()。

- A. 1 B. 3 C. 4 D. 8

17. 能发现两位错误并能纠正 1 位错的编码是()。

- A. CRC 码 B. 海明码 C. 偶校验码 D. 奇校验码

18. 在 CRC 中,接收端检测出某一位数据错误后,纠正的方法是()。

- A. 请求重发 B. 删除数据
C. 通过余数值自行纠正 D. 以上均可

19. 在大量数据传送过程中,常用且有效的检验法是()。

- A. 海明码校验 B. 偶校验 C. 奇校验 D. CRC

20. 设待校验的数据为 D8 ~ D1 = 10101011, 若采用海明校验,其海明码为()(设海明码具有一位纠错能力, P13 采用全校验);若采用 CRC, 且生成多项式为 10011, 则其 CRC 码为()。

- A. 0 1010 0 101 1 1 11 B. 0 1000 0 111 1 1 11
C. 10101011 1010 D. 10101010 1011