

计算机组成

[第2版]

王健 朱怡健 朱敏 主编

原理

COMPUTER ORGNIZATION

东南大学出版社

计算机组成原理

(第2版)

王 健 朱怡健 朱 敏 主编

东南大学出版社

·南京·

内 容 提 要

本书比较全面地论述了计算机的基本工作原理、内部组织结构及各部件的设计技术和互连方式等。全书共分9章,具体内容主要有:计算机功能部件及其组织,机器内部的数据表示,指令系统及寻址方式设计,运算器、存储器的组织结构及设计技术,CPU功能及控制器的实现方式,总线结构,常用外部设备的工作原理,输入输出组织等。本书内容完整、丰富,既对计算机组成的基本原理、方法有较详尽的介绍,又反映了当代计算机组成发展的最新技术。

本书可以作为计算机专业本科或专科的计算机组成原理课的教科书,也可作为工科类研究生及相关工程技术人员参考书。

图书在版编目(CIP)数据

计算机组成原理/王健主编. —2版. —南京:东南大学出版社,2004.1

ISBN 7-81089-458-7

I. 计... II. 王... 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(2003)第 117522 号

计算机组成原理(第2版)

出版发行 东南大学出版社
出 版 人 宋增民
社 址 南京市四牌楼2号(邮编:210096)
电 话 (025)83794844 (025)83362442(传真)
印 刷 南京玉河印刷厂
开 本 787mm×1092mm 1/16
印 张 19
字 数 474 千
版 次 2004年2月第2版 2004年2月第1次印刷
定 价 25.00 元

* 未经本社授权,本图书内任何文字图片不得以任何方式转载、演绎,违者必究。

* 东大版图书若有印装质量问题,请直接向发行科调换,电话:025-83792327。

再 版 前 言

《计算机组成原理》是计算机专业的一门主干课程,在计算机硬件课程系列中起着承上启下的作用。

本书在第1版(1994年)的基础上,根据近十年来计算机发展的现状以及编者多年来的教学实践的积累,在内容安排上作了较大的调整。例如,在中央处理器一章增加了现代计算机中普遍采用的流水线技术、超标量技术等,增加了一章专门介绍总线,以 Pentium 和 PowerPC 代替原先的 PDP-11 作为典型样机等。本书延续了第1版的写作风格,强调整机概念,站在计算机设计者的角度,从系统程序员所能见到的硬件特性入手,由外向内,从整体到部分,详尽介绍了计算机各个功能部件的组成原理、设计思想及其在整机中所起的作用。由于计算机技术发展很快,从微机到超级计算机,新技术层出不穷。限于篇幅及教科书的特点,本书主要着重对基本概念、基本原理以及基本技术的阐述,适当汲取目前流行的一些新技术,特别是微型计算机上普遍采用的新技术。

本书内容全面,系统性强,基本概念清楚,深入浅出,注重对学生的计算机基本理论和技能的培养。每一章都附有适量习题,便于读者对本章内容理解及应用。以下为本书导读:

第1章为本书的导论,从整体上介绍了计算机的结构和组成以及工作原理,简述了计算机的发展历程,并对计算机的性能指标评价及设计作了简单介绍。

第2章主要介绍了数据在计算机中的表示方法,着重比较了带符号数在机器中的几种表示方法的特点,另外还介绍了浮点数、十进制数、字符、多媒体信息的编码表示,最后对计算机中常用的几种校验码作了简单讨论。

第3章对计算机的主要执行部件——运算器的组成与工作原理作了详尽的介绍,着重讨论了实现算术运算的各种算法及其逻辑实现。

第4章介绍了计算机的存储系统及其设计,重点阐述了计算机系统的主存构成 SRAM、DRAM 的特征及设计与实现方法,讨论了存储系统的分层结构和主要性能设计技术,如多模块交叉存取、Cache 和虚拟存储器的原理、实现方法等。另外对计算机系统的重要辅存如磁盘等性能指标、工作原理也作了较为详细的介绍。

第5章介绍了指令系统的基本概念以及在计算机系统中的地位 and 作用,并给出了指令系统的完整设计过程和方法,特别强调了 RISC、CISC 两种风格指令系统的区别。

第6章介绍了 CPU 的结构组成以及功能实现的方法,给出了控制器的时序部件的构成和微操作信号产生部件的两种实现技术,并对现代处理器中普遍采用的一些提高性能的主要技术作了简明的阐述。

第7章介绍了总线及其互连方式的基本概念,总线控制的方法,以及总线设计技术等,并简要叙述了现代计算机中普遍采用的一些典型总线。

第8章介绍了计算机系统中的外围设备的分类及特点,对常用的一些输入输出设备的工作原理及其实现技术作了简明扼要的叙述。

第9章主要讨论了输入输出系统中的数据传送控制方法,如直接程序控制,中断控制,DMA控制,通道和I/O处理机控制的原理及实现,并对I/O接口的类型、功能、编址方式等内容也有清楚的描述。

本书由东南大学计算机系王健主编,系统结构教研室杨全胜、徐造林、任国林、朱怡健、朱敏、吴强老师参加了编写。所有参加本书编写的老师都担任过《计算机组成原理》课程或相关课程、实验课的教学工作,并从事过该门课程的教学研究、改革实践以及实验仪器的研制,积累了丰富的经验,并取得了多项教学成果。但限于编者水平及能力,书中仍难免存在一些不足之处,恳请广大读者批评指正。

编 者

2003年11月

目 录

1 概述

1.1 计算机的功能和结构	(1)
1.2 计算机的发展	(3)
1.3 计算机系统组成及工作过程	(6)
1.4 计算机系统性能指标及性能设计	(8)
习题 1	(11)

2 数据的表示

2.1 数字化信息编码	(12)
2.2 数值数据的编码表示	(13)
2.3 非数值数据的编码表示	(29)
2.4 数据校验码	(33)
习题 2	(38)

3 运算器与运算方法

3.1 运算器的组成与功能	(40)
3.2 加法器及定点加减法运算	(43)
3.3 定点乘法运算及实现	(51)
3.4 定点除法运算及实现	(62)
3.5 浮点运算	(69)
习题 3	(76)

4 存储系统

4.1 存储器概述	(79)
4.2 半导体随机存储器	(83)
4.3 半导体只读存储器	(96)
4.4 多体交叉存储器	(99)
4.5 高速缓冲存储器	(101)
4.6 虚拟存储器	(109)
4.7 辅助存储器	(113)
习题 4	(123)

5 指令系统

5.1 指令系统概述	(125)
5.2 操作数类型及存储方式	(130)
5.3 指令系统功能设计	(135)
5.4 寻址方式	(140)
5.5 指令字格式设计	(146)

5.6 指令系统举例	(151)
5.7 指令系统的发展	(158)
习题 5	(159)
6 中央处理器	
6.1 CPU 的结构和功能	(162)
6.2 控制器的功能和组成	(167)
6.3 组合逻辑与 PLA 控制	(173)
6.4 微程序控制	(177)
6.5 流水线处理器	(186)
6.6 超标量处理器	(195)
6.7 Pentium 和 PowerPC	(200)
习题 6	(203)
7 总线及总线互连结构	
7.1 总线的基本概念	(207)
7.2 总线控制	(210)
7.3 总线结构	(221)
7.4 总线接口	(225)
7.5 总线标准	(227)
7.6 总线设计需注意的问题	(238)
习题 7	(240)
8 外围设备	
8.1 外围设备的分类与特点	(243)
8.2 输入设备	(244)
8.3 显示器	(247)
8.4 打印机	(254)
习题 8	(260)
9 输入输出系统	
9.1 I/O 接口	(261)
9.2 I/O 数据传送控制方式	(267)
9.3 程序中断方式	(269)
9.4 直接存储器存取(DMA)方式	(280)
9.5 通道和 I/O 处理机方式	(285)
9.6 外部接口	(291)
习题 9	(296)
参考文献	(298)

1 概 述

1.1 计算机的功能和结构

1.1.1 什么是计算机

计算机是一种快速电子计算机器,它接收输入的数字信息,按照内部存储的指令序列去处理,并将产生的结果输出。指令是指示计算机硬件要完成某种功能的明确的命令,指令序列称为计算机程序,内部存储装置称为计算机主存储器,简称主存或内存。

目前计算机的种类非常多,主要体现在大小、速度和价格 3 个方面的不同。人们日常所见到的绝大多数计算机都是个人计算机,简称 PC,它被广泛地应用于家庭、学校和办公室等场所。桌面(台式)计算机的主机、显示器、键盘和鼠标等很容易安置在家庭和办公室的桌面上。手提笔记本计算机最适宜在旅途中进行文字处理工作,它将处理器、显示器和键盘等集成在一个单独的装置中。高性能的图形工作站尽管和台式机的大小差不多,但它的计算能力通常比普通的台式机高出很多,主要用于工程计算特别是交互式设计领域。

在工作站之后,还有更高性能的大型计算机和超级计算机。大型计算机主要用于需要更高处理能力的中等规模和大规模的商业数据处理和存储方面。超级计算机主要用于大规模数值计算,例如天气预报和飞机设计及模拟。在大型计算机和超级计算机中,主要的功能部件经常由分离的和较大的部件构成。

另外,在许多工业控制领域、家用电器和移动设备中,大量使用单片计算机。

尽管各种计算机的规模、性能和价格相差很大,但它们的组成和工作原理都是类似的。

1.1.2 计算机的功能

计算机的基本功能主要包括以下 4 个:

- 数据处理
- 数据存储
- 数据传送
- 控制

计算机的数据处理功能包括算术运算和逻辑运算,应用于数值计算和非数值计算两个方面。早期的计算机主要用于数值计算,而现代计算机更多的应用于非数值计算方面。计算机领域所指的数据其含义是广泛的,它包含计算机所能处理的各种对象,如数值、字符、图形、图像、声音和视频等。

计算机的数据存储功能是必需的。在计算机进行数据处理时,其指令和数据都必须保存在它的主存储器(内存)中。另外,计算机还必须有长期保存数据的能力,大量数据以文件的形式保存在计算机的辅助存储器(外存)中以便检索和更新。

计算机的数据传送功能体现在两个方面。一是在计算机系统内部各个部件之间的数据流动,如 CPU 和主存以及 CPU 内部寄存器与运算器之间的数据流动;二是计算机与外界之间的数据传送,如输入/输出(I/O)和计算机通信。

上述三种功能必须是受控的,否则计算机无法按照人们的预期来工作。因此在计算机内部需要有专门的部件来完成控制功能。这种控制功能主要包括通过各种时序信号定时,产生各种基本操作信号并按某种时序发出以完成相应功能,以及控制计算机的启动和停止等。

计算机的所有功能都是通过指令实现的。指令是计算机硬件能直接识别并执行的命令,不同种类的计算机通常有着不同的指令编码,一台计算机的所有指令集合称之为指令系统。若干条指令组织成一个指令序列(程序),用来完成某项任务。程序通常都存放在存储器中,除了特殊情况外,计算机总是在存储程序的控制下工作。

大部分指令都需要处理一些源数据并产生某些结果数据,这些数据在计算机中是广义的,包括数值数据,非数值数据,甚至指令或程序在某些情况下都是数据。当然,所有数据都必须进行二进制编码。

1.1.3 计算机功能部件

计算机由 5 个功能独立的部件组成:存储器、运算器、输入部件、输出部件和控制部件(如图 1.1 所示)。输入设备接受外部的二进制编码信息并送入存储器中存储或立即送往运算器中运算,处理步骤由存储在存储器中的程序确定,最终结果通过输出设备送往外部世界,所有这些动作都由控制部件协调。图 1.1 中并未画出这些部件的连接,有关内容将在后面的章节中详细说明。

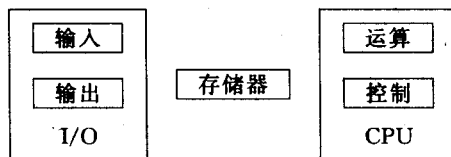


图 1.1 计算机功能部件

输入部件的基本功能是接受外部经过二进制编码的数据信息并送入计算机内部的存储器或寄存器。目前最典型的输入设备是键盘和鼠标。另外,用于图形和图像输入的数字化仪、扫描仪、摄像头,用于声音输入的话筒,工业控制中的 A/D 转换器等都是常见的输入设备。

存储器的功能是存储程序和数据。计算机的存储器分两类:主存储器(简称主存或内存)和辅助存储器(简称辅存或外存)。主存速度快,但价格高、容量小,主要用来存放正在或将要执行的程序以及相关数据。主存主要由半导体器件构成。辅存速度慢,但价格低、容量大,主要用来存放需要长期保存的程序和数据。辅存主要由磁记录设备、光设备构成,如磁盘、磁带和光盘等。

大多数计算机操作都由算术和逻辑部件(ALU)完成。ALU 的输入来自于主存和 CPU 中的寄存器,输出也到主存和寄存器。使用寄存器是因为其速度要比主存快得多。ALU、寄存器和相关线路构成了计算机的运算器。

输出部件的功能是将计算机处理的结果送往外界。和输入部件相反,输出部件常常要把二进制数值数据转换成字符、图形、图像、声音和其他一些模拟信号。典型的输出部件是显示器和打印机。另外,用于图形和图像输出的绘图仪,用于声音输出的音箱,工业控制中的 D/A 转换器等都是常见的输出设备。

存储器、运算器和输入输出设备存储和处理信息并完成输入和输出工作。所有这些操作都必须以某种方式协调起来,控制部件就是完成该项任务的。控制部件是计算机的神经中心,它发送各种控制信号到计算机的其余部件并监视它们的状态。控制部件的另一个重要任务就是定时和同步,它通过产生各种时序信号来保证各部件能够有序的工作。

由于运算器和控制器常常被做在同一个芯片中,因此它们被统称为中央处理器(机)即 CPU。CPU 和主存通常都放在同一个机箱中,人们把它们统称为主机。

1.1.4 计算机结构、组织和实现

计算机结构(Computer Architecture)和计算机组织(Computer Organization)(本书称为计算机组成)是描述计算机系统的两个重要概念。计算机结构是指那些对程序员(机器语言或汇编语言)可见的系统属性,这些属性直接影响到程序的逻辑执行。计算机组织是指实现其结构规范的操作部件以及它们的互连方式。

计算机结构主要研究指令集、寄存器定义和组织、各种数据类型表示的位数、I/O 机制,以及内存访问技术等,是计算机硬件子系统的概念结构及功能特性。计算机组织主要研究那些对程序员透明的硬件细节,如控制信号、数据通路宽度的确定、各部件连接方式及匹配、计算机和外设的接口以及使用的存储器技术等。计算机结构决定系统中的软硬件分配,计算机组成是对计算机结构的逻辑实现。

例如,在计算机指令系统中是否包含浮点运算指令是结构设计的问题。而浮点运算指令是通过现成的定点运算部件实现还是用专门的浮点单元实现则是组织的问题。通常一个计算机制造商提供的系列产品都具有相同的结构和不同的组织,从而使不同型号的机器具有不同的性能和价格。

计算机实现(Computer Implementation)是指计算机组成的物理实现。例如,CPU、主存的物理结构,芯片的集成度、速度、工艺、电源、冷却、装配、外观设计等都属于实现的范畴。

本书主要介绍计算机组织方面的内容,着重讨论有关计算机组织中各部件的组成技术、互连方式和工作原理。当然,随着计算机技术的不断发展,不仅计算机的组成技术更新很快,对计算机结构的影响也很大。组织、结构和实现的关系越来越密切,三者之间的界限变得越来越模糊,特别在微型计算机中表现的就更加明显,因此本书也会涉及一些有关计算机体系结构方面的内容。

1.2 计算机的发展

1.2.1 第一代计算机:电子管,ENIAC,冯·诺依曼机

1946 年,美国宾夕法尼亚大学为军队的弹道研究实验室设计和制造了 ENIAC (Elec-

tronic Numerical Integrator and Computer, 电子数字积分器和计算机), 它是世界上第一台通用电子数字计算机。

ENIAC 体积庞大, 重 30 吨, 占地 15000 平方英尺, 采用了 18000 多个电子管, 功耗约为 140kW。但它比当时的机械计算机要快得多, 每秒钟能做 5000 次加法。ENIAC 采用十进制进行数据表示和计算。ENIAC 的主要缺点是存储量太小, 至多只能存放 20 个字长为 10 位的十进制数; 另外, 它采用“外插式”编程, 即通过设置开关和插拔电缆来实现程序的编制和输入, 使用极不方便。

为克服 ENIAC 的不足, 应用数学家冯·诺依曼提出了计算机设计和工作的新思想。冯·诺依曼思想的主要内容有:

- (1) 计算机应由控制器、计算器、存储器和输入/输出五大功能部件构成。
- (2) 以运算器为中心, 采用二进制进行数据的表示和计算。
- (3) 计算机指令应采用二进制编码并和数据存储在同一记忆装置中, 即存储程序。每条指令被顺序从存储器中取出执行, 是一个指令驱动的串行工作方式。
- (4) 存储器是线性编址的一维结构, 每个存储单元位数固定。存储器中存放的指令和数据从其本身是无法区分的, 它们都是以二进制编码表示的。
- (5) 指令由操作码和地址码组成。操作码指定操作类型和操作数的数据类型, 地址码指明操作数所在存储器或寄存器的位置。

冯·诺依曼的核心思想是存储程序, 它不仅解决了速度匹配问题, 还使得程序的逻辑选择和数值计算一样快, 从而使计算机成为真正能够自动工作的机器。直到今天, 世界上绝大多数计算机仍然采用冯·诺依曼的结构, 所以它们都被称为冯·诺依曼机。

第一代计算机尚无软件的概念, 没有操作系统, 没有高级语言。因此, 只有专业人员才能使用计算机。

1.2.2 第二代计算机: 晶体管

晶体管是由硅材料制成的固态器件, 它与电子管相比, 体积小、功耗低、价格便宜。晶体管的使用是第二代计算机的标志。

第二代计算机除了逻辑元件采用晶体管之外, 存储器件也有了改进。计算机的内存采用磁芯存储器, 外存采用磁鼓和磁带。

计算机软件也有了很大的发展, 出现了多种高级语言, 如 FORTRAN、COBOL 等。计算机在很多领域特别是商业领域有了较为广泛的应用。

1.2.3 第三代计算机: 集成电路, 微电子技术

在第二代计算机发展过程中, 随着晶体管数目越来越多, 整个制造过程越来越复杂, 成本也越来越高。这给计算机工业的发展带来了很多麻烦。

从 1958 年开始, 微电子技术的诞生给计算机工业带来了革命性的机遇。从根本上说, 计算机组织由门电路、存储元件以及它们之间的互连结构组成。电子管、晶体管都是分立元件, 它们通过引脚连到印刷电路板上, 使得计算机硬件的制作过程非常复杂。集成电路将大量晶体管、电阻、导线等都安装在一块晶片上, 然后封装它并加上引脚形成集成块。许多这

样的集成块再安装在一个印刷电路板上形成更为复杂的电路。

根据集成块所能容纳的晶体管数目的多少,集成电路的发展经历了小规模集成电路(SSI),中规模集成电路(MSI),大规模集成电路(LSI),很大规模集成电路(VLSI),超大规模集成电路(ULSI)和巨大规模集成电路(GSI)几个时期。集成度的提高使得计算机硬件成本不断下降,体积更小,速度更快,更可靠。

第三代计算机采用的是中小规模集成电路,其逻辑元件和存储器均由集成电路实现。这一时期还有一个重要特点,采用多处理器并行结构的大型/巨型机和价廉物美的小型机同时发展。并行技术、流水线技术、数据库技术等新型技术的提出和应用对计算机的发展影响巨大。

1.2.4 第四代计算机:VLSI,微处理器

从20世纪70年代开始伴随着LSI和VLSI的迅猛发展,计算机工业的发展也有了长足的进步。集成电路芯片的集成度已从70年代初的每片几千个晶体管到现在的千万个晶体管。价格越来越低,速度也越来越快,这使得计算机已渗透到当今社会的各个角落。

第四代计算机除了主要器件采用VLSI这一显著特点之外,微处理器技术的出现和发展也是计算机发展史上里程碑式的事件。从1971年Intel的4位微处理器4004开始,到目前的32位和64位Pentium和PowerPC,微处理器技术使得计算机最终走向办公室和家庭。

由于军事、商业、空间技术等大规模和高强度计算应用的迫切需要,以并行处理计算为基础的超级计算机也得到了很大发展。目前超级计算机的主流产品大多采用大量的高性能微处理器组成大规模并行处理系统,其峰值速度已达到每秒几十万亿次。

从20世纪90年代开始,计算机网络技术特别是Internet技术的迅猛发展和普及,是计算机发展史上的又一次划时代的革命。全世界的计算机都可以通过因特网连在一起,使信息流通更加方便、快捷。电子商务、电子政务、协同办公、远程教育等新型工作方式开始步入我们的生活中。

随着后PC时代的到来,消费电子、计算机、通信(简称3C)一体化趋势日趋明显。数字化社会的呼声使嵌入式系统(Embedded System)日益受到市场和厂家的关注,嵌入式设备越来越普及,这也促进了普适计算(Pervasive Computing)的发展。

1.2.5 新一代计算机:新材料,非冯·诺依曼机

第四代计算机以后对计算机的不同时代划分的标准意见不一致。通常对新一代计算机(或称五代机)的划分并不是以器件的集成度而划分的。由于半导体器件的集成度受分子、原子的直径限制以及散热问题难以解决,人们正在搜寻速度更快、集成度更高的材料来生产元器件,如生物芯片、分子器件、超导材料、量子芯片等。

冯·诺依曼结构的计算机是指令驱动的串行执行的计算机,对现实世界中的大量并行性质的处理具有先天的不足。多年来,人们也在不断探索非冯·诺依曼结构的新型计算机,如数据驱动、并行推理等智能化计算机。

1.3 计算机系统组成及工作过程

1.3.1 计算机的软硬件系统

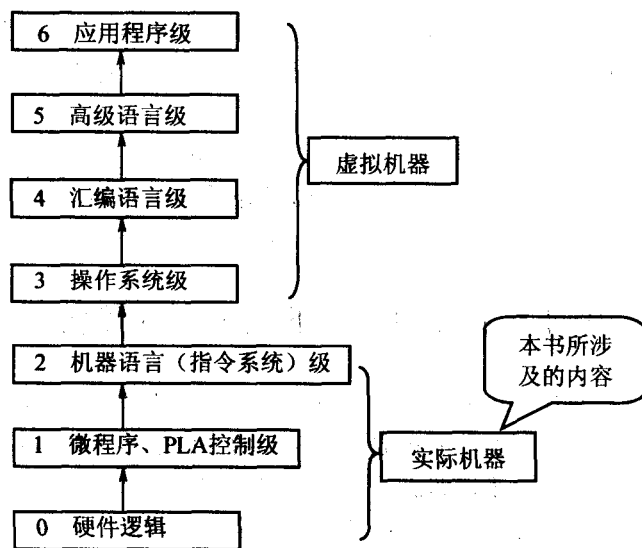
计算机系统包含硬件子系统和软件子系统。硬件是指组成计算机的各个部件及其连接的物理实体。软件是指程序以及相关的文件。

不同的人所关心的计算机硬件是不一样的。最终用户(End User)所看见的计算机硬件只是主机、显示器、键盘、鼠标、打印机等,而对于计算机内部的结构和组成、工作原理等不必关心;专业人员(Professional)更关心机器内部的结构和组成、工作方式、性能等,如 CPU、内存、总线、Cache 等。计算机设计者(Designer)主要关心计算机组成原理和实现方法,如芯片设计、寄存器设计、部件设计等。

计算机软件通常分为系统软件和应用软件。系统软件包括操作系统、语言处理程序(编译、解释程序等)、数据库管理系统(DBMS)、网络协议软件等。有时又将 DBMS、网络协议软件以及一些工具软件称为支撑软件。应用软件是由用户或专业软件公司专门开发的某个特定领域的应用程序。

硬件是计算机系统的物质基础,软件是计算机系统的应用基础。好比“躯体”和“灵魂”,两者相辅相成,缺一不可。在技术发展的今天,计算机的硬件和软件正朝着相互渗透、相互融合的方向发展。原来硬件实现的功能,可以由软件来实现。反之,原来由软件实现的功能,也可以由硬件实现。对于程序员来说,硬件和软件在逻辑上是等价的。软硬件功能划分应从整个系统的效率、速度、价格、资源利用等诸多方面综合考虑。

图 1.2 是一个计算机系统按功能划分的多层次结构。



1.3.2 计算机中各功能部件的互连结构——总线

计算机是由一组相互之间需要通信的部件或模块(CPU、存储器、I/O)组成的。将这些模块之间连接起来的所有数据通路被称为互连结构(Interconnection Structure)。

总线是计算机中普遍采用的互连结构,其主要特征是共享传输介质。在计算机系统中有多种总线,它们在各个层次提供部件间的数据通路,如 CPU 内部总线、系统总线和局部总线等。

总线通常包括数据总线、地址总线和控制总线。总线的根数各不相同,每根线能够传送一位二进制数。

数据总线是系统中各模块间传递数据的通道,典型的数据总线包含 8,16,32,64 根线。线的根数称为数据总线的宽度,它是决定系统总体性能的关键因素,反映了处理器的数据吞吐量。

地址总线用于指明数据传送的源地址和目的地址。地址总线的宽度决定了系统能够拥有的最大主存空间,但并非每台机器一定要配备到最大空间。例如,Pentium 常用 32 根地址线,其最大主存空间可达到 4G,但实际上采用 Pentium 处理器的计算机目前主存容量的典型配置是 128M 到 512M。除了对主存的访问,地址线一般也用作对 I/O 端口的访问。

控制线用于控制数据传送的方式、方向以及定时或应答等。如存储器读写、I/O 读写、总线和中断的请求和应答、复位等。控制线的数目取决于总线的类型及具体的机器配置。

总线的结构有单总线、双总线、三总线等。过去在微型计算机中几乎无例外地都采用单总线结构,这主要是为了简化结构、提高灵活性和降低成本。但随着 I/O 设备的数量不断增加,性能不断提高,使得总线的传输延迟加大,总线竞争加剧,总线成为系统的瓶颈。因此,现代计算机中多采用分层次多总线结构。

1.3.3 计算机的工作过程

计算机的工作过程就是执行程序的过程。程序由一组指令构成,在执行前调入主存。CPU 按照指令存放的顺序(转移指令除外)一条条取出并执行,直到程序被执行完。因此,要了解计算机的工作过程,就必须清楚指令的执行过程。

图 1.3 给出了一台假想计算机中 CPU 内部组成以及和主存的连接。我们考虑一条加法指令的执行过程:

ADD NUM, R₀

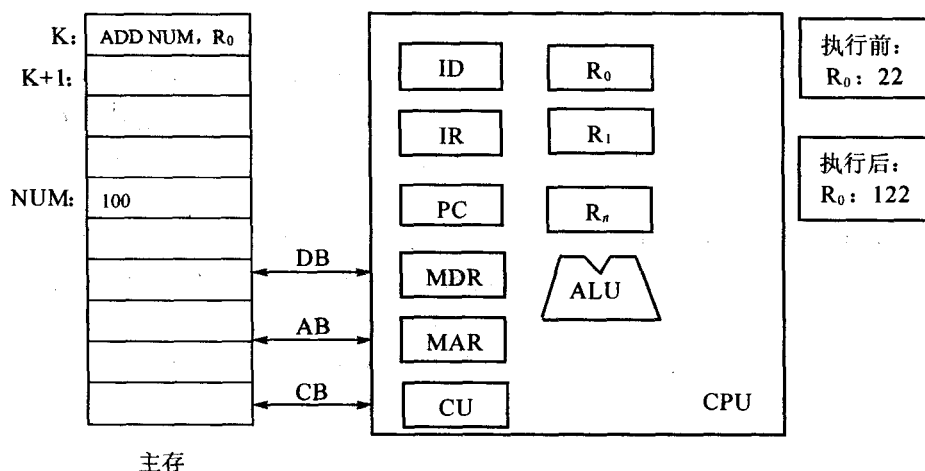
这条指令的功能是将主存单元 NUM 中的数 100 和 CPU 中寄存器 R₀ 的内容 22 相加,结果再放入 R₀ 中。假设指令占据一个存储字,存放在主存的 K 单元中。

指令的执行过程如下:

(1) CPU 将 PC 的内容 K 送入 MAR, MAR 通过地址总线 AB 将 K 送往存储器, CU 通过控制总线 CB 发送读信号 Read 到存储器。PC 的内容加 1 变为 K+1。

(2) 存储器对接收到的地址进行译码,找到 K 单元,将 ADD 指令取出并通过数据总线 DB 送往 CPU 的 MDR。

- (3) MDR 将其存放的指令码送往 IR, IR 将指令送往 ID 译码。
- (4) CPU 将 NUM 送入 MAR, MAR 通过地址总线 AB 将 NUM 送往存储器, CU 通过控制总线 CB 发送读信号 Read 到存储器。
- (5) 存储器对接收到的地址进行译码, 找到 NUM 单元, 将 100 取出并通过数据总线 DB 送往 CPU 的 MDR。
- (6) MDR 将其存放的 100 送往 ALU 的一个输入端。
- (7) R_0 将其存放的 22 送往 ALU 的另一输入端, CU 给 ALU 一个加法信号, ALU 执行加法运算。
- (8) 结果由 ALU 送往 R_0 , 本条指令执行完毕。
- (1)、(2)、(3)步为取指令周期, 对于相同的 CPU, 取指令过程是完全相同的。(4)到(8)步为执行指令周期, 不同的指令其执行过程是不同的。一条指令执行完, CPU 将根据 PC 的内容(K+1)取出下一条指令执行, 直到整个程序执行完。



ID: 指令译码器; IR: 指令寄存器; PC: 程序计数器
MDR: 存储器数据寄存器; MAR: 存储器地址寄存器
 $R_0 \dots R_n$: 寄存器组; CU: 控制单元

图 1.3 CPU 组成以及和存储器的连接

1.4 计算机系统性能指标及性能设计

1.4.1 计算机系统性能指标

1) 字长

CPU 字长是指 CPU 一次所能处理的二进制位数。字长是计算机系统最重要的指标之一, 字长越长, 表明 CPU 所能处理的数据精度越高, 处理速度越快。目前典型微机的字长为 32 位和 64 位。在某些机器中还有可变字长, 如半字、全字和双字等。CPU 字长越长, 价格也越高。一般情况下, CPU 的寄存器宽度和数据总线的宽度等于字长。

2) 主频

CPU 工作的节拍是由计算机的主时钟所控制。主时钟不断产生固定频率的时钟脉冲, 此固定频率就是 CPU 的主频。主频越高, CPU 工作速度就越快。目前微处理器如 Intel Pentium4 的主频已达到几个 G。

3) 速度

对计算机运行速度的衡量有多种方法, 比较常用的有 MIPS (Million Instructions Per Second) 和 MFLOPS (Million Floating - Point Operations Per Second)。

对于一个给定的程序, MIPS 定义为

$$\text{MIPS} = \frac{\text{指令条数}}{\text{执行时间} \times 10^6}$$

指令条数一般指加、减运算等简短指令。

对于一个给定的程序, MFLOPS 定义为

$$\text{MFLOPS} = \frac{\text{浮点操作次数}}{\text{执行时间} \times 10^6}$$

4) 主存容量

主存是和 CPU 直接交换数据的存储器。主存容量越大, 和速度较慢的辅存之间的信息交换次数就越少, 机器的处理能力也就越强。最大主存容量由 CPU 的地址总线的根数决定。如目前典型微处理器的地址线数目为 32 根, 则最大主存空间是 4G。

5) RASIS 特性

RASIS 特性是指可靠性 (Reliability)、可用性 (Availability)、可维护性 (Serviceability)、完整性 (Integrity) 和安全性 (Security)。其中平均无故障时间 MTBF (可靠性) 和平均修复时间 MTTR (可维护性) 是衡量计算机功能特性的常用指标。

6) 兼容性

兼容性 (Compatibility) 一般是指同一个软件不加修改就可以在三台计算机上运行。兼容性有向上兼容 (同类型但不同档次的机器之间) 和向前兼容 (同系列但不同时期的机器之间) 两种。兼容性可以保护用户的软件投资不受或少受损失。

1.4.2 计算机系统性能设计

随着计算机硬件技术的飞速发展, 今天人们只要花很少的钱就可以买到和过去大型机性能相当的机器。例如, 微机已可以用于诸如图像处理、语音识别、视频会议等复杂且性能要求较高的应用场合。

1) 处理器速度

按照著名的摩尔定律, 微处理器的性能每 3 年要提高 4~5 倍。要能更好地配合芯片的发展速度, 充分利用处理器硬件潜在的性能, 计算机设计者必须研究更加复杂、更加先进的技术。当代处理器普遍采用的一些先进技术主要有:

- 流水线技术
- 分支预测 (Branch Prediction)
- 数据流分析 (Data Flow Analysis)

● 推测执行(Speculative Execution)

这些技术的实现使得处理器的功能更加强大,并且能够充分利用处理器的速度潜能。

2) 性能平衡

由于计算机系统除了处理器以外,还有其他许多功能部件。当别的部件速度跟不上处理器的速度时,就会造成计算机性能失衡。因此,有必要对计算机的结构和组成做适当的调整,以使各功能部件之间的能力匹配。

处理器和主存之间的接口是整个计算机中最关键的数据通路,如果此通路的速度或存储器的速度和处理器的速度相差太大,就会严重影响处理器的效率。在当代计算机设计中有很多技术来解决这个问题,如高速缓存、分层次高速多总线技术等。

同样,I/O 设备也存在着和处理器速度差异过大的问题。解决的办法有接口缓冲、暂存机制和高速总线,在分布式系统和多机并行系统中采用 I/O 处理机。

3) 性能评价

计算机执行应用程序所需的总时间是衡量计算机性能最重要的指标。因为应用程序通常用高级语言来写,所以性能受编译程序将高级语言程序翻译成机器语言所采用的方法的影响。性能还受机器语言指令集(指令系统)的设计和计算机组成硬件设计的影响。要使计算机的性能达到最优,必须综合地考虑这三者间的平衡设计。本书主要讨论机器语言指令集和硬件的设计。

程序由一组排序的指令构成,每条指令在处理器中执行,都将会细分为若干小控制步(详见第 6 章)。每一个控制步在控制器的控制下在一个处理器时钟节拍内完成。假定一个高级语言程序通过编译程序生成一个可执行的机器语言程序,要完成该程序需要执行 N 个机器指令。 N 并不等同于经过编译产生的机器语言程序中的机器指令数,因为某些指令将会被重复执行,例如循环,还有些指令可能一次也不会被执行,如条件转移。假定每条指令要执行的控制步的数量平均为 S ,处理器主频为 R 。则程序执行的总时间 T 为

$$T = \frac{(N \times S)}{R}$$

T 相对于用户来讲比 R 更重要。构建一个好的机器指令集和编译器对计算机的性能影响巨大。一个高质量的编译器可使一个给定的应用程序的 N 数尽可能少。通常带有简单指令的指令集要比带有复杂指令的指令集的 S 数要少。但复杂指令又可使 N 数降低。

上述性能指标虽能在一定程度上反映计算机的性能,但并不全面。因为这些指标大多反映的是处理器、主存的性能,但计算机系统还有大量的外部设备。另外机器运行过程中,除了应用程序外,系统程序在空间和时间上也有不小的开销。因此,通常采用不同层次的基准测试程序(benchmark)对计算机系统进行比较客观的性能评价。

目前主要有三种测试程序。

- (1) 综合基准程序:人为编制的为体现平均执行的一段测试程序。
- (2) 核心程序:从典型的实际应用程序中抽取的一些关键循环程序段。
- (3) 实际应用程序:如各种 C 的编译程序。

当前性能评价技术主要有两种方法。

- (1) 测量技术:通过某种测量设备和测试程序直接从系统中获得评价数据。
- (2) 模型技术:对被评价的系统建立某种模型,在模型上获得评价数据。