

计算机组成与体系结构

主讲: 张骏鹏, 赵庆行

西安电子科技大学

人工智能学院

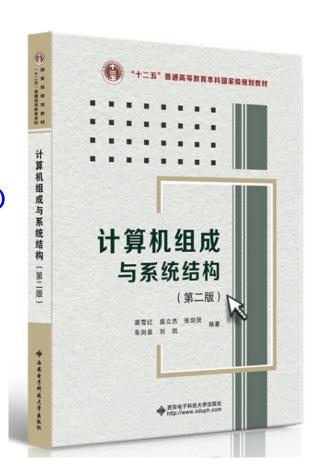
课程要求



- 总成绩=期末成绩×70%+期中成绩
 - ×10%+实验成绩×10%+平时成绩
 - $\times 10\%$
- 教材: 计算机组成与系统结构 (第二版)

作者:裘雪红; 出版社:西安电子科技大

学出版社,出版时间:2020年08月





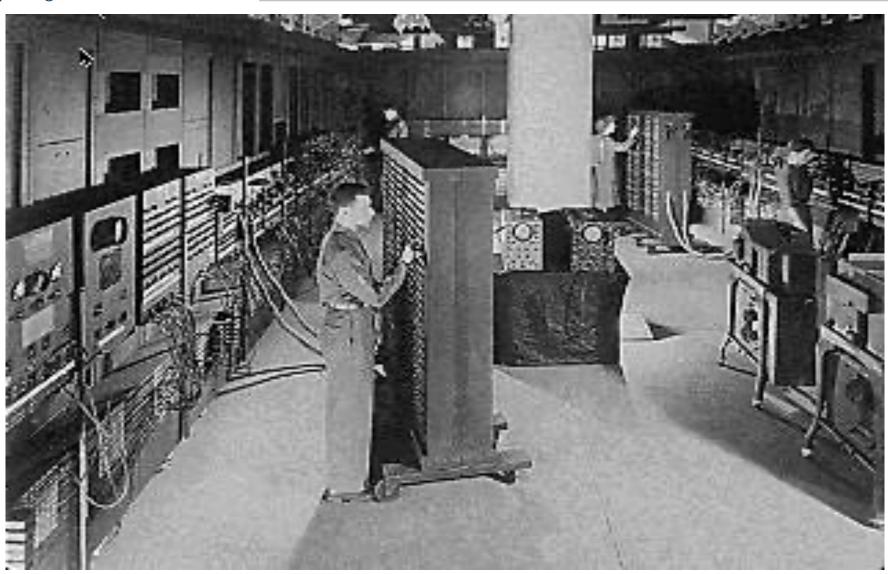


第1代—电子管计算机(1946--1957年代)

- 这一代计算机由电子管、电磁继电器等器件构成。
- 软件主要使用机器语言编程。
- 典型机器有ENIAC、IBM 701等。
- 1946年美国宾夕法尼亚大学莫尔电机学院与美国军方阿伯丁 弹道实验室试制成功一台被称为ENIAC(Electronic Numerical Integrator and Computer)的电子数字积分和计 算机。







世界上第一台实用电子计算机 ENIAC(1946)





第一台实用电子计算机(ENIAC)

60秒钟射程的弹道计算时间由原来的20分钟缩短到30秒



- 5000次加法/秒
- 重量30英吨
- 占地167m2
- 17468只电子管
- 1500个继电器
- 功耗174KW



第2代—晶体管计算机 (1958--1964年代)

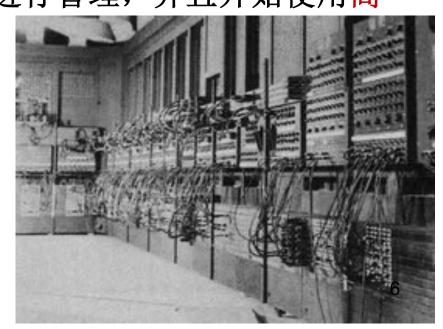
- 1954年,贝尔实验室使用800只晶体管组装了世界上第一台晶体管计算机--TRADIC。
- 这一代计算机采用晶体管、磁芯存储器等构成。
- 软件上有了监控程序对计算机进行管理,并且开始使用高

级语言。

• 这个时期的计算机有很多种,

如DEC PDP-1、IBM 7030、

Univac LARC等。





第3代—中小规模集成电路计算机(1965--1971年代)

- 1964年4月7日,IBM公司研制成功世界上第一个采用集成电路的通用计算机IBM 360系统,成为第三代计算机的里程碑。
- 这代计算机利用小规模及中规模集成电路芯片、多层印刷电路板及磁芯存储器等构成。
- 软件上,高级语言迅速发展并出现了分时操作系统。
- 在这个时期,计算机应用领域不断扩展,开始向国民经济各部门及军事领域渗透。
- 典型机器如IBM 360、370, DEC PDP-8等。





IBM System / 360





第4代—大规模和超大规模集成电路计算机(1972--2010年代)

- 大规模集成电路典型应用: 微处理器。
- 特点是:大规模、超大规模集成电路构成,主存也是由超大规模集成电路构成的半导体存储器来实现。
- 软件有DOS、Windows、Unix等。
- 典型计算机有IBM 3090、VAX 9000、PC机、苹果机等。



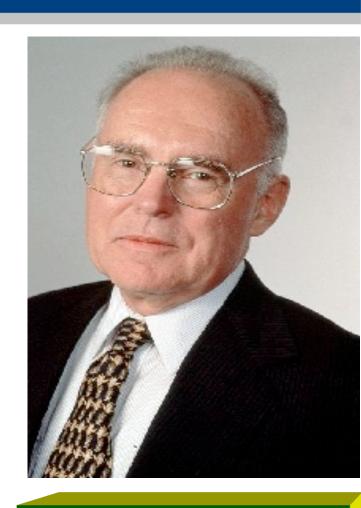
第5代—巨大规模集成电路计算机(2010--)

- 采用巨大规模、超高速集成电路构成计算机。
- 在结构上,计算机结构已从单处理器向多处理器、单机向 集群发展。既使构成微型机也采用多核处理器。
- 目前常见的是双核、4核处理器。用这样的多核处理器构成 计算机可获得很高的性能。此前,Intel公司已可以做出一 块芯片内含80个核的多核处理器,用这样的一块80核处理 器芯片构成的计算机,其速度已超过一万亿次。



摩尔定律

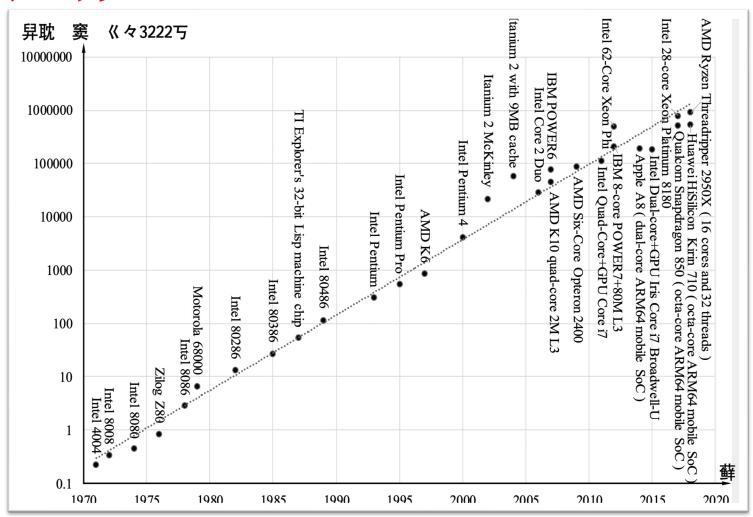
- 1965年由戈登·摩尔提出,加州理工学院的Carver Mead教授也参与了该定律的提出。
- 摩尔预言:单位面积集成电路上的晶体管数量每一年(65年)/两年(75年)就会翻番,而计算性能也将随之翻番。
- 直至2003,集成电路芯片的集成度: DRAM每18个 月、微处理器每24个月翻一倍。
- · 当前,Moore定律正在减慢,特别是2016年Intel 处 理器上集体管数量仅大约达到预测值的十分之一。



Gordon E. Moore
Intel公司的创始人之一



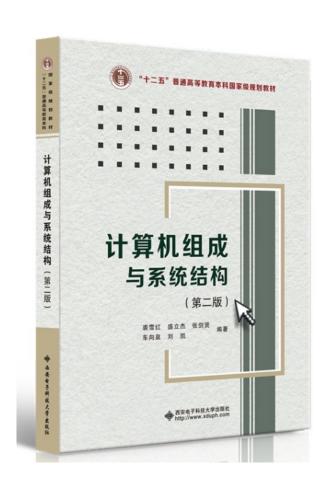
摩尔定律



12



- ▶计算机系统简介
- ▶计算机系统的层次结构
- ▶计算体系结构与组成原理
- ▶计算机硬件的主要技术指标





1. 计算机系统简介

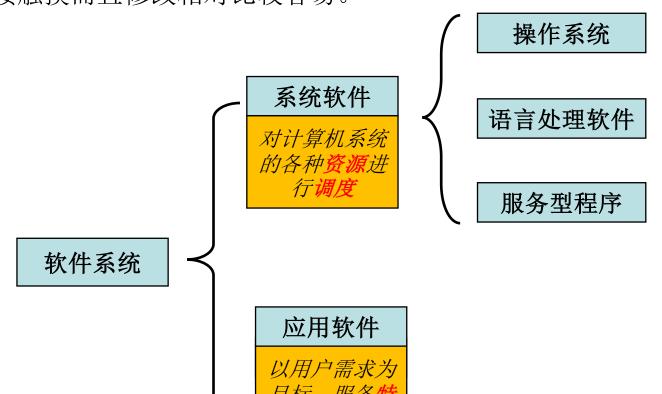
计算机系统是由"硬件"和"软件"两大部分组成的。

- ▶计算机系统中的软件系统
- ▶计算机系统中的硬件系统
 - ▶ 存储程序与冯·诺依曼计算机
 - ▶计算机的硬件基本组成
- ▶计算系统软件与硬件系统之间的关系



1.1 计算机系统——软件系统

计算机软件通常是存放在主存或外存中的二进制编码信息,不能直接触摸而且修改相对比较容易。





1.2 计算机系统——硬件系统

硬件系统是指计算机中有形的物理设备,是计算机中实际物理装置的总称。

(1) "存储程序"思想

存储程序

将程序和数据一起存放在计算机的存储器中。



程序控制

按指令地址访问存储器并取出指令,经译码依次产生 指令执行所需的控制信号,实现对计算的控制,完成 指令的功能。

"存储程序"思想奠定了现代计算机的<u>基本结构</u>,以此概念为基础的各类 计算机统称为*冯·诺依曼机*。



1.2 计算机系统——硬件系统

- (2) 冯·诺依曼计算机的主要特点
 - ▶ 指令和数据以同等地位存于存储器,并按地址访问。
 - ▶ 指令和数据以二进制数表示。
 - ▶ 指令有操作码和地址码构成。
 - ▶ 指令在存储器内按顺序存放。



1.2 计算机系统——硬件系统

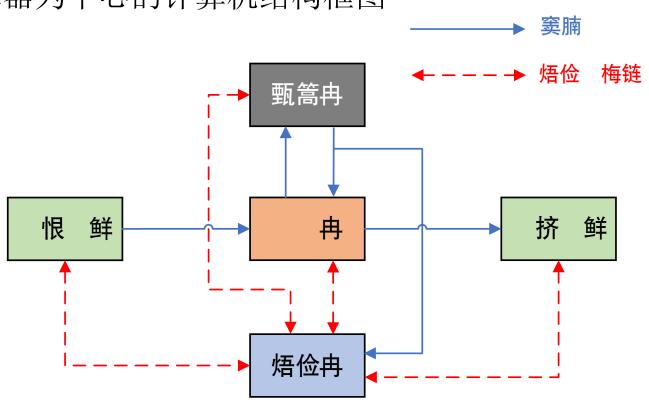
- (2) 冯·诺依曼计算机的主要特点
 - ▶ 计算机由<u>运算器</u>、<u>存储器</u>、<u>控制器</u>和<u>输入设备</u>、<u>输出设备</u>五大部件 组成;
 - ▶ 指令和数据以同等地位存于存储器,并按地址访问。
 - ▶ 指令和数据以二进制数表示。
 - ▶ 指令有操作码和地址码构成。
 - ▶ 指令在存储器内按顺序存放。
 - > 以运算器为中心。





1.2 计算机系统——硬件系统

以运算器为中心的计算机结构框图





1.2 计算机系统——硬件系统

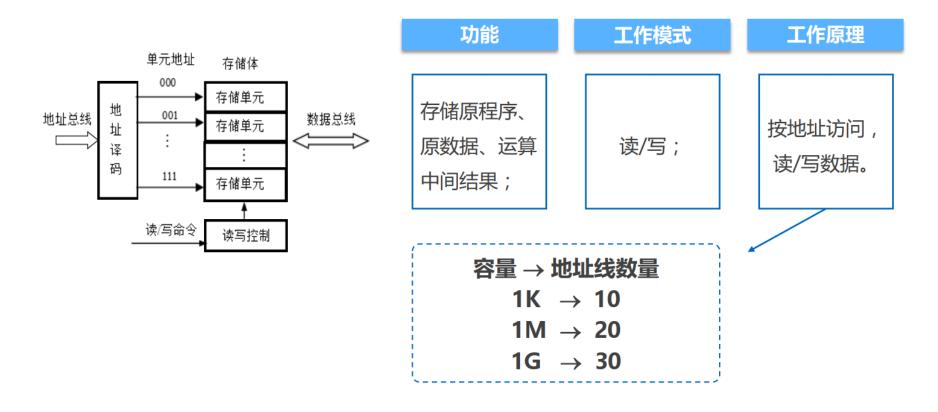
运算器是计算机的执行部件,用于进行<mark>算术运算和逻辑运算</mark>。运算器的核心是<u>算术逻辑单元</u>(Arithmetic and Logic Unit,ALU)。





1.2 计算机系统——硬件系统

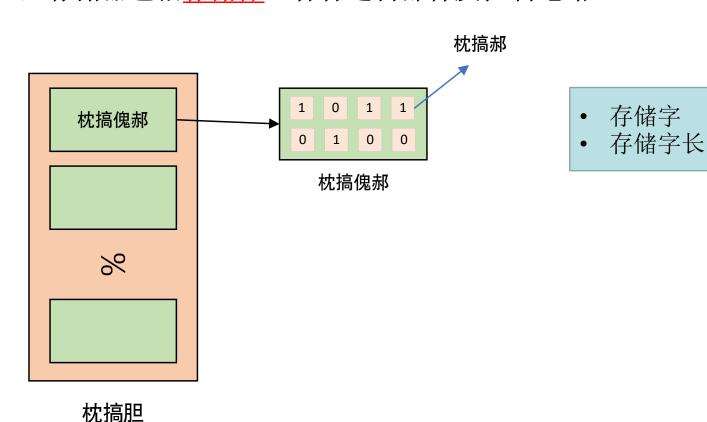
存储器分为主存储器和辅助存储器(又称外存储器)。





1.2 计算机系统——硬件系统

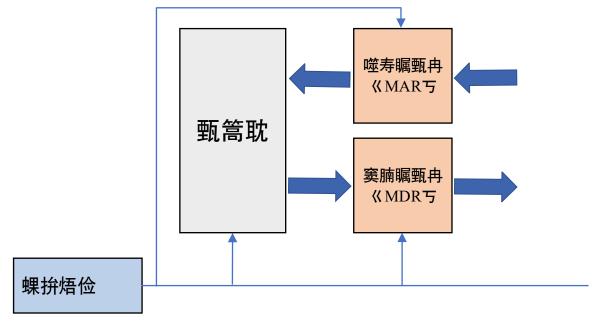
主存储器包括存储体、各种逻辑部件及控制电路。





1.2 计算机系统——硬件系统

主存储器以<u>地址存取方式</u>(按存储单元的地址进行存取)进行工作。



主存储器以<u>地址存取方式</u>(按存储单元的地址进行存取)进行工

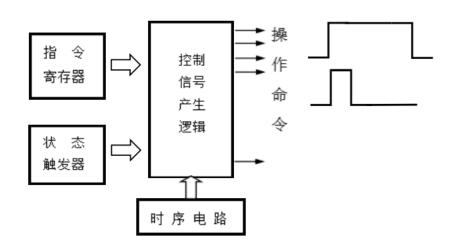
作。



1.2 计算机系统——硬件系统

控制器是计算机的指挥中心,由<u>程序计数器</u>(PC), <u>指令寄存器</u>(IR)和控制单元(CU)组成。

基本功能

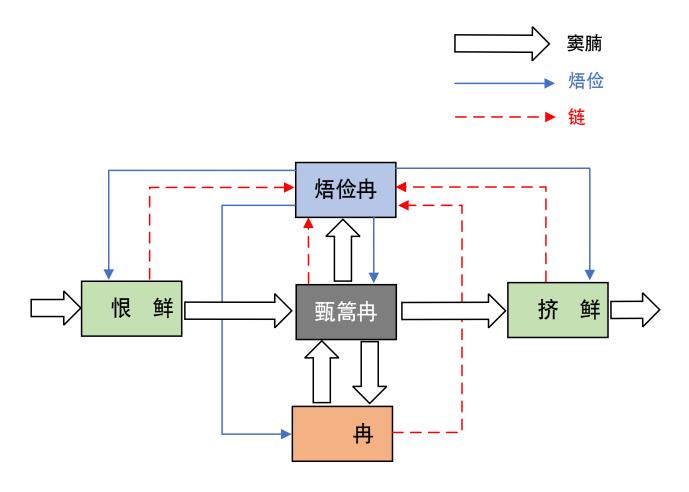


产生指令执行过程所需要的所有控制信号,控制相关功能部件执行相应操作;



1.2 计算机系统——硬件系统

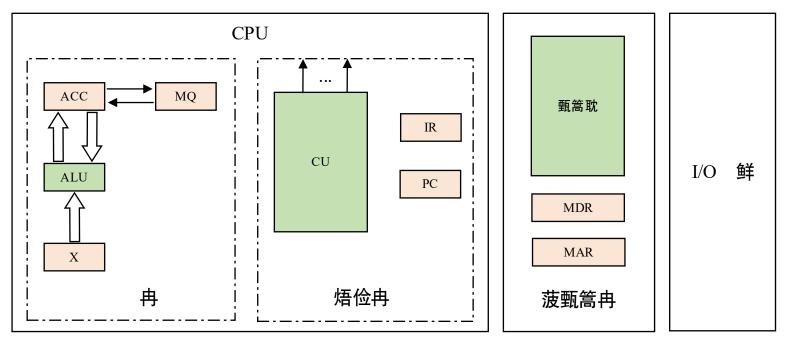
以存储器为中心的计算机结构框图





1.2 计算机系统——硬件系统

细化的计算机结构框图

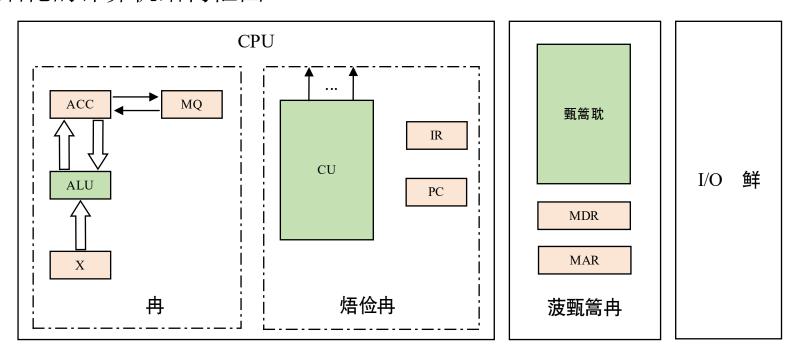


伴随大规模集成电路制作工艺的出现,在*逻辑关系*和 <u>电路结构</u>上联系紧密的 **运算器**和**控制器**往往集成在同一芯片上,称为**中央处理器**(Central Processing Unit,CPU)。



1.2 计算机系统——硬件系统

细化的计算机结构框图



CPU与主存储器合起来又可称为<u>主机</u>,I/O设备又可称<u>外部设备</u>。



1.3 计算系统软件与硬件系统之间的关系

1) 相互依存

硬件是软件运行的基础,软件的正常运行是硬件发挥作用的重要 途径。计算机系统必须要配备完善的软件系统才能正常工作,且应充 分发挥其硬件的功能。

2) 逻辑等效性

某些功能即可由硬件实现,也可由软件来实现。

3) 协同发展

软件随硬件技术的迅速发展而发展,而软件的不断发展与完善又 促进硬件的更新,两者密切地交织发展,缺一不可。

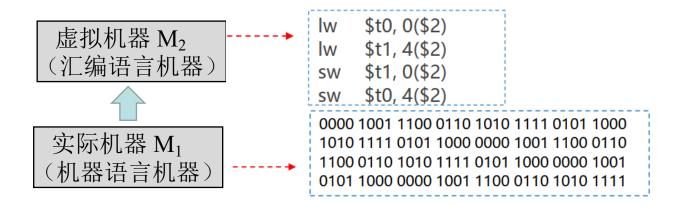


2. 计算机系统的层次结构

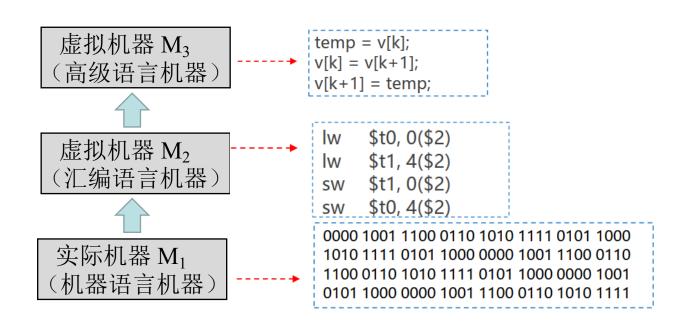
实际机器 M_1 (机器语言程序)

0000 1001 1100 0110 1010 1111 0101 1000 1010 1111 0101 1000 0000 1001 1110 0110 1100 0110 1100 0110 1010 1010 1010 1010 1010 1010 1111

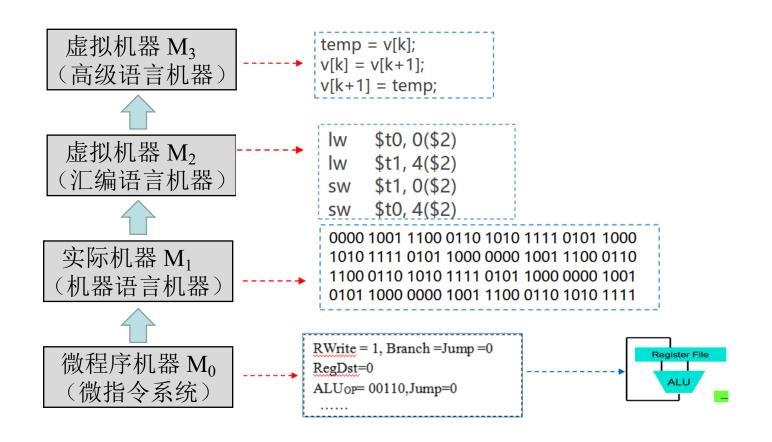




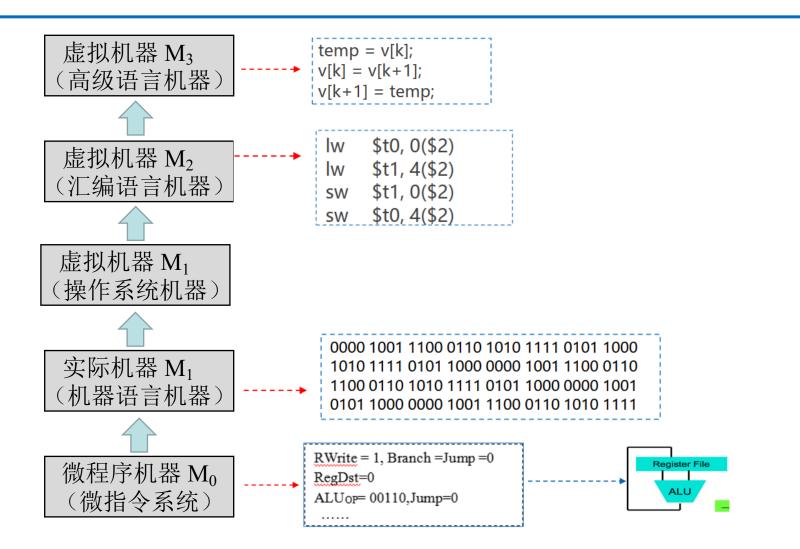














随堂测试1-1:

(单选题)

计算机只懂机器语言,而人类熟悉高级语言,故人 机通信必须借助()

- A. 编译程序
- B. 编辑程序
- C. 连接程序
- D. 载入程序





随堂测试1-2:

(单选题)

下列说法中()不正确。

- A. 高级语言的命令用英文单词来表示
- B. 高级语言的语法很接近人类语言
- C. 高级语言的执行速度比低级语言快
- D. 同一高级语言可在不同形式的计算机上执行。



随堂测试:

(单选题)

下列描述中,()是正确的。

- A. 控制器能理解、解释并执行所有的指令及存储结果。
- B. 一台计算机包括输入、输出、控制、存储及算术逻辑运算5个子系统。
- C. 所有的数据运算都在CPU的控制器中完成。
- D. 以上答案都正确。



随堂测试1-3:

(单选题)

计算机的算术逻辑单元和控制单元合称为()

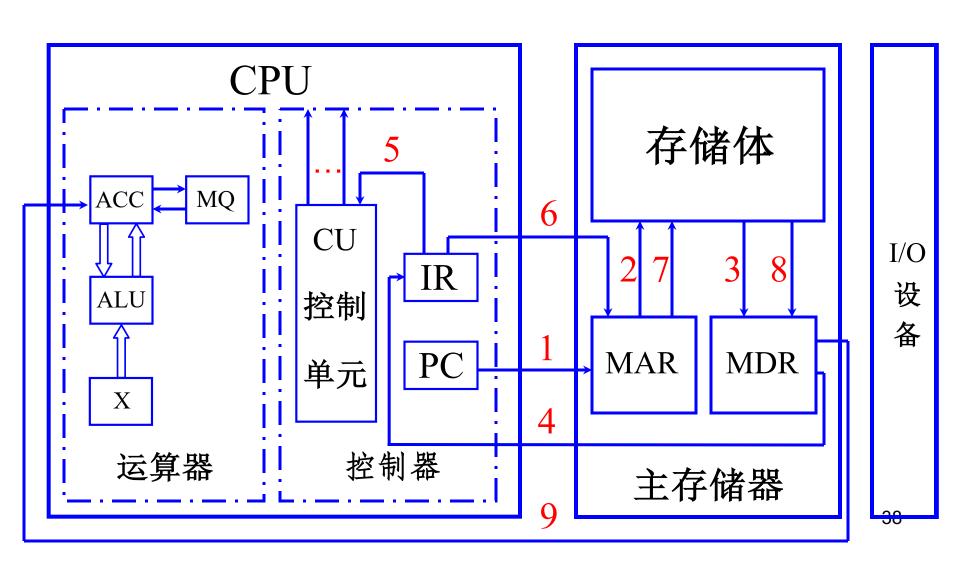
A. ALU

B. UP

C. CPU

D. CAD

以取数指令为例





随堂测试1-4:

(单选题)

用以指定待执行指令所在地址的是()

- A. 指令寄存器
- B. 数据计时器
- C. 程序计数器
- D. 累加器



3. 计算体系结构与组成原理

(1) 计算机体系结构(Computer architecture): 由程序员所看到的计算机系统的属性,即概念性结构及功能特性。主要包括:

- > 数据的表示形式;
- > 寻址方式;
- > 内部寄存器组;
- > 指令集:
- > 中断系统;

- > 处理器工作状态及其切换;
- > 存储系统;
- > 输入/输出结构;
- > 信息保护及特权;
- > 高性能设计等。



指令集体系结构

- IBM公司在设计System/360大型机时,首次引入了"指令集体系结构"的概念,将编程所需了解的硬件信息从硬件系统的具体实现中抽象出来。
- 处理器支持的指令和指令的字节级编码称为指令集体系结构 (Instruction-Set Architecture, ISA), 也称为指令系统结构 (Instruction System Architecture, ISA)。



ISA是软件和硬件的分界面,是软件和硬件之间接口的完整定义。

- 软件(程序)是由ISA规定的"指令"组成的,软件设计者面向ISA进行编程,开发出的软件可以不经修改直接在相同ISA的计算机系统上运行。
- 编译器设计者需要知道ISA规定了哪些指令,它们是如何编码的,以便实现将高级语言程序源代码转换为处理器能够识别并执行的机器语言代码。
- 处理器设计者的任务是实现ISA规定的指令、功能和相应的硬件模块。



ISA规定了计算机中程序员可见的所有组件及操作,包括:

- ① 指令集:处理器可执行的指令的集合。
 - 指令格式、操作种类以及每种操作对应的操作数的相应规定
 - _ 数据类型,即指令可以接受的操作数的类型
 - 寻址方式,即指令获取操作数的方式

② 软件可见的处理器状态:

- 寄存器组,即寄存器的个数、名称(编号)、长度和用途,包括通用寄存器和特殊用途寄存器
- 指令执行过程的控制方式,包括程序计数器、条件码定义以及每条 指令对状态的影响



ISA规定了计算机中程序员可见的所有组件及操作,包括:

③存储模型:

- 主存组织,即主存最大寻址空间和编址方式
- 字节次序,即操作数在存储空间存放时按照大端还是小端方式存放
- 存储保护
- 虚拟存储器的管理方式
- 输入输出接口的访问与管理方式

④系统模型:

- 处理器状态
- 特权级别
- 中断和异常的处理方式



指令集架构可以以很多方式进行划分,常见的

- CISC(Complex instruction set computing)复杂指令集架构
 - CISC是一种为了便于编程和提高记忆体访问效率的晶片设计体系。
 - · 一般CISC计算机所含的指令数目至少300条以上,有的甚至超过500条.
 - 以Intel, AMD的X86 CPU为代表
- RISC(Reduced instruction set computing)精简指令集架构
 - · 以ARM,IBM Power为代表



指令集架构可以以很多方式进行划分,常见的

- CISC(Complex instruction set computing)复杂指令集架构
 - 以Intel, AMD的X86 CPU为代表
- RISC(Reduced instruction set computing)精简指令集架构
 - · 以ARM, IBM Power为代表





CISC:

X86: 1978 Intel,善于处理大数据,IP掌握在Intel和AMD手中

RISC:

- ① ARM: 1983,移动(手机)市场,处理快数据为主,目前也使用在便携笔记本中,IP大部分掌握在ARM公司。
- ② MIPS (Microprocessor without Interlocked Pipeline Stages): MIPS, 2021转身投入RISC-V
- ③ RISC-V: 2010,加州大学伯克利分校,开源。当需要同时兼顾数据传输速度与传输量时,RISC-V表现出了较强的优势。模块化、指令数目少、免费开源。





3. 计算体系结构与组成原理

(2) 计算机组成(Computer organization)

- ▶ 计算机体系结构的逻辑实现,包括物理机器级内的数据流和控制流的 组成以及逻辑设计等。
- ▶ 它着眼于物理机器级内各事件的排序方式与控制方式,各部件的功能 以及各部件的联系。



3. 计算体系结构与组成原理

计算机组成的设计主要包括如下诸方面的内容:

- > 数据通路的宽度:
- > 专用部件的设置(如乘除 > 控制机构的组成方式: 法专用部件、浮点运算专 用部件等):
- > 各功能部件的并行程度:
- > 各种操作的相容性与互斥

- 性:
- > 缓冲与排队技术的应用:
- ▶ 预估、预判方法:
- > 高可靠性技术等等。



3. 计算体系结构与组成原理

(3) 计算机实现(Computer implementation)

- 计算机实现就是指计算机组成的物理实现。包括处理机、主存等部件的物理结构,器件的集成度和速度,模块、插件、底板的划分与连接,信号传输,电源、冷却及整机装配技术等。
- ▶ 它着眼于器件技术和微组装技术,其中器件技术在实现技术中占主导作用。利用具体的集成电路芯片、电子元器件、部件、插头、插座等等,根据计算机组成的逻辑设计,实现物理计算机。



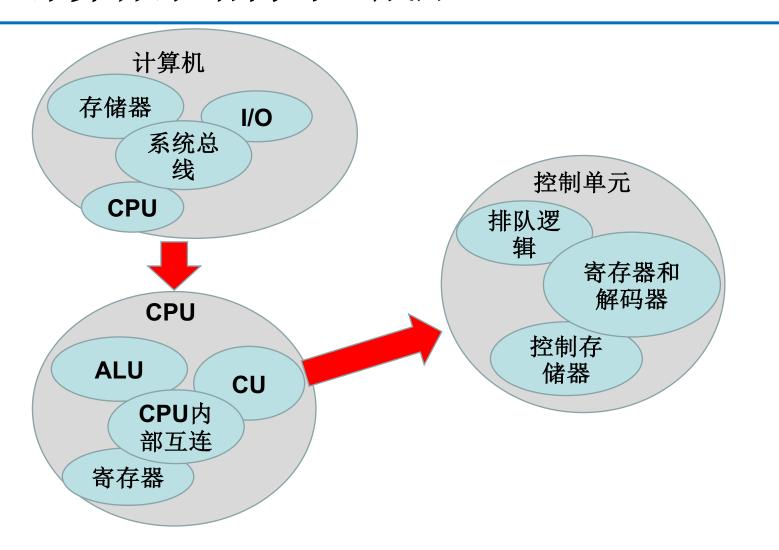
3. 计算体系结构与组成原理

体系结构、组成、实现的关系:

- 具有相同计算机体系结构(如指令系统相同)的计算机因为速度要求 不同等因素可以采用不同的计算机组成。
- 一种计算机组成可以采用多种不同的计算机实现。取决于性能价格 比的要求与器件技术的现状。



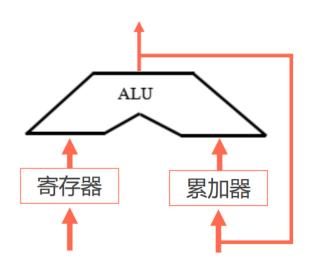
3. 计算体系结构与组成原理





计算机的性能描述

1. 机器字长: 指CPU一次能处理的数据位数



- 一般与内部寄存器的位数相等。
- 字长越长,表示数据的范围就越大, 精确度就越高。
- 直接影响加法器(或ALU)、数据 总线以及存储字长的位数
- 目前常见的有32位和64位。



2.存储容量

存储器的容量包括主存容量和辅存容量。

① 主存容量是指主存中存放二进制的总位数。即:

存储容量=存储单元个数×存储字长

- · 存储器地址寄存器(MAR)的位数反映了存储单元的个数,
- · 存储器数据寄存器(MDR)的位数反映了存储字长。
 - 如MAR为16位, 2¹⁶=65536, 表示此存储体内有65536个存储单元(即2⁶K个存储字, 1K=1024=2¹⁰)
 - MDR为32位,表示存储容量为2¹⁶×32=2Mb(1M=2²⁰)



2. 存储容量

- 现代计算机中常以字节来描述容量的大小,一个字节(记作B)已被定义为8位二进制代码。
 - 存储容量为2Mb,可以表示为2²¹b=2¹⁸B= 2⁸KB,记作
 2¹⁸B或256KB。
- ② <mark>辅存容量通常用字节数表示</mark>,例如,某机辅存 (如硬盘)的容量为80GB(1GB=1024M=2³⁰)



3. 运算速度

决定CPU的运 算能力大小的 因素包括哪些?

CPU的主频 CPU的核数 CPU的缓存

- 1) 主频f/ 时钟周期T
- 主频f: 指CPU内核工作的时钟频率,即CPU内数字脉冲信号振荡的速率。
- 时钟周期T: 也称节拍周期,是计算机中最基本的、最小的时间单位。
- f与T的关系: 互为倒数。
 - 如*f*=100MHz时, T=10ns; *f*=1GHz时, T=1ns; *f*=4GHz
 时, T=0.25ns。



3. 运算速度

- 2) CPI (Clock Cycles Per Instruction)
- 执行一条指令(平均)需要的时钟周期数(即T周期的个数)

 $= \sum ($ 各类指令的CPI×该类指令的比例)

CPI一般和哪些 因素有关? 程序的编写 指令集体系结构 计算机的组成与系统结构



3. 运算速度

2) CPI (Clock Cycles Per Instruction)

· 例如: 某计算机指令系统中各类指令所占比例及CPI 如下表所示, 求程序的CPI。

指令类型	СЫ	指令比例
算术和逻辑	1	60%
Load/Store	2	18%
转移	4	12%
Cache缺失访存	8	10%





3. 运算速度

2) CPI (Clock Cycles Per Instruction)

· 例如: 某计算机指令系统中各类指令所占比例及CPI 如下表所示, 求程序的CPI。

指令类型	СЫ	指令比例
算术和逻辑	1	60%
Load/Store	2	18%
转移	4	12%
Cache缺失访存	8	10%

解: CPI=1×60%+2×18%+4×12%+8×10%=2.24



3. 运算速度

- 3) MIPS (Million Instructions Per Second)
- · 每秒钟CPU能执行的指令总条数(单位: 百万条/秒)

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

$$= \frac{\text{指令条数}}{(\text{所有指令cpu时钟周期数之和}) \times 10^6}$$

$$= \frac{f}{\text{CPI} \times 10^6}$$
全性能公式



例如:某计算机主频为1GHz,在其上运行的目标代码包含2×10⁵条指令,分4类,各类指令所知比例和各自CPI如下表示,求程序的MIPS。

指令类型	СЫ	指令比例
算术和逻辑	1	60%
Load/Store	2	18%
转移	4	12%
Cache缺失访存	8	10%



例如:某计算机主频为1GHz,在其上运行的目标代码包含2×10⁵条指令,分4类,各类指令所知比例和各自CPI如下表示,求程序的MIPS。

指令类型	СЫ	指令比例
算术和逻辑	1	60%
Load/Store	2	18%
转移	4	12%
Cache缺失访存	8	10%

解:根据CPU全性能公式: MIPS=
$$\frac{f}{CPI \times 10^6}$$

$$CPI=1\times60\%+2\times18\%+4\times12\%+8\times10\%=2.24$$

MIPS=
$$\frac{f}{\text{CPI} \times 10^6}$$
=1×10⁹/(2.24×10⁶)=446.4



随堂测试1-5:

(单选题)

通常称"容量为640K的存储器"是指下列()

A、640×103字节的存储器

B、640×10³位的存储器

C、640×210位的存储器

D、640×210字节的存储器



随堂测试1-6:

(单选题)

程序P在计算机M上的执行时间是20s,编译优化后, P执行的指令数减少到原来的70%,而CPI增加到原来的1.2倍,则P在M上的执行时间是()。

- A. 8.4s
- **B.** 11.7s
- C. 14s
- D. 16.8s





随堂测试1-6:

(单选题)

程序P在计算机M上的执行时间是20s,编译优化后, P执行的指令数减少到原来的70%,而CPI增加到原来的1.2倍,则P在M上的执行时间是()。

A. 8.4s B. 11.7s C. 14s D. 16.8s

答案解析:

设原来指令条数为x,那么原CPI就为20/x。 经过编译优化后,指令条数减少为0.7x,CPI增加到24/x。 因此,现在P在M上的执行时间为:指令条数





Amdahl定律

- 计算机领域中被公认的一个重要定律。它定义了串行系统并行化后的加速比的计算公式和理论上限。
- 由IBM公司的计算机体系结构师、IBM360系列机的主要设计者吉恩·阿姆达尔在1967年发表的论文中首先提出的。
- 内容: 计算机系统中某部件由于采用某种更快的执行方式使系统性能改进后,整个系统性能的提高与该执行方式的使用频率或占总执行时间的比例有关。



Amdahl定律

Amdahl定律定义了加速比(speedup), 即

> 加速比 = 改进后的系统性能改进前的系统性能 改进前的系统执行时间 改进后的系统执行时间

即加速比就是计算机改进后比改进前快了多少倍

应用:改善"系统瓶颈"问题,用于系统改进之后性能提升 的测试。 67



计算机系统的加速比取决于两个因素:

- 可改进比例 f_e : 可改进部分在原系统执行时间中所占的比例
 - 例如,程序的执行时间为100s,可改进的部分是其中的20s,则 $f_e = 0.2$ 。可见, f_e 总是小于或等1的。
- 部件加速比r。可改进部分改进后性能提高的程度
 - 例如,某部件改进后,执行时间由原来的20s减少到5s,则 部件加速比 r_e =20/5=4。可见, r_e 一般是大于1的。



Amdahl定律

假设改进前的系统执行时间为 T_0 ,改进后的系统执行时间为 T_n 则 f

 $T_n = T_0 (1 - f_e + \frac{f_e}{r_e})$

若加速比用 S_p 表示,则有

$$S_p = \frac{1}{(1 - f_e) + \frac{f_e}{r_e}}$$

多个部件可同时改进的情况下,Amdahl定律可表示为

$$S_p = \frac{1}{(1 - \sum f_e) + \sum (\frac{f_e}{r_e})}$$



例:某计算机系统的某一部件的处理时间为总处理时间的40%,该部件改进后的部件加速比为10,则改进后系统的加速比 S_p 为多少?



例:某计算机系统的某一部件的处理时间为总处理时间的40%,该部件改进后的部件加速比为10,则改进后系统的加速比 S_p 为多少?

解: 由题意可知, $f_{\rm e} = 0.4$, $r_{\rm e} = 10$, 则

$$S_p = 1/[(1-0.4)+0.4/10] \approx 1.56$$

由计算可见,既使某一部件改进后的加速比已达10倍,但该部件仅影响到系统执行时间的一部分,对整个计算机系统的贡献有限,所以,改进后系统的加速比只有1.5倍左右。



$$S_p = \frac{1}{(1 - f_e) + \frac{f_e}{r_e}}$$

- 当 $(1-f_e)=0$ 时,Max $S_p=n$; 当 $(1-f_e)=1$ 时,Min $S_p=1$;
- 当 $\mathbf{n} \to \infty$ 时, $\lim \mathbf{S_p} \to 1/(1-f_e)$,这也是加速比上限。

例如,若串行代码占整个代码的20%,则并行处理的总体性能 提供不可能超过5倍。



Amdahl定律

结论

Amdahl定律告诉我们:

- 系统的改进不应是局部的,即便局部改进达到极致,对系统整体性能的提升也是有限的。
- 要想达到高并行性能,就必须考虑整体系统及程序。



例 若计算机系统有三个部件a、b、c是可改进的,它们的部件加速比分别为30、30、20。它们在总执行时间中所占的比例分别是30%、30%、20%。试计算这三部件同时改进后系统的加速比。

解: 多个部件可同时改进的情况下, Amdahl定律可表示为:

$$S_p = \frac{1}{(1 - \sum f_e) + \sum (\frac{f_e}{r_e})}$$

将已知条件代入上式,计算出 $S_p \approx 4.35$



随堂测试1-7:

(单选题)

已知计算机A的时钟频率为800MHz,假定某程序在计算机A上运行需要12s。现在硬件设计人员想设计计算机B,希望该程序在B上的运行时间能缩短为8s,使用新技术后可使B的时钟频率大幅度提高,但在B上运行该程序所需的时钟周期数为在A上的1.5倍。那么,机器B的时钟频率至少应为()才能达到所希望的要求?

- A, 800MHz
- **B**、1.2GHz
- C. 1.5GHz
- **D** \ 1.8GHz



课程介绍

