

东南大学计算机学院

计算机组织与结构

主讲教师： 徐造林

第二章 计算机系统概论

2.1 计算机的定义、功能和结构

2.1.1 什么是计算机

- ◆ **计算机**：按照内部存储的**指令**序列，对数字化**信息**进行自动高速**处理**、**存储**、**传送**、**控制**的装置。
- **指令**：指示计算机硬件完成某种功能的明确的命令；
- **信息**：有用的数据，有多种不同类型，其表现手段可以采用**数字化形式或模拟量形式**；
- **运算**：包括**算术运算**和**逻辑运算**，要求自动与高速；

- **处理**：对信息进行搜索、识别、变换，甚至联想、思考和推理等等。

◆ **计算机的特性**可以归纳为高速、通用、准确、智能四大特性。

◆ **计算机**可以分成**模拟计算机**与**数字计算机**两类：

▲ **模拟计算机：**

- 对**连续变量**进行运算或解算装置，变量可以是连续变化的电压、电流或电荷。进行加、减、乘、除、微分、积分等数学运算由相应的运算部件完成。

- 模拟计算机工作具有连续性、并行性和实时性；且操作简便，适用于连续系统的实时仿真。
- 主要缺点是受元器件精度限制与运算放大器零点漂移的影响，整机精度远低于数字计算机。

▲ 数字计算机：

- 对用离散符号表示的数据或信息自动进行处理的电子装置。
- 就是通常所说的计算机；计算机中采用二进制编码信息来表示数值、字符、指令或其它控制信息。
- 数字计算机其运算部件由高速的电子元器件组成，因此速度快、精度高，应用更广泛。

- ◆ 计算机按其用途来分可以分成**专用机**和**通用机**两类
 - ▲ **专用机**是专门用于某种用途，它对于特定用途而言最经济、最快速、最有效；如近些年高速发展的各种**嵌入式系统**也属于此范畴。
 - ▲ **通用计算机**按其规模、性能和价格来分，又可分为巨型机、大型机、小型机、工作站、微型机等多种类型。

1) 巨型机 (Supercomputer) :

- 计算机族中速度最快、性能最高、技术也最复杂、价格也是最贵的一类计算机;
- 分4代, 分别是单指令流多数据流 (SIMD) 的阵列处理机, 具有流水线结构的向量机 (VP), 多指令流、多数据流 (MIMD) 的共享主存多处理机系统 (MP), 大规模并行处理系统 (MPP) 。

2) 大型机 (Large-scale Computer、main-frame) :

- 使用当时代的先进技术构成的一类高性能、大容量计算机（但性能与价格指标均低于巨型机），它代表该时期计算机技术的综合水平。
- 大型机的处理机系统可以是单处理机、多处理机或多个子系统的复合体。

3) 小型机 (minicomputer) :

- 是一种规模与价格均介于大型机与微型机之间的一类计算机。

4) 工作站 (workstation) :

- 以个人计算环境和分布式网络计算环境为基础、其性能高于微型机的一类多功能计算机;
- 为特定应用领域的人员提供一个具有友好人机界面的高效率工作平台。具有高速的定点和浮点运算能力, 很强的处理多媒体信息的能力。如[工程设计](#)。

5) 微型机 (minicomputer) :

- 以微处理器为中央处理器而组成的计算机系统;

◆ 流行微型计算机（PC）的发展

- （1）采用**Intel8088**处理器的IBM PC和IBM PC/XT为第1代微型计算机。
- （2）采用**Intel80286**处理器的IBM PC/AT为第2代微型计算机。
- （3）采用**Intel80386**处理器的微型计算机为第3代微型计算机。
- （4）采用**Intel80486**处理器的微型计算机为第4代微型计算机。

(5) 采用**Pentium**处理器的微型计算机为第5代微型计算机。

- Intel公司的奔腾芯片有多个序列：

经典奔腾（Classical Pentium）；

高能奔腾（Pentium Pro）；

多能奔腾（Pentium MMX）；

奔腾2（PentiumII）；

奔腾3（PentiumIII）；

奔腾4（PentiumIV）。

- AMD公司的**64位处理器**。

- Intel公司的**多核处理器**

。

2.1.2 计算机的功能与应用

◆ 计算机的基本功能主要包括

- 数据运算、处理
- 数据存储
- 数据传送
- 控制

◆ 数据运算、处理功能

- **运算功能**：算术运算功能和逻辑运算，应用于数值计算和非数值计算两个方面；
- **处理对象**：数值、字符、图形、图像、声音和视频等。

◆ 数据存储功能

- **主存储器**：保存指令和数据；
- **辅助存储器**：以文件的形式保存大量数据信息。

◆ 数据传送功能

- 内部数据流动：CPU和主存以及CPU内部寄存器与运算器之间的数据流动；
- 外部数据传送：输入/输出（I/O）和计算机通信。

◆ 控制功能

- 控制器：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；
- 指令编码、指令系统：一台计算机的所有指令集合。

◆ 计算机应用经历三个阶段

- 第一阶段，这时计算机所处理的大都是科学计算和工程计算问题；
- 第二阶段，这时计算机开始在企业中应用，数据处理问题日益增多；
- 第三个阶段，这时，计算机进入各行各业和千家万户。

◆ 计算机应用领域

1. 语言与文字的处理。
2. 计算机图形学与数字图象处理。
3. 多媒体技术（JPEG、MPEG、MMX技术的处理器）。
4. 计算机辅助技术（CAD、CAM、CAI、CAT）。
5. 计算机信息系统。
6. 计算机控制。

- ◆ 计算机应用技术的发展方向为巨型化、微型化、集成化、网络化、智能化、多媒体化。

2.1.3 计算机功能部件



图2.1 现代计算机功能部件

1.1.4 计算机结构、组织和实现

◆ 计算机结构与计算机组织的辨析

- 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。
- ◆ 计算机实现指计算机组成的物理实现。

2.2 计算机的发展历程

一、电子计算机的诞生

◆ 第一台电子计算**ENIAC** (**Electronic Numerical Integrator and Computer**) 于1946年在美国诞生。

- ① 每秒**5000**次加法运算;
- ② 每秒**50**次乘法运算;
- ③ 平方和立方计算;
- ④ **Sin**和**Cos**函数数值运算;
- ⑤ 其它更复杂的计算。

- 其逻辑元件采用电子管，存储器件为磁鼓，典型逻辑结构为定点运算；用低级语言编程，无操作系统。
- 代表人物：英国科学家艾兰·图灵和美籍匈牙利科学家冯·诺依曼诞生。
- 冯·诺依曼计算机模型。
 - 1) 计算机由运算器、存储器、控制器和输入/输出五个部件组成；
 - 2) 存储器以二进制形式存储指令和数据；
 - 3) 存储程序工作方式；
 - 4) 五部件以运算器为中心进行组织。

- 冯·诺依曼计算机组成

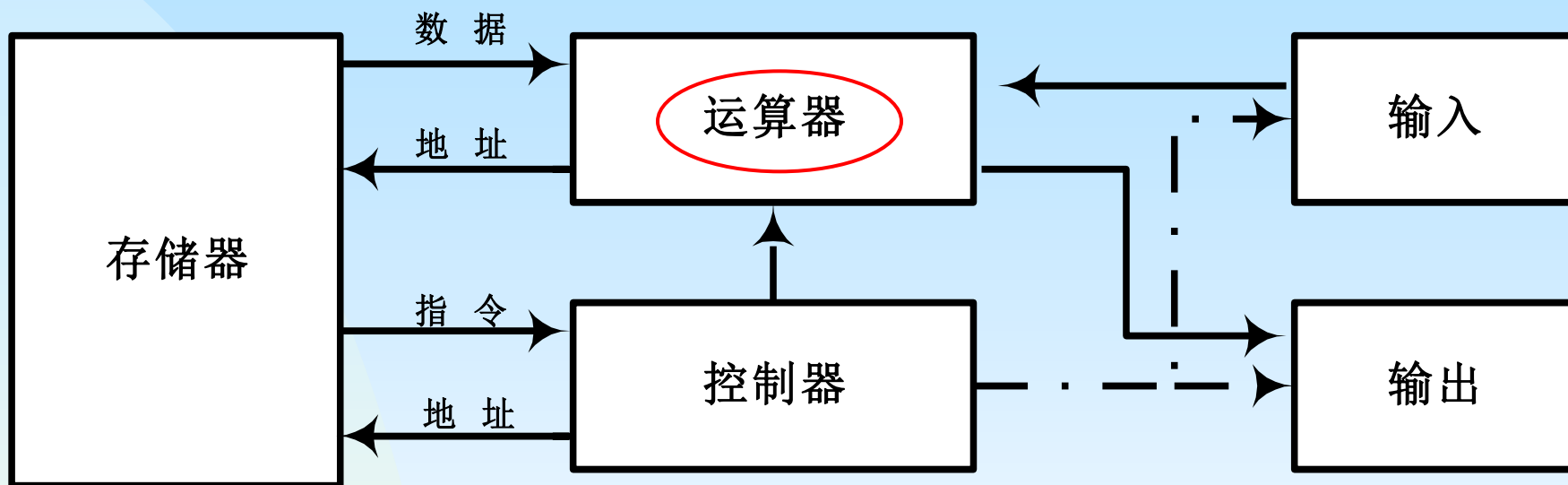


图2.2 计算机基本组成

- 现代计算机中心部件的转移

二、第二代计算机（50年代中后期到60年代中）

- 逻辑元件采用晶体管，其内存储器由磁芯构成，磁鼓与磁带成为外存储器。
- 逻辑结构实现了浮点运算，并提出了变址、中断、I/O处理等新概念。出现了多种高级语言及其编译程序。

三、第三代计算机（60年代中到70年代中）

- 逻辑元件与存储器均由**集成电路**实现；
- 硬件上，**微程序控制**、**高速缓存**、**虚拟存储器**、**流水线**等先进计算机技术开始使用；
- 出现**操作系统**，它具有资源调度、**人/机通信**和输入输出控制等功能；
- 另一大特点是大型/巨型机与小型机同时发展。

四、第四代计算机（70年代中期开始—）

- ◆ 大规模（LSI）和超大规模（VLSI）集成电路及微处理器为这一代计算机的典型特征。
- 摩尔（Golden Moore）定律：“由于硅技术的不断改进，每18个月，集成度将翻一翻，速度将提高一倍，其价格将降低一半”。

表 1.1 Intel CPU 芯片性能比较

芯片名	年代	集成度 (晶体管)	主频时钟 (MHz)	线宽 (μm)	数据总线宽 (位)	地址总线宽 (位)
4004	71	2000	2	2	4	20 24 32 32
8080	74	8000	4	1.5	8	
8086	78	3 万	5-8	1.5	16	
80286	84	13 万	10	1-1.5	16	
80386	85	27 万	33	1-1.5	32	
80486	89	100 万	35-40	1	32	
Pentium	93	300 万	60-150	0.6	64	32
ntium Pro (P6)	95	550 万	150-200	0.6	64	32
P II (P6+MMX)	97	750 万	300-450	0.35	64	32
PIII	99	800 万	600-1000	0.25	64	32
P4	2000	1000 万	1.4GHZ	0.18	64	32

- 并行处理技术的研究与应用及众多巨型机的产生。
- 第四代机时期的一个重要特点是**计算机网络**的发展与广泛应用。

五、新一代计算机

- ◆ 新器件和**非冯·诺依曼**结构已成为**新一代计算机的公认标志**。
- 神经元、生物芯片、分子电子器件、超导计算机、量子计算机。
- 日本第五代计算机系统研究计划（1982-1992年），确立**并行推理**将是未来信息处理的核心。

六、我国计算机的发展

◆ 中国古代对计算机理论与计算工具方面曾作出过杰出贡献

① **二进制数的位**：阴爻(yao)对应0，阳爻对应1，八卦为3个爻的集合；

② **十进制计数系统**：

- 殷墟甲骨文和周代青铜器上的铭文记载：用1-9的9个符号和表示十、百、千、万位值的4个符号来十万以内的自然数；

③ **筹算**：利用算筹作为运算工具；

④ **珠算**：以算盘作为工具，在元代已广泛使用。

- ◆ 我国一、二、三代机的推出比世界上一、二、三代机晚了整整一代，但这是自力更生研制出来的。
- ◆ 国防科技大学先后于1983年和1992年研制成巨型机系统银河I和II，运算速度都超过亿次。均配有操作系统、高级语言编译程序等系统软件。

- ◆ 1995年5月，中科院计算所国家智能计算机研究中心研制的“曙光1000”大规模并行处理机诞生。该机峰值速度可达每秒25亿次单精度浮点运算，内存容量为1000MB，结点机间总通信容量为每秒4800MB。
- ◆ 21世纪初，“曙光3000超级服务器”问世。由70台节点计算机组成，280个处理器，峰值浮点运算速度达每秒4032亿次，内存容量160GB。
- ◆ 联想集团最新研制新型超级并行计算机。

2.3 计算机系统组成及工作过程

2.3.1 计算机系统的层次结构

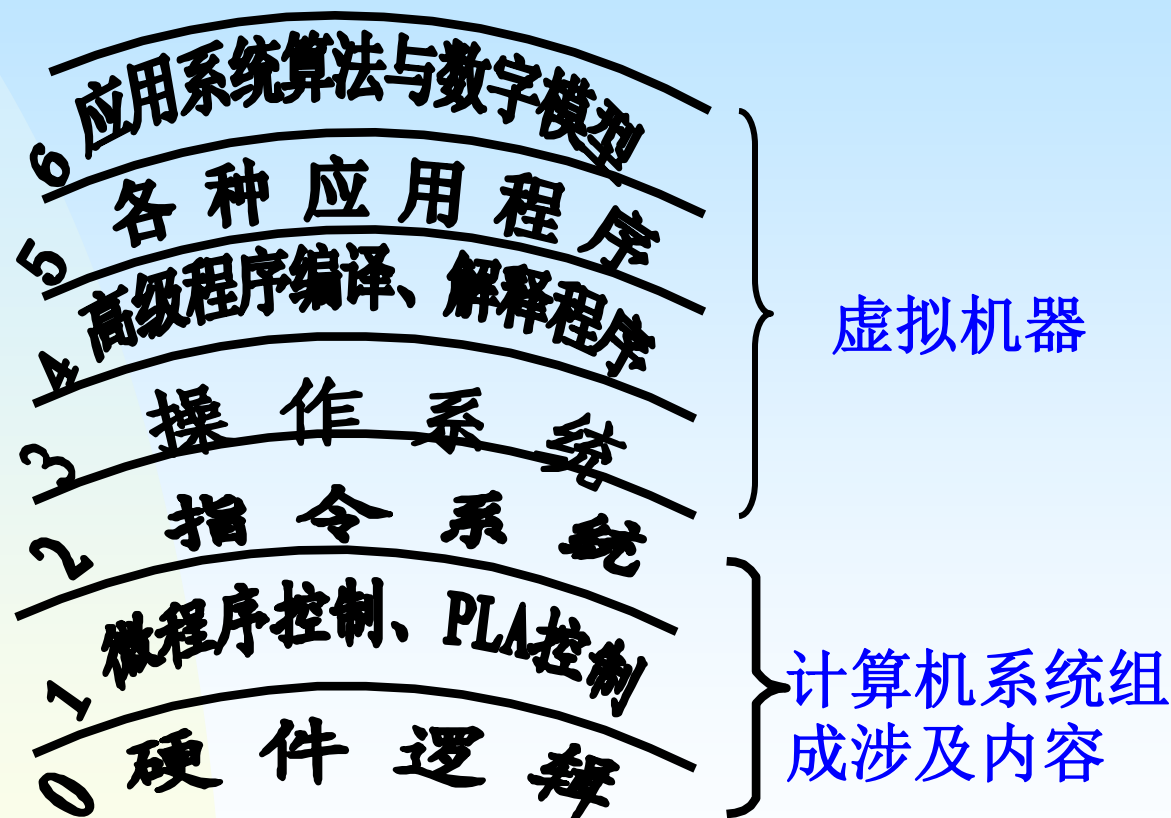


图2.3 计算机系统层次结构

◆ 本书讨论的范围涉及第0、1、2共3层

▲ 主要内容如下：

1. 高速的算术、逻辑运算方法及ALU的逻辑设计；
2. 高速的指令执行过程及指令部件的设计与实现；
 - 是采用组合逻辑技术、或微程序设计技术，还是PLA技术；
 - 是复杂指令集计算机（CISC），还是精简指令集计算机（RISC）；

3. 提高存储器容量与速度的方法，即如何解决“**CPU-Cache-MM-外存**”之间的匹配问题；
4. 高效率的**输入/输出方法**、组织，以及它们之间的互联技术；
5. 计算机五大部件（**运算器、控制器、存储器、输入和输出设备**）之间的相互作用、高效**接口**（总线）。

2.3.2 计算机中各功能部件的互连结构—总线

- **互连结构**：连接部件或模块（CPU、存储器、I/O）模块之间连接起来的所有数据通路。
- **总线**：计算机中普遍采用的互连结构，因为它可共享传输介质。
- **总线**有CPU内部总线、系统总线和局部总线等；包括**数据总线**、**地址总线**和**控制总线**。
- **总线的结构**有单总线、双总线、三总线等。

2.3.3 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。
- 考查指令 **ADD NUM, R0** 的执行过程。

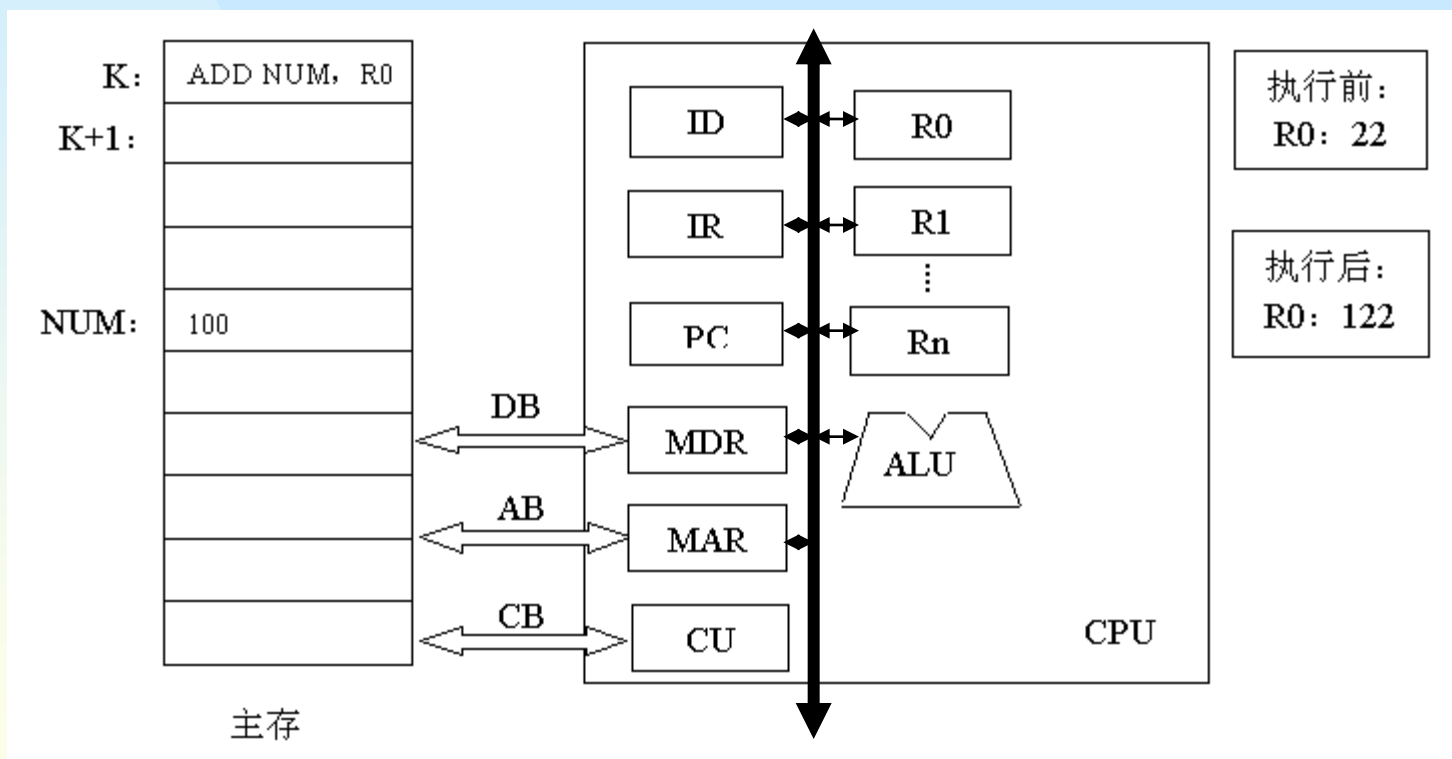


图2.4 CPU组成以及和存储器的连接

2.4 计算机系统性能指标及性能设计

2.4.1 计算机系统性能指标

***系统性能**：指在计算机硬件上运行的计算机软件的性能

***机器字长**：指CPU一次能处理的二进制位数。

└→指ALU一次能处理的

n位CPU——指机器字长为n个二进制位的CPU；

例如，Core 2 CPU为64位CPU

对系统性能的影响——机器字长越长，数据处理性能越好；

(∵应用数据长度>机器字长时，需分次运算)

对其它硬件的影响——直接影响ALU、REG长度，

间接影响存储字长、数据总线位数

* **CPI**：执行每条指令所需的平均时钟周期数。

***响应时间**：指一个任务从**任务输入到结果输出的总时间**，

即 $T_{\text{响应}} = T_{\text{CPU}} + T_{\text{I/O等待}}$, $T_{\text{CPU}} = T_{\text{运算}} + T_{\text{MEM}} + T_{\text{I/O}}$

↑
优化：CPU转去执行其它程序 → **多任务**计算机系统

CPU时间— $T_{\text{CPU}} = I_N \times \text{CPI} \times T_C = \left(\sum_{i=1}^n I_i \times \text{CPI}_i \right) \times T_C$

n 为**指令系统**的指令种类数， CPI_i 为**第*i*种指令**所需时钟周期数；

I_N 为**程序**执行的总指令数(≠程序包含的指令数，但有关联)；

I_i 为第*i*种指令的执行次数， CPI 为**每条指令平均**所需时钟周期数；

T_C 为**时钟周期**，是机器主频的倒数

T_{CPU} 与**程序算法、编译程序、指令系统、 CPI_i 及 T_C** 均相关

特点—反映了**单任务计算机系统的软硬件总体性能**

例1—某计算机主频为2GHz，指令系统只有I类指令及II类指令，指令长度分别为1B及2B，指令执行时间分别为5个及8个时钟周期。程序A的大小为2MB，其中30%为I类指令；程序A执行时，所有指令中共有20%的指令需重复执行30次，在这些重复执行的指令中，有40%为II类指令。求完成程序A的CPU时间。

解：程序A包含的I类指令数= $(2\text{MB} \times 30\%) / 1\text{B} = 0.6 \times 2^{20}$ 条，

II类指令数= $(2\text{MB} \times 70\%) / 2\text{B} = 0.7 \times 2^{20}$ 条；

重复执行的指令数= $(0.6 \times 2^{20} + 0.7 \times 2^{20}) \times 20\% = 0.26 \times 2^{20}$ 条；

执行的I类指令数= $0.6 \times 2^{20} \times 80\% + 0.26 \times 2^{20} \times 60\% \times 30$ 条，

II类指令数= $0.7 \times 2^{20} \times 80\% + 0.26 \times 2^{20} \times 40\% \times 30$ 条；

$$T_{\text{CPU}} = 5.16 \times 2^{20} \times 5 \times T_{\text{C}} + 3.68 \times 2^{20} \times 8 \times T_{\text{C}}$$

$$= (25.8 + 29.44) \times 2^{20} \times T_{\text{C}} = 55.24 \times 2^{20} \div (2 \times 10^9)$$

$$\approx 29\text{ms}$$

***吞吐量：** 又称吞吐率，指单位时间内能处理的工作量，
即 吞吐量 = n 个任务的总工作量 \div n 个任务的总时间

特点—反映了多任务计算机系统的软硬件总体性能

表示—因工作量无统一定义，通常用MIPS及MFLOPS代替

△MIPS(每秒百万次指令)

$$\text{MIPS} = \frac{\text{程序执行的指令条数}}{\text{程序执行时间} \times 10^6} \rightarrow \text{MIPS}_{\max} = \frac{1}{\text{CPI} \times T_C \times 10^6} = \frac{\text{主频} f}{\text{CPI} \times 10^6}$$

缺点： 不能反映指令系统功能强弱(可用相对MIPS法)

△MFLOPS(每秒百万次浮点运算)

$$\text{MFLOPS} = \frac{\text{程序执行的浮点操作次数}}{\text{程序执行时间} \times 10^6} \rightarrow \text{MFLOPS}_{\max} = \frac{\text{时钟频率} f}{\text{CPI} \times 10^6}$$

缺点： 只反映了浮点操作能力，不能反映系统整体性能

***存储系统容量**：存储系统能存储的二进制字的总位数。

* **RASIS特性**：可靠性、可用性、可维护性、完整性和安全性。

- **常用衡量指标**：平均无故障时间**MTBF**（可靠性）、平均修复时间**MTTR**（可维护性）。

***兼容性**：同一个软件不加修改就可以在两台计算机上运行。

2.4.2 计算机系统性能设计

1、冯·诺依曼模型计算机的性能瓶颈

- ***CPU-MEM瓶颈**：CPU与MEM的速度差距越来越大
CPU速度—4~5倍/3年，得益于VLSI技术、器件技术；
MEM速度—1~2倍/3年，受限于容量-速度-成本的矛盾。
解决方案—需从结构与技术方面解决两者速度匹配问题
- ***指令串行执行瓶颈**：指令逐条执行导致性能受限
解决方案1—快速串行处理，性能受器件技术的限制，改进效果不佳；
解决方案2—并行处理，性能受串行程序指令间相关的限制，需从结构与技术方面解决指令间相关问题

2、性能平衡设计

重点解决CPU-MEM瓶颈，减少访存延迟、或提高访存效率

***增设Cache：**减少访存延迟，减少访问主存频率；

***增加总线数据宽度：**提高访存效率，以减少访存次数；

***采用多级总线：**主存连接在最快速总线上，减少访存延迟

3、CPU性能设计

重点解决指令串行执行瓶颈，提高并行性、或减少并行性能的损失

***开发并行技术：**操作级流水→操作级并行→指令级并行；

(流水线) (SSE) (超标量流水)

***采用数据流技术：**按序执行→乱序执行；

(程序顺序) (指令就绪顺序)

***使用转移预测技术：**按预测方向先执行，猜错时再回头执行

4、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。
- 常采用不同层次的基准测试程序对计算机系统进行比较客观的性能评价。
 - 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。

第2章作业： P22—4、 6、 7、 9