

体系结构 第一章

by xjc

第一章 计算机系统结构的基本概念

1.0 单周期和多周期CPU比较

同步计算机：计算机系统有一个**时钟**，计算机所有的动作都**以这个时钟为基准**。

单周期CPU：

- 指一条指令的执行在一个时钟周期内完成，然后开始下一条指令的执行
- 即一条指令用一个时钟周期

多周期CPU：

- 指一条指令的执行在多个时钟周期内完成
- 将原来一个时钟周期，分成多个更小的时钟周期，在此期间内完成一条指令

1.1 引言

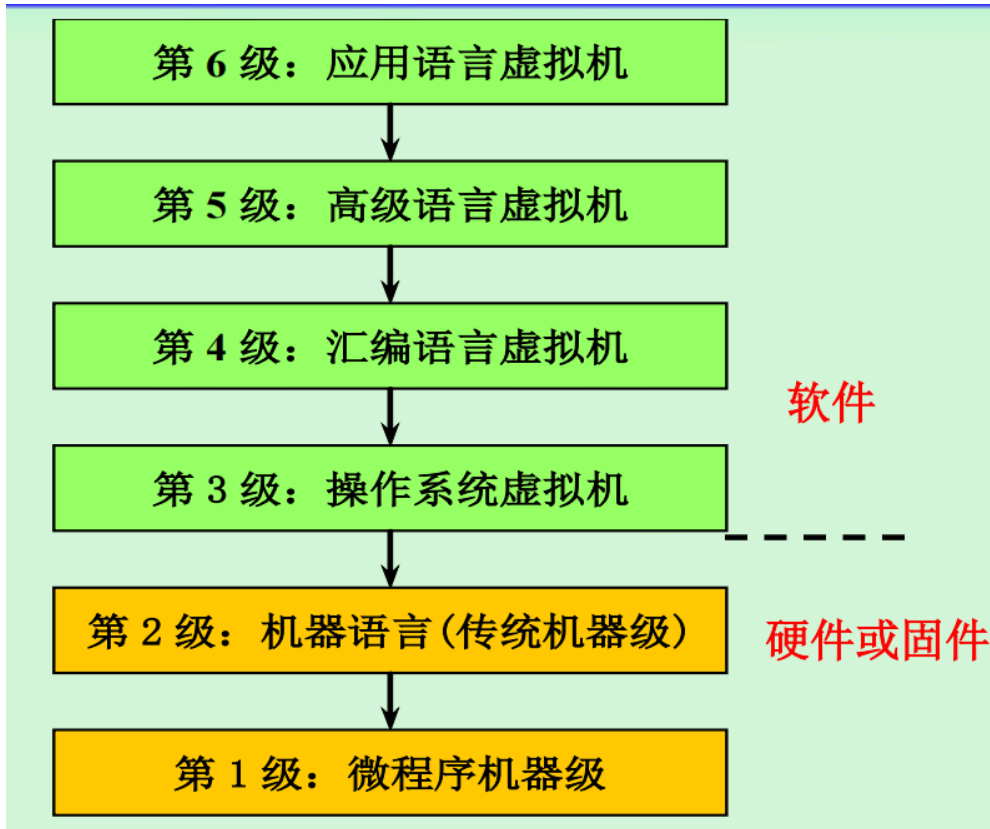
1. 第一台**通用电子计算机**诞生于**1946年**
2. 计算机技术飞速发展的原因（2个，硬软件）
 - **硬件**：计算机**制造技术**的发展
 - **软件**：计算机**体系结构**的创新
3. 计算机的四个发展阶段（时间点，了解即可）
 - **1946开始25年**：两种因素
 - **20世纪70年代~80年代**：**大规模集成电路**和**微处理器**出现，**集成电路**为代表的制造技术发展
 - **20世纪80年代中期**：**RISC结构**出现，系统结构不断更新
 - **2002年以来**：存在**功耗**问题、**指令级并行性**难以开发、**存储器访问速度**提高缓慢
4. 计算机系统结构的重大转折（一句话）
从单纯依靠**指令级并行**转向**开发线程级并行和数据级并行**

1.2 计算机系统结构的概念

1.2.1 计算机系统的层次结构

1. **计算机系统**：计算机系统=硬件/固件+软件

2. 计算机系统的层次结构：计算机语言从低级向高级发展，从计算机语言角度将计算机系统按功能划分为多级层次结构（每层以一种语言为特征）



3. 虚拟机：由软件实现的机器
4. 语言实现的两种技术
- 翻译：N+1级程序全部→N级程序后，执行N级程序
 - 解释：N+1级指令一条一条译码→一条一条执行N级指令
解释时间>翻译时间，但占用存储空间少

1.2.2 计算机系统结构的定义

1. 经典定义：程序员所看到的计算机属性，即概念性结构与功能特性
2. Amdahl提出的系统结构：传统机器语言级程序员所看到的计算机属性，包括指令集结构、组成、硬件
3. 计算机系统结构的实质：确定计算机系统中软硬件的界面（界面之上是软件实现的功能，界面之下是硬件/固件实现的功能）

1.2.3 计算机组成和计算机实现

1. 计算机系统结构：计算机系统的软硬件的界面，即机器语言程序员所看到的计算机属性
(我建议你可以把1.2.2全写)
2. 计算机组成：计算机系统结构的逻辑实现（包含物理机器级中的数据流和控制流的组成以及逻辑设计等）
3. 计算机实现：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)

层层递进的关系, 一种体系结构可以有多种组成, 一种组成可以有多种物理实现

4. 系列机: 由同一厂家生产的具有相同系统结构、但具有不同组成和实现的一系列不同型号的计算机

1.2.4 计算机系统结构的分类

(1) 冯氏分类法

1. 分类依据: 依照系统的最大并行度分类
2. 最大并行度的概念: 计算机系统在单位时间内能够处理的最大的二进制位数
最大并行度不是机器字长, 在机器字长基础上还要乘以同一时间能够处理的字数

(2) Flynn分类法

1. 分类依据: 指令流和数据流 的多倍性
2. 部分概念
 1. 指令流: 执行的指令序列
 2. 数据流: 指令流调用的数据序列
 3. 多倍性: 系统受限部件上, 同时同一执行阶段的指令或数的最大数目
3. 四种分类:
 1. 单指令流单数据流 (SISD)
 2. 单指令流多数据流 (SIMD)
 3. 多指令流单数据流 (MISD)
 4. 多指令流多数据流 (MIMD)

1.3 定量分析技术

1.3.1 以经常性事件为重点

定义: 对经常发生的情况重点对待, 采用优化方法的原则进行选择, 以得到更多的总体上的改进优化的概念:

- 分配更多的资源
- 达到更高的性能
- 分配更多的电能, 等

1.3.2 Amdahl定律

1. 定义: 加快某部件执行速度所能获得的系统性能加速比, 受限于该部件执行时间/系统总执行时间
2. 两个因素:
可改进比例 ≤ 1 : 加速时间/总时间

部件加速比 ≥ 1 : 改进前时间/改进后时间

3. 计算公式 (基本原理+推论)

$$\text{公式1: 加速比} = \frac{\text{系统性能}_{\text{改进后}}}{\text{系统性能}_{\text{改进前}}} = \frac{\text{总执行时间}_{\text{改进前}}}{\text{总执行时间}_{\text{改进后}}}$$

$$\text{公式2: 加速比} S = \frac{1}{(1 - \text{可改进比例}) + \frac{\text{可改进比例}}{\text{部件加速比}}}$$

4. 例题 (2题, 见ppt p37 p38)

1.3.3 CPU性能公式

1. 三个影响性能的参数

1. **时钟周期时间**: 一个周期的运行时间
2. **CPI**: Count per instruction, 每个指令的周期
3. **IC**: Instruction count, 指令的个数

2. 基本计算公式

$$CPU\text{时间} = \text{执行程序时钟周期数} \times \text{时钟周期时间} = CPI \times IC \times \text{时钟周期时间}$$

3. 细化的计算公式

CPI_i : 第 i 种指令的执行周期数; IC_i : 第 i 种指令的个数

$$CPU\text{时钟周期数} = \sum_{i=1}^n (CPI_i \times IC_i)$$

$$\text{有效}CPI = \frac{\text{时钟周期数}}{IC} = \sum_{i=1}^n (CPI_i \times \frac{IC_i}{IC})$$

4. 其他计算公式, 注意书上例题第一次课p42

$$MIPS = \frac{f}{CPI}$$

$$\text{程序执行时间} = \frac{\text{总时钟周期数}}{f}$$

1.3.4 程序的局部性原理

定义: 程序执行时所访问的存储器地址分布不是随机的, 而是相对的簇聚

两种局限性:

1. 程序的**时间局限性**: 程序即将用到的信息很可能就是目前正在使用的信息
2. 程序的**空间局限性**: 程序即将用到的信息很可能与目前正在使用的信息在空间上相邻或者临近

1.5 计算机系统结构中并行性的发展

1.5.1 并行性的概念

1. 定义：计算机系统在同一时刻或者同一时间间隔内进行多种运算或操作
 1. 同时性：两个或两个以上的事件在同时时刻发生
 2. 并发性：两个或两个以上的事件在同一时间间隔内发生
2. 分类
 1. 处理数据的角度，并行性等级分类（从低到高）
 1. 字串位串：同时一个字的一位
 2. 字串位并：同时一个字的全部位
 3. 字并位串：同时许多字的同一位（位片）
 4. 全并行：同时许多字的多位
 2. 执行程序的角度，并行性等级分类（从低到高）
 1. 指令内部并行：单指令内微操作
 2. 指令级并行：并行两条以上指令
 3. 线程级并行：并行两个以上线程
 4. 任务级或过程级并行：并行两个以上过程或任务
 5. 作业或程序级并行：并行两个以上作业或程序

1.5.2 提高并行性的技术途径

三种途径

1. 时间重叠：多个处理过程在时间上相互错开，轮流重叠地使用同一套硬件设备的各个部件
2. 资源重复：重复设置硬件资源
3. 资源共享：多个任务按一定时间轮流使用同一套硬件设备

1.5.3 单机系统中并行性的发展

(1) 主导作用——时间重叠

实现时间重叠的基础：部件功能专用化

(2) 资源重复

应用：多体存储器、多操作部件（实现指令级并行）、阵列处理机（并行处理机）

(3) 资源共享

原理：资源共享的概念实质上是，使用单处理机模拟多处理机的功能，形成所谓虚拟机的概念

应用：分时系统

1.5.4 多机系统中并行性的发展

(1) 3种类型的多处理机

遵循时间重叠、资源重复、资源共享原理，发展出3种不同的多处理机

同构型多处理机、异构型多处理机、分布式系统

(2) 耦合度

定义：反映多机系统中各机器之间的物理连接的紧密程度和交互作用能力的强弱

分类：紧密耦合系统（直接耦合系统，可共享主存）、松散耦合系统（间接耦合系统，可共享外存）

(3) 功能专用化（实现时间重叠）

专用外围处理机、专用处理机、异构型多处理机系统

(4) 机间互联

容错系统、可重构系统、同构型多处理机系统