

CH1 计算机概要与技术

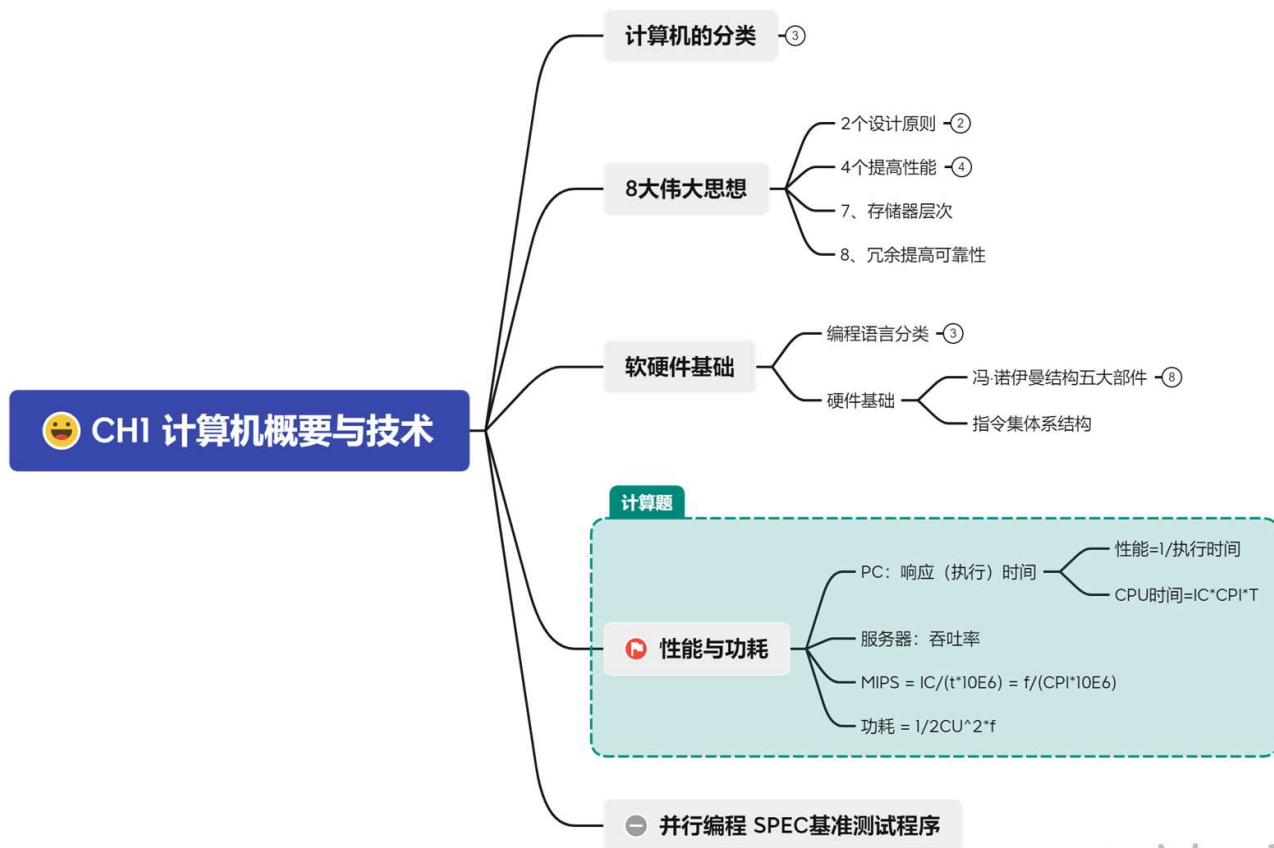
课程基于

《计算机组成与设计：硬件/软件接口》5e

Patterson & Hennesy 著

站 翼云图灵

章节导图



第一部分

计算机的分类

计算机的分类

传统3类：PC、服务器、嵌入式

后PC时代2类：WSC、PMD

储存容量单位

B站 翼云图灵

计算机的分类

传统上，按照应用，计算机可以分为：

个人计算机（PC）

服务器

适用于执行大负载任务，多用户通过网络并行访问

高端服务器被称为超级计算机

嵌入式计算机是数量最多的一类计算机

进入后PC时代：

往小的方向发展为：个人移动设备（PMD）

往大的方向发展为：仓储规模计算机（Warehouse-Scale Computer, WSC），大型数据中心的云计算服务器



存储容量

数量级前缀

两级之间有 2^{10} (1024) 倍的数量关系

千K	兆M	吉G	太T	拍P	艾E	泽Z	尧Y
1024	1024e2	1024e3	1024e4.....				
thousand	million	billion	trillion.....				

B (Byte) 字节

b (bit) 位

1字节=8位

字节常用于储存容量，位常用于传输速度

i5+GTX3060+8G+512G~B/b?

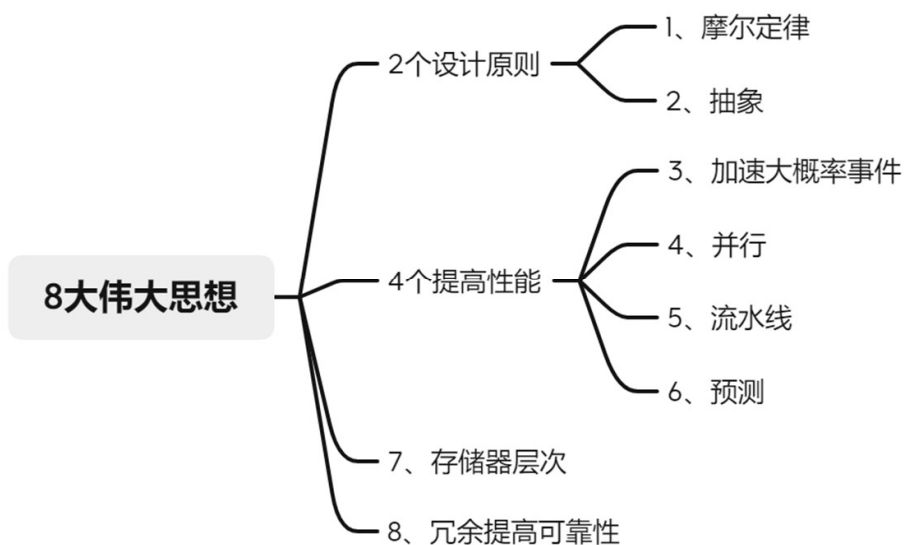
100兆的宽带B/b?



B站 冀云图灵

第二部分

计算机系统结构中的 8个伟大思想



2个设计原则

1、摩尔定律

单芯片上的集成度（单位面积硅片能容纳的晶体管数量）每18个月翻一番

提出于1965

芯片设计要依据设计开始/完成? 时的工艺水平

2、抽象

省略低层次的细节，提供简化的模型



站 翼云图灵

4个提高性能的方法

3、加速大概率事件

一台计算机绝大部分时间用来打游戏，改进游戏性能还是其他性能更能提高整体性能？

定量分析：Amdahl定律

改进后的执行时间 = 改进影响的执行时间 / 改进量 + 改进不影响执行时间

$$t_{\text{改进}} = t_{\text{相关}}/S + t_{\text{无关}}$$

4、并行

同时发生，典型例子是多核处理器

5、流水线

类比工业中的流水线

6、预测

如果

- 1) 预测失误后恢复的代价不高；
 - 2) 预测有相当的准确率
- 则可以猜测哪些操作会发生，并提前开始执行



B站 翼云图灵

存储器层次与冗余

7、存储器层次

(见第三部分 软硬件接口)

8、冗余提高可靠性

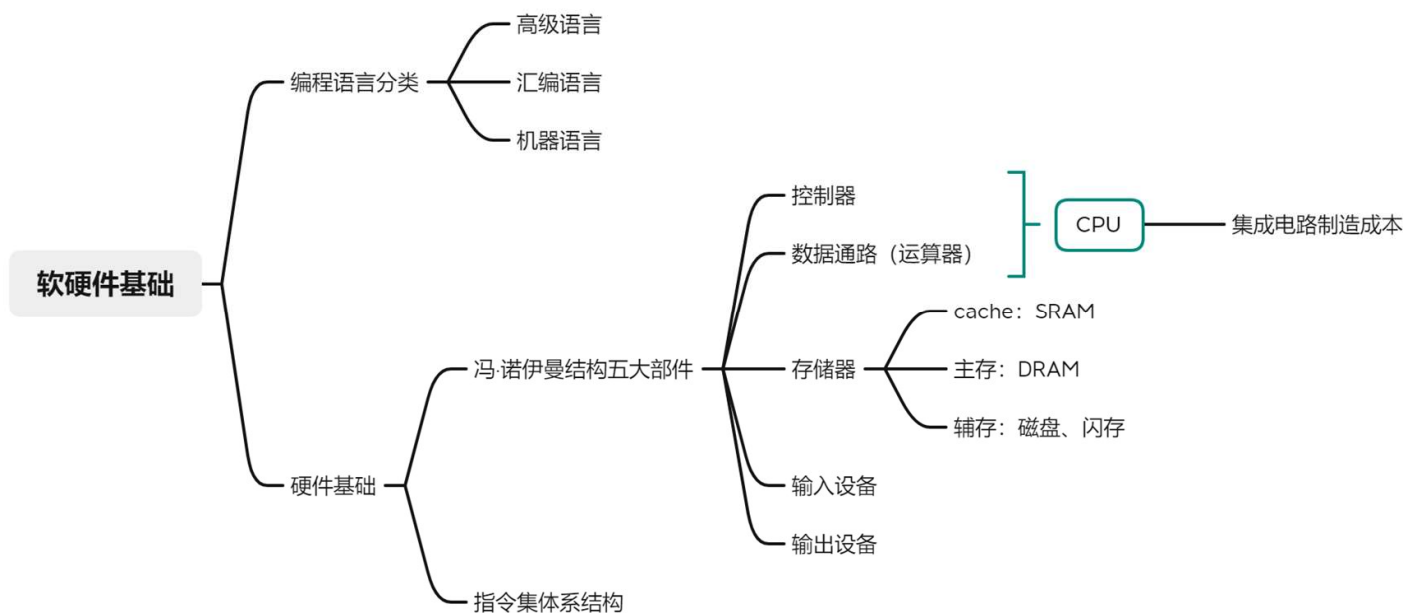
物理器件的“备胎”，失效了就换上



B站 翼云图灵

第三部分

软硬件基础



编程语言分类



机器指令

00000011101001001011001111110010

汇编语言和由它汇编出的机器语言一一对应



汇编语言

指令助记符、寄存器标号

addi \$s1, \$s1, 1

一条高级语言语句可以编译出好几条汇编语言

高级语言

语句接近自然语言，提高了程序的可移植性

cnt = cnt + 1;

机器语言依赖于机器，所以被称作低级语言

汇编语言呢？高级语言呢？

冯·诺依曼结构

冯结构中的五大经典部件：

外围的①输入设备、②输出设备，合称I/O设备

输入设备：键盘鼠标、扫描仪

输出设备：显示器、打印机

混合I/O设备：触摸屏、网卡

存放程序和数据③存储器

缓存cache ~ SRAM

主存/内存 ~ DRAM

辅存/外存/二级存储器 ~ 闪存、磁盘

*只有放入内存的程序才能被CPU执行



快但是贵，容量很小（3MB）

比较快也比较贵，容量中等（8~16GB）

慢但便宜，容量大（基本都是512GB起步）

冯·诺依曼结构

组成CPU的两大部件：④运算器（数据通路）、⑤控制器

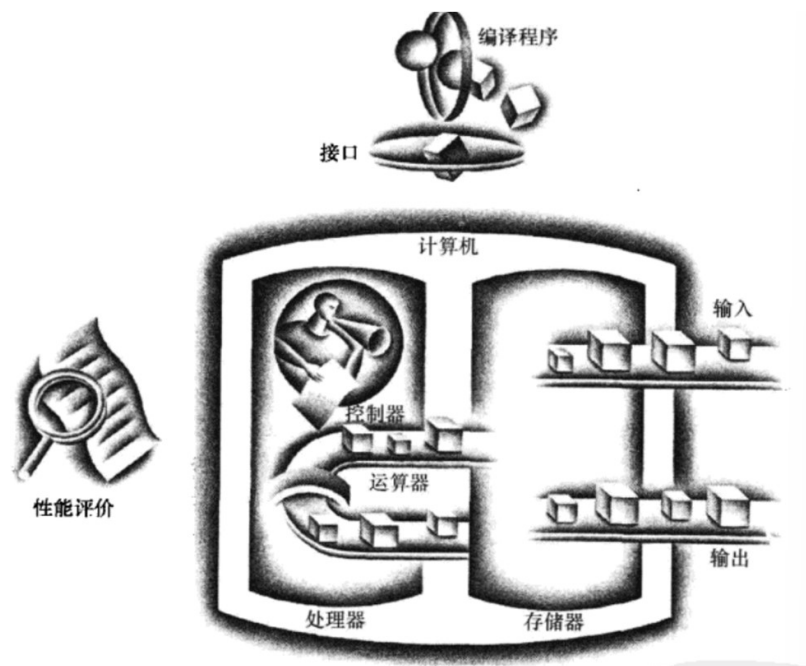


图 1-4 组成计算机的五个典型部件

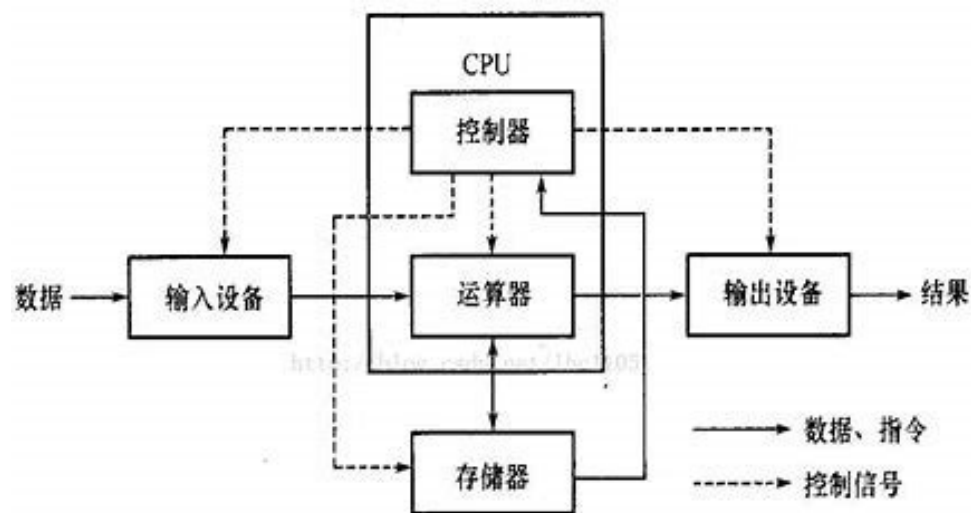


图 1.1 冯·诺依曼结构(以运算器为中心)

存储器层次结构

CPU的主频在3.5GHz，内存条的主频2666MHz，通常是谁等待谁？

解决方法：存储器层次（思想7）

顶端：快贵小 → 底端：慢廉大

第一级：cache（和cash同音），采用SRAM技术

第二级：主存/内存，采用DRAM技术

SRAM、DRAM断电后数据很快消失，为易失性存储器

第三级：辅存，过去常用磁盘，现逐渐被闪存代替

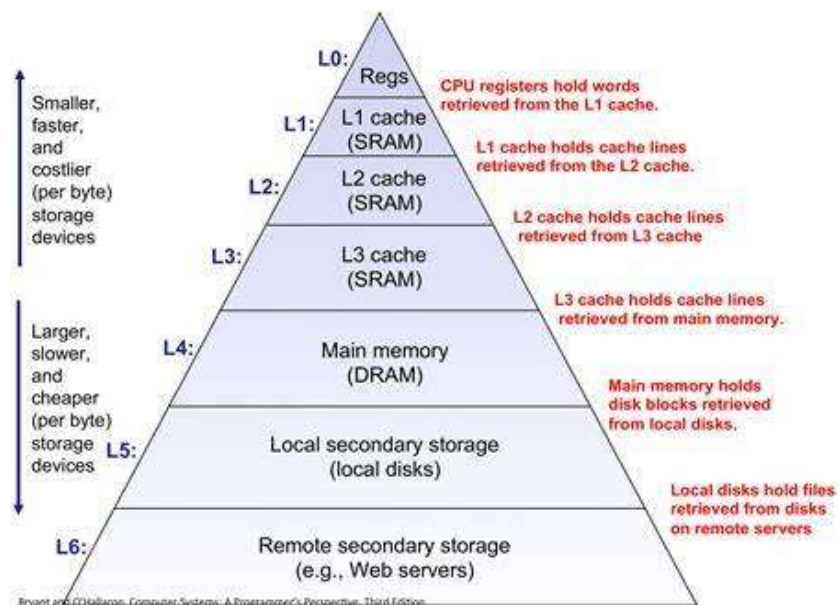
磁盘、闪存断电后不丢失数据，为非易失性存储器

*cache位于CPU中

在CPU中塞入cache，好处是：

1) 物理距离近，和两大部件通信迅速

2) 只要cache能把CPU即将要用的程序、数据，从内存中复制过来，就能缓解速度矛盾



指令集体系结构

苹果的M1系列芯片为什么性能高且续航持久？

CPU的指令集体系结构（架构，ISA）根本上不一样

即，可以执行的指令集合大不相同

一个x86，为了和二三十年前的软件保持兼容，指令只增不减，越来越庞大繁杂

一个ARM，设计思想主张精简，能效出色

教材介绍一个和ARM设计思路相近的架构：MIPS

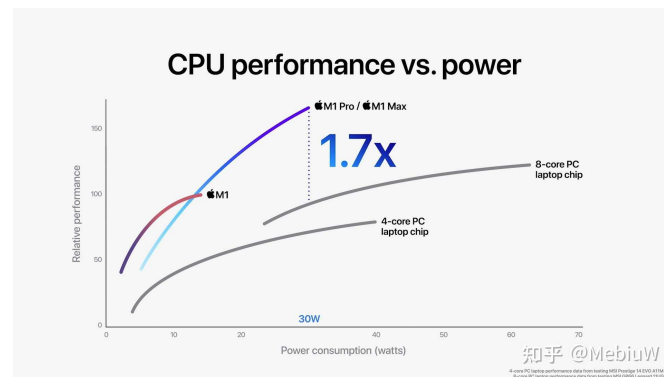
MIPS和ARM都属于RISC架构

（Reduced Instruction Set Computer，精简指令集计算机）

x86是典型的CISC架构

（Complex.....，复杂.....）

两位作者都是RISC技术的重要奠基人，对x86架构非常唾弃



B站 翼云图灵

集成电路制造

晶体管：受电信号控制的简单开关

集成电路 (Integrated Circuit, IC)：由许多个晶体管组成的芯片



制造具体过程不太重要，记住结论：

成品率越高、产量越高、芯片面积越小，芯片的制造成本越低

B站 翼云图灵

复习题

复习题是用来碎碎念回忆复习的，不是用来写的！

打*的问题在考试中不太重要或者比较复杂，请酌情跳过

- 1、传统上计算机分为哪三类？后PC时代有哪两类计算机？
- 2、KB MB GB之上有哪些存储容量单位？
- *3、计算机设计的八个伟大思想是？
- 4、Amdahl定律的公式为？
- 5、高级语言经过什么过程变成汇编语言？汇编语言又经过什么变成机器语言？
- 6、冯·诺依曼结构由哪五大基本部件组成？CPU主要有哪两大部件？还有少量的哪个部件？
- 7、存储器层次中，最快贵小的、其次的、最慢廉大的是什么？分别采用什么技术？
- 8、Intel和苹果的处理器的架构不同，是什么东西不同？
- 9、成品率、芯片面积、产量如何影响集成电路（IC）的制造成本？

第四部分

性能与功耗

计算题

性能与功耗

PC: 响应 (执行) 时间

性能=1/执行时间

CPU时间=IC*CPI*T

服务器: 吞吐率

MIPS = $IC / (t * 10E6) = f / (CPI * 10E6)$

功耗 = $1/2 CU^2 * f$

B站 翼云图灵

评价性能的指标

鲁大师、安兔兔这些跑分软件的原理？

SPEC CPU基准测试程序，是一个比较权威的跑分软件

个人计算机更看重**响应时间**（又称**执行时间**）：任务从开始到完成的时间

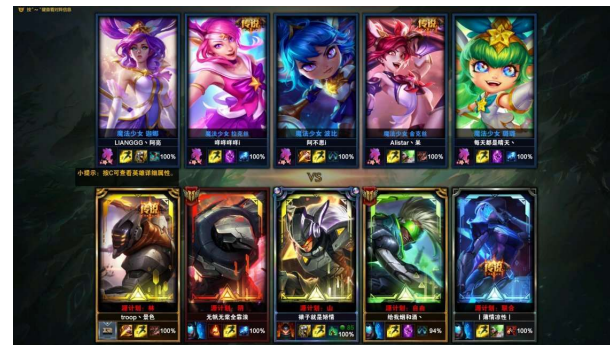
加载进入游戏需要XXX秒钟.....

服务器更看重**吞吐率**：单位时间内完成的任务数量

一秒钟能把网页发送给XXX个用户.....

课程主要讨论**CPU执行时间**

$$\text{性能} = 1 / \text{执行时间}$$



计算机A的性能是计算机B的几倍？→计算机A的执行时间是B的几分之一？

e.g. 运行同一个程序，计算机A用15秒，计算机B用10秒，谁的性能更高？它的性能比另外一台高多少？

B站 翼云图灵

CPU经典性能公式

CPU时钟周期长度 T (CPU干任何事的时间都是一个时钟周期的整数倍)
干一件事, 即执行一个程序, 需要多少个时钟周期?

$$\text{CPU执行时间} = \text{时钟周期数} \times \text{时钟周期长} \quad \text{CPUtime} = \text{cycles} \times T$$

任何程序都要以机器语言指令的形式, 才能被计算机看懂
所以, **执行程序本质上就是执行机器指令!**

所以, 时钟周期数 = 机器指令的条数 \times 每条机器指令占用的时钟周期数

*不同机器指令需要的周期数不尽相同

所以, 每条机器指令占用的时钟周期数 (cycles per instruction) 是一个平均值
即**总周期数/总指令数**

$$\text{时钟周期数} = \text{指令数} \times \text{平均每条指令所需周期数} \quad \text{cycles} = \text{IC} \times \text{CPI}$$

综上, 得到经典性能公式

站 翼云图灵

CPU经典性能公式

CPU执行时间 = 指令数 x CPI x 时钟周期长 CPUtime = IC x CPI x T

CPU的时钟频率f（主频）：一秒钟有多少个时钟周期

$$T = 1 / f$$

所以也能写成

CPU执行时间 = 指令数 x CPI / 主频 CPUtime = IC x CPI / f

主频的单位通常是吉赫兹GHz（G在此处代表 10^9 ）

e.g. 主频为4GHz的CPU，每个时钟周期的长度是多少秒？

某程序共有3亿条指令，其中1亿条的CPI为3，2亿条的CPI为6，整个程序的CPI是多少？

该CPU执行这个程序需要多长时间？

CPU参数表					
处理器型号	i7-9700k	i7-9700	i7-8700	i7-9700kF	i7-9700F
光刻	14 nm	14 nm	14 nm	14 nm	14 nm
内核数	8	8	6	8	8
基本频率	3.60 GHz	3.00 GHz	3.20 GHz	3.60 GHz	3.00 GHz
缓存	12MB	12MB	12MB	12MB	12MB
TDP	95 W	65 W	65 W	95 W	65W
内存类型	DDR4-2666	DDR4-2666	DDR4-2666	DDR4-2666	DDR4-2666
处理器显卡	英特尔® 超核显卡 G30			无 需配置独立显卡	
傲腾™ 内存支持	是				

B站 翼云图灵

CPU经典性能公式为什么靠谱

经典性能公式考虑了指令数IC、平均每条指令需要周期数CPI、时钟周期长度T（也就是主频f的倒数）**全部三个因素**使用**CPU执行时间**这个靠谱的指标来衡量性能

使用什么**编程语言**和**编译程序**、采取什么**算法**，影响指令数和CPI两个因素
采用什么**指令集体系结构**，影响指令数、CPI、主频全部三个因素
它们都会影响CPU性能

还有一个衡量性能的指标**MIPS** (million instructions per second, 百万指令每秒)

$$\text{MIPS} = \text{IC} / (\text{CPUtime} \times 10^6) = f / (\text{CPI} \times 10^6)$$

因为只考虑了主频（周期长）和CPI两个指标，很受作者唾弃

功耗

在移动设备和WSC中，电能是关键资源

但随着CPU集成度越来越高，晶体管的数量剧增，功耗越来越大、散热越来越难

CMOS（互补型金属氧化半导体）是当前占统治地位的集成电路技术

其中，晶体管的能耗主要来自动态能耗，即晶体管打开/关闭的短暂动态过程

即0→1或1→0的过程

一个晶体管经历一次0 → 1 → 0或1 → 0 → 1的开关过程

能耗 = 负载电容 x 电压²

开关频率与CPU主频相关

所以，功耗公式为

$$\text{功耗} = \frac{1}{2} \times \text{负载电容} \times \text{电压}^2 \times \text{开关频率} \quad P = \frac{1}{2} C U^2 * f$$

并行化 基准测试程序

CPU设计面临**功耗墙**问题

从2002年起，单核处理器性能增长速度从每年约52%下降到22%，功耗问题“功不可没”
因此，从单处理器向**多核处理器**转变成为近年计算机提升性能的主要手段

同时，硬件变成了多处理器

要求程序员重写代码去支持CPU的**并行**（同时执行）特性

显式并行编程应运而生，但仍有很多需要克服的问题

设有一个单核处理器A，还有一个每个核心都和A相同的四核处理器B，
它们运行同一个程序，B的CPU时间能减少到A的1/4吗？

SPEC是多家计算机销售商共同资助的组织，目的是为现代计算机提供基准测试程序

SPEC CPU基准测试程序用于测试CPU性能，取各项指标得分的几何平均数

站 翼云图灵

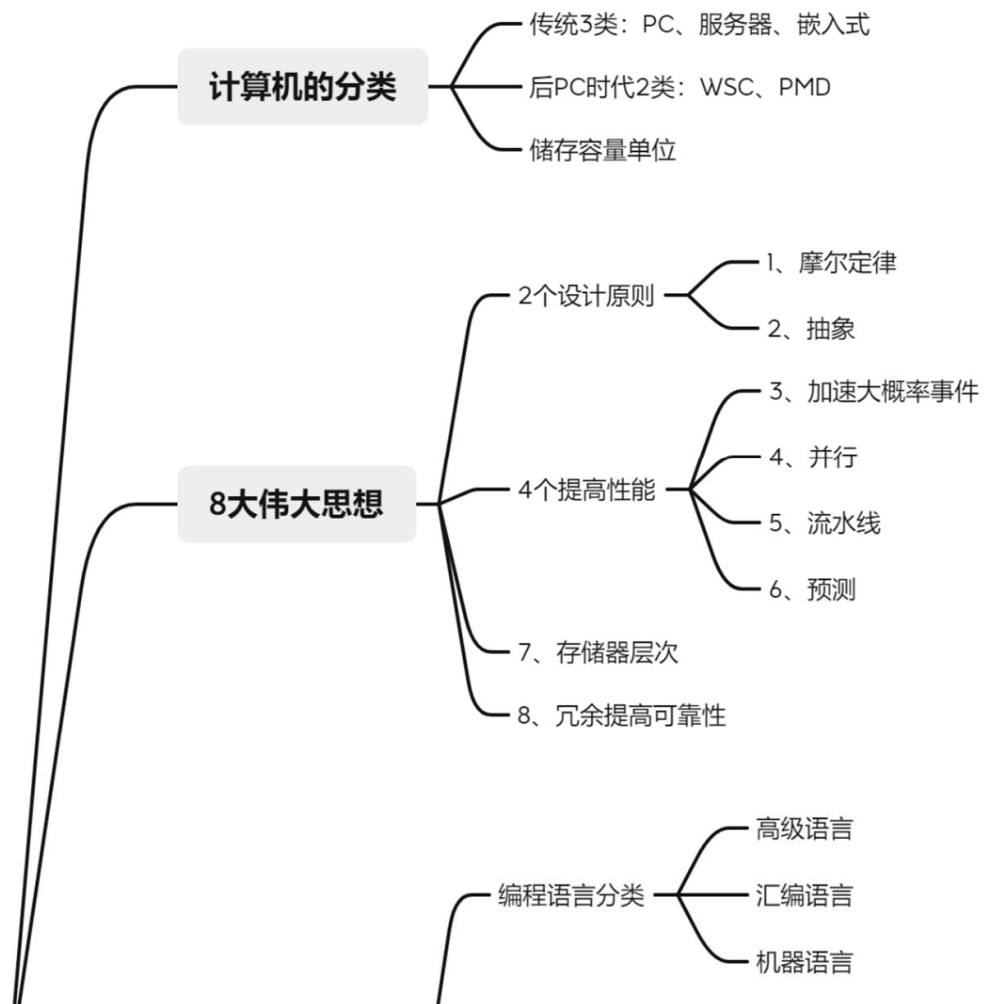
复习题

- 1、PC和服务端分别更看重哪项性能评价指标？
- 2、CPU经典性能公式计算的是什么指标？公式是什么（注意有两个联系密切的版本）？
- *3、我们是如何得出经典性能公式的？
- 4、哪四个东西会通过影响性能公式中的某些项，从而影响性能？
- 5、MIPS怎么计算？这个性能指标为什么不靠谱？

全章复习

CH1 计算机概要与技术

B站 翼云图灵



😄 CH1 计算机概要与技术

软硬件基础

硬件基础

冯·诺伊曼结构五大部件

控制器

数据通路 (运算器)

存储器

输入设备

输出设备

cache: SRAM

主存: DRAM

辅存: 磁盘、闪存

CPU

集成电路制造成本

指令集体系结构

计算题

性能与功耗

PC: 响应 (执行) 时间

性能=1/执行时间

CPU时间=IC*CPI*T

服务器: 吞吐率

MIPS = IC/(t*10E6) = f/(CPI*10E6)

功耗 = 1/2CU^2*f

$$\begin{aligned} \text{mps} &= \frac{IC}{t \times 10^6} \\ &= \frac{f}{CPI \times 10^6} \end{aligned}$$

$$\frac{1}{2} CU^2 f$$

并行编程 SPEC基准测试程序

2024 6 4 10 16