讲座2

时钟:任何可以开关的东西;用于计时和同步

时钟周期:例如: 100 MHz 总线频率:每秒 1 亿个周期: 100 * 10^6

1 GHz 中央处理器 每秒 10 亿次上升和下降边缘

每纳秒一个上升沿

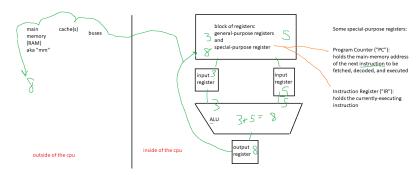
我们可能可以每纳秒开始一条指令

常规有线电视节目的时钟周期是 30 分钟……这意味着节目可以每半小时开始。这并不意味着节目时长为半小时。

电影院的 电影周期 是 5 分钟……这意味着放映可以在五分钟时开始 (例如 7:05, 7:10, 等等)。 这并不意味着放映时间是 五分钟。



冯·诺依曼数据通路: CPU 内部的CPU



第 63-65 页:加快指令执行的愿望清单:

- 1.所有指令应由硬件直接执行 复杂且昂贵
- 2. 尽可能快速/频繁地发布指令物理限制,复杂,昂贵
- 3.说明应该易于解码相同大小的说明,向后兼容性
- 4.只有 LOAD 和 STORE 指令应该引用 mm无法避免访问主存储器
- 5. 提供大量注册 昂贵

总体目标:防止CPU饥饿

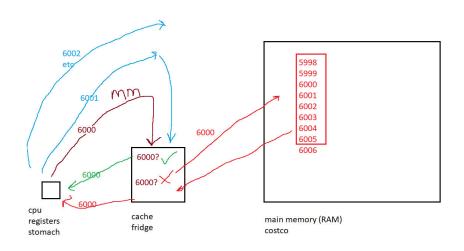
RISC 与 CISC 架构:

精简指令集计算机: 更少且更简单的指令

复杂指令集计算机: 越来越多的复杂指令: 例如。 直接在硬件中实现最常见的指令 (无需解

释),甚至是复杂的指令

我们的计算机是混合型的: 主要是 RISC, 部分是 CISC



- 1. the cpu wants one single word. it issues a request to MM for that word
- 2. the cache is checked first -if the cache has that word, then it's returned to the cpu -if the cache does not have that word, then that word PLUS ITS NEIGHBORS are fetched from mm, and the requested word is forwarded to the cpu

Programs tend to run in order. If the CPU is requesting word 6000 right now, words near 6000 are likely to be requested next/soon

Therefore bring to the cache, which has faster access times

在计算机编程中,程序往往按线性顺序运行。如果 CPU 请求单词 6000,那么<u>很可能</u>很快会<u>请求</u>与 6000 相关的单词。这被称为**空间局部性原则**。由于空间局部性原则,每当内存单词被请求时,邻近的单词也会被提取到缓存中,因为缓存能更快地为CPU提供单词。类比:如果你听到人们谈论老虎伍兹,你可以猜测他们可能会在不久的将来(2 分钟内?):高尔夫、作弊、头衔、运动、耐克。另一个类比:如果你在商店看到有人购物篮里有玉米饼,你可以猜测里面还会有其他物品:牛肉、香菜、奶酪、酸奶油

还有一个**时间局部性原则**,这意味着现在请求的同一个词<u>可能</u>会很快再次被请求<u>。由于时间局部性原则,缓存会尽量保留最近的词(</u>不让它们被驱逐)。类比:你听到走廊里有人在谈论老虎伍兹。你可以猜测他们在不久的将来会说的一些词:老虎伍兹。

定义:解释器:任何软件程序(即虚拟机),它执行另一个程序的指令。其输出是另一个更简单运行的程序。但是,这个过程需要时间。这比让硬件运行原始(更复杂、更高级)程序便宜,也比让硬件进行转换(解释)便宜。

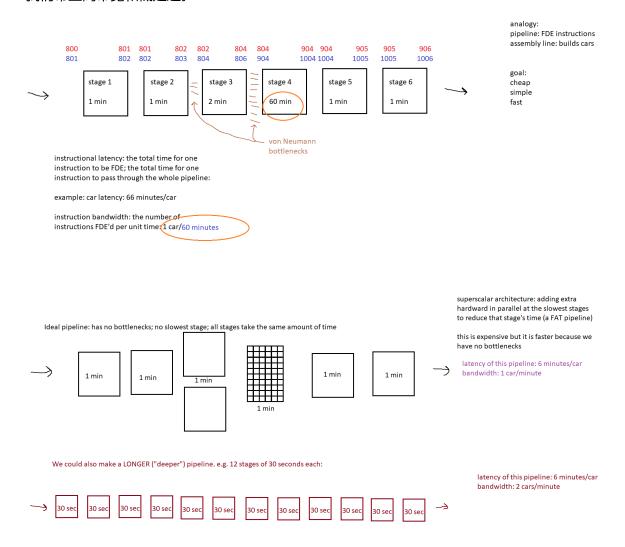
在 下节课之前,阅读教科书的第 65-69 页。还要阅读第 73-78 页

2

定义:管道是多个硬件单元在<u>并行</u>和串行中<u>获取、解码和执行</u>指令,以最大化指令带宽(以指令 FDE/秒为单位;例如 MIPS)。

指令延迟:从开始到完成获取、解码和执行一条指令的总时间(一条指令通过管道的总时间).

我们希望高带宽和低延迟。



Netflix 的延迟和带宽: 带宽高还是低? 高。

低带宽可能表现为:卡顿/缓冲、延迟、低分辨率

高延迟还是低延迟? 高延迟可能表现为: 开始观看电影的时间很长

过山车: 你想要高延迟还是低延迟? 你想在游乐设施上花一些时间……所以你想要更高的延迟。延迟是指在过山车上完成一个完整循环所需的时间。你想要高带宽还是低带宽? 更高的带宽意味着每单位时间内有更多的人完成游乐设施。通过拥有很多车(\$\$\$\$) 你可以获得高带宽和高延迟。

阵容=瓶颈

游乐园: 许多游乐设施排队时间长, 游乐时间短

迪士尼乐园:有一个非常长的游乐设施,排队时间却非常短:它是一个小小的世界。12-15分钟。只有花很多钱才能实现。

问题:

四个阶段的管道带宽和延迟分别是多少,阶段耗时为 0.01 毫秒、250,000 纳秒、30 微秒和 2 毫秒?

答案: 延迟: 0.01 毫秒 = 10 微秒 (微秒 = microseconds) + 250000 纳秒 = 250 微秒 +

30 微秒 = 30 微秒 + 2 毫秒 = <u>2000 微秒</u>

= 2290 美元

= 2.29 毫秒 / 指令

带宽: 1条指令/最慢阶段...1条指令/2毫秒

2