

9 高速缓冲存储器 (Cache)

任桐炜

2021年10月21日



南京大学
NANJING UNIVERSITY

教材对应章节

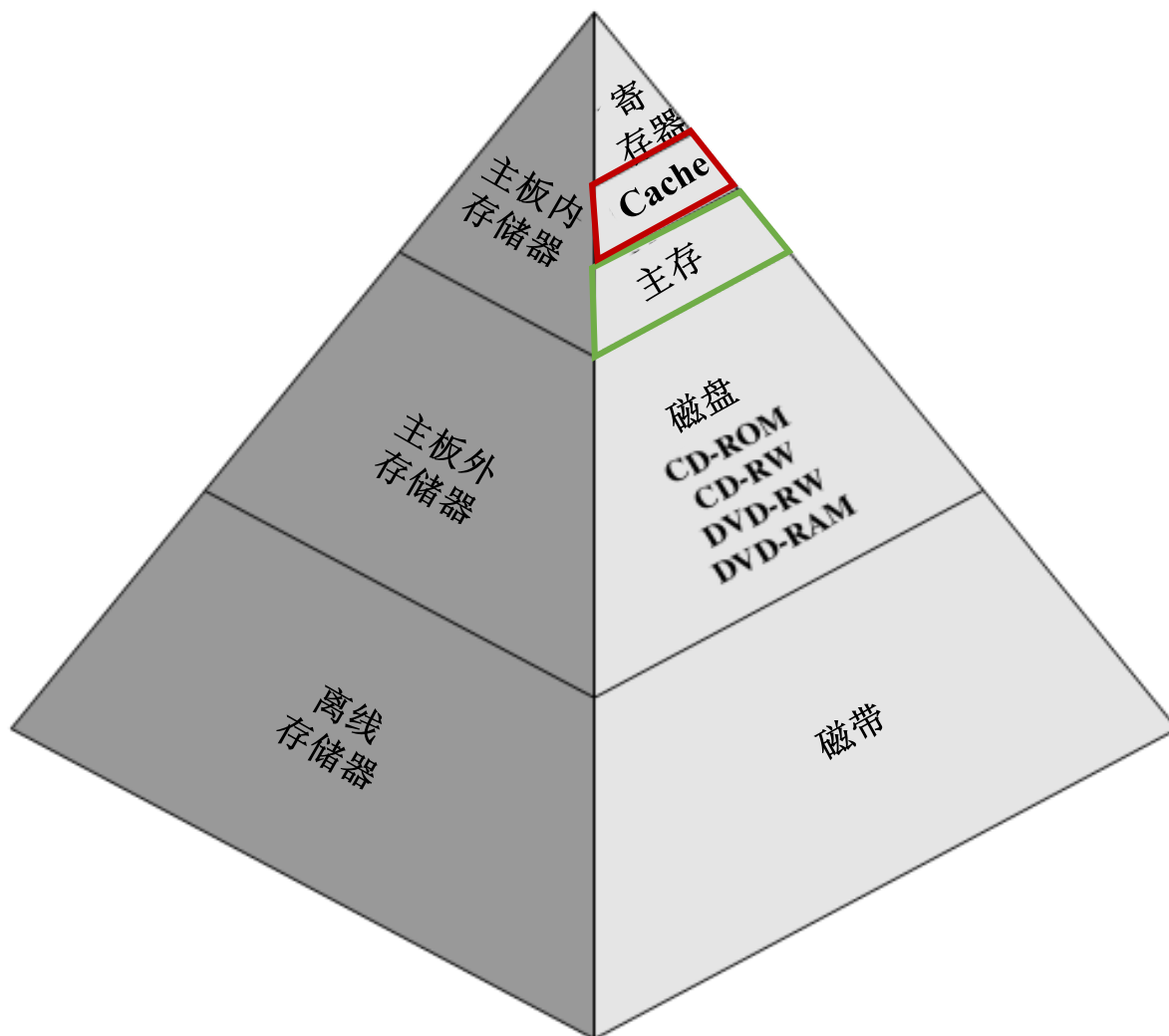


第7章 存储器分层体系结构



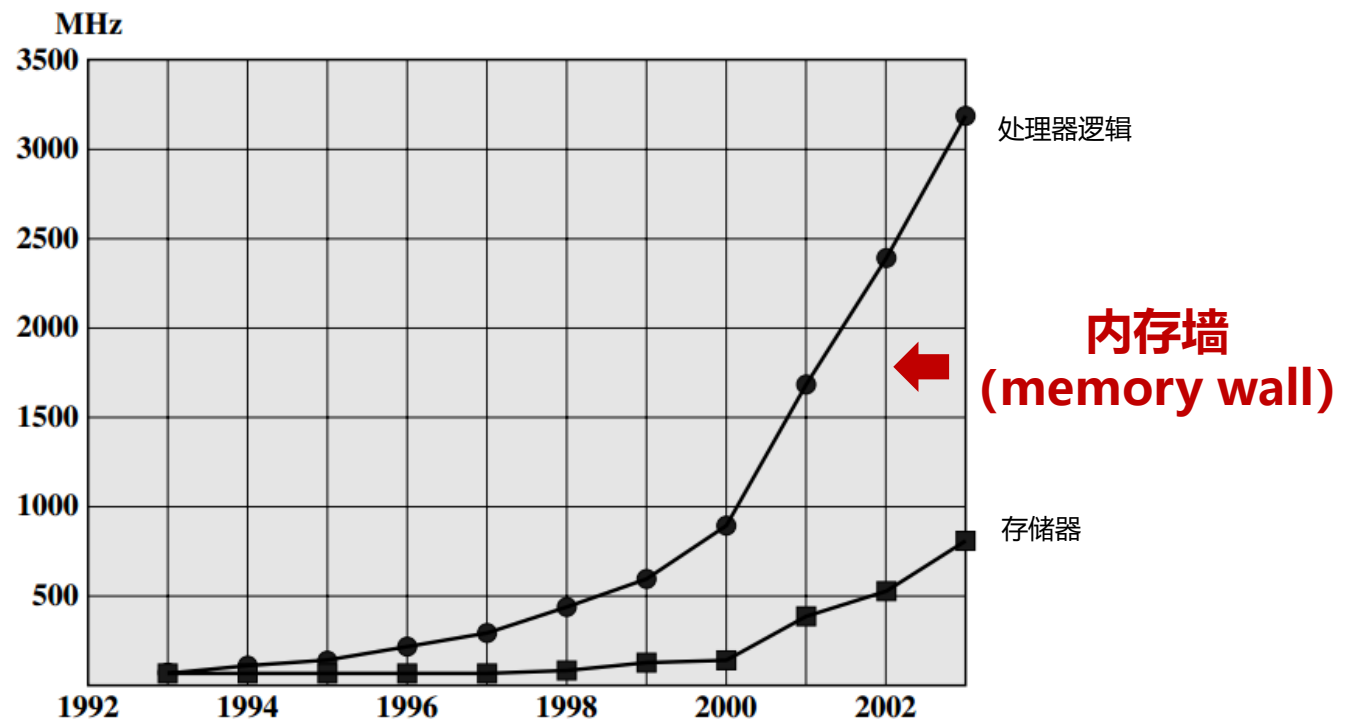
第4章 cache存储器

存储器层次结构



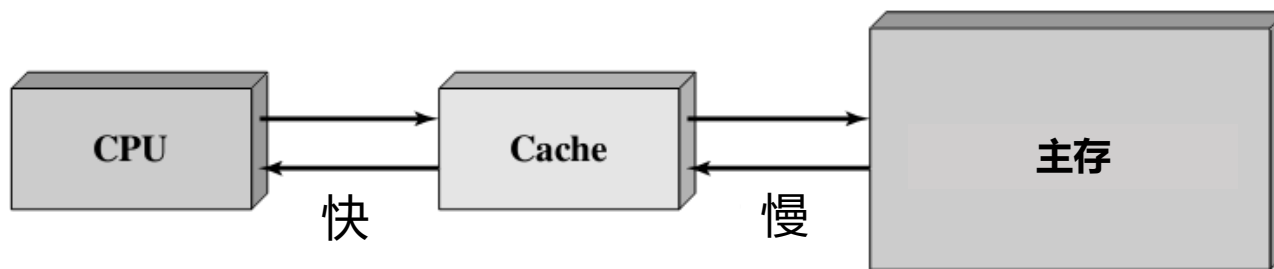
内存墙

- 问题：CPU的速度比内存的速度快，且两者差距不断扩大



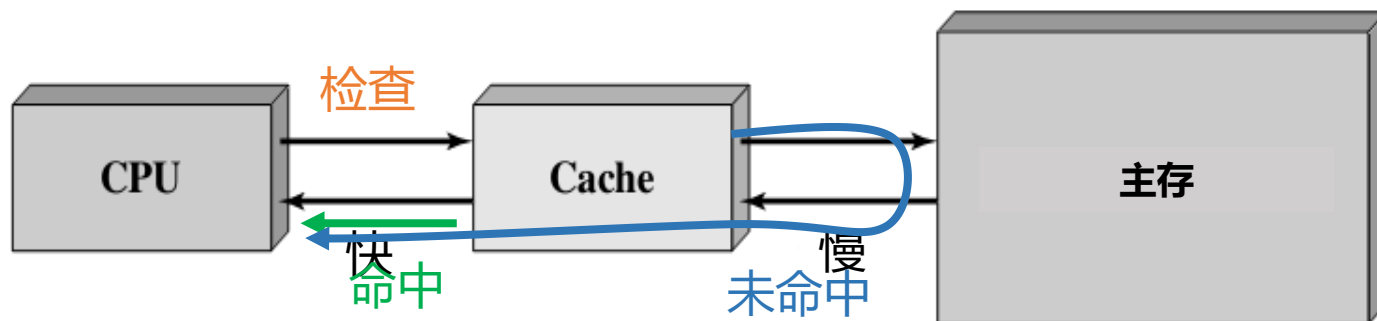
Cache的基本思路

- 解决内存墙带来的CPU和主存协作问题
 - 在使用主存（相对大而慢）之余，添加一块小而快的cache
 - Cache位于CPU和主存之间，可以集成在CPU内部或作为主板上的一个模块
 - Cache中存放了主存中的部分信息的副本



Cache的工作流程

- **检查 (Check)** : 当CPU试图访问主存中的某个字时, 首先检查这个字是否在cache中
- 检查后分两种情况处理:
 - **命中 (Hit)** : 如果在cache中, 则把**这个字**传送给CPU
 - **未命中 (Miss)** : 如果不在cache中, 则将主存中包含这个字**固定大小的块 (block)** 读入cache中, 然后再从cache**传送该字**给CPU



问题

- 如何判断是命中还是未命中?
- 如果未命中, 为什么不直接把所需要的字从内存传送到CPU?
- 如果未命中, 为什么从内存中读入一个块而不只读入一个字?
- 使用Cache后需要更多的操作, 为什么还可以节省时间?



命中和未命中的判断

- 冯·诺伊曼体系的设计
 - CPU通过**位置**对主存中的内容进行寻址，不关心存储在其中的内容
- Cache通过**标记 (tags)** 来标识其内容在主存中的对应**位置**



问题

- 如何判断是命中还是未命中?
- 如果未命中, 为什么不直接把所需要的字从内存传送到CPU?
- 如果未命中, 为什么从内存中读入一个块而不只读入一个字?
- 使用Cache后需要更多的操作, 为什么还可以节省时间?



程序访问的局部性原理

- 定义：
 - 处理器频繁访问主存中相同位置或者相邻存储位置的现象（维基百科）
- 类型
 - **时间局部性**：在相对较短的时间周期内，重复访问特定的信息（也就是访问相同位置的信息）
 - **空间局部性**：在相对较短的时间周期内，访问相邻存储位置的数据
 - **顺序局部性**：当数据被线性排列和访问时，出现的空间局部性的一种特殊情况
 - 例如：遍历一维数组中的元素



局部性原理的示例

- 时间局部性

```
int factorial = 1;  
for (int i = 2; i <= n; i++) {  
    factorial = factorial * i;  
}
```

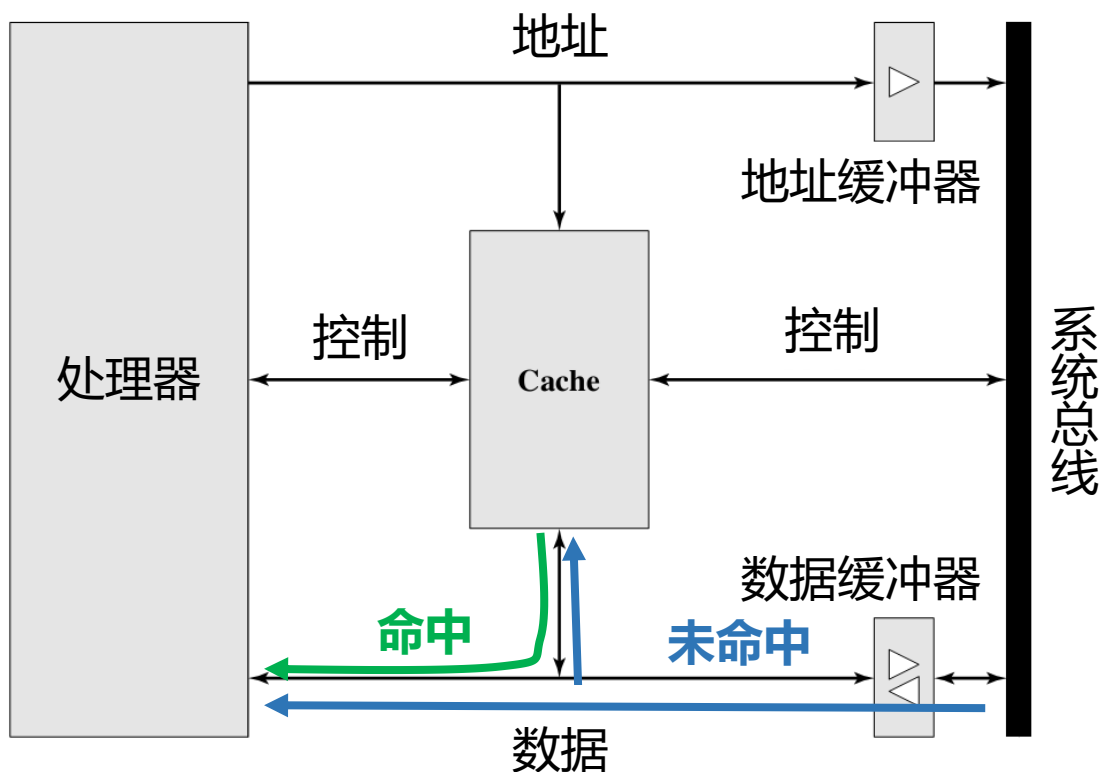
- 空间局部性

```
for (int i = 0; i < num; i++) {  
    score[i] = final[i] * 0.4 + midterm[i] * 0.3 + assign[i] * 0.2 + activity[i] * 0.1;  
}
```



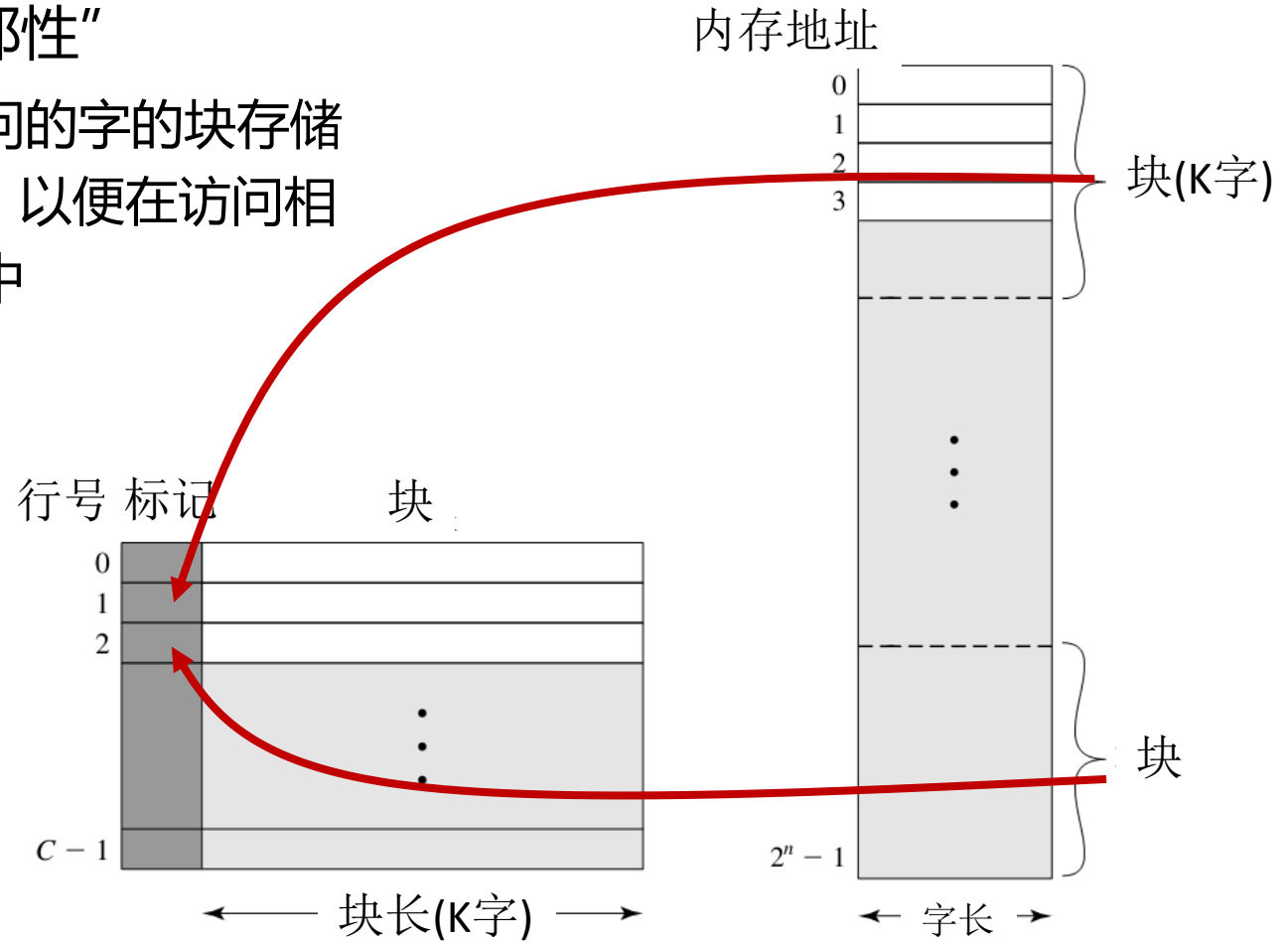
向Cache传送内容

- 利用“时间局部性”
 - 将未命中的数据在返回给CPU的同时存放在Cache中，以便再次访问时命中



传送块而不是传送字

- 利用“空间局部性”
 - 将包含所访问的字的块存储到Cache中，以便在访问相邻数据时命中



问题

- 如何判断是命中还是未命中?
- 如果未命中, 为什么不直接把所需要的字从内存传送到CPU?
- 如果未命中, 为什么从内存中读入一个块而不只读入一个字?
- 使用Cache后需要更多的操作, 为什么还可以节省时间?



平均访问时间

- 假设 p 是**命中率**, T_C 是cache的访问时间, T_M 是主存的访问时间, 使用cache时的平均访问时间为

$$\begin{aligned}T_A &= p \times T_C + (1 - p) \times (T_C + T_M) \\&= T_C + (1 - p) \times T_M\end{aligned}$$

- 命中率 p 越大, T_C 越小, 效果越好
- 如果想要 $T_A < T_M$, 必须要求

$$p > T_C / T_M$$

- 难点: cache的容量远远小于主存的容量



Cache的设计要素

- Cache容量
- 映射功能
- 替换算法
- 写策略
- 行大小
- Cache数目



谢谢

rentw@nju.edu.cn



南京大學
NANJING UNIVERSITY