

本节主题

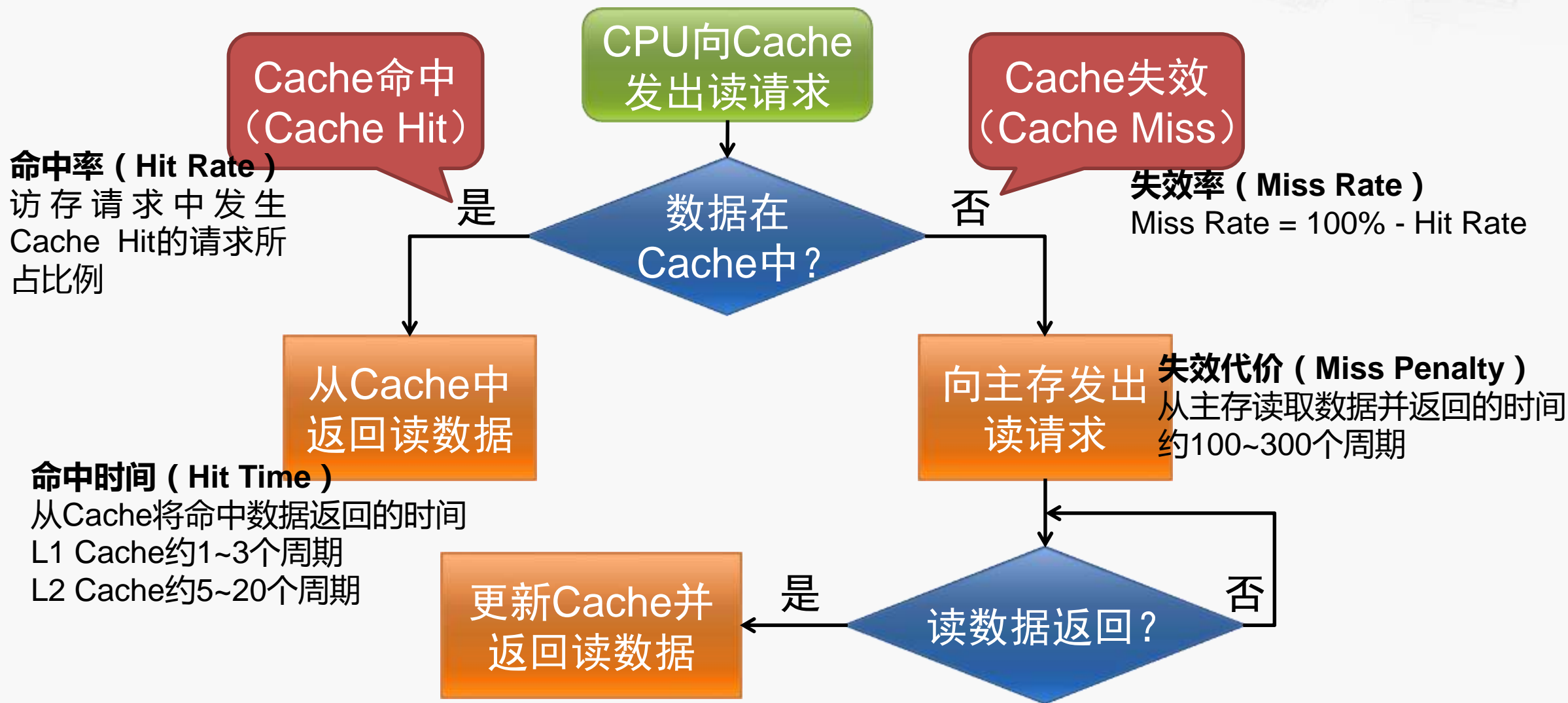


高速缓存的 设计要点

北京大学·慕课
计算机组成
制作人：陆俊林



Cache的访问过程



平均访存时间

平均访存时间Average Memory Access Time
= Hit Time + Miss Penalty × Miss Rate

1. 降低Hit Time

2. 减少Miss Penalty

3. 降低Miss Rate

减少平均访存时间的三个主要途径

命中率为97%时

平均访存时间： $3 + 300 \times 3\% = 12$

命中率为99%时

平均访存时间： $3 + 300 \times 1\% = 6$

命中率提高2%，访存性能提高一倍

注：设命中时间为3个周期，失效代价为300个周期

Cache失效原因

❏ 义务失效 (Compulsory Miss)

- 第一次访问某一数据块
- 也称为冷启失效 (Cold Start Miss)

无法有效避免

❏ 容量失效 (Capacity Miss)

- Cache无法保存程序访问所需的所有数据块

可通过增加
Cache容量缓解

❏ 冲突失效 (Conflict Miss)

- 多个存储器位置映射到同一Cache位置

如何解决





Cache的映射策略

存储器

地址

数据块

000H
010H
020H
030H
040H
050H
060H
070H
080H
090H
0A0H
0B0H
0C0H
0D0H
0E0H
0F0H
100H
....

0
1
2
3
4
5
6
7

直接映射Cache

二路组相联Cache

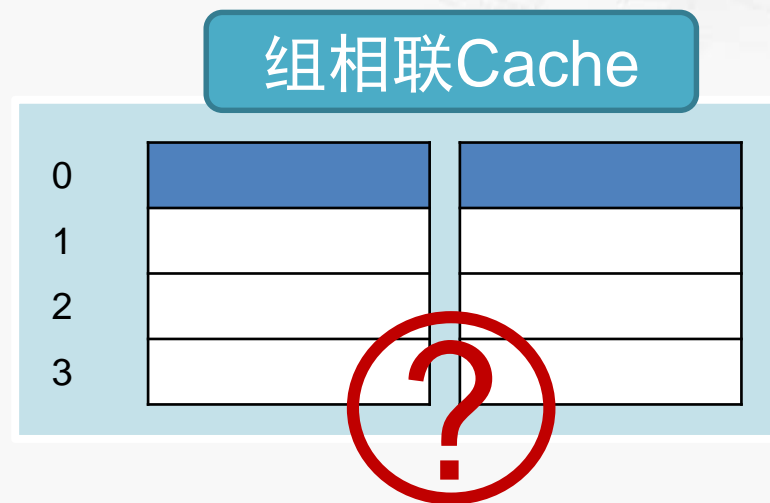
0
1
2
3

四路组相联Cache

0
1

常见的Cache替换算法

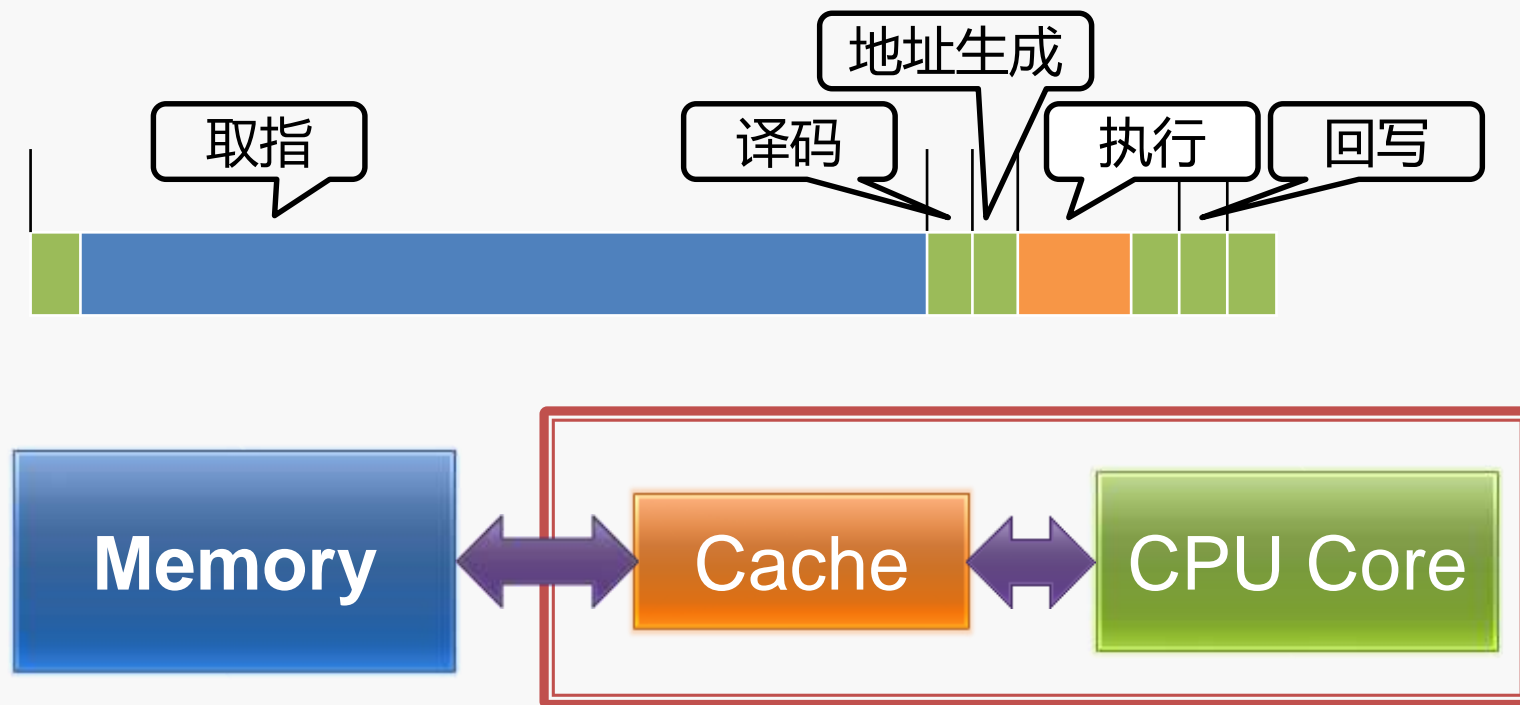
- ❏ 随机 (Random)
 - 硬件随机选择一个Cache块替换
- ❏ 轮转 (Round-Robin)
 - 按照预先设定的顺序依次选择Cache块替换
- ❏ 最近最少使用 (LRU)
 - 硬件记录访问历史信息
 - 选择距离现在最长时间未被访问的Cache块替换
- ❏



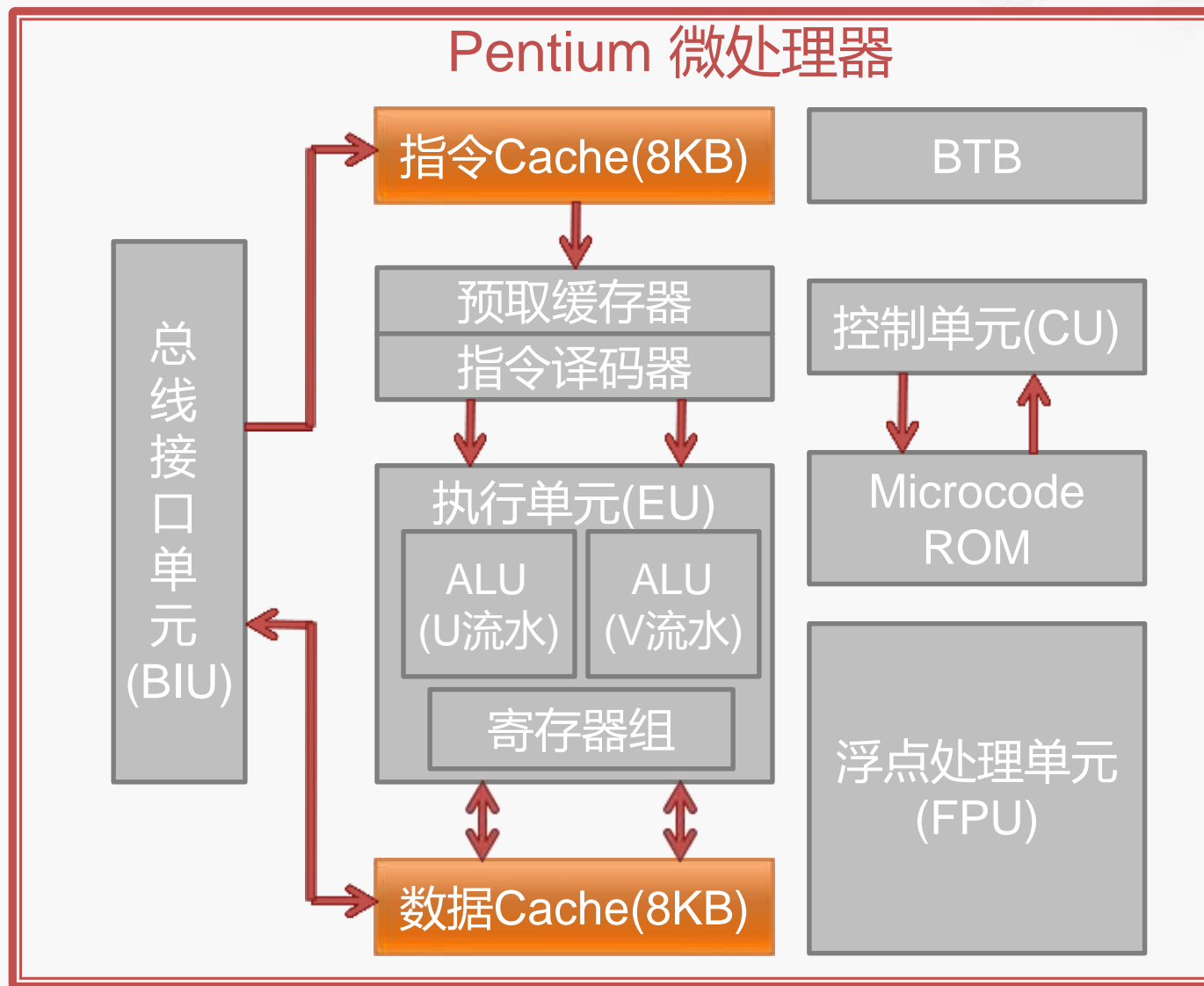
发生冲突时，如何选择替换的块

80486的高速缓存

- 80486采用指令和数据共用的8KB Cache
- 缺陷：指令和数据的局部性相互影响
 - 例如：一个数据密集的程序会很快占满Cache，几乎没有空间用于指令



Pentium的高速缓存



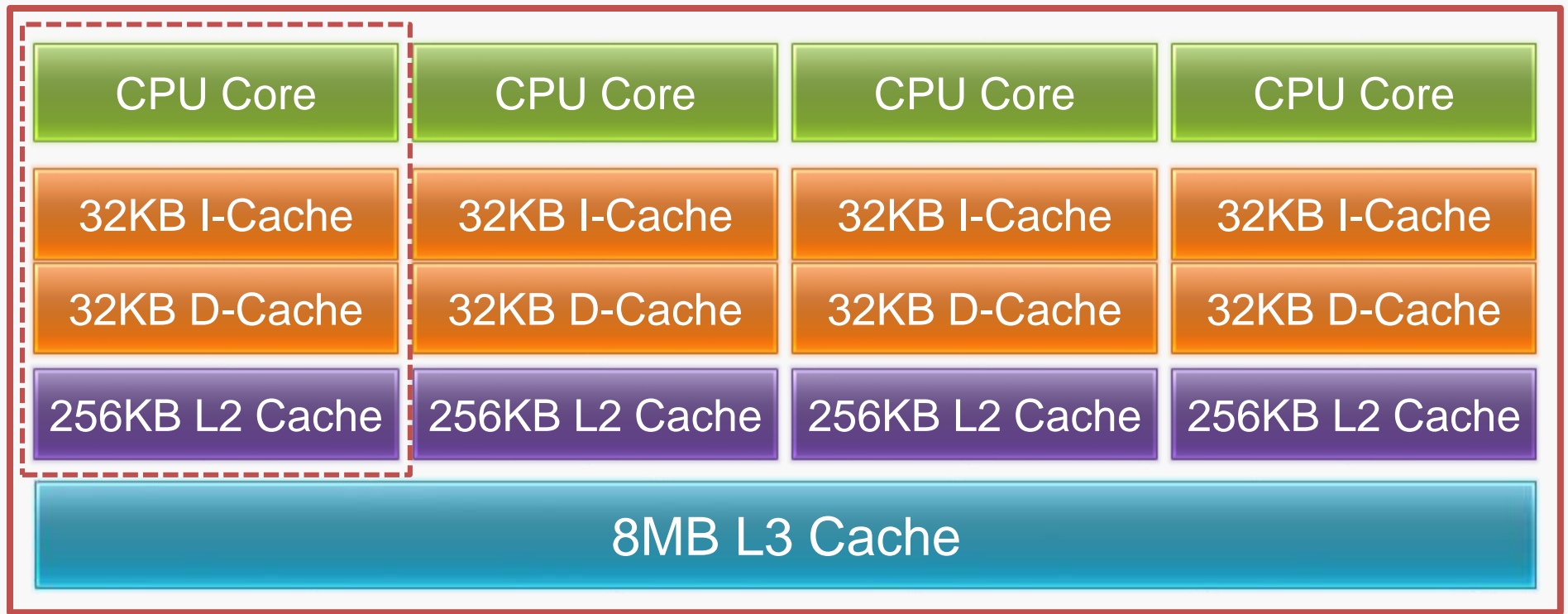


Core i7的多级高速缓存

L1 Cache : 指令和数据分离, 各32KB
8路组相联, 命中时间4个周期

L2 Cache : 统一的指令和数据, 共256KB
8路组相联, 命中时间11个周期

L3 Cache : 统一的指令和数据, 多核共享, 共8MB
16路组相联, 命中时间30~40个周期



本节小结



高速缓存的设计要点

北京大学·慕课
计算机组成
制作人：陆俊林

