

本节主题

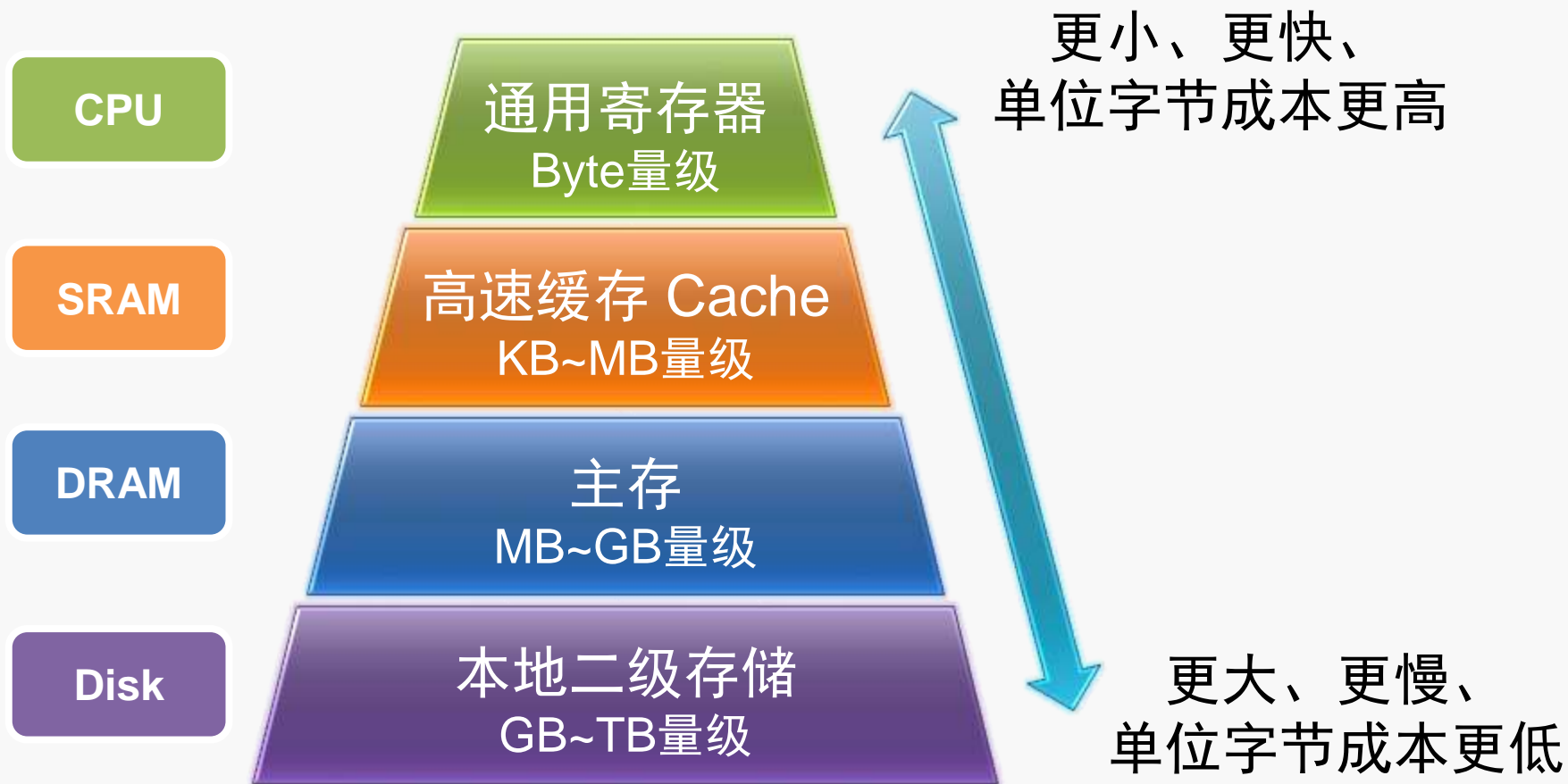


高速缓存的 工作原理

北京大学·慕课
计算机组成
制作人：陆俊林



存储层次结构 (Memory Hierarchy)



程序的局部性原理



- ④ 这是一个经验性的结论
 - 计算机程序从**时间**和**空间**都表现出“局部性”
- ④ **时间局部性** (Temporal Locality)
 - 最近被访问的存储器单元 (指令或数据) 很快还会被访问
- ④ **空间局部性** (Spatial Locality)
 - 正在被访问的存储器单元附近的单元很快会被访问

```
for (i=0; i<1000; i++)  
    for (j=0; j<200; j++)  
        sum += a[i][j];
```

典型程序段

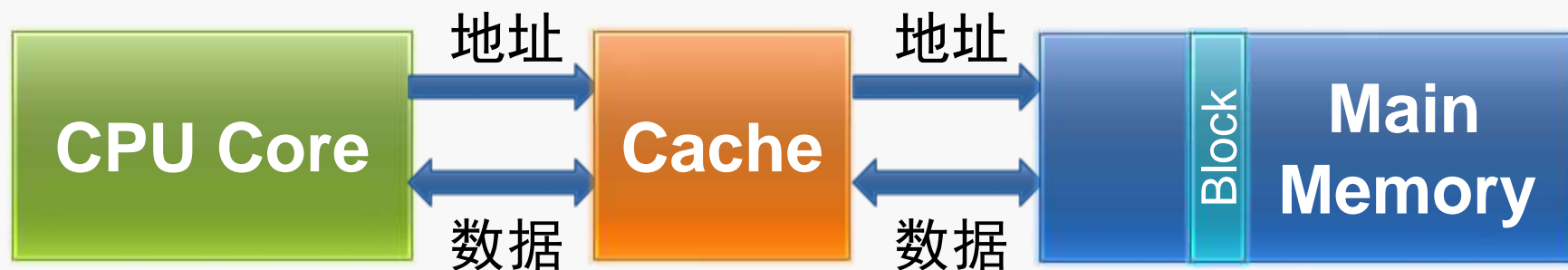
Cache的基本原理

Cache对空间局部性的利用

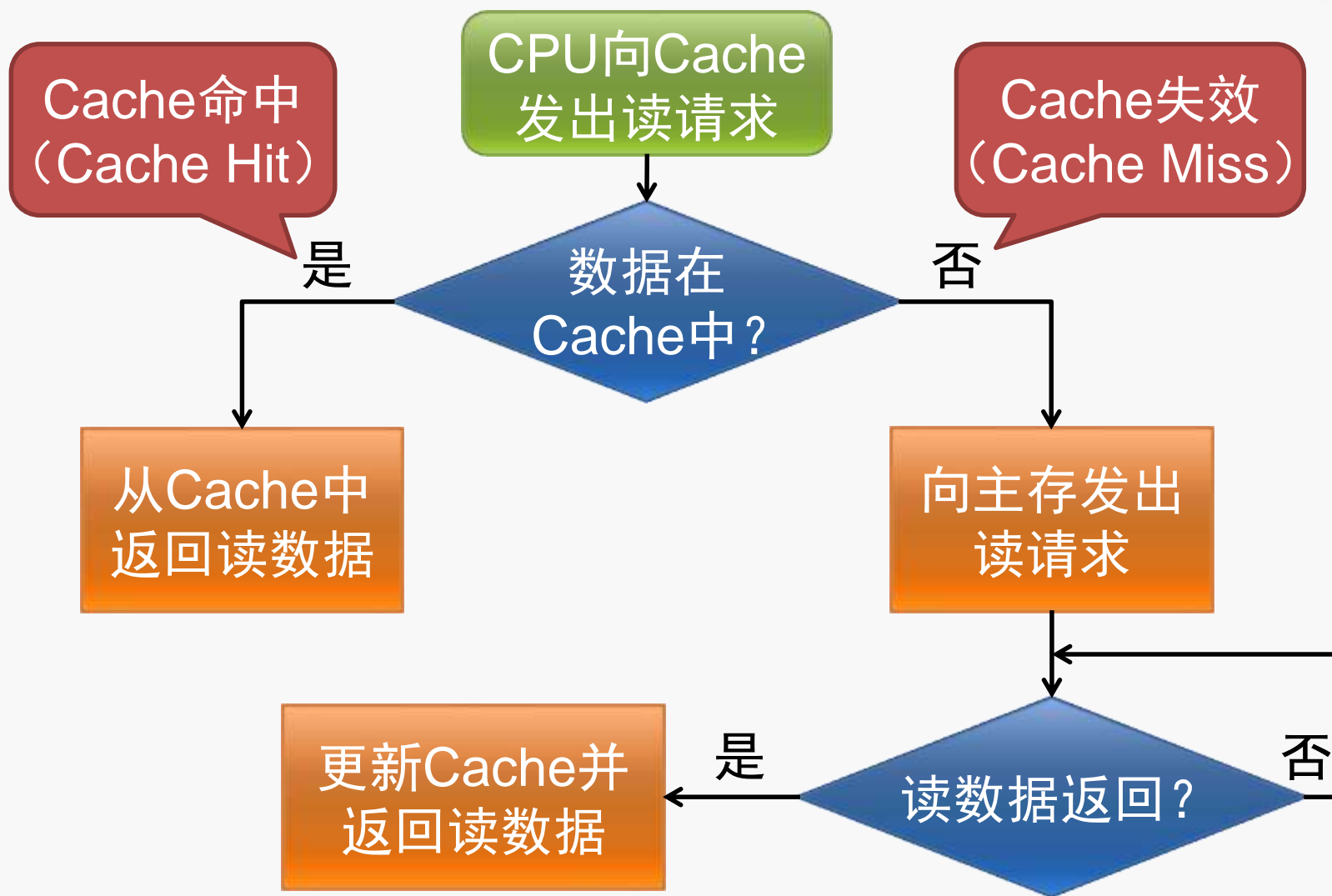
- 从主存中取回待访问数据时，会同时取回与位置相邻的主存单元的数据
- 以数据块（Block）为单位和主存进行数据交换

Cache对时间局部性的利用

- 保存近期频繁被访问的主存单元的数据



Cache的访问过程





Cache组织结构示例

	有效位	标签	数据					
表项0	0							
表项1	0		这是一个高速缓存行（Cache Line）					
表项2	0							
表项3	0							
表项4	0							
表项5	0							
.....	0		每行可存放16字节的数据块（Block）					
表项15	0		字节0	字节1	字节2	字节3	字节15



Cache读操作过程示例

```
MOV  AL, [2011H]
MOV  BL, [4011H]
MOV  CL, [3732H]
MOV  DL, [401FH]
```

注：为简化描述，不考虑段基址

	有效位	标签	数据					
表项0	0							
表项1	0							
表项2	0							
表项3	0							
.....								
表项15	0		字节0	字节1	字节2	字节3	字节15



Cache读操作过程示例（1）

1. 未命中

MOV AL, [2011H]
MOV BL, [4011H]
MOV CL, [3732H]
MOV DL, [401FH]

	有效位	标签	数据					
表项0	0							
表项1	0							
表项2	0							
表项3	0							
.....								
表项15	0		字节0	字节1	字节2	字节3	字节15

Cache读操作过程示例（1） 第1条指令完成

```
MOV AL, [2011H]
MOV BL, [4011H]
MOV CL, [3732H]
MOV DL, [401FH]
```

1. 未命中，读主存地址2010H，分配表项1，返回A1H

MOV AL, [2 0 1 1 H]

表项0

表项1

表项2

表项3

.....

表项15

有效位	标签	数据					
0							
1	20H	A0H	A1H	A2H	A3H	AFH
0							
0							
.....							
0		字节0	字节1	字节2	字节3	字节15

Cache读操作过程示例（2）

```
MOV AL, [2011H]
MOV BL, [4011H]
MOV CL, [3732H]
MOV DL, [401FH]
```

1. 未命中，读主存地址2010H，分配表项1，读出A1H

2. 未命中

MOV BL, [4 0 1 1 H]

表项0

表项1

表项2

表项3

.....

表项15

有效位	标签	数据					
0							
1	20H✗	A0H	A1H	A2H	A3H	AFH
0							
0							
0		字节0	字节1	字节2	字节3	字节15

Cache读操作过程示例（2） 第2条指令完成

```
MOV AL, [2011H]
MOV BL, [4011H]
MOV CL, [3732H]
MOV DL, [401FH]
```

1. 未命中，读主存地址2010H，分配表项1，读出A1H

2. 未命中，读内存地址4010H，替换表项1，读出B1H

MOV BL, [4 0 1 1 H]

	有效位	标签	数据					
表项0	0							
表项1	1	40H	B0H	B1H	B2H	B3H	BFH
表项2	0							
表项3	0							
.....								
表项15	0		字节0	字节1	字节2	字节3	字节15

Cache读操作过程示例（3）

```
MOV  AL, [2011H]
MOV  BL, [4011H]
MOV  CL, [3732H]
MOV  DL, [401FH]
```

1. 未命中，读主存地址2010H，分配表项1，读出A1H

2. 未命中，读内存地址4010H，替换表项1，读出B1H

3. 未命中

	有效位	标签	数据					
表项0	0							
表项1	1	40H	B0H	B1H	B2H	B3H	BFH
表项2	0							
表项3	0							
.....								
表项15	0		字节0	字节1	字节2	字节3	字节15

Cache读操作过程示例（3） 第3条指令完成

```
MOV AL, [2011H]
MOV BL, [4011H]
MOV CL, [3732H]
MOV DL, [401FH]
```

1. 未命中，读主存地址2010H，分配表项1，读出A1H

2. 未命中，读内存地址4010H，替换表项1，读出B1H

3. 未命中，读内存地址3730H，分配表项3，读出C2H

	有效位	标签	数据					
表项0	0							
表项1	1	40H	B0H	B1H	B2H	B3H	BFH
表项2	0							
表项3	1	37H	C0H	C1H	C2H	C3H	CFH
.....								
表项15	0		字节0	字节1	字节2	字节3	字节15

Cache读操作过程示例（4） 第4条指令完成

MOV AL, [2011H]

MOV BL, [4011H]

MOV CL, [3732H]

MOV DL, [401FH]

1. 未命中，读主存地址2010H，分配表项1，读出A1H

2. 未命中，读内存地址4010H，替换表项1，读出B1H

3. 未命中，读内存地址3730H，分配表项3，读出C2H

4. 命中表项1，读出BFH

表项0

表项1

表项2

表项3

.....

表项15

有效位	标签	数据					
0							
1	40H	B0H	B1H	B2H	B3H	BFH
0							
1	37H	C0H	C1H	C2H	C3H	CFH
0		字节0	字节1	字节2	字节3	字节15

Cache的写策略



④ “Cache命中” 时的写策略

- ① **写穿透** (Write Through) : 数据同时写入Cache和主存
- ② **写返回** (Write Back) : 数据只写入Cache , 仅当该数据块被替换时才将数据写回主存

④ “ Cache失效” 时的写策略

- ① **写不分配** (Write Non-Allocate) : 直接将数据写入主存
- ② **写分配** (Write Allocate) : 将该数据所在的块读入Cache后 , 再将数据写入Cache

写返回

写分配

写穿透

写不分配

本节小结



高速缓存的 工作原理

北京大学·慕课
计算机组成
制作人：陆俊林

