

# 《计算机组成原理》

## 作业4

2023

33920212204567 任宇

# 4-1

1. 存取时间：指计算机系统中存储器或外部设备读写数据所需的时间，包括数据传输时间和访问延迟时间。
2. 存取周期：指在计算机系统中执行一次读写操作所需的时间。
3. 存储器带宽：是指单位时间里存储器所存取的信息量，也称为存储器在单位时间内读出/写入的位数或字节数。
4. 存储单元：存储器中可存放一个字或若干字节的基本单位。
5. 边界对齐的数据存放：就是让数据地址按照其数据类型大小的整数倍进行储存，目的在于使得数据能以最少的次数连续读取。
6. 大端存储：将多字节数据类型的高位字节（MSB）存储在低地址处，低位字节（LSB）存储在高地址处。
7. 小端存储：将多字节数据类型的高位字节（MSB）存储在高地址处，低位字节（LSB）存储在低地址处。
8. 静态存储器：简称SRAM，是一种计算机内部使用的高速随机存取存储器，用于存储CPU的缓存、寄存器和其他重要数据。

9. 动态存储器：简称DRAM，是一种计算机内部使用的高速随机存取存储器，用于存储操作系统、应用程序和其他数据，需要频繁地刷新数据。
10. 刷新：指周期性地读取并重新写入动态存储器（DRAM）中的数据，以保持数据正确性的一个过程。
11. 刷新周期：指动态存储器（DRAM）进行一次完整刷新操作所需的时间。
12. 字扩展：将所有存储芯片的数据线和读写控制线各自并联，同时分别与CPU的数据线和读写控制线连接。
13. 位扩展：将多个一位存储体的地址线并联在一起工作即可得到多位的数据。
14. 多体交叉存储器：是一种高性能计算机存储器架构，它由多个存储体和交叉开关网络组成。
15. 高速缓冲存储器：是一种用于提高计算机处理速度的存储器，它位于主存储器和CPU之间，通常具有更快的访问速度和更小的容量。
16. 双端口存储器：是一种具有两个独立数据输入/输出端口的存储器，可以同时进行两个不同的访问操作。每个端口都能够独立地读取或写入存储器中的数据，并且不会相互影响。

17. 相联存储器：相联存储器是根据存储数据的内容来查找数据，而不是根据数据地址。
18. 时间局部性：是指在一段时间内，计算机程序访问的数据和指令往往会集中在某些区域，而这些区域可能会被反复访问。
19. 地址映射：指将逻辑地址转换为物理地址的过程。
20. 直接相联映射：是一种高速缓存的映射方式，它将主存地址空间的每个块映射到缓存中唯一的一组行。具体来说，对于一个给定的主存块，它只能被映射到缓存中一行位置上，而这个位置是通过主存地址进行哈希计算得出的。
21. 全相联映射：将主存地址空间的每个块都可以映射到缓存中的任意一行。
22. 组相联映射：它将主存地址空间划分为多个大小相等的块，每个块称为一个“组”。每个组中可以容纳多个缓存行。
23. 命中率：指在特定时间段内缓存访问中成功地从缓存中读取数据或指令的比例。
24. 虚拟存储器：是一种计算机内存管理技术，它将主存储器和磁盘等外部存储器结合起来，使得操作系统可以为每个进程提供一个独立的、连续的地址空间。

- 25. 页框号：指在操作系统中用来表示物理内存地址的一种标识符。
- 26. 页表（慢表）：将虚拟地址映射到物理地址，并记录了这个映射关系。
- 27. 页表项：在使用分页技术实现虚拟内存时，每个进程都维护自己的页表，而页表项则用来描述页面与物理地址之间的映射关系。
- 28. TLB（快表）：是计算机处理器中用于加速虚拟地址到物理地址转换的缓存。它通常是一个小而高速的硬件缓存，存储了最近访问过的一部分页表项。
- 29. LRU算法：近期最少使用算法是将近期内最久未被访问的行淘汰。
- 30. LFU算法：最不经常使用算法将被访问次数最少的cache行淘汰。
- 31. Cache一致性：是指多个缓存副本之间的数据一致性。
- 32. 写回法：当CPU对cache写命中时，只修改cache的内容而不立即写入主存，只有当此行被替换出cache时才将数据写回主存。
- 33. 写穿法：当cache写命中时，同时对cache和主存中的同一数据块进行修改。

## 4-2

(1) [2010] 下列有关 RAM 和 ROM 的叙述中, 正确的是 A。

I. RAM 是易失性存储器, ROM 是非易失性存储器

II. RAM 和 ROM 都采用随机存取方式进行信息访问

III. RAM 和 ROM 都可用作 cache

IV. RAM 和 ROM 都需要进行刷新

A. 仅 I 和 II

B. 仅 II 和 III

C. 仅 I、II 和 IV

D. 仅 II、III 和 IV

答: A, 原因如下:

ROM 是只读存储器, 不能作为 cache, DRAM 需要刷新, 而 ROM 不需要刷新

(2) [2014] 某容量为 256MB 的存储器由若干  $4M \times 8$  位的 DRAM 芯片构成, 该 DRAM 芯片的地址引脚和数据引脚总数是 A。

A. 19

B. 22

C. 30

D. 36

答: A, 原因如下:

$4M \times 8 \text{ 位} = 2^{22} \times 8$ , 需要 22 根地址线, 而 DRAM 采用地址复用技术, 地址线数为 11 根, 再加上 8 根数据线,  $11 + 8 = 19$



## 4-2

(3) [2009] 某计算机主存容量为 64KB，其中 ROM 区为 4KB，其余为 RAM 区，按字节编址。现要用  $2\text{KB} \times 8$  位的 ROM 芯片和  $4\text{KB} \times 4$  位的 RAM 芯片来设计该存储器，则需要上述规格的 ROM 芯片数和 RAM 芯片数分别是 D。

A. 1、15      B. 2、15      C. 1、30      D. 2、30

答：D，原因如下：

ROM 芯片数 =  $(4\text{KB} \times 8) / (2\text{KB} \times 8) = 2$  片

RAM 范围是 60KB，RAM 芯片数 =  $(60\text{KB} \times 8) / (4\text{KB} \times 4) = 30$  片

(4) [2010] 假定用若干个  $2\text{KB} \times 4$  位的芯片组成一个  $8\text{K} \times 8$  位的存储器，则地址 0B1FH 所在芯片的最小地址是 D。

A. 0000H      B. 0600H      C. 0700H      D. 0800H

答：D，原因如下：

需要用到字位扩展，需要用到 13 根地址线，对高 2 位地址译码进行片选

0B1FH = 0 1011 0001 1111

最小地址为 0 1000 0000 0000 = 0800H

## 4-2

(5) [2018] 假定 DRAM 芯片中存储阵列的行数为  $r$ 、列数为  $c$ ，对于一个  $2\text{KB} \times 1$  位的 DRAM 芯片，为保证其地址引脚数最少，并尽量减少刷新开销，则  $r$ 、 $c$  的取值分别是 C。

A. 2048、1

B. 64、32

C. 32、64

D. 1、2048

答：C，原因如下：

行列数差值要尽量小，且行数应该选少的

(6) [2019] 假定一台计算机采用 3 通道存储器总线，配套的内存条型号为 DDR3-1333，即内存条所接插的存储器总线的工作频率为 1333MHz、总线宽度为 64 位，则存储器总线的总带宽大约是 B。

A. 10.66GB/s

B. 32GB/s

C. 64GB/s

D. 96GB/s

答：B，原因如下：

$3 \times 8 \times 1333\text{MB/s}$  约等于  $32\text{GB/s}$



## 4-2

(7) [2015] 某计算机使用 4 体交叉编址存储器，假定在存储器总线上出现的主存地址（十进制）序列为 8005、8006、8007、8008、8001、8002、8003、8004、8000，则可能发生访存冲突的地址对是 D。

A. 8004 和 8008      B. 8002 和 8007      C. 8001 和 8008      D. 8000 和 8004

答：D，原因如下：

用低2位地址进行片选，模块序号等于访存地址%存储器交叉模块数  
当相邻的4次访问中出现相同存储模块时会发生冲突。

(8) [2015] 下列存储器中，在工作期间需要周期性刷新的是 B。

A. SRAM      B. SDRAM      C. ROM      D. FLASH

答：B，原因如下：

DRAM使用电容，由于电容的物理特性，需要周期性刷新

## 4-2

(9) [2011] 下列各类存储器中, 不采用随机存取方式的是 B。

A. EPROM

B. CDROM

C. DRAM

D. SRAM

答: B, 原因如下:

CDROM即光盘, 使用顺序存取方式

(10) [2012] 下列关于闪存 (Flash Memory) 的叙述中, 错误的是 A。

A. 信息可读可写, 并且读、写速度一样快

B. 存储元由 MOS 管组成, 是一种半导体存储器

C. 掉电后信息不丢失, 是一种非易失性存储器

D. 采用随机访问方式, 可替代计算机外部存储器

答: A, 原因如下:

闪存写入时需要先擦除原来的数据, 所以写比读慢

(11) [2017] 下列关于数组 **a** 的访问局部性的描述中, 正确的是 A。

- A. 时间局部性和空间局部性皆有
- B. 无时间局部性, 有空间局部性
- C. 有时间局部性, 无空间局部性
- D. 时间局部性和空间局部性皆无

答: A

(12) [2009] 某计算机的 **cache** 共有 16 块, 采用二路组相联映射方式 (即每组 2 块)。每个主存块大小为 32B, 按字节编址。主存 129 号单元所在主存块应装入的 **cache** 组号是 C。

- A. 0
- B. 1
- C. 4
- D. 6

答: C, 原因如下:

cache 一共有 16 块, 每组两块, 即分为 8 组, 组索引字段应该为 3 位, 每个主存块大小为 32B, 块内偏移地址是 5 位, 129 即 0 100 00001  
对应的是第 4 块

## 4-2

(13) [2012] 假设某计算机按字编址, cache 有 4 行, cache 和主存之间交换的块大小为 1 个字。若 cache 的内容初始为空, 采用二路组相联映射方式和 LRU 替换策略。访问的主存地址依次为 0、4、8、2、0、6、8、6、4、8 时, 命中 cache 的次数是 C。

A. 1

B. 2

C. 3

D. 4

答: C, 原因如下:

主存按照字编址, cache 块大小就是一个字, 主存地址中没有块内偏移字段, cache 分为 2 组, 每组两行, 主存地址最低位就是 cache 组索引。

(14) [2015] 假定主存地址为 32 位, 按字节编址, 主存和 cache 之间采用直接相联映射方式, 主存块大小为 4 个字, 每个字 32 位, 采用写回的方式, 则能存放 4K 字数据的 cache 的总容量至少是 C 位。

A. 146K

B. 147K

C. 148K

D. 158K

答: C, 原因如下:

按字节编址, 块大小为  $4 \times 32 \text{ 位} = 16 \text{ B}$ , 块内偏移地址为 4 位, cache 容量为 4K 字, 则 cache 有  $4 \text{ KB} \times 4 \text{ B} / 16 \text{ B} = 1 \text{ K}$  行, cache 行索引为 10 位, 标记字段为  $32 - 10 - 4 = 18$  位

cache 总容量 = cache 大小  $\times$  行数, cache 总容量 =  $(16 \times 8 + 1 + 18 + 1) \times 1 \text{ K} = 148 \text{ K 位}$

(15) [2014] 采用指令 cache 与数据 cache 分离的主要目的是 D。

- A. 降低 cache 的缺失损失
- B. 提高 cache 的命中率
- C. 降低 CPU 平均访存时间
- D. 减少指令流水线资源冲突

答：D

(16) [2015] 假定编译器将赋值语句 “ $x=x+3$ ;” 转换为指令 “add xaddr,3”，其中，xaddr 是 x 对应的存储单元地址。若执行该指令的计算机采用页式虚拟存储管理方式，并配有相应的 TLB，且 cache 使用写穿的方式，则完成该指令功能需要访问主存的次数至少是 B。

- A. 0
- B. 1
- C. 2
- D. 3

(17) [2010] 下列命中组合情况中

答：B，原因如下：

分为取数、运算和写回三步，取数时读取 xaddr 可能由于 TLB 命中、cache 命中而不需要访问主存，写穿法需要访问主存，所以至少访问一次



## 4-2

(17) [2010] 下列命中组合情况中，一次访存过程中不可能发生的是 D。

A. TLB 未命中, cache 未命中, Page 未命中  
B. TLB 未命中, cache 命中, Page 命中  
C. TLB 命中, cache 未命中, Page 命中  
D. TLB 命中, cache 命中, Page 未命中

答: D

(18) [2013] 某计算机主存地址空间大小为 256MB，按字节编址。虚拟地址空间大小为 4GB，采用页式存储管理方式，页面大小为 4KB，TLB（快表）采用全相联映射，有 4 个页表项，内容如表 4.12 所示。

表 4.12 4 个页表项的内容

有效位	标记	页框号	...
0	FF180H	0002H	...
1	3FFF1H	0035H	...
0	02FF3H	0351H	...
1	03FFFH	0153H	...

则对虚拟地址 03FFF180H 进行虚实地址转换的结果是 A。

A. 0153180H      B. 0035180H      C. TLB 缺失      D. 缺页

答: A, 原因如下:

按字节编址，页面大小为 4KB，页内地址共 12 位。地址空间大小为 4GB，虚拟地址 32 位，前 20 位为页号。虚拟地址 03FFF180H，页号为 03FFFH，页内地址为 180H，所以物理地址为 0153180H



(19) [2019] 下列关于缺页处理的叙述中, 错误的是 D。

- A. 缺页是在地址转换时 CPU 检测到的一种异常
- B. 缺页处理由操作系统提供的缺页处理程序完成
- C. 缺页处理程序根据页故障地址从外存读入所缺失的页
- D. 缺页处理完成后执行发生缺页的指令的下一条指令

答: D, 原因如下:

缺页处理完成后应该回到发生缺页的指令继续执行

(1) 计算机系统中采用层次化存储体系结构的目的是什么？

答：计算机系统中采用层次化存储体系结构的主要目的是在不同的存储设备之间实现数据的快速访问和高效利用。层次化存储体系可以利用程序局部性的原理，全方位优化存储系统各项性能指标。

(2) 为什么在存储器芯片中设置片选输入端？

答：存储扩展时需要用到多个存储芯片，设置片选输入端可以方便选择正确的存储器芯片来进行数据访问。

(3) 动态MOS存储器为什么要刷新？如何刷新？

答：动态MOS存储器之所以需要刷新，是因为它们使用的是基于电容的存储单元，这些电容会逐渐漏电，导致存储的数据逐渐丧失。为了避免数据丢失，动态MOS存储器需要定期进行刷新操作。刷新方式分为集中刷新、异步刷新和分散刷新3种，由于存储器采用行列结构，所以刷新时先按行进行。

(4) 试述多体交叉存储器的设计思想和实现方法。

答：多体交叉存储器是一种高效的存储器结构，其设计思想是将大容量的存储器划分为多个小块，每个小块成为一个存储体，以提高存储器的并行性和整体性能，可以分为高位多体交叉和低位多体交叉两种。高位多体交叉主要用于存储容量扩展，存储地址高位译码片选存储模块，统一存储体内的地址是连续的；低位多体交叉则使用低位译码片选存储模块。

(5) 为什么说cache对程序员是透明的？

答：因为在程序员看来数据是在寄存器，内存和辅存之间交换的，感觉不到cache的存在，也无法操控cache。

(6) 直接相联映射方式下为什么不需要用替换算法？

答：因为每个主存块只能映射到高速缓存中的唯一一个位置。在直接相联映射的高速缓存结构中，每个主存块都映射到了唯一的高速缓存行，当指令或数据被访问时，先计算出其所对应的高速缓存地址，然后直接查询该地址上的缓存行是否与所需的块匹配，如果匹配，则命中；否则，发生缺失。由于每个主存块只能映射到唯一的高速缓存行中，因此不需要进行替换操作。

(7) 为什么要考虑cache的一致性？

答：在多处理器系统中，每个处理器都有自己的缓存，这些缓存独立地存储着各自使用的数据。当多个处理器同时访问同一块内存时，就可能出现缓存不一致的情况。当一个处理器修改了数据后，其他处理器的缓存中并没有及时更新该数据，导致不同处理器读取到的数据不一致。这种情况会导致程序执行结果的错误或者异常。

(8) 替换算法有哪几种？它们各有什么优缺点？

答：先进先出算法替换最先被载入行，其算法系统开销小，但无法利用程序局部性，cache命中率不高。

最不经常使用算法替换访问次数最少的cache行，每个cache行都需要设置一个计数器，硬件成本较高，不足是不能严格反映近期访问情况。

近期最少使用算法是将近期内最久未被访问的行淘汰，为此，每行也都需要设置一个计数器，这种方法的cache命中率较高，但难点在于快速比较多行计数器。

随机替换算法就是从特定的行中随机地选取一行进行替换，无法利用程序局部性，命中率一般，但是其硬件成本最低。

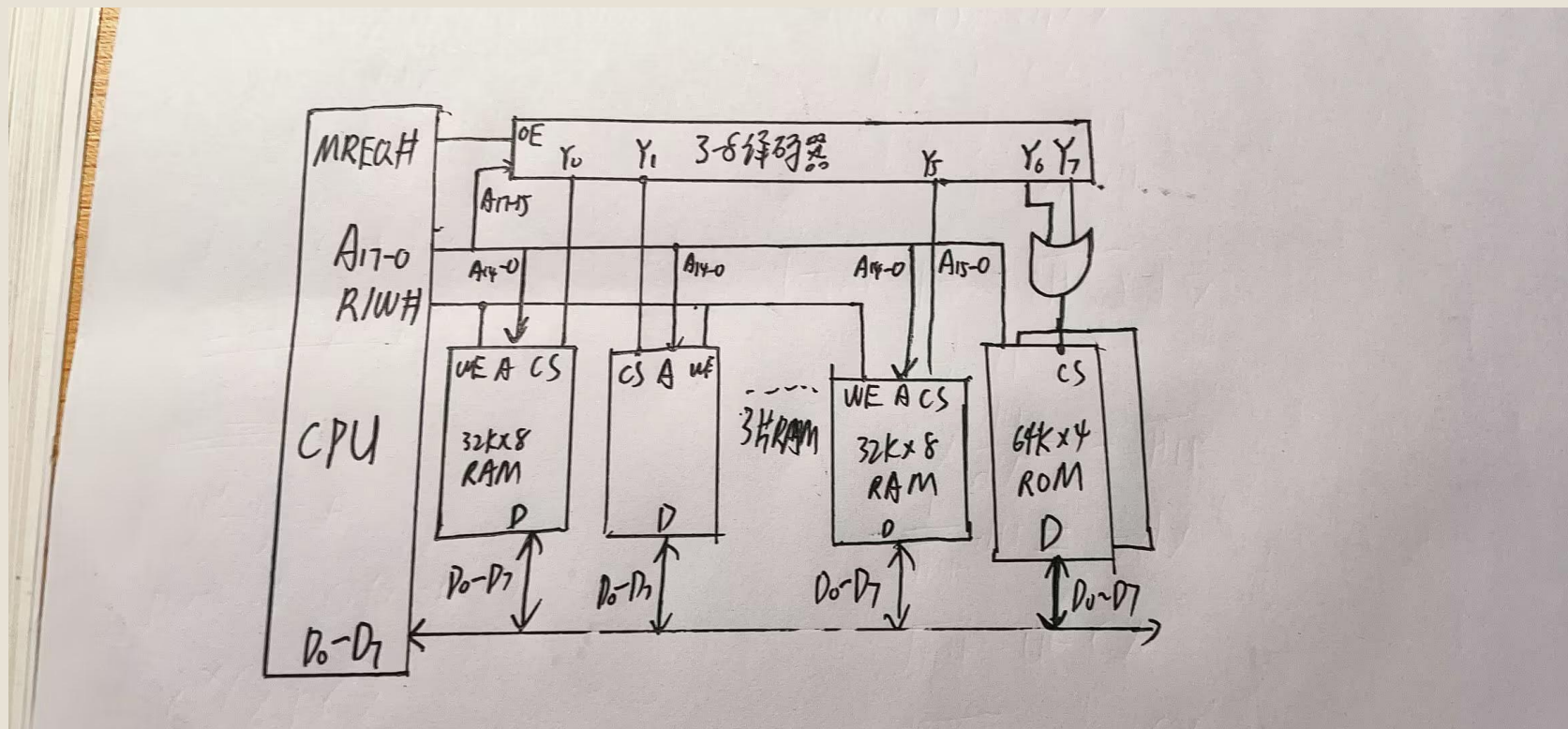
# 4-6

4.6 用  $32K \times 8$  位 RAM 芯片和  $64K \times 4$  位 ROM 芯片设计  $256K \times 8$  位存储器。其中，从  $30000H$  到  $3FFFFH$  的地址空间为只读存储区，其他为可读、可写存储区。完成存储器与 CPU 的连接。

答:  $256KB = 2^{18}B$

$30000H$  到  $3FFFFH$  地址空间作为只读存储区，使用 2 片  $64K \times 4$  位 ROM 芯片做位扩展

$00000H$  到  $2FFFFH$  地址空间为 RAM 区间，容量为  $192KB$ ，需要 6 片  $32K \times 8$  位 RAM 芯片做字扩展：





奇结构及其与 CPU 连接的框图。

✓4.8 用  $64\text{K} \times 1$  位的 DRAM 芯片构成  $1\text{M} \times 8$  位的存储器，若采用异步刷新，每行刷新间隔不超过  $2\text{ms}$ ，则产生刷新信号的间隔时间是多少？假设读写周期为  $0.5\mu\text{s}$ ，若采用集中刷新方式，则存储器刷新一遍最少要用多少个读写周期？CPU 的“死”时间为多少？

答：64KB  $\times$  1位的DRAM芯片是256列  $\times$  256行，如果采用异步刷新，需要将2ms分为256个时间段即  $2\text{ms}/256=7.8125\mu\text{s}$ ，每个时间段的最后  $0.5\mu\text{s}$  用于刷新行，因此产生刷新信号的间隔时间是  $7.8125\mu\text{s}$ 。

若采用集中刷新方式，存储器刷新一遍至少需要256个读写周期  
“死时间”为  $256 \times 0.5\mu\text{s}=128\mu\text{s}$



# 4-11

在  $1\mu s$  内至少访问存储器一次，采用哪种刷新方式比较合适？信号的产生周期是多少？

4.11 设 cache 的容量为  $2^{14}$  块，每块是一个 32 位字，主存容量是 cache 容量的 256 倍，其中有表 4.13 所示的数据（地址和数据均采用十六进制表示）。

表 4.13 主存数据分布情况

地址	数据	地址	数据
000000	87568536	01FFFC	4FFFFC68
000008	87792301	FFFFF8	01BF2460
010004	9ABEFCDD		

将主存中这些数据装入 cache 后，cache 各块中的数据内容及相应的标志是什么？

(1) 全相联映射；(2) 直接相联映射；(3) 四路组相联映射。

答：

全相联映射：主存地址22位，块内偏移地址2位

直接相联映射：区地址8位，区内行地址14位，字地址2位

四路组相联映射：块内偏移2位，组索引12位，标记10位

Cache行	标志	数据
0	000000	87568536
1	000002	87792301
2	004001	9ABEFCDD
3	007FFF	4FFFFC68
4	3FFFFE	01BF2460

Cache行	标志	数据
0000	00	87568536
0002	00	87792301
0001	01	9ABEFCDD
3FFF	01	4FFFFC68
3FFE	FF	01BF2460

Cache组	标志	数据
000	000	87568536
002	000	87792301
001	004	9ABEFCDD
0FFF	007	4FFFFC68
0FFE	3FF	01BF2460

# 4-13

4.13 某计算机的主存容量为 4MB, cache 容量为 16KB, 每块包含 8 个字, 每字为 32 位, 映射方式采用四路组相联。设 cache 的初始状态为空, CPU 依次从主存第 0,1,2,...,99 号单元读出 100 个字 (每次读一个字), 并重复此操作 10 次, 替换算法采用 LRU 算法。

(1) 求 cache 的命中率。

(2) 若 cache 比主存快 10 倍, 分析采用 cache 后存储访问速度提高了多少。

答: (1) 一块8个字, 因此读出100个字被分配在13个块内。Cache中能存放的主存块数为 $16KB / (8 \times 4B) = 512$ 块, 因为是四路组相联, 所以cache包含组块数 $512 / 4 = 128$ 组, 组索引为7位。

13块数据被填入后足够存放, 不会被调出, 所以每次访问除了第一个访问不命中以外, 其余均可命中, 10次操作访问内存1000次, 其中不命中的次数只有13次

命中率 =  $(1000 - 13) / 1000 = 98.7\%$

(2) 设访问cache一个数据单元的时间为 $t$ , 则访问主存时间为 $10t$

不采用cache, 即 $10t \times 1000 = 10000t$

采用cache, 即 $10t \times 13 + 987 \times t = 1117t$

提高了  $(10000t - 1117t) / 1117t = 7.95$ 倍

4.14 假定某数组元素按行优先顺序存放在主存中，则在以下两段伪代码 A 和 B 中，分析下列问题。

(1) 两段代码中对数组访问的时间局部性和空间局部性。

(2) 变量 sum 的时间局部性和空间局部性。

(3) for 循环体对指令访问的时间局部性和空间局部性。

```
int sum_array_A(int a[M][N])
int i,j,sum=0;
for(i=0;i<M;i++)
    for(j=0;j<N;j++)
        sum+=a[i][j];
return sum;
```

```
int sum_array_B(int a[M][N])
int i,j,sum=0;
for(i=0;i<N;i++)
    for(j=0;j<M;j++)
        sum+=a[j][i];
return sum;
```

4.15 主存容量为 8MB，缓存容量为 2GB，公平管理地址...

(1)

(2)

(3)

(4)

4.1

回答下:

(

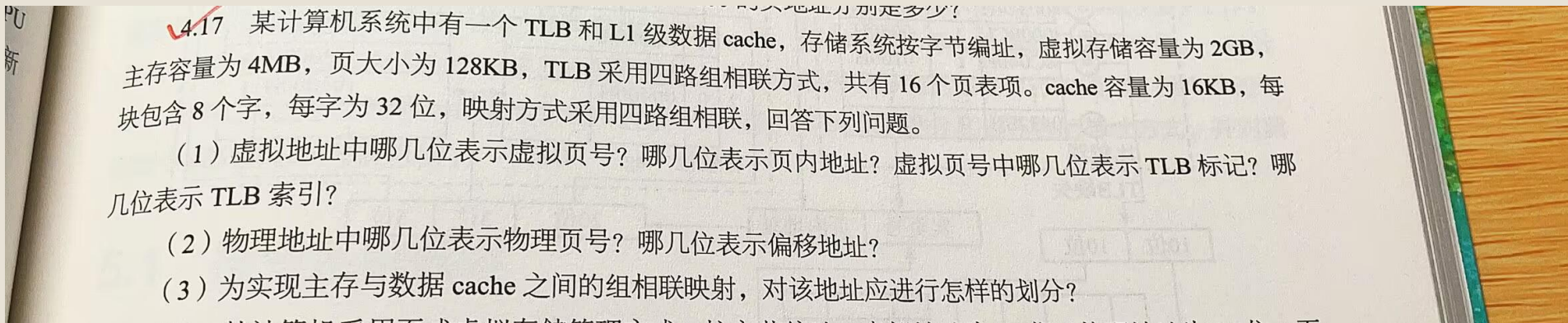
答：(1) 数组在内存中按照行优先存放，而数据按块从主存映射到cache中。程序A按照行优先顺序访问，有良好的空间局部性，但是每个数组元素只用一次，不存在时间局部性。程序B按照列优先顺序访问，空间局部性不佳，也不存在时间局部性。

(2) sum是单个变量，不存放空间局部性，但是sum在每次循环中都会被使用到，具有良好的时间局部性。

(3) for指令循环执行，因此对指令的访问具有良好的时间局部性，另外，循环体中的机器指令一般也是按照顺序执行，具有一定的空间局部性。



# 4-17



答：（1）虚拟存储页面数=2GB/128KB=2<sup>14</sup>，因此页号位数为14位，页大小为128KB，所以页内地址为17位，即虚拟地址高14位表示虚页号，低17位表示页内偏移地址。

TLB采用四路组相联，16个页表项，每组4路，共4组，所以组索引是2位，虚页号的高14位中，高12位为TLB标记，低2位为TLB组索引。

（2）物理内存4MB=2<sup>22</sup>B，地址线22根，其中高5位表示物理页号，低17位表示页内偏移地址。

（3）cache块大小为32B，块内字节偏移地址为5位，采用四路组相联，cache分成16KB/（4x32B）=128组，组索引7位，剩下的22-7-5=10位作为标记，即

标记	组索引	块内偏移地址
10	7	5