

4.1 解释下列名词

存取时间：从计算机存储器中读取或写入数据所需的时间。通常包括了存取存储单元的时间、数据传输时间等。

存储单元：是计算机中用于存储数据或指令的最小单元。在存储器中，存储单元通常是以字节为单位，而在较低级别的存储器（如缓存）中，存储单元可以是以比特或字为单位。

刷新：通常指的是存储器中的数据重新写入存储介质以防止数据丢失或损坏。

地址映射：指的是将逻辑地址或虚拟地址转换为物理地址的过程。在计算机系统中，逻辑地址通常由 CPU 生成，物理地址则是存储器单元上实际的地址。

命中率：指的是 cache 在给定的时间段内，所请求的数据在缓存中已经存在的比率。

大端存储：一种数据存储方式，最高有效字节存储在最低的内存地址中；最低有效字节存储在最高内存地址中。

页表：虚拟内存管理中的一种数据结构，用于将虚拟地址映射到物理地址。它记录了每个虚拟页面与实际物理内存中页面的映射关系。

LRU：最近最少使用算法，它根据数据的使用频率来确定哪些数据应该被保留在缓存中，哪些数据应该被替换出去。

写回法：一种缓存写入策略，只有在缓存中的数据被修改后才会被写回到主存储器。该策略可以减少对主存储器的写入次数，提高系统性能。

4.2 选择题

(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

分析：RAM 是易失性存储器，ROM 是非易失性存储器，且都采用随机存取方式进行信息访问。ROM 一般不可用作 cache，因为 ROM 是只读存储器 (Read Only)。只有 RAM 需要进行刷新，因为 RAM 是易失性存储器。故选 A。

(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

分析： $4M = 4 \times 2^{20}B = 2^{22}B$ ，按照字节寻址，地址引脚应该有 22 根，但 DRAM 采用地址复用技术，地址引脚是原来的一半，所以是 11 根，再加上数据引脚 8 根，总共 19 根，故选 A。

(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 $2KB \times 8$ 位的 ROM 芯片和 $4KB \times 4$ 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 (D)

RAM 芯片数分别是。

- A. 1、15 B. 2、15 C. 1、30 D. 2、30

分析：ROM 区所需是 4KB，且 ROM 规格为 $2KB \times 8$ 位，因此只需进行字扩展，需要 $4/2 = 2$ 片 ROM。对于 RAM，由于 RAM 规格是 $4KB \times 4$ 位，需要对其进行字位同时扩展，需要 $((64 - 4) \times 8) / (4 \times 4) = 30$ 片 RAM。故选 D。

(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

分析: $2\text{K} \times 4$ 位芯片组成 $8\text{K} \times 8$ 位存储器, 需要芯片数量为 $(8 \times 8) / (2 \times 4) = 8$ 片。由于 $8 = 2^3$, 所以需要 3 位片选信号。片内为 2K , 而 $2\text{K} = 2 \times 2^{10}\text{B} = 2^{11}\text{B}$, 所以片内地址是 11 位。而由于先进行位扩展再进行字扩展, 所以两个芯片位扩展后变成了八位, 这样的八位芯片需要 4 组, 所以最终只用在片选地址位数的基础上加上 2 位即可。根据地址 0B1FH , 对应二进制为 $0\ 1011\ 0001\ 1111$, 高三位片选信号为 010 , 表示 2, 因此对应第 2 片芯片, 该芯片的起始地址为 $0\ 1000\ 0000\ 0000$, 即 0800H , 故选 D。

(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

分析: 在 DRAM 分时复用的情况下, 芯片引脚个数取决于行地址线和列地址线中的较大值, 对于一个 $2\text{K} \times 1$ 位的 DRAM 芯片, 总共需要 11 条地址线, 只有当一个取 5、另一个取 6 时才能使引脚数最小, 而 DRAM 的刷新开销取决于行数, 因此行地址线为 5, 列地址线为 6, 所以行数为 $2^5 = 32$, 列数为 $2^6 = 64$, 故选 C

(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

分析: 总线宽度为 64 位, 即 8 字节, 则总带宽为 $8 \times 1333 \times 3 = 31.992$, 故选 B

(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

分析: 根据模块序号 = 访存地址 % 存储器交叉模块数, 即计算“访存地址 % 4”可以得到上述序列的模块序号分别为 1、2、3、0、1、2、3、0、0。发生访存冲突的规则是给定的访存地址在相邻的 4 次访问中出现在同一个存储模块内, 因此 8004 和 8000 对应的模块号均为 0, 说明这两次访问出现在同一模块内且相邻, 会发生冲突, 故选 D

(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

分析: DRAM 采用电容存储, 所以每隔一段时间要刷新一次, 否则存储的信息就会丢失, 故选 B

(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

分析: RAM 一般都采取随机存取方式, CD 排除。EPROM 时可擦除可编程 ROM, 芯片可重复擦除和写入, 采用随机存取方式, 故选 B

(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

分析：闪存的信息确实可读可写，但是读写速度并不一样快，故选 A

(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

分析：时间局限性指一旦一条指令执行了，在不久的将来他可能会被再次执行；空间局限性是一旦一个存储单元被访问，那么它附近的存储单元也会很快被访问。因此对于数组 a 的访问一般都有时间局限性和空间局限性，故选 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

分析：由于共有 16 块 cache，且采用二路组相联映射，可以分组为 $16 / 2 = 8$ 组。每个主存块大小为 32B，且按字节编址，因此第 129 号单元所在 cache 应为 $129 / 32 + 1 = 5$ 块 cache，由于 cache 从零开始编号，所以 cache 组号应为 $5 - 1 = 4$ ，故选 C

(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 (A)。

- A. 1 B. 2 C. 3 D. 4

(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

分析：一块 4 个字，一个字 32 位，所以块内字节数 $= 4 * 32 / 8 = 16B$ ，按照字节编址，块内地址需要 4 位（4 根线）。4K 容量的 cache 中有 1K 个块， $1K = 2^{10}$ ，所以需要 10 位来表示块的索引。所以标识位就是 $32 - 4 - 10 = 18$ 位，因为采用写回法，所以还有脏位，最后再加上有效位，最终总容量 $= 4K * 32 + 1K * (18 + 1 + 1) = 148K$ ，故选 C

(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

分析：举个例子，比如第一步在进行数据 cache 的运算 $x + y$ 时，第二步在指令 cache 这一步时，如果此时这条指令需要前面的运算结果，就需要等数据 cache 运算完，才能进行指令 cache 运算，这样就产生了数据相关的冲突，故选 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

分析：取指令之后会根据 x 的地址去内存中取出 x 的数值，若指令所在的页面位于 TLB 中，则不用去访问内存，直接根据 TLB 的实页号去找相应的页。若要寻找的页位于 cache 中，则无需访问内存，直接取出即可。取出 x 并完成运算后，通过写穿法进行写入，需要将 cache 和内存一起写入，所以必须要访问一次内存，因此最少次数为 1. 故选 B

(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

分析：TLB 中存放了一些经常用到的页表（副本），而页表 Page 是用来翻译物理地址的，翻译完成后用来查询内存中的指令和数据，而 cache 存放的是内存的副本。因此，TLB 命中，页表 Page 一定命中，故选 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

分析：虚拟存储地址空间是 4GB， $4GB = 2^2 * 2^{30} = 2^{32}$ ，则虚地址长度为 32 位，主存地址空间是 256MB， $256MB = 2^8 * 2^{20} = 2^{28}$ ，所以主存地址长度为 28 位。页面大小是 4KB， $4KB = 2^2 * 2^{10} = 2^{12}$ ，则页内地址长度为 12 位。而虚拟地址 03FFF180H 是页内地址，其中 03FFFH 为虚页号，查询 TLB 发现该页在主存中，实页号为 0153H，所以虚实地址变换后的结果为 0153180H，故选 A。

(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

分析：缺页中断和普通中断的区别是缺页中断是在指令执行过程中，而普通中断是在指令执行结束完成的一小段中断检测间隔内，所以处理完缺页中断应该回到发生缺页的指令，故选 D。

4.3 简答题

(2) .为什么在存储芯片中设置片选输入端？

答：主要目的是允许 CPU 选择与其通信的特定存储芯片，因为在计算机中往往不能只通过一片存储芯片就满足计算机的存储需求，通常需要用多个存储芯片进行字位扩展。当 CPU 需要访问内存中的内容时，就需要知道应该访问哪个芯片，所以就需要在存储芯片中设置片选输入端。

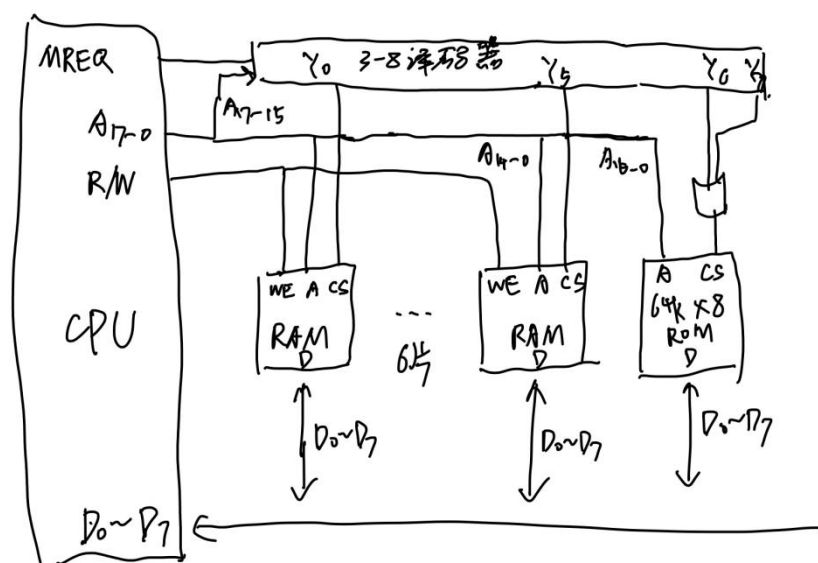
(6) .直接相联映射方式为什么不需要使用替换算法？

答：因为在直接相联映射方式下，每个主存块只能映射到 cache 中的一个特定块。也就是说当我们需要将一个主存块加载到 cache 中时，如果该 cache 块已经被其他主存块映射，我们的替换策略就是直接将该 cache 块替换为待加载的主存块，并不需要进行选择。

(7) .为什么要考虑 cache 的一致性？

答：cache 一致性问题是由多个 CPU 或处理器共享同一块主存储器引起的，由于每个处理器都有自己的 cache，当一个处理器修改了主存中的数据时，它的 cache 中的数据就和主存储器中的数据不一致了。若其他处理器不知道这个修改，就会导致处理器中的数据不一致，引发各种错误和问题。所以，为了保证多个处理器共享一块主存储器时的数据一致性，就需要考虑 cache 一致性。当一个处理器修改了主存储器中的数据时，就必须通知其他处理器更新它们的 cache，来保证数据的一致性。

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。



4.10 用 $16K \times 1$ 位的 DRAM 芯片构成 $64K \times 8$ 位的存储器，设存储器的读写周期为 $0.5\mu s$ ，要使 CPU 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？若每行刷新闻隔不超过 $2ms$ ，该方式下刷新信号的产生周期是多少？

答：由于存储器读写周期为 $0.5\mu s$ ，要使 CPU 在 $1\mu s$ 内至少访问存储器一次，采用分散刷新或异步刷新方式比较合适。若每行刷新闻隔不超过 $2ms$ ，则周期为 $2ms / 128 = 15.625\mu s$

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	9ABEFC00		

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。

	cache行	标志	数据
全相联	0	000000	87568536
	1	000002	87792301
	2	004001	9ABEFC00
	3	007FFF	4FFFFC68
	4	3FFFFFFE	01BF2460
直接相联	0000	00	87568536
	0002	00	87792301
	0001	01	9ABEFC00
	3FFF	01	4FFFFC68
	3FFE	FF	01BF2460
组相联	000	000	87568536
	002	000	87792301
	001	004	9ABEFC00
	0FFF	007	4FFFFC68
	0FFE	3FF	01BF2460

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。

答：由于每个字块有 8 个字，初态为空，因此 CPU 读第 0 号单元时未命中，必须访问主存，并将改字所在的主存块调入 cache 第 0 组的任一块内，接着 CPU 读入 1-7 单元时均命中。同理，CPU 读 8, 16, ..., 88, 96 时均未命中。可见，CPU 在连续读入 100 个字中共有 13 次未命中，后面 10 次循环读的全部命中，命中率为： $(100 * 10 - 13) / (100 * 10) = 0.987 = 98.7\%$

（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：设 cache 周期为 t ，则主存周期为 $10t$ ，没有 cache 的访问时间为 $10t * 100 * 10$ ，有 cache 的访问时间为 $t * (100 * 10 - 13) + 10t * 13$ 。则提高的倍数为： $(10t * 100 * 10) / (t * (100 * 10 - 13) + 10t * 13) = 8.95$

4.16 某页式虚拟存储器共 8 页，每页为 1KB，主存容量为 4KB，页表如表 4.14 所示。

表 4.14 虚拟存储器页表

虚页号	0	1	2	3	4	5	6	7
实页号	3	2	1	2	3	1	0	0
装入位	1	1	0	0	1	0	1	0

（1）失效的页有哪几页？

答：失效的页分别是 2、3、5、7

（2）虚地址 0、3028、1023、2048、4096、8000 的实地址分别是多少？

答：0、3028、1023、4096 的实地址分别为 3072、2004、4095、0（均为 16 进制）。2048 与 8000 出现了缺页。

4.18 某计算机采用页式虚拟存储管理方式，按字节编址，虚拟地址为 32 位，物理地址为 24 位，页大小为 8KB；TLB 采用全相联映射；cache 数据区大小为 64KB，按二路组相联方式组织，主存块大小为 64B。存储访问过程的示意图如图 4.57 所示。请回答下列问题。

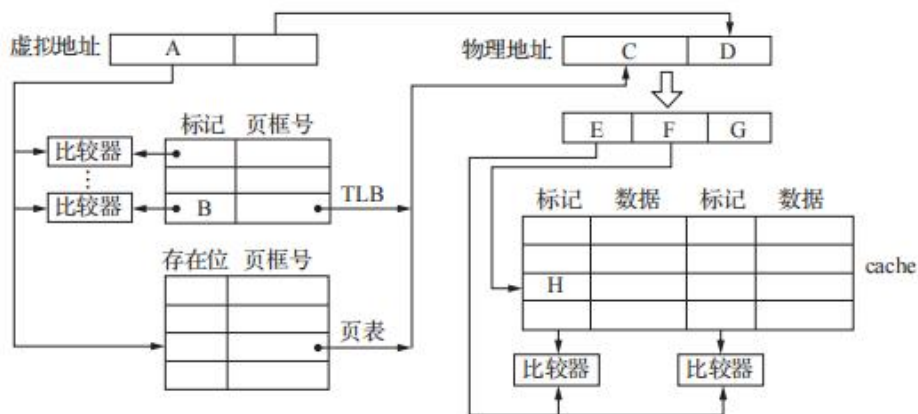


图 4.57 存储访问过程示意图

(1) 图中字段 A ~ G 的位数各是多少？TLB 标记字段 B 中存放的是什么信息？

答：A~G 的位数分别为 19、19、11、13、9、9、6，TLB 中标记字段 B 的内容是虚页号，表示该 TLB 项对应哪个虚页的页表项

(2) 将块号为 4099 的主存块装入 cache 中时，映射的 cache 组号是多少？对应 H 字段的内容是什么？

答：映射的 cache 组号为 3，对应的 H 字段内容为 0 0000 1000B

(3) cache 缺失处理的时间开销大还是缺页处理的时间开销大？为什么？

答：因为缺页处理需要访问磁盘，而 cache 缺失之邀访问主存，cache 缺失带来的开销小，而处理缺页的开销大

(4) 为什么 cache 可以采用写穿策略，而修改页面内容时总是采用写回策略？

答：因为采用写穿策略时需要同时写快速存储器和慢速存储器，而写磁盘比写主存慢得多，所以 cache 可以采用写穿策略，而虚存则应采用写回策略

4.19 某计算机采用页式虚拟存储管理方式，按字节编址。CPU 进行存储访问的过程如图 4.58 所示。

回答下列问题。

(1) 主存的物理地址占多少位？

答：物理地址映射到 cache 后的地址位数不变，从图中可以看出映射到 cache 后的物理地址占 $20 + 3 + 5 = 28$ 位

(2) TLB 采用什么映射方式？TLB 是用 SRAM 还是用 DRAM 实现？

答：采用全相联映射方式，可以把页表内容调入任一块空 TLB 项中，TLB 中每项都有一个比较器，没有映射规则，只要空闲就行。TLB 采用的是 SRAM 实现，读写速度更快，多用于容量较小的高速缓冲存储器

(3) cache 采用什么映射方式？若 cache 采用 LRU 替换算法和写回策略，则 cache 每行中除数据

(Data)、tag 和有效位外，还应有哪些附加位？cache 总容量是多少？cache 中有效位的作用是什么？

答: cache 中每组有两行, 采用 2 路组相联映射; 应该还有替换算法控制位和一致性维护位; 主存字块标记占 20 位, 组号占 3 位, cache 中有 $2^3 = 8$ 组, 每组两行, cache 中有 $8 * 2 = 16$ 行, 行内地址占 5 位, 计算机按字节编址, cache 大小为 $2^5 * 1B = 32B$; Cache 每行标记位数=主存字块标记位数+有效位数+替换算法控制位+脏位位数=(20+1+1+1)bit=23bit。Cache 每行数据位数=32B=32×8bit=256bit。Cache 总容量=(Cache 每行标记位数+Cache 每行数据位数)×Cache 行数=(23bit+256bit)×16=558B;

(4) 若 CPU 给出的虚拟地址为 0008 C040H, 则对应的物理地址是多少? 是否在 cache 中命中? 说明理由, 若 CPU 给出的虚拟地址为 0007 C260H, 则该地址所在主存块映射到的 cache 组号是多少?

答: 虚拟地址 0008C 040H, 前半部分为虚页号, 后半部分为页内地址, 查找 TLB 中 Tag 为 0008CH 的项, 有效位为 1, 实页号为 0040H, 得到其物理地址为 0040 040H; 物理地址高 20 位对应主存子块标记, 所以物理地址 00400 40H 的主存字块标记为 00400H, 查 cache 表中 Tag 为 00400H 的项, 左边有一个 00400H 满足, 但有效位为 0, 不命中, 右边不存在 00400H, 也不命中, 所以访问 cache 不会命中; 物理地址的低 12 位和虚拟地址低 12 位相同, 对于虚拟地址为 0007C260H, 低 12 位为 260H = 0010 0110 0000B。根据物理地址的结构, 物理地址的后八位 0110 0000B 的前三位 011B 是组号, 因此该地址所在的主存映射到 cache 组号为 011B = 3。

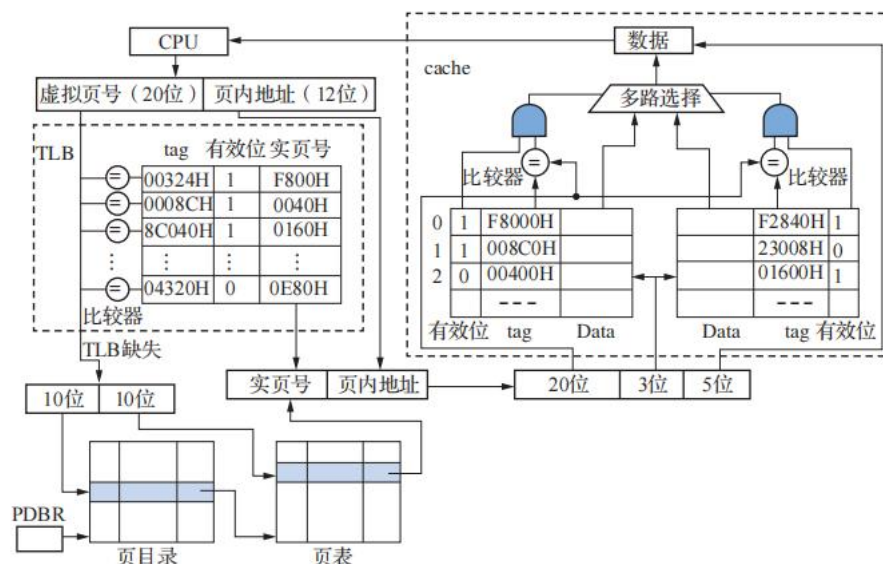


图 4.58 页式虚拟存储器访问过程示意图