

习题 6.7

用GCD测试法判断下面的循环中是否存在循环携带的真数据相关。

```
for(i = 2; i <= 100 ; i += 2)  
    a[i] = a[i-1];
```

在这个循环中 $a=1$ ， $b=0$ ， $c=1$ ， $d=-1$ ，这样 $\text{GCD}(a,c)=1$ ， $d-b=-1$ ，由于前者可以整除后者，但由于该循环的步长为2，而 i 仅和 $i-1$ 有关，因此不存在循环携带的真数据相关。

习题 6.8

下面这段循环完成点积运算，寄存器F2的初始值为0.试结合使用循环展开和基本指令调度技术，消除其中的所有流水线“空转”周期。假设流水线延迟如表所示，分支指令也会带来1个“空转”周期。

产生结果的指令	使用结果的指令	延迟(时钟周期数)
浮点计算	另一个浮点计算	3
浮点计算	浮点store(S.D)	2
浮点load(L.D)	浮点计算	1
浮点load(L.D)	浮点store(S.D)	0

```
loop:  L.D      F0,0(R1)
      L.D      F4,0(R2)
      MUL.D    F0,F0,F4
      ADD.D    F2,F0,F2
      DADDUI   R1,R1,#-8
      DADDUI   R2,R2,#-8
      BNE     R1,R3,loop
```

原程序周期数：
每对元素所需的时钟周期数=12，
其中空转数=5；

新程序

```
loop:  L.D  F0,16(R1)      ;F0 ← A(i+2)
        L.D  F4,16(R2)      ;F4 ← B(i+2)
        L.D  F6,8(R1)       ;F6 ← A(i+1)
        MUL.D F0,F0,F4      ;F0 ← A(i+2) × B(i+2)
        L.D  F8,8(R2)       ;F8 ← B(i+1)
        L.D  F10,0(R1)      ;F10 ← A(i)
        MUL.D F6,F6,F8      ;F6 ← A(i+1) × B(i+1)
        ADD.D F2,F0,F2      ;F2 ← F2 + A(i+2) × B(i+2)
        L.D  F12,0(R2)      ;F12 ← B(i)
        DADDUI R1,R1,-24    ;R1 ← R1-24
        MUL.D F10,F10,F12   ;F10 ← A(i) × B(i)
        ADD.D F2,F6,F2      ;F2 ← F2 + A(i+1) × B(i+1)
        DADDUI R2,R2,-24    ;R2 ← R2-24
        BNE  R1,R3,loop     ;若 R1 ≠ R3, 循环
        ADD.D F2,F10,F2     ;F2 ← F2 + A(i) × B(i)
```

新程序周期数：每对元素所需的时钟周期数=16/3=5.3，
其中空转数=1/3=0.3

习题 7.9

- 假设在3000次访存中，第一级cache不命中110次，第二级cache不命中55次。试问：在这种情况下，该cache系统的局部不命中率和全局不命中率各是多少？

- 解：

第一级cache不命中率(全局和局部)是 $110/3000$ ，即3.67%

第二级cache的局部不命中率是 $55/110$ ，即50%

第二级cache的全局不命中率是 $55/3000$ ，即1.83%

习题7.10

给定以下的假设，试计算直接映象Cache和两路组相联Cache的平均访问时间以及CPU的性能。由计算结果能得出什么结论？

- (1)理想Cache情况下的CPI为2.0，时钟周期为2ns，平均每条指令访存1.2次；
- (2)两者Cache容量均为64KB，块大小都是32字节；
- (3)组相联Cache中的多路选择器使CPU的时钟周期增加了10%；
- (4)这两种Cache的失效开销都是80ns；
- (5)命中时间为1个时钟周期；
- (6)64KB直接映象Cache的失效率为1.4%，64KB两路组相联Cache的失效率为1.0%。

习题7.10(解答)

- 平均访问时间=命中时间+失效率×失效开销
- 平均访问时间1-路= $2.0+1.4\% * 80=3.12\text{ns}$
- 平均访问时间2-路= $2.0*(1+10\%)+1.0\% * 80=3.0\text{ns}$
- 两路组相联的平均访问时间比较低
- $\text{CPUtime} = (\text{CPU执行} + \text{存储等待周期}) * \text{时钟周期}$
- $\text{CPU time} = \text{IC} (\text{CPI执行} + \text{总失效次数}/\text{指令总数} * \text{失效开销}) * \text{时钟周期}$
- $= \text{IC} ((\text{CPI执行} * \text{时钟周期}) + (\text{每条指令的访存次数} * \text{失效率} * \text{失效开销} * \text{时钟周期}))$
- $\text{CPU time 1-way} = \text{IC}(2.0*2+1.2*0.014*80)=5.344\text{IC}$
- $\text{CPU time 2-way} = \text{IC}(2.2*2+1.2*0.01*80)=5.36\text{IC}$
- 相对性能比: $5.36/5.344=1.003$
- 直接映象cache的访问速度比两路组相联cache要快1.04倍, 而两路组相联Cache的平均性能比直接映象cache要高1.003倍。因此这里选择两路组相联。

习题7.11

- 伪相联中，假设在直接映象位置没有发现匹配，而在另一个位置才找到数据（伪命中）时，需要1个额外的周期，而且不交换两个Cache中的数据，失效开销为50个时钟周期。试求：
- 推导出平均访存的时间公式。
- 利用（1）中得到的公式，对于2KBCache和128KBCache，重新计算伪相联的平均访存时间。请问哪一种伪相联更快？
- 假设 2KB直接映象Cache的总失效率为0.098，2路相联的总失效率为0.076；
- 128KB直接映象Cache的总失效率为0.010，2路相联的总失效率为0.007。

习题7.11(解答)

- 不管作了何种改进，失效开销相同。不管是否交换内容，在同一“伪相联”组中的两块都是用同一个索引得到的，因此失效率相同，即：
失效率_{伪相联} = 失效率_{2路}。
- 伪相联cache的命中时间等于直接映象cache的命中时间加上伪相联查找过程中的命中时间*该命中所需的额外开销。
- 命中时间_{伪相联} = 命中时间_{1路} + 伪命中率_{伪相联} × 1
- 交换或不交换内容，伪相联的命中率都是由于在第一次失效时，将地址取反，再在第二次查找带来的。
- 因此 伪命中率_{伪相联} = 命中率_{2路} - 命中率_{1路} = (1 - 失效率_{2路}) - (1 - 失效率_{1路}) = 失效率_{1路} - 失效率_{2路}。交换内容需要增加伪相联的额外开销。
- 平均访存时间_{伪相联} = 命中时间_{1路} + (失效率_{1路} - 失效率_{2路}) × 1
- + 失效率_{2路} × 失效开销_{1路}
- 将题设中的数据带入计算，得到：
- 平均访存时间2Kb = 1 + (0.098 - 0.076) * 1 + (0.076 * 50) = 4.822
- 平均访存时间128Kb = 1 + (0.010 - 0.007) * 1 + (0.007 * 50) = 1.353
- 显然是128KB的伪相联Cache要快一些。

习题 7.12

假设采用理想存储器系统时的基本CPI是1.5，主存延迟是40个时钟周期；传输速率为4B/时钟周期，且cache中50%的块是修改过的。每个块中有32B，20%的指令是数据传送指令。并假设没有写缓存，在TLB不命中的情况下，需要20时钟周期，TLB不会降低cache命中率。CPU产生指令地址或cache不命中时产生的地址有0.2%没有在TLB中找到。

(1)在理想TLB情况下，计算均采用写回法16KB直接映像混合Cache、16KB两路组相联混合cache和32KB直接映像混合cache机器的实际CPI

(2)在实际TLB情况下，用(1)的结果，计算均采用写回法16KB直接映像混合Cache、16KB两路组相联混合cache和32KB直接映像混合cache机器的实际CPI

习题7.12

(1) 假设TLB不命中率=0

$$\begin{aligned}\text{均摊不命中开销} &= \text{不命中率} \times \{[40 + 32B/4B + 0 \times 20] + 50\% \times [40 + 32B/4B + 0 \times 20]\} \\ &= \text{不命中率} \times \{48 + 50\% \times 48\} = \text{不命中率} \times 72\end{aligned}$$

$$\text{实际CPI} = 1.5 + 1.2 \times \text{不命中率} \times 72 = 1.5 + \text{不命中率} \times 86.4$$

带入3种Cache结构的不命中率得：

Cache结构	不命中率	实际CPI
16KB直接混合映像	0.029	4.0056
16KB两路混合映像	0.022	3.4008
32KB直接混合映像	0.020	3.2280

习题7.12

(2) 假设TLB不命中率=0.2%

均摊不命中开销=不命中率 \times {[40+32B/4B+0.2% \times 20]+50% \times [40+32B/4B+0.2% \times 20]}
=不命中率 \times {48.04+50% \times 48.04}=不命中率 \times 72.06

实际CPI2=1.5+1.2 \times 不命中率 \times 72.06=1.5+不命中率 \times 86.472

带入3种Cache结构的不命中率后得

Cache结构	不命中率	实际CPI
16KB直接混合映像	0.029	4.0077
16KB两路混合映像	0.022	3.4024
32KB直接混合映像	0.020	3.2294

习题7.14

- 假设一台计算机具有以下特性：
- 95%的访存在Cache中命中；
- 块大小为两个字，且失效时整个块被调入；
- CPU发出访存请求的速率为 10^9 字/秒；
- 25%的访存为写访问；
- 存储器的最大流量为 10^9 字/秒（包括读和写）；
- 主存每次只能读或写一个字；
- 在写回策略时，Cache中有30%的块被修改过；
- 写失效时，Cache采用写分配法。
- 现欲给计算机增添一台外设，为此想先知道主存的频带已经使用了多少。试对于以下两种情况计算主存频带的平均使用比例。
 - (1) 写直达Cache；
 - (2) 写回法Cache。

习题7.14(解答)

- 采用按写分配
 - (1) 写直达cache访问命中，有两种情况：
 - 读命中，不访问主存；
 - 写命中，更新cache和主存，访问主存一次。
 - 访问失效，有两种情况：
 - 读失效，将主存中的块调入cache中，访问主存两次；
 - 写失效，将要写的块调入cache，访问主存两次，再将修改的数据写入cache和主存，访问主存一次，共三次。上述分析如下表所示。

访问命中	访问类型	频率	访存次数
Y	读	$95\% \times 75\% = 71.3\%$	0
Y	写	$95\% \times 25\% = 23.8\%$	1
N	读	$5\% \times 75\% = 3.8\%$	2
N	写	$5\% \times 25\% = 1.3\%$	3

$$\begin{aligned} & \text{一次访存请求最后真正的平均访存次数} \\ & = (71.3\% \times 0) + (23.8\% \times 1) + (3.8\% \times 2) + (1.3\% \times 3) = 0.35 \\ & \text{已用带宽} = 0.35 \times 10^9 / 10^9 = 35.0\% \end{aligned}$$

习题7.14(解答)

- （2）写回法cache访问命中,有两种情况:
读命中,不访问主存;
写命中,不访问主存。采用写回法,只有当修改的cache块被换出时,才写入主存;
访问失效,有一个块将被换出,这也有两种情况:
如果被替换的块没有修改过,将主存中的块调入cache块中,访问主存两次;
如果被替换的块修改过,则首先将修改的块写入主存,需要访问主存两次;然后将主存中的块调入cache块中,需要访问主存两次,共四次访问主存。

访问命中	块为脏	频率	访存次数
Y	N	$95\% \times 70\% = 66.5\%$	0
Y	Y	$95\% \times 30\% = 28.5\%$	0
N	N	$5\% \times 70\% = 3.5\%$	2
N	Y	$5\% \times 30\% = 1.5\%$	4

所以:

一次访存请求最后真正的平均访存次数

$$= 66.5\% \times 0 + 28.5\% \times 0 + 3.5\% \times 2 + 1.5\% \times 4 = 0.13$$

$$\text{已用带宽} = 0.13 \times 10^9 / 10^9 = 13\%$$

习题8.11

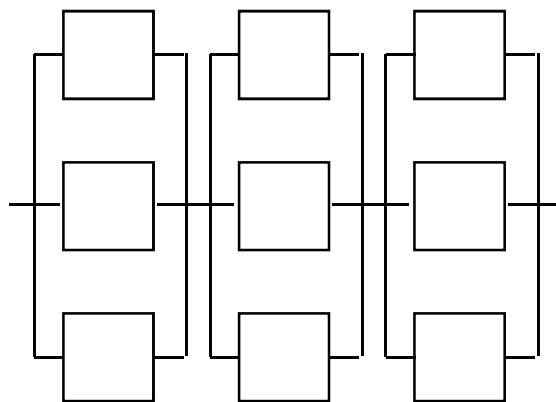
- 假设在一个计算机系统中：
- (1) 每页为32KB，cache块大小为128B
- (2) 对应新页的地址不在cache中，CPU不访问新页中的任何数据；
- (3) Cache中95%的被替换块将再次被读取，并引起一次不命中；
- (4) cache使用写回方法，平均60%的块被修改过；
- (5) I/O系统缓冲能够存储一个完整的cache块
- (6) 访问或不命中在所有cache块中均匀分布；
- (7) 在CPU和I/O之间，没有其他访问cache的干扰；
- (8) 无I/O时，每100万个时钟周期内有18000次不命中；
- (9) 不命中开销是40个时钟周期，如果被替换的块被修改过，则再加上30个周期用于写回主存；
- (10) 假设计算机平均每200万个周期处理一页
- 试分析I/O对于性能的影响有多大

习题8.11（解）

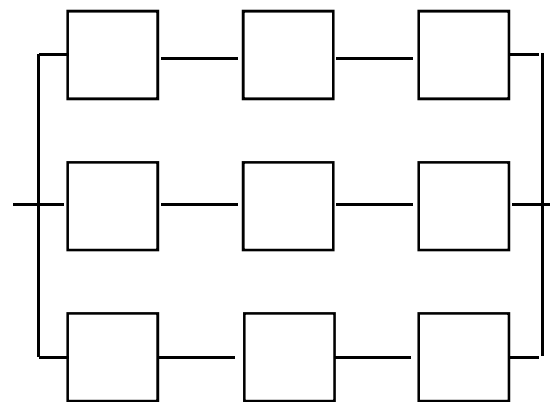
- 每个页有 $32\text{KB}/128=256$ cache块
- 因为cache和内存之间按块传输，所以I/O传输本身不产生cache缺失，但是它可能要替换cache中的有效块。如果这些被替换块中有60%是被修改过，将需要 $(256 \times 60\%) \times 30 = 4608$ 个时钟周期将这些被修改过的块写回主存。
- 这些被替换出去的块中，有95%的后续需要访问，从而产生 $95\% \times 256 = 244$ 次不命中，将再次发生替换。由于被替换的244块中数据是从I/O直接写入cache中，因此所有块都为被修改块，需要写回主存（因为CPU不会直接访问从I/O来的新页中的数据），所以他们不会立即从主存中调入cache，需要时间是 $244 \times (40 + 30) = 17080$ 个时钟周期。
- 没有I/O时，每一页平均使用200万个时钟周期，cache不命中36000次，其中60%被修改过，所需的处理时间为
- $(36000 \times 40\%) \times 40 + (36000 \times 60\%) \times (40 + 30) = 2088000$ 时钟周期
- I/O造成的额外性能损失比例为：
- $(4608 + 17080) / (2000000 + 2088000) = 0.53\%$
- 即产生大约0.53%的性能损失

8.12 混联系统可靠度

假定每个部件的正常工作时间服从指数分布，它们的不可靠度都是 10^{-3} ，比较并串联系统(a)和串并联系统(b)的可靠性哪一个更高。



(a) 并串联系统



(b) 串并联系统

8.12 混联系统可靠度(解：)

可靠度： $R(t)$ ，不可靠度： $F(t)$ ，关系式： $R(t) + F(t) = 1$ 。

在各单元相同情况下，用 n 代表串联倍数， m 代表并联倍数。

①串并联系统：先串联、后并联

可靠度 $R(t) = 1 - [1 - R_i^n(t)]^m$ 。

②并串联系统：先并联、后串联

可靠度 $R(t) = [1 - (1 - R_i(t))^m]^n$ 。

公式用途：RAID 0+1是串并联系统，RAID 1+0是并串联系统

• $F = 10^{-3}$, $R = 0.999$

(1)并串联系统 $(1 - (10^{-3})^3)^3 = (0.9999999999)^3 = 0.9999999997$

(2)串并联系统 $1 - (1 - (0.999)^3)^3 = 0.9999999997$

所以，(1)并串联系统可靠度更大

9.9、设函数的自变量是十进制数表示的处理机变化。现有32个处理器，编号为0、1、...、31。

(1)分别计算下列互连函数

$\text{Cube}_2(12)$ $\sigma(8)$ $\beta(9)$ $\text{PM2I}_{+3}(28)$ $\text{Cube}_0(\sigma(4))$ $\sigma(\text{Cube}_0(18))$

(2)用 Cube_0 和 σ 构成混洗交换网（每步只能用 Cube_0 和 σ 一次），网络直径是多少？

从5号处理机发送数据到7号处理机，最短路径要经过几步？列出经过的处理机编号

(3)采用循环移数网络构成互连网，网络直径是多少？结点度是多少？与2号处理机距离最远的是几号处理机？

解：

(1) $\text{Cube}_2(12) = \text{Cube}_2(01100\text{B}) = 01000\text{B} = 8$

$\sigma(8) = \sigma(01000\text{B}) = 10000\text{B} = 16$

$\beta(9) = \beta(01001\text{B}) = 11000\text{B} = 24$

$\text{PM2I}_{+3}(28) = \text{PM2I}_{+3}(11100\text{B}) = 11100\text{B} + 01000\text{B} \bmod 2^5 = 4$

$\text{Cube}_0(\sigma(4)) = \text{Cube}_0(\sigma(00100\text{B})) = 01001\text{B} = 9$

$\sigma(\text{Cube}_0(18)) = \sigma(\text{Cube}_0(10010\text{B})) = 00111\text{B} = 7$

(2) 2^5 个结点的混洗交换网的直径是 $2n-1=2*5-1=9$;

从5号处理机(00101B)发送到7号处理机(00111B)最短路径经过6步

包括5步左移和1步求反：

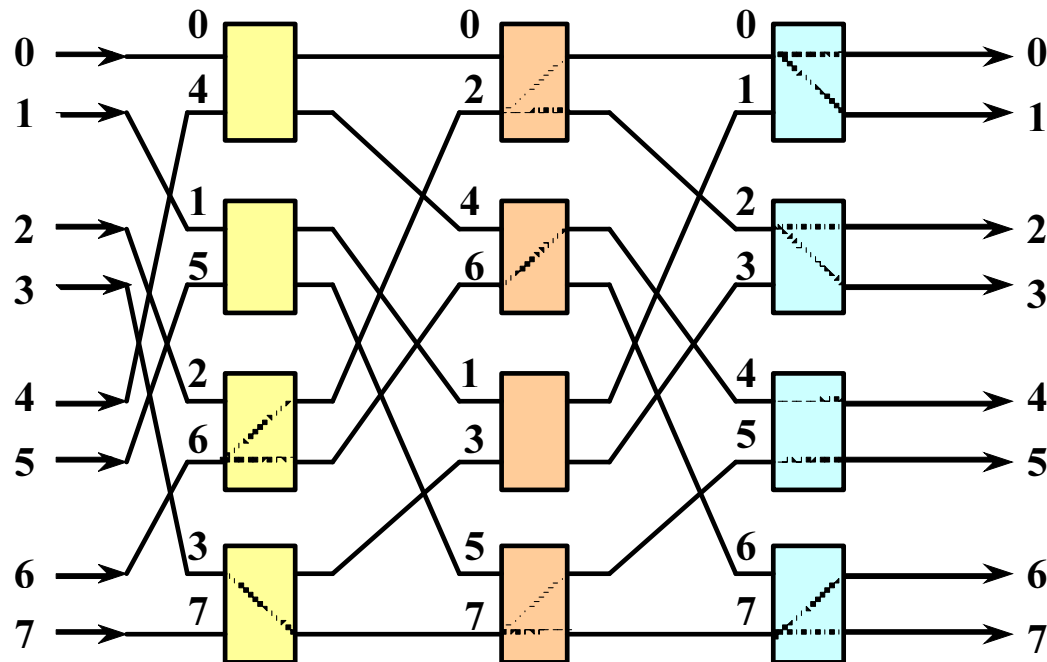
00101B -> 01010B -> 10100B -> 01001B -> 10010B -> 10011B -> 00111B

(3) 网络直径是3；结点度是 $2n-1 = 2*5-1 = 9$ ；12或16，其他对称节点也

是。

习题9.13

- 用一个 $N=8$ 的三级Omega网络连接8个处理机（ $P_0 \sim P_7$ ），8个处理机的输出端分别依次连接Omega的8个输入端 $0 \sim 7$ ，8个处理机的输入端分别依次连接Omega的8个输出端 $0 \sim 7$ ，如果处理机 P_6 要把数据播送到处理机 $P_0 \sim P_4$ ，处理机 P_3 要把数据播送到处理机 $P_5 \sim P_7$ ，那么，Omega网络能否同时为它们的播送要求实现连接，画出实现播送的Omega网络的开关状态图。



习题10.6

- 一个具有32台处理机的系统，对远程存储器访问时间是2000ns。除了通信以外，假设计算中的访问均命中局部存储器。当发出一个远程请求时，本地处理机挂起。处理机的时钟周期是10ns，假设指令基本的CPI为1.0（假设所有访存均命中cache）。对于下述两种情况：
 - （1）没有远程访问；
 - （2）0.5%的指令需要远程访问
- 试问前者比后者快多少？

解：远程访存时钟周期为 $2000/10=200$

则 有远程访问情况

$$CPI_2 = CPI_1 + p \times C = 2$$

因此前者比后者快2倍

习题10.9

- 采用排队锁和**fetch-and-increment**重新实现栅栏同步，并将它们分别与采用旋转锁实现的栅栏同步进行性能对比。

- 解： **fetch-and-increment(count);**
- **if(count==total){** //进程全部到达
- **count = 0;** //重置计数器
- **release=1;** //释放进程
- **}**
- **else{** //还有进程未到达
- **spin(release=1);** //等待信号
- **}**
- 当有N个处理器时，上述代码执行**fetch-and-increment**操作N次，当访问释放操作的时候，有N个**cache**未命中。当最后一个处理器到达栅栏条件后，**release**补置为“1”，此时有N-1个**cache**未命中（对于最后一个到达栅栏的处理器，当它读**release**的时候，将在主存中命中）。所以，共有 $3N-1$ 次总线传输操作。如果有10个处理器，则共有29次总线传输操作，总共需要2900个时钟周期。



课程结束...