

5. 考虑一个深度流水线处理器，无分支指令时其基本 CPI 为 1。对于分支指令采用两种方案，方案 A 使用一个分支目标缓存（BTB），缓存缺失代价为额外 3 个周期，缓存命中但预测错误的代价为额外 4 个周期，缓存命中且预测正确则无分支代价。假设这个 BTB 的命中率为 90%，预测正确率为 90%。方案 B 不使用分支预测，分支代价固定为额外 2 个周期。假设分支频率为所有指令的 15%，则处理器采用方案 A 比采用方案 B 快多少？

$$\text{方案A: } CPI_a = 0.85 + 0.15 \left(0.1 \times 4 + 0.9 \times 0.1 \times 5 + 0.9 \times 0.9 \times 1 \right) \\ = 1.099$$

$$\text{方案B: } CPI_b = 0.85 + 0.15 \times 3 = 1.3$$

$$S = \frac{CPI_b}{CPI_a} = 1.183$$

\therefore A 的速度是 B 的速度的 1.183 倍
快了 18.3%

12. 考虑如下的代码片段：

	li	a0,0
	li	a4,10000
0xe44:	addi	a1,a,0
Loop:	addi	a3,a,2
	<u>rem</u>	a2,a1,a3
0xe44:	bne	a2,a,0,Rem2

$$\begin{aligned}a_0 &= 0 \\a_4 &= 0 \\a_1 &= 0 \\a_3 &= 2\end{aligned}$$

if (a1 % a3 != 0) j to Rem 2

```

Rem2: addi    a3,a0,5      93
      rem     a2,a1,a3      if(a1)
0xe84: bne     a2,a0,End   //B2
          #...CodeB
End:   addi    a1,a1,1
0xec0: bne     a1,a4,Loop  //B3

```

$$\text{if } (a_1 \% a_3 \neq 0) \quad a_1 = a_1 + 1$$

- 1) 写出与该汇编代码功能一致的 C 语言代码。
 - 2) 无分支预测时，上述代码中的三条 bne 指令发生跳转的比例分别是多少？
 - 3) 引入一个静态分支预测器，该预测器对向前跳转总是给出“跳转”预测，对向后跳转总是给出“不跳转”预测，则上述代码中的三条 bne 指令的预测准确率分别是多少？

int i=0
while (i < 10000)
{ if (i % 2 == 0) # code A
 if (i % 5 == 0) # code B
 if

$$2) B_1: \text{t13y } 50\% \quad (i\%_2 = 0) \\ B_2: \text{t13y } 80\% \quad (i\%_5 = 0)$$

B₃: 正向 99.9% (从 1 到 10000, 共 9999 次跳转)

3) B. 短期预测呈北转，准确率50%

B₂ 這是預期跳離，惟兩手及

β_3 等效吸收系数 100%

```

        li      a0,0
        li      a4,10000
        addi   a1,a0,0
Loop:  addi   a3,a0,2
        rem    a2,a1,a3
0xe44: bne   a2,a0,Rem2 //B1
        #...CodeA

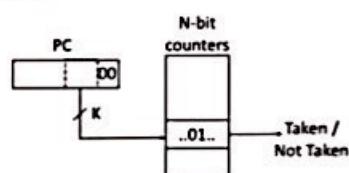
Rem2:  addi   a3,a0,5
        rem    a2,a1,a3
0xe84: bne   a2,a0,End //B2
        #...CodeB

End:   addi   a1,a1,1
0xec0: bne   a1,a4,Loop //B3

```

- 写出与该汇编代码功能一致的C语言代码。
- 无分支预测时，上述代码中的三条**bne**指令发生跳转的比例分别是多少？
- 引入一个静态分支预测器，该预测器对向前跳转总是给出“跳转”预测，对向后跳转总是给出“不跳转”预测，则上述代码中的三条**bne**指令的预测准确率分别是多少？

13. 仍考虑题 12 中的代码片段，现引入局部预测器，如下图所示。该预测器使用 PC 的第 $[K+2:3]$ 共 K 位索引一张预测器表，该表的每个表项是一个 N -bit 的计数器，计数器的最高位用于预测是否跳转（1 为跳转，0 为不跳转），并根据实际跳转结果更新计数器的值（跳转自增 1，增至 2^N-1 后不再变化；不跳转自减 1，减至 0 后不再变化）。假设所有计数器的初始值均为 0。



- 1) 要保证上述代码片段被映射到不同的局部预测器, K 的最小值是多少?
 - 2) 要使该预测器对三条 bne 指令的预测准确率均不低于题 12 中的静态预测器, N 的最小值是多少?
 - 3) 对上述给出的最小 N, 在程序稳态时, 三条 bne 指令的预测准确率分别是多少?

2) B₁: FTFTFT · FT

计数 $D(0|0)$ = 0 | $N=1$: 1 位随机数 $\overline{000}$

答：DD DD DD DD 0100 0100 01 N=2：倍爾率 50%，易知 N≥3 的倍爾率是 100%。
∴ FFTTFT FFTTFT FFTTFT

准确率约 60%

= 2 00 00 01 01 11 11 10 11 11 11 11 10 11 11 11 11 若对这段代码多调用，将输出什么？

3. 000 000 100 010 011 100 011 100 100 110 111 110 111 110 111 111

易 N>2时 精确度会有所下降

$\beta_3 : \text{FFFFFFFFFF}$... T

易系 $N \geq 2$ 时准确率必大于 0.5%

$$\text{No. } N_{\min} = 2$$

$$3) f_1 = 50\%$$

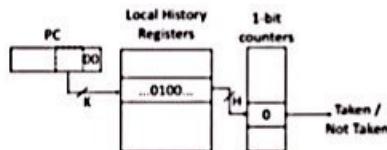
$$\beta_2 = 80\%$$

B₂: 99.99%

1) $0xe44 : (1110\text{ }00010000)_2$
 $0xe84 : (1110\text{ }1000\text{ }0100)_2$
 $0xec0 : (1110\text{ }1100\text{ }0000)_2$

14. 仍考虑题 12 中的代码片段，现引入局部分支历史，如图所示。该预测器使用 PC 的第 $[K+2]:3$ 共 K 位索引一个局部分支历史表，其每个表项是一个 H 位的局部分支历史。

该 H 位的历史被进一步用于索引一张单比特计数器构成的预测表（1 为跳转，0 为不跳转）。计数器会根据实际跳转结果进行更新。



假设 K 的值足够大，使得上述代码片段中的不同分支会被映射到局部分支历史表中的不同位置，则为了使得三条 bne 指令都能在程序稳态时被完全准确地预测，H 的最小值是多少？

$$H=1: "0" \rightarrow 1 \} \rightarrow \text{跳转}$$

$$H=2: "1" \rightarrow \text{可得1, 也可得0, X}$$

B_1 : i%2 ≠ 0 则跳 $\rightarrow 010101 \dots 0101$

B_2 : i%5 ≠ 0 则跳 $\rightarrow 01110101 \dots 011111$ $H=3, 111 \rightarrow 1$ 也得 0, \checkmark

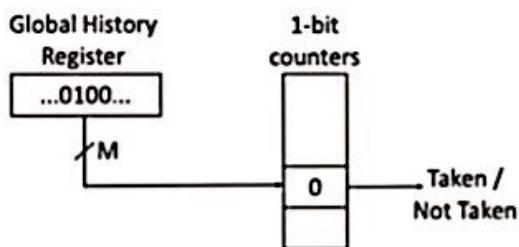
$H=4$: "0111", "110", "110", "1011" $\rightarrow 1, 111 \rightarrow H=4 \checkmark$

B_2 : i%5 == 0 则跳 $\rightarrow 111111 \dots 1110$

H 位代表的逻辑：状态机 前 H 次正确

综上 H 最小值为 4 (至少包含最近 4 次预测)

15. 仍考虑题 12 中的代码片段，现引入全局分支历史，如下图所示。该预测器拥有一个 M 位的 GHR，记录了程序中任意分支的跳转历史。当一个新分支被执行时，跳转分支使得 GHR 左移 1 位并在末位写入 1，未跳转分支则使得 GHR 左移 1 位并在末位写入 0。GHR 被用于索引一张单比特计数器构成的预测表（1 为跳转，0 为不跳转）。计数器会根据实际跳转结果进行更新。



为了使得三条 bne 指令都能在程序稳态时被完全准确地预测，M 的最小值是多少？

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T
F	T	T	T	T	F	T	T	T	F	T	T	T	T	F	
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
B, B ₂ , B ₃															
FFT															
TTT															
F77															
TTT															
F77															
TF7															
FT7															
TTT															
FT7															
TT7															
FFT															
TTT															
FT7															
TT7															
FT7															
TFT															

手写注释：(F=0, T=1)

分析：从手写注释可以看出，程序在某些点上无法准确预测。例如，在第 15 位时，GHR 的值是 1111111111111111，对应的计数器值是 1。然而，由于 GHR 的宽度 M 只有 12 位，因此在第 15 位时，GHR 的值实际上是 1111111111111110。因此，计数器的值应该是 0，而不是 1，导致预测错误。因此，M 的最小值应该是 12。

可知 $M \geq 12$

16. 在实际应用中常常能遇到如下的代码场景：

```
for (int i=0;i<P;i++){           //outer loop  
    for (int j=0;j<Q;j++){   //inner loop  
        //SomeCode  
    }  
}
```

视进入循环体执行为“跳转”，不进入为“不跳转”。现有两种分支预测器方案：方案 A 使用题 13 中的预测器结构 ($N=1$)，方案 B 使用题 14 中的预测器结构 ($H=Q$)。假设 $Q > 2$ 且内循环体 (inner loop) 内部没有分支指令、外循环体 (outer loop) 针对变量 i 的分支总是被预测器忽略、预测器的 K 值足够大、所有计数器的初始值均为 0。试分析当 P 和 Q 满足什么数值关系时，方案 A 的预测准确率优于方案 B？

内循环：TTTTTTTT...TF TTTT...TF
 $\underbrace{\quad\quad\quad}_{Q-1} \quad \downarrow \text{1次} \quad \underbrace{\quad\quad\quad}_{Q-2}$
0 1 1 1 1 1 1 1 ... 1 0
方案 A：预测准确率 $\frac{Q-2}{Q}$

方案 B：第一轮只有最后一跳出循环时正确；
第二次，第三次也只有一跳出循环时正确一次；从第三轮开始均正确。

预测准确率 $\frac{(P-2)Q+2}{PQ}$

$$\frac{Q-2}{Q} > \frac{(P-2)Q+2}{PQ} \Rightarrow Q > P+1$$

17. 考虑如下的指令序列：

Loop: lw a4,0(a3)
addi a3,a3,4
addi a1,a1,-1
B1: beqz a4,B2
addi a2,a2,1
B2: bnez a1,Loop

(1) $B_1: FTFTFTFTFT$ 错4次
 $B_2: TTTTTTTTTF$ 错1次

global history

(2)



$B_1: FTFTFTFT$
 $B_2: TTTTTTTTTF$

$B_1:$ 错4次

$B_2:$ 如果两个分支机初值
则错1次

假设 $a1$ 初值值为 n ($n > 0$), $a2$ 初始值为 0, $a3$ 初始值为 p (p 为一个指向 32 位整型数组首地址的指针)。

- 1) 假设处理器使用 2 位局部预测器, 分支 $B1$ 和 $B2$ 映射到不同的预测器表项。若 $n=8$ 且数组的数据模式为 $p[] = \{1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0\}$ 。则上述代码执行过程中一共会发生多少次错误预测?
- 2) 现引入 1 位的全局分支历史, 1)中的其他假设不变, 则上述代码执行过程中一共会发生多少次错误预测?
- 3) 若改为 2 位的全局分支历史表, 1)中的其他假设不变, 则上述代码执行过程中一共会发生多少次错误预测?
- 4) 比较上述结果, 分析在该情境中全局分支历史表的位数对预测准确率有怎样的影响?
当 n 非常大时, 上述哪种预测器表现最好?
- 5) 当数组 $p[]$ 的数据模式变为在 0 和 1 之间以均等概率随机取值时, 4 中的结论有什么变化?

2位全局分支历史，1位全局分支历史 01235
F T T T T T T T T T T F T T T T T T T F
错5次

4) 全局

5) 两种预测结果差不

18. 解释为什么即使在顺序的 5 级 RISC 流水线中，指令引发的异常也可能会乱序产生？为了支持精确的异常处理，流水线是如何做到对乱序产生的异常进行（按程序顺序的）顺序处理的？

① 因为每条指令执行的周期不同，且指令与指令之间还会存在数据依赖、控制冲突等不可控因素。

② 指令在 WB 阶段必须保证按照源程序的顺序进行写回。这样如果有指令产生中断或异常，处理器可以禁止在程序顺序上位于这条指令之后的指令的写回操作。也就是说如果顺序写回。当指令被发射时，处理器在重排序缓冲区中保留了记录，当指令执行完成后需要写回时，若在它之前尚未写回的指令，则该指令继续在重排序缓冲区等待。

这样流水线一旦发生中断需要回滚，没有执行写回操作的指令不会对处理器执行精确的顺序造成影响。

20. 考虑一个拥有浮点单元的单发射乱序处理器，该处理器包含以下假设：
- 处理器的浮点单元包含一个 2 运算周期的加法器、一个 10 运算周期的乘法器，和一个单执行周期的浮点加载/存储单元，加法和乘法器均是完全流水化的。
 - 当发生写回冲突时，更早的指令会获得优先写回权。
 - 浮点指令的结果只能在写回阶段完成后被其他指令使用，整型指令的结果则可以前

馈。

- 处理器使用寄存器重命名，从 T0、T1、T2 起有不受限制的重命名寄存器可用。
- 译码级每周期可以将至多 1 条重命名后的指令添加到 ROB 中，指令通过 ROB 顺序提交且每周期至多提交 1 条指令。指令能够被提交的最早时间是完成写回后的下一个周期。

① 忽略前端取指，指令经过译码、发射、执行和写回后即可完成执行并提交。

现考虑如下的指令序列：

11:	fld	f1, 5(a0)	1
12:	fmul.d	f2, f1, f0	10
13:	fadd.d	f3, f2, f0	2
14:	addi	a0, a0, 8	2
15:	fld	f1, 5(a0)	1
16:	fmul.d	f2, f1, f1	10
17:	fadd.d	f2, f2, f3	2

$$T_3 = 6 + 8$$

1) 如果 ROB 的深度是无限的，将下表补充完全。（部分结果已给出）

	周期				操作码	目标	源 1	源 2
	Decode (ROB enqueue)	Issue	WB	Committed				
11	0	1	2	3	fld	T0	a0	-
12	1	3	13	14	fmul.d	T1	T0	f0
13	2	14	16	17	fadd.d	T2	T1	f0
14	3	19	16	18	addi	T3	a0	-
15	4	5	6	17	fld	T4	T3	-
16	5	19	29	20	fmul.d	T5	T4	Ty
17	6	30	32	33	fadd.d	T6	T5	T2

2) 如果 ROB 仅容纳 2 条指令，当一条指令提交后的下一周期该条目可以被新指令占据。重新将下表补充完全。（部分结果已给出）

	周期				操作码	目标	源 1	源 2
	Decode (ROB enqueue)	Issue	WB	Committed				
11	✓ 0	1	2	3	fld	T0	a0	-
12	✓ 1	3	13	14	fmul.d	T1	T0	f0
13	✓ 4	14	16	17	fadd.d	T2	T1	f0
14	15	16	18	19	addi	T3	a0	-
15	18	19	20	21	fld	T0	T3	-
16	20	21	31	32	fmul.d	T1	T0	T0
17	22	32	34	35	fadd.d	T4	T1	T2