

2013-2014 学年秋冬学期高级计算机体系结构试卷

学号_____ 姓名_____ 分数_____

一、选择题（共 10 题，每题 2 分，共 20 分）

1、根据指令流和数据流的多倍性可以将计算机体系结构分为四大类。其中以阵列处理机和向量处理机为代表的是（ ），以多机计算机系统为代表的是（ ）。

A. SISD, SIMD B. SIMD, MIMD C. MISD, MIMD D. MIMD, SIMD

2、在流水线中，利用简单硬件技术“转发（Forwarding）”可以解决一些数据冒险问题，从而减少流水线停顿，但是并非所有数据冒险都可以通过转发方式处理。下列指令序列中，不能利用转发技术完全消除停顿的是（ ）。

A. DADD R1, R2, R3	B. LD R1, 0(R2)
DSUB R4, R1, R5	DSUB R4, R1, R5
C. DADD R1, R2, R3	D. LD R4, 0(R1)
LD R4, 0(R1)	DADD R1, R2, R3

3、下列语句间存在的指令间相关性有____、____，在基本流水线（MIPS）中会产生竞争的有____、____。

DADDIU R1,R3,# -8
BNE R1,R2,LOOP
SUB R4,R4,#8

A、RAW, WAR; RAW, WAR	B、真相关, 反相关; RAW, 转移
C、真相关, 控制相关; RAW, 转移	D、RAW, WAW; RAW, WAW

4、下列哪个选项正确包含了记分牌全部限制因素：

- ①指令间可用并行数少
- ②积分卡的项数有限
- ③功能单元数目和类型
- ④存在反相关和输出相关
- ⑤未能利用转发
- ⑥无法解决 RAW

A. ①②③④⑥ B. ②③⑤⑥ C. ①②③④⑤ D. ①③④⑤⑥

5、Tomasulo's 硬件调度算法有两个主要优点，一是____竞争检测逻辑，二是消除了____竞争的阻塞。

A、集中, WAW 和 WAR	B、分散, WAW 和 WAR
C、集中, WAW 和 RAW	D、分散, WAW 和 RAR

6、通过硬件动态预测转移指令的行为可以减少转移代价，2 位转移预测缓冲器是硬件动态预测技术之一。假设 2 位转移预测缓冲器的初始值为 0，某循环体共循环 10 次，若该循环

体只在前 9 次循环实际转移, 最后 1 次实际不转移, 则 2 位预测缓冲器的命中率为 ()。若该循环体在第一次循环实际转移, 且实际转移行为间隔变换一次, 则 2 位预测缓冲器的命中率为 ()。

- A. 90% , 50%
- B. 70% , 20%
- C. 70% , 50%
- D. 90% , 20%

7、多发射处理器的目标是允许在一个时钟周期内发射多条指令。两种基本的多发射技术有（ ），（ ）。其中（ ）每时钟周期发射固定数目的指令，（ ）主要采用硬件检测竞争。

- A、多处理器，超标量；超标量，多处理器
B、VLIW，超标量；VLIW，超标量
C、超标量，VLIW；超标量， VLIW
D、超标量，VLIW；VLIW，超标量

8、以下说法正确的是:

- ①集中式共享存储器系统结构所有的处理器访问存储器的时间不一致
- ②在分布式共享存储（DSM）系统中，任何一个处理器都能够通过引用地址的方式访问任意节点上的存储器
- ③在消息传递多处理器系统中，不同处理器中相同的物理地址分别指向两个不同存储器中的不同位置
- ④NUMA（非均匀存储器访问）的访问时间取决于数据字在存储器中的位置

- A. ①② B. ①②③④ C. ②③④ D. ①③④

9、出现多处理机 cache 不一致的原因有：共享可写的数据、进程迁移和 I/O 传输。有以下两个不同处理器的程序 T1 和 T2 共享同一个可写的数据，T1 与 T2 分别使用缓存 cache-1 和 cache-2，且采用写回式写回缓存。内存与缓存的初始值如下表所示。这两个进程按下表顺序执行之后，内存中保存的数据为（ ）。

Program T1	Program T2	Cache-1	Cache-2	内存初始值	程序执行顺序
ST X,1	LD Y,R1	X=	X=	X=0	T1 完成程序执行
ST Y,10	ST Y',R1	Y=	X'=	Y=5	Cache-1 写回 X
	LD X,R2		Y=	X'=	T2 执行结束
	ST X',R2		Y'=	Y'=	Cache-2 写回 X'与 Y'
					Cache-1 写回 Y

- A. $X=1, Y=10, X'=1, Y'=5$ B. $X=0, Y=10, X'=1, Y'=5$
C. $X=1, Y=5, X'=1, Y'=10$ D. $X=0, Y=5, X'=1, Y'=5$

10、在基于目录的 cache 一致性协议中，已知数据块可能处于以下 3 种状态：共享（shared）、未缓存（uncached）、独占（exclusive）。则在数据块处于共享状态时可能的目录请求有：

- ①读缺失②写缺失③数据写回

- A. ①②③ B. ①② C. ②③ D. ①③

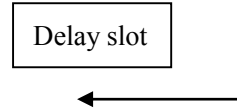
二、简答题（共5题，每题4分，共20分）

1、简述流水线的三类冒险，并指出数据冒险有哪几类，以及解决数据冒险的方法。

2、简述分支延时槽的原理，并根据分支延时槽的原理优化下述代码。

ADD R1, R2, R3

If R2 = 0 then



3、简述硬件投机机制的原理和 ROB 的作用，并指出 ROB 与普通的 tomasulo 算法中的保留站（reservation stations）的功能区别。

4、简述 tomosulo 算法和记分牌算法的异同。

5、试解释并行性的含义。

三、计算题（共 2 题，每题 10 分，共 20 分）

1、（1）概述 Amadahl 定律

（2）假定指令中的 FP（浮点）运算频率=25%，FP 平均 CPI=5.0，平均非 FP 的 CPI=2.33，FPSQR（浮点开方）运算频率 5%， FPSQR 的平均 CPI=20，假定有两种方法提高性能，分别是：

- a. 将 FP 中 FPSQR 的 CPI 减少到 2
- b. 将 FP 的 CPI 减少到 2.5，

试计算这两种方案的 CPI，并计算出较好的方法的加速比，结果保留两位有效数字。

（3）现有 100 个处理器，为了达到 50 倍的加速比，试计算所需要的并行度，结果保留四位有效数字。

2、设指令间的相关性参数如下表，假定采用一个标准 5 级整数流水线，这些功能单元被完全流水化或复制。试分析计算下列问题：

- （1）计算该循环在未进行任何调度时迭代一次需多少时钟。
- （2）采用软件流水方式编译优化下列循环，使其循环内的竞争最少。
- （3）计算优化后该循环迭代一次需多少时钟。

```
LOOP:  L.D      F0, 0(R1)
        ADD.D   F4, F0, F2
        S.D     0(R1), F4
        DADDUI  R1, R1, #-8
        BNEZ    R1, LOOP
```

前操作指令	后继相关指令	延迟时钟
FP ALU 操作	FP ALU 操作	3
FP ALU 操作	Store (双字)	2
Load (双字)	FP ALU 操作	1
Load (双字)	Store (双字)	0

四、分析题（共 3 题，共 40 分。第 1 题 15 分，第 2 题 10 分，第 3 题 15 分）

1、(m,n) 相关分支预测器利用最近执行的 m 个分支的行为从 2^m 个预测器中作出选择，这些预测器都是 n 位预测器。现有一个 (2,2) 相关分支预测器共 8K 位。

(1) 在该相关分支预测器中有多少项？

(2) 画出这个相关预测器的硬件框图。

(3) 假设全局转移缓存和每个转移预测器的初始值都为 0，a 初始值为 1，利用上述 (2,2) 相关分支预测器，下列程序连续执行 5 次时命中率是多少？

```
if(a==0)
    a=1;
if(a==1)
    a=2;
```

```
Reg[R1]=a;
    BNEZ    R1, L1;
    DADD    R1, R0, #1;
L1: DADD    R3, R1, #-1;
    BNEZ    R3, L2;
    DADD    R1, R0, #2
L2: ...
```


2013-2014 学年秋冬学期高级计算机体系结构试卷分析

一、选择题

1、B。SIMD 计算机属于并行结构计算机，一条指令可以同时多个数据进行运算。SIMD 计算机由单一的指令部件控制，按照同一指令流的要求，为多个处理单元分配各不相同的数据并处理。SIMD 计算机以阵列处理机和向量处理机为代表。MIMD 计算机属于并行结构计算机，多个处理单元根据不同的控制流程执行不同的操作，处理不同的数据。MIMD 计算机是能够实现指令、数据作业、任务等各级全面并行计算的多机处理系统。

2、B。本题考查对 forwarding 技术的理解。Forwarding 技术是解决流水线中的部分数据冒险问题的重要硬件技术，但由于 load 指令只有在 MEM 周期结束之后才能得到数据，所以即使利用转发也需要一个停顿之后才能得到数据。本题中 B 选项中 LD 与 DSUB 存在数据相关，且不能利用转发技术完全消除停顿。A、C 选项都能利用转发技术消除停顿，而 D 选项不存在数据相关。答案选 B。

3、C。

DADDIU R1,R3,#-8①

BNE R1,R2,LOOP②

SUB R4,R4,#8③

①②之间存在真相关，产生 RAW 竞争。②存在控制相关，产生转移竞争。

4、C。

5、B。相较简单的方案而言，Tomasulo 方案有两个优势：1) 冒险检测逻辑的分散；2) 消除可能产生的 WAW 和 WAR 冒险的停顿。第一个优势源于分布式保留站和 CDB 的使用，第二个优势（消除 WAR 和 WAW）是利用保留站来重命名寄存器，并在操作数可用时，立即将其存储在保留站中。

6、C。本题考查对 2 位转移预测缓冲器的理解。2 位预测器只有在连续预测错误两次之后才会修改预测方向。当循环体在前 9 次实际转移，最后一次实际不转移时：预测器开始预测不转移，前两次都预测失败，连续失败两次后，2 位预测器预测转移，所以 3~9 次预测成功，第 10 次实际不转移，预测错误，故命中率为 70%。当循环体第一次实际转移且实际转移行为间隔变换一次时：预测器每两次命中一次，故命中率为 50%。答案选 C。

7、B。多发射技术有 superscalar 超标量方法和 VLIW 超长指令字。超标量主要采用硬件检测竞争，VLIM 采用编译构成可并行执行的指令包，每个周期始终发射固定数目的指令。

8、C。现有的 MIMD 机器根据存储器组织方式可以分为两类：集中式共享存储器系统结构和分布式存储器系统结构。集中式共享存储器结构只有单一存储器结构，对每个处理器而言

都是对等的，每个处理器访问的时间都相同，所以也被称为对称（共享存储器）多处理器系统（SMP）或均匀存储器访问（UMA）。分布式存储器多处理器系统的每个节点包含处理器、存储器、输入输出系统和互联网络的接口。

根据处理器间传递数据所用的方法，有两种不同的系统结构。分布式共享存储器系统（DSM）和消息传递多处理器系统。DSM 将物理上分离的存储器作为逻辑上共享的地址空间进行寻址，所以任何一个处理器都能够通过引用地址的方式访问任意节点上的存储器，但是其访问时间取决于数据字在存储器中的位置，所以也被称为 NUMA（非均匀存储器访问）。消息传递多处理器系统的地址空间由多个私有的地址空间组成，这些私有地址空间在逻辑上是分散的，并且不能被远程处理器寻址。

9、A。本题考察共享可写数据时引起的 cache 不一致，写回式和回写式两种缓存写入方式的理解。采用写回式缓存时缓存更新的数据不会立即反应到内存中。本题中，T1 完成程序执行时 cache1 中的 $X=1$ ， $Y=10$ 。Cache1 写回 X 后，内存中的 $X=1$ 。T2 执行结束时 cache2 中的 $X=1$ ， $X'=1$ ， $Y=5$ ， $Y'=5$ 。Cache2 写回 X' 与 Y' 后内存中 $X=1$ ， $Y=5$ ， $X'=1$ ， $Y'=5$ 。Cache1 写回 Y 后内存中 $X=1$ ， $Y=10$ ， $X'=1$ ， $Y'=5$ 。故答案为 A。

10、B。

二、简答题

1、流水线的三类冒险分别是：

（1）结构冒险：当硬件在指令重叠执行中不能支持指令所有可能的组合时发生资源冒险。

（2）数据冒险：在同时执行的指令中，一条指令依赖于前一条指令的数据而得不到时发生的冒险。

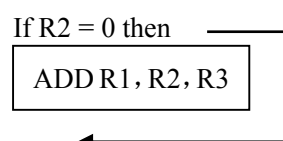
（3）控制冒险：流水线中的转移指令或其他改写 PC 的指令造成的冒险。

其中有 3 类数据冒险：RAW（写后读）：指令 j 试图在指令 i 写一个数据之前读取它，这时 j 会读到错误的值，RAW 对应于数据的真相关；WAW（写后写）：指令 j 试图在指令 i 写一个数据之前写该数据，留下的值将会是指令 i 的结果，WAW 对应于输出相关，只在特定类型的流水线中才发生；WAR（读后写）：指令 j 试图在指令 i 读一个数据之前写该数据，这时指令 i 会错误的读出新值，WAR 对应于反相关，不会发生在静态流水线之中。

解决数据冒险的方法有：双跳（double bump）；停顿（stall）；转发（forwarding）；指令重排序（instruction reorder）。

2、**分支延时槽**的原理：引入分支延迟槽的目的主要是为了提高流水线的效率。流水线中，分支指令执行时因为确定下一条指令的目标地址一般要到第 ID 级以后，在目标确定前流水线的取指级是不能工作的，即整个流水线就“浪费”（阻塞）了一个时间片，为了利用这个时间片，在体系结构的层面上规定跳转指令后面的一个时间片为分支延迟槽（branch delay slot）。位于分支延迟槽中的指令总是被执行，与分支发生与否没有关系。这样就有效利用了一个时间片，消除了流水线的“气泡”。

调整后的代码：



3、（1）硬件投机机制的原理：

基于硬件的投机技术实质上是综合了下述三种技术的一种集成技术:①应用动态转移预测技术选择投机指令②应用投机技术达到在控制相关性消除以前就执行指令③应用动态调度技术来调度程序基本块的不同组合。实际上就是动态投机与动态调度相结合的一种技术。

(2) ROB 的作用:

重构序缓存 (ROB) 相当于一个额外的虚拟存储器, 相当于 tomasulo 算法中的保留站、load buffer 和 store buffer 等的功能。重构序缓存在指令完成操作之后直到交付这段时间里保存该指令的结果, 作为其他指令操作数的源。

(3) ROB 与 RS 的区别:

在 tomasulo 算法中, 当指令完成写结果的操作后, 所有的后继指令都将从寄存器文件中读取结果。而在推测技术中, 只有在指令提交之后寄存器文件才会被更新, 即在指令执行到指令提交这段时间之内, 由 ROB 提供操作数。

4、(1) 核心思想相同之处:

两者消除 RAW 竞争的思想相同。

Tomasulo 方法采用了记分牌方法的动态调度的核心思想, 多条指令处于发射状态, 等待条件成熟, 可以不按顺序执行。

(2) 核心思想不同之处:

Tomasulo 方法通过寄存器换名过程可以消除 WAR 和 WAW 竞争。

记分牌方法能检测 WAR 和 WAW 竞争, 一旦检测到存在 WAR 和 WAW 竞争, 通过插入停顿周期来解决这一竞争。所以, 记分牌方法不能消除 WAR 和 WAW 竞争。

(3) 检测竞争和控制指令执行方式的不同:

Tomasulo 方法检测竞争和控制指令执行两方面功能是通过分布在每一功能单元的保留站来进行的, 因此 Tomasulo 方法是一种分布式方法。

记分牌方法的上述功能是通过统一的记分牌来实现的, 因此记分牌方法是一种集中式方法。

(4) 写结果的方法不同:

Tomasulo 方法直接将功能单元输出的结果送往需要该结果的所有保留站, 而不必经过寄存器这一中间环节。

记分牌方法是将结果写入寄存器, 因而可能造成等待这一结果的指令都出现停顿现象, 之后, 所有相关指令的功能单元在读 FP 寄存器时又可能出现竞争现象。

5、并行性是指计算机系统具有可以同时进行运算或操作的特性, 在同一时间完成两种或两种以上工作。它包括同时性与并发性两种含义。同时性指两个或两个以上事件在同一时刻发生。并发性指两个或两个以上事件在同一时间间隔发生。是通过对有限物理资源强制行驶多用户共享以提高效率。

三、计算题

1、

1) Amdahal 定律定义了使用某一特定功能所获得的加速比。加速比取决于下面两个因素: 原计算机计算时间中可升级部分所占的比例; 通过升级执行模式得到的改进, 也就是说在为整个程序使用这一执行模式时, 任务的运行速度会提高多少倍。

2) 由公式
$$CPI_{original} = \sum_{i=1}^n CPI_i * \left(\frac{IC_i}{Instruction\ count} \right)$$
$$= (5 \times 25\%) + (2.33 \times 75\%) = 3$$

DADDUI	R1, R1, #-8	2
ADD.D	F4, F0, F2	3
L.D	F0, 0(R1)	4
DADDUI	R1, R1, #-8	5

每执行一次启动代码需要 5 个时钟周期

#循环代码:

loop: S.D	16(R1), F4 ;存到 M[i]	1
ADD.D	F4, F0, F2;M[i-1]	2
LD	F0, 0(R1) ;取 M[i-2]	3
BNEZ	R1, loop	4
DADDUI	R1, R1, #-8	5

每执行一次循环代码需要 5 个时钟周期

#结束代码:

ADD.D	F4, F0, F2	1
S.D	16(R1), F4	2
Stall		3
S.D	8(R1), F4	4

每执行一次结束代码需要 4 个时钟周期

所以当迭代 n 次时，总共需要 $9 + 5n$ 个时钟周期。所以每迭代一次需要的时钟周期数

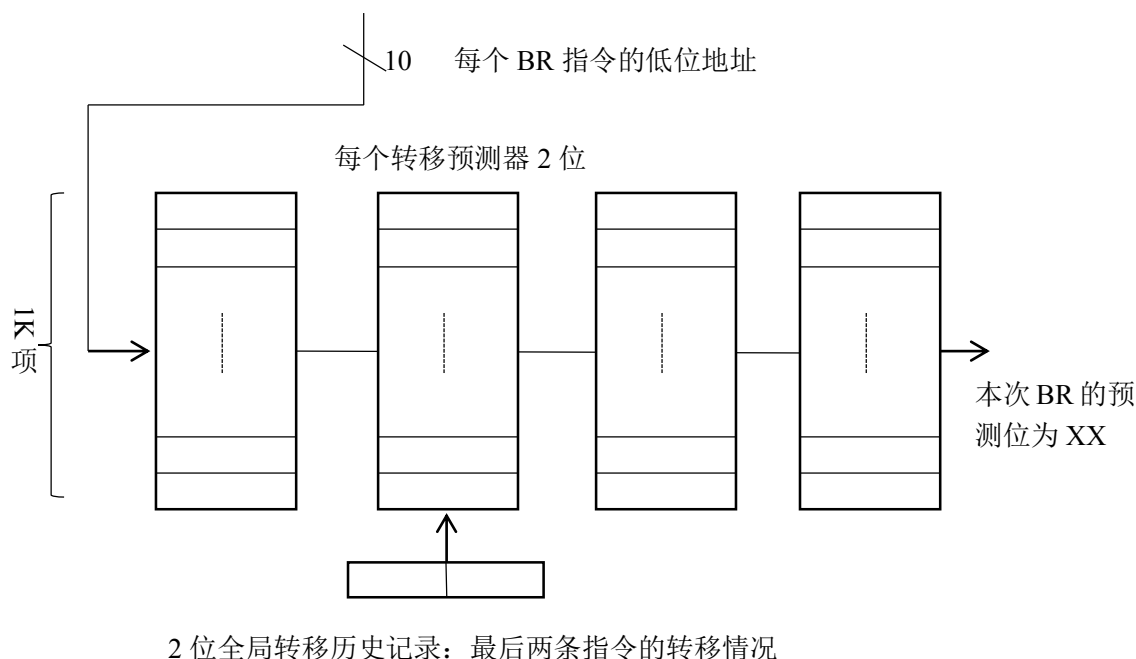
$$(9 + 5n) / n = 5 + 9/n$$

四、分析题

1、本题要求对关联预测器和 2 位预测器有较好的理解，能综合运用关联预测器与 2 位预测器的结合。

(1) $2^2 \times 2 \times$ 由分支选中的预测项数=8K，则由分支选中的预测项数=1K。

(2) 这个相关预测器的硬件框图如下：



(3) 程序执行时的数据变化如下图所示，故命中率为 $4/10=40\%$ 。

次数	a	实际是否转移	全局转移历史	Br1 对应预测器组				是否命中	a	实际是否转移	全局转移历史	Br2 对应预测器组				是否命中
				0	1	2	3					0	1	2	3	
1	1	是	00->01	0->1	0	0	0	否	1->2	否	01->10	0	0	0	0	是
2	2	是	10->11	1	0	0->1	0	否	2	是	11	0	0	0	0->1	否
3	2	是	11	1	0	1	0->1	否	2	是	11	0	0	0	1->2	否
4	2	是	11	1	0	1	1->2	否	2	是	11	0	0	0	2->3	是
5	2	是	11	1	0	1	2->3	是	2	是	11	0	0	0	3	是

2、本题考查基于 Tomasulo 动态调度算法的硬件投机技术的综合应用。要求对 Tomasulo 动态调度算法与 ROB 有较好的理解。实现推测之后的关键思想在于允许指令乱序执行，但强制它们循序提交，以防止在指令提交之前采取任何不可挽回的动作，这也是之所以添加 ROB 的原因。

当 DIV.D 准备好提交时，DIV.D 之前的指令已经提交，虽然 ADD.D 已经完成执行，但因为要按序完成指令，所以 ADD.D 还不能提交。由于没有保留站处于繁忙状态，故保留站没有示出。各硬件中的状态如下表所示：

重排序缓冲器					
项目	繁忙	指令	状态	目的地	值
1	否	L.D F6, 32(R2)	提交	F6	Mem[32+Regs[R2]]
2	否	L.D F2, 44(R3)	提交	F2	Mem[44+Regs[R3]]
3	否	MUL.D F0,F2,F4	提交	F0	#2*Regs[F4]
4	否	SUB.D F8,F2,F6	提交	F8	#2-#1

5	是	DIV.D F10, F0, F6	写结果	F10	#3/#1
6	是	ADD.D F6, F8, F2	写结果	F6	#4+#2

FP 寄存器状态										
字段	F0	F1	F2	F3	F4	F5	F6	F7	F8	F10
ROB#							6			5
繁忙	否	否	否	否	否	否	是	否	否	是

3、cache 状态转移流程表如下所示：

	P1			P2			BUS				MEM	
step	Stat	Addr	Value	Stat	Addr	Value	Stat	Proc	Addr	Value	Addr	value
P2:Write 20 to A1				Excl.	A1	20	WrMs	P2	A1			
P1:Write 40 to A2							WrMs	P1	A2			
	Excl	A2	40	Inv.			WrBr	P2	A1	20	A1	20
P2:Read A2				Shar.	A2		RdMs	P2	A2		A1	20
	Shar.	A2	40				WrBk	P1	A2	40	A1、 A2	20、40
				Shar.	A2	40	RdDa	P2	A2	40	A1、 A2	20、40
P1:Write 30 to A2	Excl.	A2	30	Inv.			WrMs	P1	A2		A1、 A2	20、40
P1:Write 50 to A1							WrMs	P1	A1			
	Excl.	A1	50				WrBk	A2	30		A1、 A2	20、30