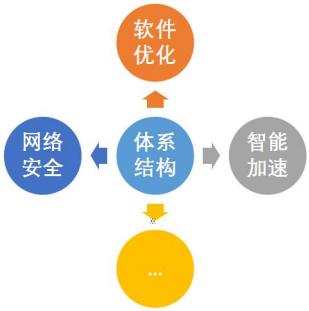
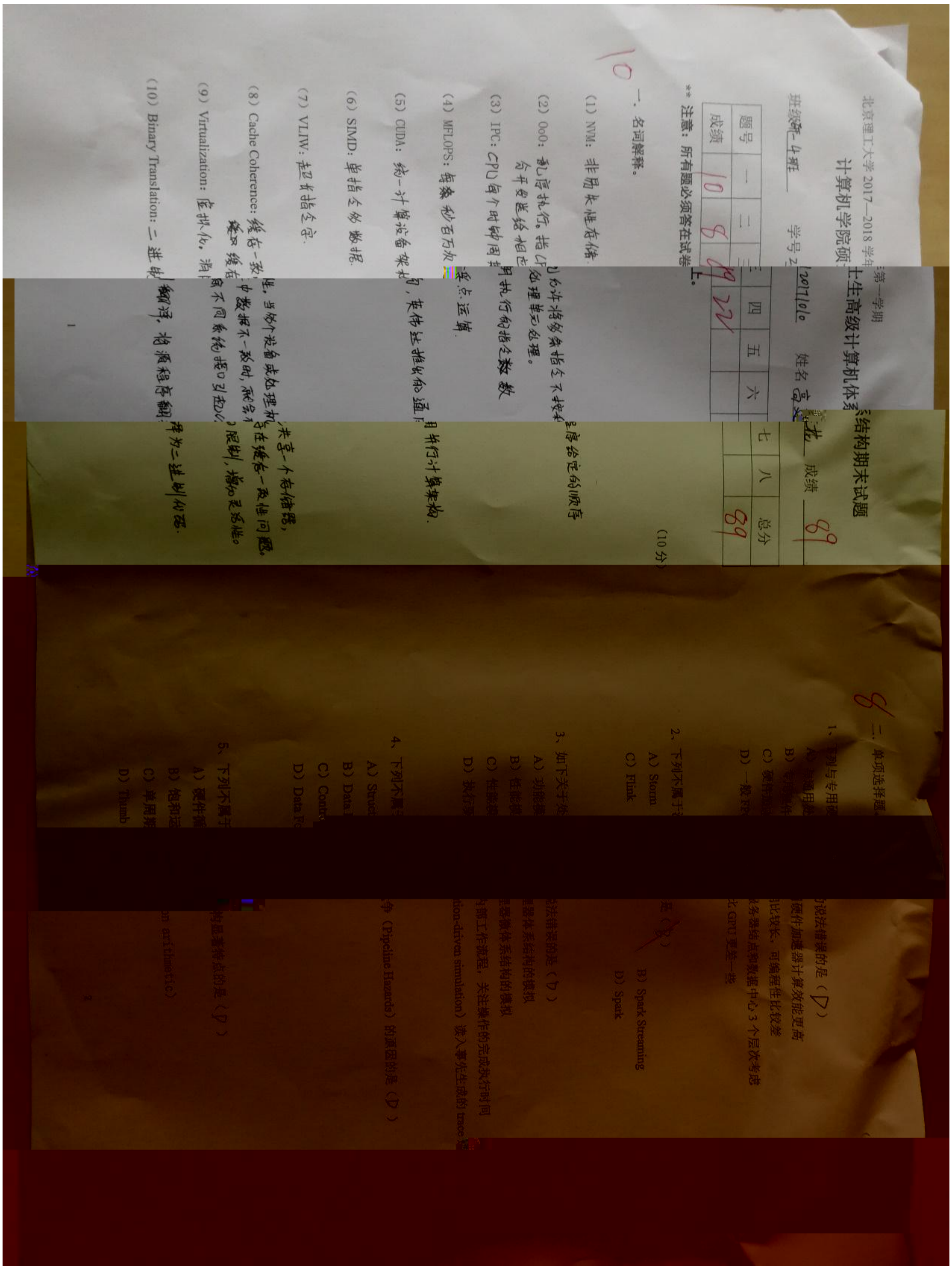
研究生精品课程简介

0700037 30


安全、软件优化、
最新技术：
非易失性存储、虚拟化、云
基础理论：
典型架构、存储模型、


软件优化
网络安全
体系结构
智能加速
...

1



北京理工大学 2017—2018 学年第一学期

计算机学院硕士生高级计算机体系结构期末试题

班级 研-4班　学号 2017010…　姓名 高…花　成绩 89

题号	一	二	三	四	五	六	七	八	总分
成绩	10	8	49	22					89

** 注意：所有题必须答在试卷上。

10 一、名词解释。（10 分）

（1）NVM：非易失性存储

（2）OoO：乱序执行，指 CPU 允许将多条指令不按程序给定的顺序分开发送给相应处理单元处理。

（3）IPC：CPU 每个时钟周期执行的指令数

（4）MFLOPS：每秒百万次浮点运算

（5）CUDA：统一计算设备架构，英伟达推出的通用并行计算架构。

（6）SIMD：单指令多数据

（7）VLIW：超长指令字

（8）Cache Coherence：缓存一致性，当多个设备或处理机共享一个存储器，缓存中数据不一致时，就会存在缓存一致性问题。

（9）Virtualization：虚拟化，消除不同系统接口引起的限制，增加灵活性。

（10）Binary Translation：二进制翻译，将源程序翻译为二进制代码。

1

8 二、单项选择题。

1、下列与专用…说法错误的是（D）

A）…通用…硬件加速器计算效能更高

B）…比较长，可编程性比较差

C）硬件…服务器结点和数据中心 3 个层次考虑

D）一般 FP…比 GPU 更差一些

2、下列不属于…是（B）

A）Storm　B）Spark Streaming

C）Flink　D）Spark

3、如下关于…说法错误的是（D）

A）功能模…器体系结构的模拟

B）性能模…器微体系结构的模拟

C）性能模…内部工作流程，关注操作的完成执行时间

D）执行…tion-driven simulation）读入事先生成的 trace

4、下列不属…争（Pipeline Hazards）的原因的是（D）

A）Struct…

B）Data…

C）Contr…

D）Data F…

5、下列不属于…的显著特点的是（D）

A）硬件…

B）饱和运…on arithmetic）

C）单周期…

D）Thumb

三、简答题。

1、假设变量 A 和 B 的初始值均为 0，当表格中的代码在三个处理器上执行时，R1 的值可能会是什么，并解释为什么。

P1	P2	P3
A=1		
	if (A==1) B = 1	
		if (B == 1) R1=A

答：① R1=1，P1执行A=1，P2执行if语句，得到B=1，P3执行if语句，得到R1=1

② R1=0，P1，P2，P3各自有cache，并将A和B的值存在各自cache中。当P1将cache中A的值改为1，并将A的更新传给P2，P2执行B=1，并将B的更新传给P3，此时，有可能P3还未收到P1中A的更新，但收到P2中B的更新，即A=0，B=1，所以R1=0。

2、如下表给出了多种存储介质的参数对比，请简要说明 PRAM（又称为 PCM）的工作原理，并从系统设计的角度来看如何在新的计算机系统中使用 PRAM？

表 1　多种存储介质参数对比[*]

Metrics	SRAM	DRAM	Flash(NOR)	Flash(NAND)	FeRAM	MRAM	PRAM	RRAM	STT-RAM
	Existing Products						Prototype		
Non-Volatile	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cell Size/F^2	50~120	6~10	10	5	15~34	16~40	6~12	6~10	6~20
Read Time/ns	1~100	30	10	50	20~80	3~20	20~50	10~50	2~20
Write/Erase Time/ns	1~100	15	1μs/10ms	1ms/0.1[illegible]	50	3~20	60~120	10~50	2~20
Endurance	1016	1016	105	105	1012	>1015	108	108	>1015
Write Power	Low	Low	Very High	Very High	Low	High	High	Low	Low
Other Power Consumption	Current Leakage	Refresh Current	None	None	None	None	None	None	None
High Voltage Required/V	No	3	6~8	16~20	2~3	3	1.5~3	1.5~3	<1.5

答：(1) PRAM的工作原理：利用硫系玻璃在不同温度下所表现出不同状态的特性，通过使用电脉冲控制PRAM介质单元的温度来改变单元中介质的电阻值，利用不同的反光特性和电阻值来表征逻辑值。

(2) 从上表中可以看出，与目前较为常用的DRAM相比，PRAM具有非易失性、低功耗、高集成度等优势，但其写和读取延时间较长，写延迟长、功耗较高，耐久性较差，寿命有限。从替代者上看，PRAM可以替代DRAM，或和DRAM协同作用，来弥补其高功率、低集成度、易失等不足。另外PRAM也可以尝试作为cache，主存等，但需要做一些技术改变来改善其缺点。

3、请简要说明并比较 OoO、SIMD 和 VLIW 三种指令级并行技术的各自特点和不同。

答：OoO (Out of Order) 乱序执行，是指CPU允许将多条指令不按程序规定的顺序分开发送给相应电路处理单元处理。

SIMD (Simple Instruction Multiple Data) 单指令多数据，通过并行使用多个相同功能的处理单元，达到数据级并行的技术。

VLIW (Very Long Instruction Word) 超长指令字。通过并行执行多条没有依赖的指令，达到指令级并行的技术。

不同：OoO乱序执行，SIMD数据级并行，VLIW指令级并行。

4、试比较三个性能度量方法："时钟频率"、"MIPS" 和 "执行时间" 各自有什么优缺点。

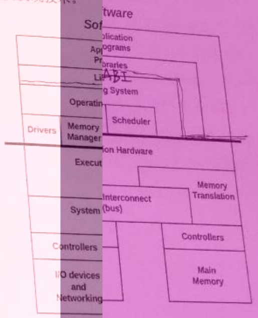
答：时钟频率：优点：easy to measure，consistent（一致性），independent

缺点：is non-linear，不可靠。

MIPS：优点：可重复，容易测量，独立性。缺点：非线性，不可靠，不满足一致性。

执行时间：优点：线性，可靠，可重复，容易测量，满足一致性，独立。缺点：不完全可重复。

5、进程级 QEMU 虚拟机被广泛使用，请在图中标注进程级虚拟机在系统层次中的位置，并说明 QEMU 虚拟机的主要实现技术。



答：进程级虚拟机在ABI (ISA + system calls) 层实现。

QEMU虚拟机的主要实现技术：

二进制翻译

寄存器映射

SPC to TPC 映射表

动态翻译

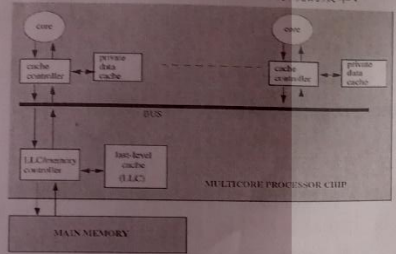
控制流优化

条件代码优化

模拟管理器。

4

6、如下所示为一多核处理器体系结构，假设该处理器采用基于总线监听的 Cache 一致性协议，请给出 Cache 控制器（图中 Cache controller）的 MSI 协议状态转换图，并说明如何对该协议进行改进以提高 Cache 系统的效率？



答：Cache控制器的MSI协议状态转换图：

改进提高Cache系统的效率.

① 增加E(Exclusive)状态，当Cache得到的数据是由主存发送，此时Cache即为E状态，增加了Cache-Cache传送，提高了Cache的效率，降低了延迟，减少了总线通信阻塞。

② 增加O(Owner)状态。~~[illegible]~~

7、请从存储顺序一致性（Memory Consistency Model）的角度说明，如下 Java 代码并行在多个处理器内核上执行时存在什么样的问题，以及如何修复该问题。

```
class Foo{
    private Helper helper = null;
    public Helper getHelper(){
        if(helper == null)
            synchronized(this){
                if(helper == null)
                    helper = new Helper();
            }
        return helper;
    }
}
```

答：对于 helper = new Helper()；使用new总共有两个步骤，一是开辟空间，此时helper已非空，二是调用构造函数初始化对象。两个步骤之间会有一定的时间间隔，如果步骤二是在后面执行，那么在步骤一和步骤二之间去读取 helper，此时的处理器会读到一个还未初始化，读到的helper即时helper非空，但未完全初始化的脏数据。

修复该问题，可在getHelper方法中定义一局部变量，通过 new Helper()来初始局部变量，然后将其赋值给 helper，从而避免了上述问题。

四．综合应用。 （24 分）

1、请列举自己研究领域的一个计算密集型的典型算法，说明使用 CPU、GPU 和硬件加速器哪种平台实现更合理一些，并解释为什么。

答：快速多极子算法：通过计算电磁在物体上的散射场，来得到物体形状，电磁波携带[illegible]信息。

12

该算法的核心都为是远/相互作用的聚集，转移，发散三个步骤，有很大计算量，以及近相互作用填充。因为其层与层之间有数据依赖，不能完全并行，而且不同层的计算量不同，负载不均衡。

对于上述算法，我认为使用 CPU+GPU 并行更合理一些，通过设计合适的任务调度算法，来将任务不同计算量的任务分别发送给CPU和GPU，大计算量的任务在众核 CPU 上，小计算量以及核心的调度在CPU上运行，充分利用其控制逻辑。

2、如下为一指令序列，请分析这些指令之间的 WAW，WAR 和 RAW 相关性，并以此为例说明如何通过寄存器重命名实现多条指令并行执行。

10

(1) add	r3, r2, r3
(2) sub	r2, r1, r3
(3) mult	r1, r3, r1
(4) add	r2, r3, r1
(5) add	r2, r1, r3

答：WAW(Write-After-Write)：(2)→(4)，(4)→(5)

~~WAR~~ WAR(Write-After-Read)：(1)→(2)，(2)→(3)，~~(3)→(4)~~ (1)→(4)，(1)→(5)

RAW(Read-After-Write)：(1)→(2)，(1)→(3)，(1)→(4)，(1)→(5)，(3)→(4)，(3)→(5)

寄存器重命名

(1) add P_4, P_2, P_3

(2) sub, P_5, P_1, P_4

(3) mult P_6, P_4, P_1

(4) add P_7, P_4, P_6

(5) add P_8, P_6, P_4

	Y_1	Y_2	Y_3
(0)	P_1	P_2	P_3
(1)	P_1	P_3	P_4
(2)	P_1	P_5	P_4
(3)	P_6	P_5	P_4
(4)	P_6	P_7	P_4
(5)	P_6	P_8	P_4

多条指令并行执行

{ (1)
(2) (3)
(4) (5)