ADVANCED COMPUTER ARCHITECTURE
FINAL EXAM
8 JAN 2016

Run LaTeX again to produce the table

Student Name: _____

Student ID: _____

QUESTIONS

1. (10 points) 假设某计算机分别以 2 和 5 MFLOPS 的速率生成 20% 和 80% 的计算结果。那么哪种改进更为有利: 提高 2 MFLOPS 速率, 或者提高 5 MFLOPS 速率? 请解释你的选择。(A machine produces 20% and 80% of its results at the rates of 2 and 5 MFLOPS, respectively. What is more advantageous: to improve the 2 MFLOPS rate, or to improve the 5 MFLOPS rate? Please explain your decision.)

Solution:

Time for 2 MFLOPS part:

$$0.2/2 = 0.1$$

Time for 5 MFLOPS part:

$$0.8/5 = 0.16$$

5 MFLOPS part is the common case, improve 5 MFLOPS part.

2. (10 points) 请分析某计算机上两种改进的效果: 把占总执行时间 25% 的浮点除法操作提高为原来的 5 倍, 或者把占总时间 75% 的全体浮点操作提高 50%。假设这两种改进相互独立, 而且成本相同。(Which

change is more effective on a certain machine: speeding up 5-fold the floating point division operation only, which takes up 25% of execution time, or speeding up 50% all floating point operations, which take up 75% of total execution time? Assume that the cost of accomplishing either change is the same, and the two changes are mutually exclusive.)

Solution:

To improve floating div by 5 times:

Speedup = (0.25 + 0.75) / (0.25/5 + 0.75) = 1/0.8 = 1.25

To improve all floating ops by 50%:

Speedup = (0.75 + 0.25)/(0.75/1.5 + 0.25) = 1/0.75 = 1.33

Improving all floating ops is more effective.

3. (10 points) 假设某计算机指令集中比较和分支指令, 各占指令数 10%, 其 CPI 分别是 1 和 2。一种改进方式是合并比较和分支指令, 从而减少指令总数。合并后的比较分支指令 CPI 为 2。其它指令 CPI 为 2。但这个改进会导致时钟周期时间上升 10%。请分析这个改进是否能提高计算机性能。(Suppose a computer has Compare and Branch instructions which count for 10% of the total instruction count separately. The CPI for Compare and Branch instructions is 1 and 2 respectively. The CPI for other instructions is 2. A possible change to the computer is to replace the Compare and Branch instructions with a combined CompBranch instruction which has a CPI of 2. The consequence of this change is that the clock cycle time will increase 10%. Will this change improve the overall performance?)

Solution:

CPU time before the change:

CPUTime = CPI*IC*T = (0.1*1+0.1*2+0.8*2)*IC*T = 1.9*IC*T

CPU time after the change:

CPUTime = (1*0.1/0.9 + 2*0.8/0.9)*0.9*IC*T*1.1 = 1.87*IC*T

The change will improve the overall performance

4. (10 points) 假设某 LOAD/STORE 类型计算机有如下特点: (A LOAD/STORE machine has the characteristics shown below):

Instruction type	Frequency	CPI
ALU ops	0.43	1
Loads	0.21	2
Stores	0.12	2
Branchs	0.24	2

我们发现 20% 的 ALU 指令会直接使用一个从内存装载的操作数而之后不再用到。因此一个可能的改进是增加一条新的 ALU 指令, 该指令可以从内存读取一个操作数, 其 CPI 为 2。这个改进会导致时钟周期时间上升 20%。请分析这个改进是否会提高计算机性能。(We also observe that 20% of the ALU operations directly use a loaded value that is not used again. Thus we hope to improve things by adding new ALU instructions that have one source operand in memory. The CPI of the new instructions is 2. The only unpleasant consequence of this change is that the Clock Cycle Time will increase 20%. After all, will CPU performance increase?)

Solution:

CPU time before the change:

CPUTime = CPI*IC*T = (0.43*1+0.57*2)*IC*T = 1.57*IC*T

CPU time after the change:

CPUTime = (1*0.43*0.8/(1-0.43*0.2) + 2*0.656/(1-0.43*0.2))

(1-0.43*0.2)*IC*T*1.2 = 1.656*IC*T*1.2 = 1.9872*IC*T

The change will NOT improve the overall performance

5. (10 points) 假设某 LOAD/STORE 类型计算机有如下特点: (A LOAD/STORE machine has the characteristics shown below):

Instruction type	Frequency	CPI
ALU ops	0.5	1
Loads	0.2	2
Stores	0.1	2
Branchs	0.2	2

编译器可以优化掉 20% 的 ALU 指令, 但不能减少 Load,Store 和 Branch 指令。假设计算机时钟频率是 1 GHz。请计算优化前和优化后代码的运行时间和 MIPS。注意 MIPS 是否正确反映了性能的改进。(An optimizing compiler for the machine discards 50% of the ALU operations, although it cannot reduce loads, stores, or branches. Assuming a 1 GHz clock, what is the MIPS rating for optimized code versus unoptimized code? Does the ranking of MIPS agree with the ranking of execution time?)

Solution:

CPU time before the optimization:

CPUTime = CPI*IC*T = (0.5*1+0.5*2)*IC*T = 1.5*IC*T

MIPS before the optimization:

MIPS = f/(CPI*1M) = 1G/(1M*(0.5*1+0.5*2)) = 1000/1.5

CPU time after the change:

CPUTime = (1*0.5*0.8/(1-0.5*0.2) + 2*0.5/(1-0.5*0.2))
(1-0.5*0.2)*IC*T = 1.4*IC*T

MIPS after the optimization:

MIPS = f/(CPI*1M) = 1G/(1M*(1*0.4/0.9+2*0.5/0.9)) =
1000/1.56

CPU time decreases, so performance improves. But MIPS drops, which implies that performance degrades.

6. (10 points) 假设某计算机有内存和高速缓存两级存储层次。当所有内存访问都在高速缓存命中时,CPI 为 2。所有的数据访问都由 Load 和 Store 指令完成, 这两种指令共占总指令数 30%。如果高速缓存缺失开销是 40 个时钟周期, 缺失率是 1%。那么无高速缓存缺失相对于有高速缓存缺失而言性能要好多少?(Assume we have a machine where the CPI is 2.0 when all memory accesses hit in the cache. The only data accesses are loads and stores, and these total 30% of the instructions. If the miss penalty is 40 clock cycles and the miss rate is 1%, how much faster would the machine be if all memory accesses were cache hits?)

Solution:

CPU time with no cache miss:

CPUTime = CPI*IC*T = 2*IC*T

CPU time considering cache miss:

CPUTime = (2+0.01*40*(1+0.3))*IC*T = 2.52*IC*T

Speedup of Ideal case over Real one:

Speedup = 2.52*IC*T/(2*IC*T) = 1.26

7. (10 points) 假设有一级数为 n 的指令流水线, 请分析其相对于串行执行的理想加速比。实际中有那些因素导致流水线不能达到其理想加

速比? 有何应对方法?(Suppose there is an instruction pipeline with n stages. Please deduce its ideal speedup over sequential execution. In reality, what prevents the pipeline reaching its ideal speedup? How to deal with it?)

Solution:
Ideal speedup of pipelining over sequential execution is n .
Hazards in pipeline prevent it reaching the ideal speedup:
(1) Structural hazard (2) Data hazard (3) Control hazard
Techniques to deal with hazards:
(1) Structural design with more hardware resources (2) Data forwarding (3) Instruction scheduling (4) Branch prediction

8. (10 points) Intel Core i7 处理器的存储层次如下图所示 (Below is the diagram showing the memory hierarchy of the Intel Core i7 processor):

假设采用 64 位地址, 请计算 L3 高速缓存用于字节选择, 组选择和标志的位数。假设各级高速缓存的缺失率均为 1%,L3 高速缓存的平均延迟是 35 个时钟周期, 内存访问延迟为 200 个时钟周期, 请计算 L1 高速缓存平均访问延迟。(Suppose the processor uses 64-bit addresses, please calculate the bit widths of the Byte-selection, Set-selection and Tag fields for the L3 cache. If the miss rates for all caches are 1%, the average access delay for L3 cache is 35 clock cycles, and the main memory access delay is 200 clock cycles, please calculate the average access delay for the L1 cache.)

Solution:
Block size in L3: 64 bytes

No. of blocks in L3: 8MB/64B = 128K
No. of sets in L3: 128K/16 = 8K
So byte-selection: 6 bits
Set-selection: 13 bits
Tag: 64-6-13 = 45 bits
Average Access Delay for L1:

$$AAD = 0.99*4 + 0.01*(0.99*11 + 0.01*(0.99*35 + 0.01*200)) = 4.072$$

9. (10 points) MESI 协议是解决多处理器高速缓存一致性问题的 IEEE 标准监听协议。MESI 各指代什么? 各自含义是什么? E 和 S 的区别是什么? (MESI protocol is the IEEE standard snoopy protocol for the multi-processor cache coherence. What does MESI stands for? What is the meaning for each of the M, E, S, I? What is the difference between E and S?)

Solution:
MESI stands for Modified, Exclusive, Shared and Invalid.
Modified means the cache line has been written by its CPU.
Exclusive means the cache line is read by its CPU only.
Shared means the cache line is read by its CPU and maybe other CPUs.
Invalid means the cache line is not valid.
Difference between E and S is that E means only this cache has the data, S means the data may be present in other caches.

10. (10 points) RAID 0, 1, 5, 01 或 10 是常用的 RAID 配置。假设有若干 1TB 容量的磁盘, 请问 RAID5 配置最少需要多少块磁盘? 要保存与最少 RAID5 配置同样数量的数据,RAID0,RAID1 和 RAID01 配置

各需要多少磁盘? RAID01 配置和 RAID1 配置有什么不同? (RAID 0, 1, 5, 01 or 10 are popular RAID levels used in computer systems. Suppose we have some hard disks all of 1TB size. To setup a RAID5 disk array, how many disks do we need at least? To store same amount of data, how many disks do we need for RAID0, RAID1 and RAID01? What is the difference between the RAID01 and RAID1?)

Solution:

RAID 5 requires at least 3 disks which contains 2TB data.

RAID 0 of 2TB data requires 2 disks.

RAID 1 and 01 of 2TB data require 4 disks each.

RAID 01 is the duplication over RAID 0 array, while RAID 1 is simply the duplication of 2TB disks.