

Tentamen i kursen EDA330

Datorsystemteknik

30/5 1998

Tentamensdatum: Lördag 30/5 1998 kl. 8.45 i sal VV

Examinator: Jonas Vasell

Institution: Datorteknik

Förfrågningar: Jonas Vasell (ankn. 1689)

Lösningar: anslås tisdag 2/6 på institutionens anslagstavla utanför laboratoriet

Resultat: anslås senast fredag 12/6 på institutionens anslagstavla utanför laboratoriet

Rättningsgranskning: tid och plats anslås tillsammans med resultaten

Betygsgränser: 3: 24-35 poäng, 4: 36-47 poäng, 5: 48-60 poäng

Tillåtna hjälpmedel: inga

Allmänt: För full poäng på en uppgift krävs både ett korrekt svar och en motivering. En bra motivering är minst lika viktig som ett korrekt svar. Redovisa noggrant alla gjorda antaganden utöver de som anges i uppgiftstexten. Skriv tydligt och använd gärna figurer. Maximal poäng på varje deluppgift anges inom parentes efter uppgiftstexten.

Välgångsönskning: Lycka till!

Uppgifter (1-5):

1. Följande MIPS-program utför en beräkning som förekommer på många ställen i en viss tillämpning:

```
    slt  $v0, $a0, $a1
    beq  $v0, $zero, L1
    add  $v0, $a0, $zero
    add  $v1, $a1, $zero
    beq  $zero, $zero, L2
L1:
    add  $v0, $a1, $zero
    add  $v1, $a0, $zero
L2:
```

- Vilken funktion utför denna beräkning? (2 p)
 - Från vilket/vilka register hämtas indata (argument), och i vilket/vilka register läggs resultatet? (2 p)
 - Om programmet utförs i en pipeline som den i bilaga 1, med stall som enda möjlighet att hantera pipelinekonflikter, hur många extra cykler orsakade av pipeline-konflikter kommer det att krävas för att exekvera programmet? Antag att \$a0 från början innehåller värdet 5, och \$a1 värdet 1. (4 p)
 - Eftersom beräkningen utförs på många ställen i tillämpningen, så vill man göra en subrutin av den. Visa minsta nödvändiga tillägg till koden ovan för att göra om den till en subrutin, och skriv en så kort kod som möjligt som laddar argumentregistren med konstantvärden och anropar subrutinen. (4 p)
2. Följande deluppgifter gäller hantering av avbrott (exceptions) i allmänhet och för MIPS-processorn i synnerhet.
- Ange fyra olika typer av avbrott eller exceptions, och ange under vilka omständigheter de genereras. (4 p)
 - Utgående från den pipeline-implementering av MIPS som visas i bilaga 1, beskriv i ord och genom ritning vilka tillägg till hårdvaran som krävs för att hantera avbrott och var dessa tillägg ska göras. (4 p)
 - Beskriv vad som händer i MIPS vid ett avbrott. Utgå från implementeringen i deluppgift b. (4 p)

3. Cacheminnen kan bland annat beskrivas av följande parametrar: S = total datalagringskapacitet i antal bytes, B = blockstorlek i antal bytes, A = associativitet (t.ex. innebär 4-way set associative att $A=4$), och W = antal bitar i en fysisk adress. Vi gör dessutom följande antaganden:
- S , B , och A är jämna tvåpotenser, och $s = \log_2 S$, $b = \log_2 B$, och $a = \log_2 A$.
 - Cacheminnet är fysiskt adresserat.
 - Alla adresser är byte-adresser.
 - Random replacement används som utbytesalgoritm.
 - Write-back används som skrivningsstrategi.

Uttryck följande som en funktion av parametrarna ovan (S , s , B , b , A , a , W):

- a. Antalet block i cacheminnet. (2 p)
- b. Antalet set i cacheminnet. (2 p)
- c. Antalet bitar i en tag. (2 p)
- d. Totala antalet bitar som måste kunna lagras i cacheminnet. (3 p)
- e. Minsta tillåtna sidstorlek i antal bytes i ett virtuellt minnessystem om cache-åtkomst ska kunna påbörjas parallellt med adressöversättning. (3 p)

4. Nedan beskrivs ett datorsystem med en MIPS-processor, cacheminne, MMU (inklusive TLB för virtuellt minne), och primärminne. Beräkna CPI för processorn utgående från de uppgifter som ges om systemet. (12 p)

Processorn är en MIPS implementerad som en 5-steps pipeline med stegen Instruction Fetch, Instruction Decode, Execute, Memory Access, och Write Back. Forwarding används för att lösa upp datakonflikter så långt som möjligt. För alla typer av hopp används strategin Assume Not Taken, och hoppberäkningar (adress och eventuellt villkor) görs i ID-steget. Processorns klockfrekvens är 100 MHz, och alla instruktioner spenderar normalt en klockcykel i varje pipeline-steg.

För de program som körs är följande statistik känd: Av alla exekverade instruktioner innebär 20% minnesläsningar, 10% minnesskrivningar, och 20% någon form av hopp. 70% av hoppen tas. 50% av minnesläsningarna följs direkt av en instruktion som använder det som läses in. Avbrott under programkörningen är så ovanliga att deras eventuella inverkan på CPI kan försummas. Effekter av uppstartning av pipeline är också försumbara eftersom ett mycket stort antal instruktioner exekveras.

Systemet använder virtuellt minne, och alla adressöversättningar sker via en Memory Management Unit (MMU). För denna uppgift antas att alla adressöversättningar kan ske direkt i MMU utan sidfel eller missar i TLB. MMU behöver alltså inte kommunicera med primärminnet. De fysiska adresserna är 28 bitar breda.

Det finns två separata cacheminnen; ett för instruktioner (ICACHE), och ett för data (DCACHE). Båda arbetar med cacheblock som är 4 ord (=16 byte) stora. För ICACHE är träffsannolikheten 99%, och för DCACHE 90%. DCACHE använder write-back som skrivningsstrategi, och vid 50% av missarna i DCACHE måste ett cacheblock skrivas tillbaka till primärminnet. Cacheminna använder fysiska adresser, och har en åtkomsttid vid träff på en processorcykel (inklusive adressöversättning). Vid en miss signalerar cacheminnet att det är klart så snart hela det saknade cacheblocket hämtats in.

MMU, ICACHE, och DCACHE (samt ett gränssnitt mot en systembuss) är kopplade till primärminnet via en processor-minnebuss. Denna buss är synkron, och har samma frekvens som processorn. Bussen är 32 bitar (=1 ord) bred, och har multiplexad adress- och dataöverföring. Ett ord kan överföras per busscykel. Varje transaktion på bussen innebär att data motsvarande ett cache-block läses från eller skrivs till primärminnet. En transaktion inleds med i genomsnitt en cykels väntan på att bussen ska bli ledig följt av en cykel för arbitrering. Så snart arbitreringen är klar tilldelas kontroll av bussen en av enheterna på bussen tills transaktionen är klar. Dataöverföringen inleds med att adressen till det aktuella blocket skickas till primärminnet. Primärminnet behöver 200 ns för åtkomst av första ordet i blocket, och 20 ns för de direkt efterföljande orden. Överföring på bussen och åtkomst i primärminnet kan ske parallellt. Förfarandet är oberoende av om det är en läsning eller skrivning, skillnaden är bara i vilken riktning data skickas.

TIPS: Identifiera först de olika typer av pipelinekonflikter som kan uppstå, och beräkna sedan inverkan av var och en av konflikttyperna på CPI. Du får poäng för varje korrekt identifierad konflikttyp, och för varje korrekt beräknad inverkan på CPI.

5. Nedan följer ett antal frågor med tre svarsalternativ (1, X, 2) vardera, varav endast ett är rätt. Ställ upp svaren som en tipsrad. Varje rätt svar ger en poäng.
- Minneskoherens innebär (1) att alla tillgängliga kopior av en del av minnet alltid är lika. (X) att det bara får finnas en kopia av varje del av minnet. (2) att minnet är skrivskyddat.
 - MIMD står för (1) Magnetic In-line Memory Device. (X) Multiple Instruction Multiple Data. (2) Multiple Interrupt Masking Device.
 - Det minsta positiva tal som kan representeras i IEEE 754 single precision är c:a (1) 2×10^{-38} . (X) 2×10^{-127} . (2) 2×10^{-308} .
 - TLB (1) är en komponent som buffrar operationer i flyttalsenheter. (X) är en komponent som lagrar en del av sidtabellen för virtuellt minne. (2) är en komponent för grafikacceleration.
 - MIPS-instruktionen lui (1) laddar ett teckenlöst heltal. (X) laddar den mest signifikanta delen av en stor konstant. (2) laddar den minst signifikanta delen av en stor konstant.
 - PCI är en vanlig standard för (1) processor-minnesbussar. (X) bakplansbussar. (2) I/O-bussar.
 - Amdahls lag handlar om (1) uppsnabbningen för ett system då en viss given del av systemet görs snabbare. (X) tillväxttakten för minneskrav hos typiska tillämpningsprogram. (2) kapacitetsökningen för DRAM-kretsar.
 - Typisk storlek för ett processorchip idag är (1) 30 mm^2 . (X) 300 mm^2 . (2) 3000 mm^2 .
 - För cache-minnen betyder direktavbildat (direct mapped) att (1) varje cache-block är associerat med ett primärminnesblock. (X) varje block kan lagras var som helst i cache-minnet. (2) varje block kan lagras på exakt ett ställe i cache-minnet.
 - Snooping är (1) en teknik för att upprätthålla minneskoherens i multiprocessorsystem. (X) ett sätt att kommunicera med I/O-enheter. (2) en metod för snabb uppdatering av sidtabeller.
 - LRU är (1) en utbytesalgoritm för cache och virtuellt minne. (X) en instruktion för subrutinanrop. (2) en skrivningsstrategi för virtuellt minne.
 - Flynnns klassificering gäller (1) nätverkstopologier. (X) processortyper. (2) typer av multiprocessorsystem.

SLUT

Bilaga 1 : MIPS pipeline

