

L1 : 40% mem access  
60ns Cache miss

missPen =  $\frac{1}{.4} \times 60\text{ns} = 150\text{ns}$

clock rate:  $\frac{1}{150\text{ns}} = 6.67\text{ GHz}$   
miss rate: 30%

CPI<sub>mem</sub> =  $4 \times 0.3 = 1.2$

CPI<sub>mem</sub> =  $0.3 \times 150\text{ns} = 45\text{ns}$   
CPI = 1.2  
CPI<sub>total</sub> = 2.2

Clock rate =  $\frac{1}{45\text{ns}} = 22.2\text{ GHz}$

CPI<sub>mem</sub> =  $0.3 \times 24\text{ns} = 7.2$   
CPI<sub>mem</sub> = 1.72  
CPI<sub>total</sub> = 1.72

$\frac{1.72}{2.2} = 0.78$

Load: IFD + RR + ALU + DA + RE: 600ops  
Store: IFD + RR + ALU + DA: 520ops  
Logic/Integer: IFD + RR + ALU + RE: 480ops  
Floating Points: IFD + RR + FPU + RE: 520ops  
Branch: IFD + RR + ALU: 400ops

number stages: 5

Operations in each stage: 1 → 5, 2 → 4, 3 → 4, 4 → 4, 5 → 3

New clock rate:  $\frac{5 \times 600\text{ops}}{1.6\text{ GHz}} = 1.6\text{ GHz}$

Specs up compared to original

600ops with NO PIPEline

Speedup:  $\frac{1.6\text{ GHz}}{0.78\text{ GHz}} = 2.05$

Number of Stages: 5

Ideal = Stages × Non-pipelined

Dataflow Diagram:  
Each ALU  
Single Load  
Store Branch

F3 Assume variables f,g,h,i,j are assigned to registers \$t0 to \$t4 (in order)  
Assume that the base address of Arrays A and B are in \$s6 and \$s7  
Registers \$t0 to \$t2 are to hold temp values. Translate to C

- add \$t0, \$s6, \$s0
- add \$t1, \$s7, \$s1
- lw \$s0, 0(\$t0)
- addi \$t2, \$t0, 4
- lw \$t0, 0(\$t2)
- add \$t0, \$t0, \$s1
- sw \$t2, 0(\$t1)

- Temp0 = A[f] // Address of A + variable f, A[f] ← address of "f"  
Temp1 = B[g] // "B" "g", B[g] ← address of "g"
- f = A[f] // f = A[0]
- Temp2 = Temp0 + 1 // Temp2 = A[f+1]
- Temp0 = & Temp2 // Temp0 =
- Temp0 = Temp0 + f
- Temp2 = \*Temp1

A[f]      B[g]  
\$6=16      f=8  
\$10=24      \$6+f  
A[0] A[1] A[2] A[3] ... A[N]

$A[0] \times B[0] = A[0]$

$$\begin{aligned}
&= \frac{32KB}{7.84ms + \frac{32KB}{\min(4 \times 100MB/s, 200MB/s)}} \\
&= \frac{32KB}{7.84ms + .16ms} \\
&= 4MB/s
\end{aligned}$$

Thus the bandwidth available to RAM is  $200 - 36 - 4 = 160$  MB/s. Since each transfer is 4 B, the transfers per second is  $40 \times 10^6$  transfers/sec or 1 every 25ns.

2. (40 pts) Use the following chart to show the state of a 4 location, 2-Way associative cache, that uses LRU. If a location has a number printed in it, the address is valid, if no number appears the contents are invalid. For simplicity the computer only has 16 locations in memory. If the cache takes 5ns to access and RAM takes 60ns, what is the effective access time given the sequence?

Time	0	1	2	3	4	5	6	7	8	9	10
Lookup Address	-	2	5	6	B	5	2	2	B	C	5
Cache location 00	A										
Cache location 01	B										
Cache location 10											
Cache location 11											

Time	0	1	2	3	4	5	6	7	8	9	10
Lookup Address	-	2	5	6	B	5	2	2	B	C	5
Cache location 00	A	A	A	6	6	6	6	6	C	C	
Cache location 01	B	B	B	B	B	B	B	B	B	B	B
Cache location 10		2	2	2	2	2	2	2	2	2	2
Cache location 11			5	5	5	5	5	5	5	5	5

MR=.4

$$\begin{aligned}
BW &= \frac{\text{Data Transferred}}{\text{Time to Transfer}} \\
&= \frac{\text{Bus Width}}{4T_{Hand} + \max(2T_{Hand}, T_{RAM})} \\
&= \frac{16B}{4(10ns) + \max(2(10ns), 40ns)} \\
&= \frac{16B}{80ns} \\
&= 200MB/s
\end{aligned}$$

$$\begin{aligned}
T_{eff} &= T_{cache} + MR(T_{RAM}) \\
&= 5ns + .4(60ns) \\
&= 29ns
\end{aligned}$$

The effective transfer rate of the pages from the disks is:

$$\begin{aligned}
Rate_{Disk} &= \frac{\text{Data Transferred}}{\text{Time to Transfer}} \\
&= \frac{\text{Data Transferred}}{\text{Total Latency} + \frac{\text{Data Transferred}}{\text{Combined Disk Transfer Rate}}}
\end{aligned}$$

3. (80 pts) A pipelined RISC computer has 8 stages, and runs at 2.5 GHz. The cache has a miss rate of 1% for data and instructions, and a miss penalty of 12 ns. The system has a dynamic branch predictor that is wrong only 10% of the time. Branch errors cost 5 cycles.

- (a) What is the ideal (no stalls) speedup over a non-pipelined machine?
  - (b) What is the impact to the CPI due to cache misses on a non-memory operation?
  - (c) What is the impact to the CPI due to cache misses on a memory operation?
  - (d) What is the impact to the CPI due to branch errors on branching instructions?
  - (e) If memory operation make up 20% of the commands in a typical program and branching make up 15% of the commands, what is the average CPI?
- (a)  $n = \frac{\text{Time Without Pipeline}}{\text{Time With Pipeline}} = \frac{I \times 8}{I \times 8} \approx 8$  for large I (number of instructions).
- (b)  $\Delta CPI = \text{Miss Rate} \times \text{Miss Penalty} \times \text{Clock Frequency} = (.01)(12ns)(2.5GHz) = .3$
- (c) Twice above or .6.
- (d)  $\Delta CPI = \text{Branch Error Rate} \times (\text{BranchPenalty}) = .1 \times 5 = .5$
- (e)  $CPI_{avg} = .2(1 + .6) + .15(1 + .3 + .5) + .65(1 + .3) = .32 + .27 + .845 = 1.435$

4. (20 pts) A 32-bit virtual memory system has a 64KB page size, and 1 GB of RAM. How large is the physical page number in bits? Assuming that the each entry in the table is word aligned, how large is the lookup table in bytes?

$$64KB = 2^{16}$$

$$1GB = 2^{30}$$

So the physical page number is  $30 - 16 = 14$  bits

The table size is  $2^{32 - 16} \times 4B = 2^{16}B = 256KB$