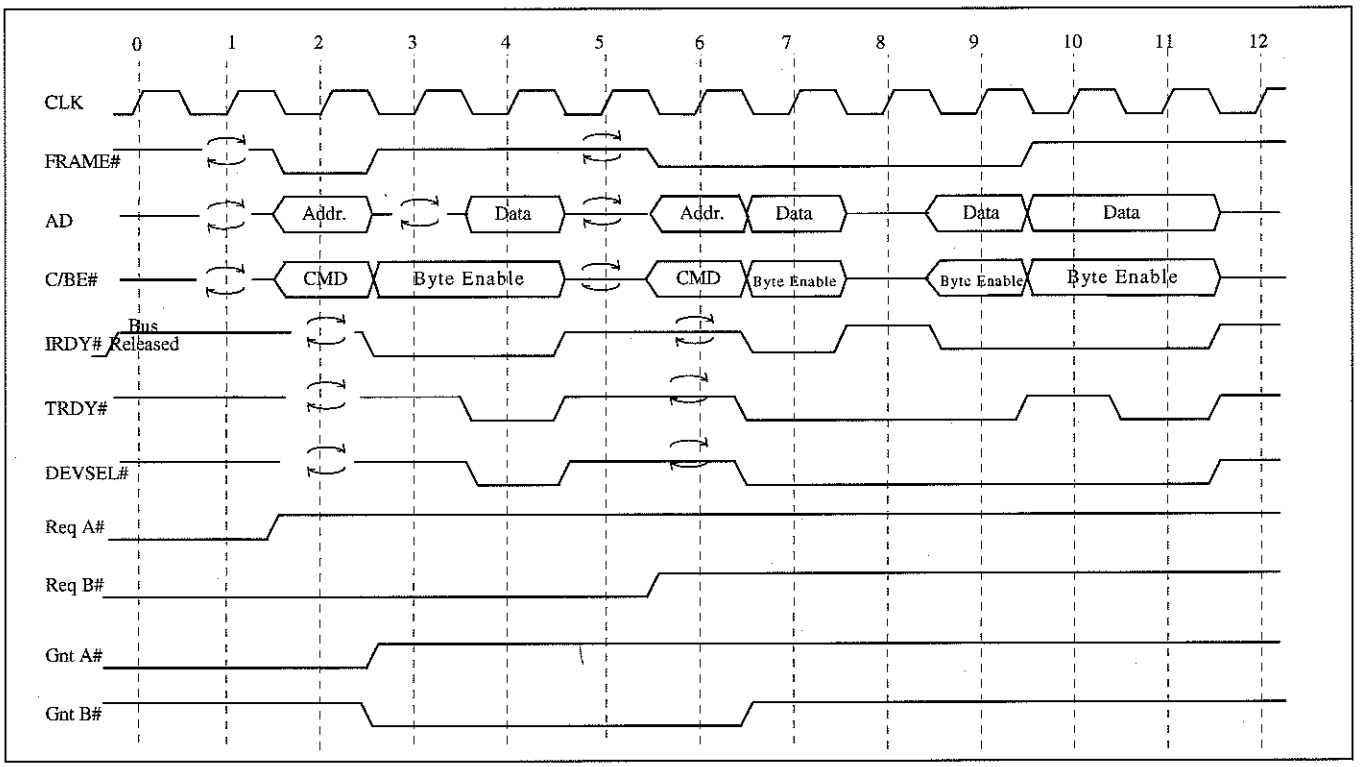
Fall 2009 Name: Mark F.

# Computer Architecture Test 2

Question 1. (21 points) Consider the following PCI timing diagram.




3 a) At what clock cycles is data read off the AD wires? 4, 7, 9, 11

2 b) Circle who controls (i.e., puts signals on) the "Req B" wire: Device A (Device B) Bus Arbitrator

2 c) Circle who controls (i.e., puts signals on) the "Gnt B" wire: Device A Device B (Bus Arbitrator)

3 d) For the second bus transaction, which device asserted the FRAME wire at clock cycle 5.5? Device B

3 e) For the second bus transaction, how does the device know when it is time to assert the FRAME wire?

DEVSEL was deasserted and the device had been grant the bus next

3 f) During the second bus transaction, why was there a delay between the first Data and the second Data?

The Initiator was not ready

g) Use the above diagram to explain, how the PCI protocol hides some of the arbitration for who gets to use the bus next with the current bus transfer.

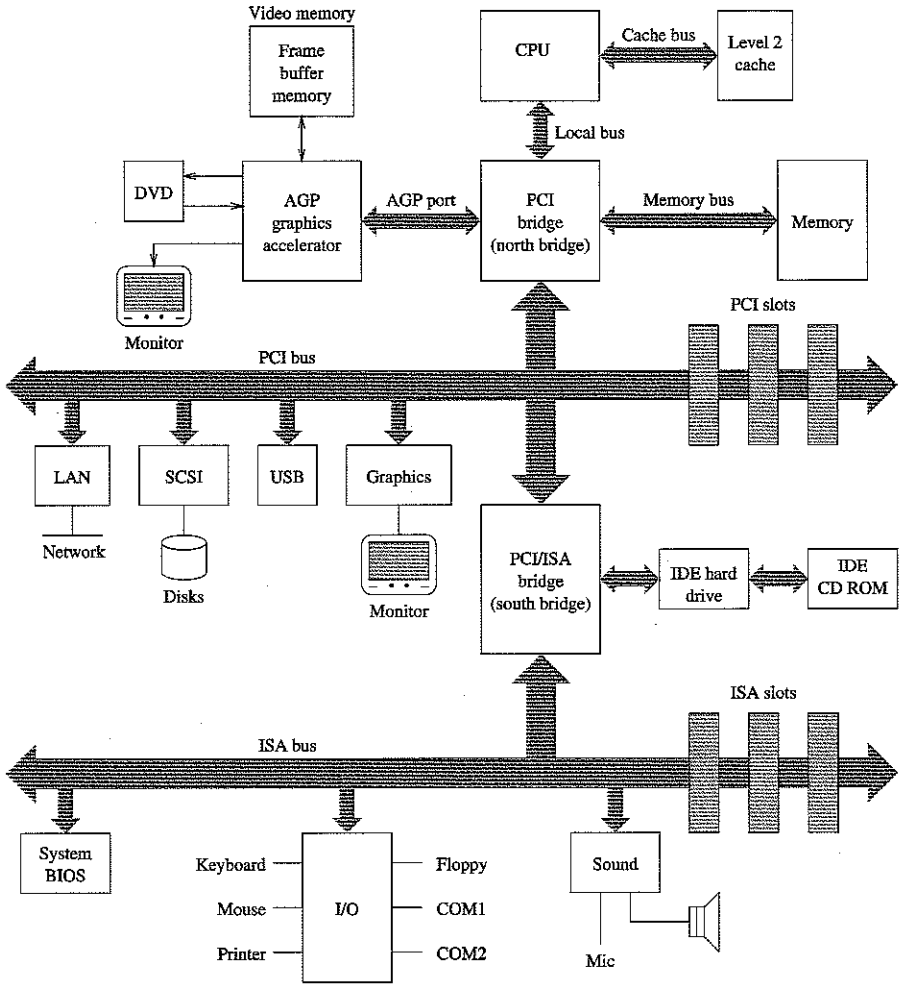
5 While device A uses the bus, the arbitrator sends the Grant signal to B at cycle 2.5. When device B sees the bus is free, i.e., DEVSEL deasserted at cycle 5.5, it starts using the bus.

Question 2. (4 points) Explain why a bus with dedicated (separate) address and data lines would NOT have a faster READ operation than a bus with time-multiplexed address and data lines (a time-multiplexed bus means that the same lines are used to send the address and data as in the PCI bus).

4 In a READ operation the address must be sent to the "target" before it can read the data return it. Since we don't need to send the address and data at the same time — Multiplexed Addr./Data lines does not slow a READ.

25

Question 3. (15 points)




As the above diagram shows, real computers utilize a hierarchy of buses.

a) What happens to the speed of the buses as you move "farther" away from the CPU?

3 The bus speed and I/O device get slower the farther away from the CPU.

b) How does a hierarchy of buses improve performance of a computer system over a single bus?

4 Each bus can be performing different I/Os at the same time.

c) What is the purpose of the "bridges" (e.g., PCI/ISA bridge) in a computer system?

4 The speed and protocol of each bus is different so the "bridges" act as protocol translators and buffers due to the different speeds.

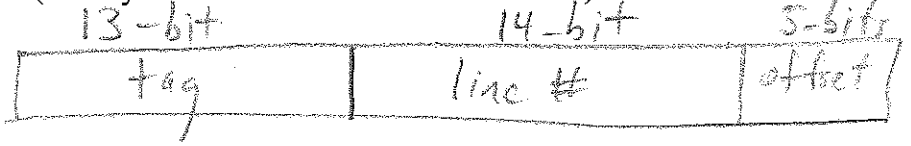
15

Question 4. (15 points) On a computer with 32-bit addresses, suppose we have a 2 GB ($2^{31}$ bytes) main memo. that is byte addressable, and a 512KB ($2^{19}$ bytes) cache with 32 ($2^5$) bytes per block.

a) How many total lines are in the cache? $\frac{2^{19}}{2^5} = 2^{14}$ lines

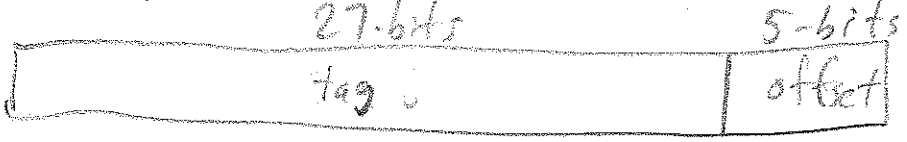
b) If the cache is direct-mapped, how many cache lines could a specific memory block be mapped to? one

c) If the cache is direct-mapped, what would be the format (number of tag bits, cache line bits, block offset bits) of the address? (Clearly indicate the number of bits in each)




d) If the cache is fully-associative, how many cache lines could a specific memory block be mapped to? any, $2^{14}$

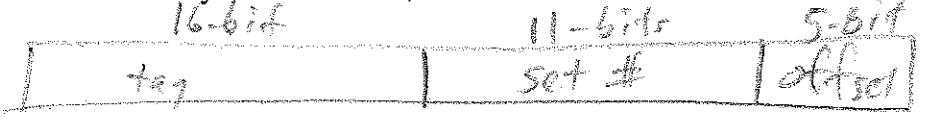
e) If the cache is fully-associative, what would be the format of the address?




f) If the cache is 8-way set associative, how many cache lines could a specific memory block be mapped to? 8

g) If the cache is 8-way set associative, how many sets would there be? $\frac{2^{14}}{2^3} = 2^{11}$ sets

h) If the cache is 8-way set associative, what would be the format of the address?




Question 5. (15 points) Consider the results of running two programs (A and B) on identical processors, except for their cache types:

| Cache Type | Program A's Execution Time | Program B's Execution Time |
|---|---|---|
| Direct-mapped cache | 20 seconds | 300 seconds |
| 2-way set-associative cache | 15 seconds | 302 seconds |
| 4-way set-associative cache | 14 seconds | 305 seconds |

a) For program A, explain why changing from a direct-mapped cache to a 2-way set-associative cache improved performance. Program A must have loop(s) that contain accesses that conflict for the same cache line within the loop body.

b) For program B, explain why changing from a direct-mapped cache to a 2-way set-associative cache did not improve performance.

Program B does not conflict for the same cache line within the loop.

Question 6. (15 points) Square-memory implementations of large memories often support fast page-modes that allow a burst of consecutive memory reads from within the same row of the square memory.

a) How does this benefit performance on a cache miss?

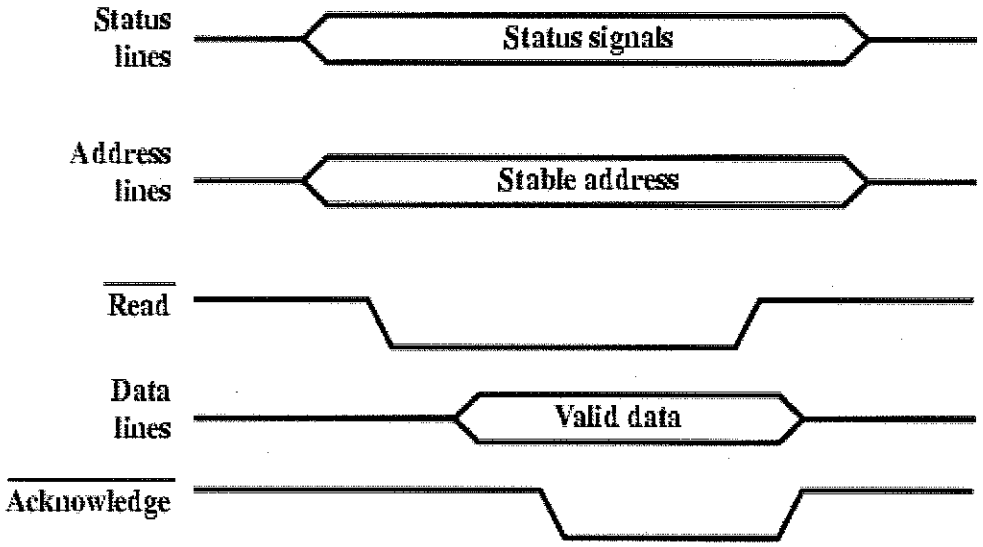
8 On a cache miss a block of memory is read into a cache line from consecutive memory. Since the fast page mode is faster than consecutive reads, the cache line can be filled faster.

b) An additional improvement allows the processor to specify a different ordering of the consecutive memory reads. For example, a processor might specify a burst of size 4 reads with offsets of 2-3-0-1 from the specified memory address. How does this benefit performance on a cache miss?

7 We could set the mode to read the word causing the miss first. As soon as this word arrives at the cache, it can be supplied to the CPU. This allows the CPU to execute while the rest of the cache line is filled.

15

Question 7. (15 points) Consider the asynchronous READ timing diagram.




a) How does the CPU know that the data is available on the Data lines? It sees the

7 Acknowledge line is asserted.

b) How is bus skew handled? The address (or Data) is put on the

8 lines before the signal that they are available via

15 Read (or Acknowledge).