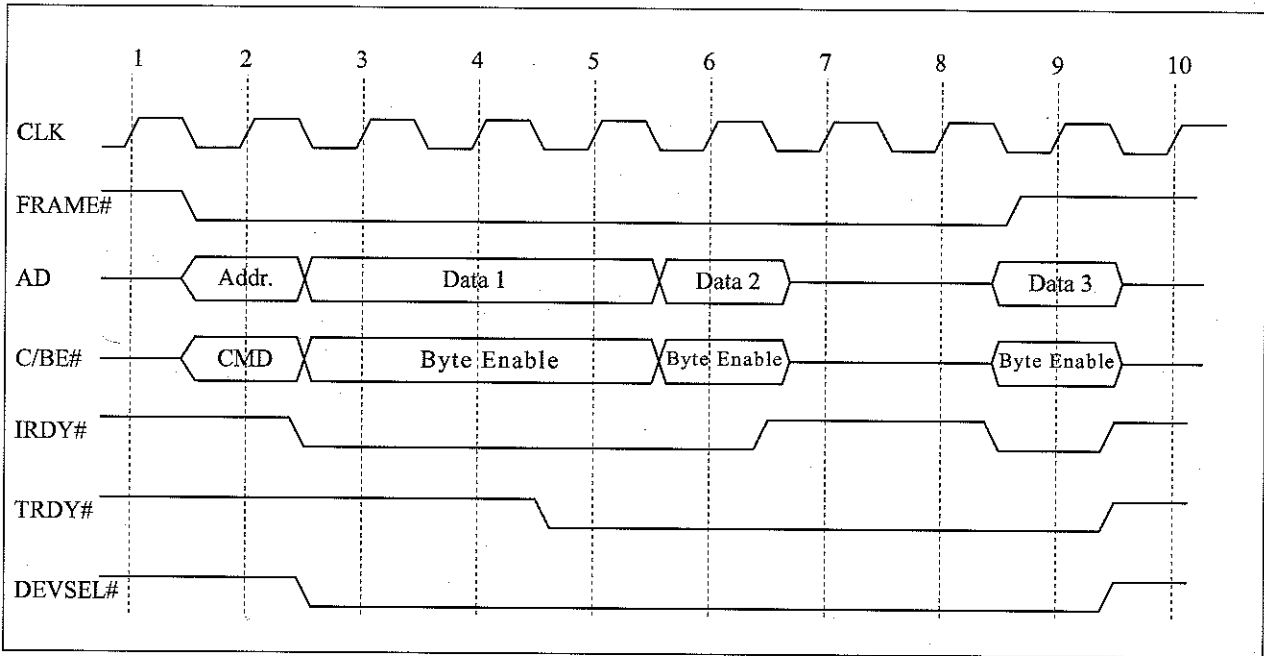


Computer Architecture Test 2

Question 1. (20 points) Consider the following PCI timing diagram.



- a) At what clock cycles is data read off the AD wires? 5, 6, 9
- b) Circle who controls (i.e., puts signals on) the "FRAME#" wire: Initiator Target Bus Arbitrator
- c) Circle who controls (i.e., puts signals on) the "IRDY" wire: Initiator Target Bus Arbitrator

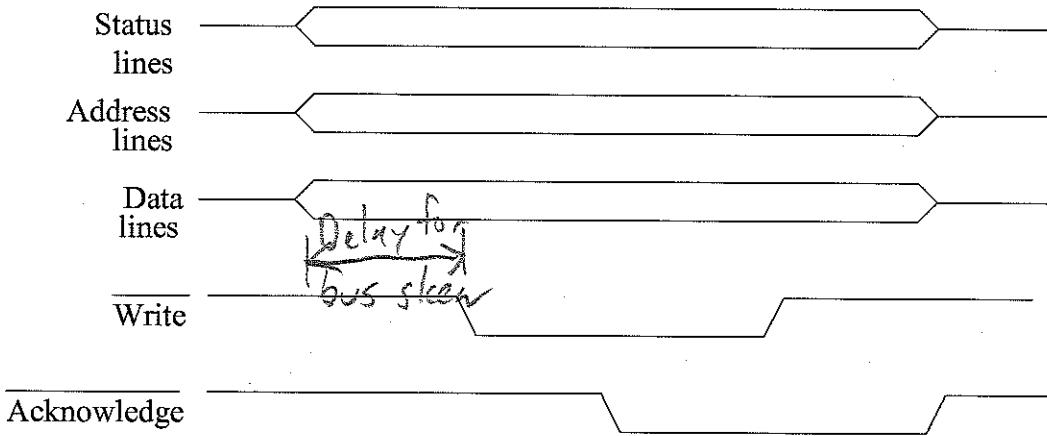
d) How does the target know there will not be a fourth piece of Data in this bus transaction?
 When Data 3 is read at clock cycle 9, the Frame is deasserted.

e) Why was there a delay between Data 2 and Data 3? The initiator was not ready for two clock cycles

Question 2. (5 points) For synchronous bus protocols (e.g., PCI bus above), how is bus skew handled?

The signals are changed in the middle of a clock cycle, and the data is read at the beginning of the next clock cycle.

Question 3. (13 points) Consider the **asynchronous WRITE** timing diagram.



a) How does the CPU know when the data has been read off of the Data lines?

The Acknowledge

Write is asserted.

b) In the above diagram, how is bus skew handled?

The data + addr. bus and allowed to stabilize before signaling that it is there by asserting the Write wire.

Question 4. (12 points) Complete the following table about the memory hierarchy.

	CPU Registers	Cache	RAM (main memory)	Harddisk
Usage - what is stored in it	Data values of current pgn.	Instr. + Data blocks	Instr. + Data for running pgn's	Pgm + Data Files for long-term storage.
Size - how much storage is there on a typical desktop	100s bytes	64KB - 2MB (L1) (L2)	512MB - 4GB	100GB - 1TB
Speed - how fast is the information accessed	1-5 ns	10 ns	50ns	12,000,000ns

Question 5. (5 points) Why is the L1 cache on a processor often split into two separate caches: a data cache and an instruction cache?

Processors are often pipelined with an instr. Fetch occurring at the same time as Data/operand fetch. Having two separate caches allow them to work at the same time.

Question 6. (5 points) Explain how a 4-way set-associative cache might improve a program's performance over a direct-mapped cache of the same size.

A program might have an instruction block and data block both mapping to the same line in a direct-mapped cache with both needed in same loop. Each iteration of the loop would cause cache misses in direct-mapped cache, but both could be in a set-associative cache.

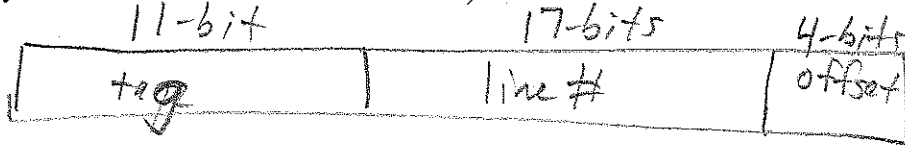
Question 7. (10 points) On a computer with 32-bit addresses, suppose we have a 1 GB (2^{30} bytes) main memory that is byte addressable, and a 2 MB (2^{21} bytes) cache with 16 (2^4) bytes per line.

a) How many total lines are in the cache?

$$\frac{2^{21} \text{ bytes}}{2^4 \text{ bytes}} = 2^{17} \text{ lines}$$

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

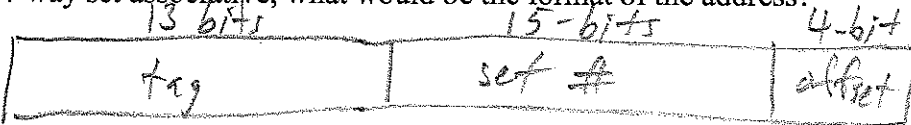
c) If the cache is direct-mapped, what would be the format (number of tag bits, cache line bits, block offset bits) of the address? (Clearly indicate the number of bits in each)



d) If the cache is 4-way set associative, how many cache lines could a specific memory block be mapped to? *4*

g) If the cache is 4-way set associative, how many sets would there be? $\frac{2^{17}}{2^2} = 2^{15}$ sets

h) If the cache is 4-way set associative, what would be the format of the address?



Question 8. (5 points) When a cache miss occurs on a memory word, a larger block of memory containing this memory word is transferred to the cache. Why does transferred a larger block than just the missed memory word typically improve cache performance?

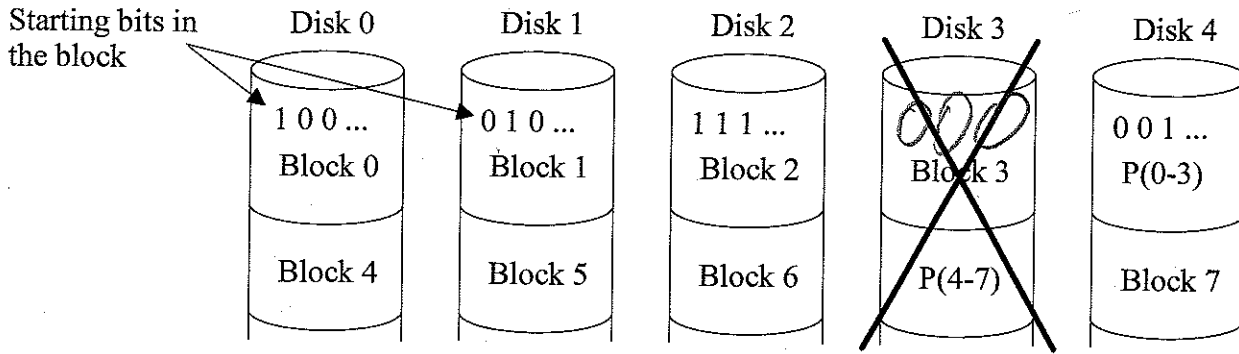
Programs typically exhibit spatial locality of reference, e.g., sequential execution of pgm instructions. The 1st instr. in the block causes a miss, but loading the larger block means later instructions with hit in the cache.

Question 9. (5 points) Large main (RAM) memories use dynamic RAM (DRAM) which stores each bit of data in a separate capacitor as opposed to static RAM (SRAM) which uses an SR-latch to store each bit of data. Why is DRAM used to implement main (RAM) memories?

DRAM requires only 1 gate per bit of storage while SRAM requires 4-5 gates per bit. Thus, DRAM memory chips can store about 4-5 times more memory.

Question 10. (9 points) Suppose we have a 5 disk RAID 5 (block-level distributed parity) array.

a) If Disk 3 crashes, reconstruct the first three bits of block 3. Assume even parity is being used.



b) Assume that Disk 3 has crashed. If you write a new block of data starting with bits 0 0 1 ... to Block 3, then can Blocks 5 and 6 be read at the same time as the write for Block 3 is occurring? (Justify your answer)

No, the P(0-3) need to be updated to reflect the new Block 3, but the only way to determine the new parity bits is to read Blocks 0, 1, and 2.

Question 11. (6 points) Suppose we have an 6 disk RAID array with each disk having a 100 MB/sec data transfer rate. Complete the following table **assuming ONE of the disks is faulty**.

RAID Level	Maximum number of concurrent, independent READs	Maximum number of concurrent, independent WRITEs	Data Transfer Rate for a single large READ
RAID 0 (no redundancy with large strips)	Not functioning	Not functioning	0 MB
RAID 1 (Mirroring with large strips)	5	3	100 MB
RAID 3 (bit-interleaved parity)	1	1	500 MB
RAID 5 (block-level distributed parity)	5	3 or 2	100 MB

Question 12. (5 points) Circle all correct answers to explain why RAID level 5 (blocked with distributed parity) is good for a database server.

- a) a database read operation typically involves one disk in the RAID array
- b) a database read operation typically involves two disks in the RAID array
- c) a database read operation typically involves all of the disks in the RAID array
- d) a database write operation typically involves one disk in the RAID array
- e) a database write operation typically involves two disks in the RAID array
- f) a database write operation typically involves all of the disks in the RAID array
- g) many database I/O operations can be performed in parallel on the RAID array
- h) a single large read operation spanning several RAID disks can boost the data transfer rate of the read
- i) if one disk in the RAID array fails, then the database server can continue to operate
- j) if two disks in the RAID array fail, then the database server can continue to operate