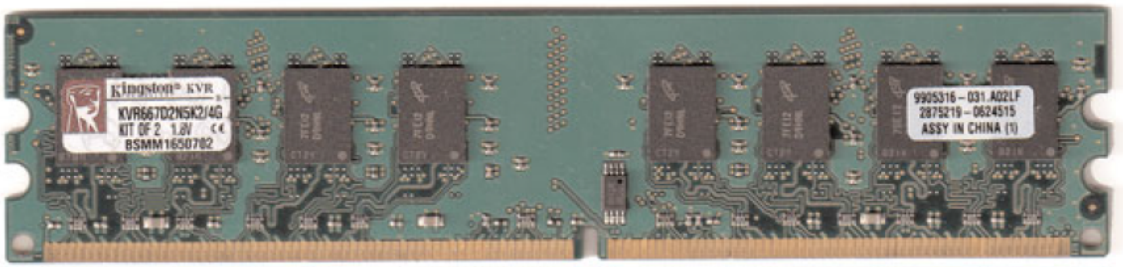
Module 10 Example Set 4

1. How does a DRAM differ from SDRAM?
SDRAM is short for synchronous DRAM. It runs in synchronization with the memory bus. Traditional DRAM used an asynchrono9us interface, which means it operated independently of the processor . Although the latency for the first unit of data is the same for both, the fact that the SDRAM signals are synchronized with the system bus signals allows SDRAM to supply additional data units within a burst transfer in fewer cycles. For example, acquiring the first or 4 data units within a burst may take 5 cycles, but each of the remaining data units within the burst would take only 1 cycle per unit for a total of 8 (5+1+1+1) cycles. Asynchronous DRAM on ther otherhand could take as much 14 cycles (5+3+3+3) to supply the same amount of data.

2. How does DDR, DDR2, DDR3 and DDR4 SDRAM differ from standard SDRAM.
Regular or standard SDRAM is often called single data rate (SDR) because it transfers a single unit of data per transfer cycle. DDR (double data rate) SDRAM provides two units of data per transfer cycle by sending data at the leading and trailing edge of the bus clock cycle. So DDR essentially operates at a frequency that is double that of the data bus frequency.
DDR2 (DDR generation 2) acheives twice the throughput of DDR memory by using differential pairs of signal wires to allow faster signaling without noise and interference problems. Its internal operating frequency is twice that of DDR and it can perform up to 1066MTps (millions of transfers per second).
Shown below is an image of a typical DDR2 memory module:



DDR3 has higher levels of performance, lower power consumption and greater reliability than DDR2. It cuts the internal clock cycle time again to half that of the DDR2 memory and employs a lower operating voltage. It provides a transfer rate of up to 2133MTps.
DDR4 takes this a step further by operating at an internal frequency twice that of DDR3 and operating at an even lower voltage level to reduce power consumption and save energy by producing less heat. It provides a transfer rate of up to 4266MTps.

3. How is the width of a memory chip and the width of the CPU-to-memory bus related to the amount of data that is obtained in response to a read operation?
The width of the cpu-to-memory bus must match the sum of the widths of the memory chips that are accessed in parallel. For example, if 4 memory chips are selected together and accessed in parallel and each chip has a width of 8, then the bus width would have to be 4*8 = 32 bits. This determines the amount of data that is transferred in response to a single memory read operation.

4. Describe the behavior of a MIPS lb (load byte) instruction?
In executing the lb instruction, the sum of the contents of the base register plus the sign-extended offset is used to identify the location in memory of the byte to be transferred. Bits 2 through 31 of the address identify the memory word to be read. The low 2 bits of the memory address (bits 0 and 1) are used as the offset to the byte within the word to be used. This byte is copied into the lower 8-bits of the result register and the upper 24 bits are filled with copies of the sign bit (the MSB) from the byte.

5. How does the behavior of a MIPS lbu (load byte unsigned) instruction differ from the behavior of a lb (load byte) instruction?
The only difference is that the upper 24 bits of the result register are filled with 0's by the lbu instruction rather than with copies of the sign bit in the loaded byte.

6. What determines the amount of data that is obtained from a memory chip in response to a single read operation?
The size of each memory cell (i.e., the width of the chip) determines how much data will be obtained from the chip in response to a single read operation.

7. What determines the amount of data the is transferred over the cpu-to-memory bus in response to a single read operation?
The width of the bus determines how much data will be transferred in response to a single read operation. The width of the bus is some multiple of the width of the memory chips. For example, if the width of each chip is 16 bits and the width of the bus is 32 bits, then two chips must be read in parallel to obtain 32 bits of data. If instead, the bus width is 16 bits, then only one chip of width 16 would be read.

8. Given the width and depth of a memory chip, how is the total storage capacity of the chip (in bytes) computed?
The storage capacity in bits is just the product of the width times the depth. The width is the number of bits per cell and the depth is the number of cells in the chip. The storage capacity in bytes is given by (width*depth)/8 since a byte contains 8 bits.