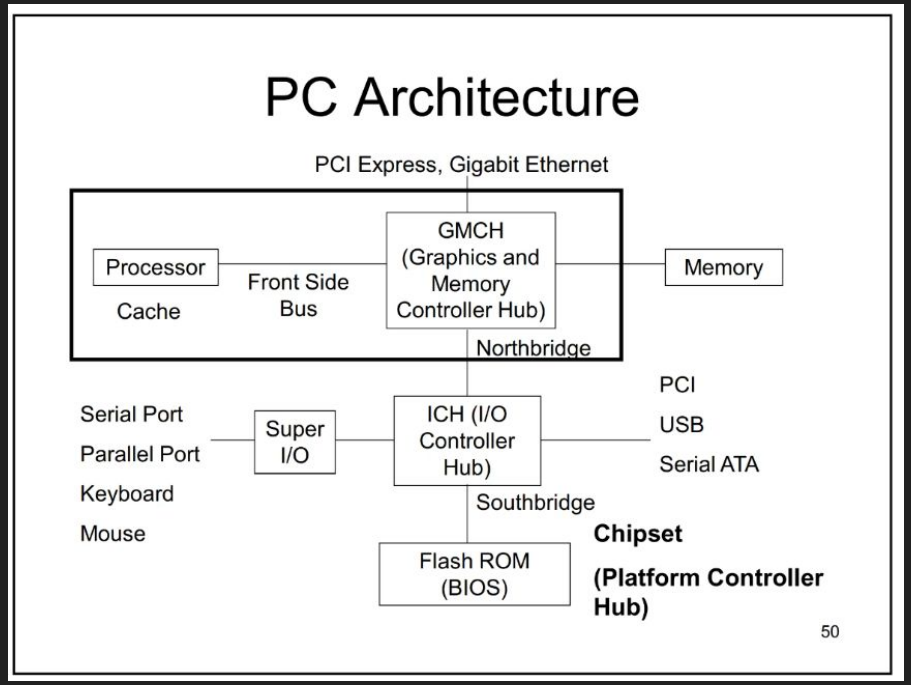
13 - Memory

CEG 4330/6330 - Microprocessor-Based Embedded Systems

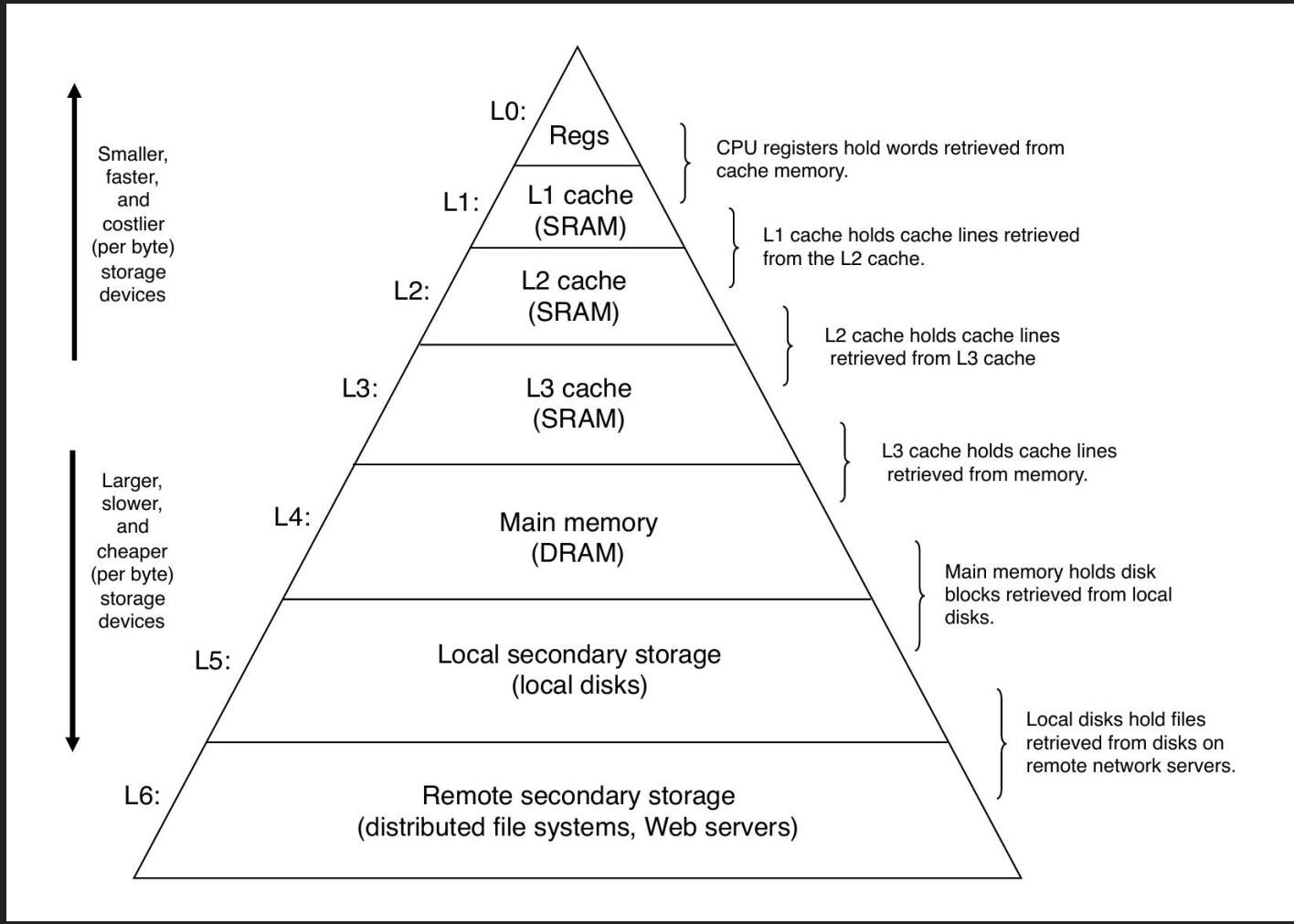
Max Gilson

PC Architecture

- A computer requires the processor, memory, and I/O to communicate with each other using busses
- Bridges are used to convert between busses
- Northbridge
  - Used to “bridge” the processor bus, PCI-E bus, and memory bus so they can all communicate
- Southbridge
  - Used to “bridge” the I/O ports so they can all communicate and eventually communicate with the processor and memory



The Great Memory Hierarchy



Locality of Reference

- Temporal Locality
  - Recently accessed data/instructions are likely to be accessed again
    - Example: In a loop, some of the variables/data will be accessed every time the loop runs
- Spatial Locality
  - Data that exists in nearby memory addresses are likely to be accessed in the future
    - Example: Instructions get accessed one instruction after the other (except for subroutines)
    - Example: Arrays may get accessed one element after the other
- In both of these cases, the cache memory becomes incredibly useful because we can access memory very quickly

Caches Exploiting Temporal and Spatial Locality

- Temporal
  - When something is accessed from memory, it’s also saved onto the cache
  - Then when the CPU needs to access it again, it’ll go to the cache instead of the slow RAM
- Spatial
  - When something is accessed from memory, the next few memory locations are saved to the cache also
  - Then if the CPU needs to access the next item in memory, it’ll go to the cache
  - The number of memory words that are accessed are called the cache line size or cache block size

SRAM vs DRAM

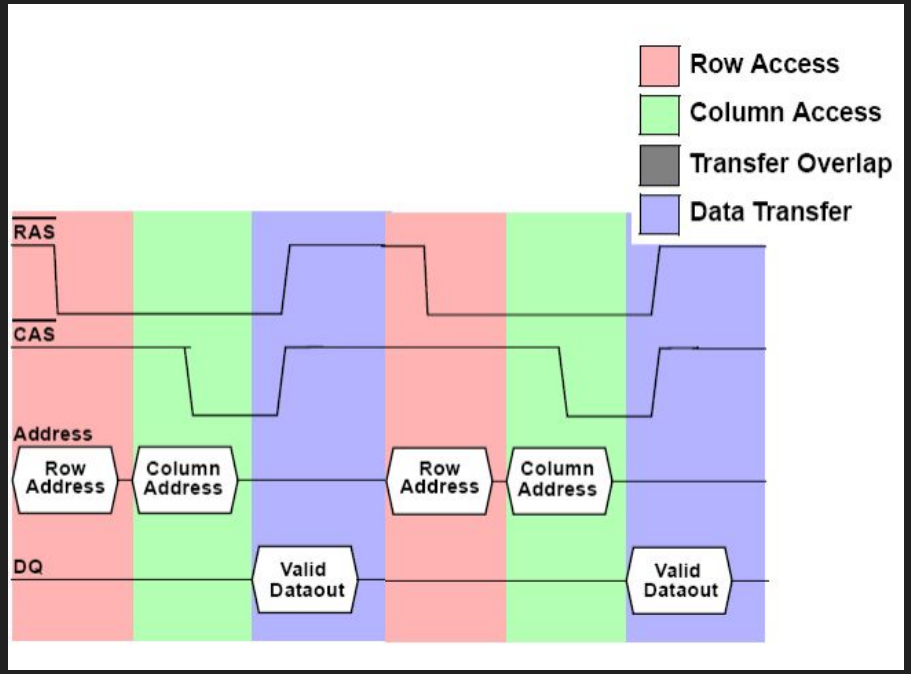
- Random Access Memory
  - Static RAM
    - Faster but more expensive
    - Used for caches
    - Volatile
  - Dynamic RAM
    - Slower but less expensive
    - Used for main memory
    - Volatile

Static RAM
➢ SRAM uses transistor to store a single bit of data
➢ SRAM does not need periodic refreshment to maintain data
➢ SRAM's structure is complex than DRAM
➢ SRAM are expensive as compared to DRAM
➢ SRAM are faster than DRAM
➢ SRAM are used in Cache memory

Dynamic RAM
➢ DRAM uses a separate capacitor to store each bit of data
➢ DRAM needs periodic refreshment to maintain the charge in the capacitors for data
➢ DRAM's structure is simplex than SRAM
➢ DRAM's are less expensive as compared to SRAM
➢ DRAM's are slower than SRAM
➢ DRAM are used in Main memory

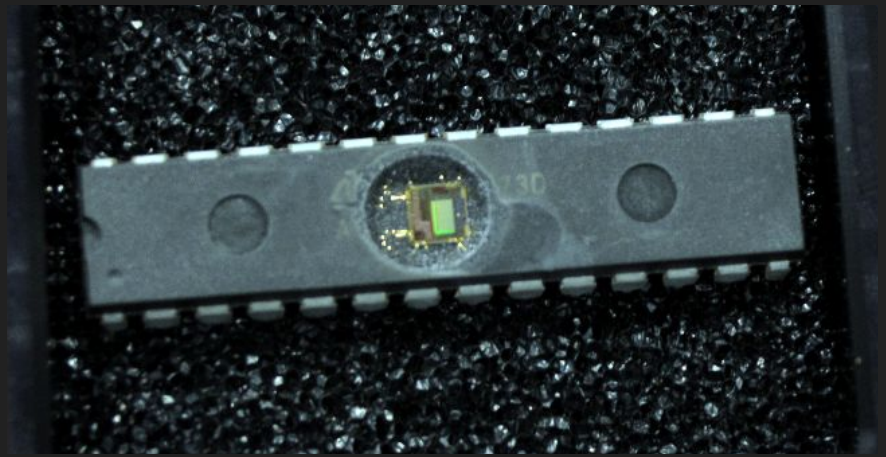
Addressing, RAS, and CAS

- Assume 1GB of memory on a chip of DRAM
  - 2^{30} possible addresses
  - Memory is in the form of a grid, with addresses and columns
  - With address multiplexing, 15 wires (smaller chip, less pins)
- RAS - Row Address Strobe
- CAS - Column Address Strobe



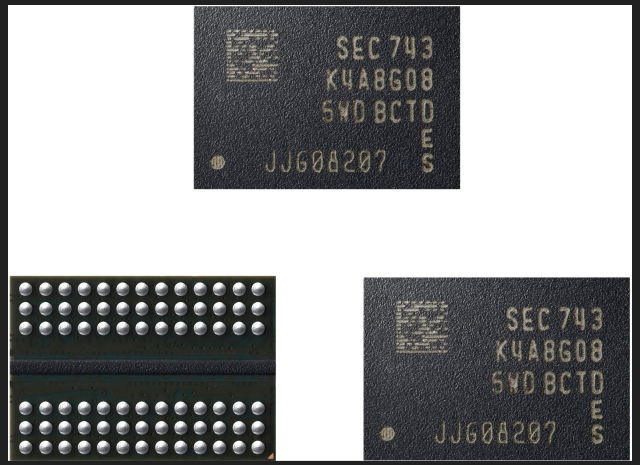
Chip Size

- Chip size is generally determined by the number/size of pins required
  - Reducing number/size of pins is a great way to shrink the size of an IC
- The IC silicon wafer inside the plastic housing is much smaller



Single DRAM IC vs DRAM Card

- It is not physically/financially possible to incorporate 16GB of DRAM into a single IC
- To achieve these memory sizes, many DRAM ICs are soldered to a circuit board that plugs in as a card
  - This is a “stick” of RAM
- Many embedded systems do not have space or do not require this amount of memory, so a single IC can be used

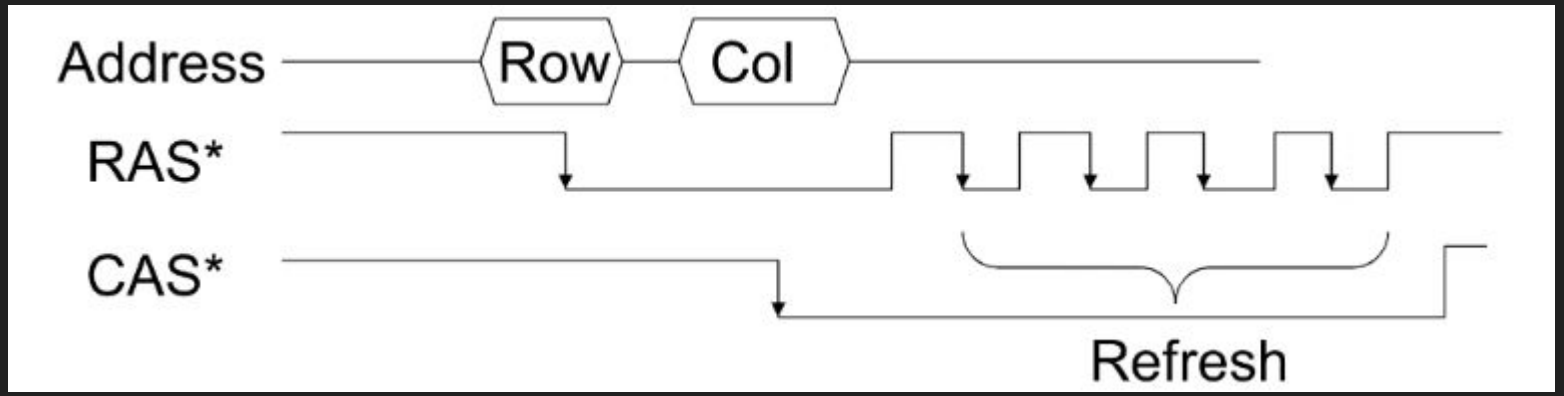


DRAM Controller

- A DRAM controller (memory controller) can be used as an interface between a processor and DRAM
  - This is not always required, many microprocessors do not require a memory controller, they interface with DRAM directly
- The CPU’s connection to the memory controller is not multiplexed (30 bits)
  - Address pins + Address strobe
- The DRAM’s connection to the memory controller is multiplexed (15 bits)
  - Address pins + Row address strobe + Column address strobe
- A memory controller is especially useful when using multiple DRAM chips
- In modern hardware, the memory controller is integrated into the CPU itself

DRAM Refreshing

- DRAM requires periodic refreshing in order to retain the memory
  - Once power is removed, the memory is gone forever
- CAS-before-RAS:



Bus Arbitration

- Bus arbitration is where multiple master devices request access and grant access to the bus
  - A simple example: every master can grant or request bus access
- This is necessary for refreshing RAM
  - A CPU might want to read from memory, but the memory controller might want to refresh
  - The memory controller can request usage of the bus periodically to refresh
  - The CPU must wait for the memory controller to grant access to the bus
- Interrupts do not work for bus arbitration because CPU requires immediate memory access when an interrupt is triggered

DDR SDRAM

- SDRAM: Synchronous Dynamic Access Memory
  - Memory that is synchronized to a clock
- DDR: Double Data Rate
  - Transfers data at both rising and falling edges of clock
- The DRAM in your laptop/desktop is most likely some form of DDR SDRAM

Dual Channel

- Multi-channeled memory allows for accessing multiple memory addresses simultaneously
- Assume:
  - Even addresses are stored in Bank 0 of memory
  - Odd addresses are stored in Bank 1 of memory
  - Using multiple banks is known as interleaved memory
- With a dual channel memory, both address 0 and address 1 can be accessed simultaneously, greatly improving memory bandwidth