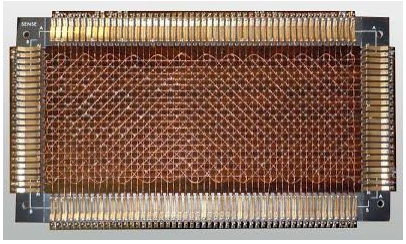
Memory Architecture

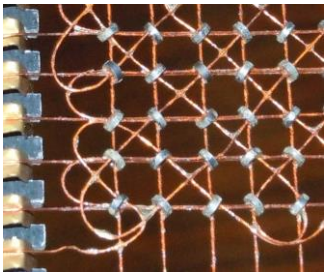
- Goal
  - Fast, Reliable, Durable, Inexpensive storage and retrieval of electronic information

Miniaturization: Smaller, faster, Cheaper, more

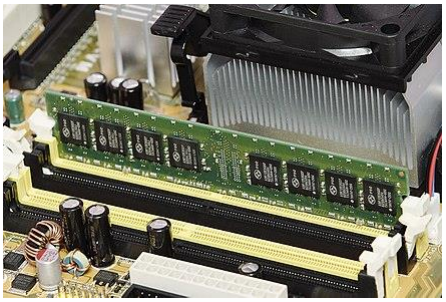
54



Core Memory Plane



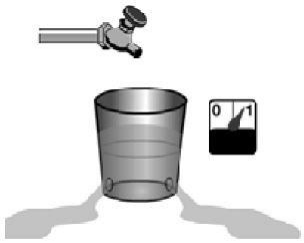
Modern RAM chips

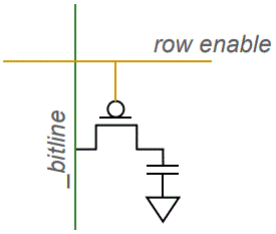


55

Memory Technology: DRAM

- DRAM (Dynamic Random Access Memory)
  - Stored value represented by a capacitor's charge state
    - 1 capacitor
    - I access transistor
  - A capacitor is like a small bucket that can store a charge
    - This charge can leak away quickly (~ms) unless the capacitor's charge is refreshed
    - This refreshing is performed either by the CPU, or by a dedicated memory controller
      - Every bit must be read and written-back 1000s of times/second

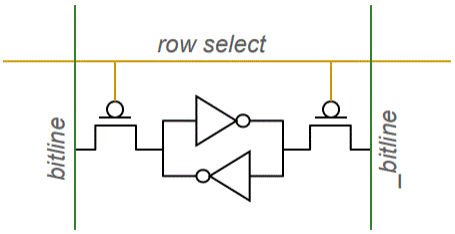




56

Memory Technology: SRAM

- SRAM (Static Random Access Memory)
  - Two cross-coupled inverters store a single bit
  - Bit maintained by a feedback path
  - 4 transistors for storage
  - 2 transistors for access
- SRAM is fast and expensive
  - Used for Cache
- DRAM is cheaper and slower
  - Used for Main Memory



57

Memory Requirements

- An ideal computer memory would have
  - Zero access time
  - Infinite capacity
  - Zero cost
  - Infinite bandwidth (for multiple parallel access)
- However,
  - Bigger -> Slower
    - It takes longer to determine locations
  - Faster -> Cost
    - SRAM vs DRAM, for example
  - Greater Bandwidth -> Cost
    - More Banks, more ports, higher frequency, faster technology

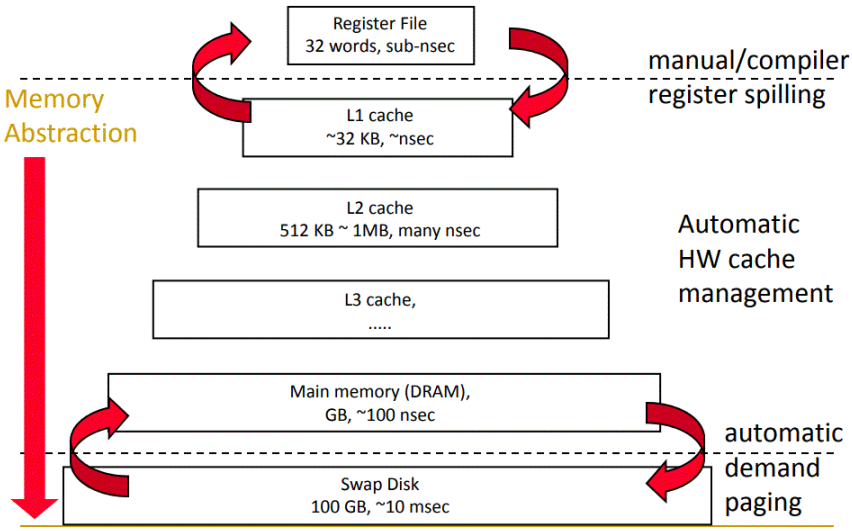
58

Why Have a Memory Hierarchy?

- We want to have both speed and size
- Both cannot be achieved with a single level
- Therefore, use multiple levels of storage (bigger and slower with distance from the processor) and ensure the most relevant data moves to the faster levels **before** it is needed.

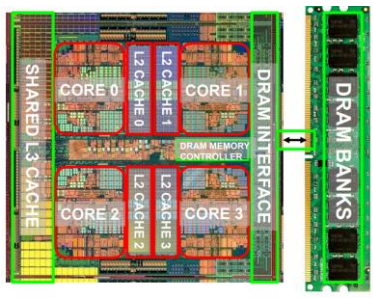
59

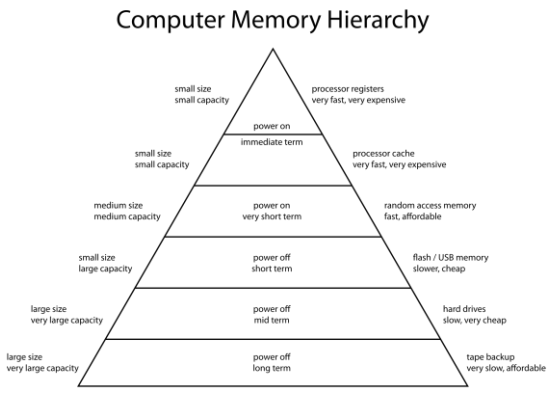
Modern Memory Hierarchy



60

Memory Hierarchy





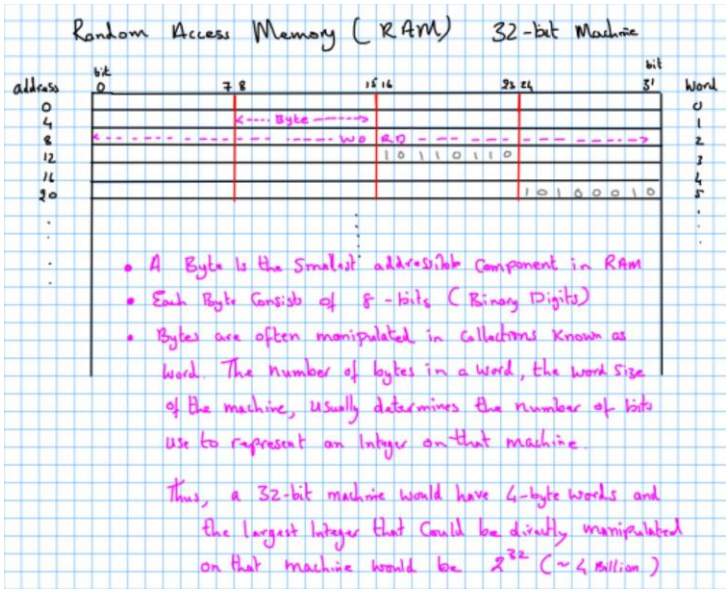
61

Random Access Memory

- Computers express memory addresses as binary numbers.
- If an address has *m* bits, the maximum number of addressable locations is 2^m
- The number of bits in the address is independent of the number of bits per location. A memory with 2^{12} location of 8 bits each and a memory with 2^{12} locations of 64 bits each need 12-bit addresses.
- Nearly all computer manufacturers have standardized on an 8-bit location, called a **byte**.
- Bytes are grouped into **words**. A computer with a 32-bit word has 4 bytes/word, whereas a computer with a 64-bit word has 8 bytes/word.
- The significance of a word is that most instructions operate on entire words, for example, adding two words together.
- Thus a 32-bit machine will have 32-bit registers and instructions for manipulating 32-bit words, whereas a 64-bit machine will have 64-bit registers and instructions for moving, adding, subtracting, and otherwise manipulating 64-bit words.

62

The Random Access Memory (RAM)



63

Error Checking

- Modern Memory is reliable
- Memory controllers check for errors only on system start up
- However, some memory chips, requiring even great reliability, have built-in error-checking and use parity checks to discover errors.

64