CSE 675.02: Introduction to Computer Architecture

Register File Design and Memory Design

Presentation E

Slides Gojko Babić

Register File

- MIPS register file includes 32 32-bit general purpose registers

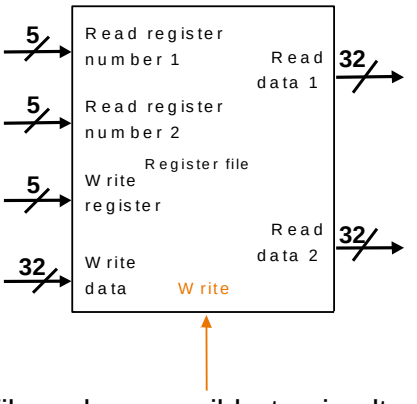


Figure B.8.7

- This register file makes possible to simultaneously read from two registers and write into one register as it is appropriate for MIPS processor.

Studying assignment: B.9

Register File Functioning

- A register file functions as follows:
  - any value provided on 5-line Read register number 1 port results in the content of the corresponding register being provided on the 32-line Read data 1 port
  - any value provided on 5-line Read register number 2 port results in the content of the corresponding register being provided on the 32-line Read data 2 port
  - on the falling edge of write line, values that appear on 32-bit Write data port are written into the register with the number specified on the 5-line Write register port.

Note that requirements for set-up time (and hold time) also apply here.

Register File Design: Read Part

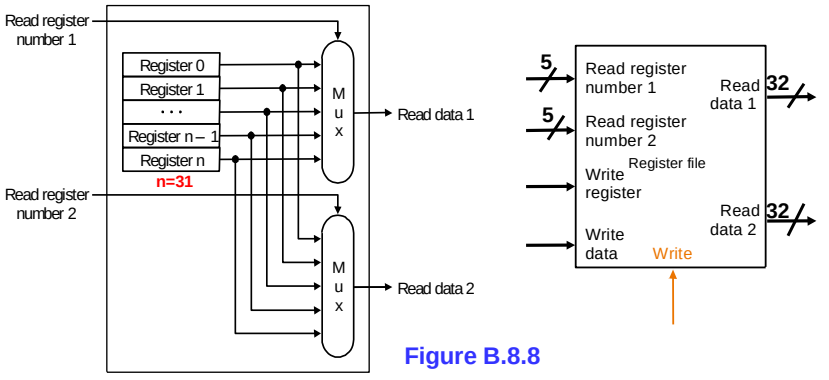
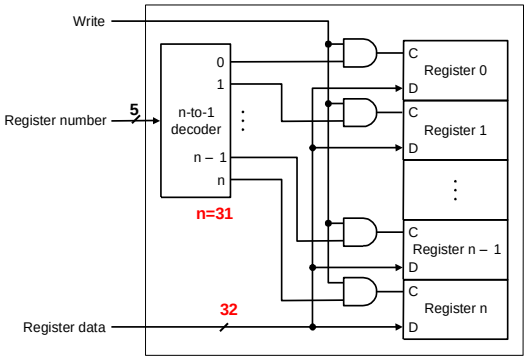


Figure B.8.8

This is the RF design at the level of registers and multiplexers.

Register File Design: Write Part



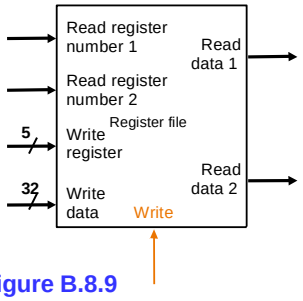


Figure B.8.9

Introduction to Memory Design

- Main memory is built in one of two technologies:
  - SRAM - Static Random Access Memory
  - DRAM - Dynamic Random Access Memory
- Both memory technologies are *volatile*
- A memory is normally built using a number of memory chips.
- Memory chips have specific configurations given as a product of two numbers, e.g.
  - **128M*1** – 128M addressable locations with 1 bit in each location, i.e. width of read/write operations is 1 bit
  - **16M*8** – 16M addressable locations with 8 bits in each location, i.e. width of read/write operations is 8 bits
- Notice that two chips above accommodate identical number of bits (128M bits).

SRAM and DRAM: 1 Bit Memory Cell

- In SRAM technology, three-state D-latch is a basic building block, i.e. basic memory cell. Internally, D-latch can have a state corresponding to 0 or 1.
- In DRAM technology, a basic memory cell is build around one capacitor coupled with one transistor. The value in the cell is stored as a charge. A charge can not be stored indefinitely and DRAM chips must be periodically refreshed. Since charge can be kept for several msec, 1-2% of time is used for refreshing.

DRAM & SRAM Characteristics

• Since 1975, the main memory has been implemented using semiconductor DRAM's.

• SRAM – technology is normally used for caches.

DRAM chip capacity had been growing at rate of about 4 times every three years, while lately growth slowed down to 2 times every two years.

Current DRAM chip capacity has reached 4GB on a single memory module with an access time in the range 40-60 nsec and a cycle time of about 80 nsec.

Access time & cycle time are two measures of memory latency:

• access time – the time between a read is requested and when the desired content arrives,

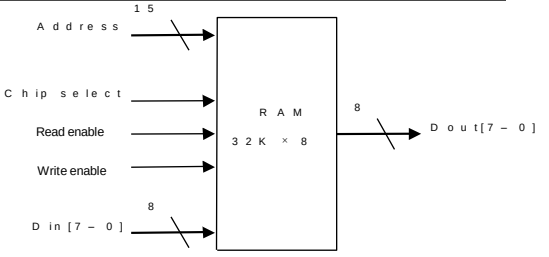
• cycle time – the minimum time between two memory requests.

DRAM & SRAM Characteristics (Cont.)

- For DRAM technology cycle time is longer than access time.
  - One of the reasons: Since a read is destructive in DRAM technology, any read has to be followed by internal write with values just read.
- For SRAM technology access time and cycle time are identical.
- In comparable technologies, SRAM cycle time is about 8 to 16 times faster than DRAM, e.g. currently 0.5-5 nsec.
- But, SRAM chip capacity (as well as density) is roughly 4 to 8 times less than that of DRAM
- Also, SRAM is more expensive, e.g. 1GB in 2004 $4,000 – $10,000 for SRAM and $100 – $200 for DRAM.
- In addition, SRAM chips have higher power consumption and power dissipation than DRAM chips.
- Thus SRAM designs are concerned with speed, while in DRAM designs the emphasis is on cost per bit and capacity.

Memory Chip Functioning

Example: 32K×8 chip
- **read and write operations are 8 bits wide**
- **there are 32K addressable locations**



- **Functioning of memory chip:**
  - **CS (Chip select) has to be set for either reading or writing**
  - **R (Read enable)=0 & W (Write enable)=0 → chip is not being accessed**
  - **R=0 and W=1 → write values at Din lines into the chip address at Address lines**
  - **R=1 and W=0 → read into Dout lines values from the chip address at Address lines**
  - **R=1 and W=1 → not allowed**

Two designs of memory chip:
- **Basic structure design**
- **Typical organization design**

Basic Structure Design of 4×2 SRAM Chip

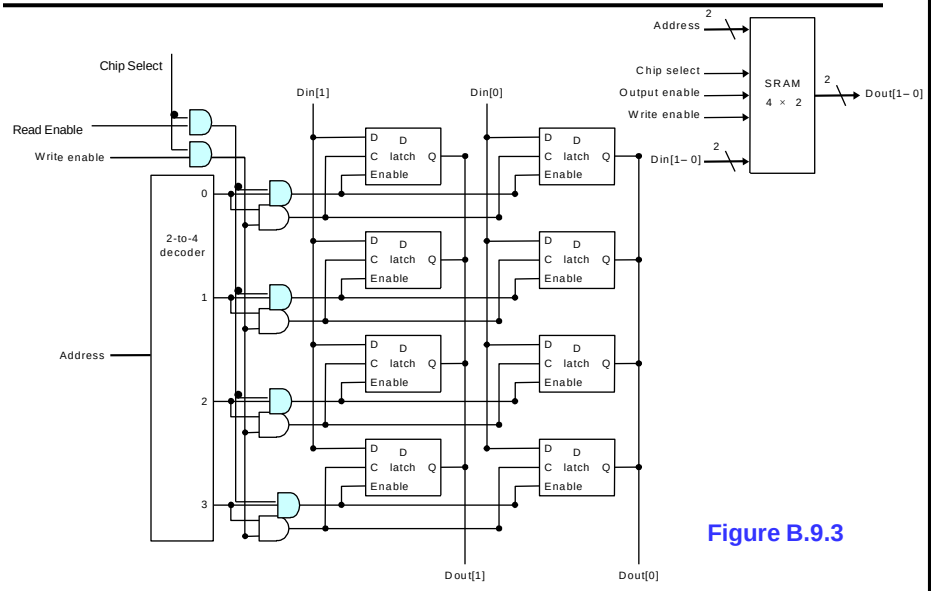
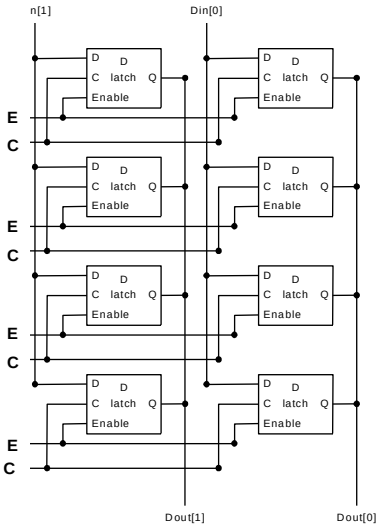


Figure B.9.3

Basic Structure Design

- The basic structure design of SRAM chip uses some ideas from the register file design e.g. the write parts in two designs are identical. The main differences are in read part design. In the memory chip with the usage of three-state D-latches a multiplexer is eliminated. E.g. for 32K*8 SRAM chip, a multiplexer with 32K inputs each input having 8 lines would be needed.
- But for the basic structure design of SRAM chip, we still need a very large decoder. E.g. for 32K*8 SRAM chip, a decoder with 15 input lines (that is not so bad) and 32K output lines (that is bad) is required.
- The typical organization design uses two level decoding that eliminates need for that very large decoder.

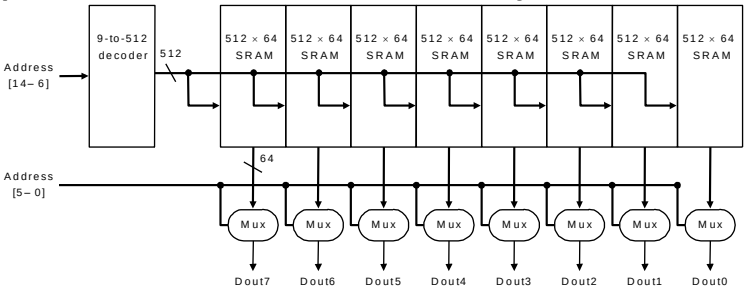
4×2 Array of D- Latches



The next example will be using 512×64 array of D-latches.

Typical Organization Design

- ***Example:*** **Design (read part only) a typical organization (i.e. two level decoding design) of 32K×8 SRAM chip that uses 512*64 arrays of D-latches.**
- **Note: Arrays used have to have a bit capacity equal to a number of addressable locations in the chip, e.g. in this example that condition is satisfied since 512*64 = 32K. A number of arrays used should be equal to the number of bits in each memory location.**



Also see Figure B.9.4 with another example of typical organization.
DRAM memory chip would have similar design.

Main Memory Specification

- A memory has identical inputs and outputs as memory chips, except that CS does not exist. But the specification of a memory should include:
  a. memory capacity (usually in bytes),
  b. memory addressability, i.e. smallest unit that has its address,
  d. width of read/write operations, i.e. a number of bits that can be read or written from/to memory.
- Operations on memory: reading from memory and writing into memory;
  - RE=0 and WE=0 → memory is not being accessed
  - RE=0 and WE=1 → writing into memory
  - RE=1 and WE=0 → reading from memory
  - RE=1 and WE=1 → not allowed