ELEN0040 - Electronique numérique

Patricia ROUSSEAUX

Année académique 2014-2015

CHAPITRE 7

Memory basics

Definitions

Memory

An electronic device capable of storing binary information and which allows the stored data to be accessed and retrieved.

- A memory is made of a collection of storage cells called memory cells together with the necessary circuits to transfer information to and from them.
- Memories are classified according to their
  - characteristics : capacity, volatility, ...
  - access type : direct (or random) vs. sequential

Essential characteristics

- Data elements :
  - bit : a single binary digit, the smallest data element (relative to one memory cell)
  - byte : a group of eight bits
  - word : a group of bytes whose size is a typical unit of access for the memory ; the memory is organized as an array of words of n bits. A word is the entity of bits that moves out and in of memory
- Capacity : the amount of information that can be stored in the memory, it is expressed as a number of words : m words $\times n$ bits, example : $64K \times 8$bits.
- Each word is identified by its address, a binary identification number of each word of the memory. For a capacity of m words, $k = \lceil \log_2 m \rceil$ address bits are required.
  The capacity if usually expressed in terms of :
  $K = 2^{10}$, $M = 2^{20}$, $G = 2^{30}$

Essential characteristics (continued)

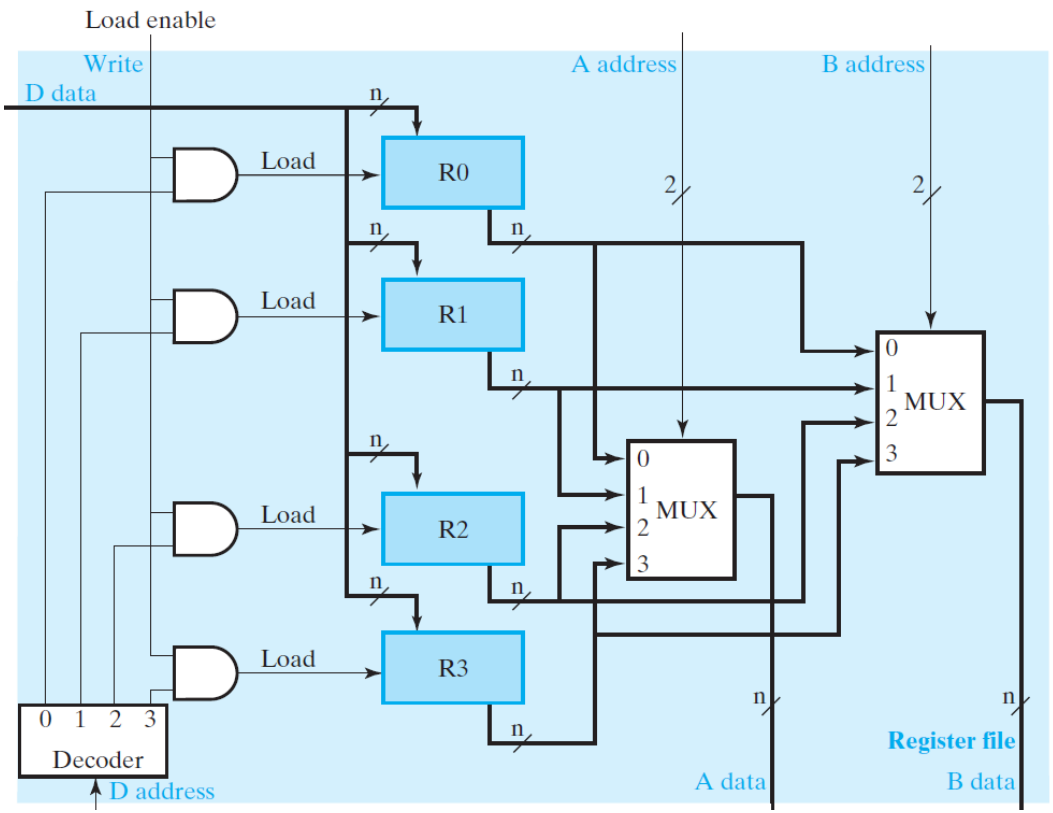
- Volatility :
  - if the information stored in memory is lost when power is turned off, the memory is said to be volatile. It needs external power to maintain the stored information.
  - Otherwise, it is said to be nonvolatile and the information is kept in memory after power is turned off.
- Two basic memory operations :
  - read the values of the data stored in one word of the memory. The data in memory is transferred to processing parts of the system
  - write data values in a word of memory. The data are transferred and kept in memory
- Access time : the time interval between the request of a read operation and the time when the read data are made available
- Cycle time : the minimum time interval between two successive memory accesses

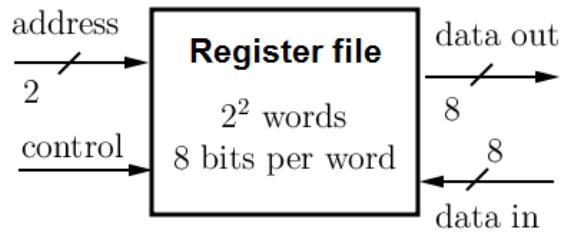
Access types

- Direct access, also called "random" access :
  - each word of information has its own address
  - each address can be directly accessed
  - the access time is independent of the information location
  - example : computer central memory
- Sequential access :
  - the access takes different times depending on the information location
  - example : hard drive where access time depends on where the desired location is relative to the current physical position
  - much slower than direct access

Register file

- We already know a type of memory : registers
- Registers are included in the datapath of digital systems
- We call Register File the part of the datapath including the registers and their associated selection circuitry
- Example, a register file with 4 8-bit registers $= 4 \times 8$ memory, 4 words and 8 bits





Characteristics :

- volatile
- sequential synchronous system
- direct access
- memory cell = one flip-flop
- Access and cycle time = one clock cycle
- very fast but biggest size (flip-flop)

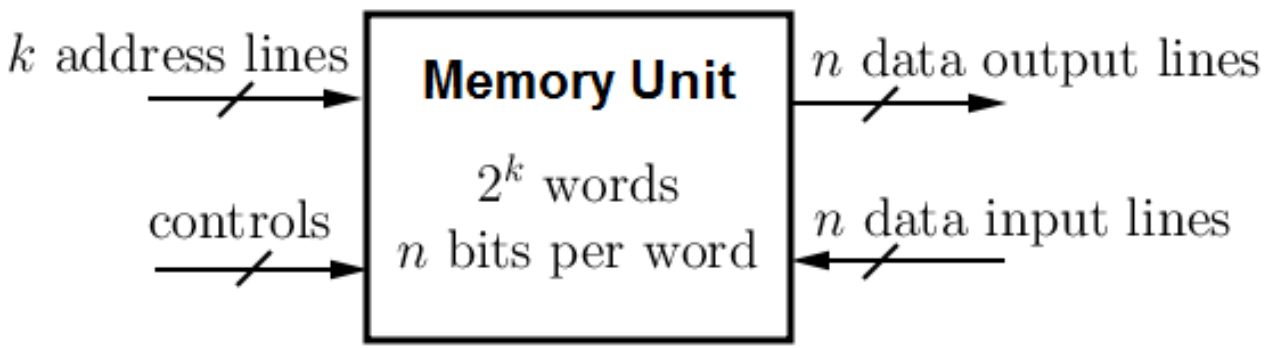
RAM and ROM memories

Memories with higher capacity are required.

- RAM or Random Access Memory ("mémoire vive") :
  - volatile memory : data are lost at power off
  - read and write operations
  - direct access
  - sequential system
  - requires : n data input and output lines, k address lines, read/write controls
- ROM or Read Only Memory ("mémoire morte") :
  - non volatile memory : information is "programmed", stored permanently and cannot be changed in normal operation
  - no write operation in normal operation
  - direct access
  - combinational system
  - requires : n data output lines, k address lines

Random Access Memory

- logically similar to register file
- stores bits using a bit storage approach that is more efficient (in terms of size) than a flip-flop
- k address lines are decoded to address 2^k words of memory
- each word is n bits
- the control signals are used to define the memory operation (read, write, no operation or more elaborate operation)
- read values are put on the n output data lines
- values to be written in memory are put on the n input data lines, which can coincide to the output lines in case of bidirectional bus



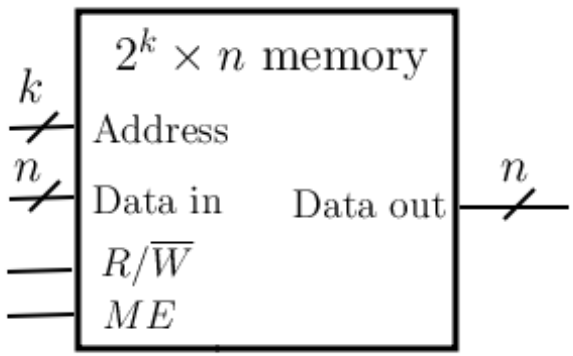
RAM memories : two technology families

- SRAM or Static RAM :
  - the binary information is stored by latches
  - the stored information is valid as long as supply power is applied
  - static behavior
- DRAM or Dynamic RAM
  - the binary information is stored in the form of electrical charges on a capacitor
  - the capacitor tends to discharge with time, dynamic behavior
  - the capacitor has to be periodically recharged
  - the DRAM cell is refreshed every few ms, reading and rewriting the information stored
- Comparison :

	SRAM	DRAM
Cost	high	moderate
Speed	high	slower
Refresh	No	Yes
Power consumption	high	low

Read and write operations

Consider a $2^k \times n$ memory. Its behavior is determined by the two control signals ME and $R/\overline{W}$.

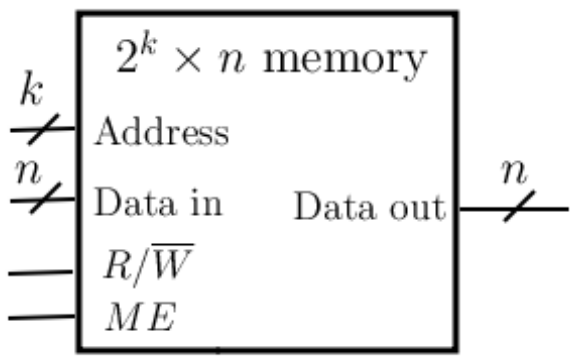


$ME = 0$		No operation
$ME = 1$	$R/\overline{W} = 1$	Read
$ME = 1$	$R/\overline{W} = 0$	Write

- ME is the memory enable control, it enables or disables the RAM
- Address specifies the location to read from or write to
- $R/\overline{W}$ selects between reading or writing
  - to read, $R/\overline{W} = 1$ and Data out will be the n-bit value stored at Address
  - to write, $R/\overline{W} = 0$ and Data is is the n-bit value to save at Address

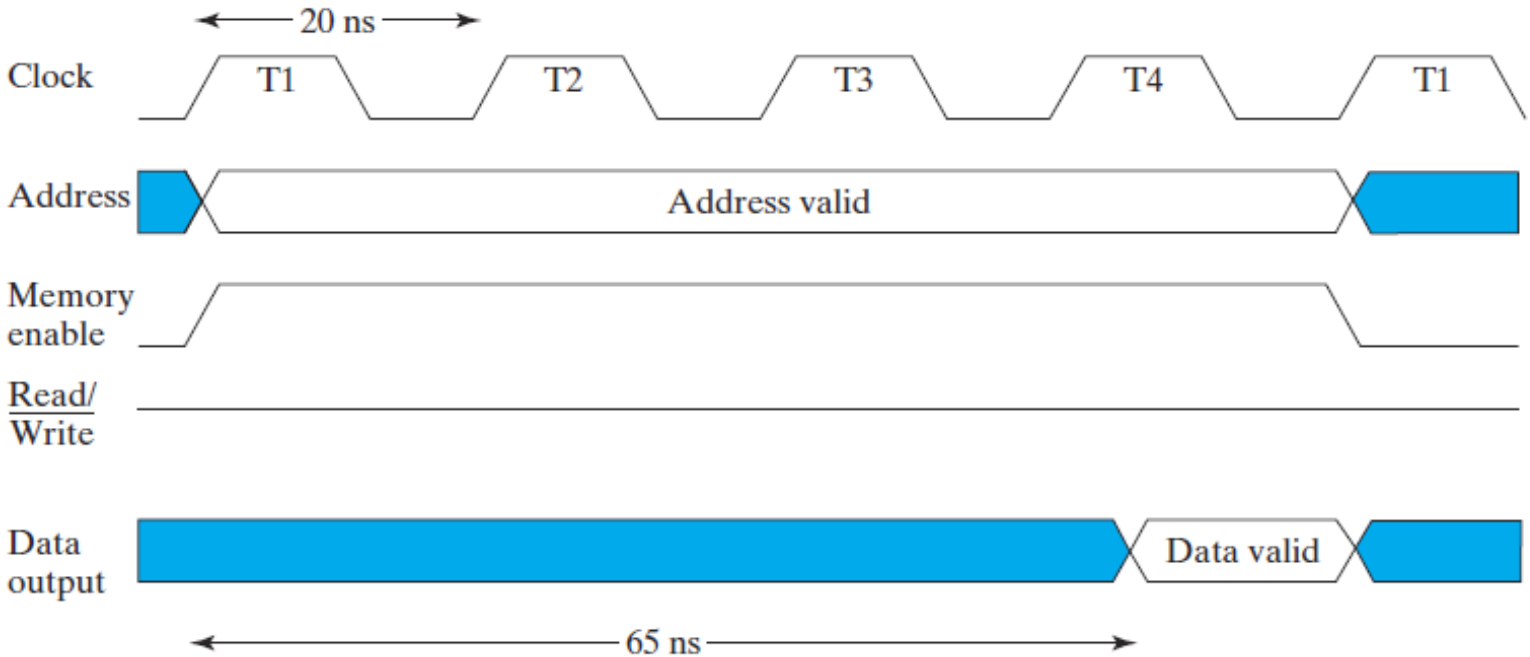
Read and write operations

- To read from this RAM, the controlling circuit must :
  - enables the operation by setting $ME = 1$
  - select the read operation by setting $R/\overline{W} = 1$
  - place a valid address on the Address input
  - the contents of the memory at Address appear on data out at the end of the access time, after transients have died out and signals stabilized
  - The input Data in is not used for read operation
- To write to this RAM, the controlling circuit must :
  - enables the operation by setting $ME = 1$
  - select the write operation by setting $R/\overline{W} = 0$
  - place the desired address to the Address input
  - place the word to be stored on the Data input
  - The output Data out is not used for write operation



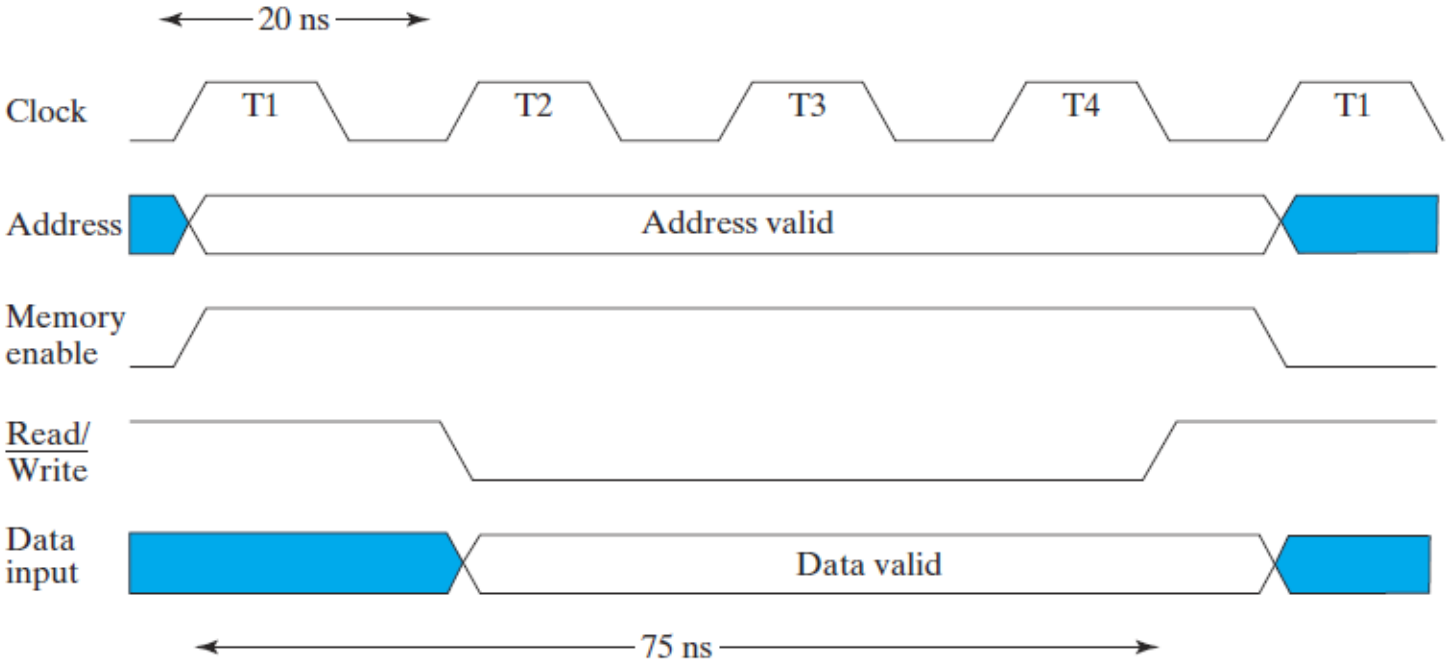
Memory operation timing

- Most basic memory are asynchronous
- Read and write operations are timed by changes in values on the control signals ME and $R/\overline{W}$. Example of read timing :
  - CPU clock period = 20 ns, access time = 65 ns, cycle time = 75 ns
  - 4 clock pulses are required for one memory request
  - the address is applied with ME set to 1, $R/\overline{W} = 1$
  - the memory places the data outputs after the access time of 65 ns
  - the data are transferred into internal registers of the CPU at the next clock pulse



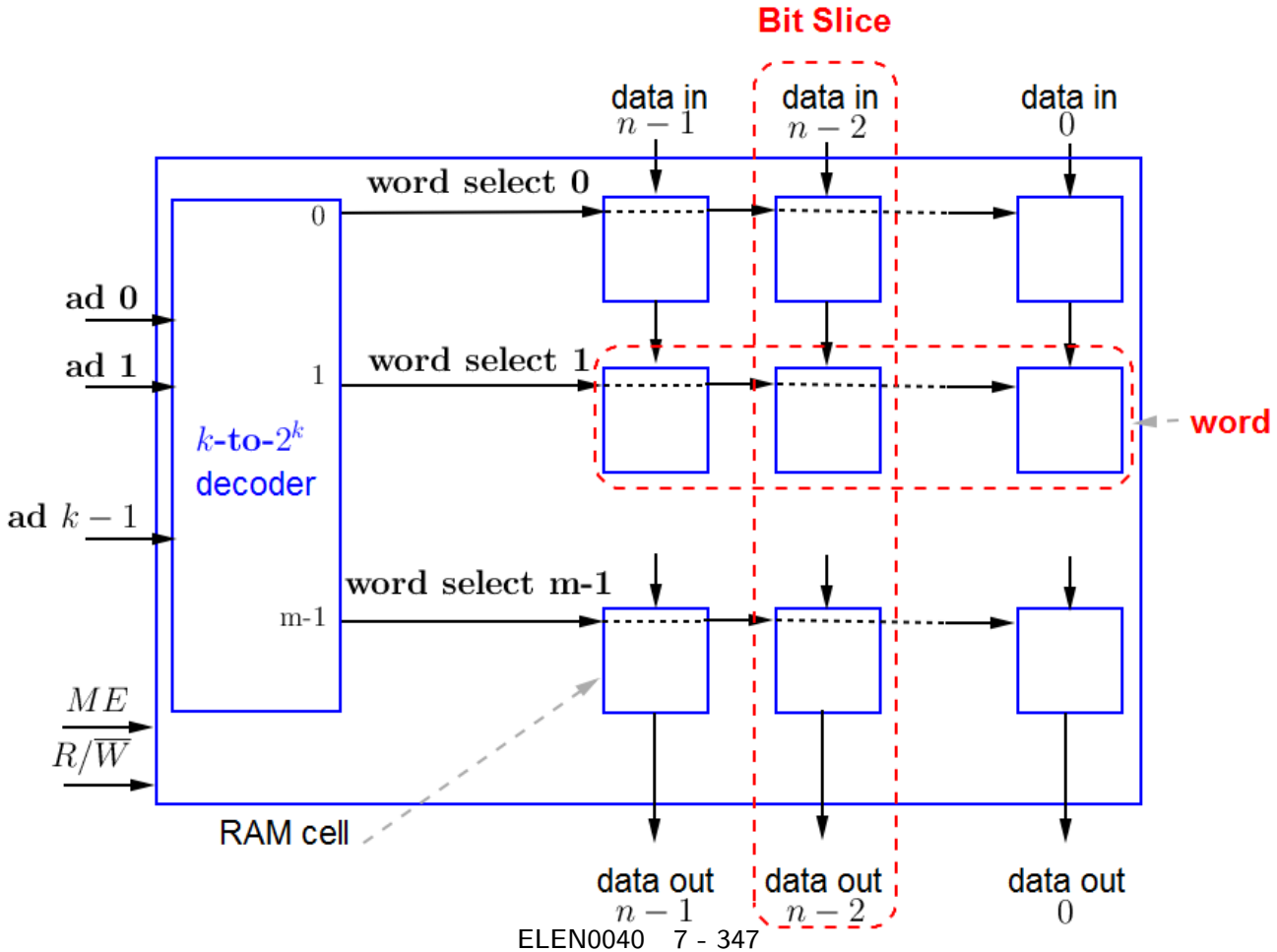
Memory operation timing

- Example of write timing
  - the address is applied first with ME set to 1, $R/\overline{W} = 1$
  - the address must be established at least a specified time before $R/\overline{W}$ changes to 0 and held for at least a specified time before $R/\overline{W}$ goes back to 1 to avoid disturbing stored contents of other addresses
  - data inputs must be established at least a specified time before and after $R/\overline{W}$ goes back to 1 to write correctly
  - ME remains at 1 during minimum the cycle time = the time required by the memory to complete its operations



Static RAM realization

- A memory is made of an association of several RAM chips plus a control logic circuit.
- The internal structure of a RAM chip of $m = 2^k$ words of n bits consists of an array of mn elementary storage cells + the required control logic circuit



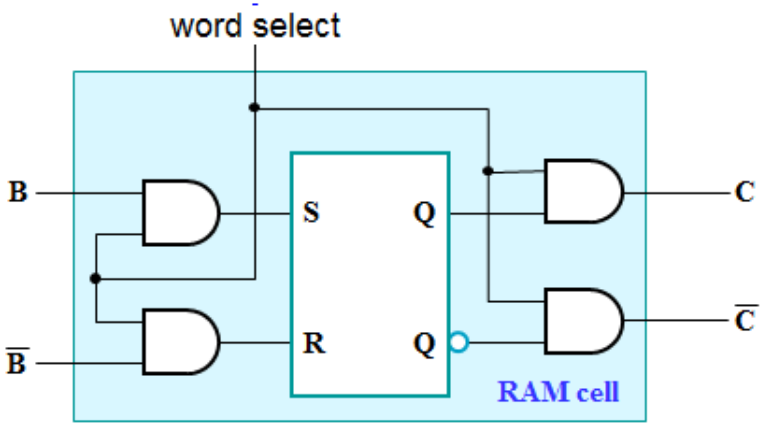
Static RAM realization

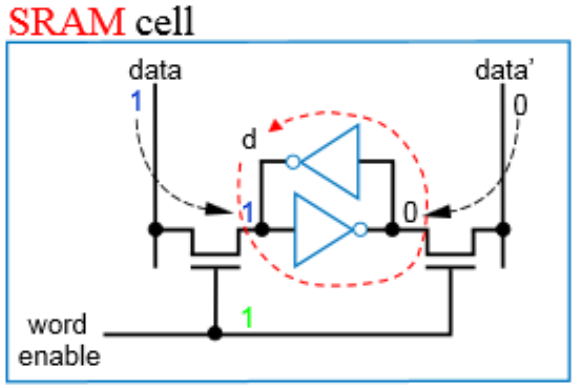
We will consider successively :

- the constitution of an SRAM cell : the basic element to store one bit
- the association of $m = 2^k$ RAM cells to build a bit slice memory of m words of one bit
- the association of n bit slices to build a memory chip of m words of n bits
- for larger memories, the association of several memory chips

SRAM cell

- The basic storage cell is an electronic circuit using 6 transistors whose logic behavior may be modeled by an SR latch.



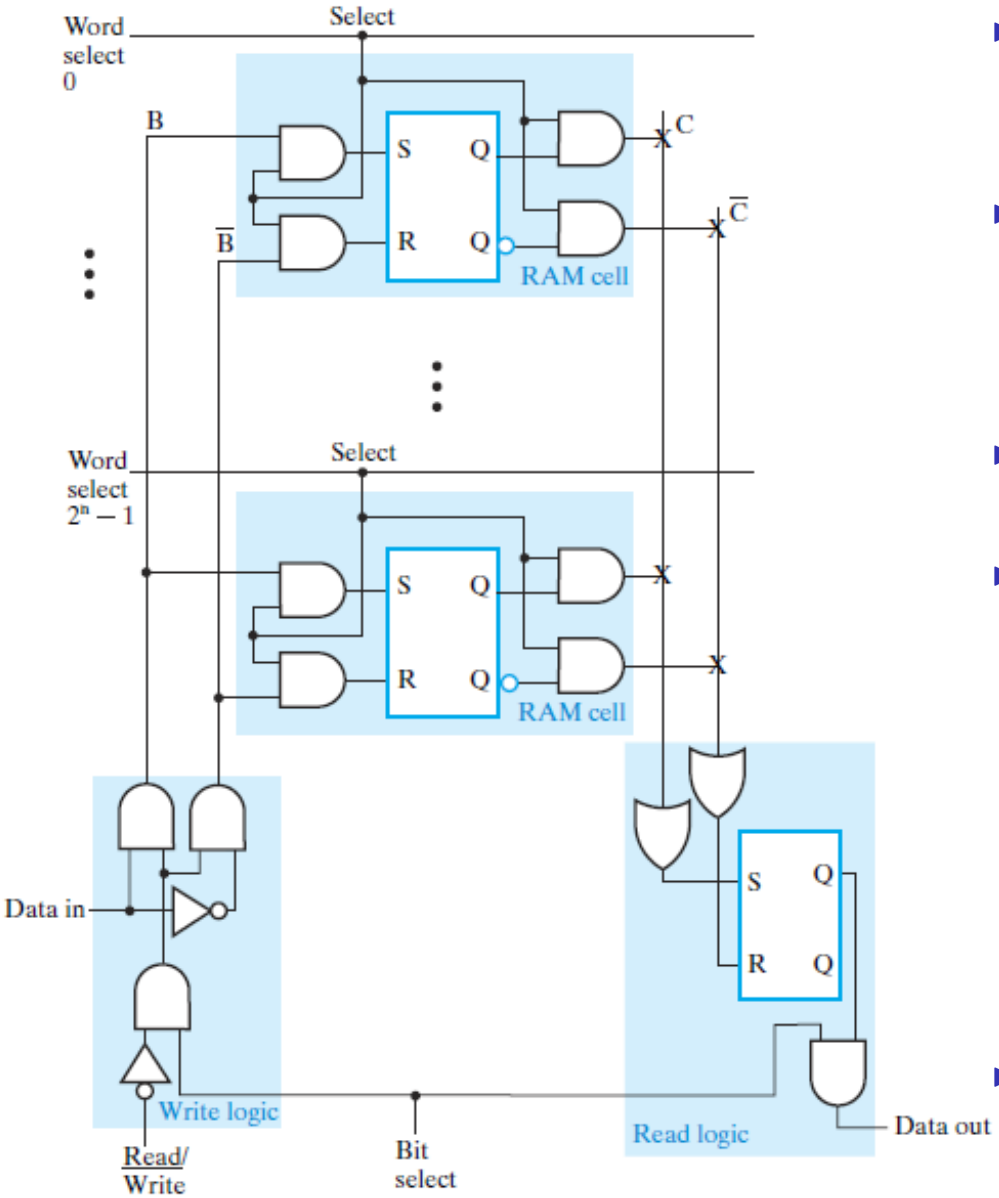


- Logic behavior :

  word select=0 $C = \bar{C} = 0$ Q holds stored value
  word select=1 $C = B$, $\bar{C} = \bar{B}$

- Read and write operations are enabled by the "word select" input
- B, $\bar{B}$ are used for a write operation
- C, $\bar{C}$ are used for a read operation

SRAM Bit Slice - Example



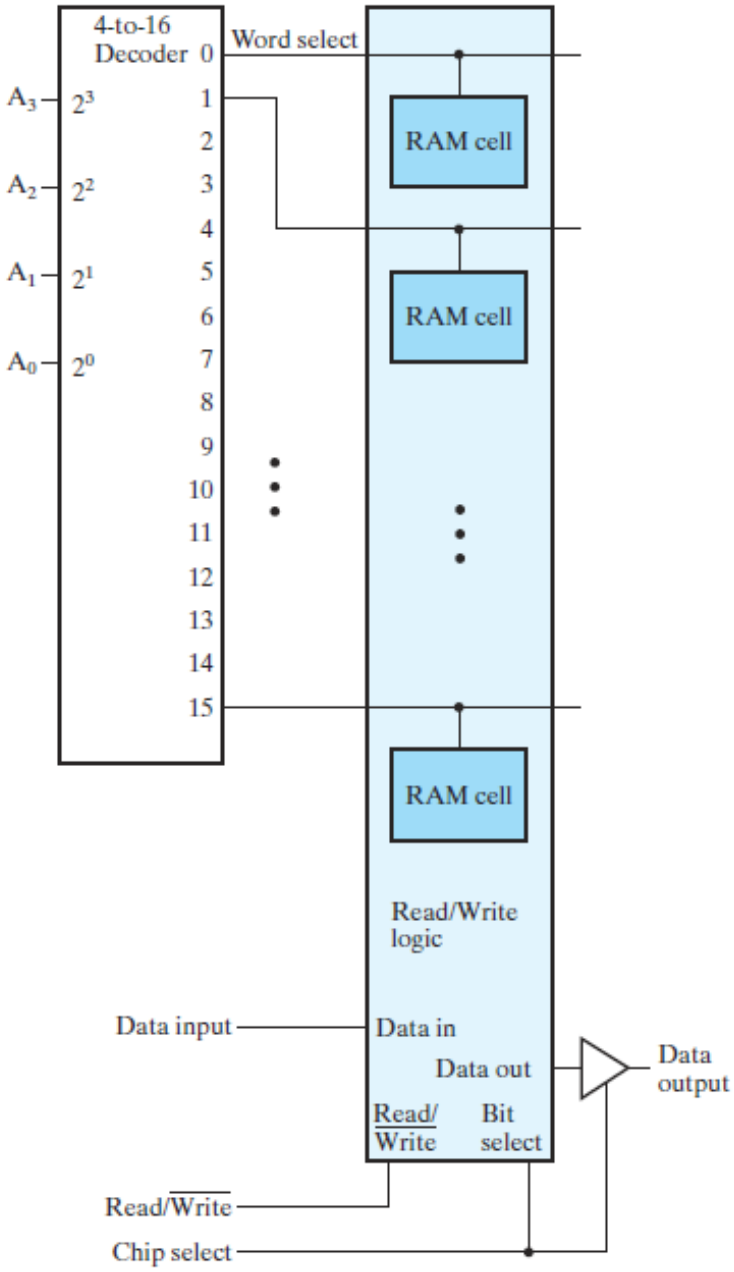
- m RAM cells
- the write logic circuit derives B and $\bar{B}$ inputs of the bit cells from the $R/\overline{W}$ control and the Data In input
- the read logic circuit provides the Data Out output from the C and $\bar{C}$ outputs of the bit cells and the $R/\overline{W}$ control
- the memory enable signal is "Bit select"
- Write Logic :

$$B = BitSelect.\overline{R/\overline{W}}.DataIn$$

$$\bar{B} = BitSelect.\overline{R/\overline{W}}.\overline{DataIn}$$

  a new value is stored if the operation is enabled (BitSelect=1) and write operation is selected ($R/\overline{W} = 0$)
- If $R/\overline{W} = 1$, B and $\bar{B}$ of each cell are 0 and memorized values are kept unchanged

$2^k \times 1$ SRAM Chip

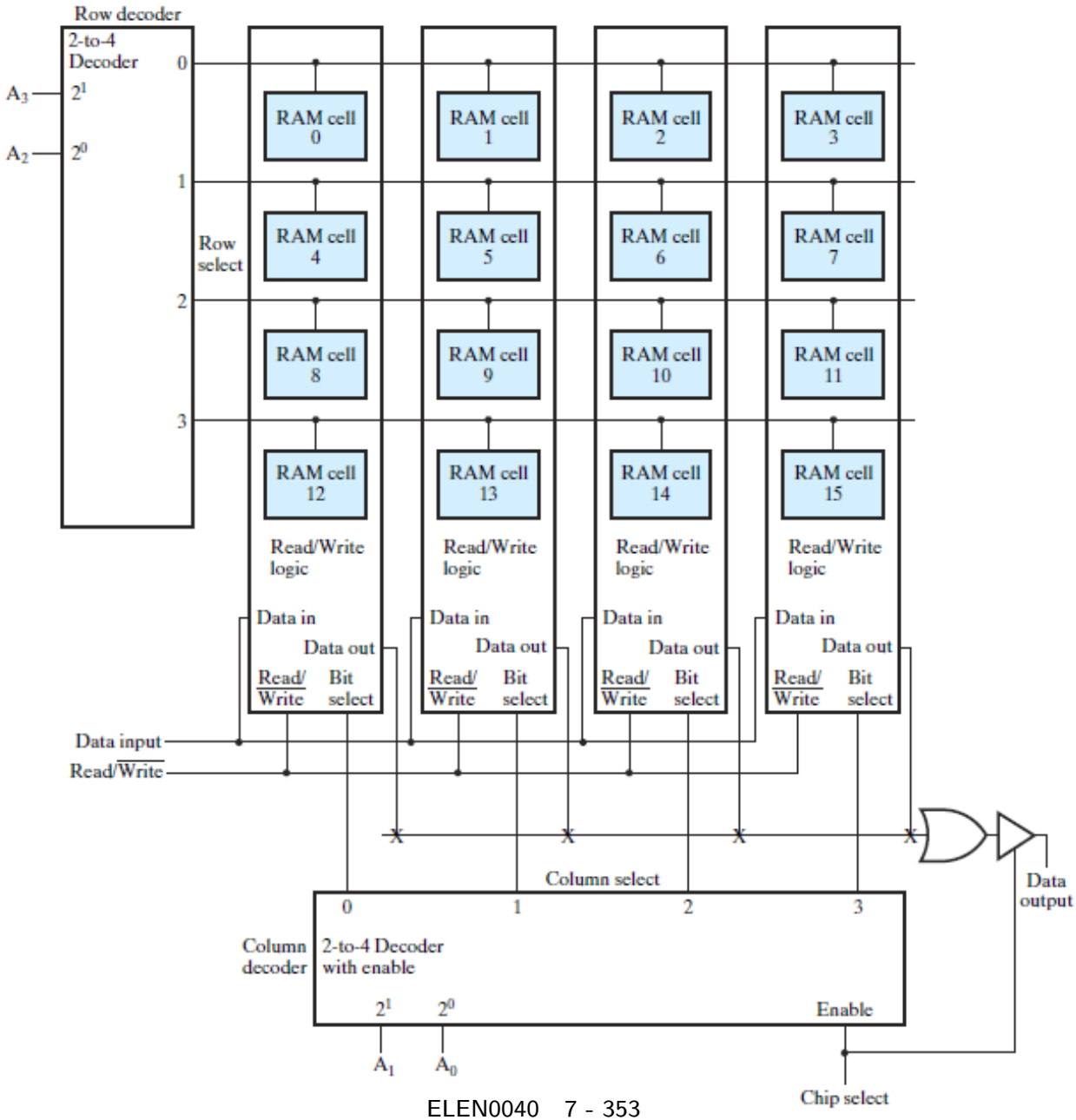


- To build a memory chip with 1-bit memory words, we need :
  - a decoder to decodes the k address lines to the 2^k Word Select signals
  - a 3-state buffer at the output to be able to combine data outputs from different memory chips (one Chip Select at 1 at a time)
  - the memory enable signal is "Chip select"
- Example of 16×1 SRAM Chip
- For words of 2 bits : connect
  - a second Bit Slice for the second data input
  - a 3-state buffer at the second data output
- ...

Using an array of memory cells

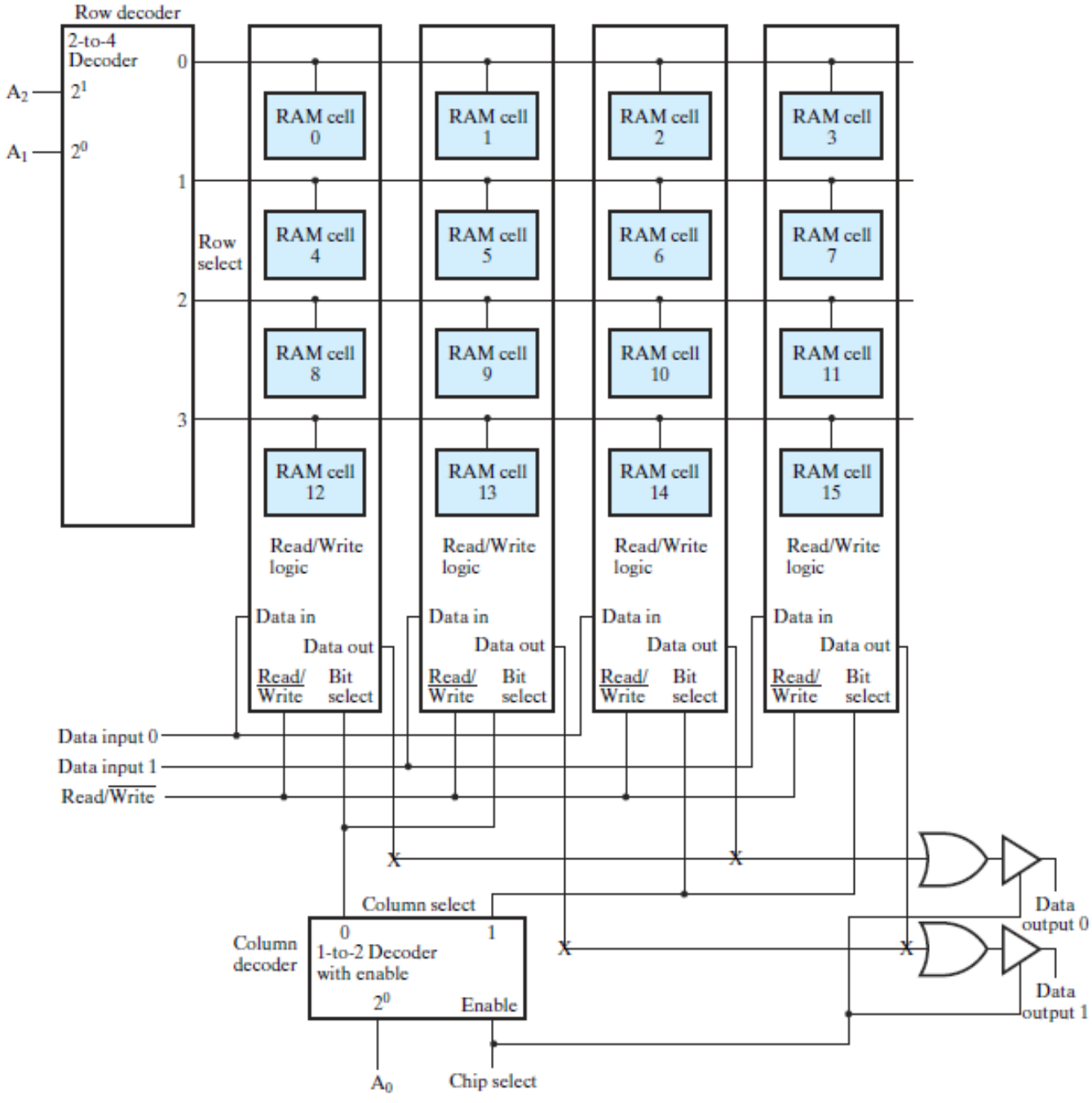
- Memory can have a large number of words
- The decoder size and the important fanouts of gates lead to an increase of the access and cycle times
- The use of a two-dimensional array of cells leads to better performances
- The cell selection uses two decoders of smaller size and several bit slices
- One cell is identified by its row and column
  - The address bits are split into two sets :
    - the half most significant bits are used to select the row
    - the half less significant bits are used to select the column, which corresponds to a bit slice
  - Word Select is replaced by Row Select
  - Bit Select is replaced by Column Select

Example : 16 × 1 RAM



RAM Chip with word length $>$ 1bit

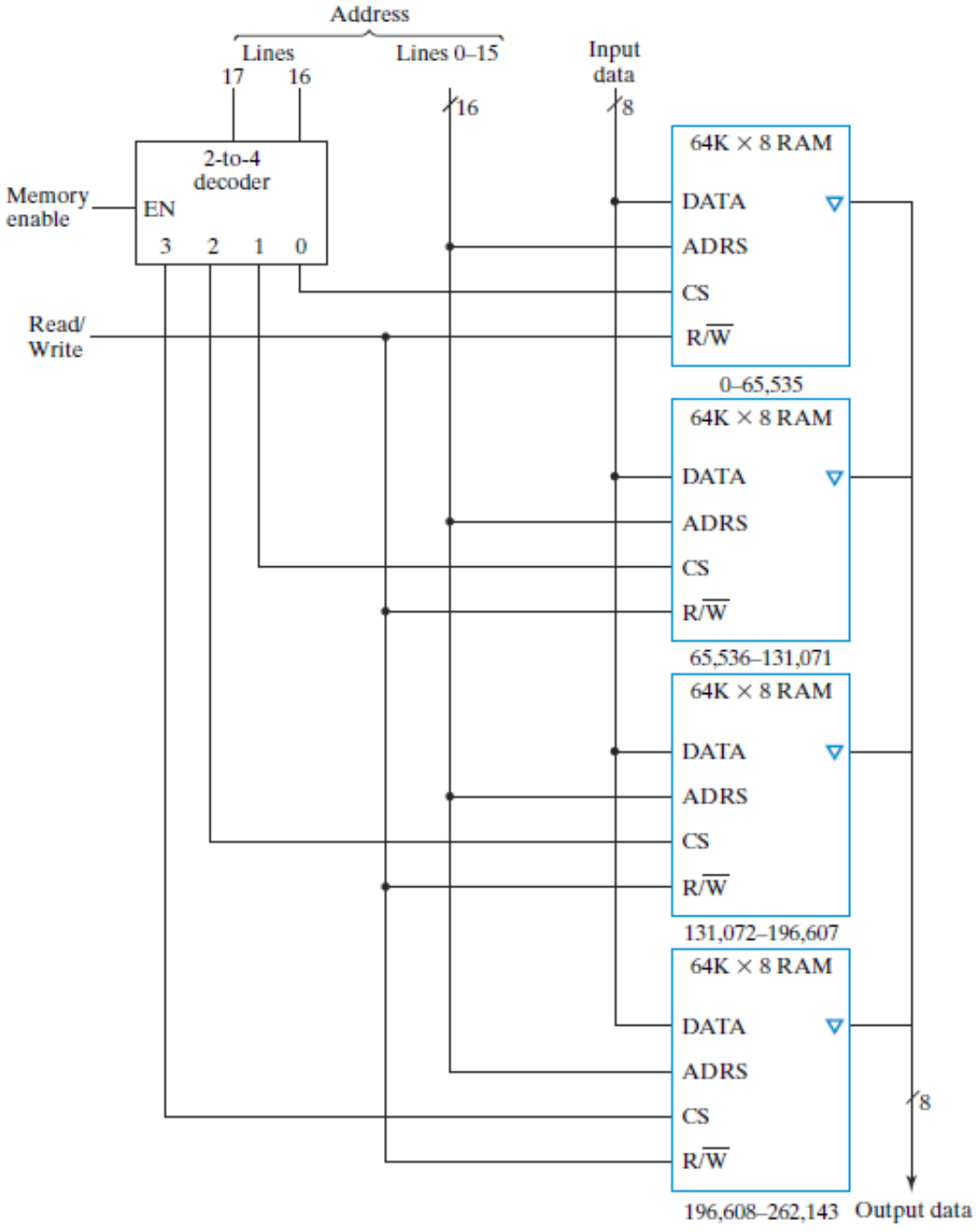
- The same RAM cell array can be used to produce a 8×2 RAM Chip
- 8 words $\rightarrow$ 3 address bits
- Row decoder : 2-to-4
- Column decoder : 1-to-2
- two Data In inputs, one for each bit of data
- each Column Select selects two columns, one for each Data In bit
- example : address 001 corresponds to cells 2,3
- two Data Out outputs


Row decoder
2-to-4
Decoder
A_2
A_1
2^1
2^0
0
1
2
3
Row
select
RAM cell 0
RAM cell 1
RAM cell 2
RAM cell 3
RAM cell 4
RAM cell 5
RAM cell 6
RAM cell 7
RAM cell 8
RAM cell 9
RAM cell 10
RAM cell 11
RAM cell 12
RAM cell 13
RAM cell 14
RAM cell 15
Read/Write logic
Data in
Data out
Read/ Write
Bit select
Data input 0
Data input 1
Read/$\overline{\text{Write}}$
Column select
Column decoder
0
1
1-to-2 Decoder with enable
2^0
Enable
A_0
Chip select
Data output 0
Data output 1

Association of SRAM Chips

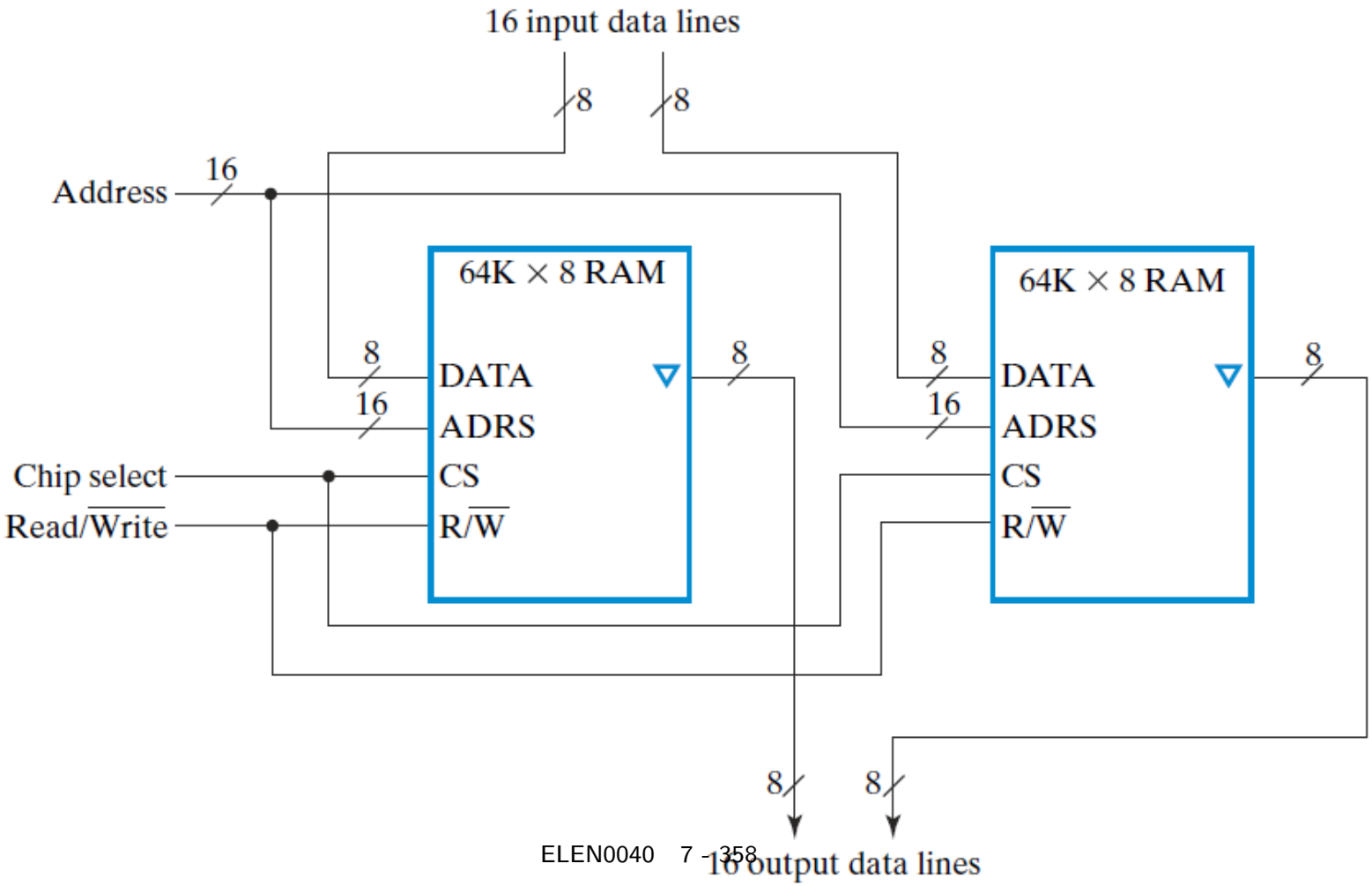
- Several RAM Chips are usually associated to :
  1. build memories of larger capacity, increase m
  2. provide a larger word length, increase n

1. Increase memory capacity
   - example : use four $64K \times 8$ RAMs to provide $256K \times 8$ RAM
   - $64K = 2^{16}$: 16 address bits, $256K = 2^{18}$, 2 additional address bits are required
   - a 2-to-4 decoder is used to provide the 4 Chip Select signals to select one RAM chip, there is only one signal active at a time
   - the data input lines and the 16 least significant address lines are shared by the 4 Chips
   - the data output lines are connected together (Chip select on the enable input of the 3-state buffers present on each Chip output)


Address
Lines
17 16
Lines 0–15
Input
data
16
8
2-to-4
decoder
Memory
enable
EN
3 2 1 0
Read/
Write
64K × 8 RAM
DATA
ADRS
CS
R/W
0–65,535
65,536–131,071
131,072–196,607
196,608–262,143
8
Output data

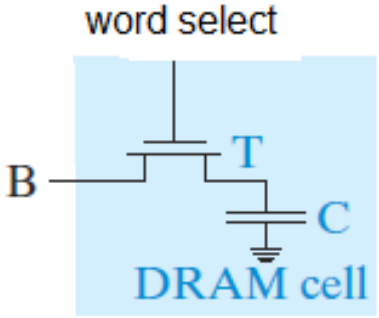
Increase the word length

2. Combine two $64K \times 8$ RAMs to build a $64K \times 16$ RAM
   - put the memory chips side-by-side
   - the address and control lines are shared
   - the two sets of 8 data input and output lines are concatenated to form the 16 output lines



Dynamic RAM - DRAM Cell

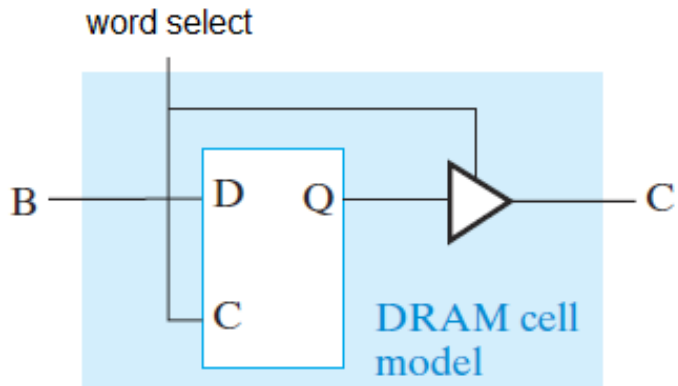
- The DRAM Cell is made of a transistor and an capacitor
- The capacitor is used to store electrical charges
  - high level of charges : logic value 1
  - low level of charges : logic value 0
- The transistor acts as a switch :
  - switch open : the value is stored
  - switch closed : a new value is transferred



- This cell needs to be refreshed periodically :
  - when the switch is open, there exists a small leakage current
  - the charge on the capacitor varies
  - the stored value will eventually be lost $\rightarrow$ refresh operation is needed
- During a read operation, the capacitor slightly charges or discharges and the amount of charges present on the capacitor at the end of the read operation does not correspond to 0 or 1 level, so that the stored value is destroyed = destructive read
- The original value has to be restored

DRAM Cell - Logic model

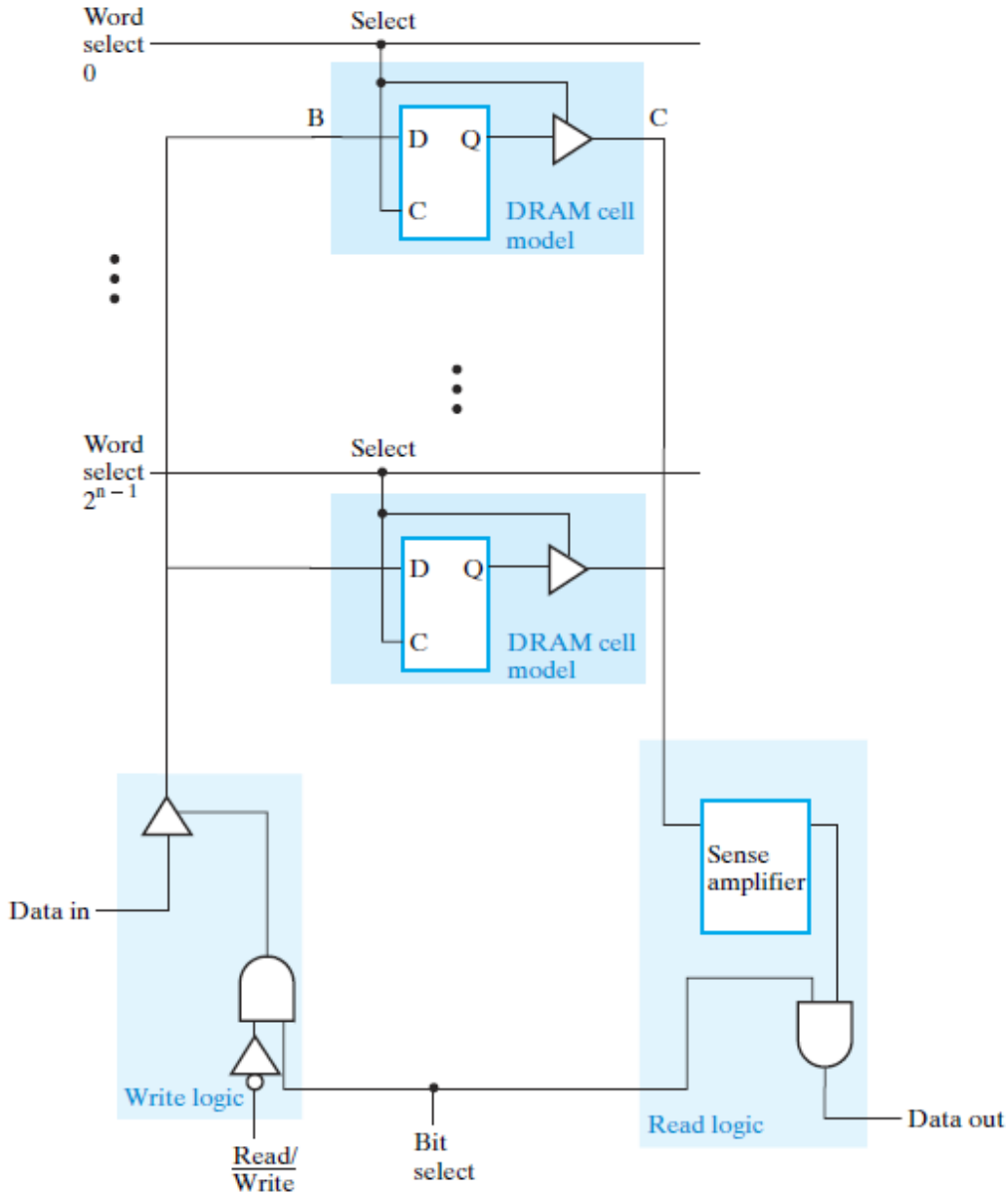
- The logic cell behavior can be modeled by a D latch and a 3-state buffer



word select=0 C = High-Z Q holds stored value
word select=1 $C = B$

- Read and write operations are enabled by the "word select" input
- The DRAM Bit Slice model is similar to that of the SRAM bit but :
  - a sense amplifier is used to adapt the small voltage change on the capacitor to the high and low voltage levels
  - a 3-state buffer is used in the read logic circuit

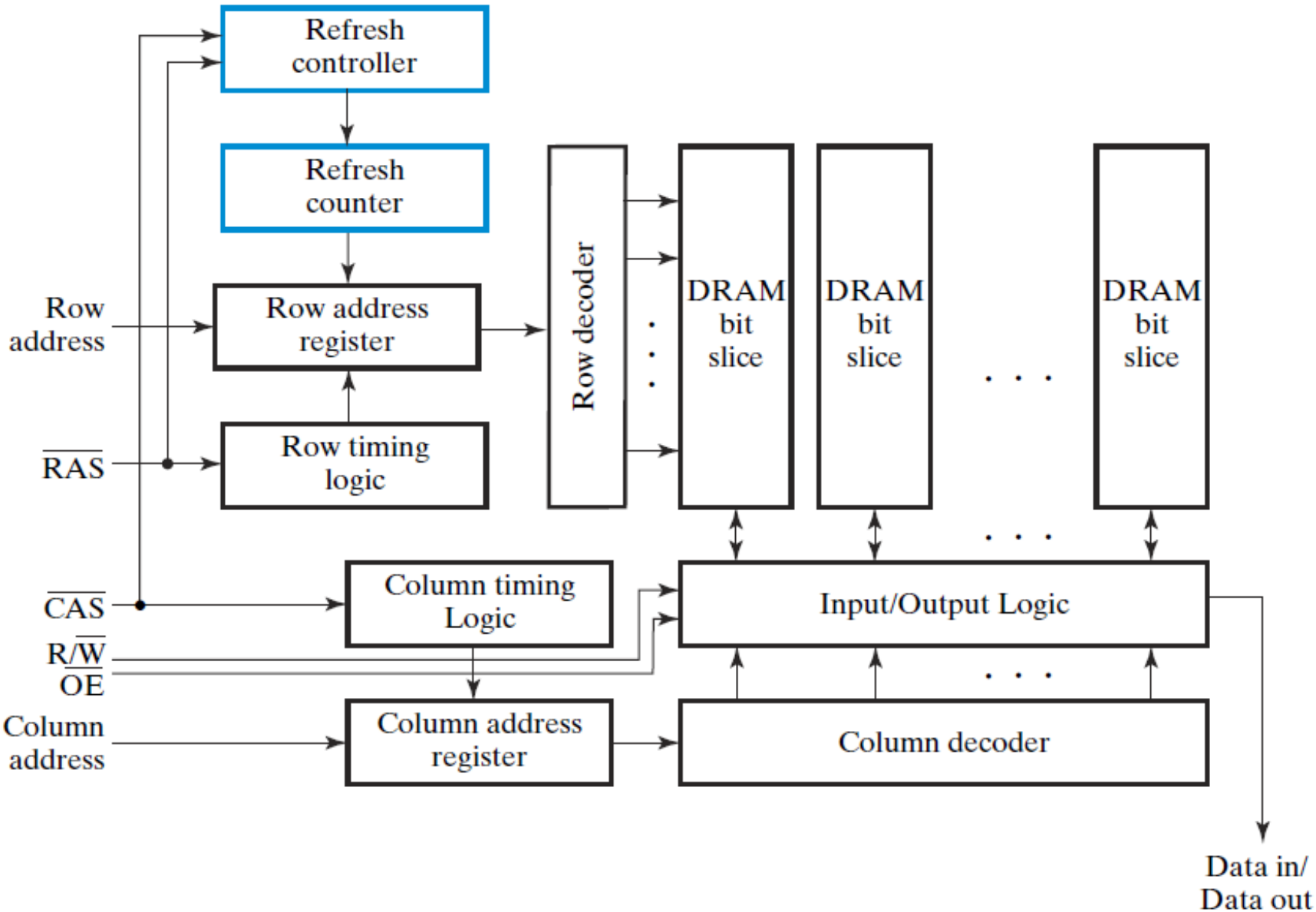
DRAM Bit Slice



DRAM Chip

- As for SRAM, The design of a DRAM chip is made from an array of mn cells
- The DRAM cell is simpler than the SRAM cell and thus the number of DRAM cells on a chip can be higher ; DRAM chips are used to build large memories
- Hence, the number of address bits is high and to limit the complexity of the selecting circuit, the row and column addresses are applied serially
- A row and a column register are added to store the address during the operations
- The control timing signals are :
  - $\overline{RAS}$ to control the loading of the row address into the row-register, the row address is loaded first
  - $\overline{CAS}$ to control the loading of the column address into the column-register
  - $R/\overline{W}$ to select between read and write operation
  - $\overline{OE}$ enables the output on Data output lines in a read operation
- The refresh logic includes a refresh controller and a refresh counter

DRAM Block Diagram



Association of DRAM Chips

- Similar procedures as for SDRAM : build larger memories with increased capacity and word length
- Addition of a DRAM Controller to :
  - separate the address into row address and column address and define the timing of their application
  - perform the periodic refresh operation
  - provide status signals to the rest of the system, indicating e.g. whether or not the memory is available or busy performing refresh,..

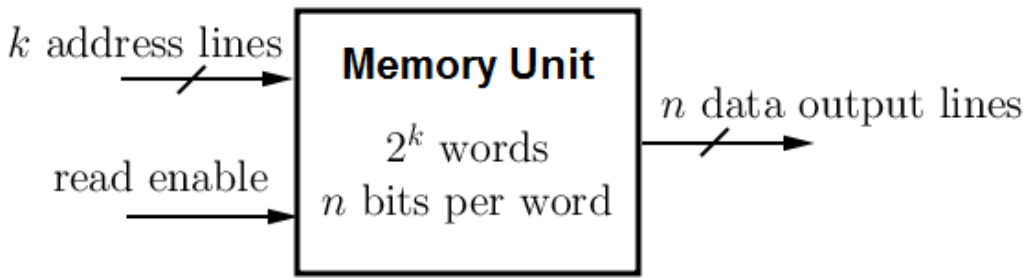
DRAM types

Technology improvements enable to increase capacity and speed at the price of more sophisticated control. Actually, three main types of performing DRAM :

- Synchronous DRAM :
  - transfers to and from the DRAM are synchronized with the system clock
  - take advantage of the row/column address separation : contiguous words can be read using the same row address
- DDR DRAM : Double Data Rate synchronous DRAM, a synchronous DRAM where
  - transfers of data occur on both edges of the clock cycle
  - the transfer rate is thus doubled : 2 data words per clock cycle
- Rambus DRAM :
  - use a dedicated bus allowing transmission of packed data between memory and the memory bus to the processor

Read Only Memory

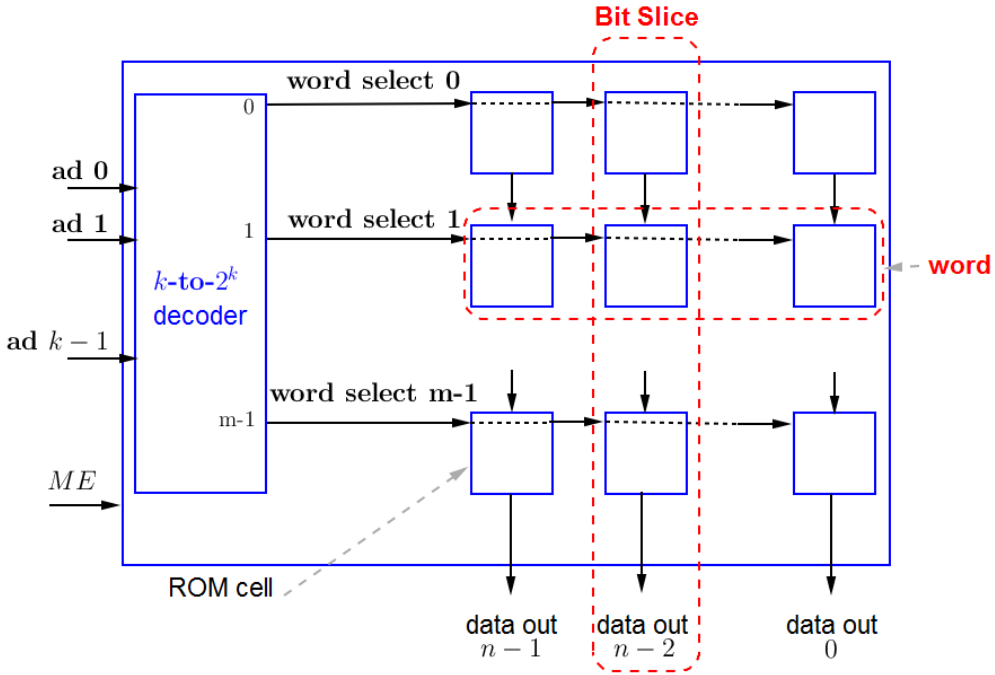
- Memory that can only be read from, not written to
  - k address lines
  - no data input lines
  - n output data lines
  - one possible control signal to enable the read operation, no $R/\overline{W}$ control input



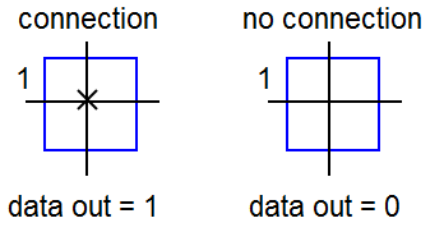
- Advantages over RAM :
  - compact : may be smaller
  - nonvolatile : save bits even when power supply is off
  - speed : may be faster (especially than DRAM)
  - used to store data or programs that won't change

ROM - internal structure

- The internal logical structure is similar to RAM, without the data input lines



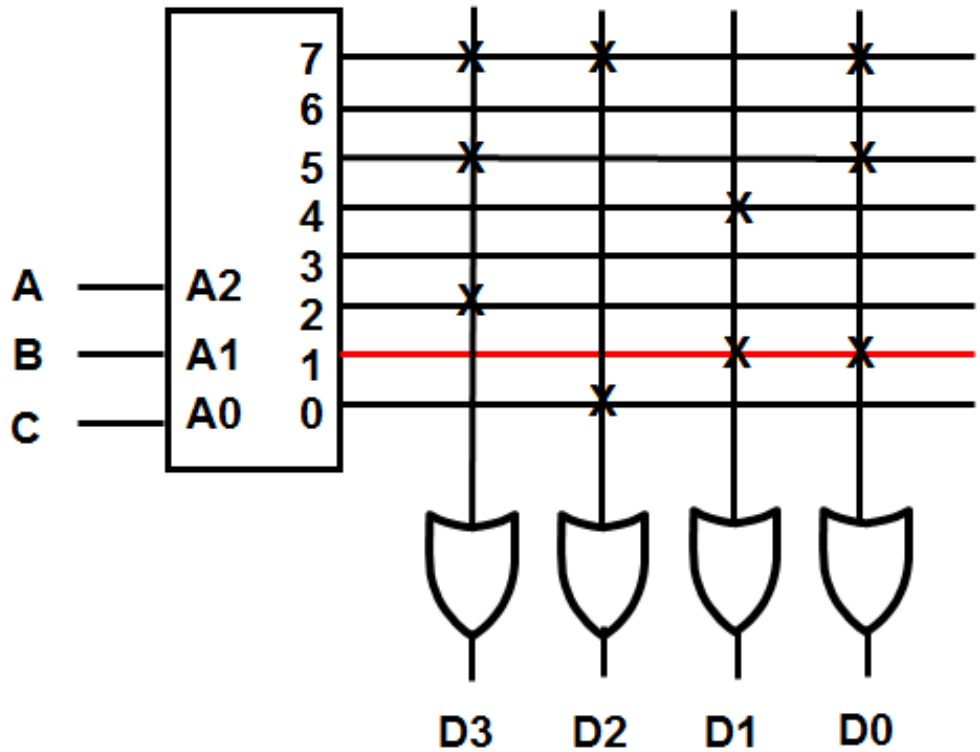
- The memory cell uses one transistor acting as a connector that can be programmed to model 0 or 1 logic value. This can be symbolically represented as follows :



ROM - Example

- a 8×4 ROM
- 3 address lines
- 4 data output lines
- Stored values

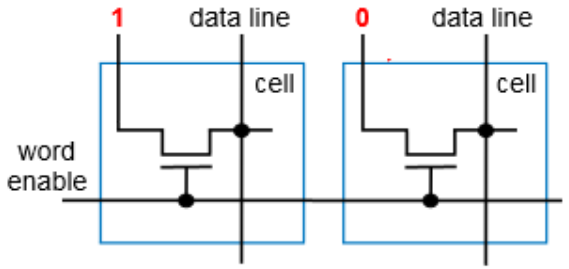
address	data
000	0100
001	0011
010	1000
011	0000
100	0010
101	1001
110	0000
111	1101



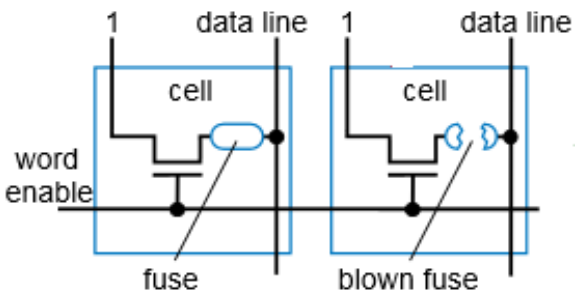
ROM types

Several types depending on the way the stored values are "programmed"

- ROM or mask-programmed ROM
  - bits are hardwired as 0s or 1s during chip manufacturing
  - word enable (from decoder) simply passes the hardwired value though the transistor
  - very fast and compact memory

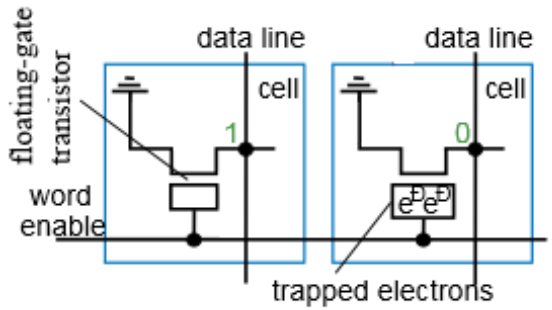


- PROM or Fused-Based Programmable ROM
  - each cell has a fuse
  - a special device, a "programmer",
    - blows fuses to code a 0 value
    - leaves fuse intact to code a 1 value
  - the memory can be programmed only once



ROM types (continued)

- EPROM or Erasable Programmable ROM
  - use a "floating-gate transistor"
  - a "programmer" device uses higher than normal voltage to "trap" electrons into the gate for cells that should store 0
  - for other cells, stored value will be 1
  - stored data can be erased by exposing the chip to a shine ultraviolet light
  - the whole memory is erased



- EEPROM or Electronically - Erasable Programmable ROM
  - similar to EPROM but erasing is done electronically
  - the erasing is done one word at a time
  - more bulky than EPROM

ROM types (continued)

- Flash Memory
  - Similar to EEPROM, but all words or large blocks of words can be erased simultaneously
  - Higher density of integration than EEPROM
  - Both types, when embedded in systems
    - can be reprogrammed to store new values
    - require bidirectional data lines and read/write control input
    - also need a busy status output to indicate that erasing is in progress
- We define that
  - RAM is readable and writable
  - ROM is read-only
- But some ROMs act almost like RAMs
  - EEPROM and Flash are in-system programmable but
  - write operation is slower than in RAMs
  - the number of writes may be limited
- Some RAMs act almost like ROMs
  - Nonvolatile RAMs have also been devised

Références

- *Logic and Computer Design Fundamentals, 4/E*, M. Morris Mano Charles Kime , Course material
  http ://writphotec.com/mano4/
- *Cours d'électronique numérique*, Aurélie Gensbittel, Bertrand Granado, Université Pierre et Marie Curie
  http ://bertrand.granado.free.fr/Licence/ue201/coursbeameranime.pdf
- *Lecture notes, Course CSE370 - Introduction to Digital Design*, Spring 2006, University of Washington,
  https ://courses.cs.washington.edu/courses/cse370/06sp/pdfs/
- *Lecture notes, Course ECE274 - Digital Logic*, Spring 2009, University of Arizona,
  http ://www2.engr.arizona.edu/ rlysecky/courses/ece274-09s/Main/Lectures.html

Terms of Use

- All (or portions) of this material © 2008 by Pearson Education, Inc.
- Permission is given to incorporate this material or adaptations thereof into classroom presentations and handouts to instructors in courses adopting the latest edition of Logic and Computer Design Fundamentals as the course textbook.
- These materials or adaptations thereof are not to be sold or otherwise offered for consideration.
- This Terms of Use slide or page is to be included within the original materials or any adaptations thereof.