

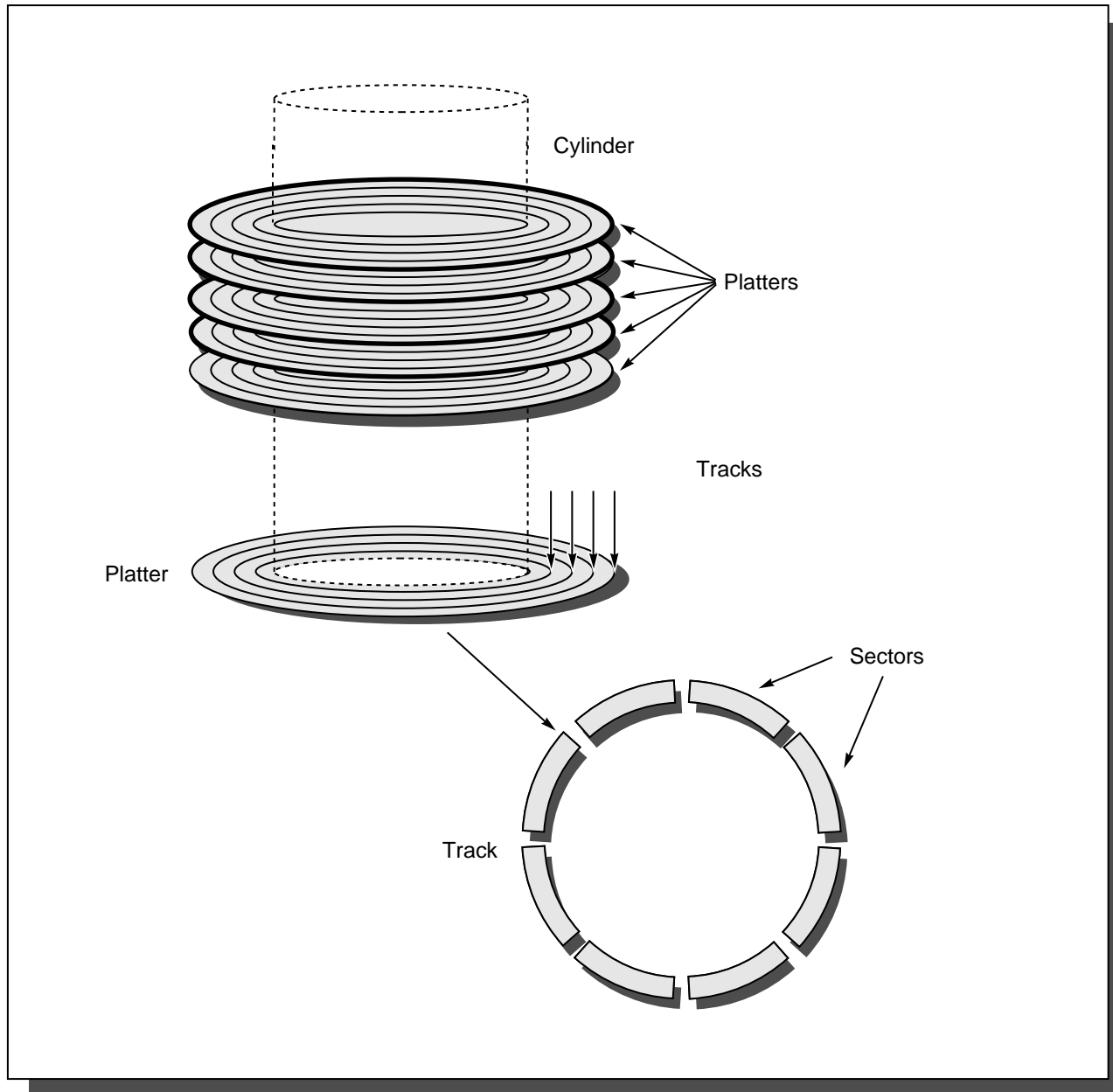
## TYPES OF MEMORY

- Non-random access
  - ▷ Latency dependent on addresses of last block and current block
  - ▷ Examples:
    - Magnetic disk (latency = seek time + rotation time)
    - Magnetic tape (completely sequential access)
    - CD-ROM
  - ▷ Several contiguous blocks take almost the same time as one block
- Random access memory (RAM)
  - ▷ Latency independent of address
  - ▷ Example: Semiconductor memory
    - Static RAM (SRAM) — Retains data as long as power is on
    - Dynamic RAM (DRAM) — Data must be refreshed periodically

## DISK MEMORY (1)

- A **hard disk** records bits as patterns of magnetization in a thin coating on the surface of a rigid platter
  - ▷ Information is read or written through a set of **heads** attached to the ends of an **arm** with multiple segments





**Disks are organized into platters, cylinders, tracks, and sectors**

**DISK MEMORY (2)**

- The latency of a magnetic disk access is given by the formula

latency = seek time + rotational delay + transfer time + controller delay  
(not counting delays due to the fact that other processes may be using the disk)

- Terminology:

- ▷ Seek time = time for the arm to move from its present track to the track where the requested data is located
- ▷ Rotational time = time for the requested sector to rotate to the position of the arm

$$\text{rotational time} \leq \frac{60}{\text{disk rpm}}$$

- ▷ Transfer time = time for data transfer from disk to main memory

$$\text{transfer time} \geq \frac{\text{amount of data in MB}}{\text{transfer rate (in MB/s)}}$$

**DISK MEMORY (3)**

- Example: latency of writing one 512-byte sector on a magnetic disk rotating at 5400 rpm, with the following parameters:

Average seek time = 12 ms

Transfer rate = 5 MB/s

Controller delay = 2 ms

$$\begin{aligned}\text{average latency} &= 12 \text{ ms} + \frac{60}{5400} \text{ s} + \frac{0.5 \times 10^{-3} \text{ MB}}{5 \text{ MB/s}} + 2 \text{ ms} \\ &= 19.7 \text{ ms}\end{aligned}$$

## RANDOM-ACCESS MEMORY (1)

- Read-only
  - ▷ Mask-programmed ROM
  - ▷ User-programmable (PROM)
- Read mostly
  - ▷ EPROM, EEPROM
- Read/write
  - ▷ Static RAM (SRAM): Semiconductor logic
    - Fast; no need for refresh cycle
    - Costly; hot
  - ▷ Dynamic RAM (DRAM): Arrays of CMOS capacitors
    - Low cost & heat dissipation; a commodity product
    - High capacity
    - Slow because of complex addressing and refresh cycle

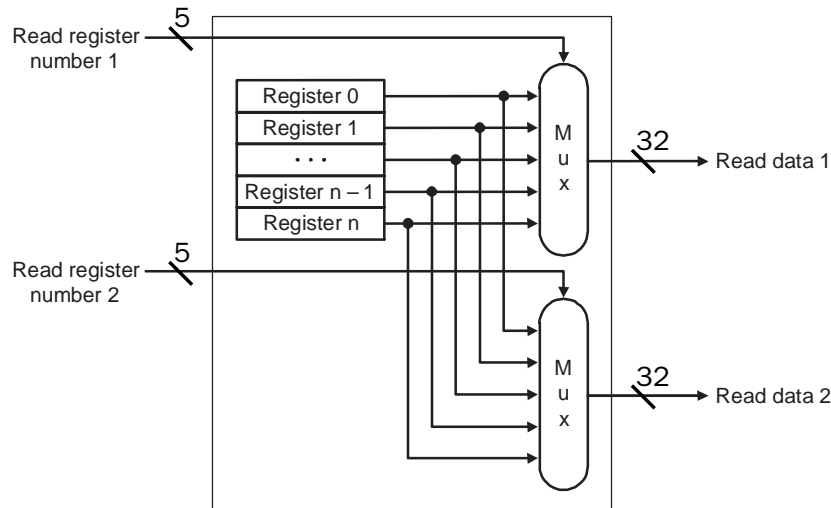
## RANDOM-ACCESS MEMORY (2)

- SRAM
  - ▷ Access time: 5–15 ns
  - ▷ Price (8/96): \$80–200/MB
  - ▷ Hierarchical design of SRAM memory:
    - Cell design
    - IC design
    - SRAM memory system design
- SRAM cell design:
  - ▷ Clocked D-latch (single or master-slave configuration)
  - ▷ Tristate output
    - Enable asserted: Normal latch operation
    - Enable deasserted: Latch acts as an unconnected wire (high-impedance state)

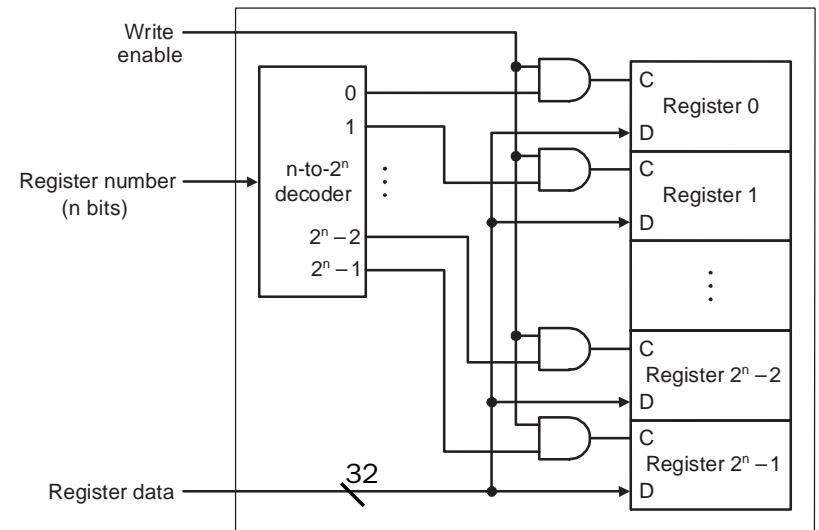
# RANDOM-ACCESS MEMORY (3)

- Memory cells are accessed as an array

▷ One-dimensional array (example: register file)



READ ACCESS

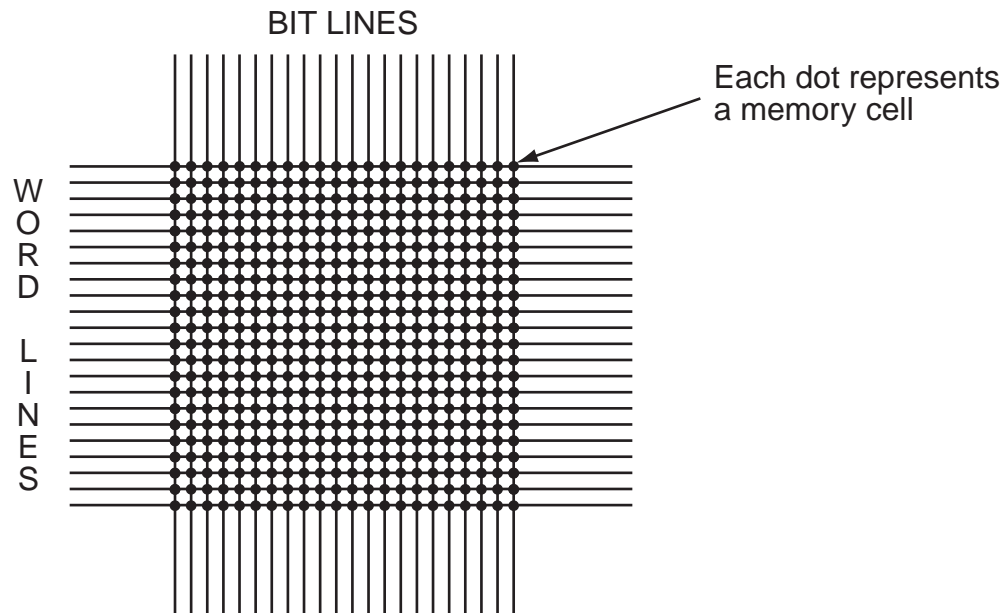


WRITE ACCESS

- Read access to a specific register is afforded through a multiplexor
  - ◇ Select signal is the register address (register number)
- Write access to a specific register is obtained by decoding the register address, which asserts that register's write enable signal

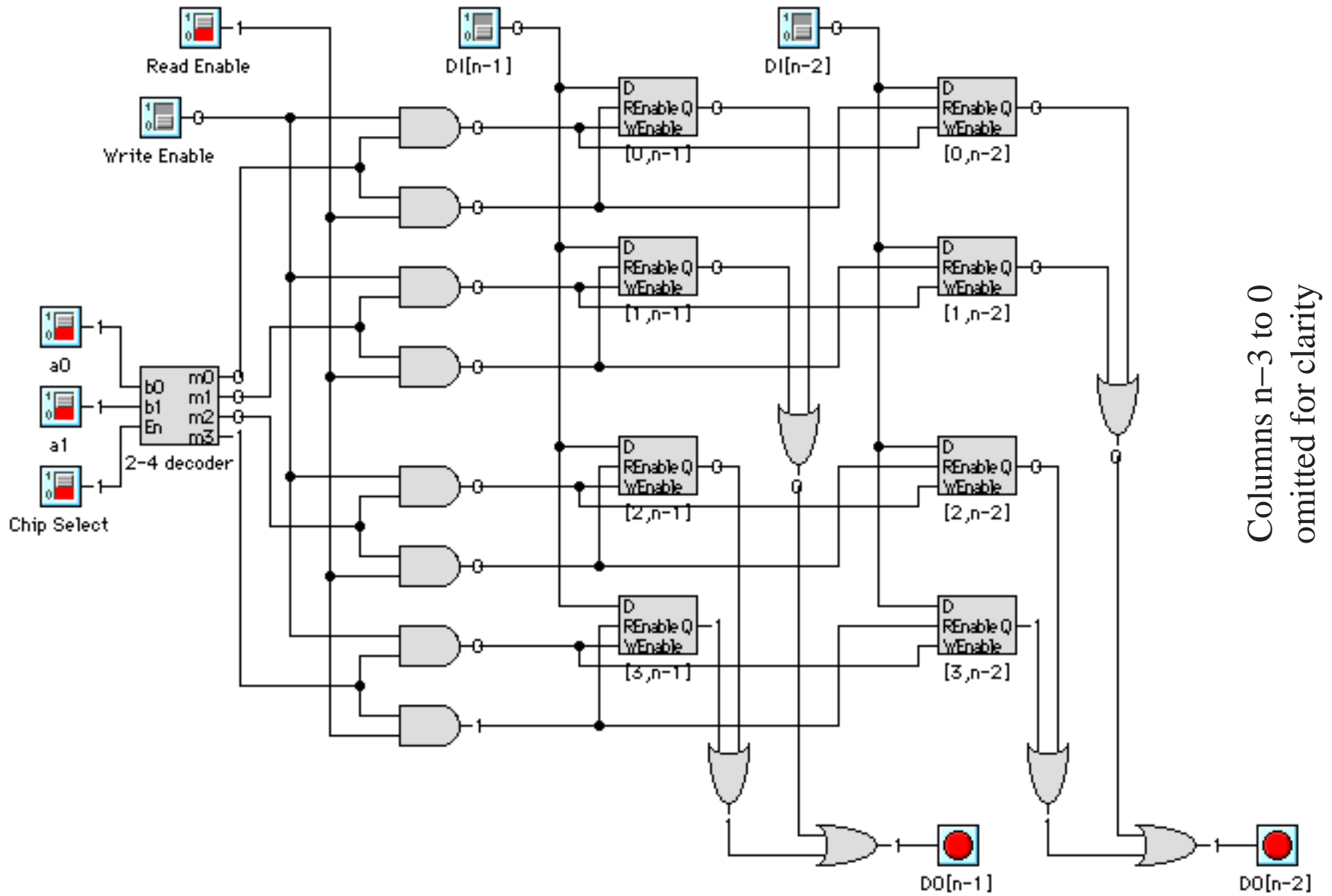
**RANDOM-ACCESS MEMORY (4)**

- Two-dimensional array of memory cells
  - ▷ Implement as a 1-dimensional array of multi-cell words
  - ▷ Each cell contains a fixed number of bits
  - ▷ Separate decoders select the row and column of the cell that is addressed
  - ▷ When a word line is enabled, data appears on all of the bit lines



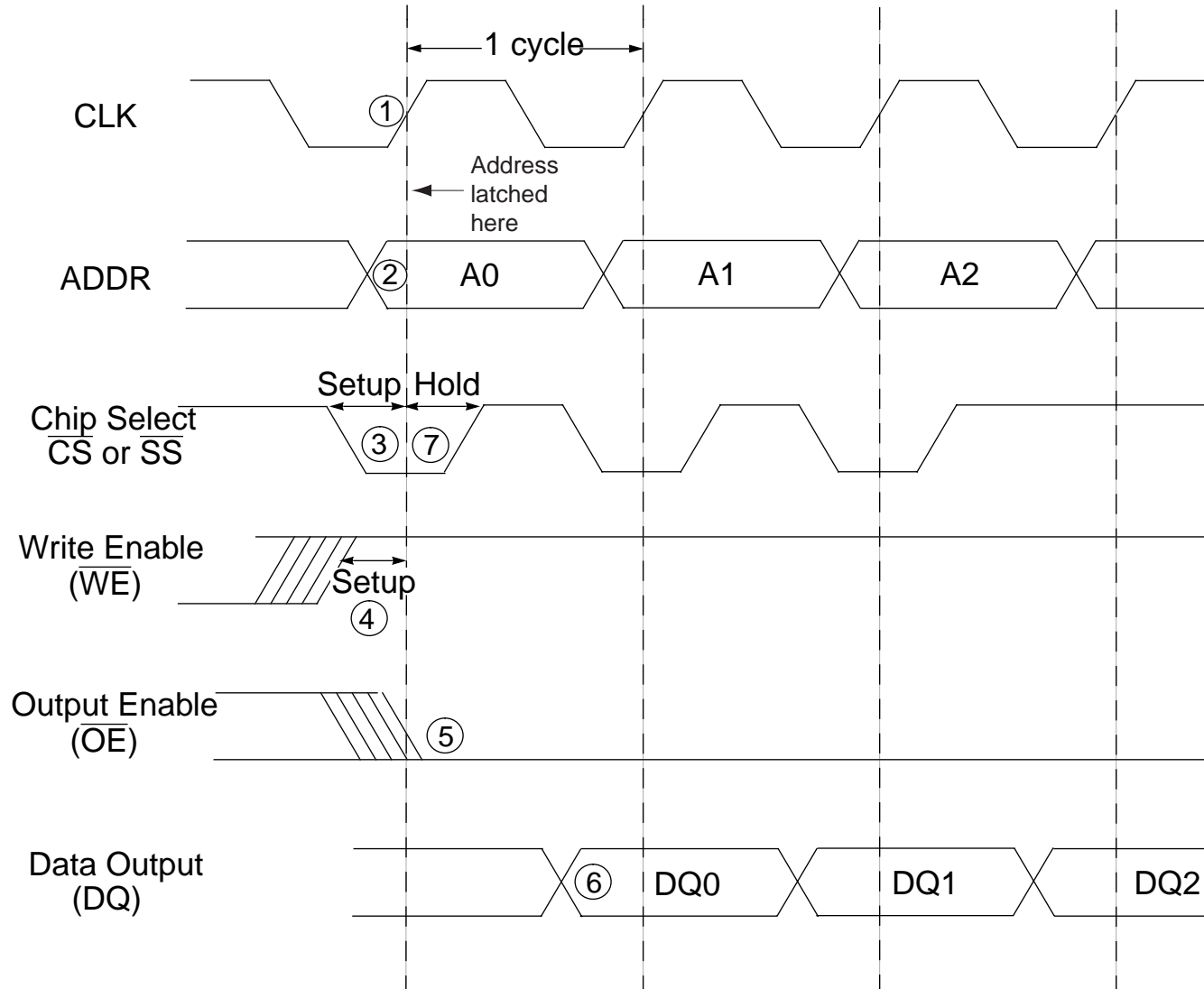
**RANDOM-ACCESS MEMORY (5)**

- On a memory IC that stores megabits of data, there are thousands of row and column select lines
  - ▷ Too many select lines to pin out
  - ▷ There are too many outputs for a multiplexor of practical size and delay
  - ▷ Solution: Use separate decoders for the rows and columns
  - ▷ Assume a chip with  $M$  rows and  $N$  columns
    - The  $M$  least significant bits of the main memory address determine the row
    - The next  $N$  bits determine the column



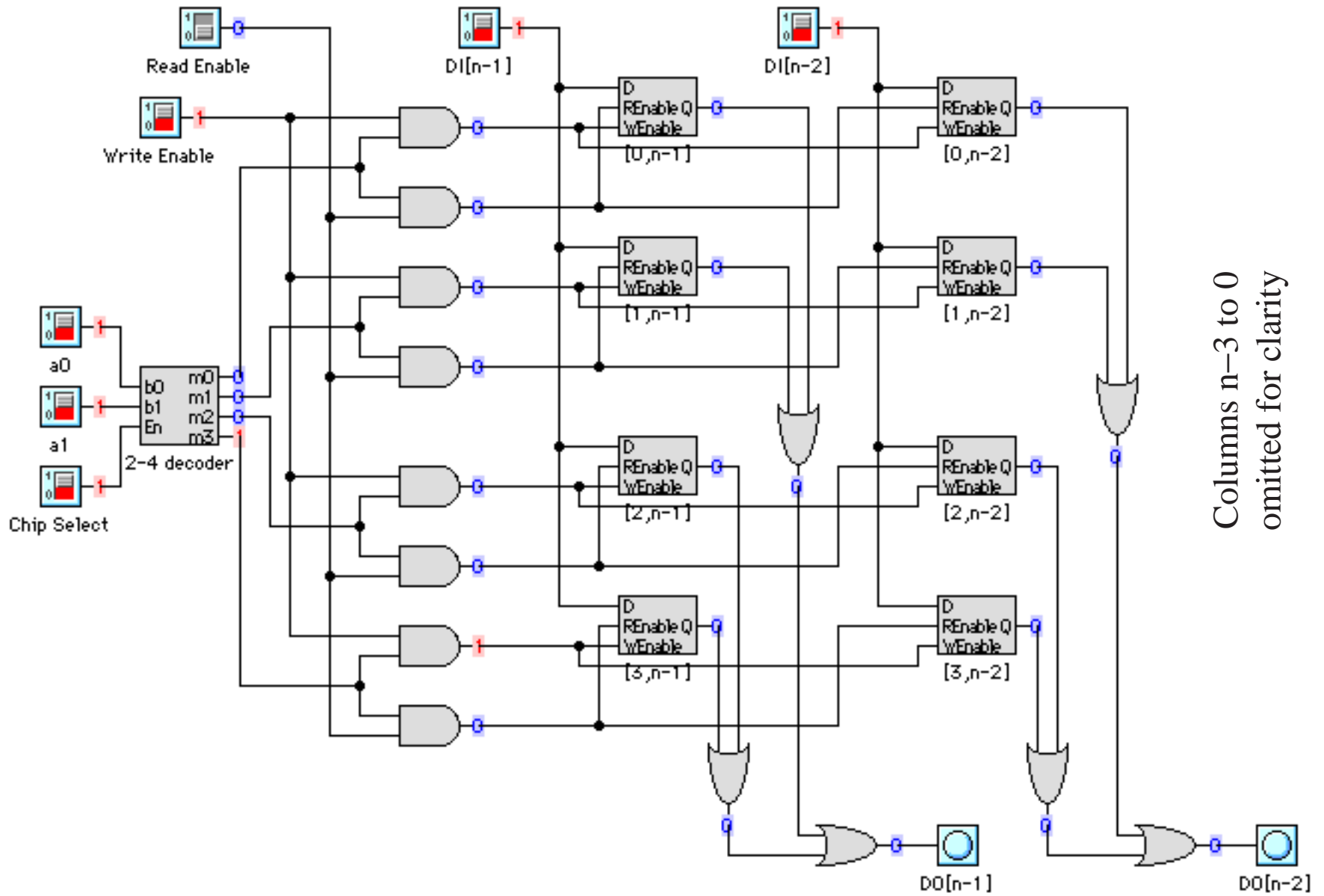
$4 \times n$  SRAM (read cycle)

# Reading from SRAM



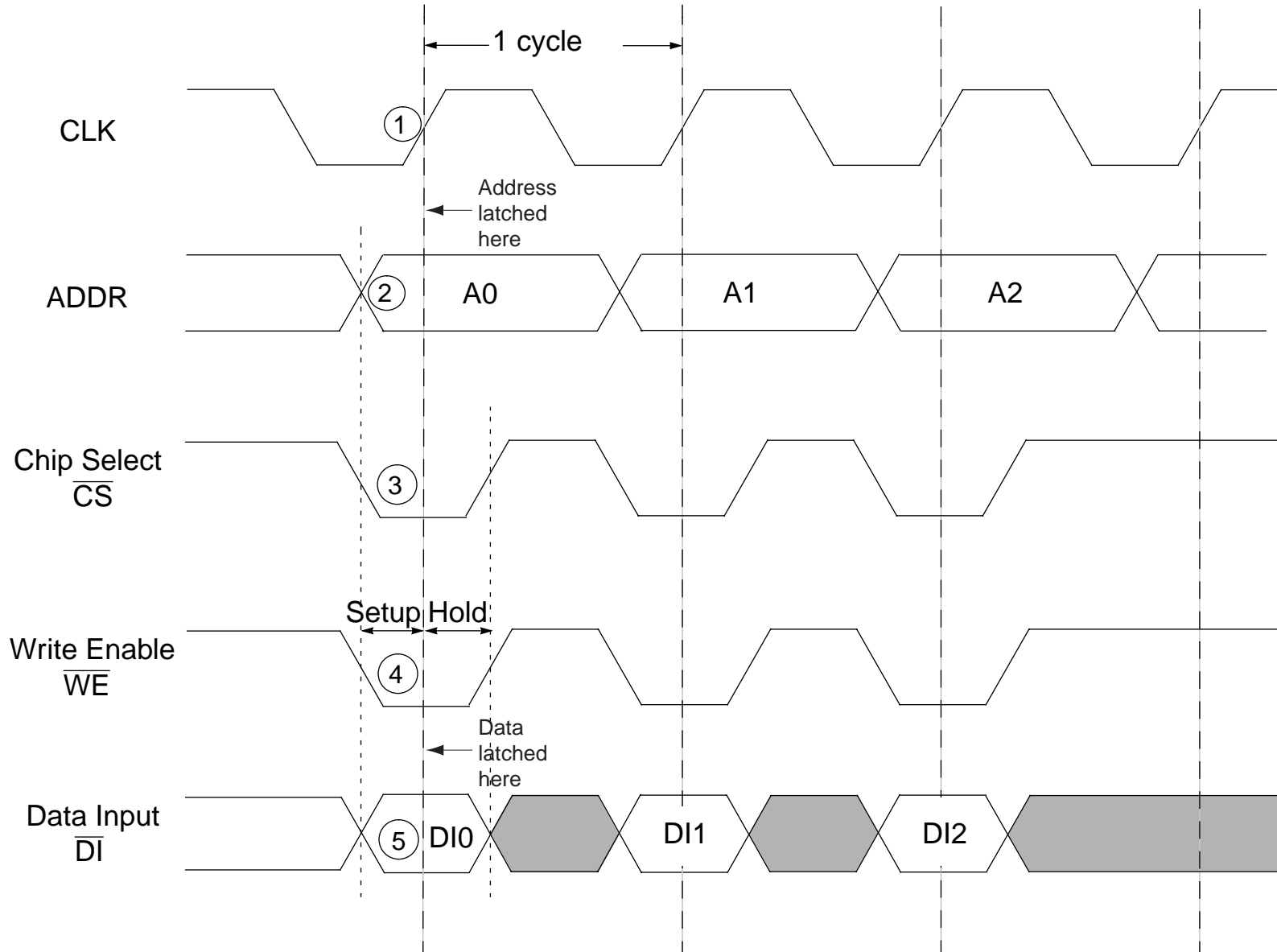
**Note:** DQ0 is the data associated with Address 0 (A0). DQ1 is the data associated with Address 1 (A1).





4 × n SRAM (write cycle)

# Writing to SRAM



**Note:** DI0 is written to the memory location with address 0 (A0).

**RANDOM-ACCESS MEMORY (6)**

## • SRAM memory system design

$$\triangleright \text{No. of rows of ICs} = \frac{\text{No. of words in memory system}}{\text{No. of words per IC}}$$

$$\triangleright \text{No. of columns of ICs} = \frac{\text{No. of bits per word in memory system}}{\text{No. of bits per word per IC}}$$

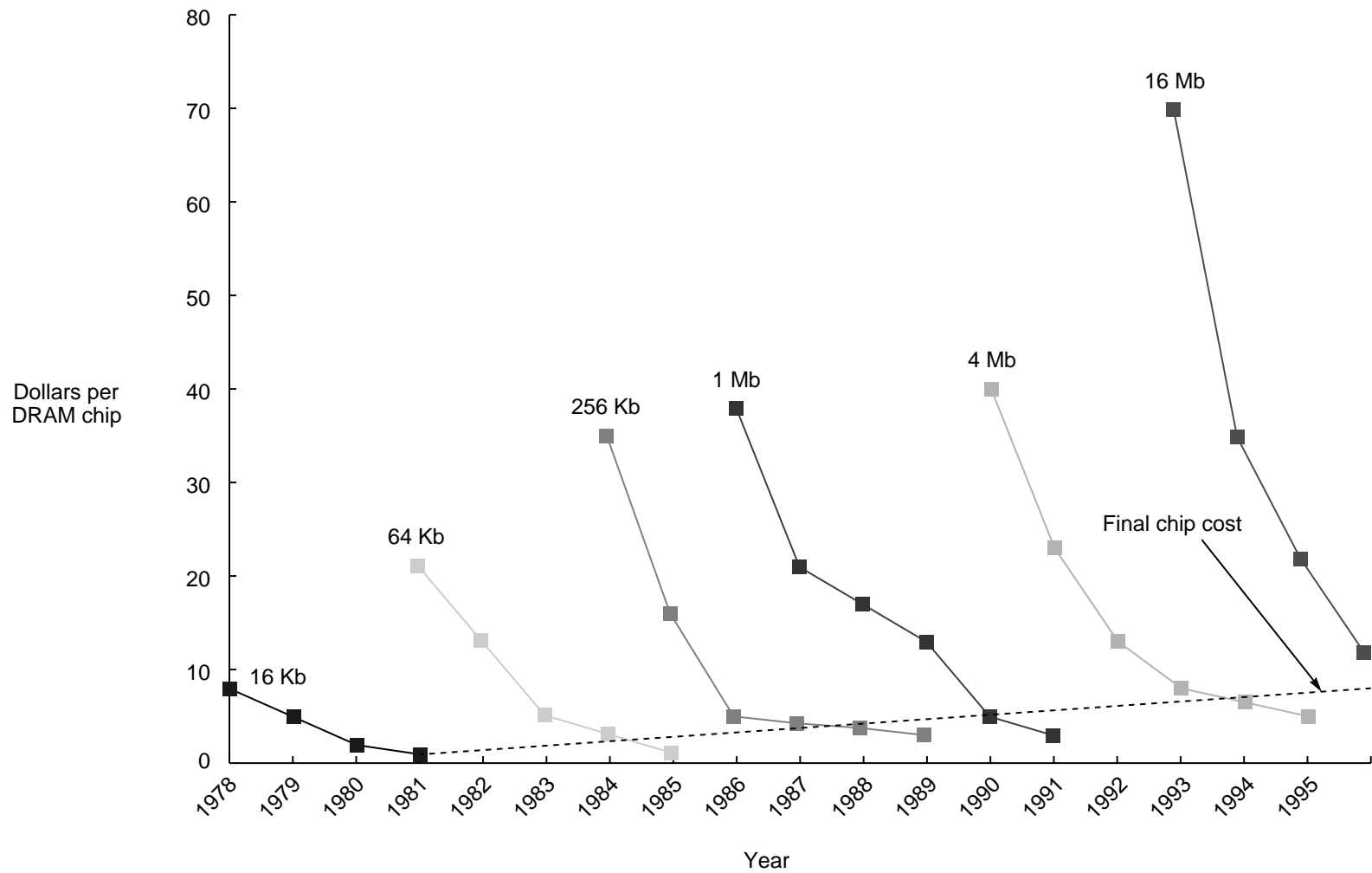
 $\triangleright$  Addressing:

- No. of rows of ICs determines the no. of address bits into the decoder

• Example: A  $4\text{M} \times 32$  bit ( $4\,194\,304 \times 32$  bit) SRAM memory system built from  $2\text{M} \times 8$  bit SRAM ICs

$$\triangleright \text{No. of rows of ICs} = \frac{4}{2} = 2 \Rightarrow 1\text{-to-}2 \text{ decoder}$$

$$\triangleright \text{No. of columns of ICs} = \frac{32}{8} = 4$$



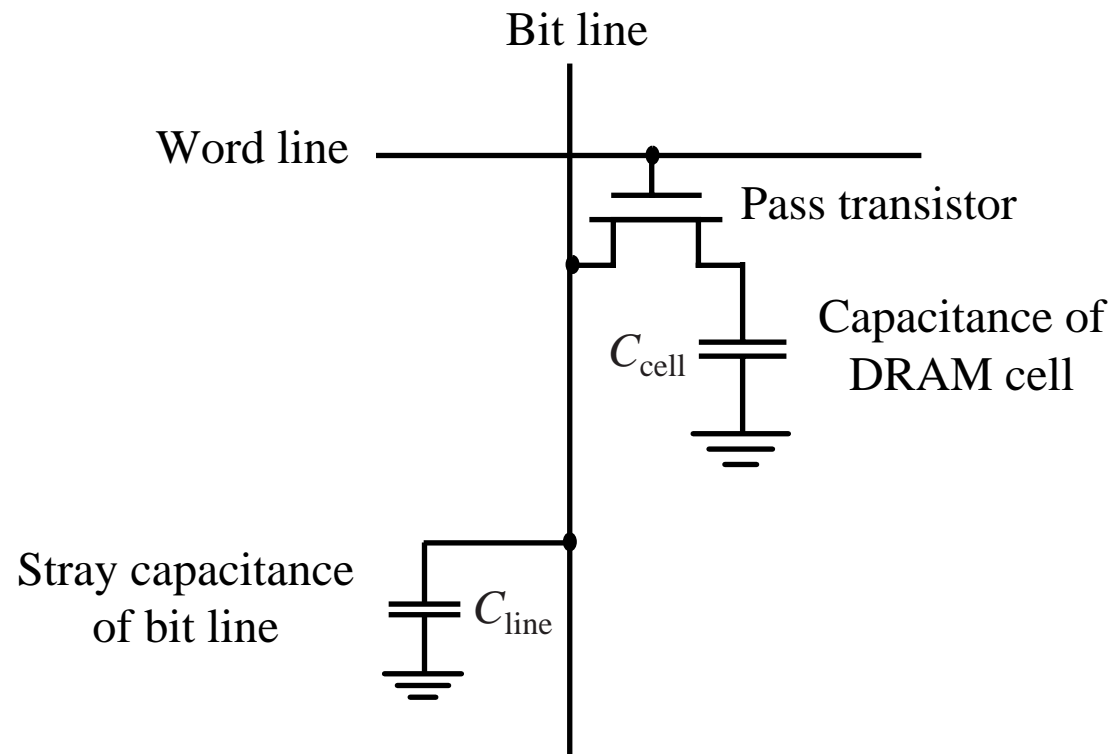
**RANDOM-ACCESS MEMORY (7)**

- DRAM memory cell design
  - ▷ MOS technology
  - ▷ At the intersection of each bit line and each word line, a capacitor  $C_{\text{cell}}$  is the memory unit; a transistor functions as a switch
    - Word line asserted: Switch open  $\Rightarrow C_{\text{cell}}$  connected to bit line
    - Write operation: Value to be written (0 or 1) asserted on bit line
      - ◊ Value = 1 (0)  $\Rightarrow C_{\text{cell}}$  charged (discharged)
    - Read operation: Bit line precharged to halfway voltage,  $V_{\text{pre}}$ 
      - ◊ Word line activated  $\Rightarrow C_{\text{cell}}$  discharges onto (charges from) line
      - ◊ Voltage change

$$\Delta V = (V_{\text{bit}} - V_{\text{pre}}) \frac{C_{\text{cell}}}{C_{\text{cell}} + C_{\text{line}}} \approx 250 \text{ mV}$$

detected by the sense amplifier for the bit line

# 1-TRANSISTOR DRAM CELL



**RANDOM-ACCESS MEMORY (8)**

## • DRAM memory cell physics

- ▷ Charge stored on a 2-conductor capacitor with potential difference  $V$  between conductors:

$$Q = CV$$

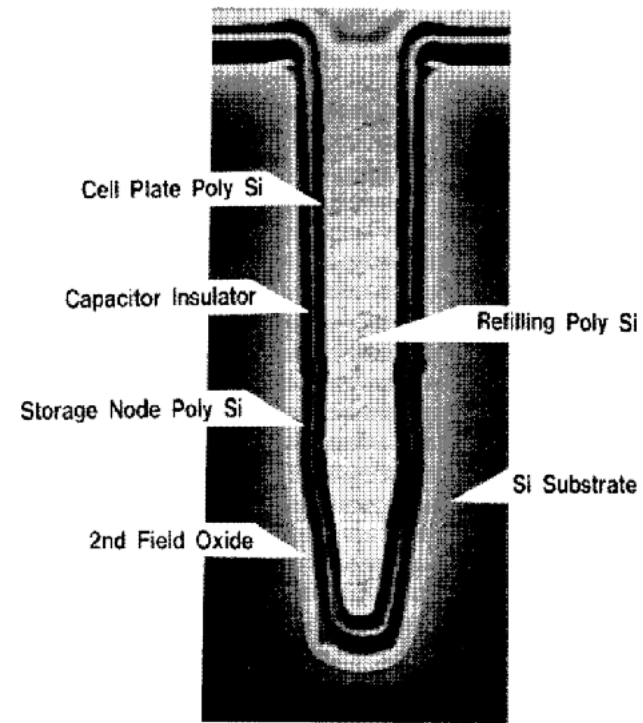
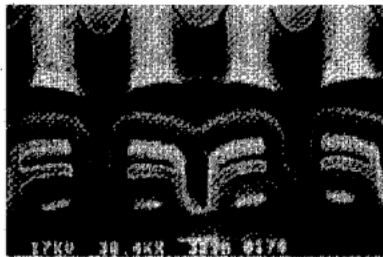
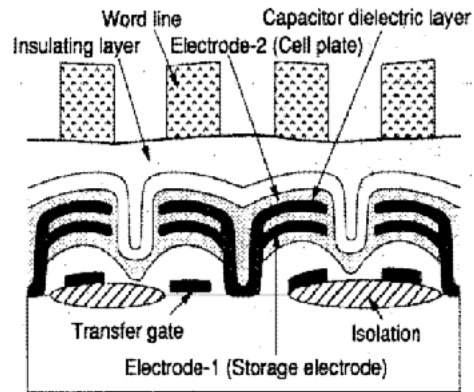
- ▷ Capacitance of two parallel plates with area  $A$  and separation  $d \ll A$ :

$$C = \frac{\epsilon A}{d}$$

- ▷ Conclusions:

- To make  $Q$  as large as possible for a given  $V$ , one must increase  $A$  and  $\epsilon$  while decreasing  $d$ 
  - ◊  $\epsilon = \epsilon_0 \epsilon_r$ , where  $\epsilon_r$  is a material property (hard to change!)
  - ◊ The area  $A$  can be increased without increasing the footprint of a memory cell by using trench or stack geometries

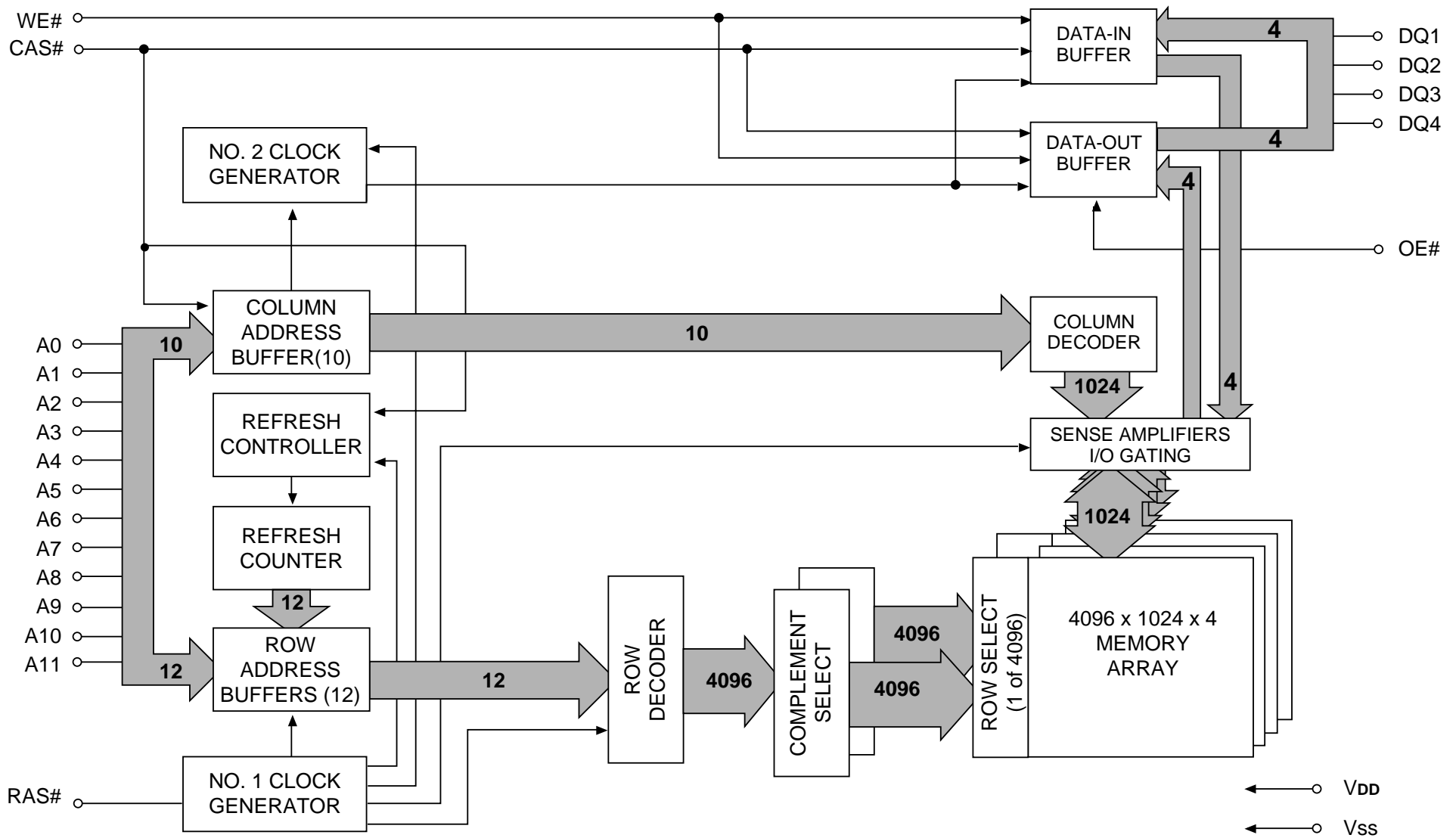
# STACK AND TRENCH DRAM CELL GEOMETRIES



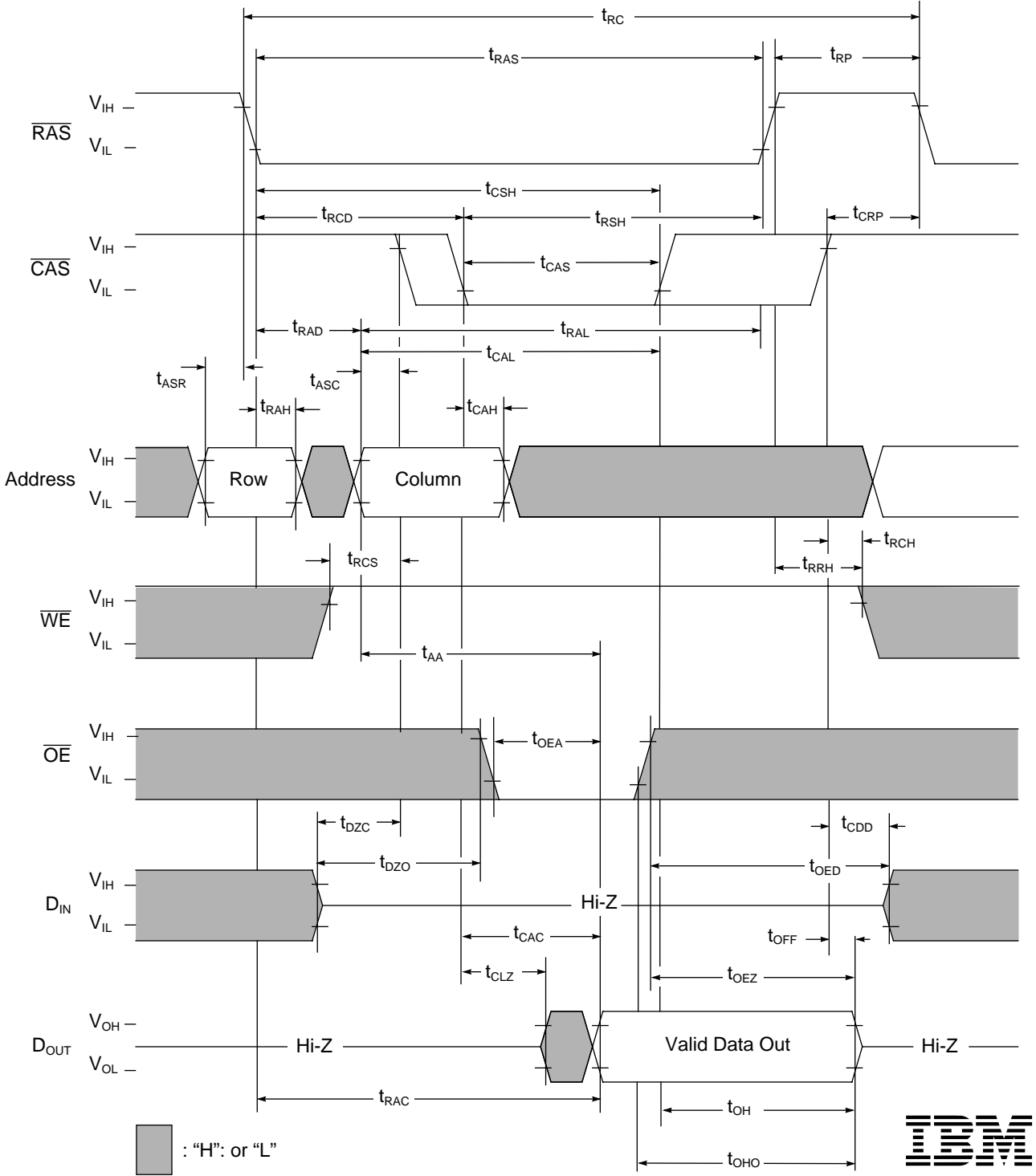
**RANDOM-ACCESS MEMORY (9)**

- DRAM IC design
  - ▷ Hierarchical organization (an array of arrays of words, etc.)
  - ▷ Tristate outputs permit construction of multi-IC memory systems
  - ▷ Data in/data out pins are shared in order to reduce pin count
    - Access mode (read or write) is selected by **Output Enable** and **Write Enable** signals
  - ▷ Row and column address pins are also shared
    - First the **Row Address Strobe (RAS)** samples the row address
    - The row address is latched
    - Then the **Column Address Strobe (CAS)** samples the column address
  - ▷ Shared I/O and address lines, plus refresh cycle
    - ⇒ DRAM initial latency  $\approx 5\text{--}10 \times$  SRAM latency

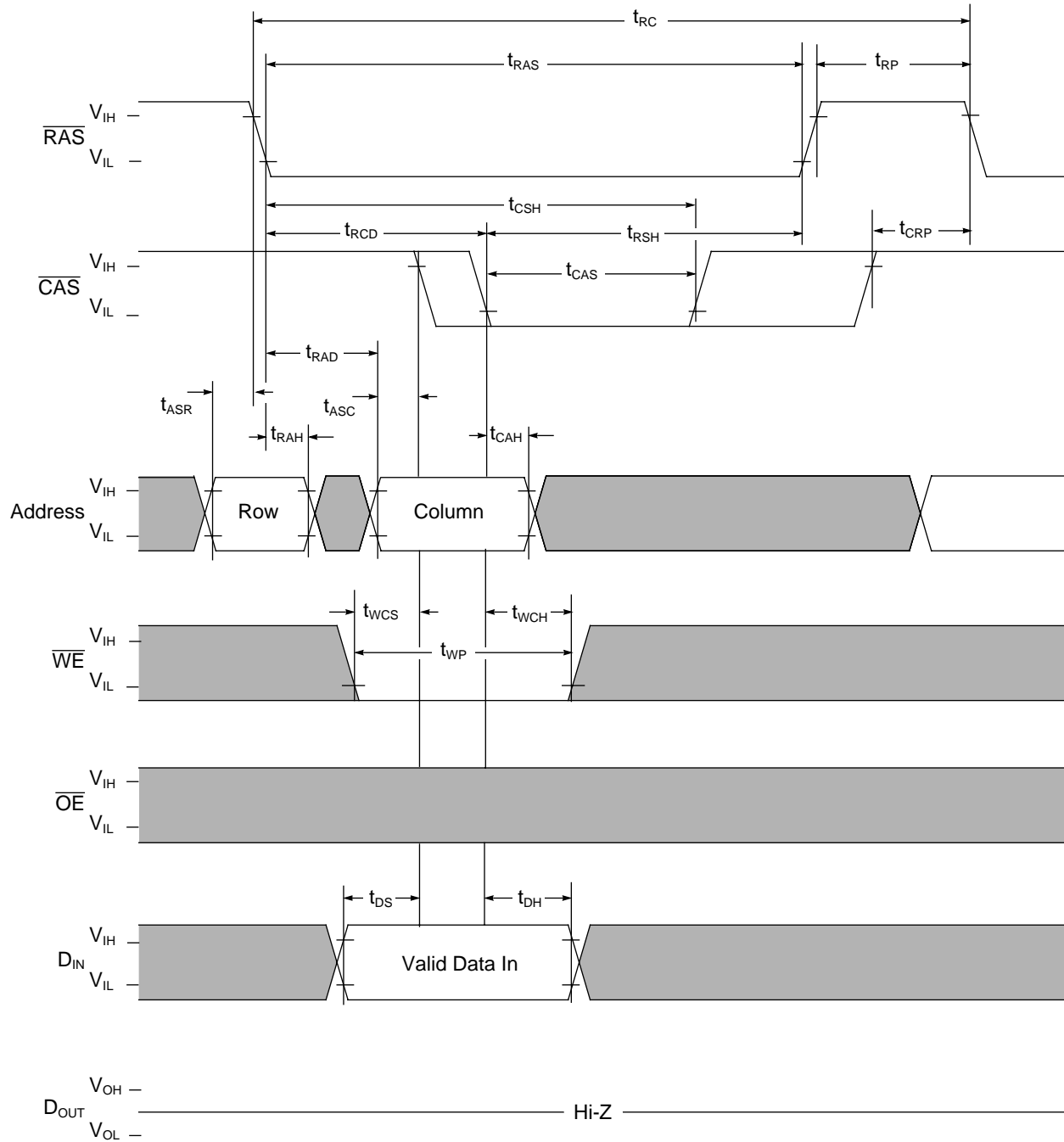
# FUNCTIONAL BLOCK DIAGRAM - 4K REFRESH



# DRAM Read Cycle



# DRAM Write Cycle (Early Write)

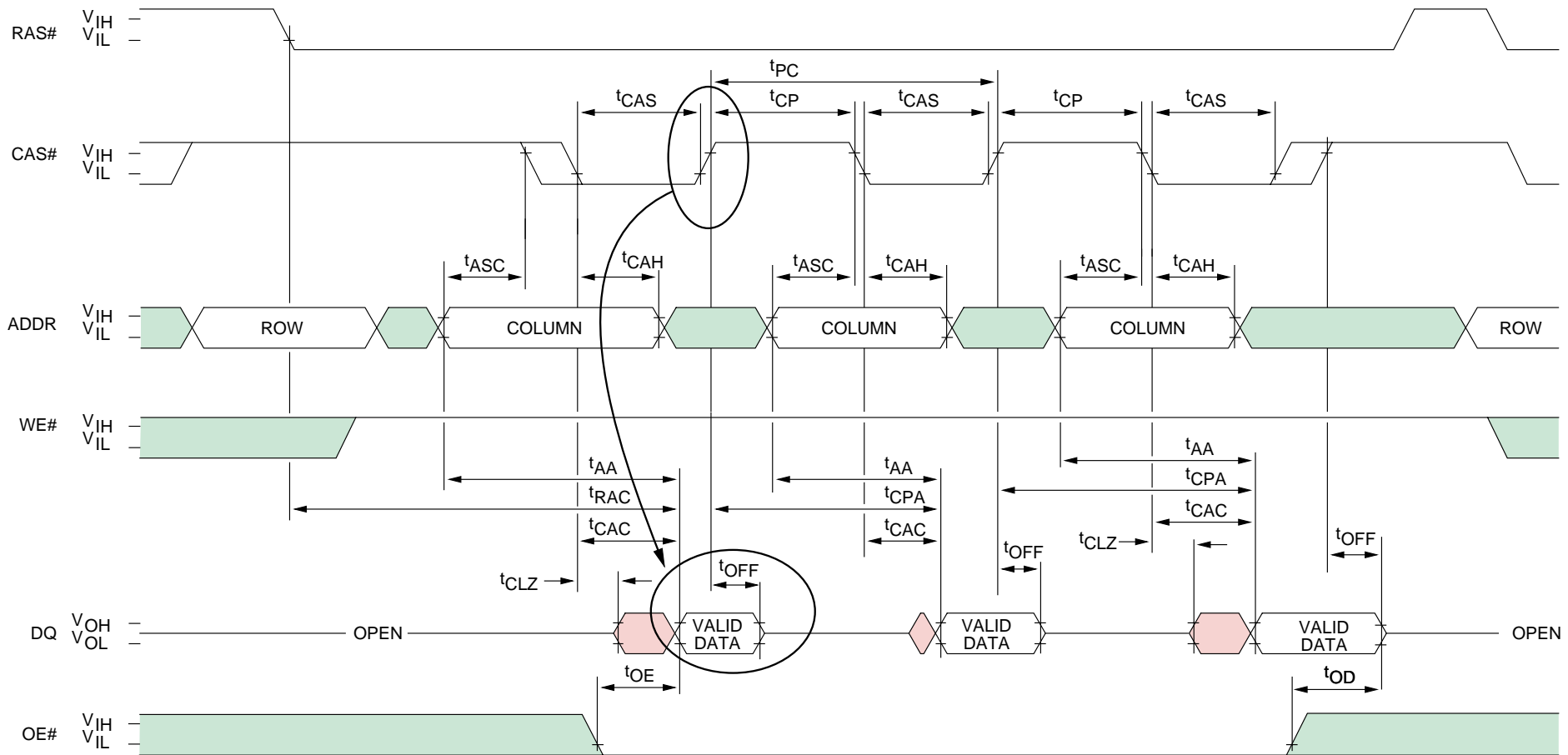


■ : "H" or "L"



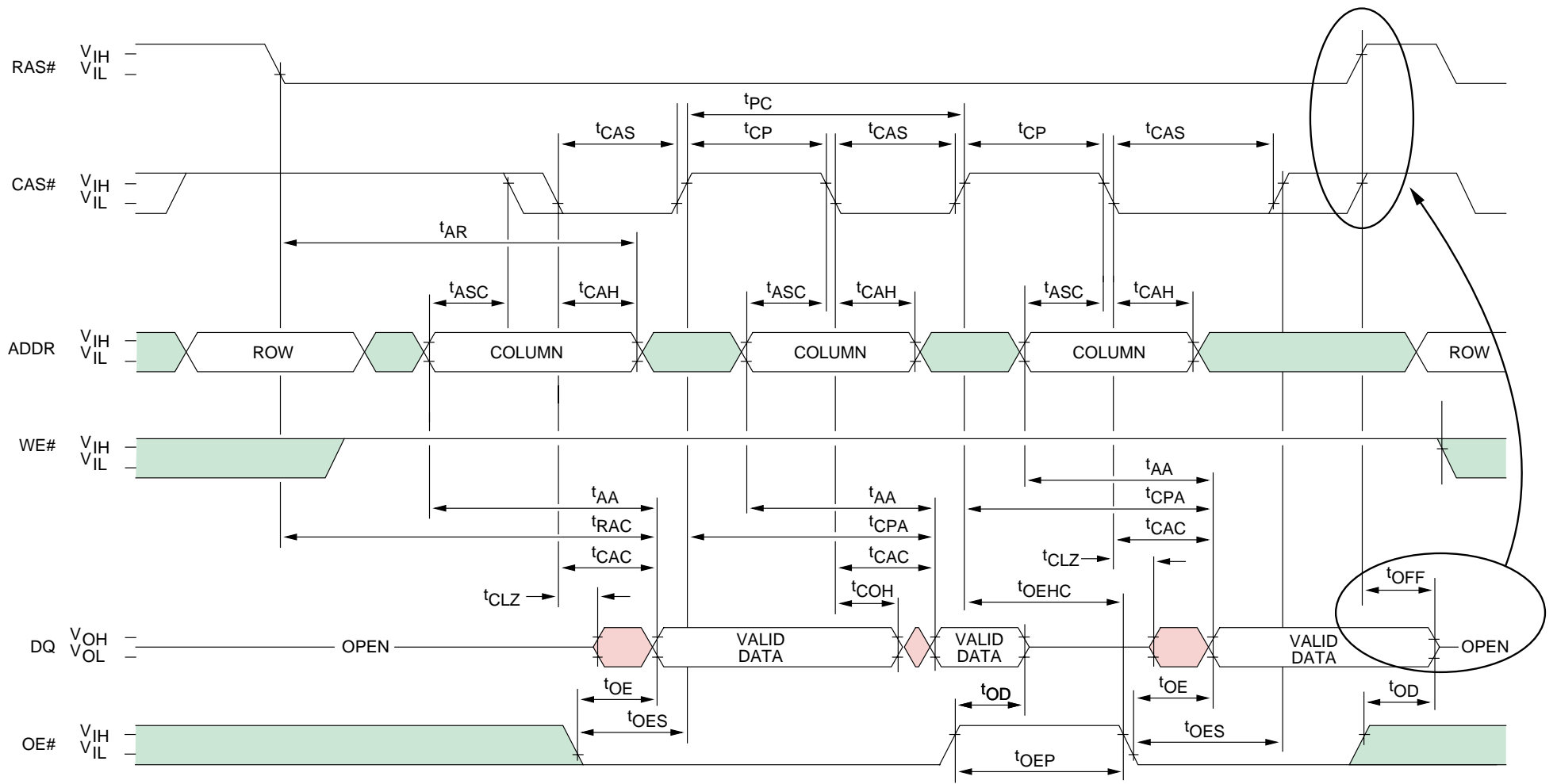
**RANDOM-ACCESS MEMORY (10)**

- Classic DRAM access mode (RAS and CAS strobed for each bit)
  - ▷ Row access time =  $t_{\text{RAC}} \approx 60 \text{ ns}$
  - ▷ Cycle time =  $t_{\text{RC}} \approx 2t_{\text{RAC}}$  = time between successive accesses
- DRAM burst modes (typically one burst = 4 contiguous accesses)
  - ▷ Fast page mode (FPM) DRAM IC has a register to hold a row
    - RAS needs to be strobed only once
    - CAS must be strobed for each cell accessed (see next slide)
  - ▷ Extended data out (EDO) is basically a more useful FPM
  - ▷ Data is not disabled when  $\overline{\text{CAS}}$  goes high (see next slide)
    - Valid data remains available until data from next access appears
  - ▷ Pipelined burst mode
    - Memory is interleaved using banks on the DRAM IC
    - The DRAM IC generates addresses to be fetched
    - Eliminates cycle time



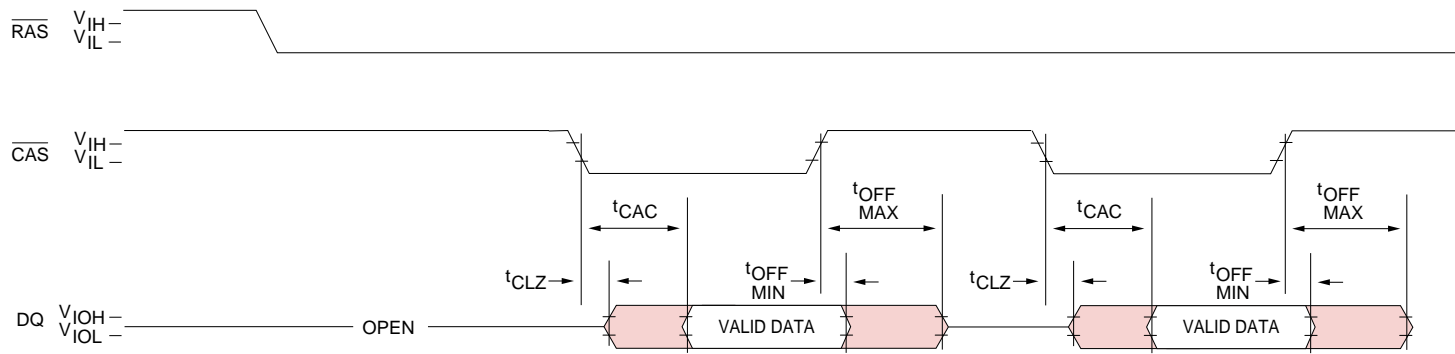
## FPM DRAM PAGE MODE CYCLE

- DON'T CARE
- UNDEFINED

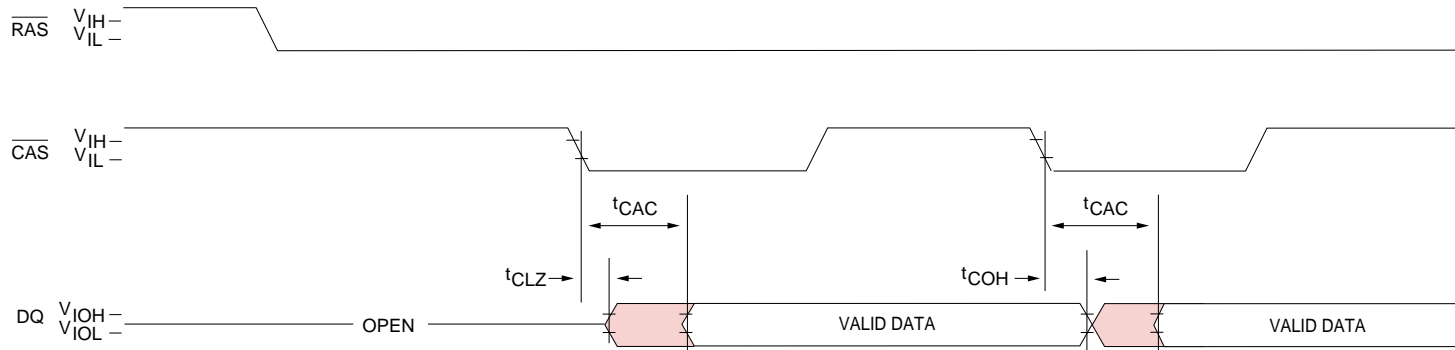


## EDO DRAM PAGE MODE CYCLE

- DON'T CARE
- UNDEFINED



**(a) Conventional FPM**

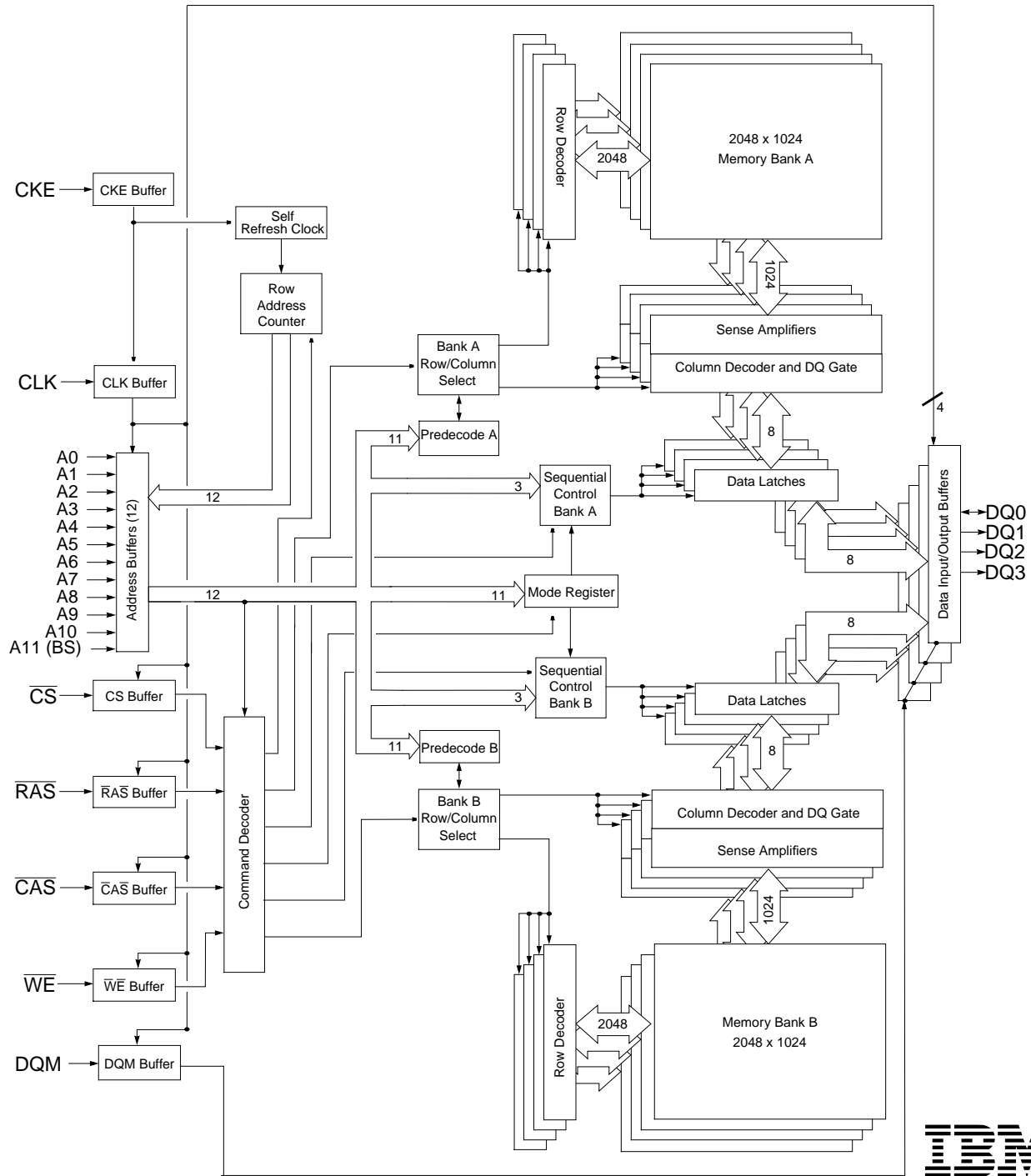


**(b) EDO**

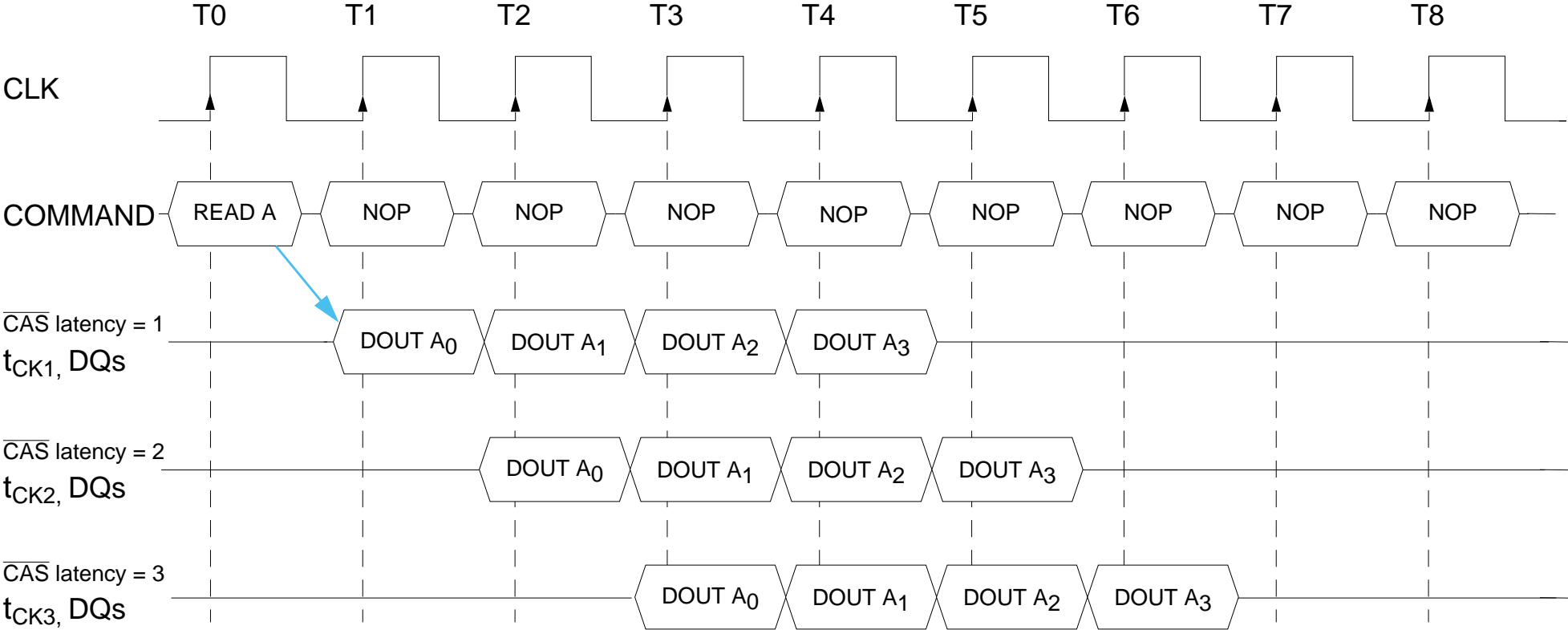
 UNDEFINED

## PHYSICAL DIFFERENCE BETWEEN FPM AND EDO

# SDRAM block diagram (2Mbit x 4 I/O x 2 banks)



# Burst read operation



**RANDOM-ACCESS MEMORY (11)**

- Because 1-transistor DRAMs store bits as charges on capacitors, the charges leak away and therefore must be refreshed periodically
- DRAM refresh cycles:
  - ▷ The Joint Electronics Design Engineering Council (JEDEC) has approved two refresh types for  $4M \times N$  DRAMs
    - One version of a  $4M \times 4$  DRAM requires 10 column-address bits and 12 row-address bits
      - ◇  $2^{12} = 4096$  (4K) rows are refreshed every 64 ms
    - The other version for the same capacity DRAM requires 11 row-address bits and 11 column-address bits
      - ◇  $2^{11} = 2048$  (2K) rows are refreshed every 32 ms
      - ◇ 4K refresh chips draw less current per refresh than 2K refresh chips, because the page depth (number of columns) is  $2^{10} = 1024$  in the 4K refresh chips vs.  $2^{11} = 2048$  in the 2K refresh chips

**SUPERCOMPUTER MEMORY SIZES**

Computer	Date	Memory Size (MB)	Technology
CDC 6600	1964	0.75	ferrite cores
CDC 7600	1969	0.25	ferrite cores
CRAY-1	1976	4	ECL
CRAY X-MP/48	1984	64	ECL
CRAY Y-MP/864	1988	512	ECL
CRAY Y-MP C916	1993	8192	BiCMOS