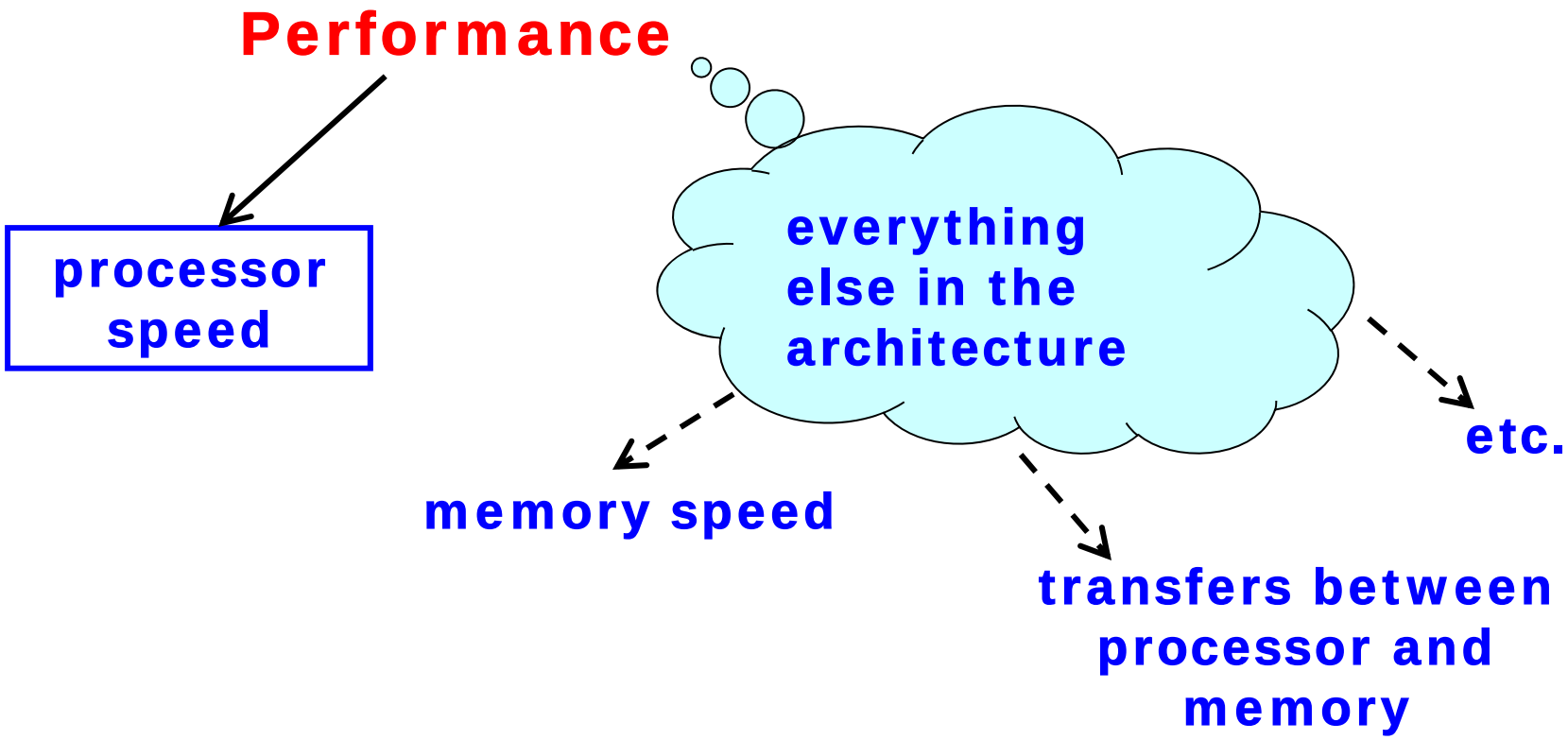
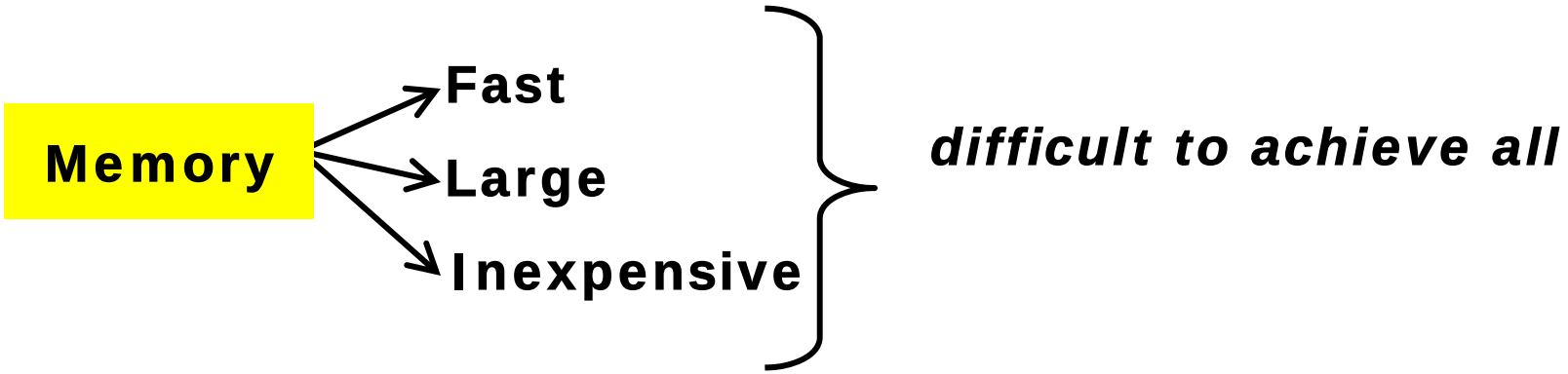
20 Memory
CSC 230

Department of Computer Science
University of Victoria

M&H: 7.1, 7.2, 7.3 (7.4 optional)
Stallings: 4.1, 5.1 (not all details), 5.3 (information only)

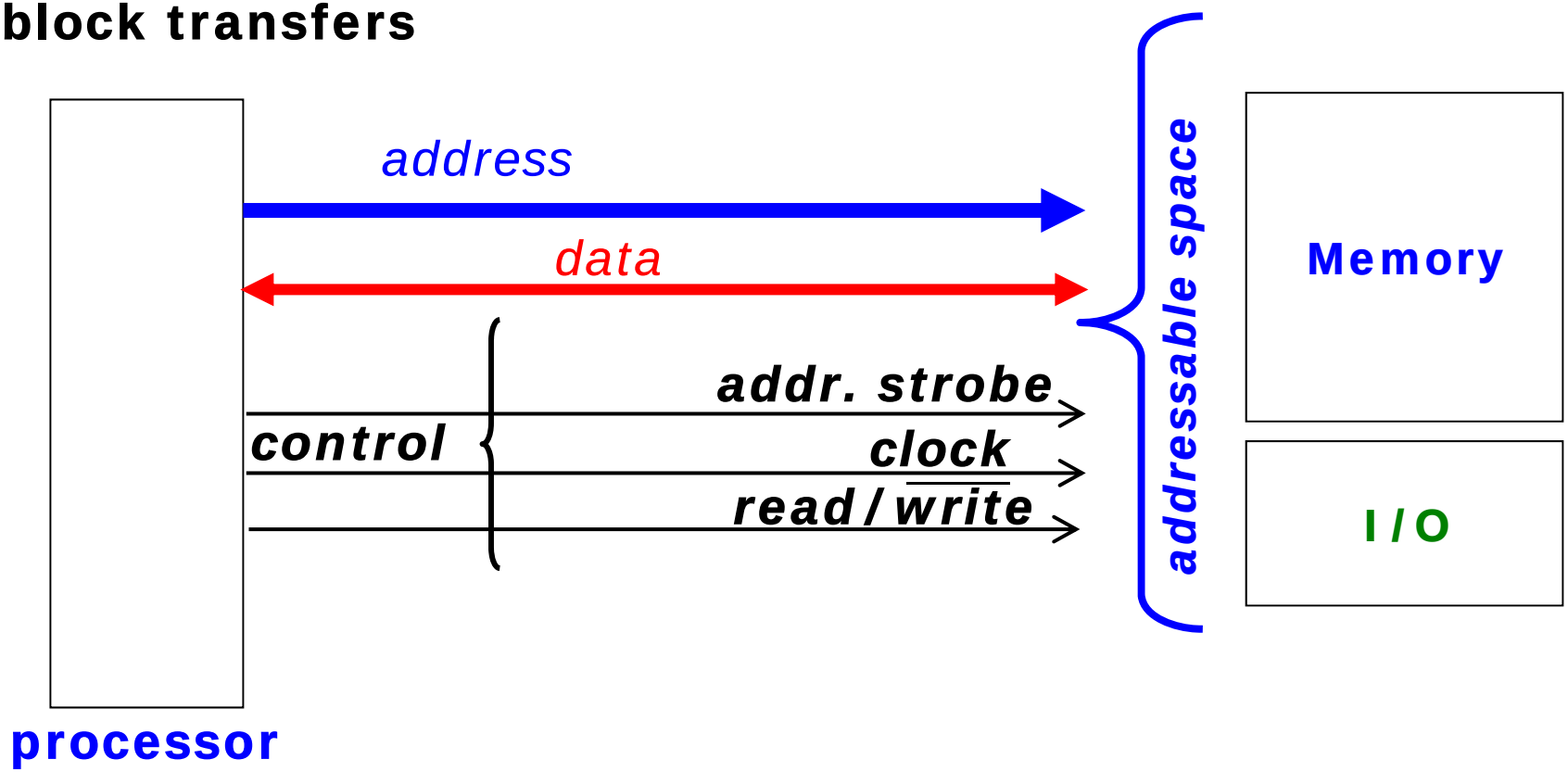

Performance
processor speed
everything else in the architecture
memory speed
transfers between processor and memory
etc.


Memory
Fast
Large
Inexpensive
difficult to achieve all

maximum sizes (some items from midterm questions)

- n-bit address bus ➔ 2^n addressable locations
- byte-addressable versus word-addressable
- little endian (Intel) versus big endian [ARM both]
- implementation is difficult
- block transfers

review



A lot of definitions (1)

- **read operation:** the transfer of information *FROM memory – also "fetch" or "get"*
- **write operation:** the transfer of information *TO* memory –also "put"
- **memory access time:** from the beginning to the end of a read or write operation
- **memory cycle time:** delay *between* the start of 2 operations ➔(can be a bottleneck)

  also: a measure of the time required for a read or write operation plus any time required before the memory is ready for the next operation

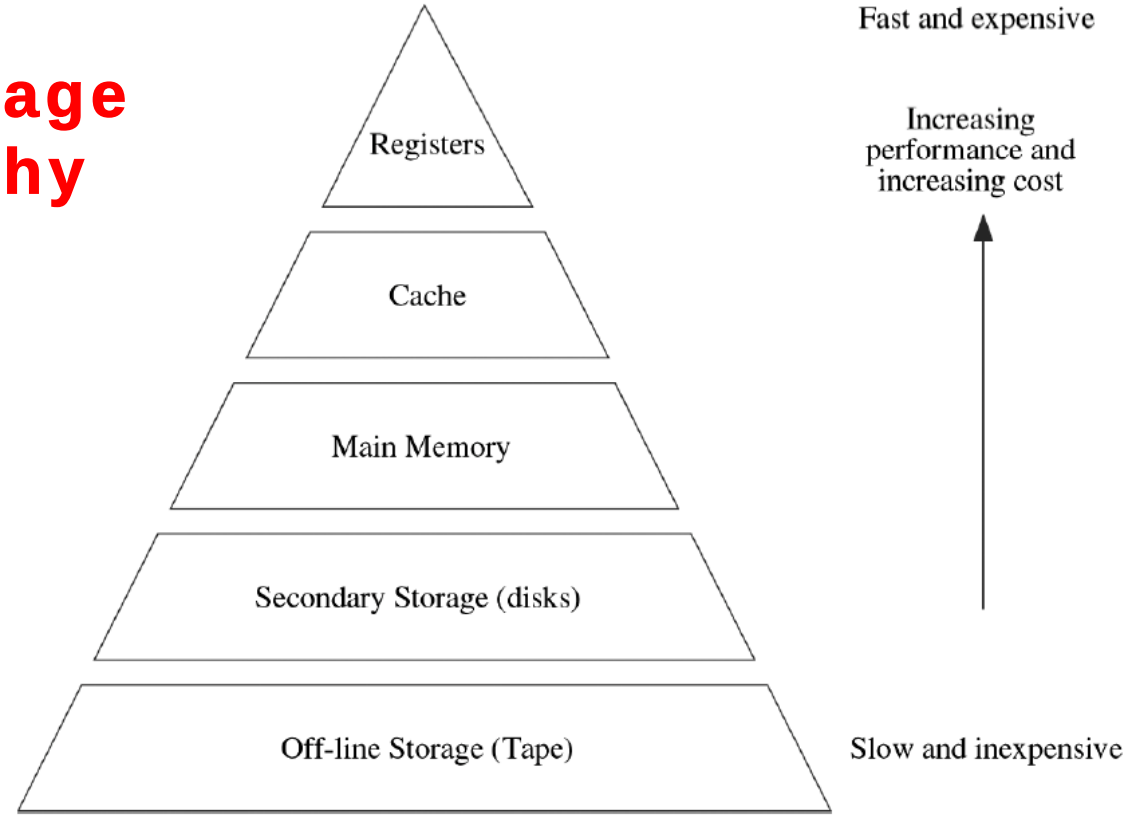
A lot of definitions (2)

- **volatile memory:** loses its contents when power is turned off
- **nonvolatile memory:** keeps its contents when power is turned off
- **capacity:** the total amount of information that can be stored in a storage device (e.g. memory)
- **density:** measure of information stored per unit of chip space

A lot of definitions (3)

- ***random access memory (RAM):*** **memory chip where the actual location has little effect on how long it takes to access information (RAM)**
- ***Read-only memory (ROM):*** **contents are fixed (ROM), burned at manufacturing time and read-only**
- **ROM Applications:** **Firmware, Bootstrap memory, Data tables**
- ***static memory:*** **semiconductor memory; the contents remain as long as power is applied**
- ***dynamic memory:*** **semiconductor memory for which the contents have to be periodically refreshed, typically every 8-64 ms**

The Storage Hierarchy



Memory type	Access time	Cost/MB	Typical amount used	Typical cost
Registers	0.5 ns	High	2 KB	–
Cache	5–20 ns	$80	2 MB	$160
Main memory	40–80ns	$0.40	512 MB	$205
Disk memory	5 ms	$0.005	40 GB	$200

Static RAM

MAIN FEATURES:

- Retain state while power is applied
- Bits stored as on/off switches
- Usually about 6 CMOS transistors per bit
- No charges (capacitance) to leak
- No refreshing needed when powered unlike dynamic RAM (DRAM),
- More complex construction
- Larger per bit

FAST

EXPENSIVE → both in area and manufacturin

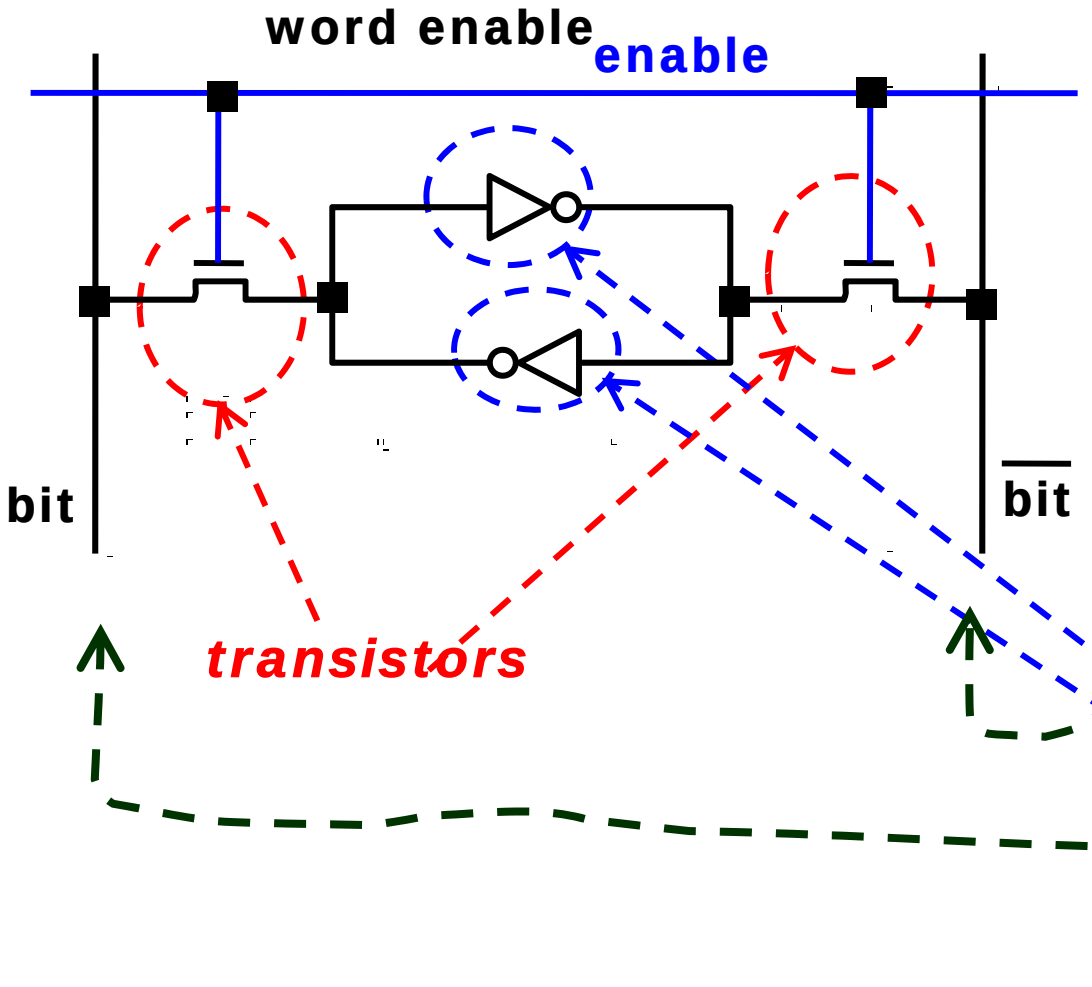
NOT VERY DENSE

used for Cache and registers

Capacitance is the ability of a body to hold an electrical charge.

Complementary metal–oxide–semiconductor (CMOS) is a technology for constructing integrated circuits.

STATIC RAM CELL

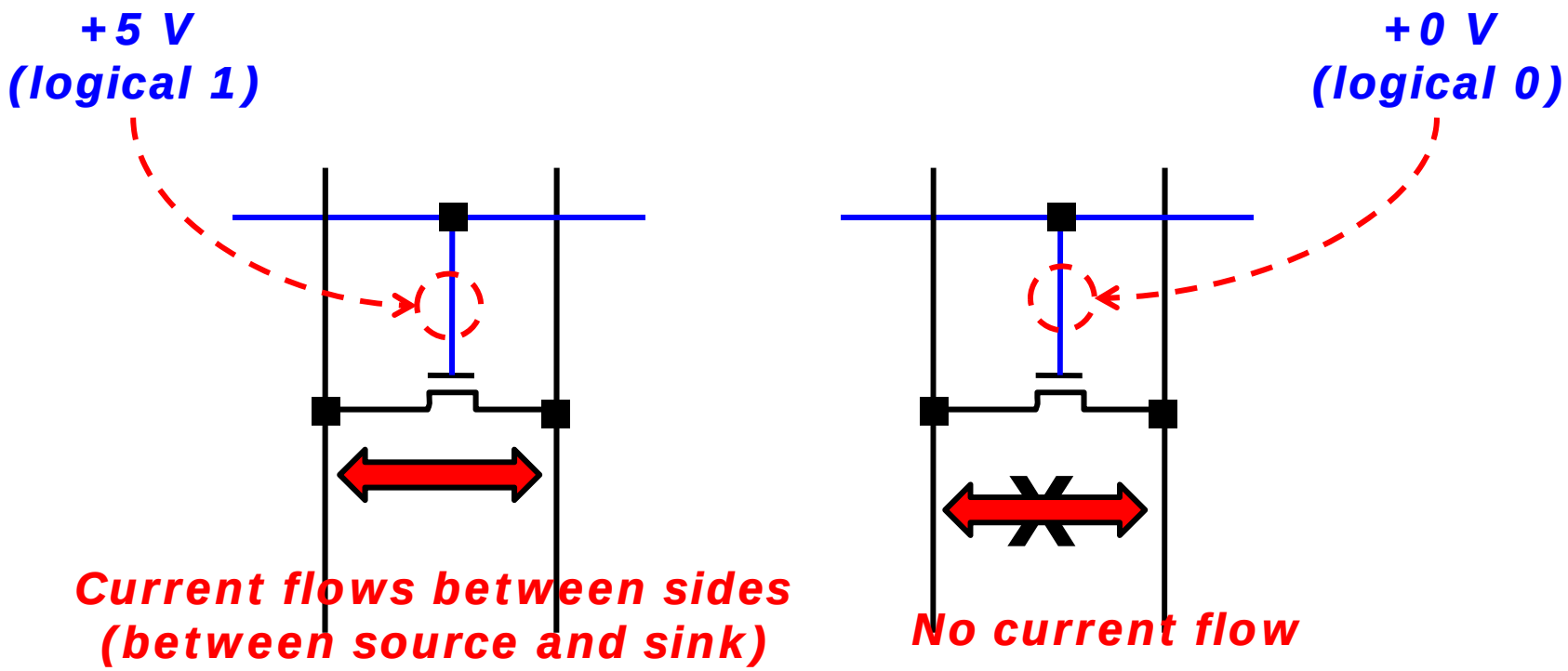


READ: sensors for "difference" in voltage between ***bit*** and ***$\overline{bit}$*** (sense amplifiers)

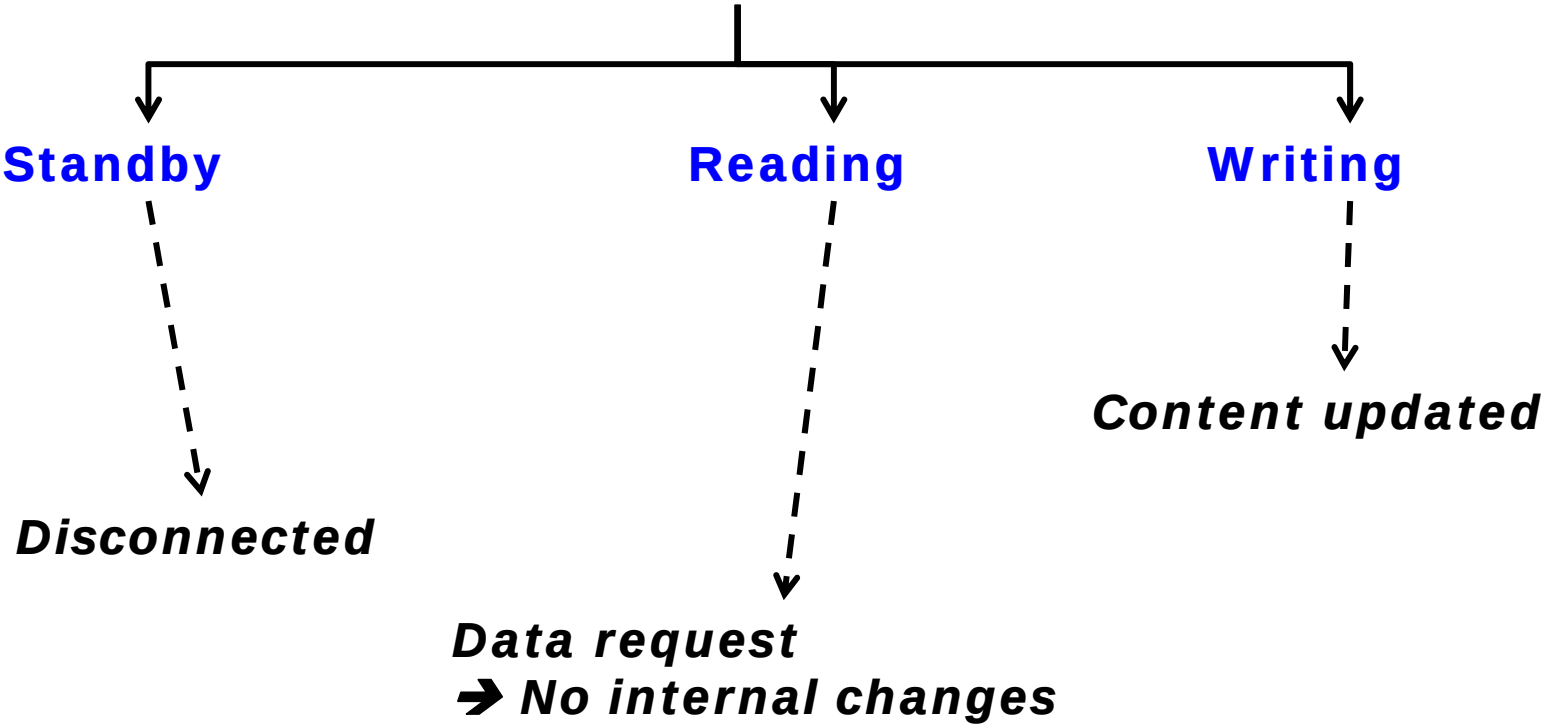
Transistor (n): when +5v ➔ current flows between sides

when +0v ➔ no current flows

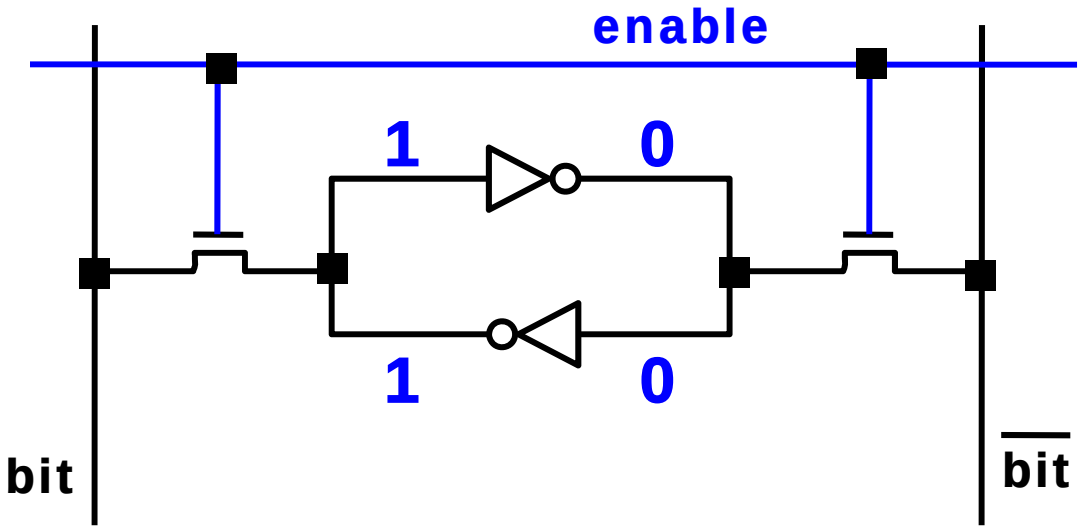
Transistor


+5 V
(logical 1)
+0 V
(logical 0)
Current flows between sides
(between source and sink)
No current flow

STATIC RAM CELL: modes



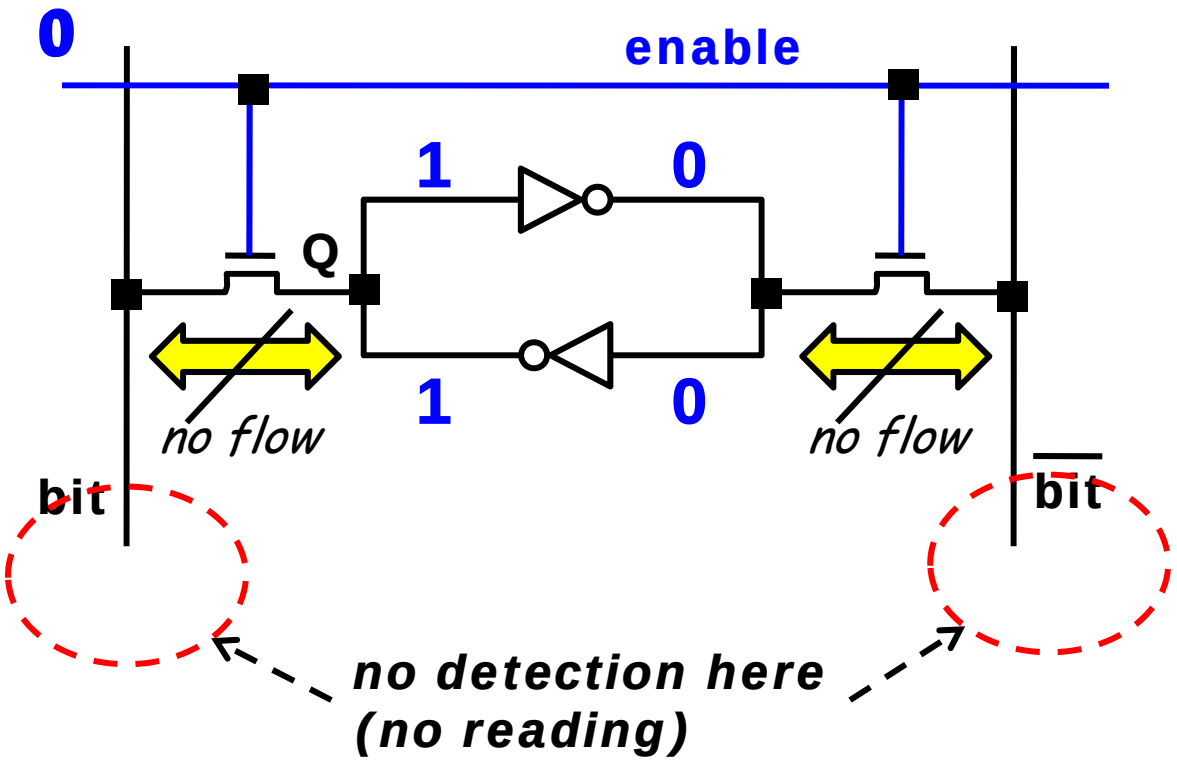
STATIC RAM CELL (example)



Note: $\overline{\text{bit}}$ is often written as bit' or as ¬bit

Assume that the content is already in there as shown

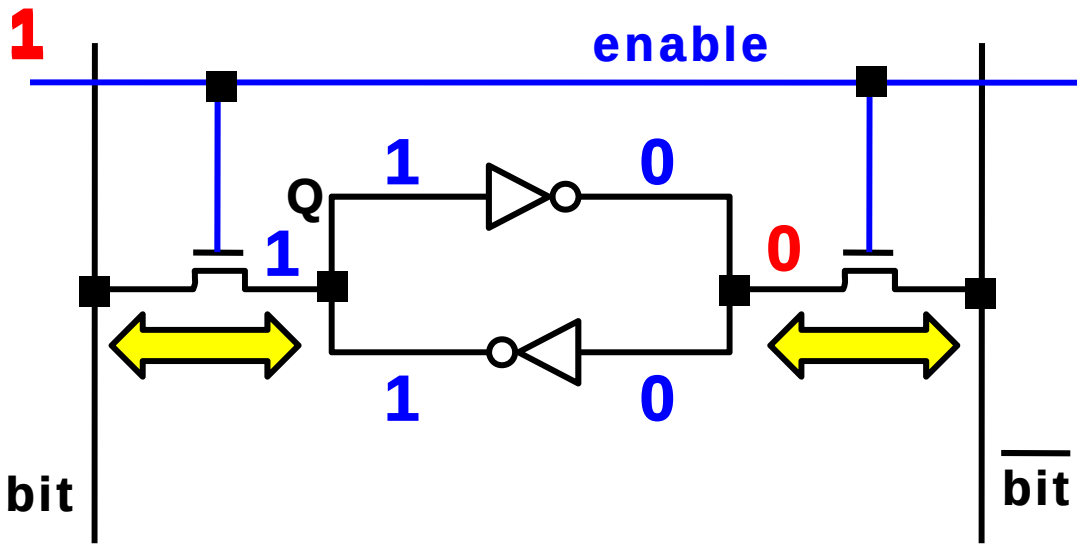
STATIC RAM CELL (Standby)



Assume that the content is already in there as shown

- ➔ As long as ENABLE = 0, state is maintained internally
- ➔ no change inside
- ➔ nothing detected outside by sense output
- ➔ as long as there is power connected

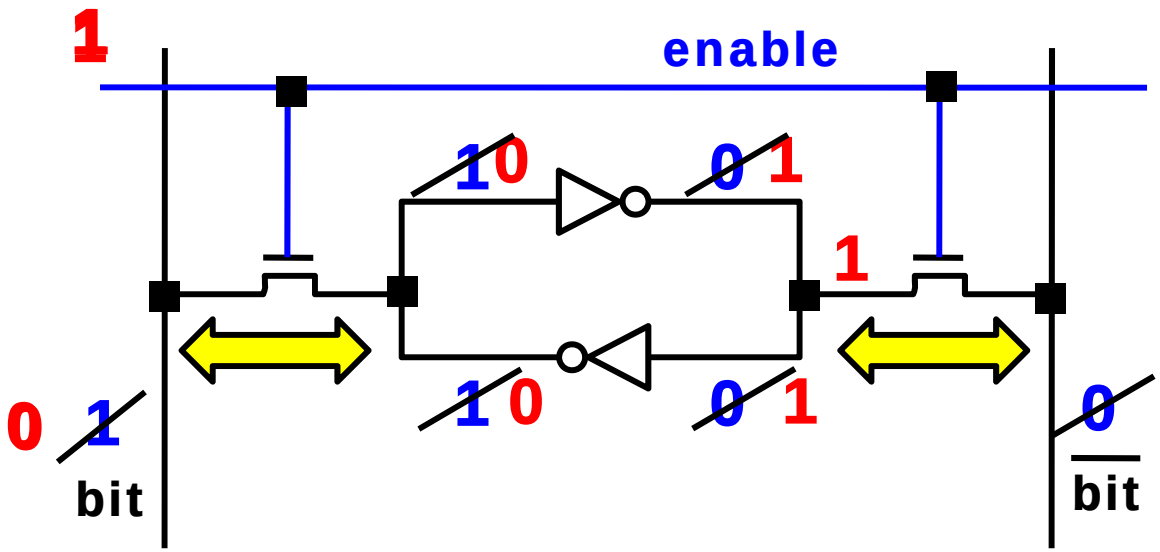
STATIC RAM CELL: reading its contents



Assume that the content is already in there as shown

1. **Place 1 on ENABLE to read**
2. **Transistors are closed ➔ flow through**
3. **State is maintained**
4. ***bit*** **and bit' are detected through the sense output**

STATIC RAM CELL: writing a “0”

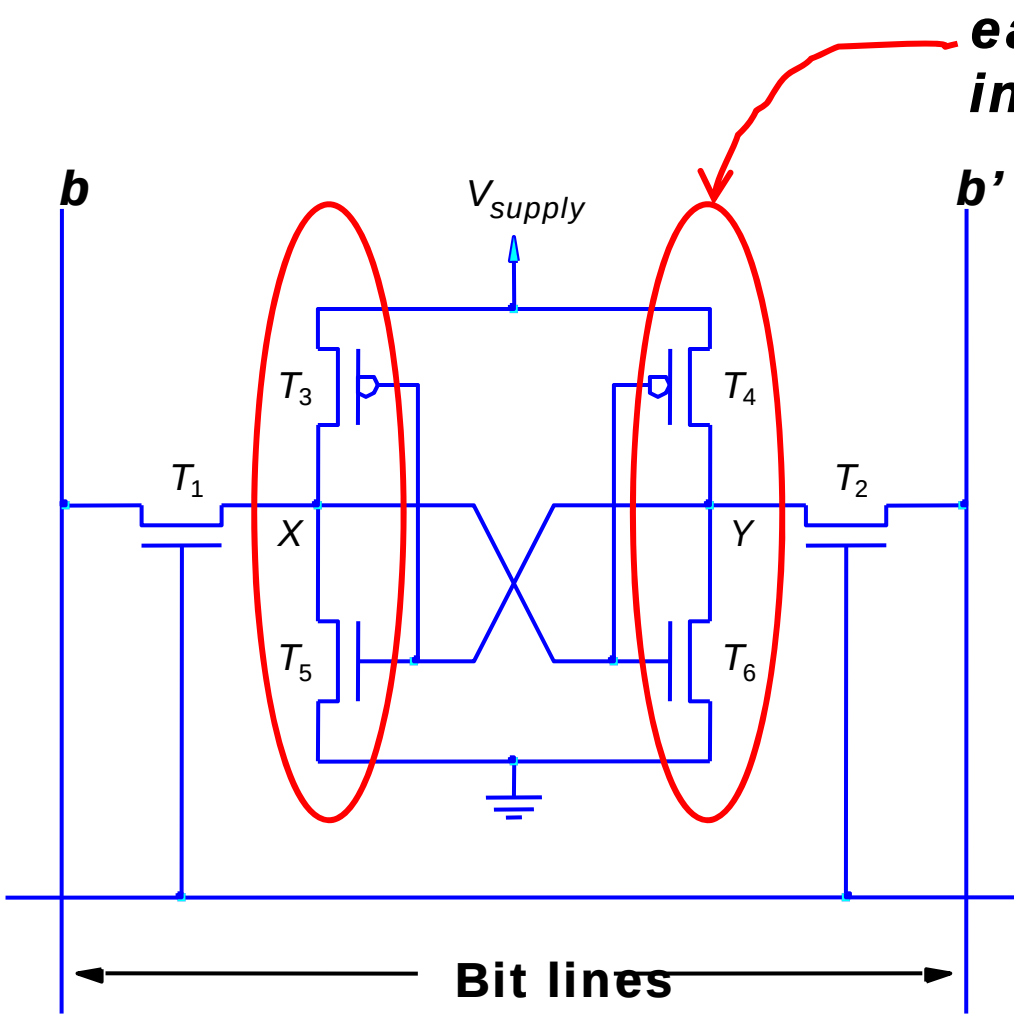


Assume that the content is already in there as shown

1. **Place 0 on *bit***
2. ***Place 1 on ENABLE***
3. **Transistors are closed ➔ flow through**
4. **State is changed**
5. ***bit* and bit’ are detected through the sense output**

SRAM

Each bit in an SRAM is stored on four transistors that form two cross-coupled inverters.



This storage cell has two stable states which are used to denote 0 and 1.

Two additional access transistors serve to control the access to a storage cell during read and write operations.

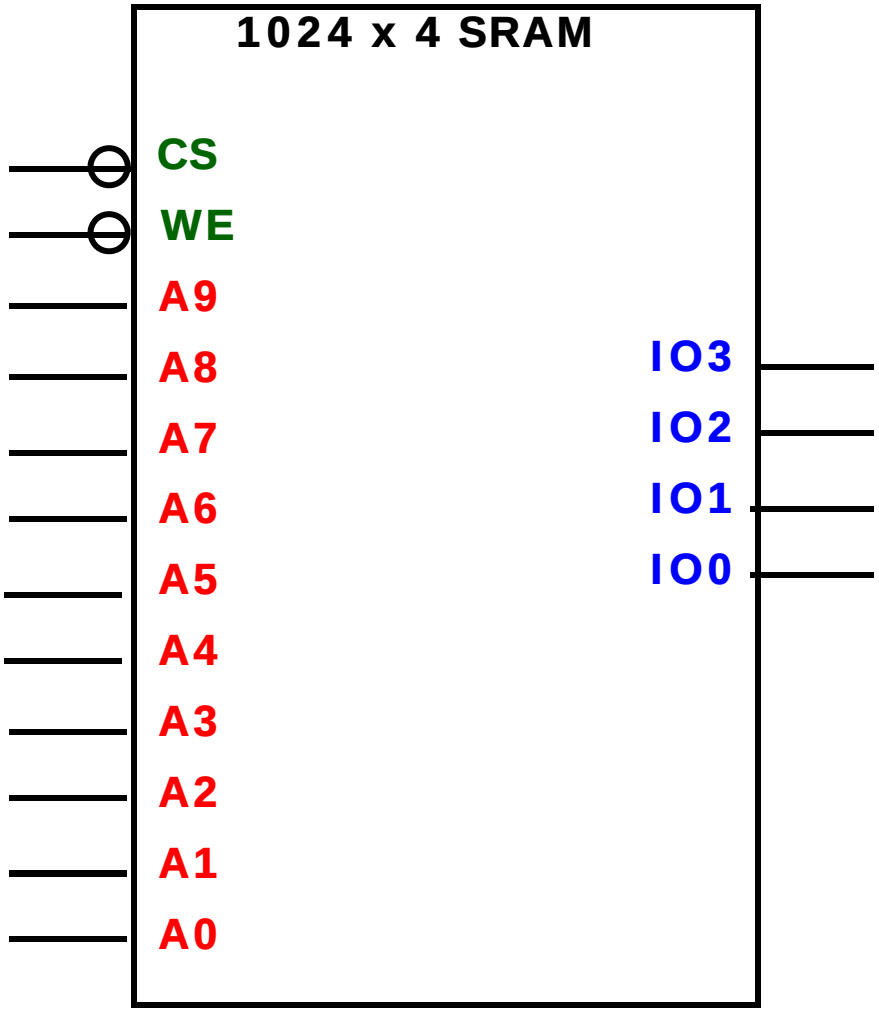
STATIC RAM: typical chip

CS:Chip Select Line
(active low)

WE:Write Enable Line
(active low)

10 Address Lines

4 Bidirectional Data Lines



Dynamic RAM

MAIN FEATURES:

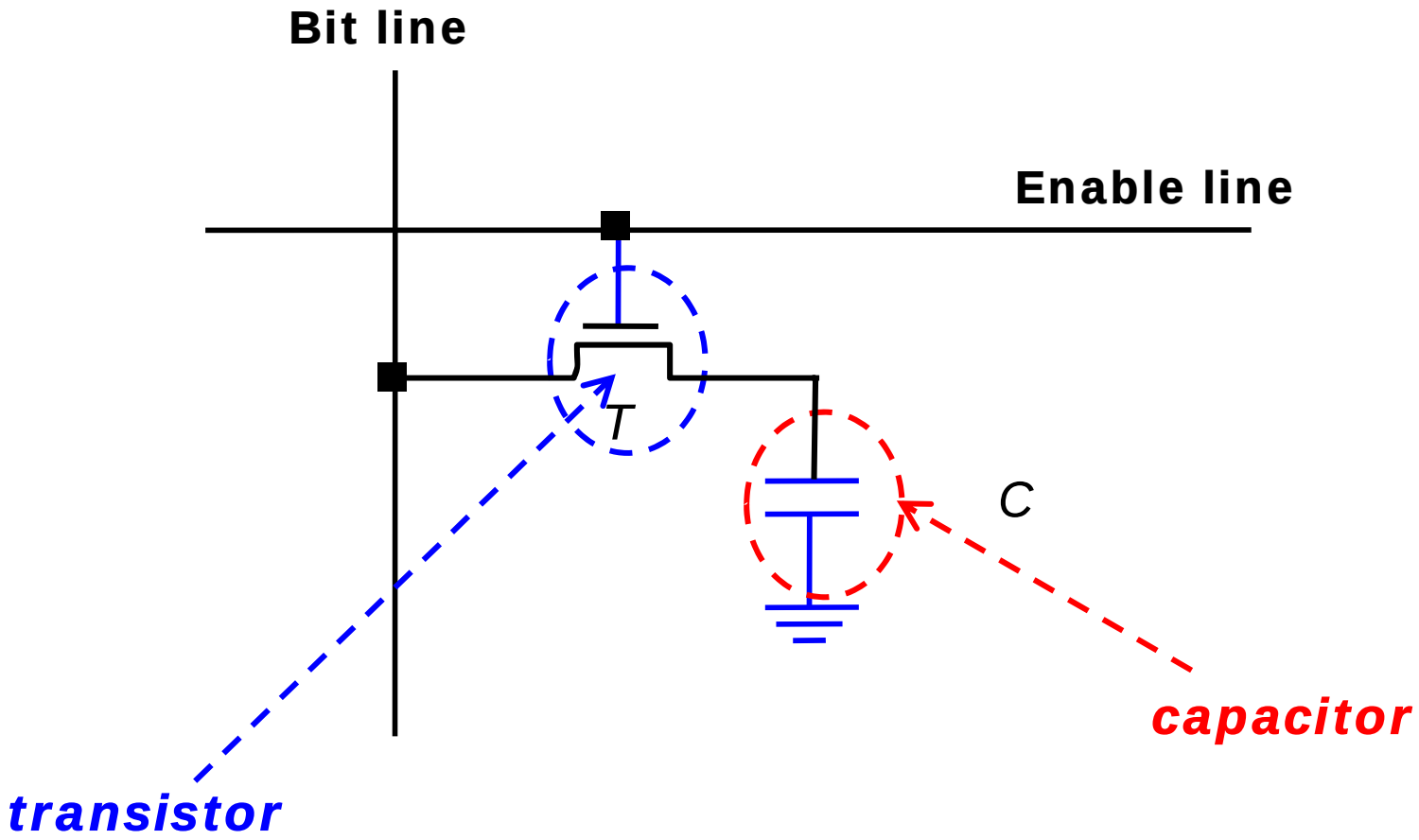
- **Information stored as charge in a capacitor**
- **Charges leak**
- **Needs refreshing even when powered**
  - **➔ Need refresh circuits**
- **Retains state for 16 to 64 ms**
- **Simpler construction**

MORE DENSITY **➔ Smaller area per bit**
➔ usually 1 transistor + 1 capacitor per bit

LESS EXPENSIVE **in manufacturing (as cost per bit)**

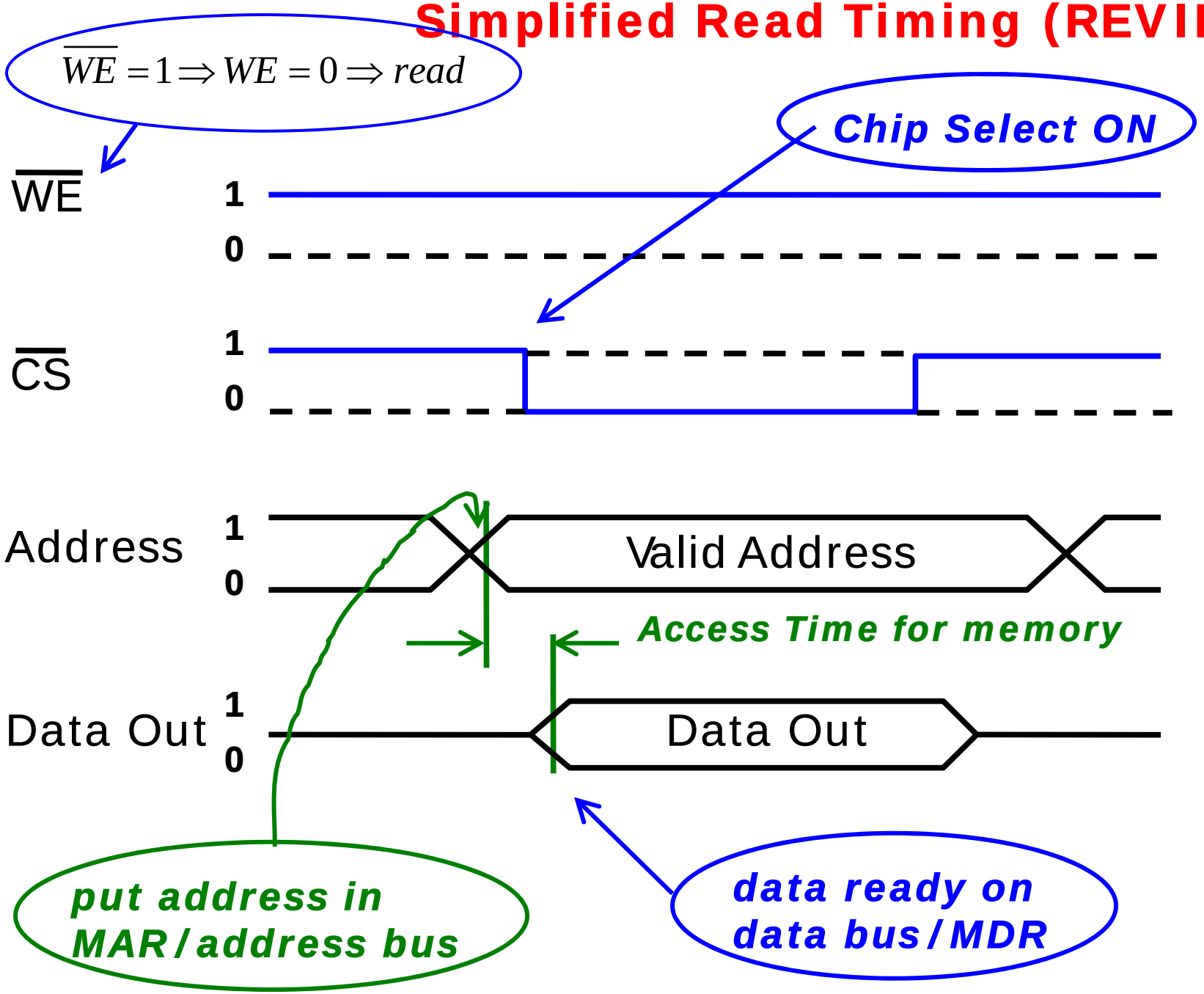
MUCH SLOWER **than SRAM**

Used in Main memory

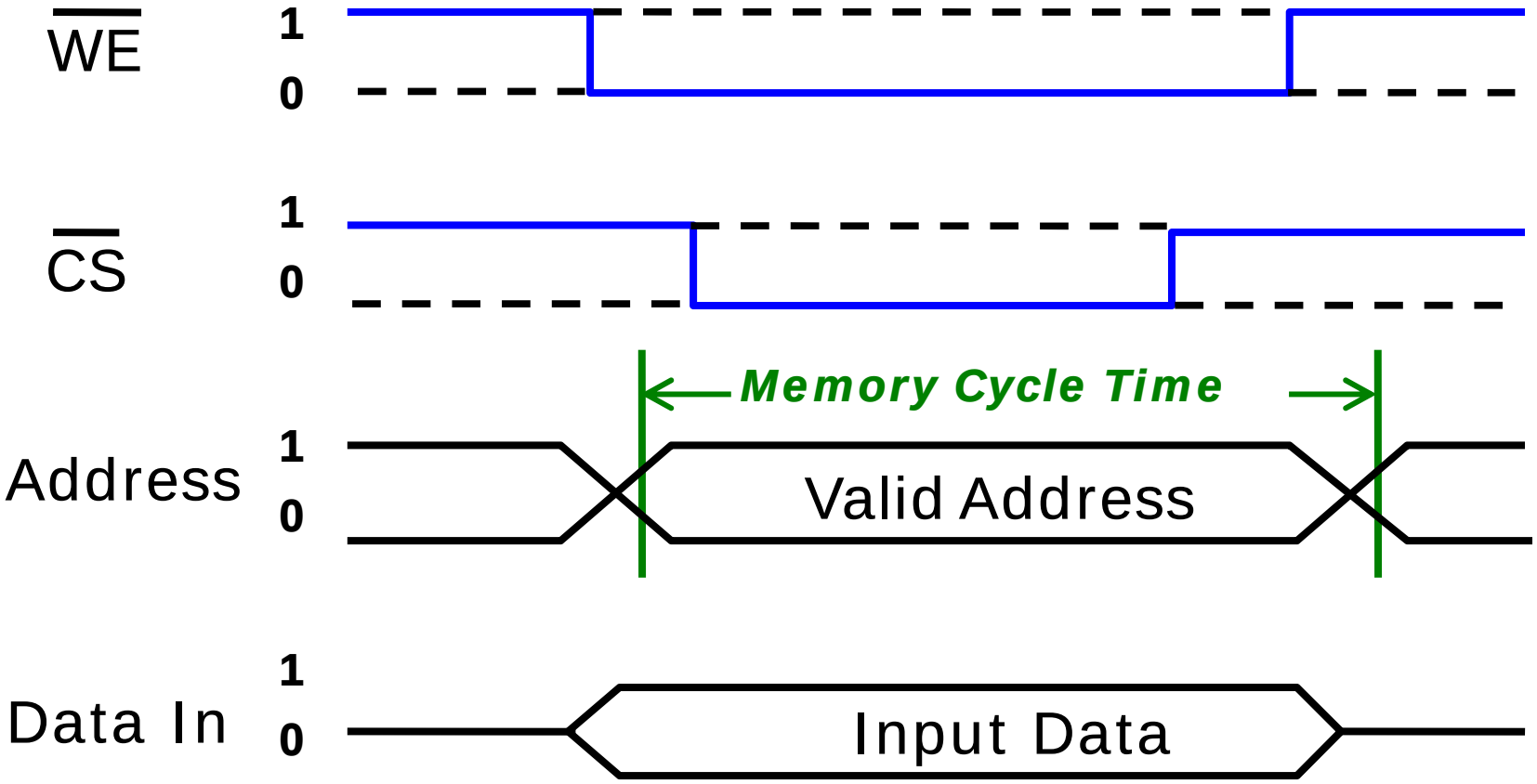


A single-transistor dynamic memory cell

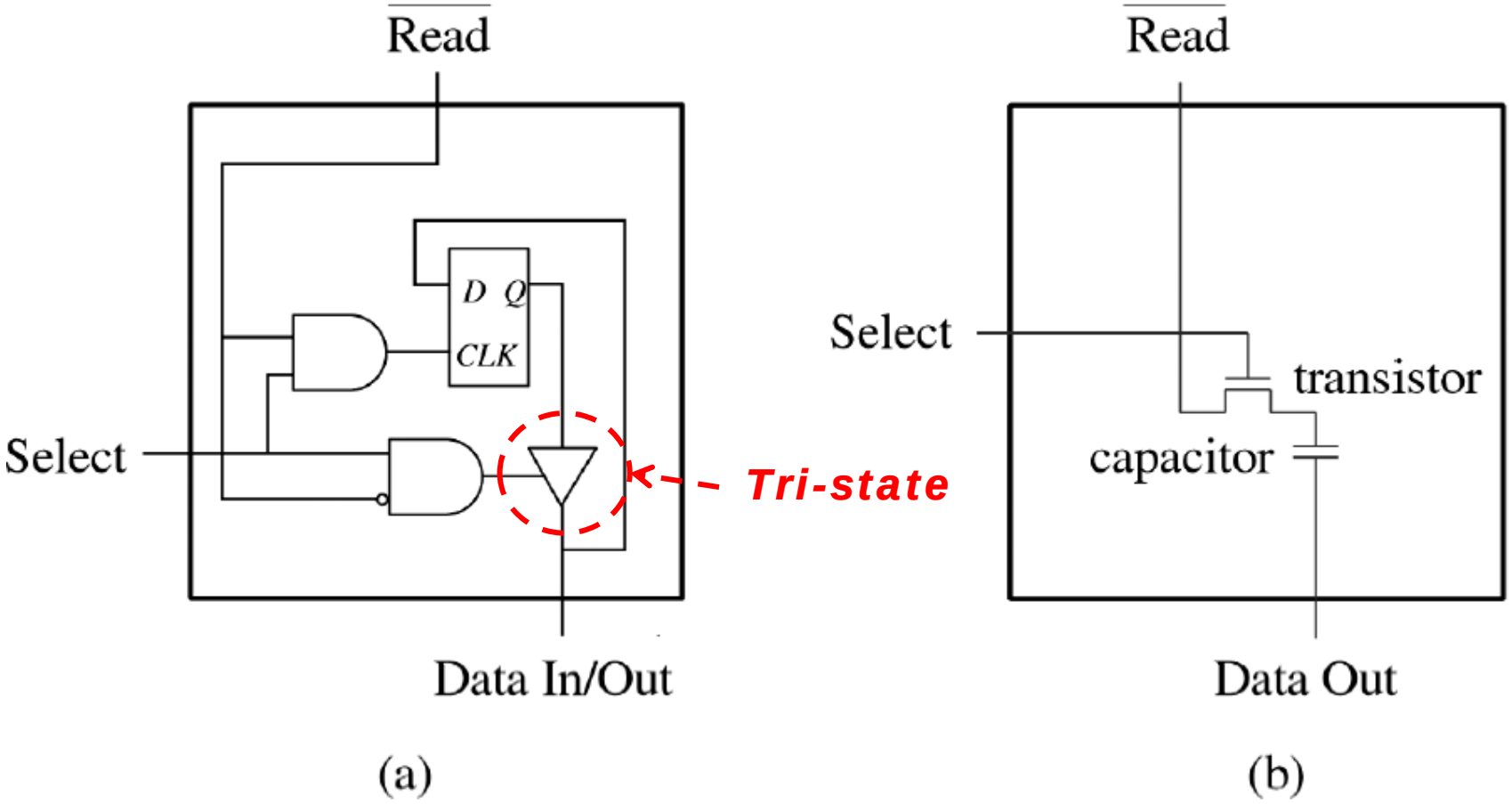
Simplified Read Timing (REVIEW)



Simplified Write Timing (REVIEW)

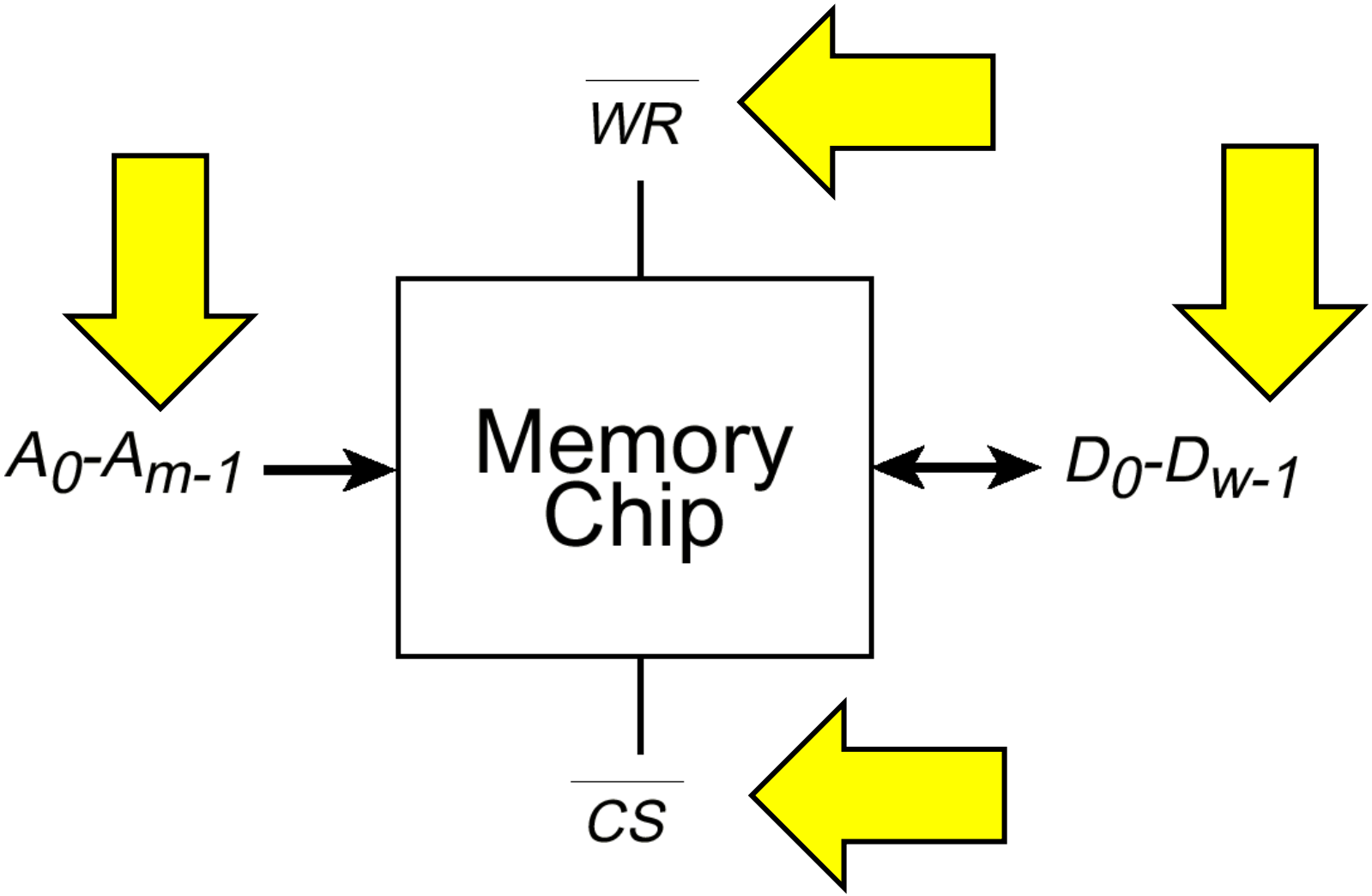


Functional Behavior of a RAM Cell



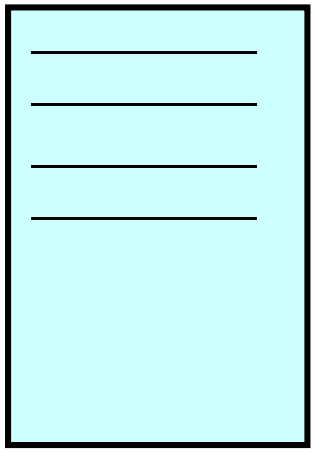
Static RAM cell (a) and dynamic RAM cell (b).

Simplified RAM Chip Pinout



Refresh Cycle for Dynamic RAM: an example

8K = 8,192 rows (each 8 bits)



assume: **4 clock cycles to access/refresh each row**

➔8192 x 4 **= 32,768 clock cycles to refresh all**

assume: 900 MHz clock cycle

➔ then time to refresh all:

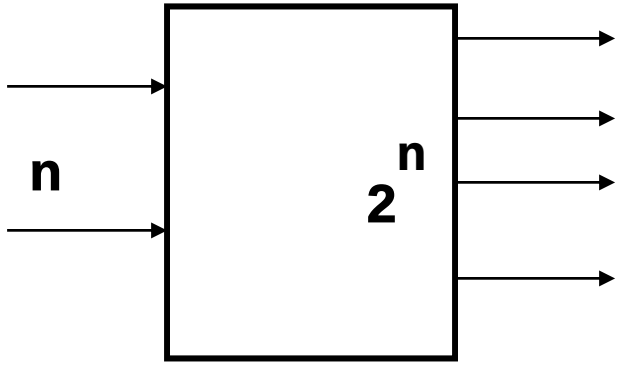
$$\frac{32,768}{900\times10^{6}} = 36.4\times10^{-6} \text{ sec}$$

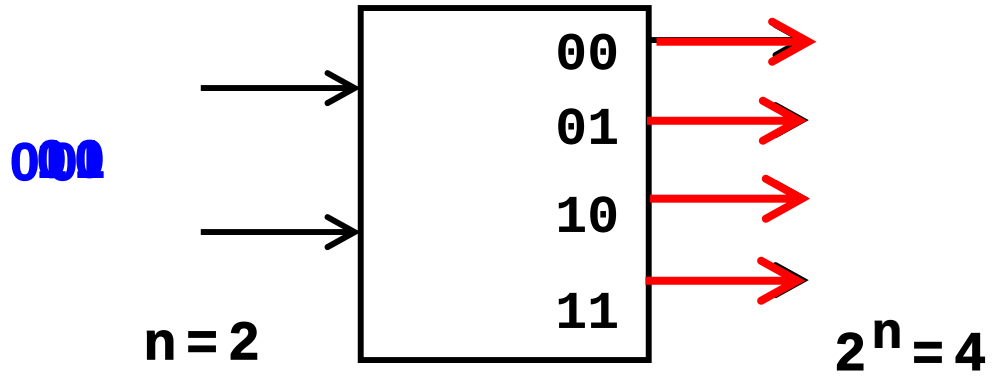
➔ that is, *0.0364 ms* at each refresh interval needed for the refresh itself

➔ then 0.0364 / 64 = 0.0005 ➔ 0.5% of memory access time is refresh overhead

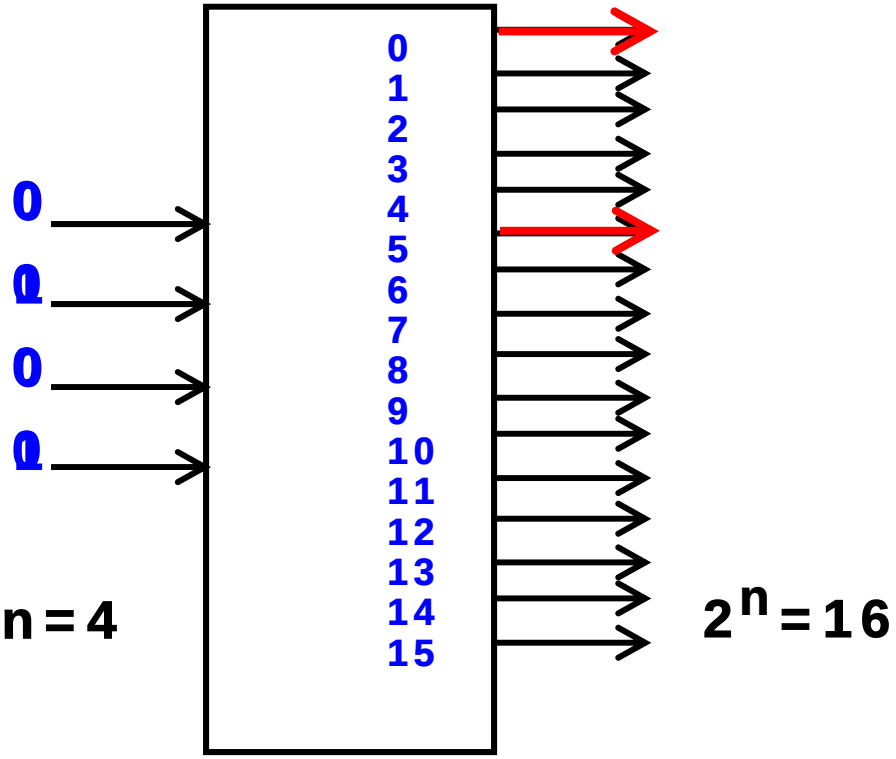
64 ms refresh interval

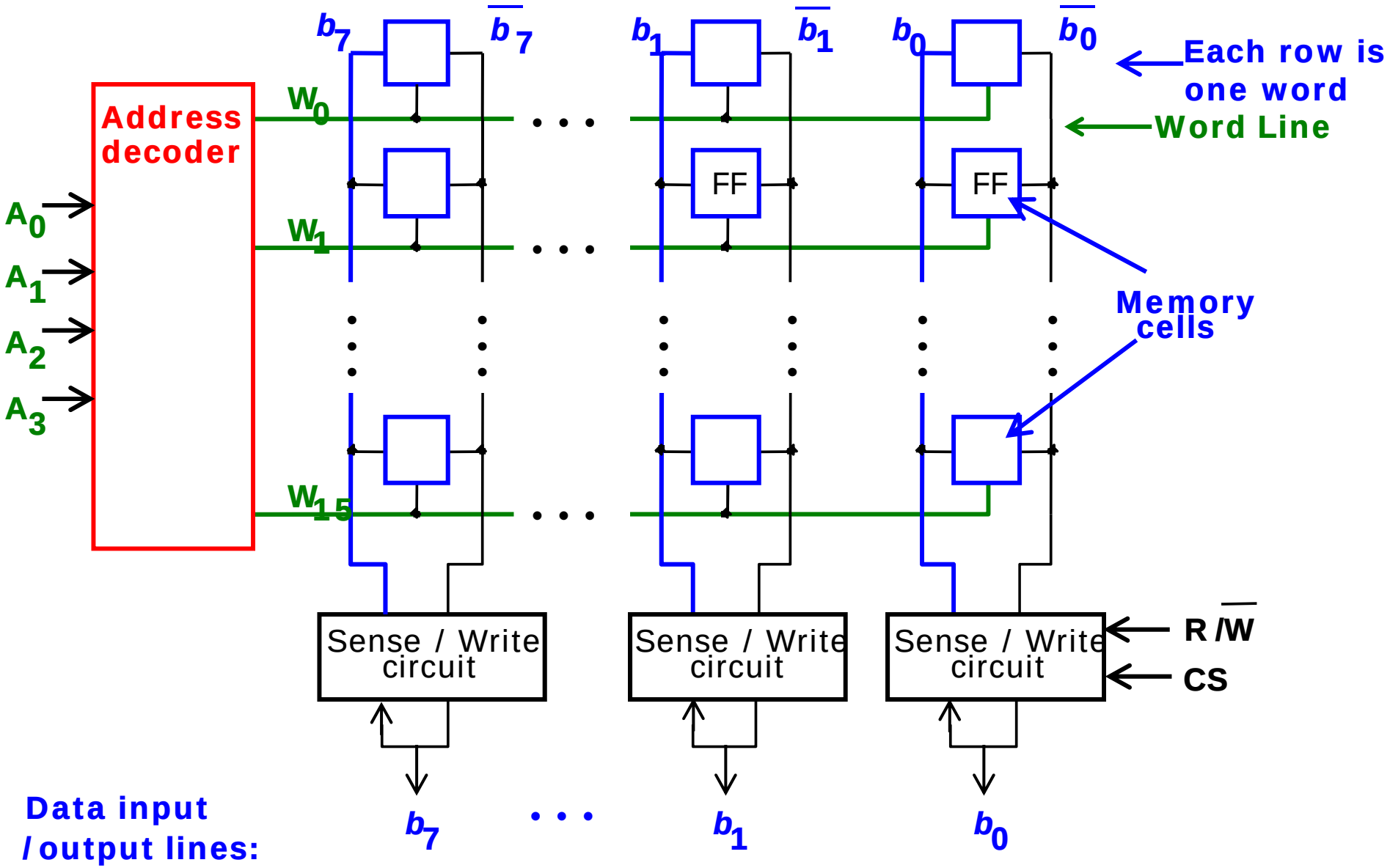
what is a Decoder?



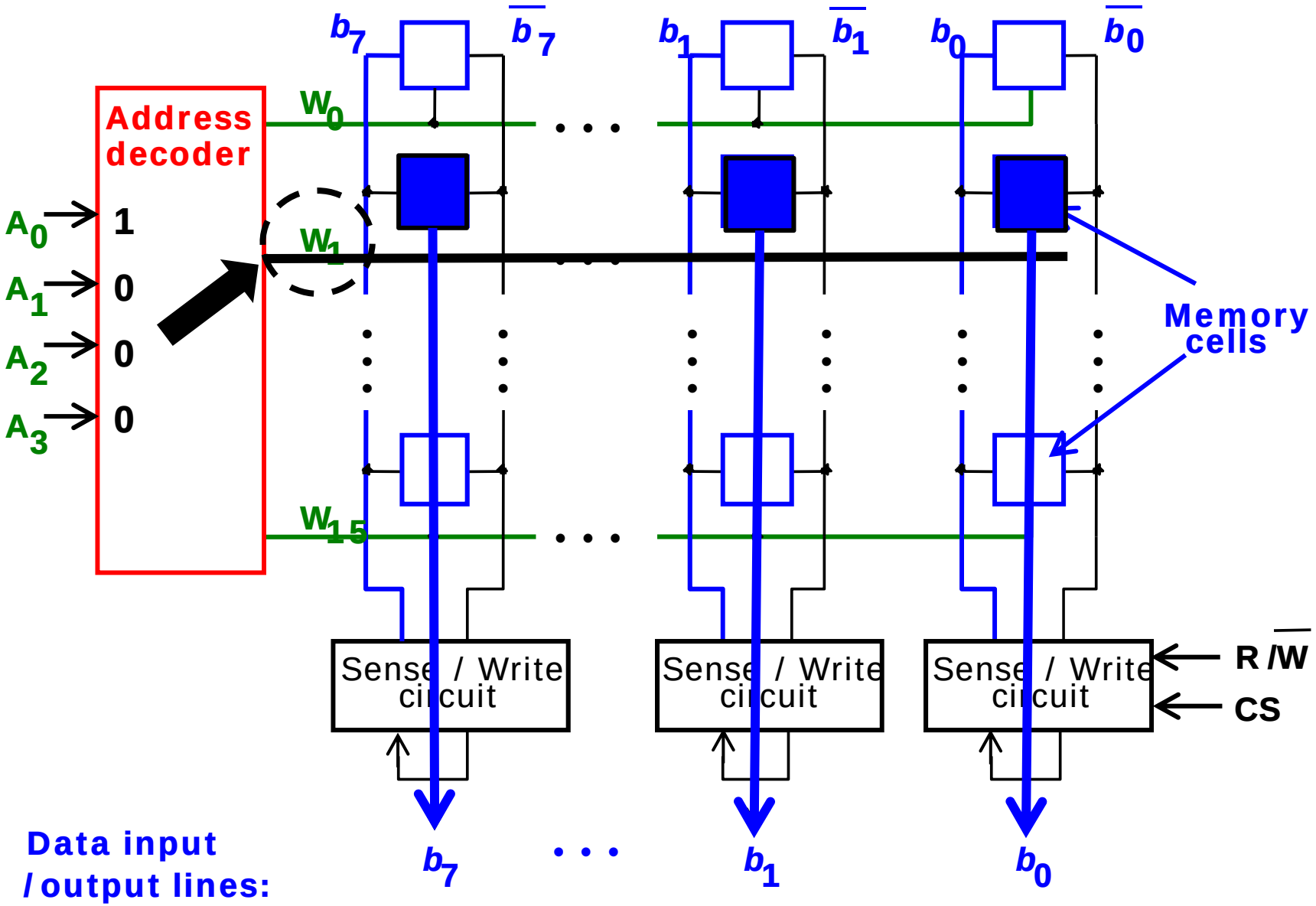


given n-bits on input lines, only 1 of 2^n output lines is activated (high or low)





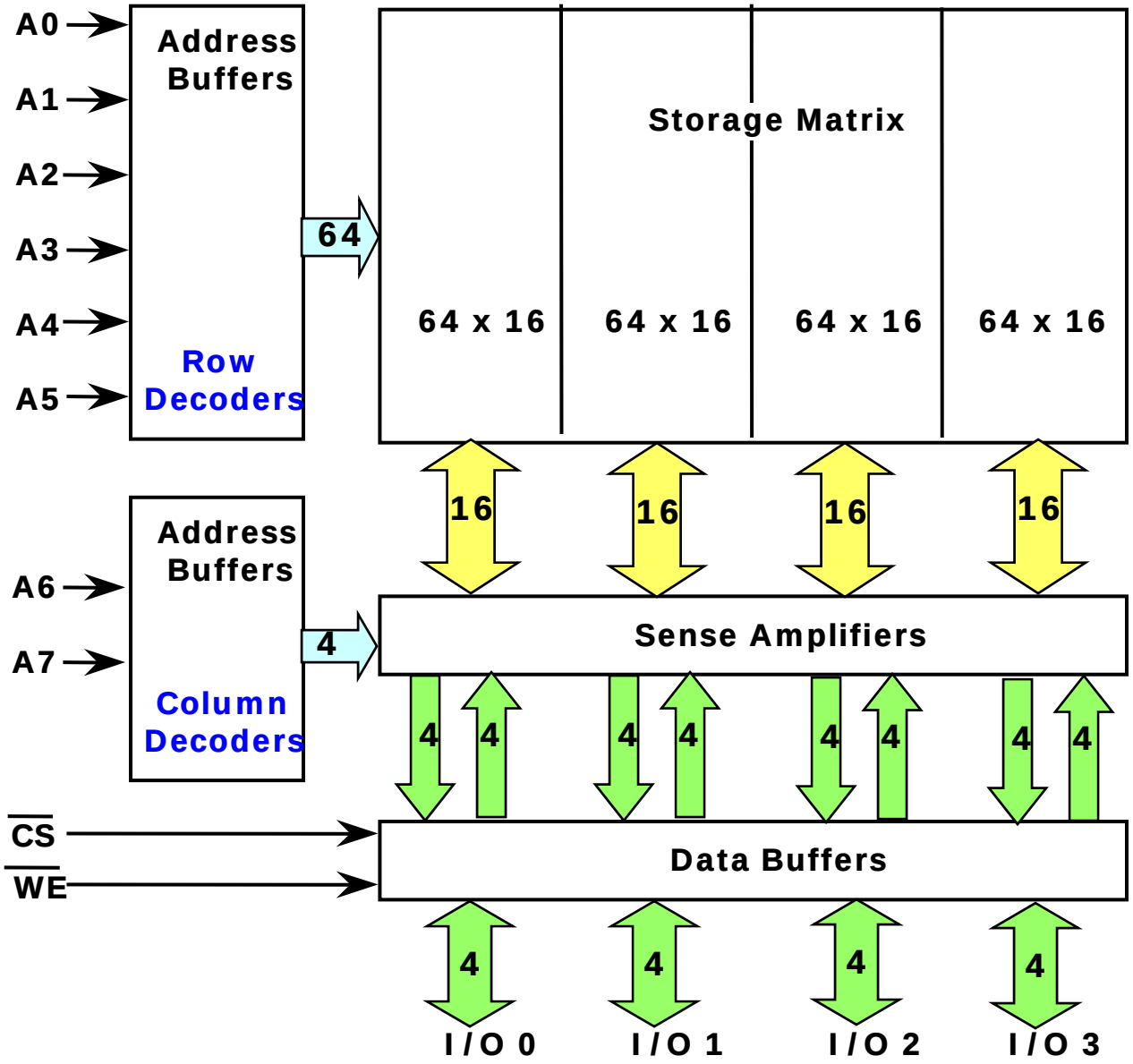
Organization of bit cells in a memory chip



Organization of bit cells in a memory chip

256 words of 16 bits

Block Diagram



Some Address bits select a row

Some Address bits select within row

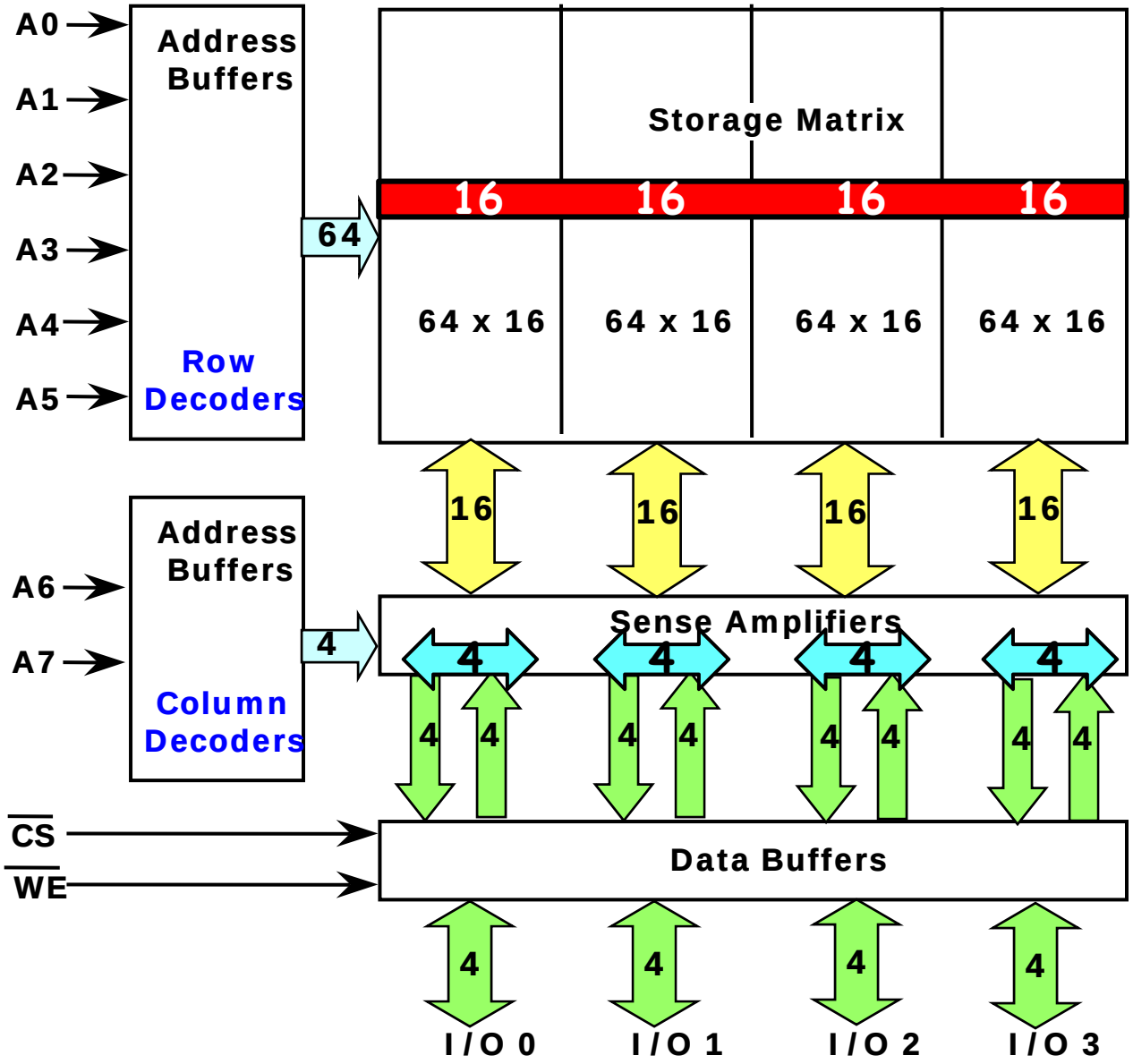
64 x 64 Square Array

OR

256 x 16 Long Rectangle Array

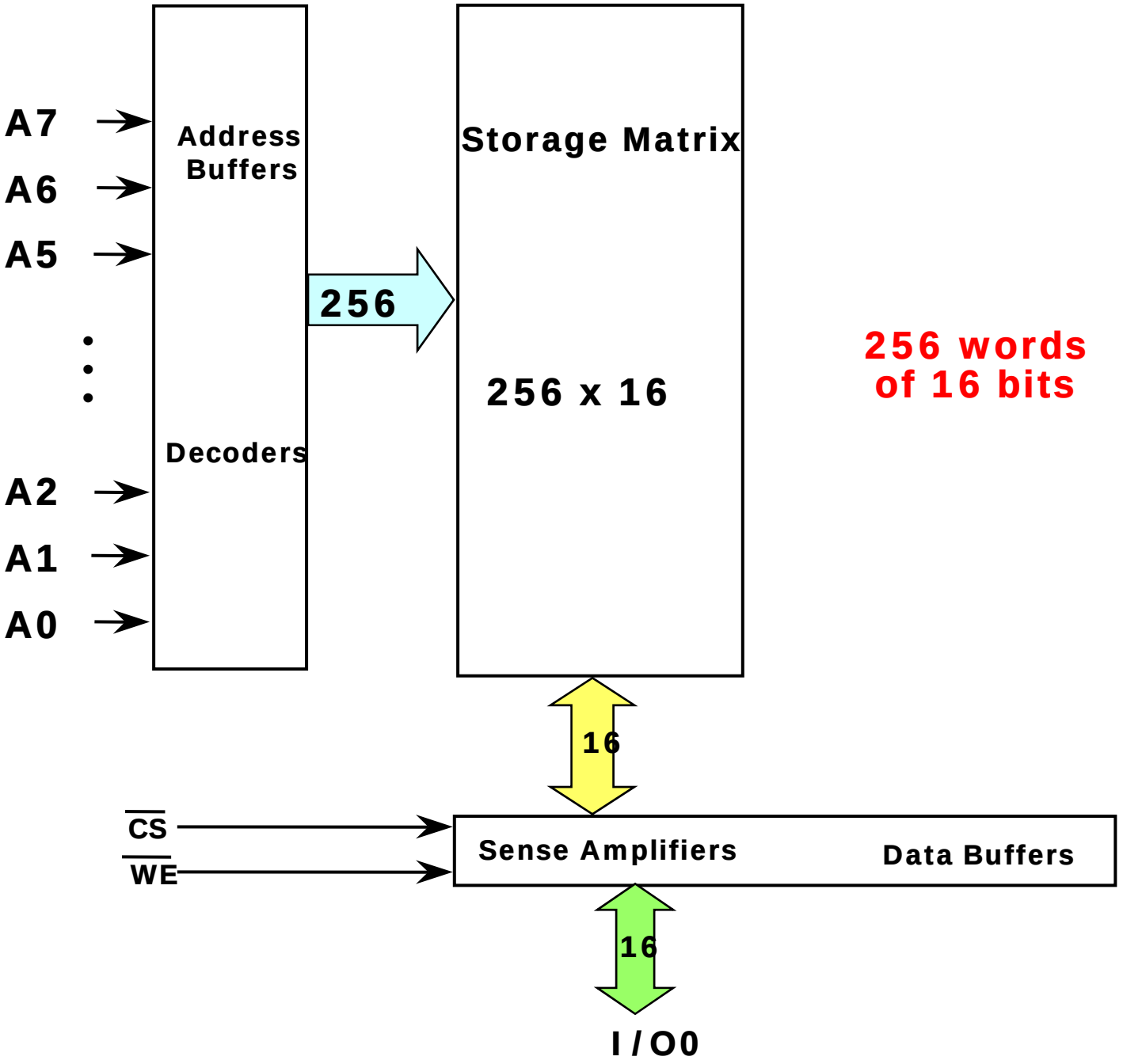
256 words of 16 bits

Block Diagram

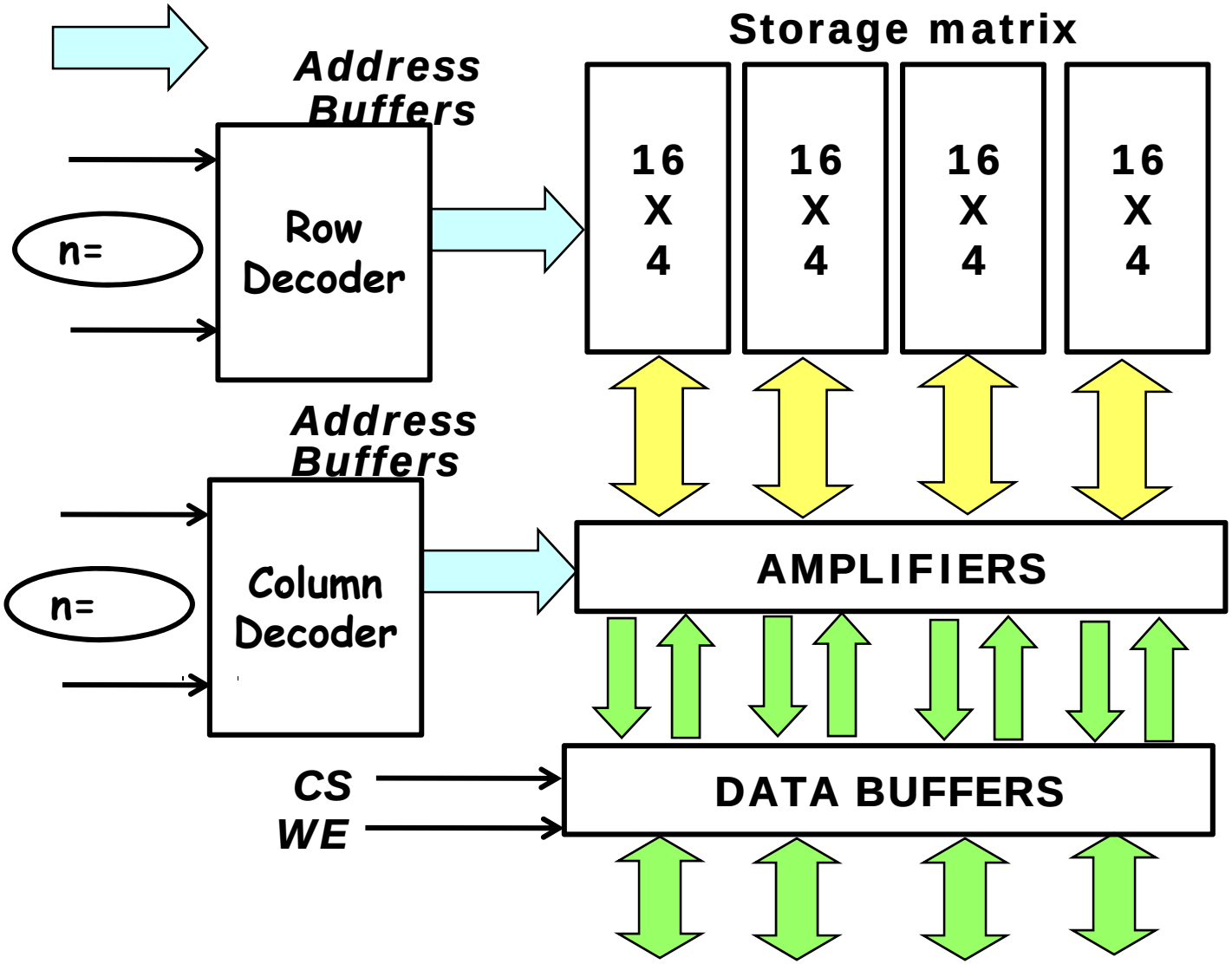


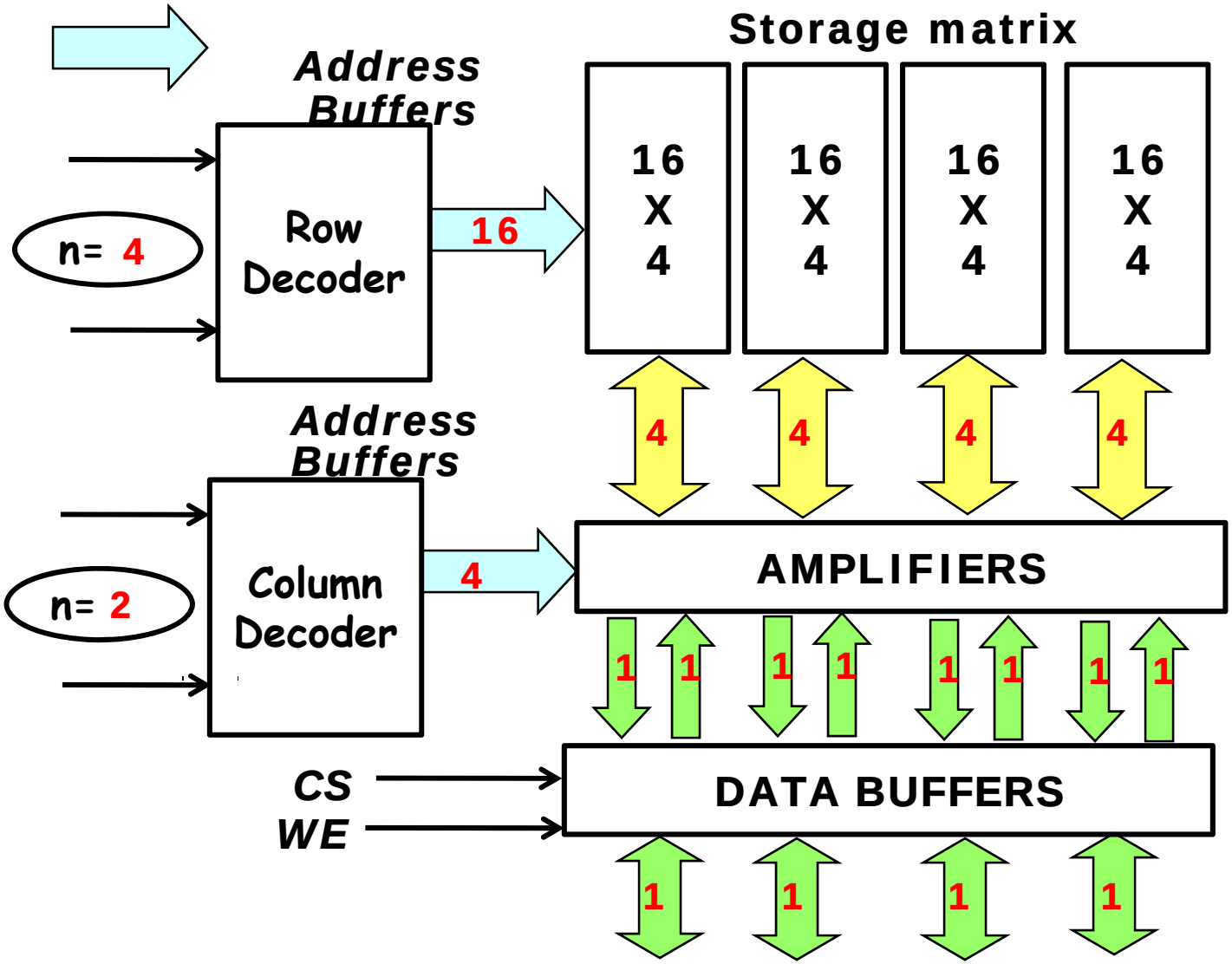
Some Address bits select a row

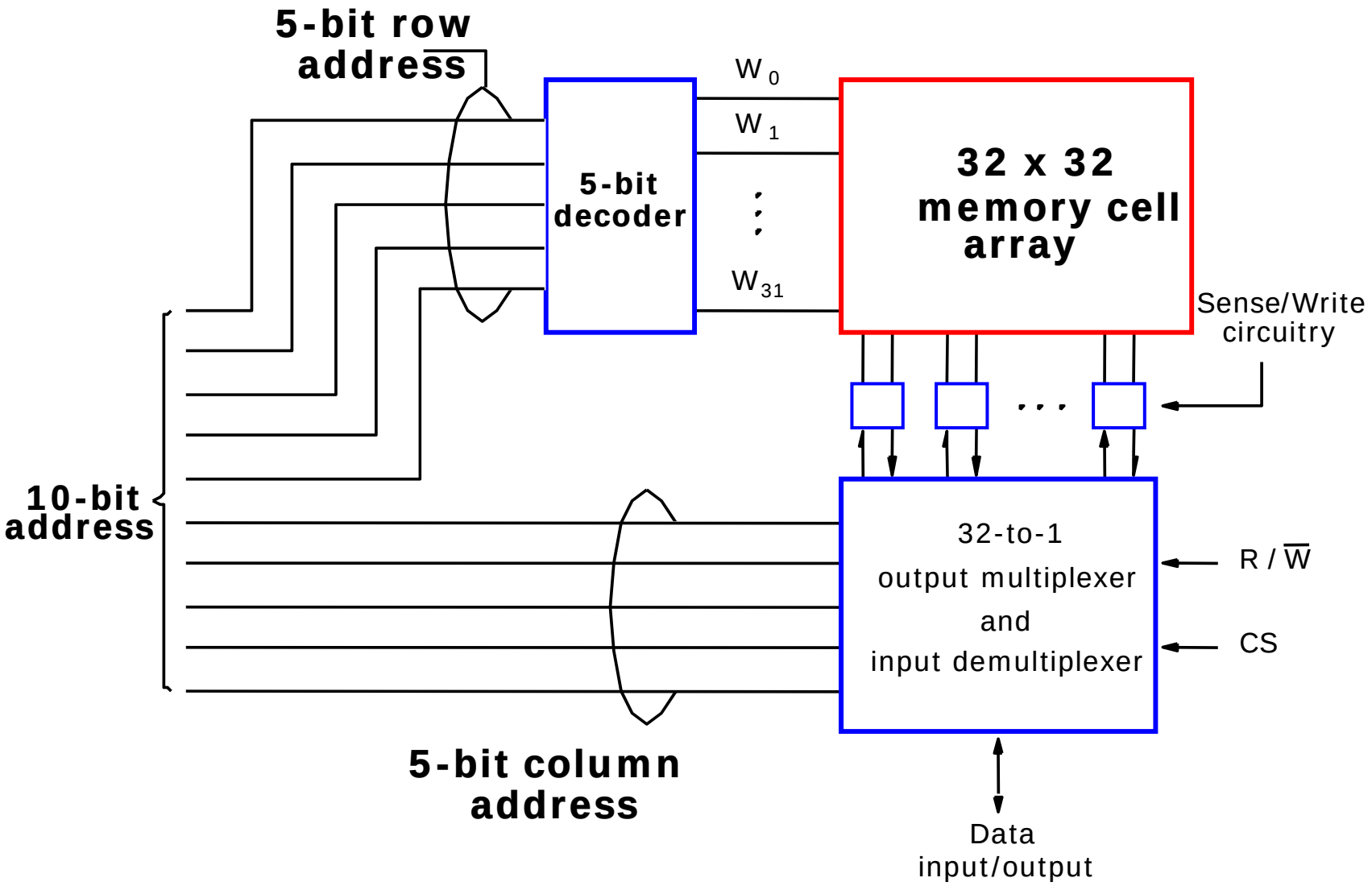
Some Address bits select within row



256 words of 16 bits

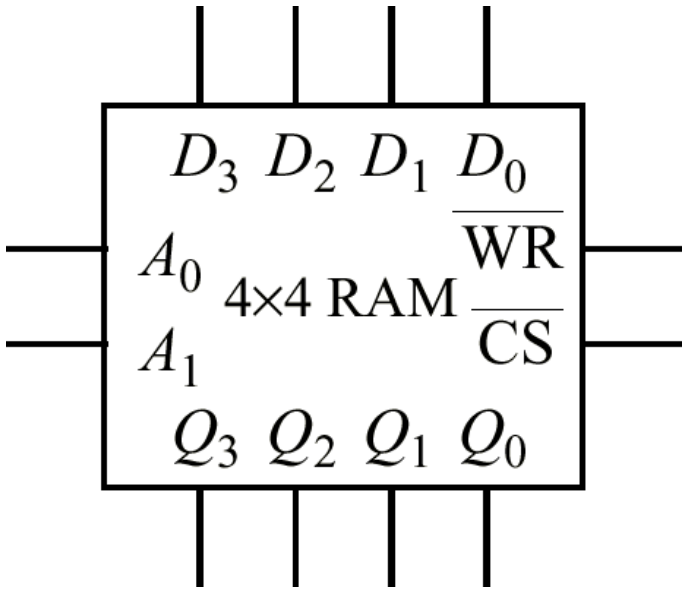

Storage matrix
Address Buffers
Row Decoder
n=
16 X 4
16 X 4
16 X 4
16 X 4
Address Buffers
Column Decoder
n=
AMPLIFIERS
DATA BUFFERS
CS
WE


Storage matrix
Address Buffers
Row Decoder
n= 4
16
16 X 4
16 X 4
16 X 4
16 X 4
4
4
4
4
Address Buffers
Column Decoder
n= 2
4
AMPLIFIERS
1
1
1
1
1
1
1
1
CS
WE
DATA BUFFERS
1
1
1
1

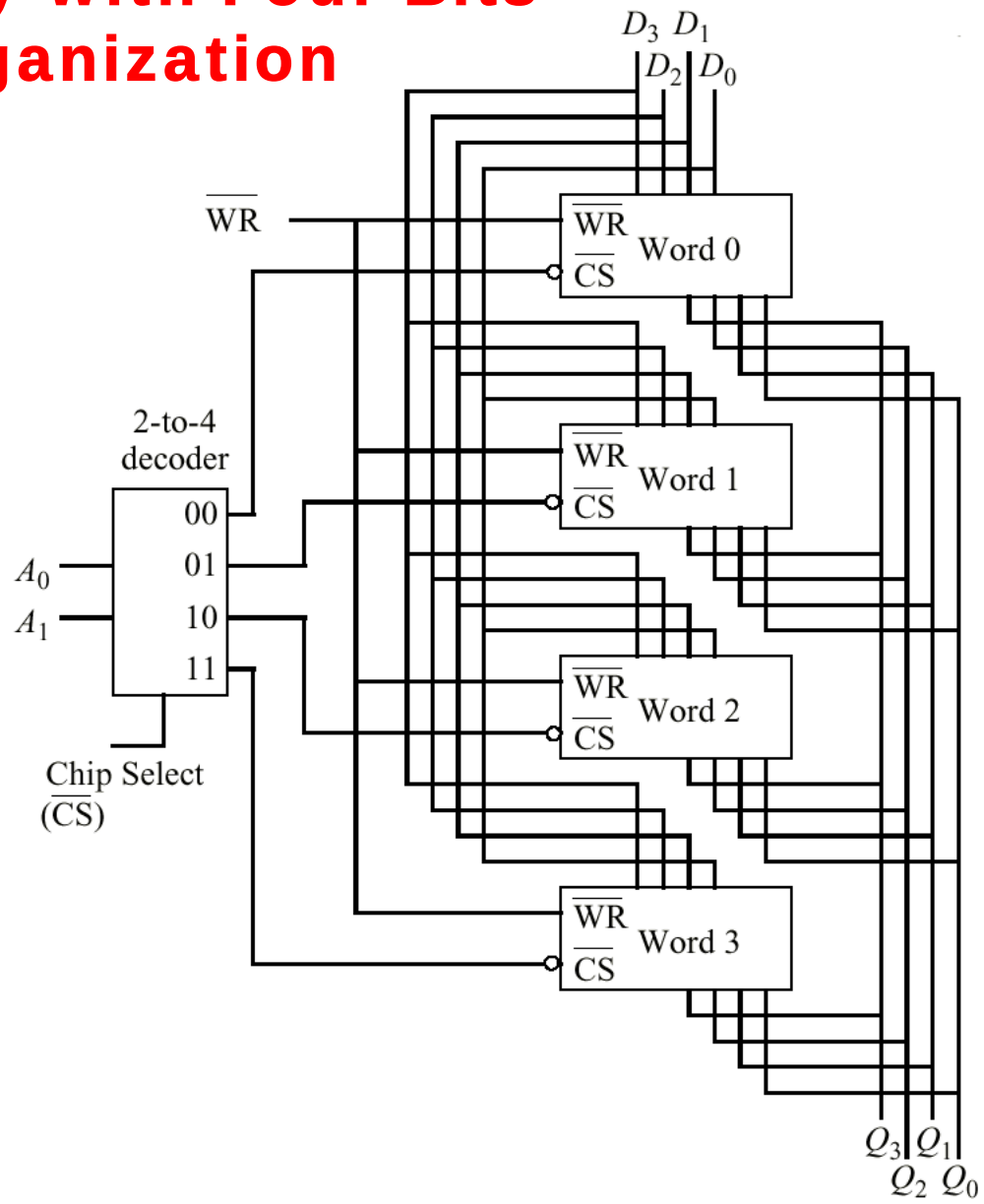


Organization of a 1K × 1 memory chip

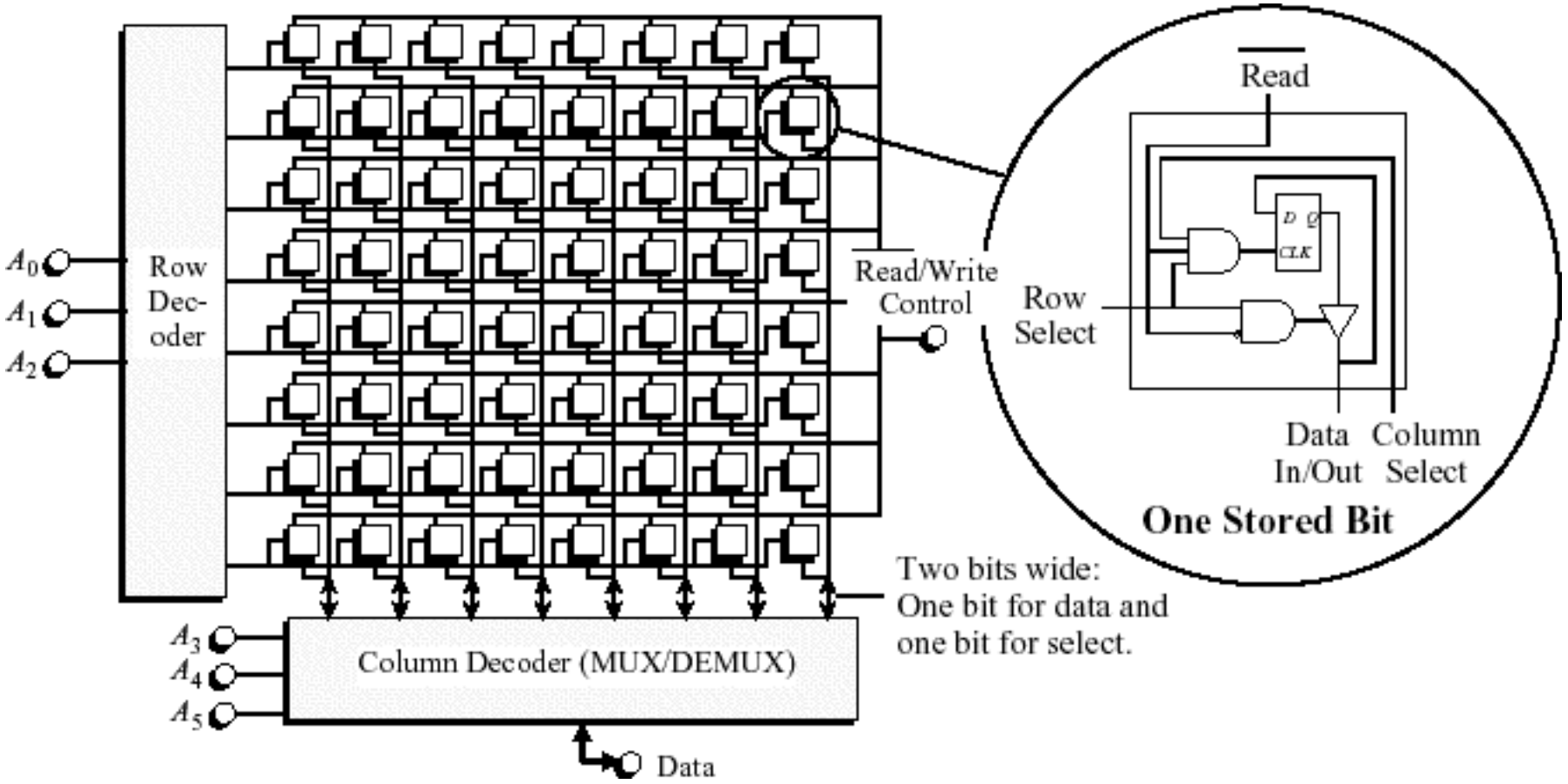
A Four-Word Memory with Four Bits per Word in a 2D Organization



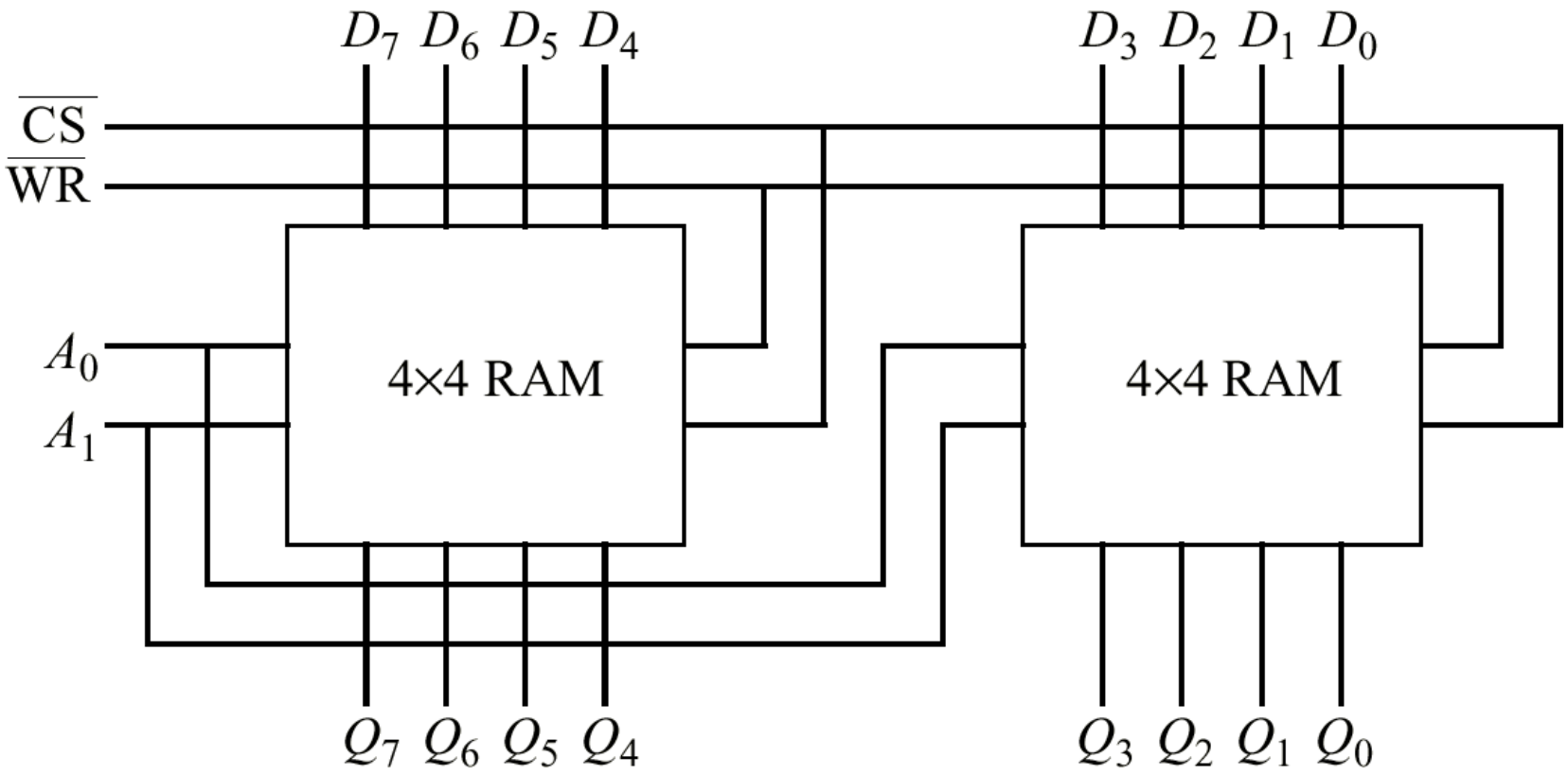
A Simplified Representation of the RAM



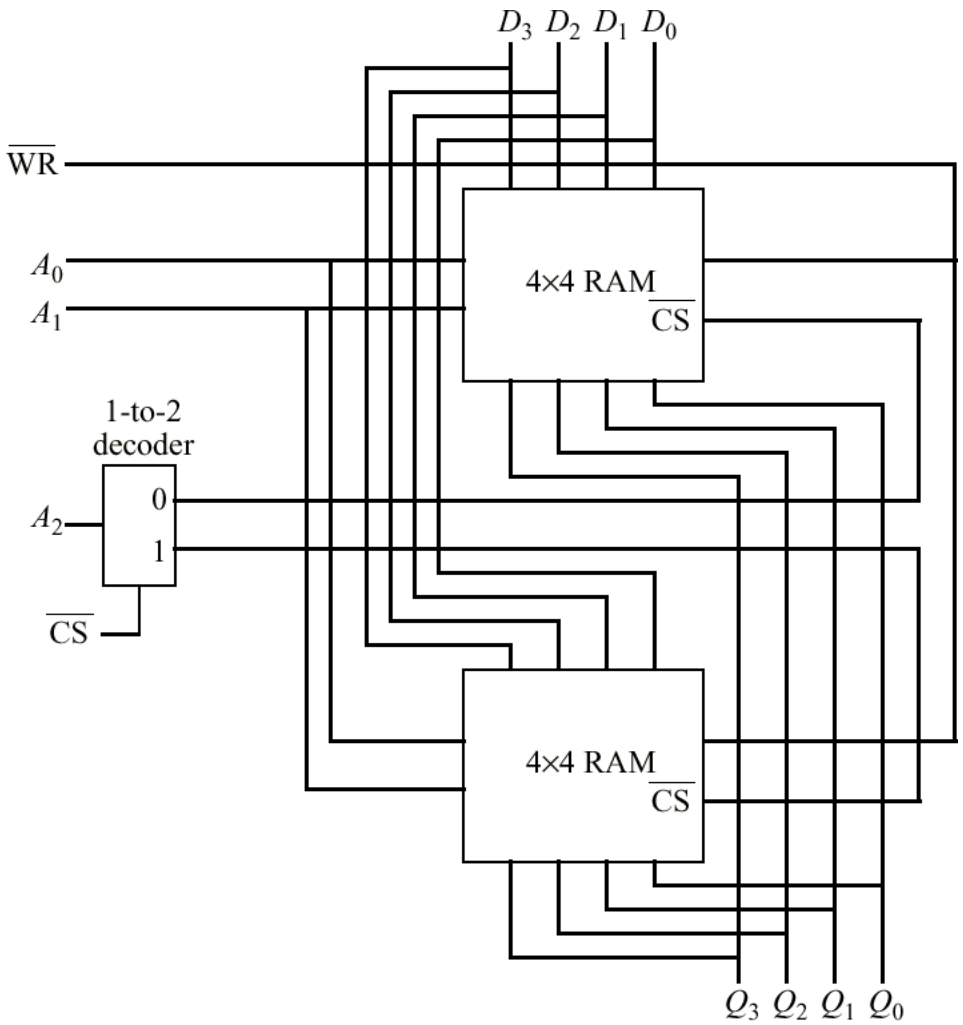
2-1/2D Organization of a 64-Word by One-Bit RAM



Two Four-Word by Four-Bit RAMs are Used in Creating a Four-Word by Eight-Bit RAM

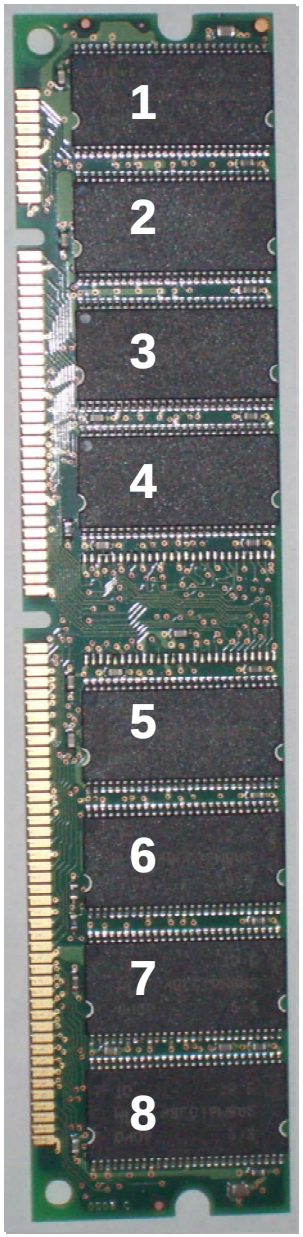


Two Four-Word by Four-Bit RAMs Make up an Eight-Word by Four-Bit RAM



Single-In-Line Memory Module

- 256 MB dual in-line memory module organized for a 64-bit word with 16 16M × 8-bit RAM chips (eight chips on each side of the DIMM).



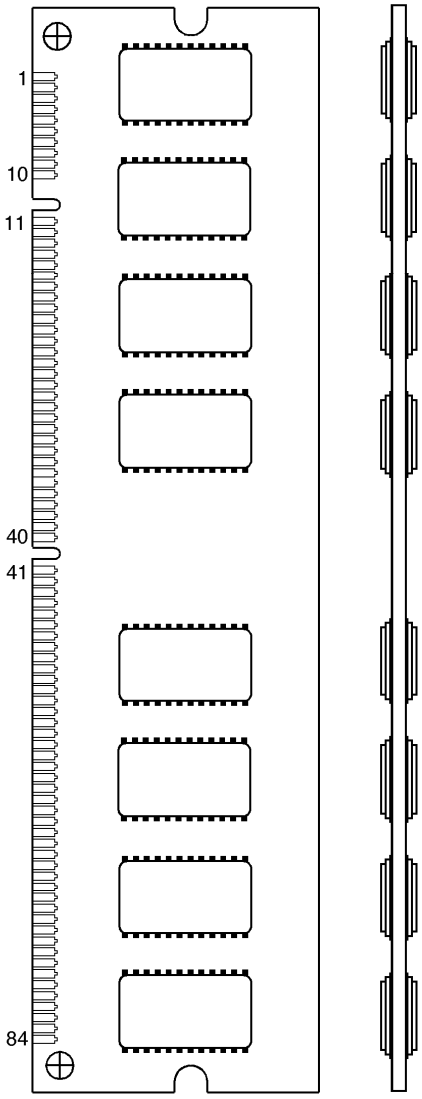
PIN NO.	NAME	PIN NO.	NAME	PIN NO.	NAME	PIN NO.	NAME
1	V_{SS}	43	V_{SS}	85	V_{SS}	127	V_{SS}
2	DQ0	44	$\overline{OE2}$	86	DQ32	128	NC
3	DQ1	45	$\overline{RAS2}$	87	DQ33	129	$\overline{RAS3}$
4	DQ2	46	$\overline{CAS2}$	88	DQ34	130	$\overline{CAS6}$
5	DQ3	47	$\overline{CAS3}$	89	DQ35	131	$\overline{CAS7}$
6	V_{DD}	48	$\overline{WE2}$	90	V_{DD}	132	NC
7	DQ4	49	V_{DD}	91	DQ36	133	V_{DD}
8	DQ5	50	NC	92	DQ37	134	NC
9	DQ6	51	NC	93	DQ38	135	NC
10	DQ7	52	NC	94	DQ39	136	NC
11	DQ8	53	NC	95	DQ40	137	NC
12	V_{SS}	54	V_{SS}	96	V_{SS}	138	V_{SS}
13	DQ9	55	DQ16	97	DQ41	139	DQ48
14	DQ10	56	DQ17	98	DQ42	140	DQ49
15	DQ11	57	DQ18	99	DQ43	141	DQ50
16	DQ12	58	DQ19	100	DQ44	142	DQ51
17	DQ13	59	V_{DD}	101	DQ45	143	V_{DD}
18	V_{DD}	60	DQ20	102	V_{DD}	144	DQ52
19	DQ14	61	NC	103	DQ46	145	NC
20	DQ15	62	NC	104	DQ47	146	NC
21	NC	63	NC	105	NC	147	NC
22	NC	64	V_{SS}	106	NC	148	V_{SS}
23	V_{SS}	65	DQ21	107	V_{SS}	149	DQ53
24	NC	66	DQ22	108	NC	150	DQ54
25	NC	67	DQ23	109	NC	151	DQ55
26	V_{DD}	68	V_{SS}	110	V_{DD}	152	V_{SS}
27	$\overline{WE0}$	69	DQ24	111	NC	153	DQ56
28	$\overline{CAS0}$	70	DQ25	112	$\overline{CAS4}$	154	DQ57
29	$\overline{CAS1}$	71	DQ26	113	$\overline{CAS5}$	155	DQ58
30	$\overline{RAS0}$	72	DQ27	114	$\overline{RAS1}$	156	DQ59
31	$\overline{OE0}$	73	V_{DD}	115	NC	157	V_{DD}
32	V_{SS}	74	DQ28	116	V_{SS}	158	DQ60
33	A0	75	DQ29	117	A1	159	DQ61
34	A2	76	DQ30	118	A3	160	DQ62
35	A4	77	DQ31	119	A5	161	DQ63
36	A6	78	V_{SS}	120	A7	162	V_{SS}
37	A8	79	NC	121	A9	163	NC
38	A10	80	NC	122	A11	164	NC
39	A12	81	NC	123	NC	165	SA0
40	V_{DD}	82	SDA	124	V_{DD}	166	SA1
41	NC	83	SCL	125	NC	167	SA2
42	NC	84	V_{DD}	126	NC	168	V_{DD}

Single-In-Line Memory Module

- Schematic diagram of 256 MB dual in-line memory module. (Source: adapted from http://www-s.ti.com/sc/ds/tm4en64kpu.pdf.)

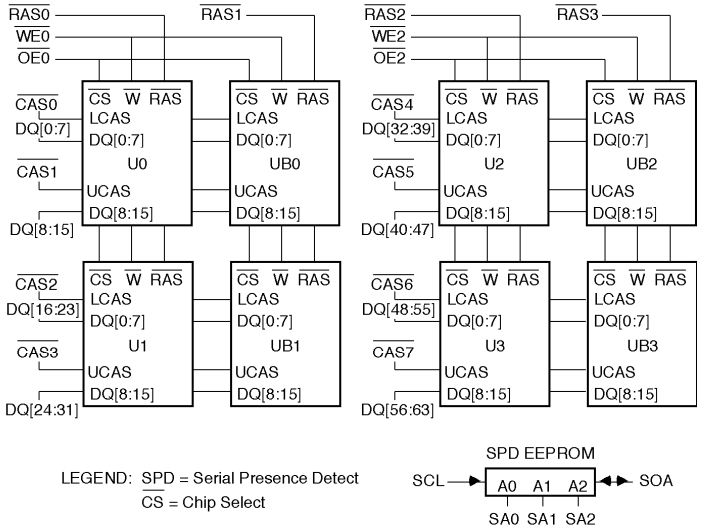
MH32EN64KPU EXTENDED-DATA-OUT DYNAMIC RAM MODULE

DUAL-IN-LINE MEMORY MODULE (TOP VIEW) SIDE VIEW



PIN NOMENCLATURE	
A[0:11]	Row Address Inputs
A[0:9]	Column Address Inputs
DQ[0:63]	Data In / Data Out
$\overline{CAS}$[0:7]	Column-Address Strobe
$\overline{RAS}$[0:3]	Row-Address Strobe
$\overline{WE0}$ and $\overline{WE2}$	Write Enable
$\overline{OE0}$ and $\overline{OE2}$	Output Enable
SA[0:2]	Serial-Presence-Detect (SPD) Device Add Input
SDA	Serial PD Address / Data
SCL	Serial PD Clock
NC	No-Connect Pin
V_{DD}	3.3-V Supply
V_{SS}	Ground

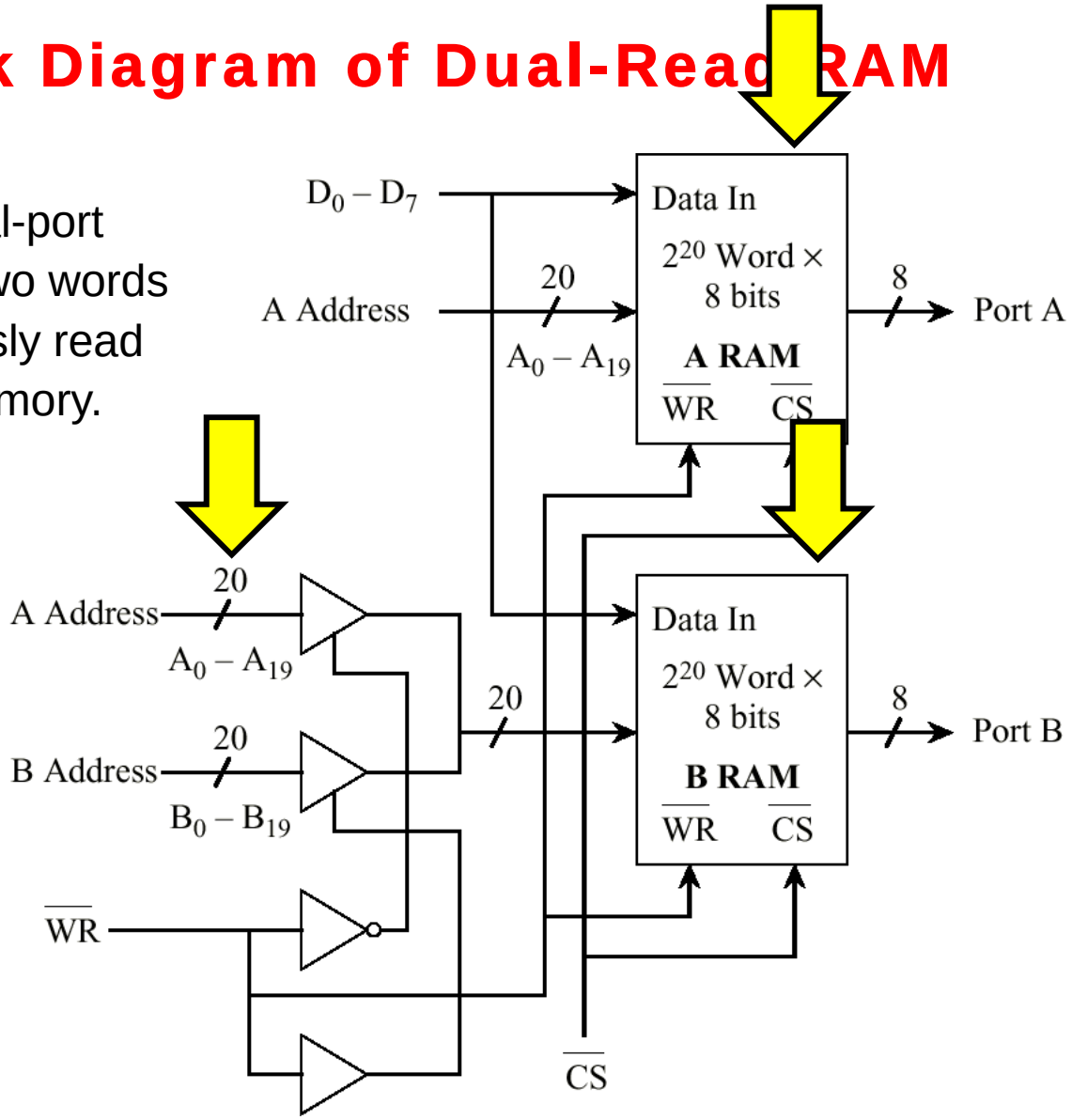
functional block diagram



Block Diagram of Dual-Read RAM

- A dual-read or dual-port RAM allows any two words to be simultaneously read from the same memory.

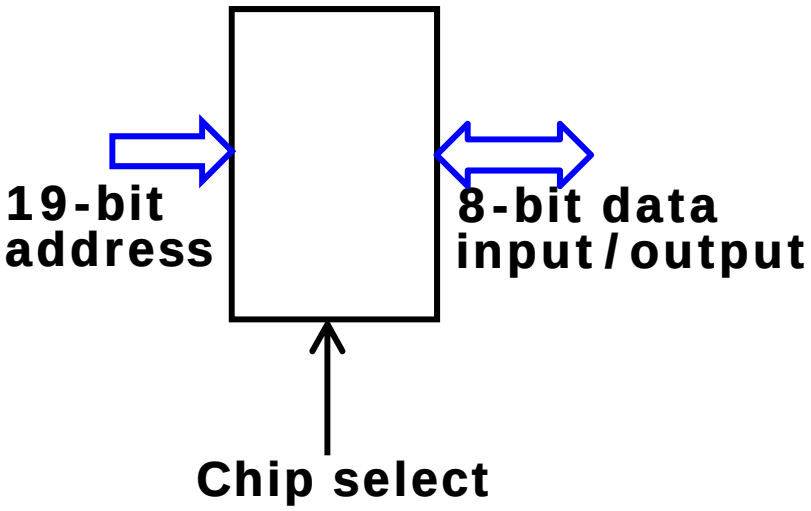
REMEMBER about DMA and need for DUAL PORT Memories?



EXAMPLE: goal is to develop a memory with 2M words of 32 bits each.

1. Structure of large memories: choose the basic chip

512K x 8 memory chip



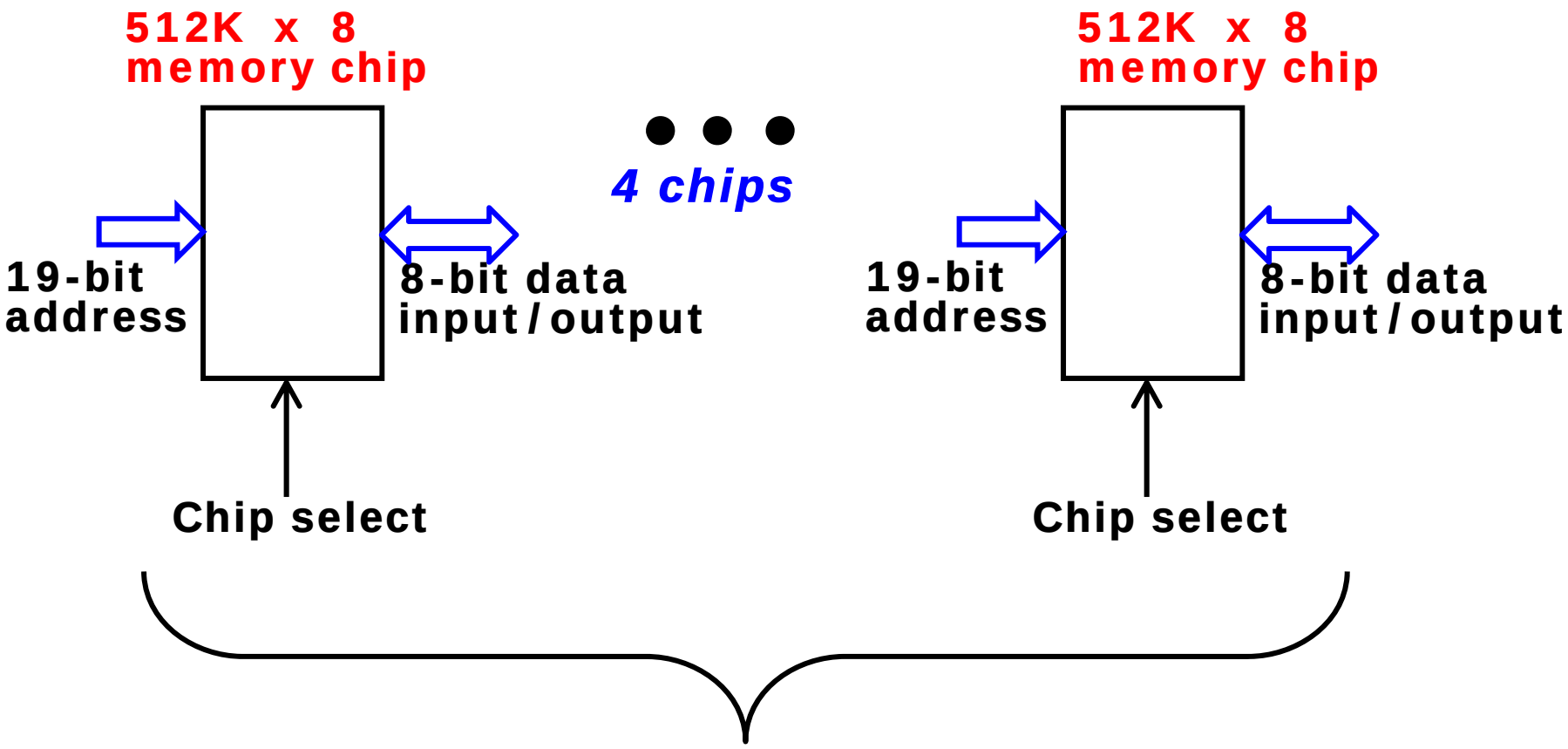
Start with 1 chip where:

512K rows

1 word of 8 bits per row

$$512K = 2^9 \times 2^{10}$$

2. Structure of large memories: put a few together in a row to get word size



Put 4 together: 512K x 4 chips = 512K words of 32 bits each

3. Structure of large memories: put a few sets together

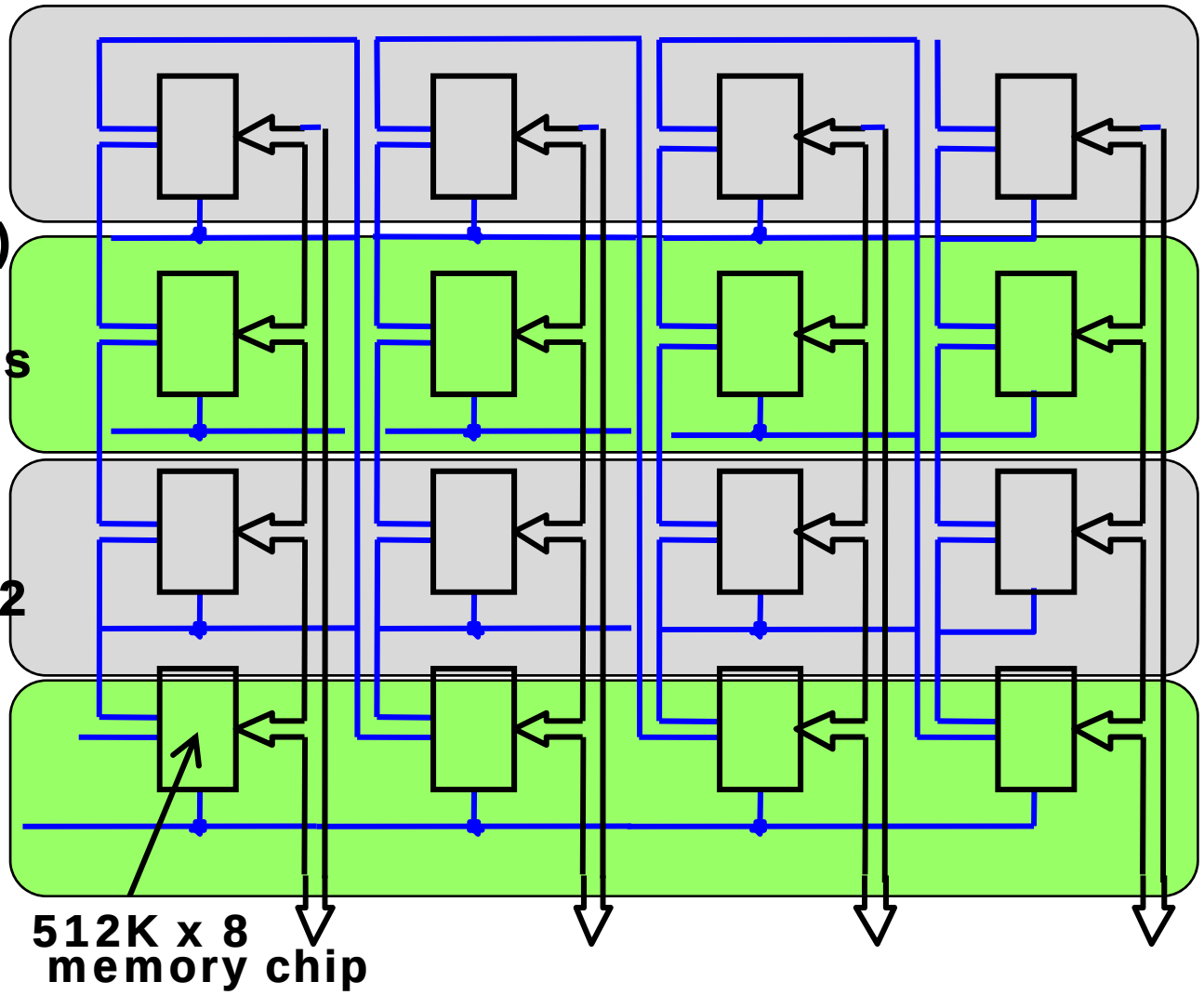
Put 4 horizontal sets together:

512K rows x 4

(512K x 4 = 1024K x 2 = 2M)

= 2M words/rows of 32 bits each

i.e 2^{21} rows of 32 bits each



Addressing?

for each row of 4 chips to make 32-bit word:

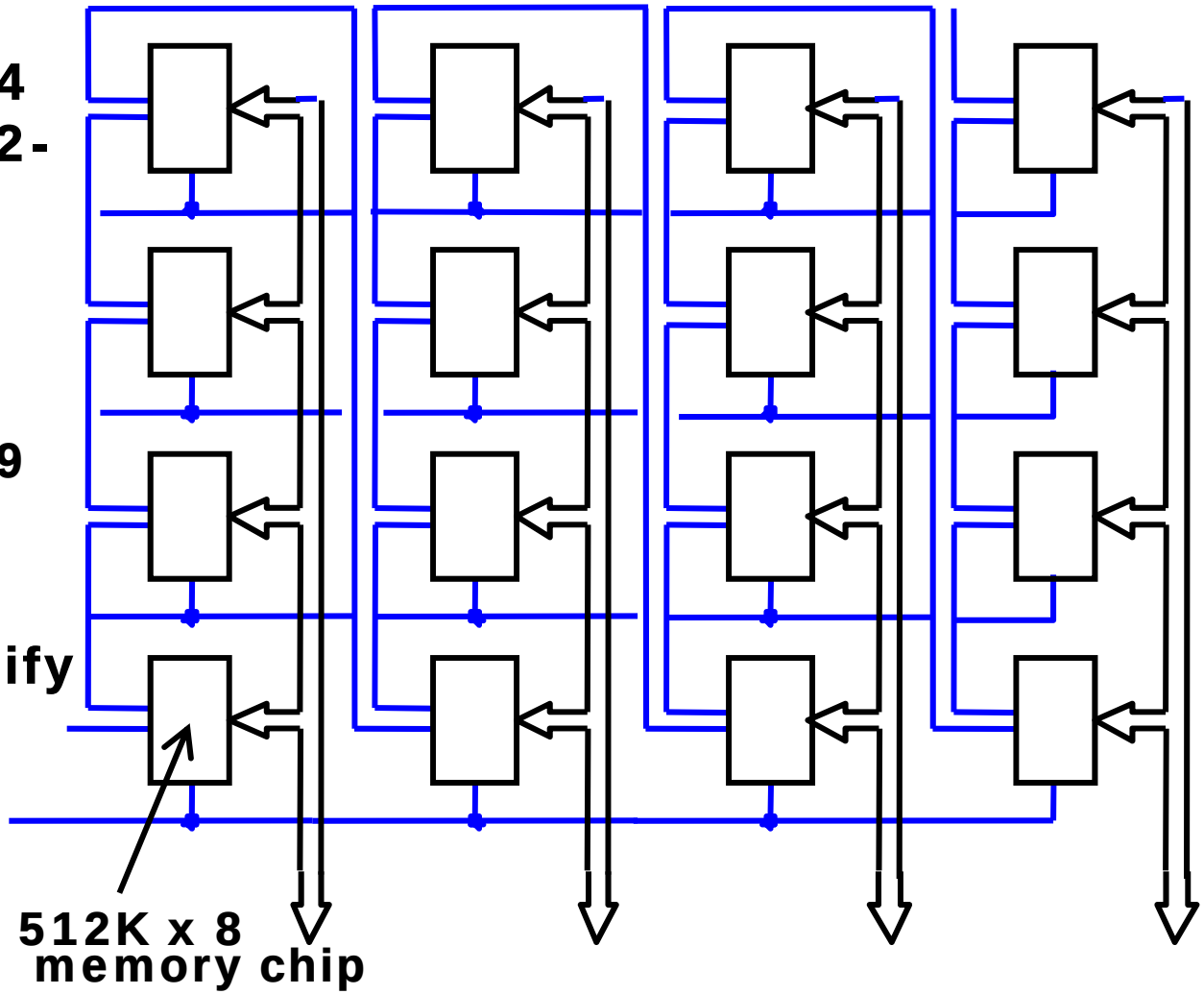
512K rows

= 512 x 1024

= 2^9 x 2^{10} = 2^{19}

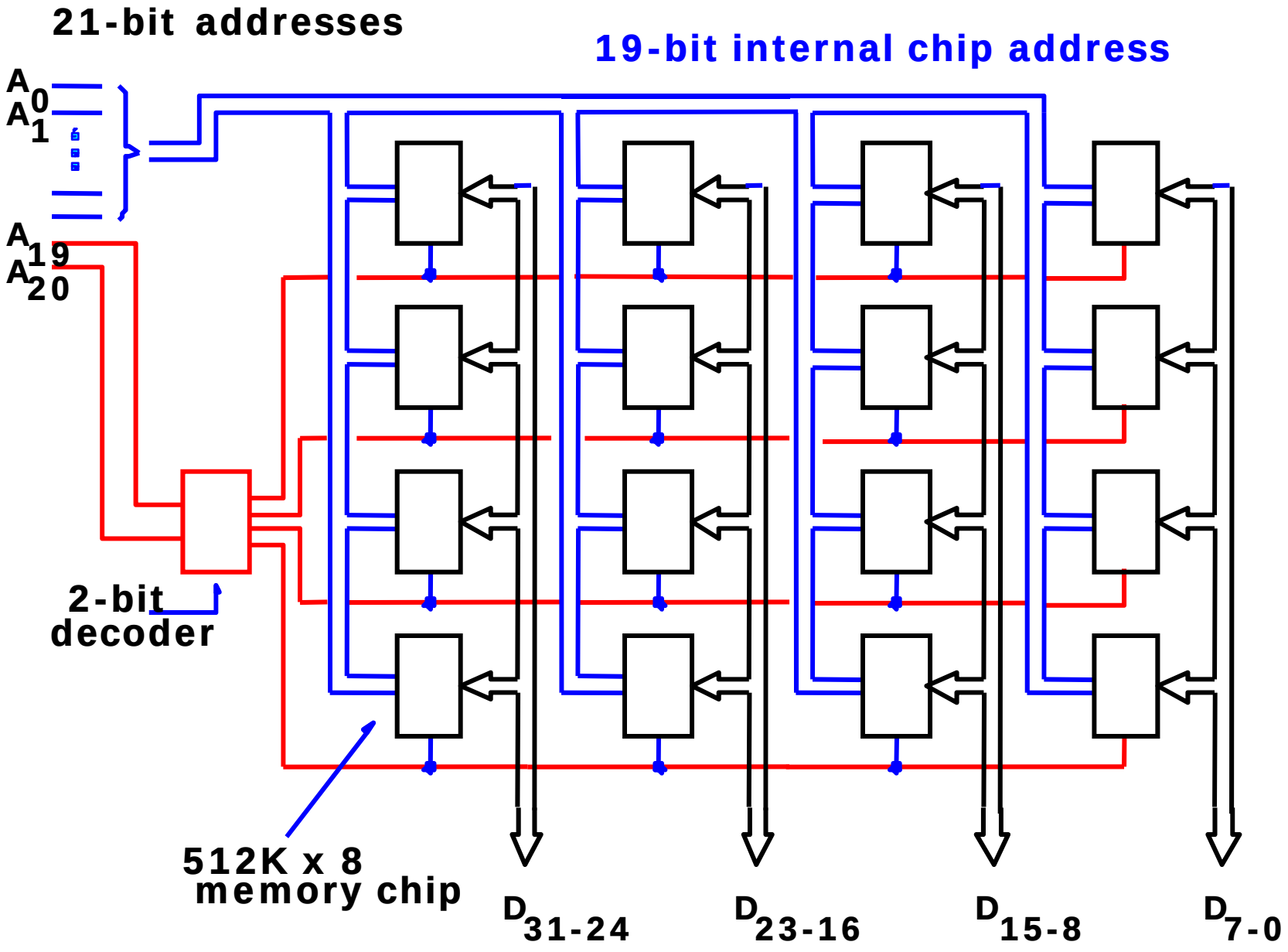
➔need 19-bit

address to identify row within each chip



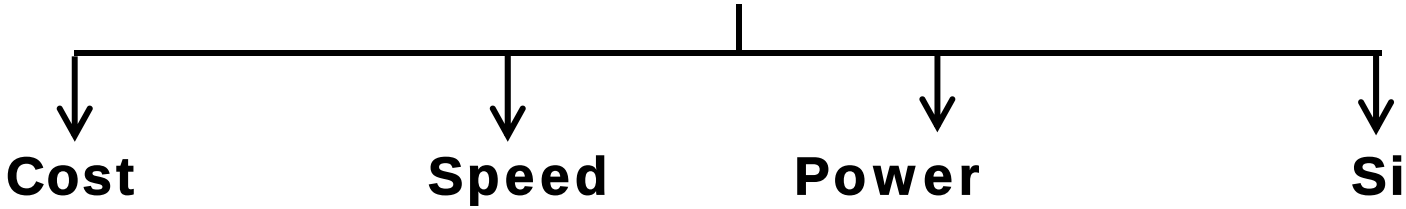
4 horizontal sets ➔ need 2-bit address to choose set

➔ total needed = 21-bit address



Organization of a 2M × 32 memory module with 512K × 8 SRAM

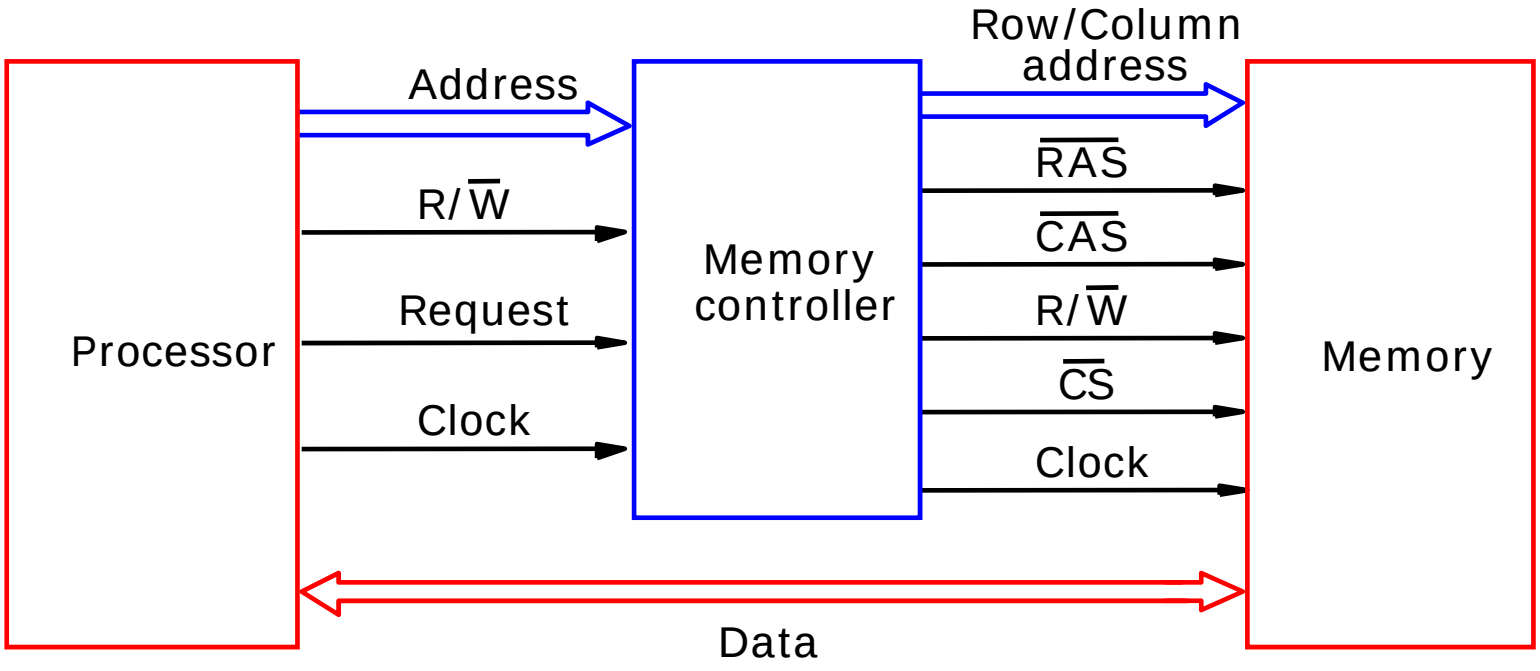
Choice of RAM



STATIC RAM ➔ **when speed is needed** ➔ **Cache**

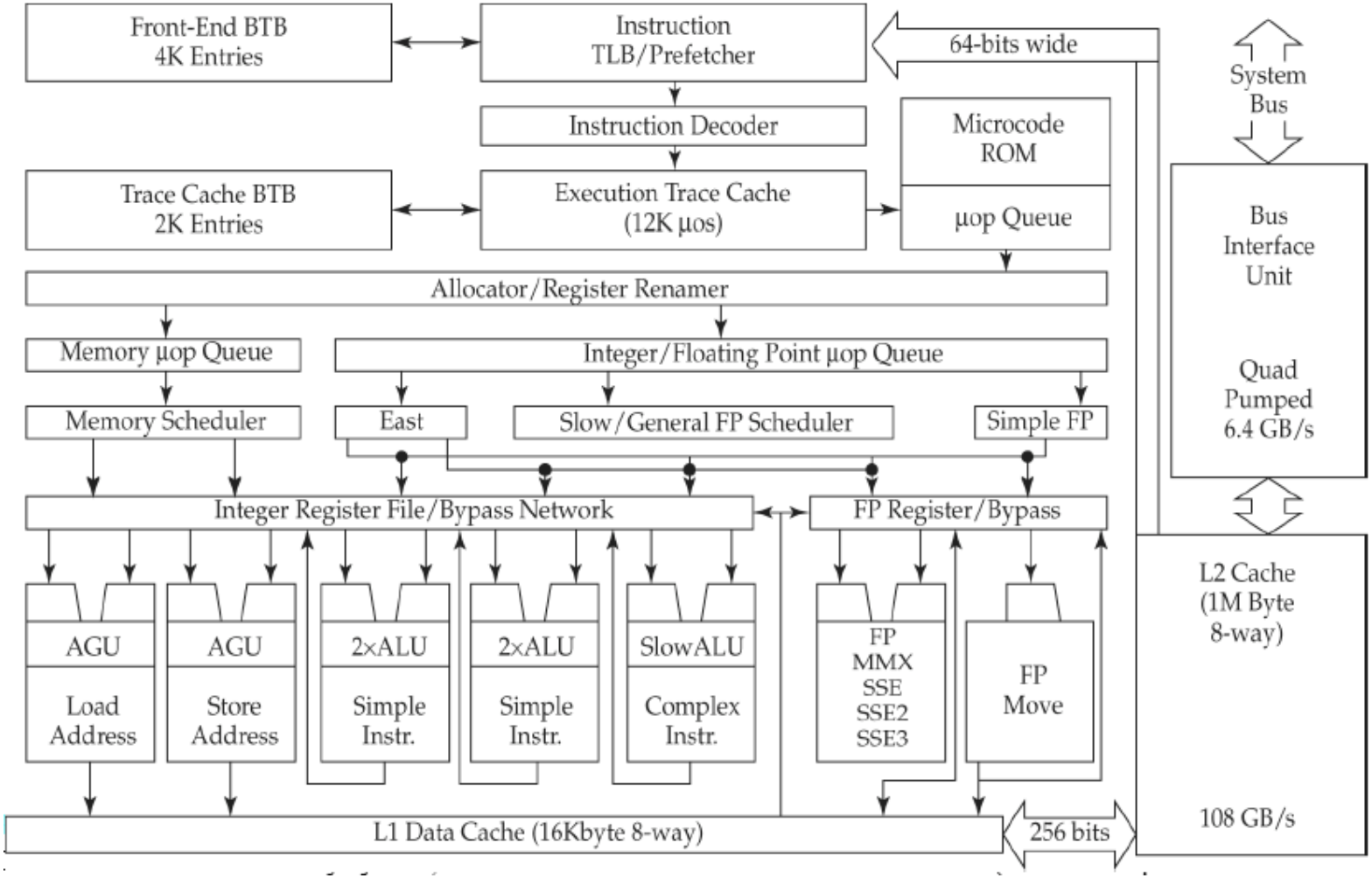
DYNAMIC RAM ➔ **density desired** ➔ **main memory**

Processor: **does not need to know and issues the same k-bit address**

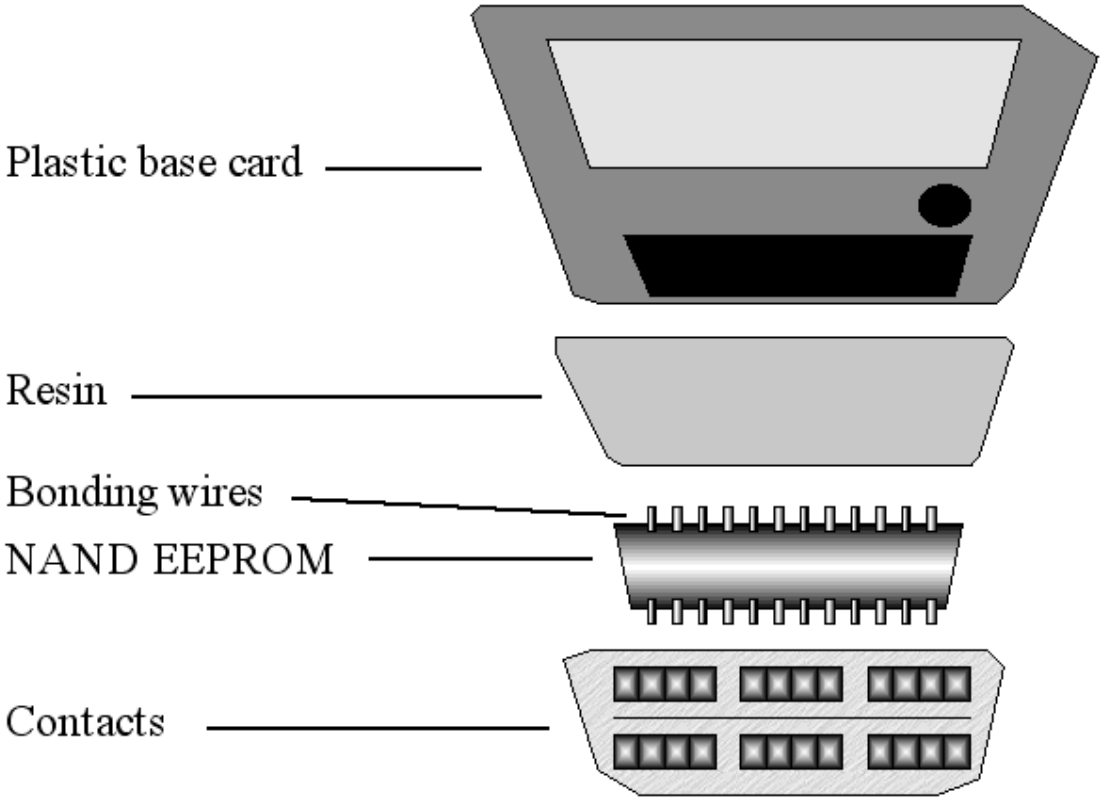


Memory technology	Typical access time	$ per GB in 2004
SRAM	0.5 – 5 ns	$4,000 - $10,000
DRAM	50 – 70 ns	$100 - $200
Disk	5,000,000 – 20,000,000 ns	$0.50 - $2

The Intel 4 Pentium Memory System



Flash Memory

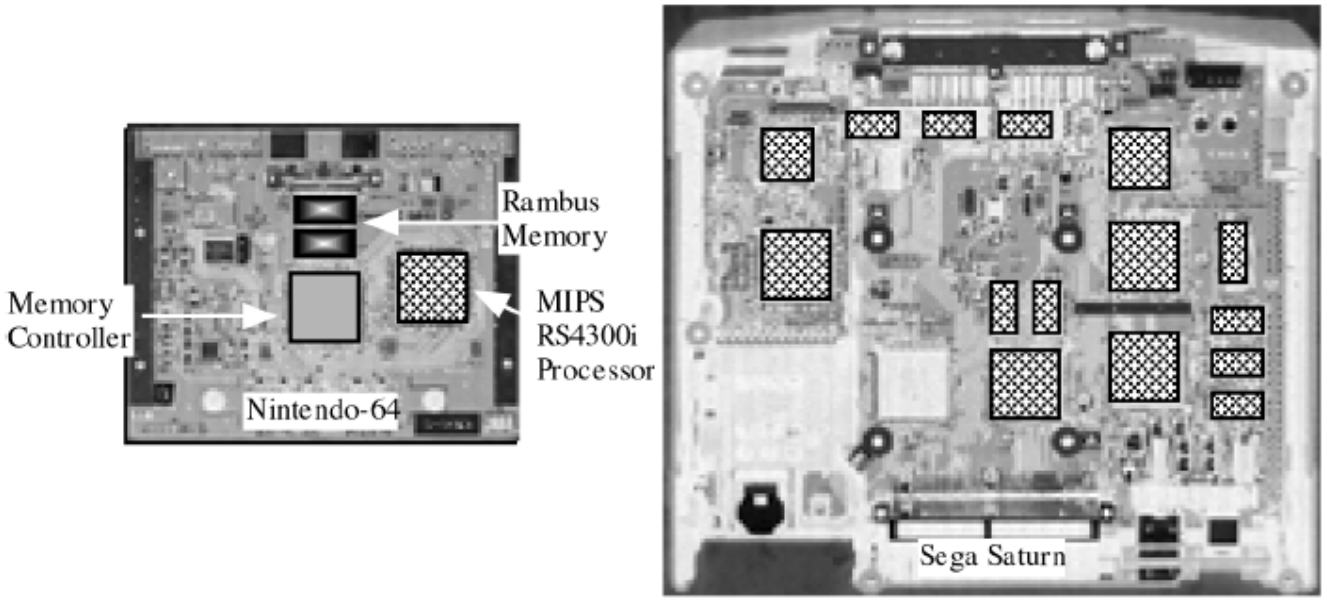


(a) (b)

- (a) External view of flash memory module and (b) flash module internals. (Source: adapted from HowStuffWorks.com.)

Rambus Memory

- Rambus technology on the Nintendo 64 motherboard (left) enables cost savings over the conventional Sega Saturn motherboard design (right).



- Nintendo 64 game console: