Chapter 1
Introduction to Memories

Jin-Fu Li
Advanced Reliable Systems (ARES) Lab.
Dept. of Electrical Engineering
National Central University
Jhongli, Taiwan

Outline

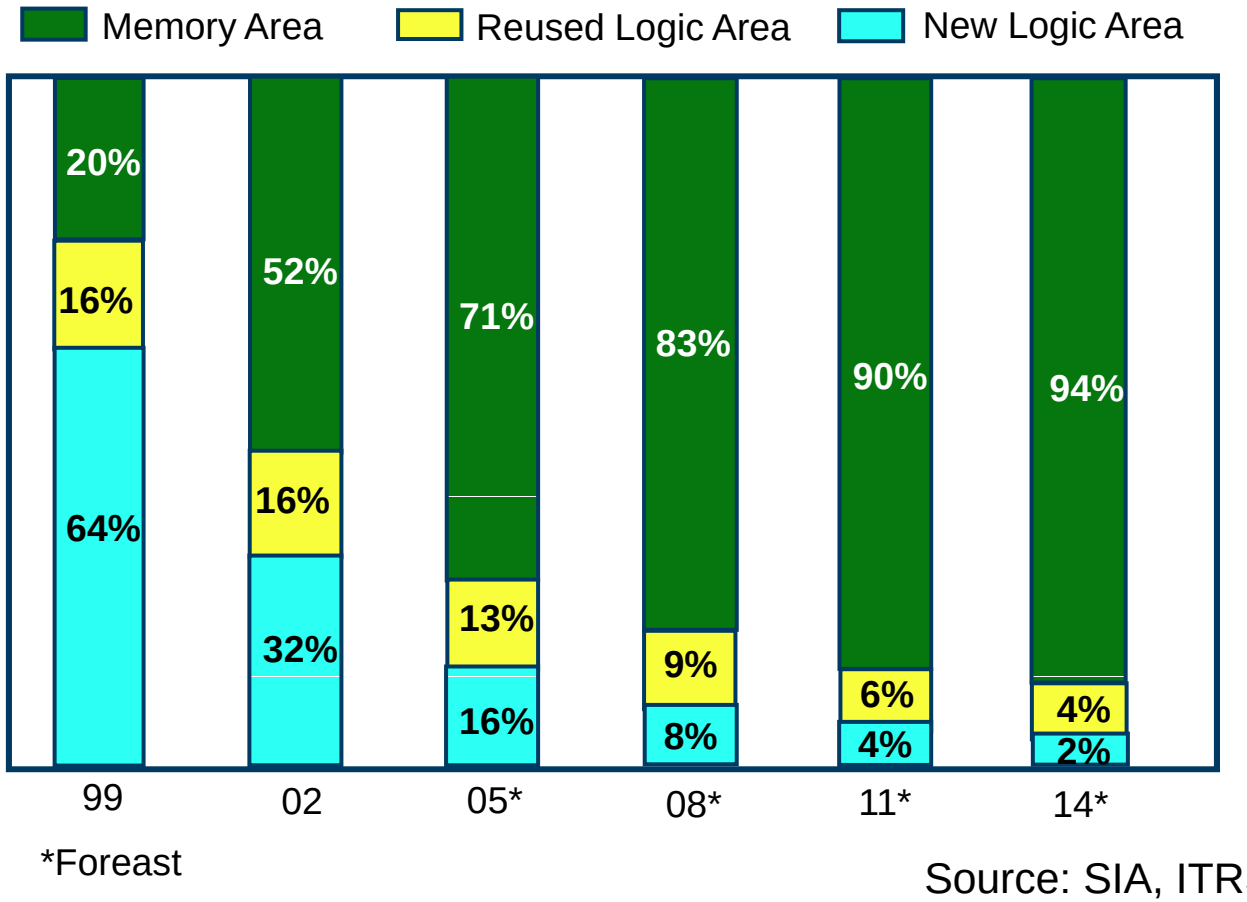
- Importance of Embedded Memories
- Overview of Memory Structures

Embedded Memory – The Key to SOC

- Embedded memory is becoming more central to integrated circuit design
- Historically, ICs were dominated by the logic functions, with memory being external
- Today, an SOC contains many memory blocks of different sizes, shapes and functionality
  - Typically, embedded memories represent about 30%~50% SOC area
- The Semiconductor Industry Association (SIA) predicts that 90% of the SOC's surface will be memory in 2011

Embedded Memory – The Key to SOC

- SOC memory continues to increase



Source: SIA, ITRS, 2000

Embedded Memory – Advantages

- Area
  - Embedding multiple memories on a single SOC reduces the amount of silicon used
- Performance
  - Embedding faster, wider memories and moving them closer to processor can increase system performance substantially
- Power
  - Embedded memories eliminate the need to drive off-chip capacitance between the stand-alone memories and other system chips
- Design reuse
  - By reusing embedded memories, system designers can significantly reduce develop time and cost

Embedded Memory – Quality

- During manufacture
  - Yield
  - Exponential yield model
  - $Y = e^{-\sqrt{AD}}$, where A and D denote the area and defect density, respectively
- After manufacture
  - Reliability
- During use
  - Soft error rate

Quality During Manufacture

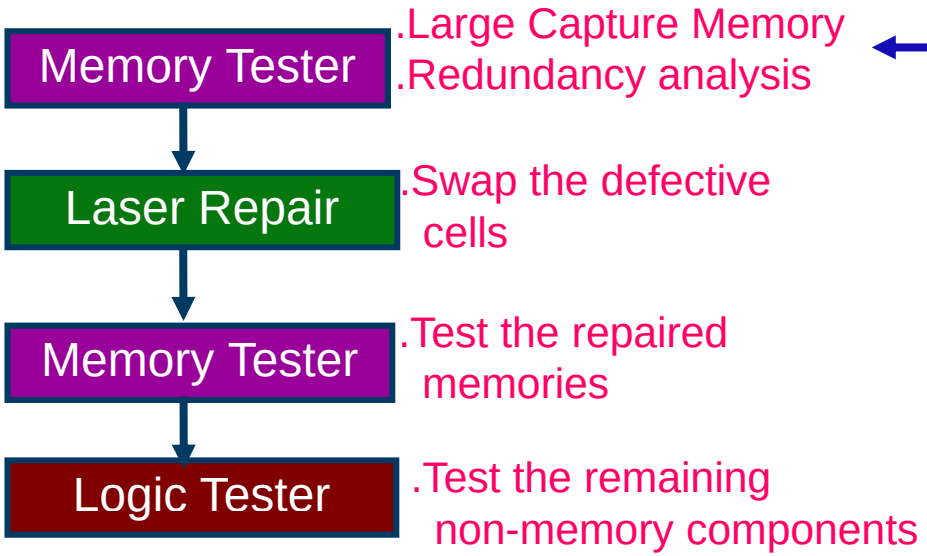
- Issues of embedded memory
  - As the density of transistors is increased, the D is increased compared to logic
  - About 2X logic for high density 6-T SRAM
- Solutions
  - Redundancy & laser repair using ATE
  - Error correction code (ECC)
- Redundancy & repair
  - Achieve yield parity with logic, or better
  - About 3% area overhead
  - Recommended over 1Mb

Quality During Manufacture

- Corrects <5 defects per Mb
- Laser repair manufacturing flow

- ECC
  - Detect/repair defects in individual words
  - 25-30% area overhead
    - Ex: 6 extra bits for single bit correction in 32 bit words
  - Latency penalty
    - At least one clock cycle latency penalty

Quality-Insurance Strategies for Memories

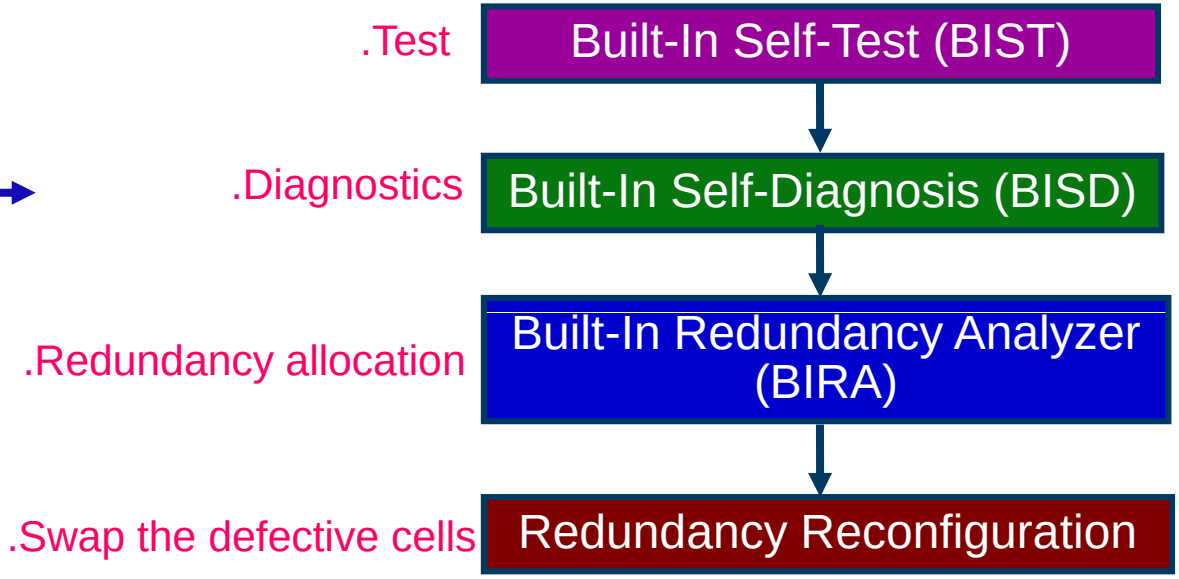


Conventional test and repair approaches

1. Long testing time
2. Expensive memory tester and logic tester are needed

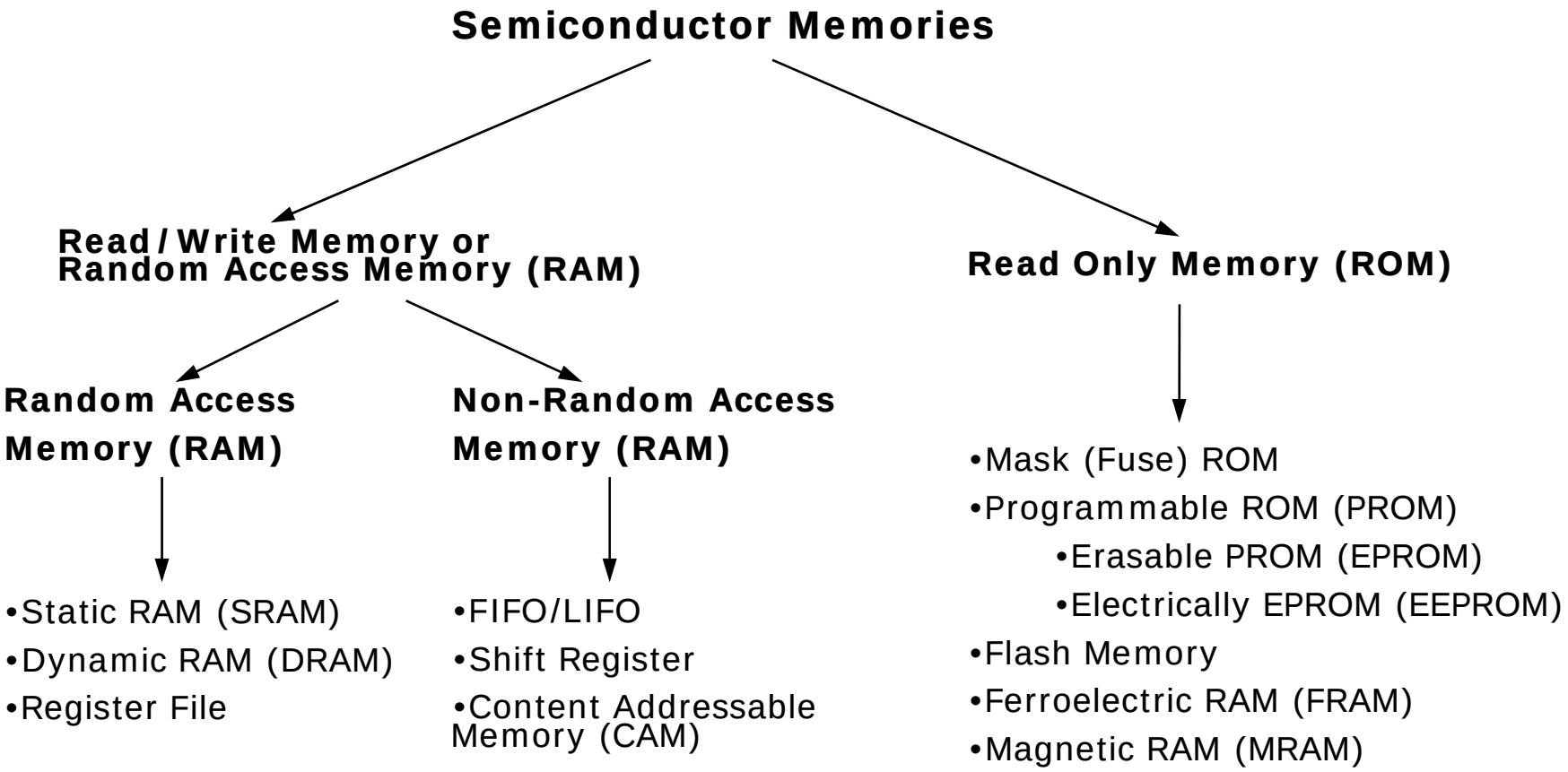
BIST & BISR approaches

1. Short testing time
2. Only cheap logic tester is needed



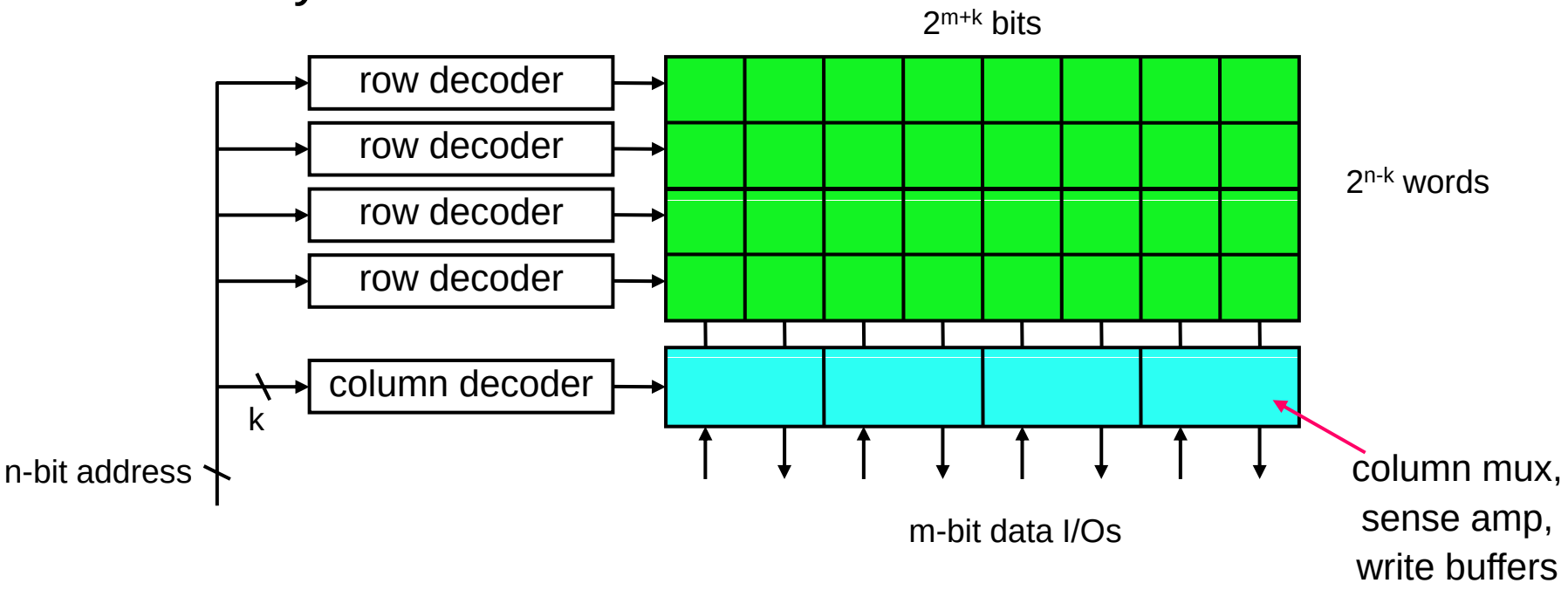
Introduction to Memories

- Overview of semiconductor memory types

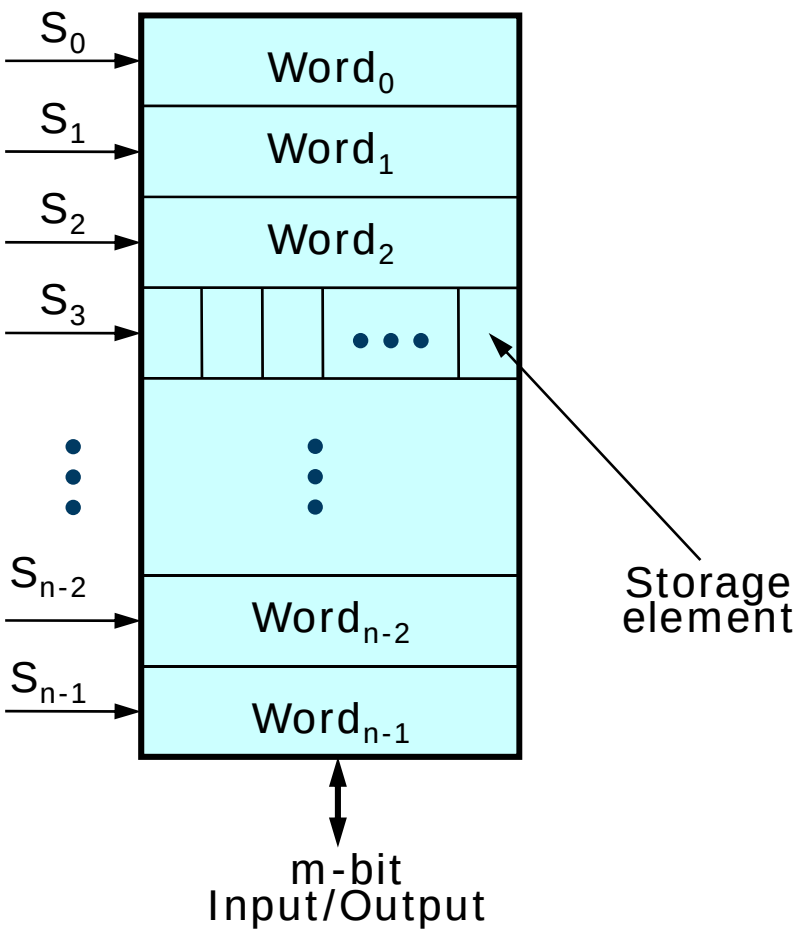


Memory Elements – Memory Architecture

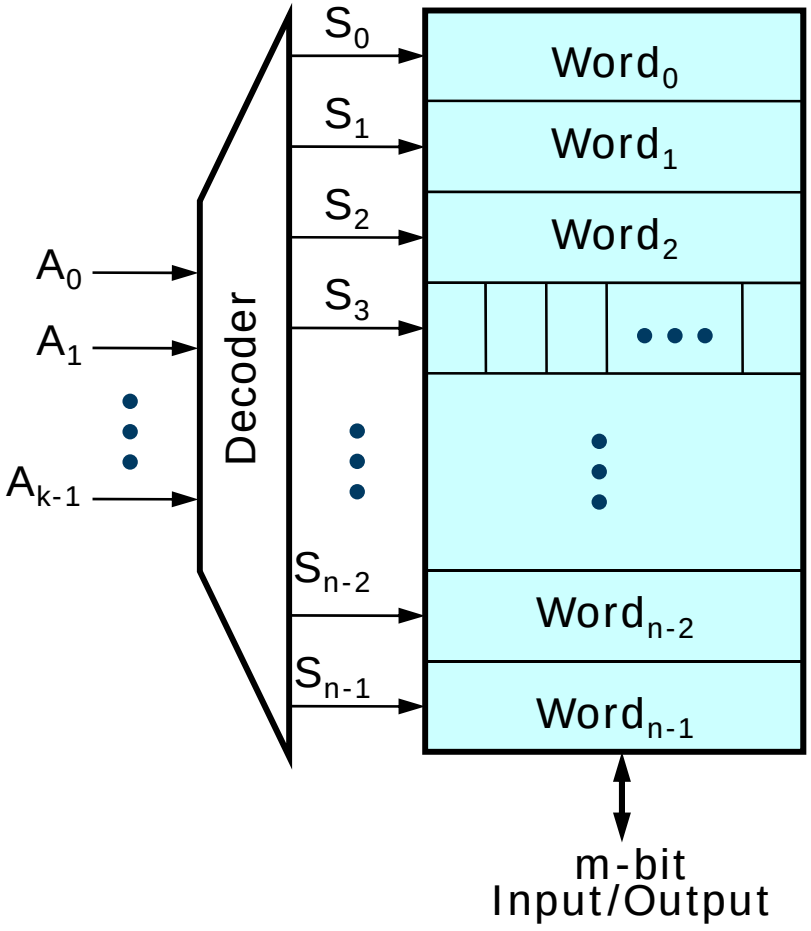
- Volatile memories may be divided into the following categories
  - Random access memory
  - Serial access memory
  - Content addressable memory
- Memory architecture



1-D Memory Architecture

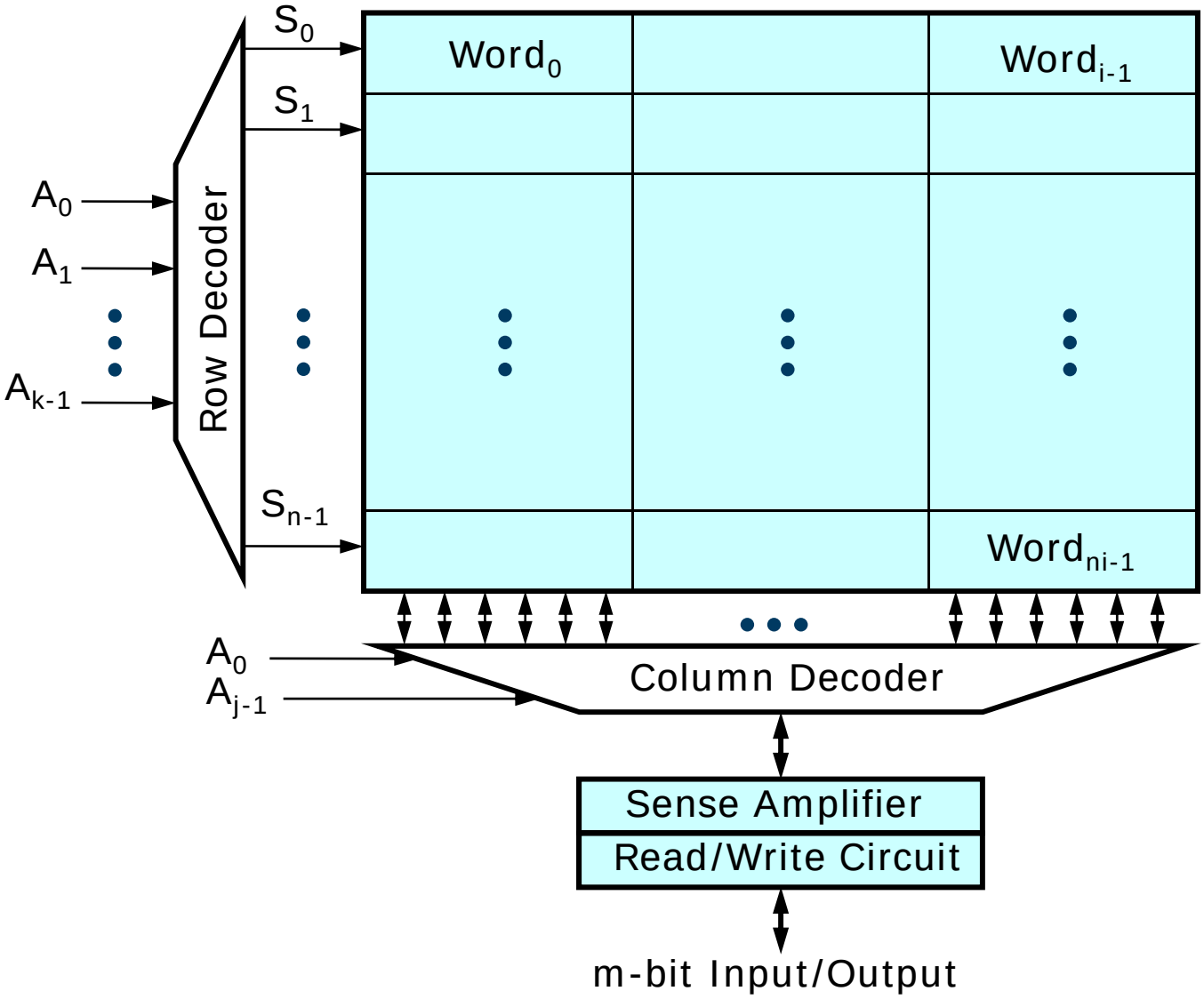


n select signals: S_0-S_{n-1}

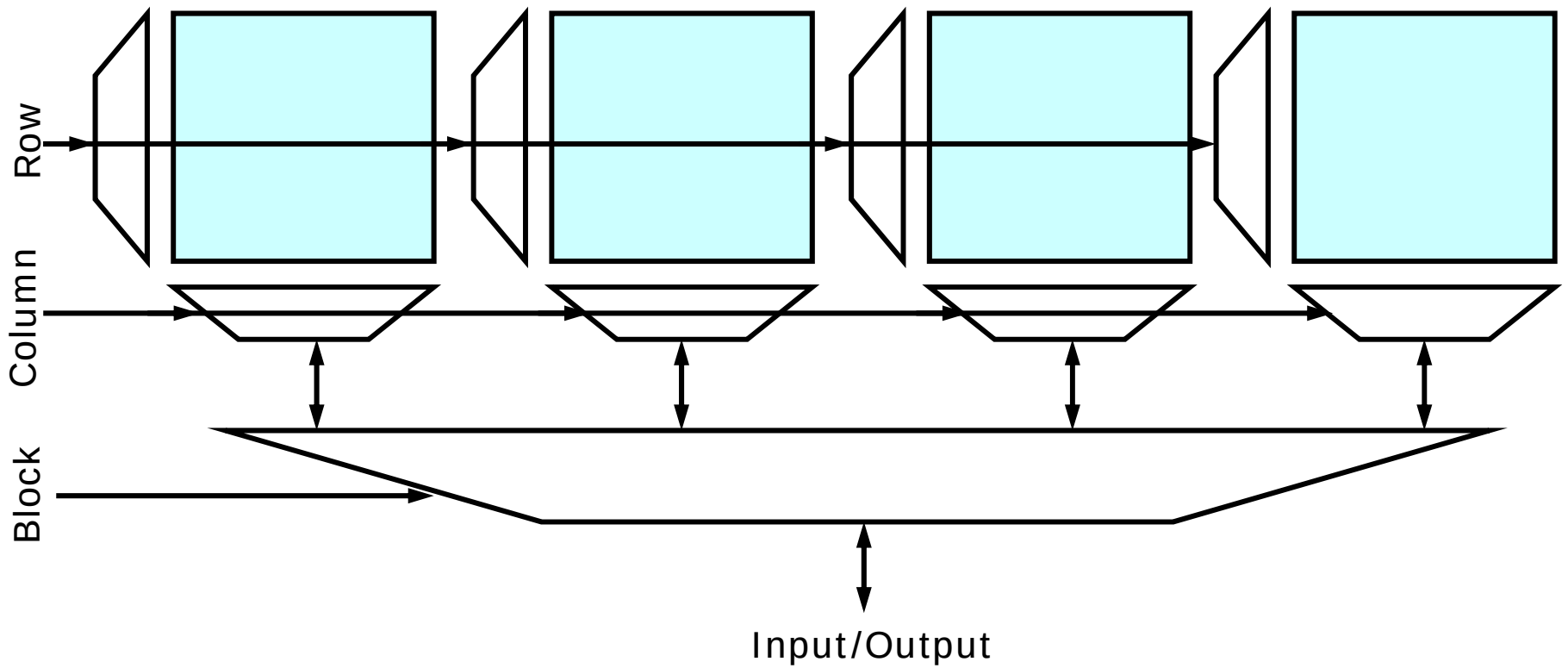


n select signals are reduced to k address signals: A_0-A_{k-1}

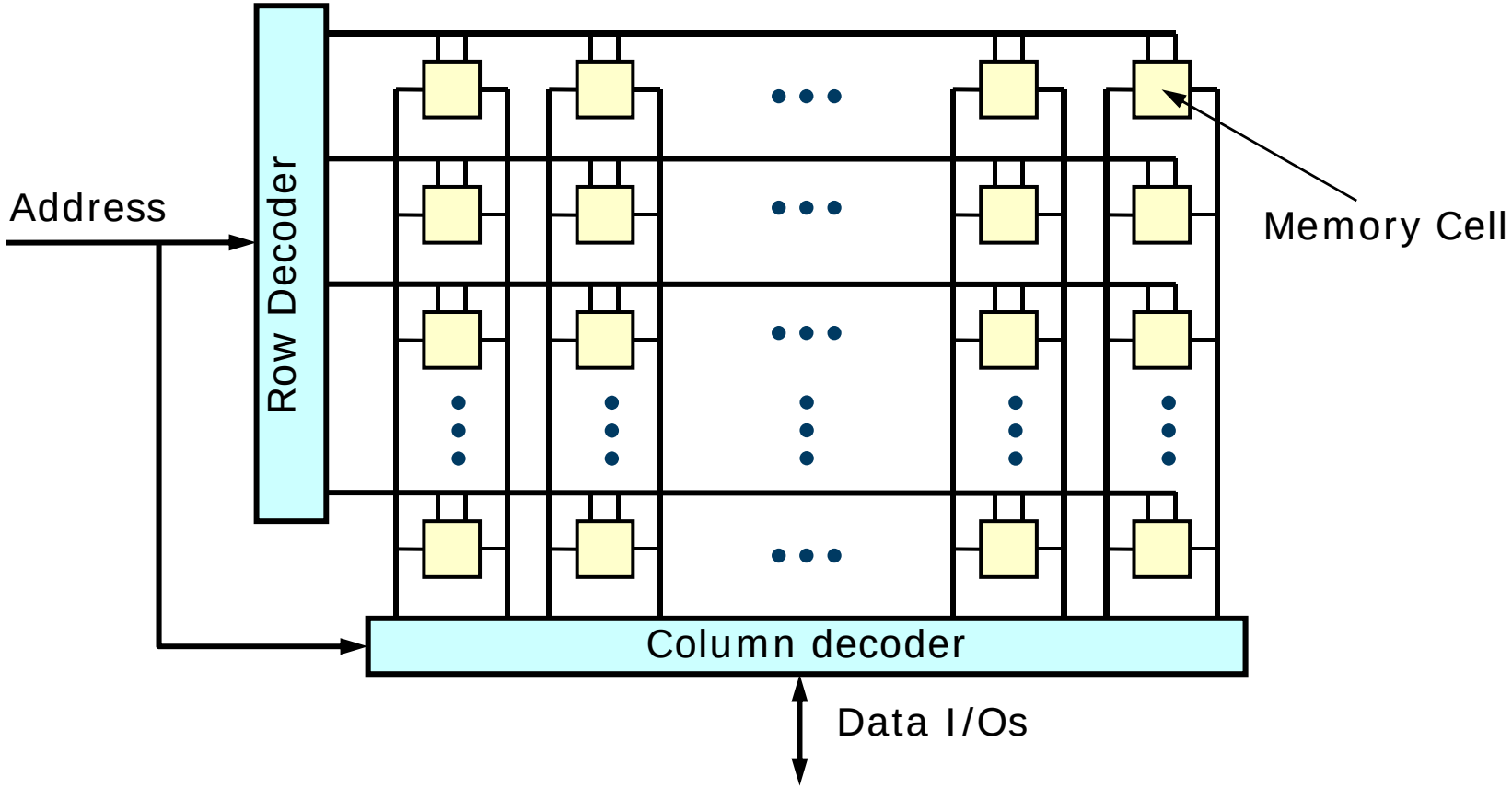
2-D Memory Architecture



3-D Memory Architecture

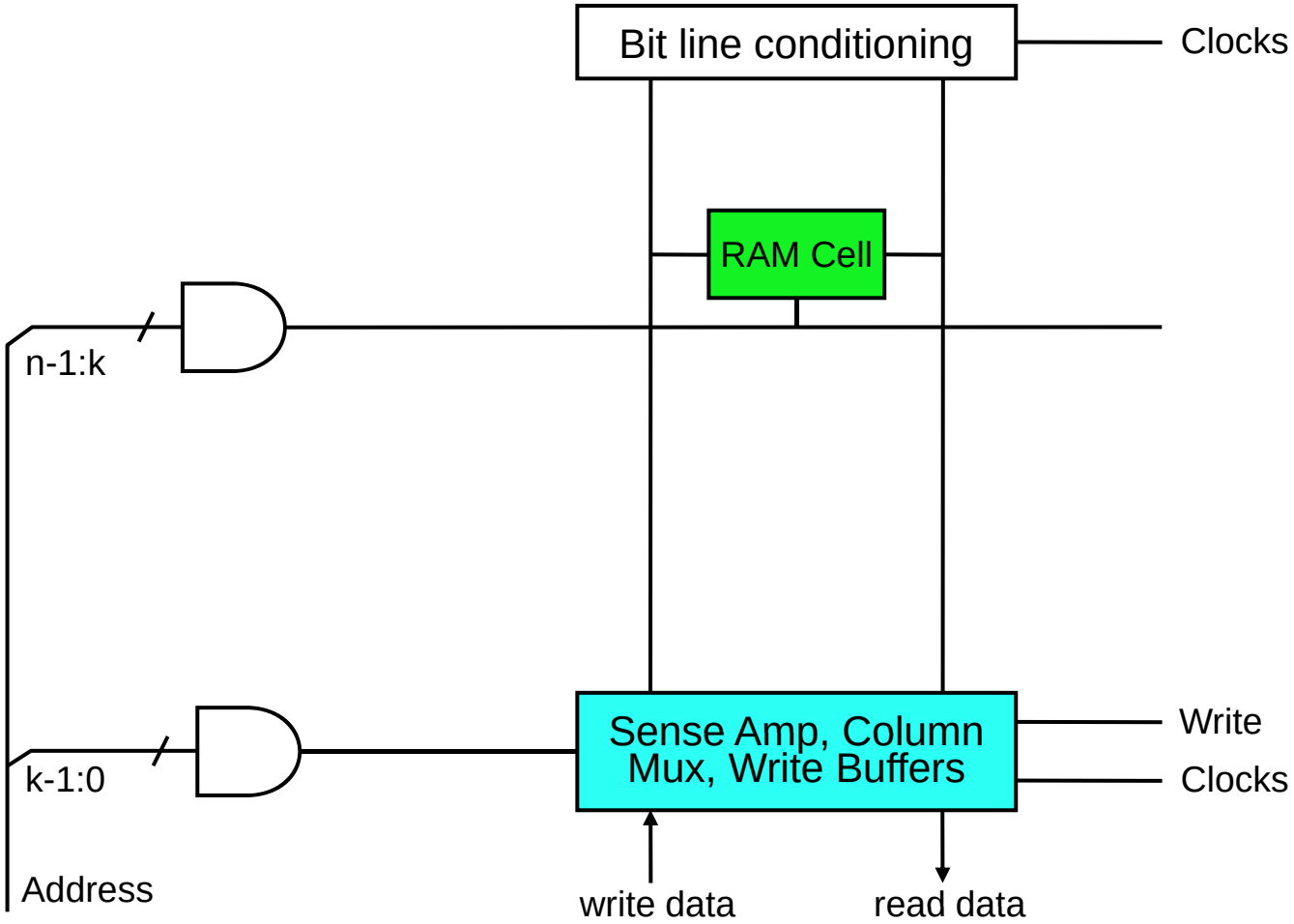


Conceptual 2-D Memory Organization



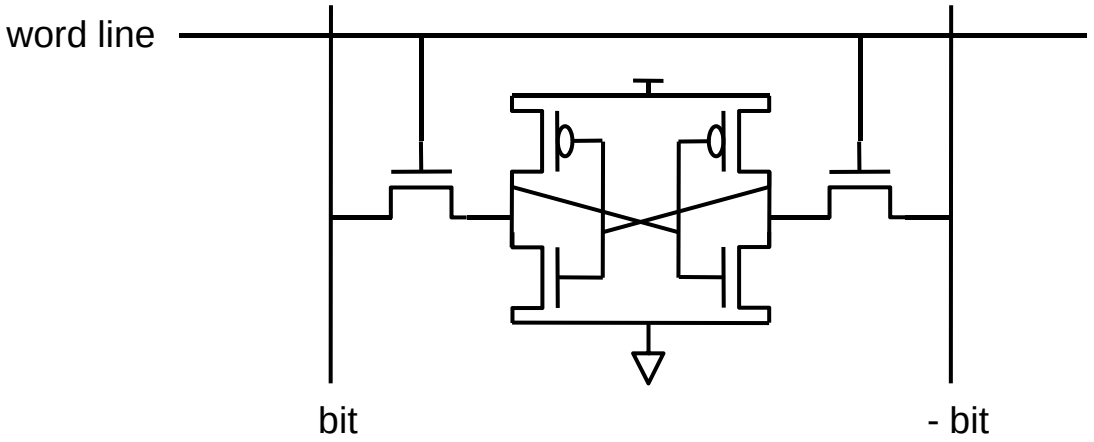
Memory Elements – RAM

Generic RAM circuit

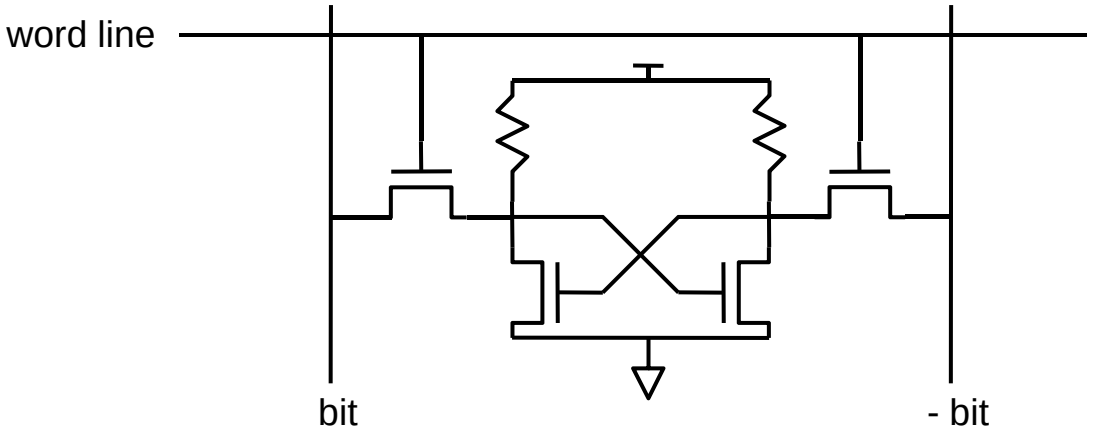


Memory Elements – RAM Cells

6-T SRAM cell

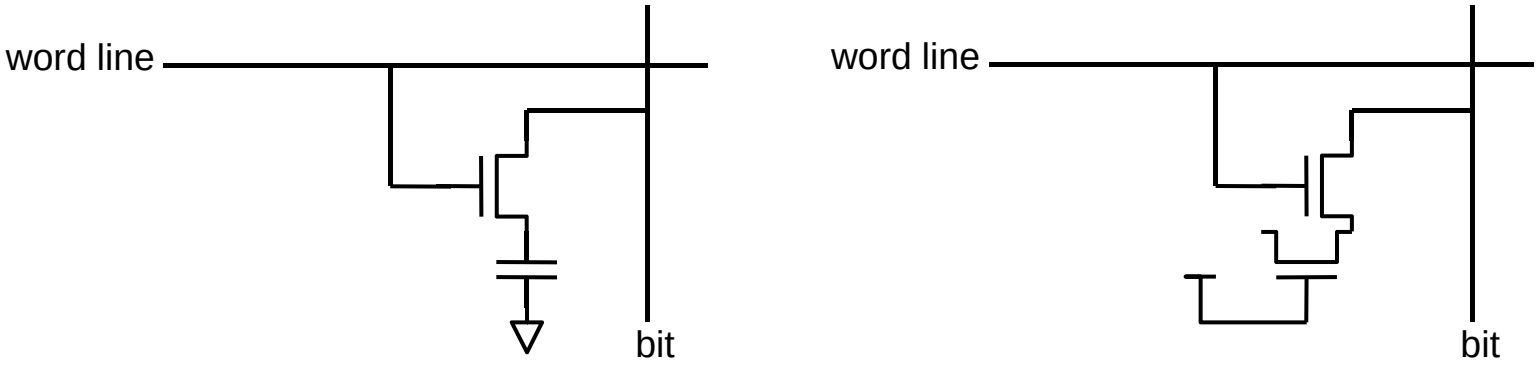


4-T SRAM cell

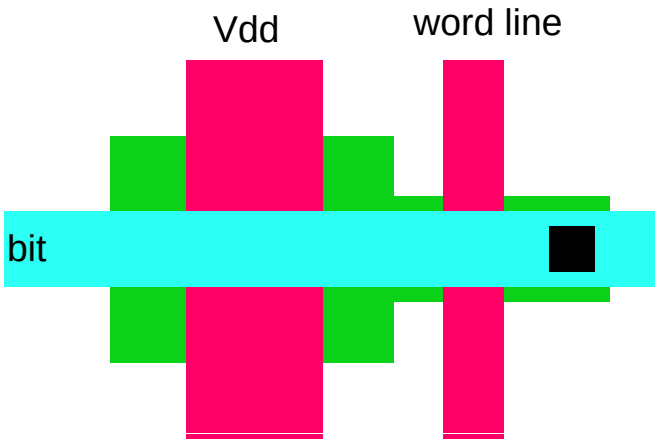


Memory Elements – RAM Cells

1-T DRAM cell

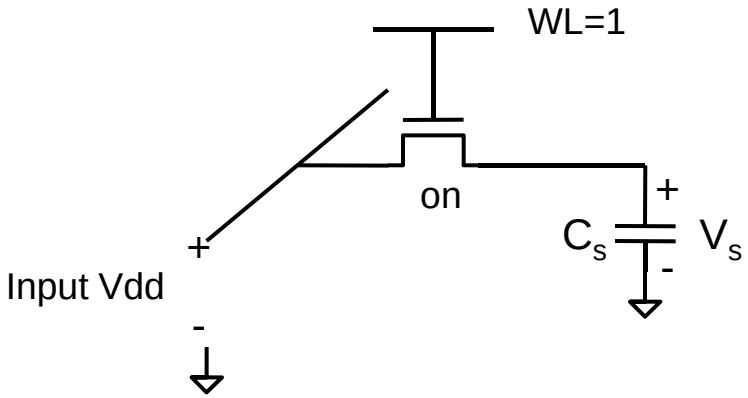


Layout of 1-T DRAM (right)

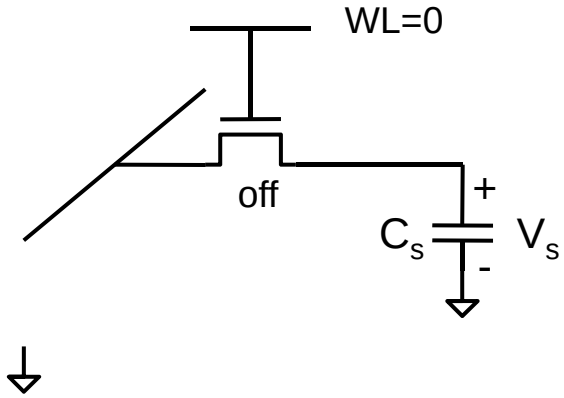


Memory Elements – DRAM Retention Time

Write and hold operations in a DRAM cell



Write Operation



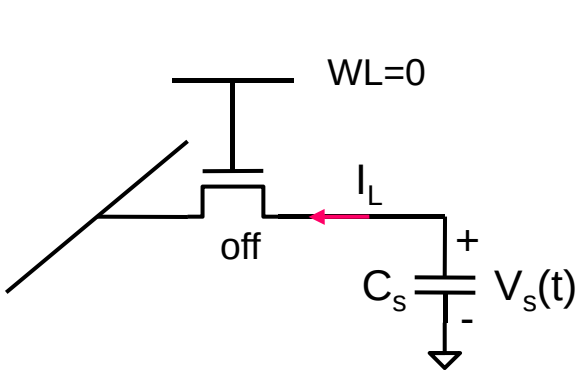
Hold

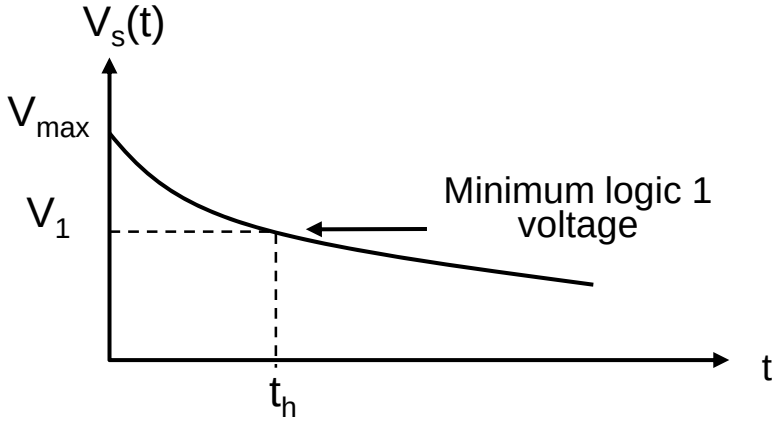
$$V_s = V_{\max} = V_{DD} - V_{tn}$$

$$Q_{\max} = C_s (V_{DD} - V_{tn})$$

Memory Elements – DRAM Retention Time

Charge leakage in a DRAM Cell





$$I_L = -\left(\frac{dQ_s}{dt}\right)$$

$$I_L = -C_s\left(\frac{dV_s}{dt}\right)$$

$$I_L \approx -C_s\left(\frac{\Delta V_s}{\Delta t}\right)$$

$$t_h = |\Delta t| \approx \left(\frac{C_s}{I_L}\right)\Delta V_s$$

Memory Elements – DRAM Refresh Operation

As an example, if IL=1nA, Cs=50fF, and the difference of Vs is 1V, the hold time is

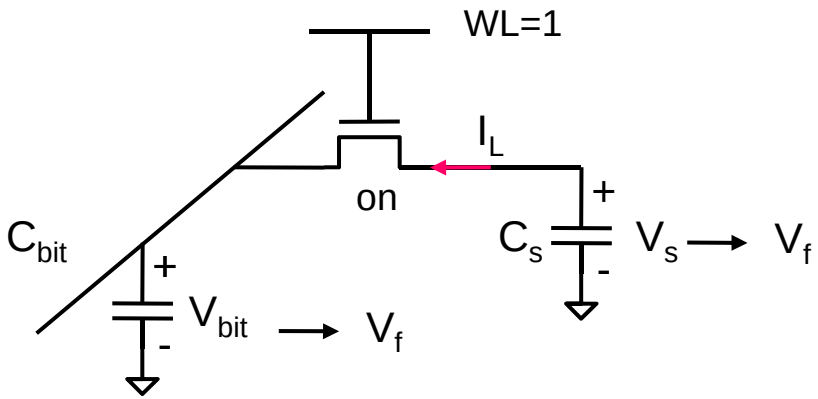
$$t_h = \frac{50 \times 10^{-15}}{1 \times 10^{-9}} \times 1 = 0.5 \mu s$$

Memory units must be able to hold data so long as the power is applied. To overcome the charge leakage problem, DRAM arrays employ a **refresh operation** where the data is periodically read from every cell, amplified, and rewritten.

The refresh cycle must be performed on every cell in the array with a minimum refresh frequency of about

$$f_{refresh} \approx \frac{1}{2t_h}$$

Memory Elements – DRAM Read Operation



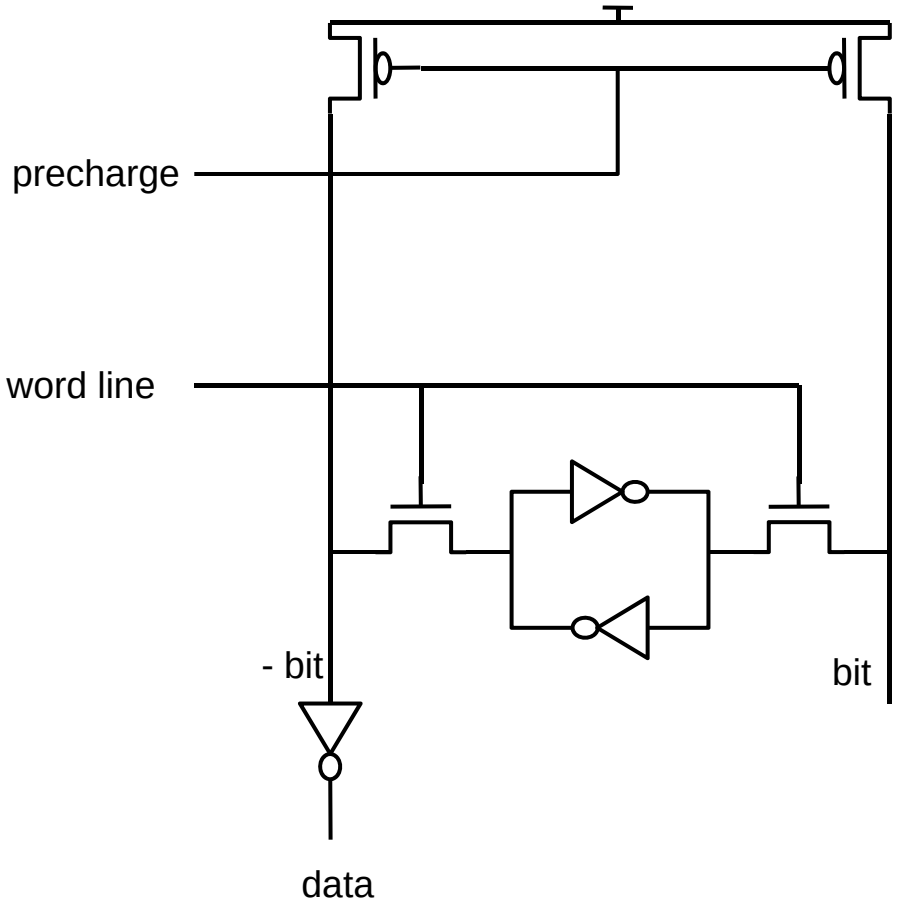
$$Q_s = C_s V_s$$

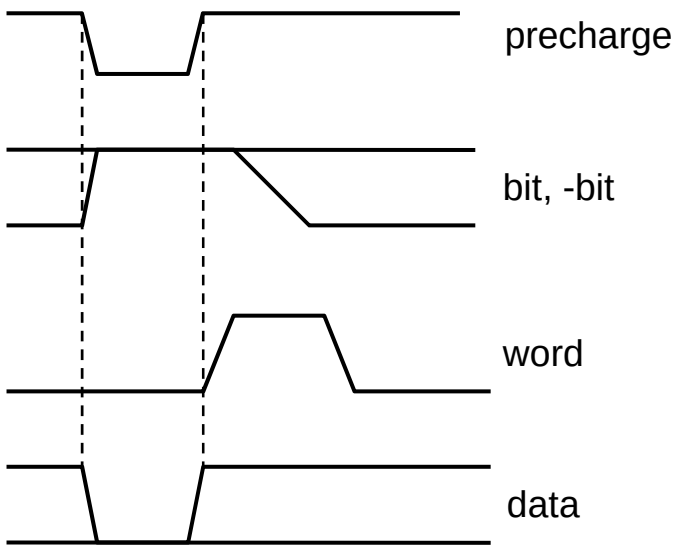
$$Q_s = C_s V_f + C_{bit} V_f$$

$$V_f = \left(\frac{C_s}{C_s + C_{bit}}\right) V_s$$

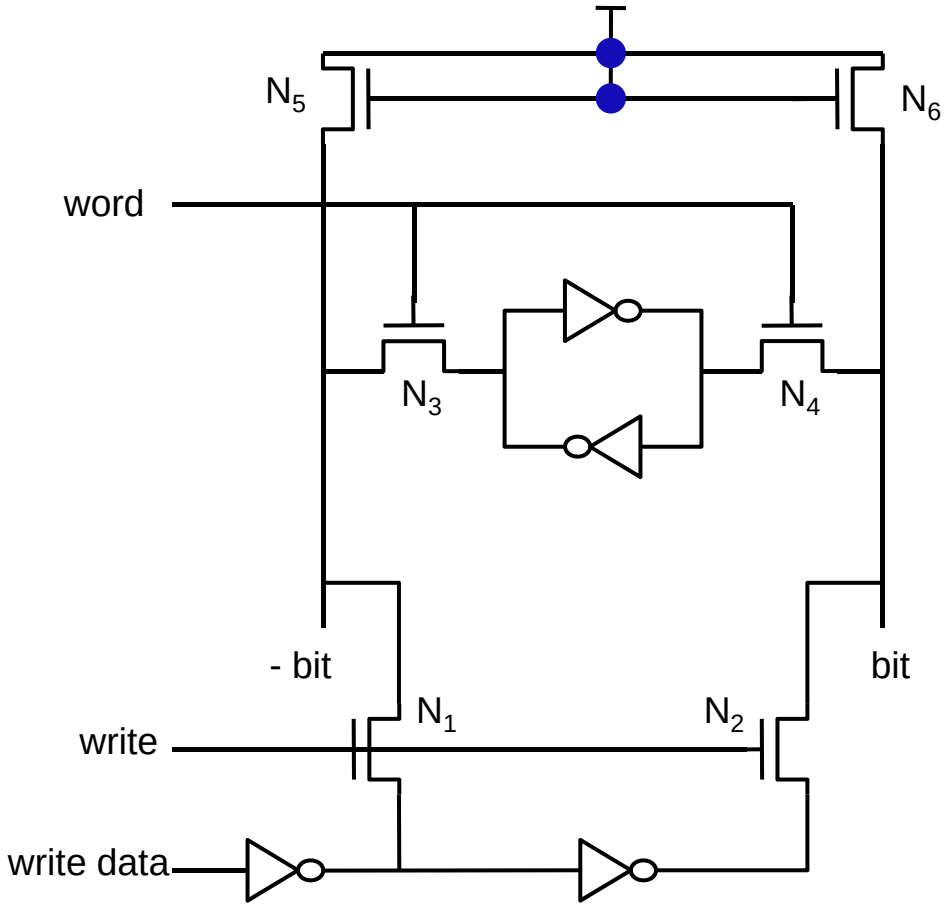
This shows that $V_f < V_s$ for a store logic 1. In practice, V_f is usually reduced to a few tenths of a volt, so that the design of the sense amplifier becomes a critical factor

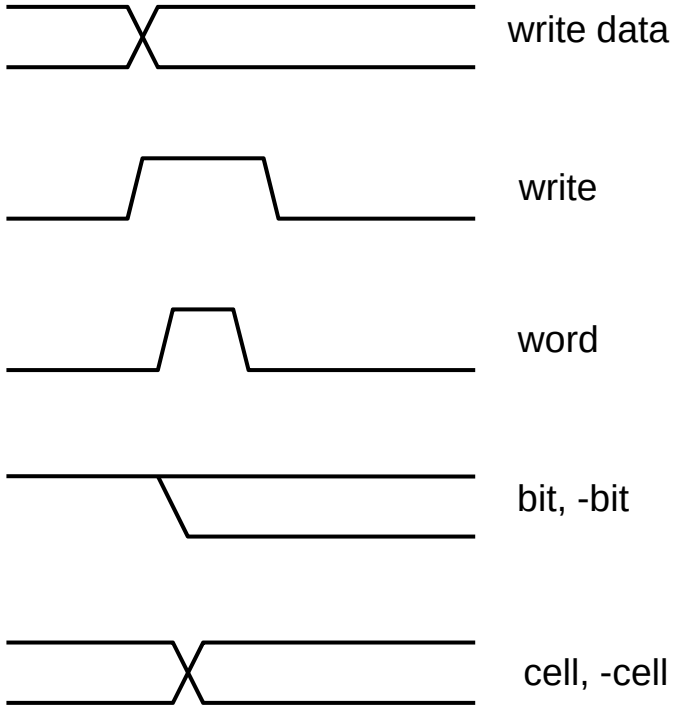
Memory Elements – RAM Read Operation



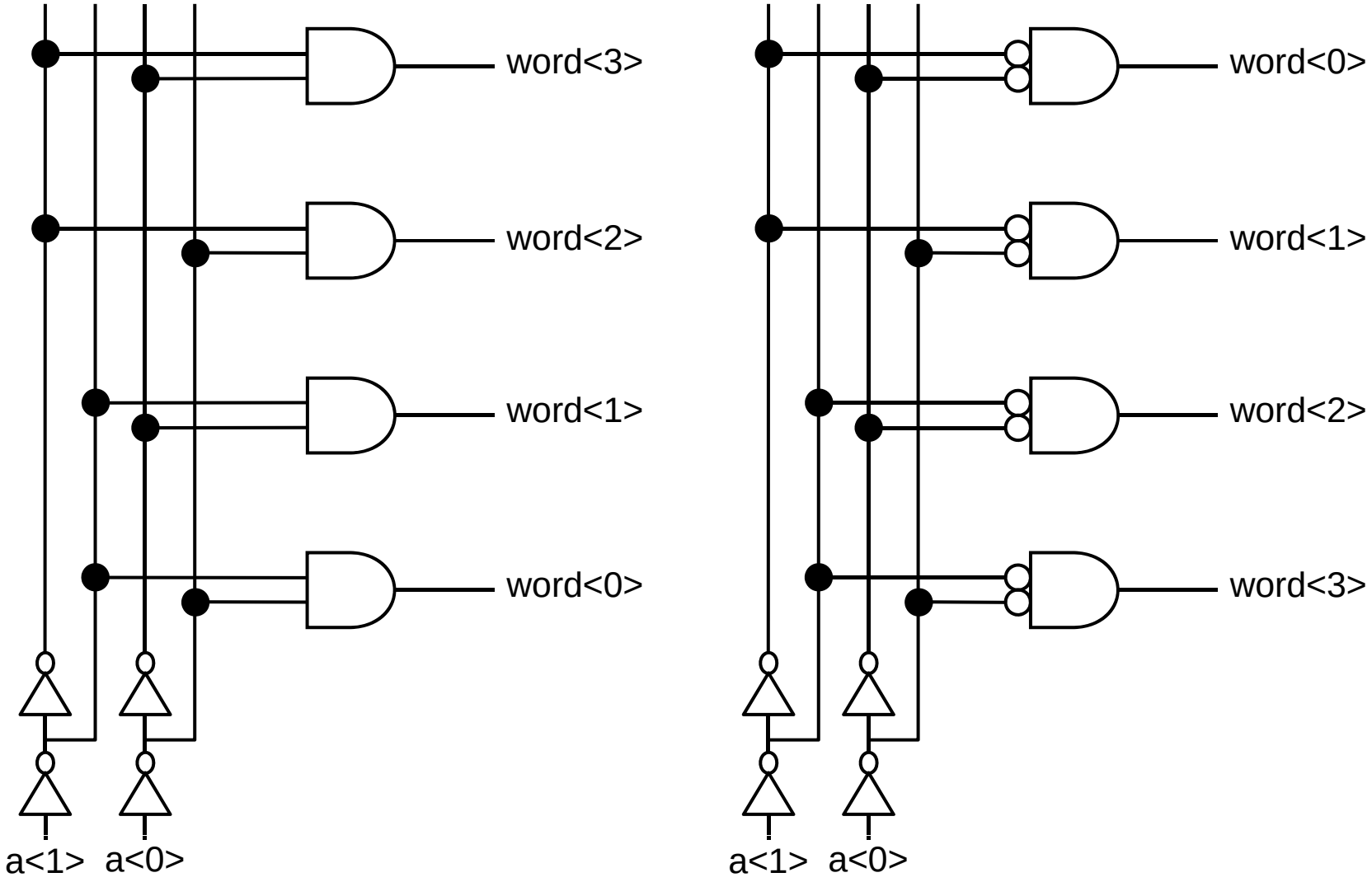


Memory Elements – RAM Write Operation



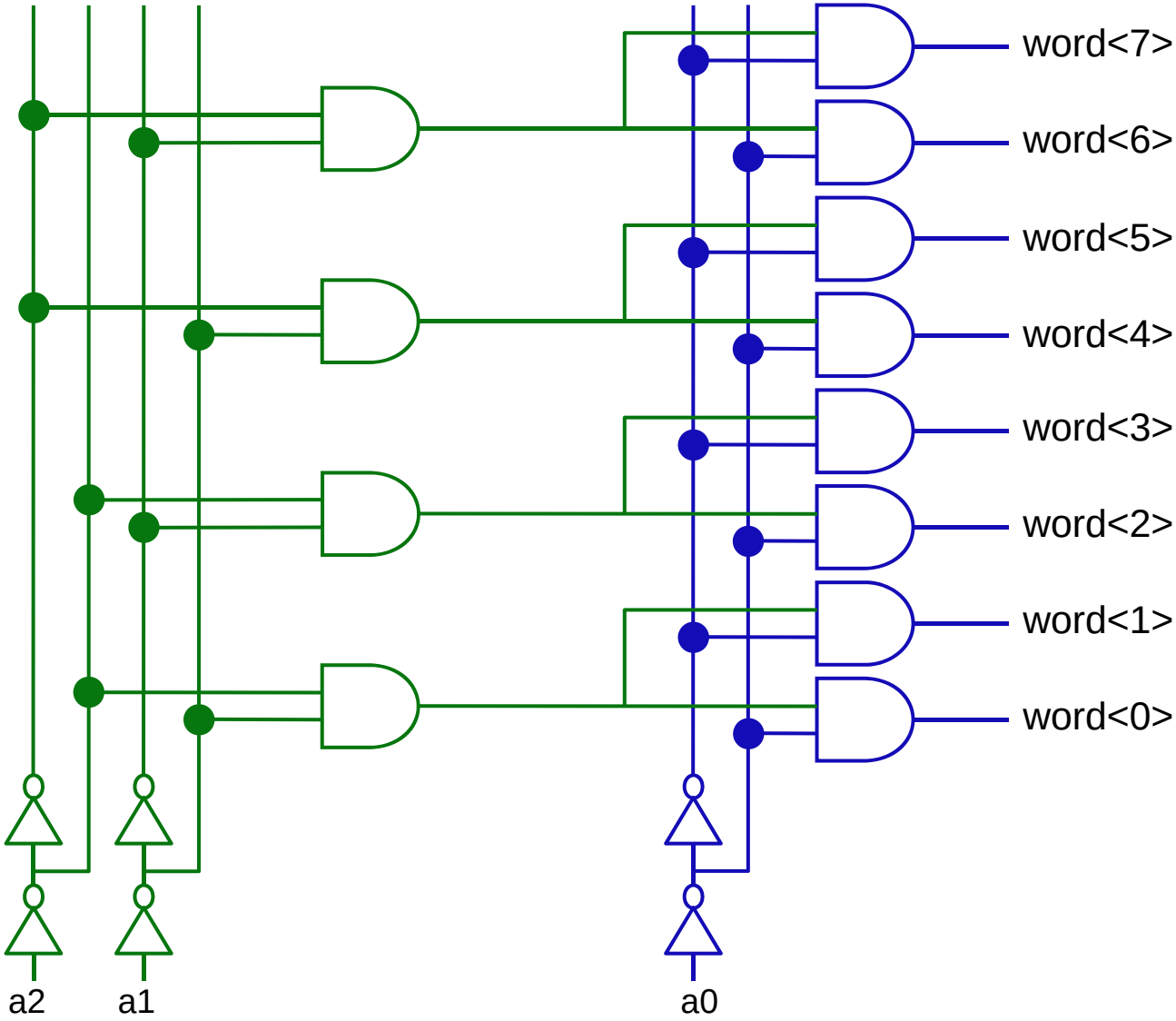


Memory Elements – Row Decoder



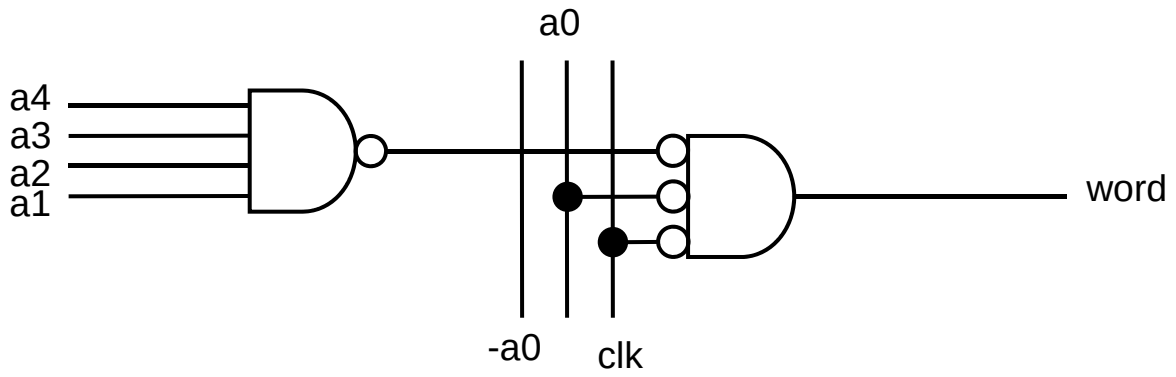
Memory Elements – Row Decoder

Predecode circuit

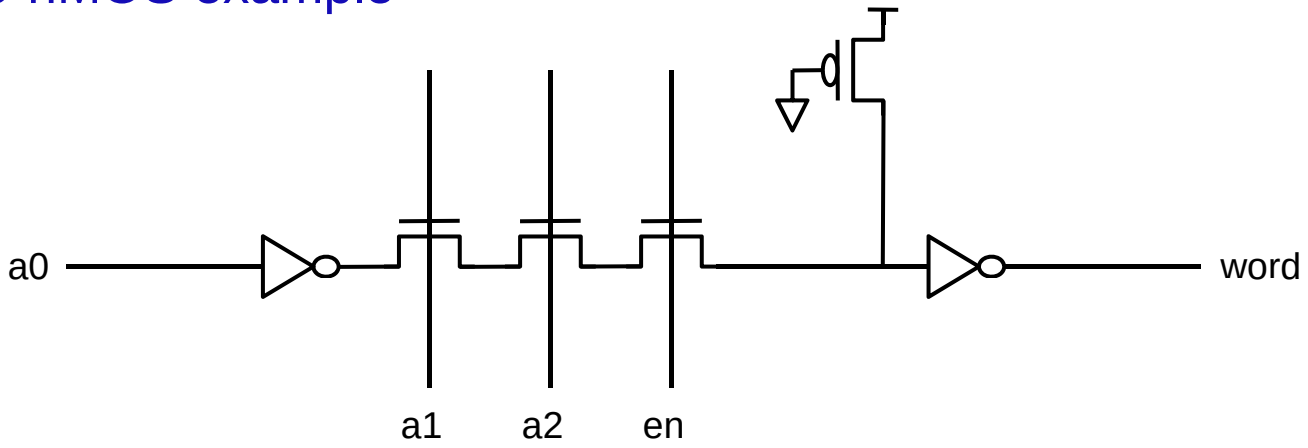


Memory Elements – Row Decoder

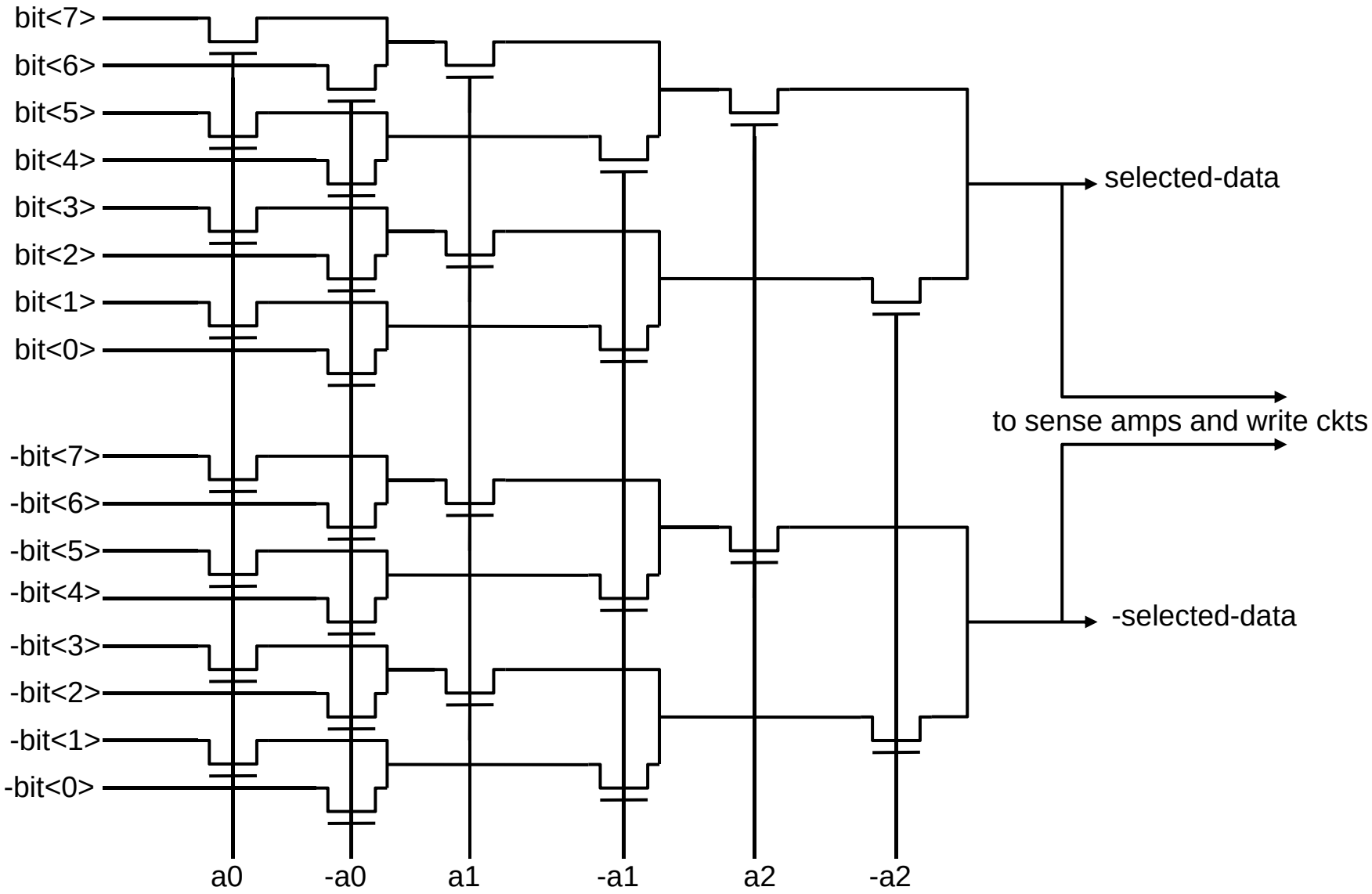
Actual implementation



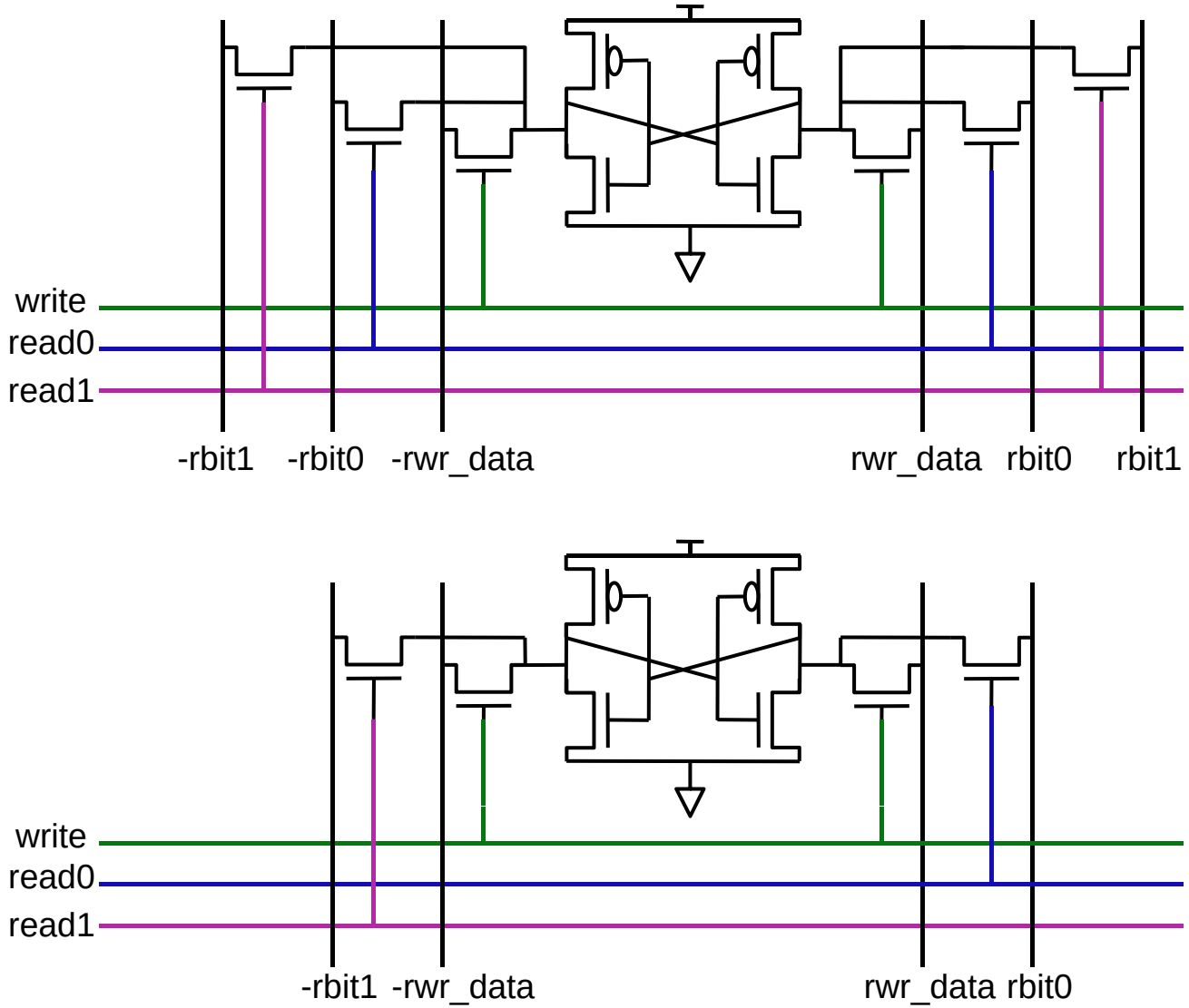
Pseudo-nMOS example



Memory Elements – Column Decoder

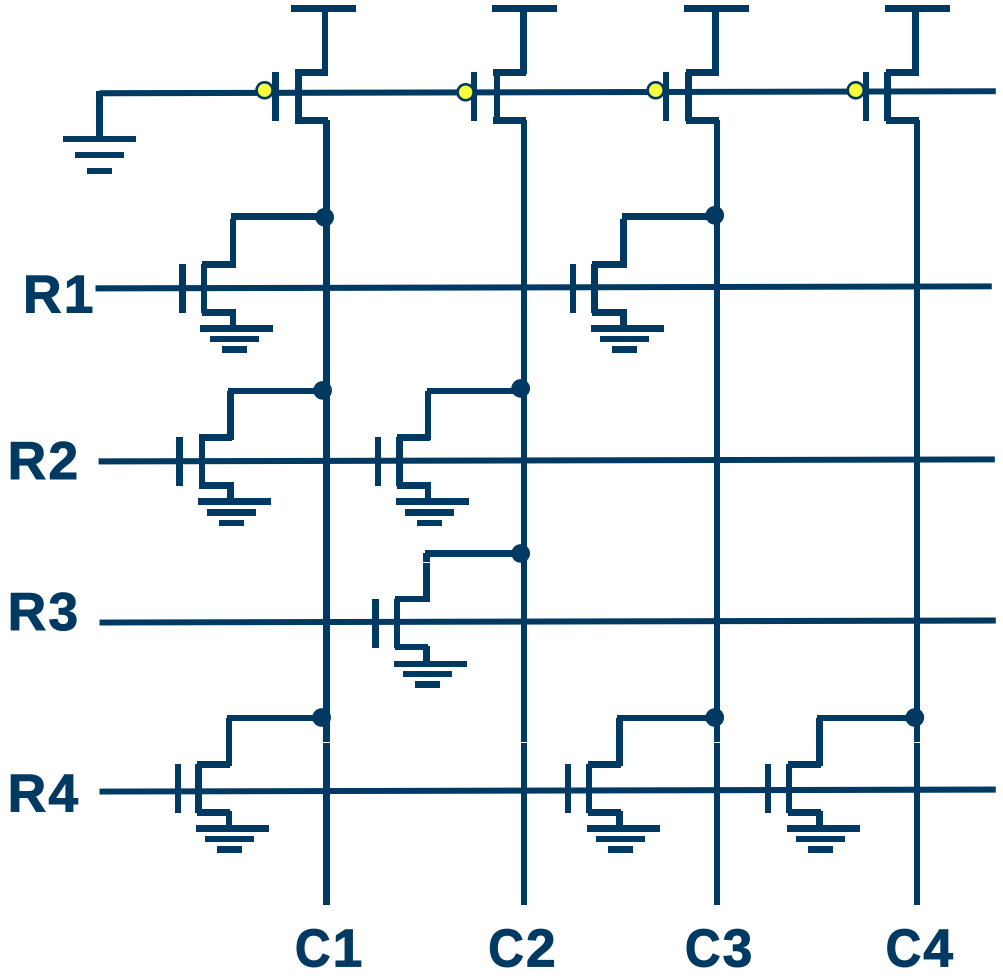


Memory Elements – Multi-Ported RAM



Memory Elements – ROM

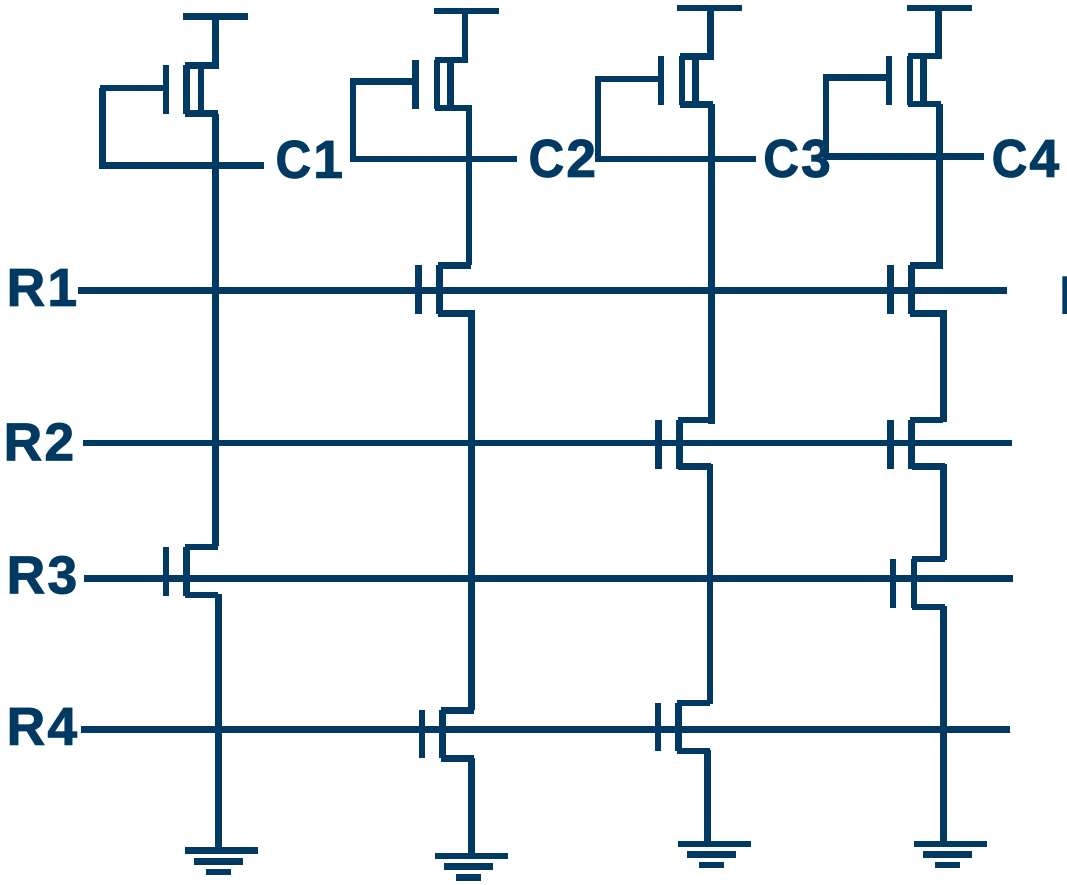
A 4x 4-bit NOR-based ROM array



R1	R2	R3	R4	C1	C2	C3	C4
1	0	0	0	0	1	0	1
0	1	0	0	0	0	1	1
0	0	1	0	1	0	0	1
0	0	0	1	0	1	1	0

Memory Elements – ROM

A 4x 4-bit NAND-based ROM array



R1	R2	R3	R4	C1	C2	C3	C4
0	1	1	1	0	1	0	1
1	0	1	1	0	0	1	1
1	1	0	1	1	0	0	1
1	1	1	0	0	1	1	0

Memory Elements – ROM

Typical ROM architecture