Digital Integrated Circuits
A Design Perspective

Semiconductor Memories
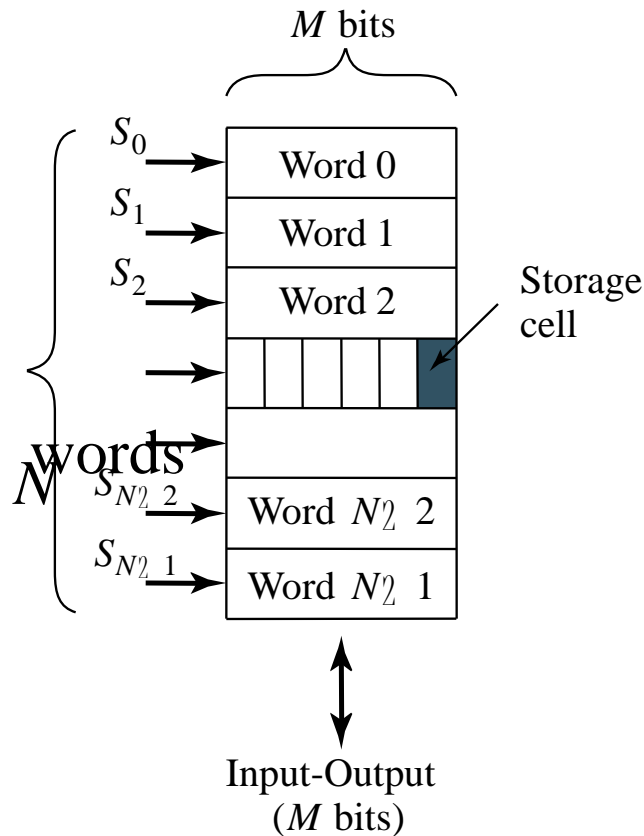
Chapter Overview

- ❑ **Memory Classification**
- ❑ **Memory Architectures**
- ❑ **The Memory Core**
- ❑ **Periphery**
- ❑ **Reliability**
- ❑ **Case Studies**

Semiconductor Memory Classification

Read-Write Memory		Non-Volatile Read-Write Memory	Read-Only Memory
Random Access	**Non-Random Access**	EPROM E^2PROM FLASH	Mask-Programmed Programmable (PROM)
SRAM DRAM	FIFO LIFO Shift Register CAM		

Memory Architecture: Decoders

M bits

S_0 → Word 0
S_1 → Word 1
S_2 → Word 2

Storage cell

N words

$S_{N \underline{2} 2}$ → Word $N \underline{2}$ 2
$S_{N \underline{2} 1}$ → Word $N \underline{2}$ 1

Input-Output
(M bits)

M bits

S_0 → Word 0
→ Word 1
A_0 → Word 2
A_1

Storage cell

$A_{K \underline{2} 1}$

Decoder → Word $N \underline{2}$ 2
→ Word $N \underline{2}$ 1

$K \; 5 \; \log_2 N$

Input-Output
(M bits)

Intuitive architecture for N x M memory
Too many select signals:
N words == N select signals
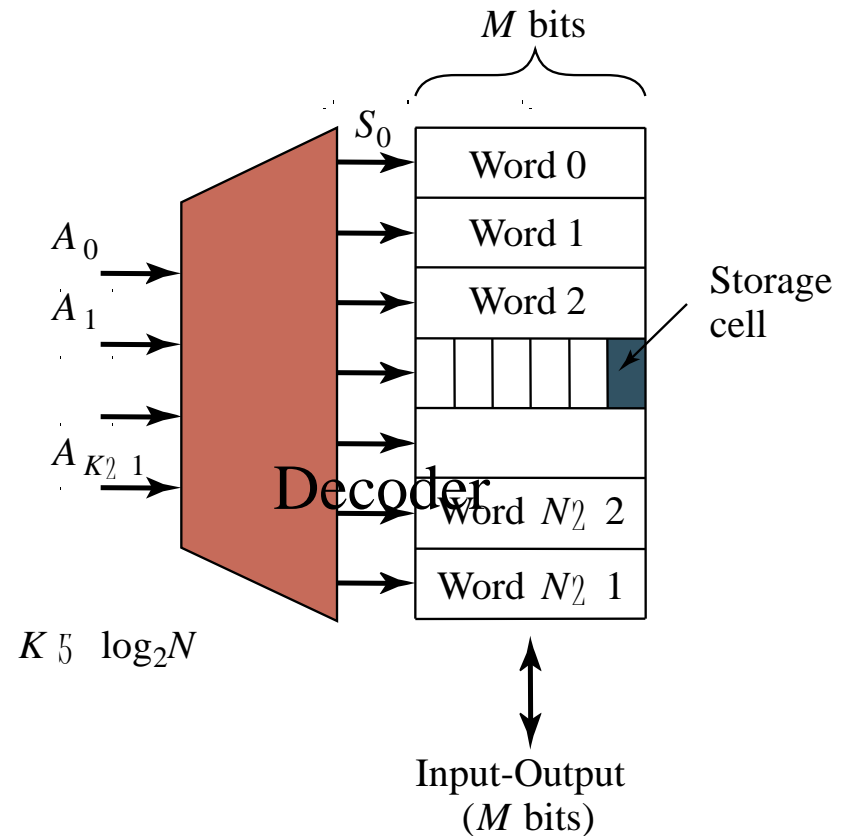
Decoder reduces the number of select signals
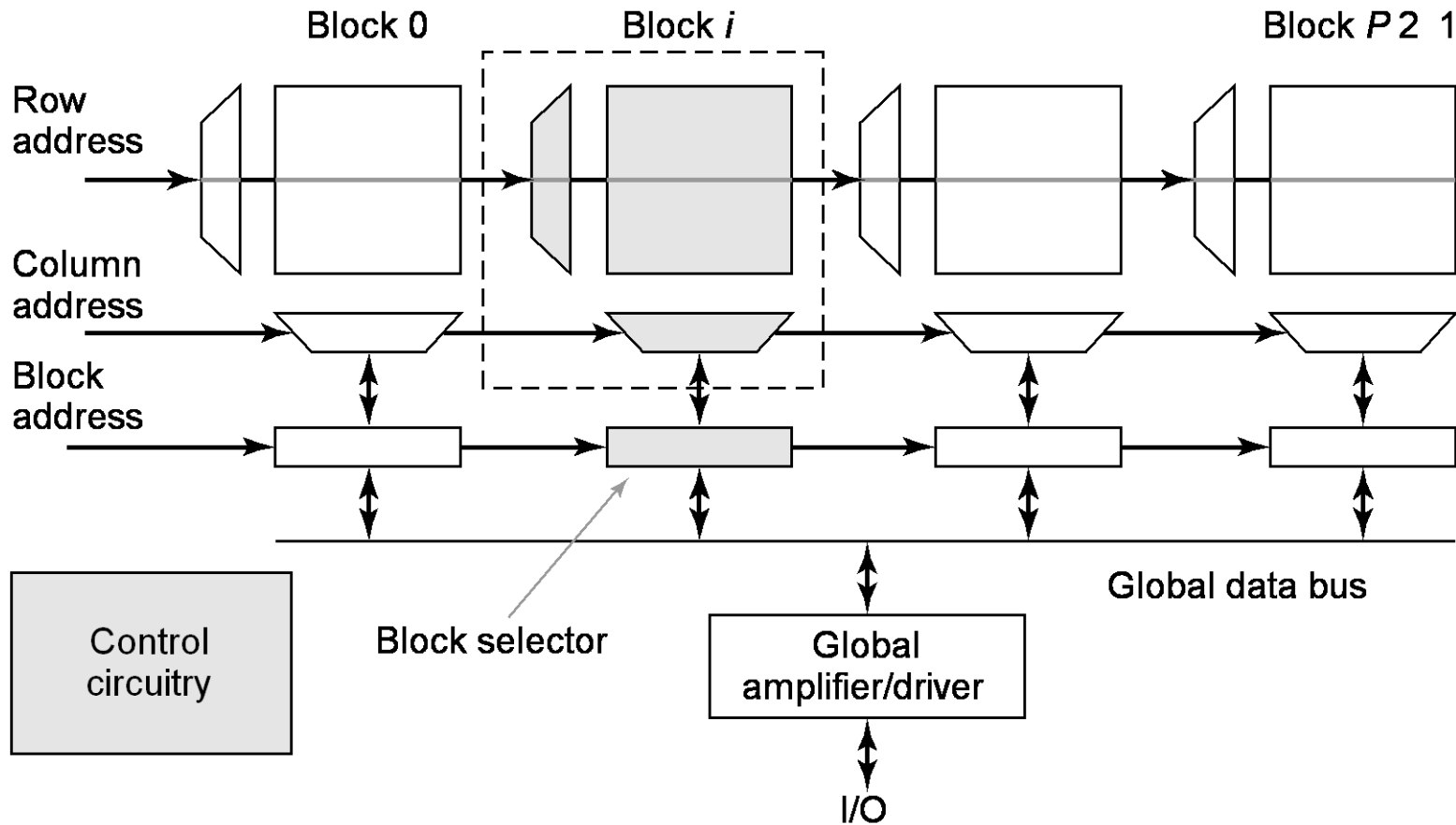$$K = log_2 N$$

Array-Structured Memory Architecture
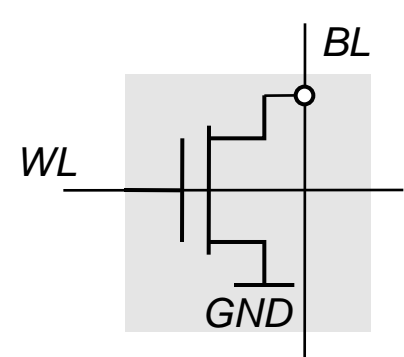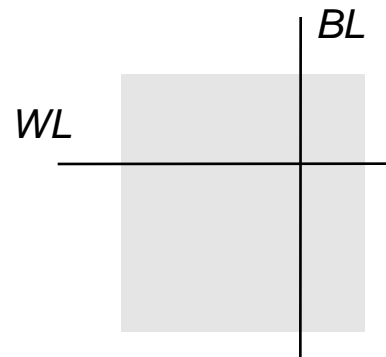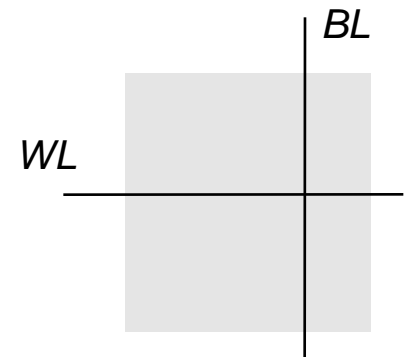
Problem: ASPECT RATIO or HEIGHT >> WIDTH
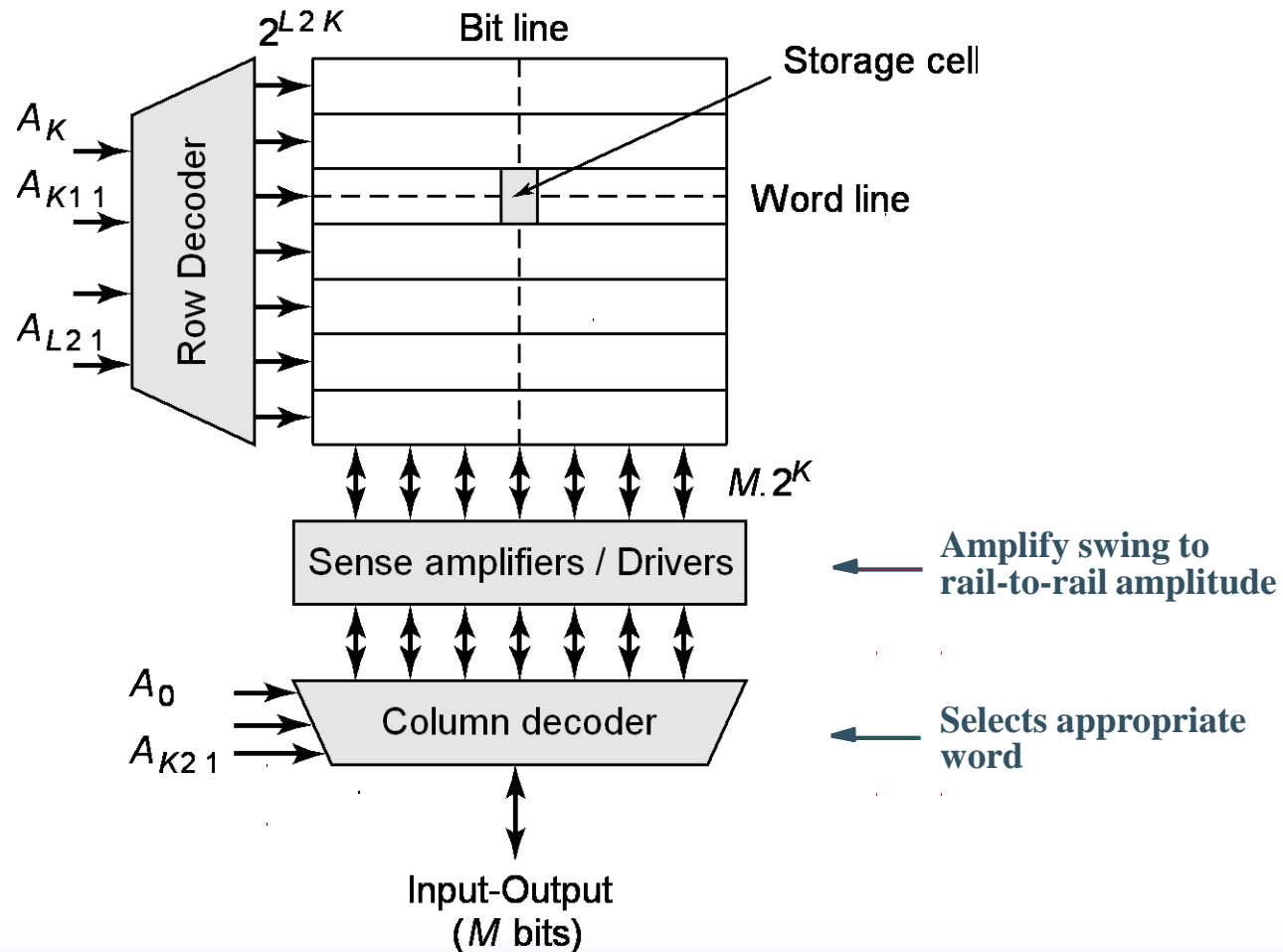
Hierarchical Memory Architecture

Block 0   Block i Block P 2 1

Row address

Column address

Block address

Control circuitry

Block selector

Global data bus

Global amplifier/driver

I/O

Advantages:
1. Shorter wires within blocks
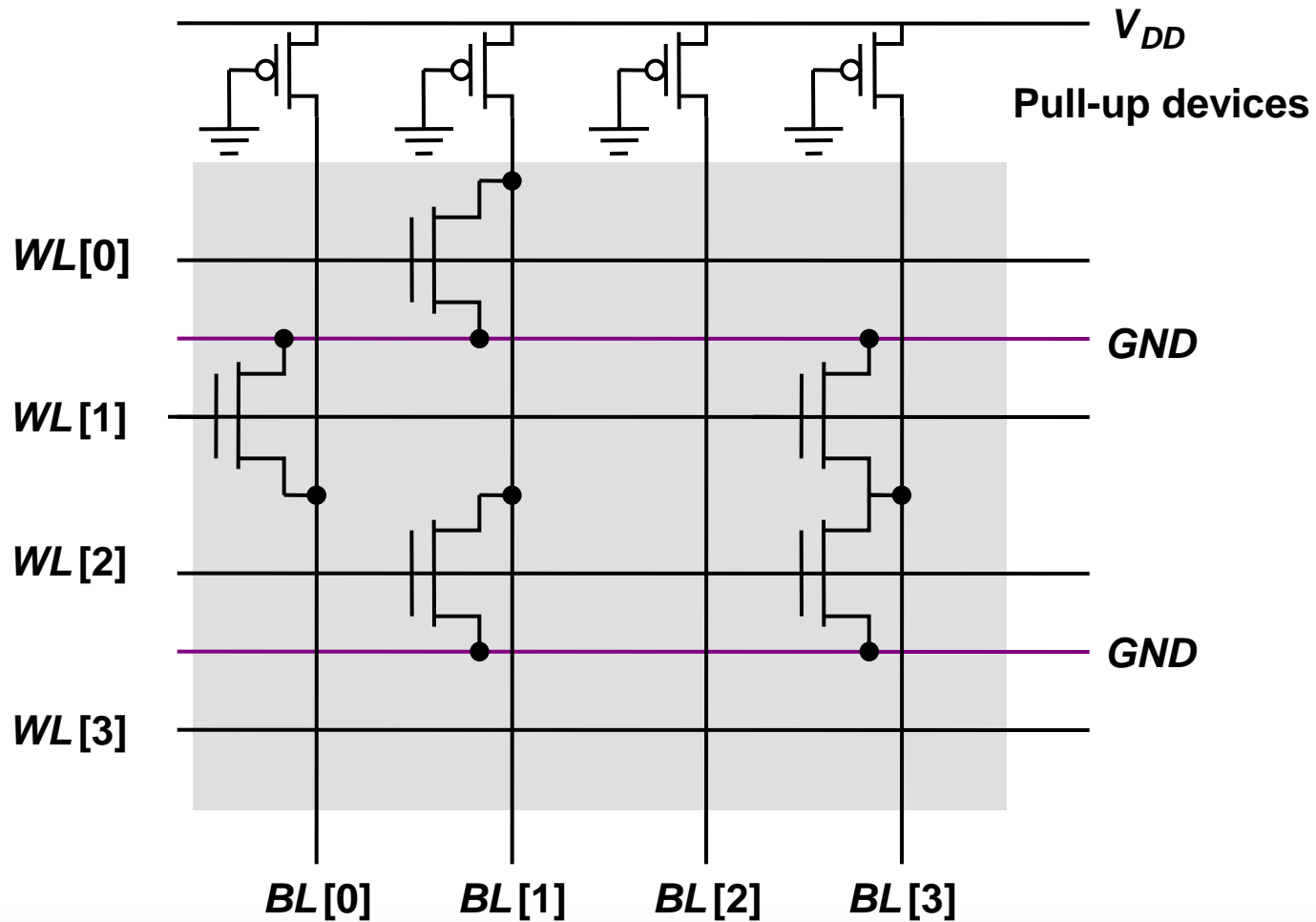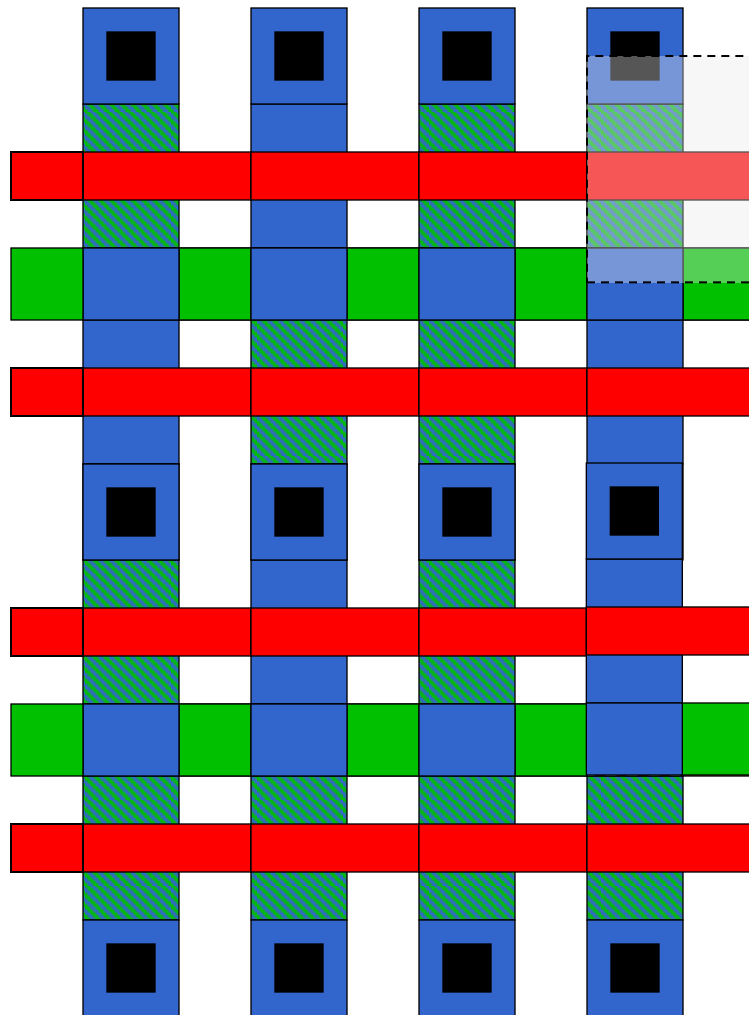**2. Block address activates only 1 block => power savings**

Read-Only Memory Cells

1

0
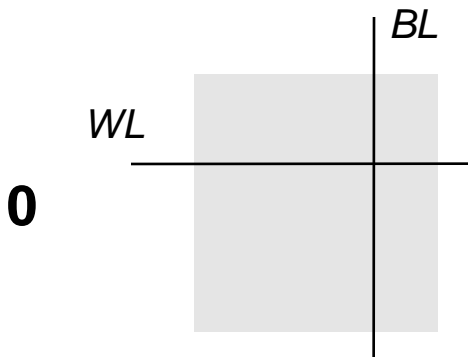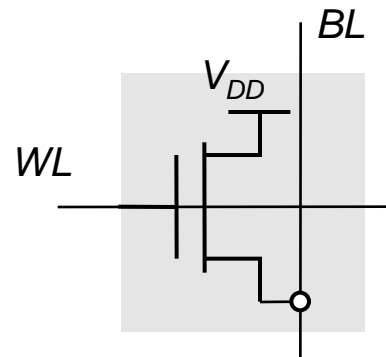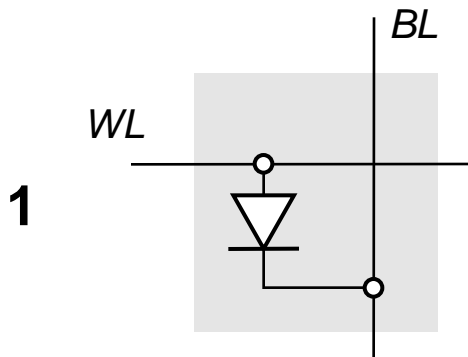
Diode ROM   **MOS ROM 1**   **MOS ROM 2**

MOS OR ROM

Pull-down loads

MOS NOR ROM

V_{DD}

Pull-up devices

WL[0]

GND
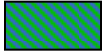
WL[1]

WL[2]

GND

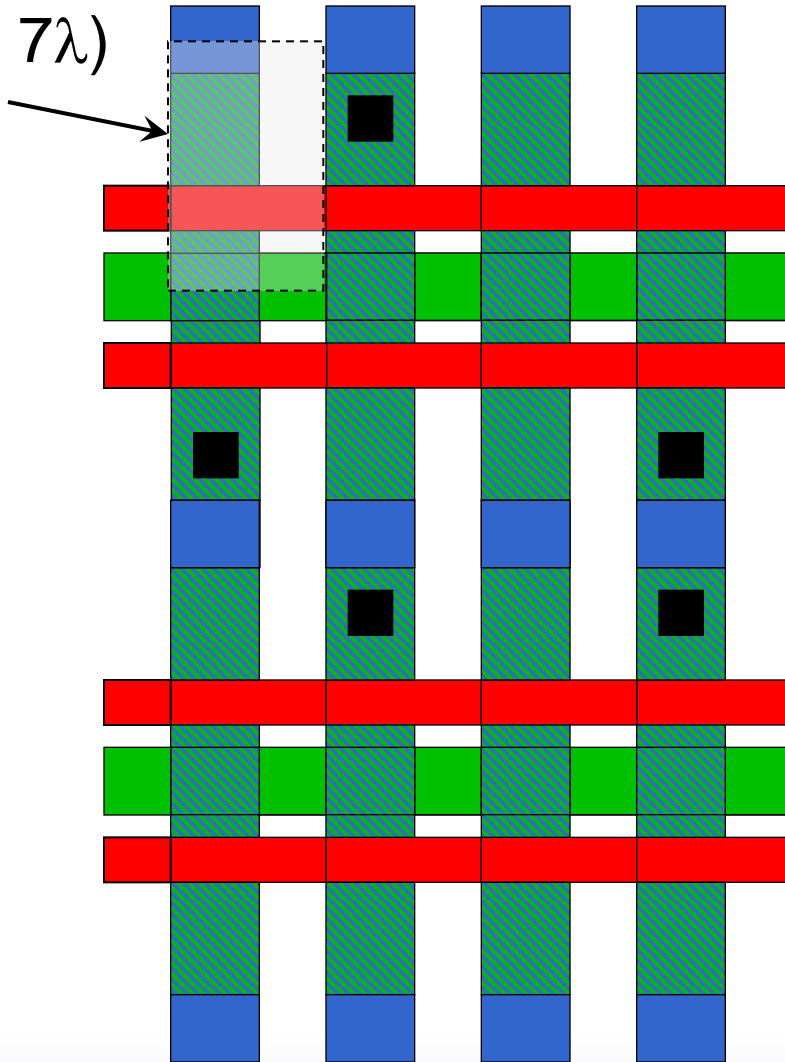WL[3]

BL[0]   BL[1]   BL[2]   BL[3]

MOS NOR ROM Layout

Cell (9.5λ x 7λ)

Programming using the Active Layer Only

- Polysilicon
- Metal1
- Diffusion
- Metal1 on Diffusion

MOS NOR ROM Layout

Cell (11λ x 7λ)

Programming using the Contact Layer Only

🟥	Polysilicon
🟦	Metal1
🟩	Diffusion
▨	Metal1 on Diffusion

MOS NAND ROM
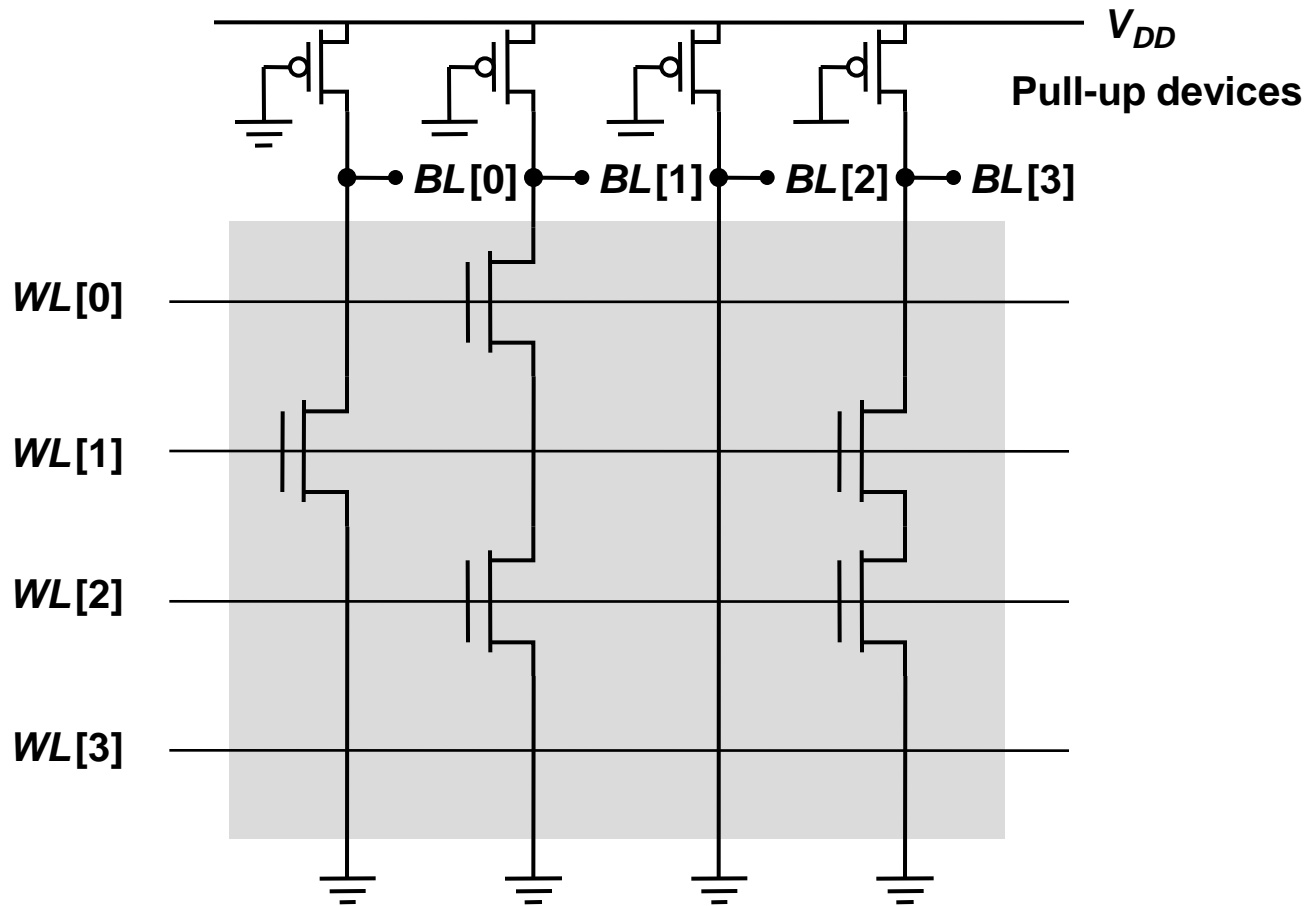
All word lines high by default with exception of selected row

Precharged MOS NOR ROM
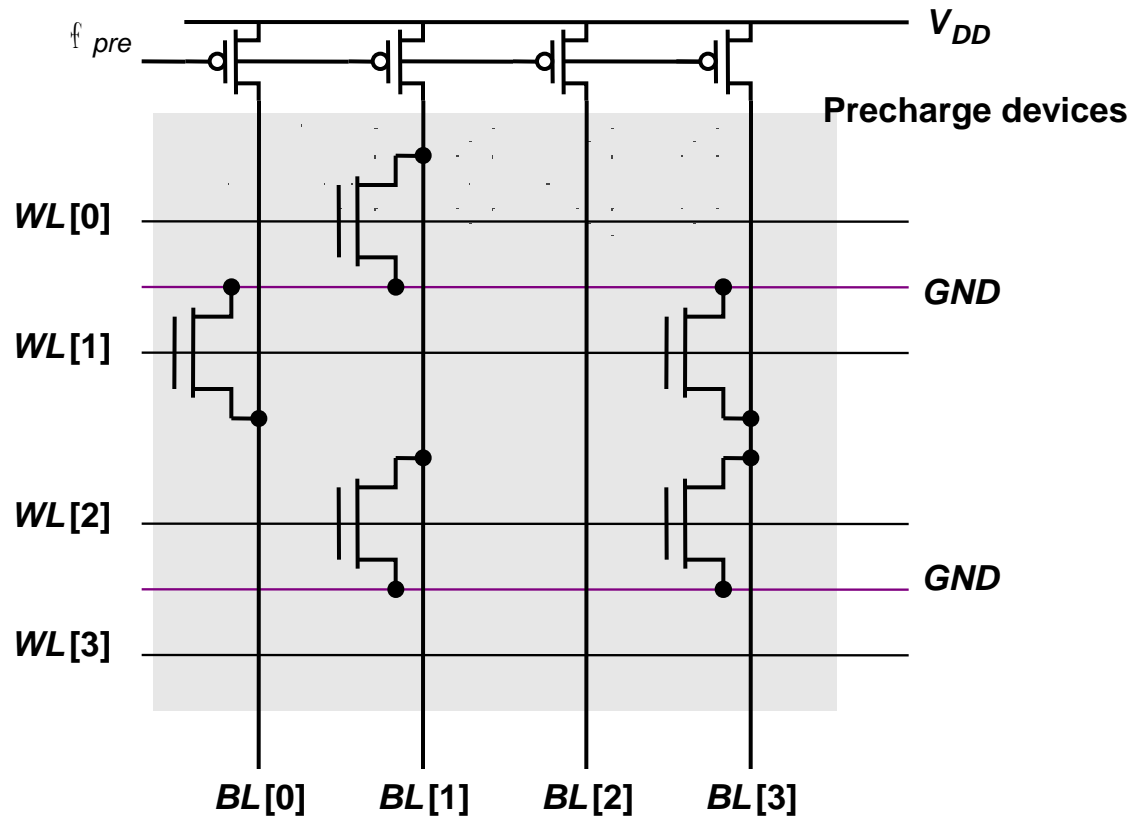
PMOS precharge device can be made as large as necessary, but clock driver becomes harder to design.

Non-Volatile Memories
The Floating-gate transistor (FAMOS)

Floating gate

Gate

Source

Drain

t_{ox}

t_{ox}

n^+

p

$n^+{}_-$

Substrate
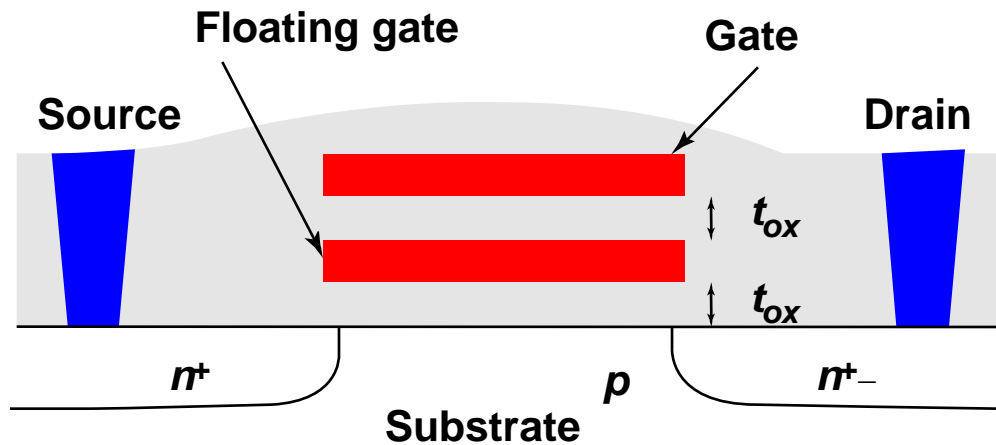
Device cross-section

D

G

S
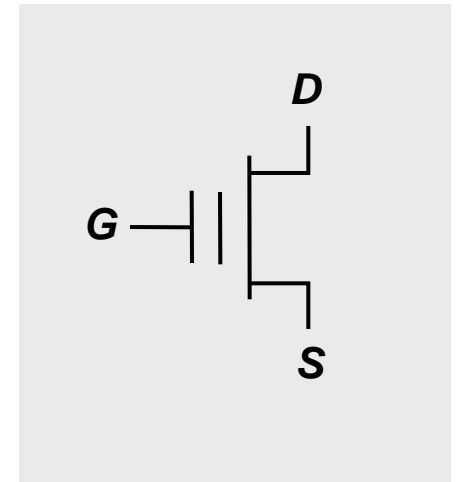
Schematic symbol

Floating-Gate Transistor Programming
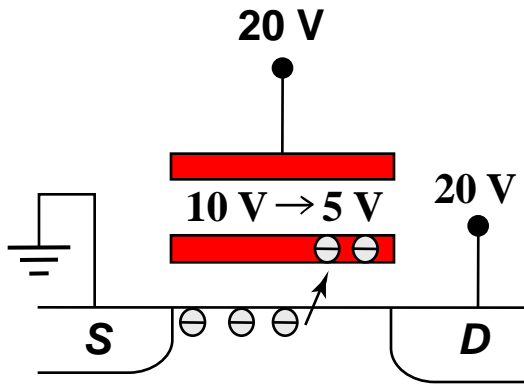
Avalanche injection
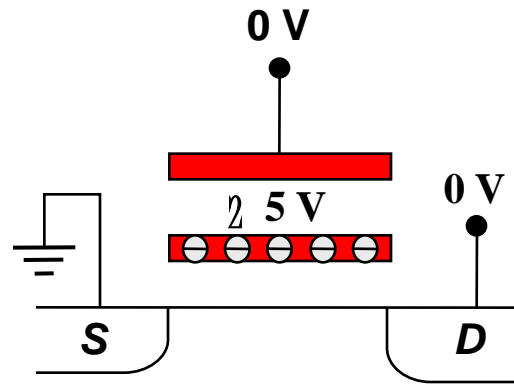
Removing programming voltage leaves charge trapped
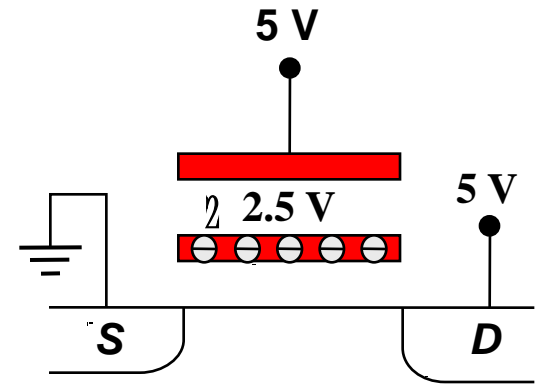
Programming results in higher V_T.

FLOTOX EEPROM

FLOTOX transistor

Floating gate

Gate

Source

Drain

20–30 nm

n^1

n^1

Substrate
p

10 nm
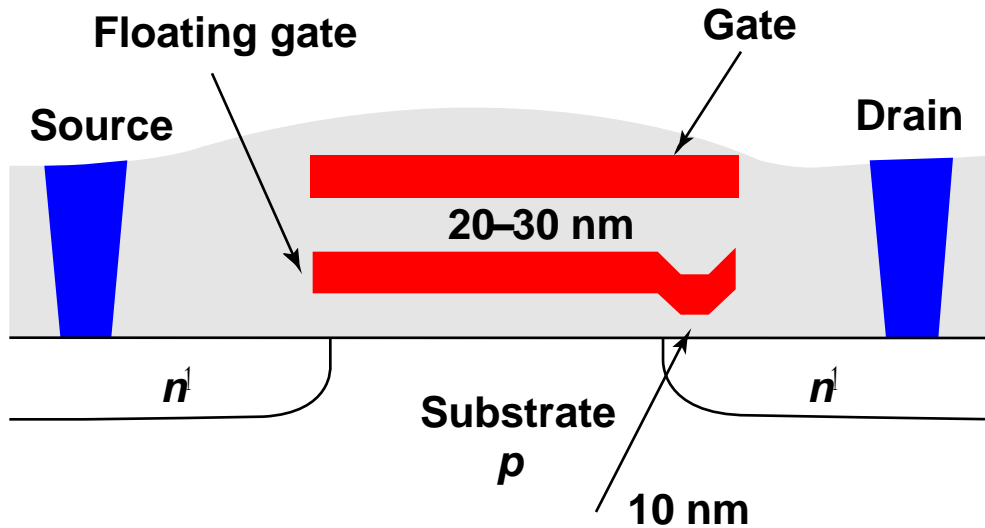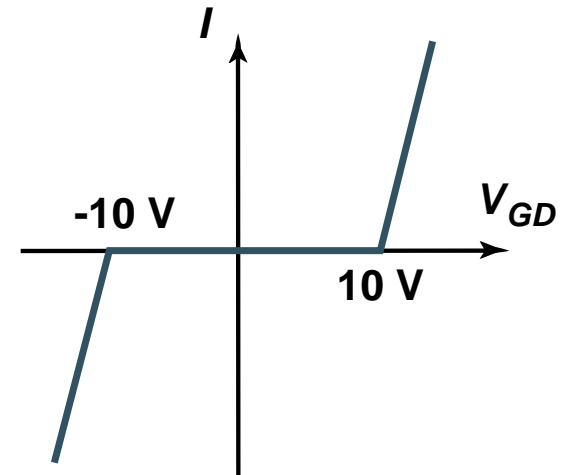
Fowler-Nordheim
I-V **characteristic**

I

V_{GD}

-10 V

10 V

EEPROM Cell

BL

WL

V_{DD}
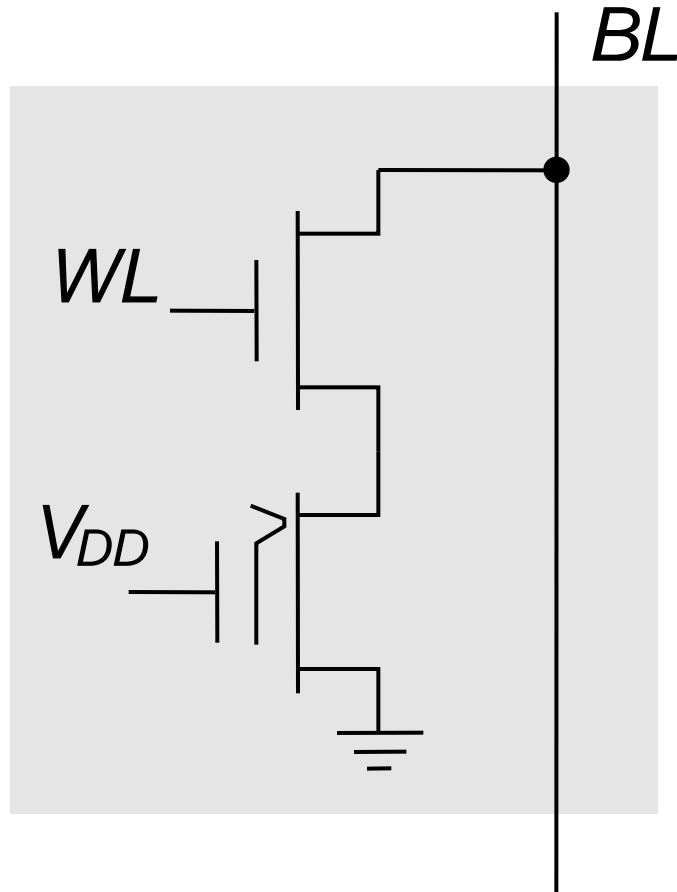
Absolute threshold control is hard
Unprogrammed transistor might be depletion
⇨ **2 transistor cell**

Flash EEPROM

Control gate

Floating gate

erasure

Thin tunneling oxide

n^1 source

n^1 drain

programming

p-substrate

Many other options …

Characteristics of State-of-the-art NVM

Table 12-1 Comparison between nonvolatile memories ([Itoh01]).
$V_{DD} = 3.3$ or 5 V; $V_{PP} = 12$ or 12.5 V.

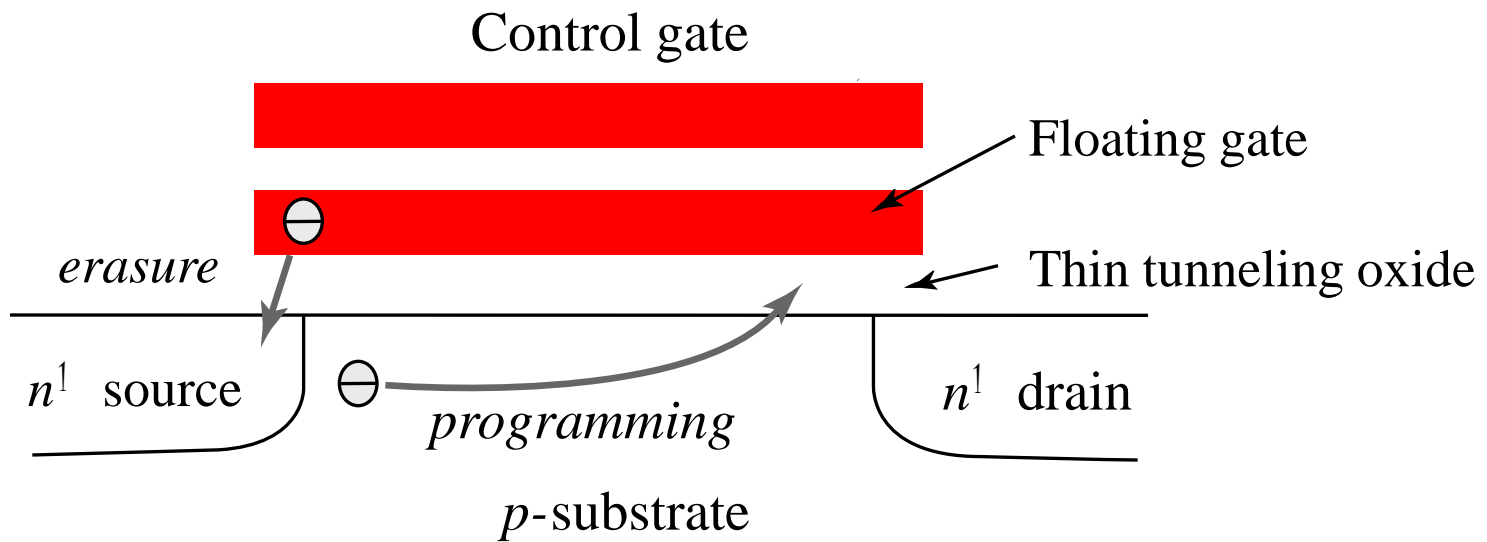|	Cell—Nr. of Transistors	Cell Area (ratio wrt EPROM)	Mechanism		External Power Supply		Program/Erase Cycles
			Erase	Write	Write	Read	
MASK ROM	1 T (NAND)	0.35–5	—	—	—	V_{DD}	0
EPROM	1 T	1	UV Exposure	Hot electrons	V_{PP}	V_{DD}	~100
EEPROM	2 T	3–5	FN Tunneling	FN Tunneling	V_{PP} (int)	V_{DD}	10^4–10^5
Flash Memory	1 T	1–2	FN Tunneling	Hot electrons	V_{PP}	V_{DD}	10^4–10^5
			FN Tunneling	FN Tunneling	V_{PP} (int)	V_{DD}	10^4–10^5

Read-Write Memories (RAM)

❑ **STATIC (SRAM)**

Data stored as long as supply is applied
Large (6 transistors/cell)
Fast
Differential
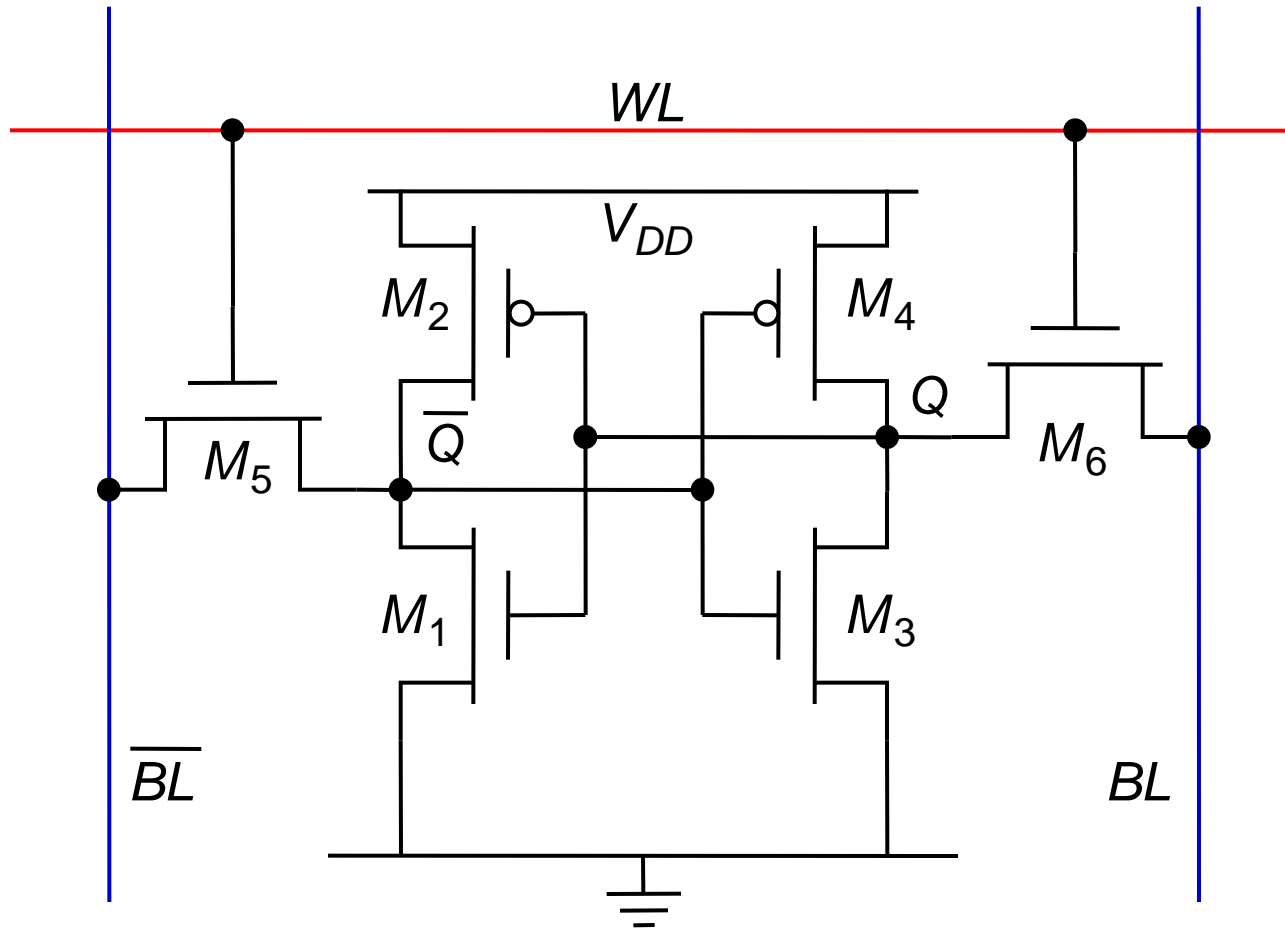
❑ **DYNAMIC (DRAM)**

Periodic refresh required
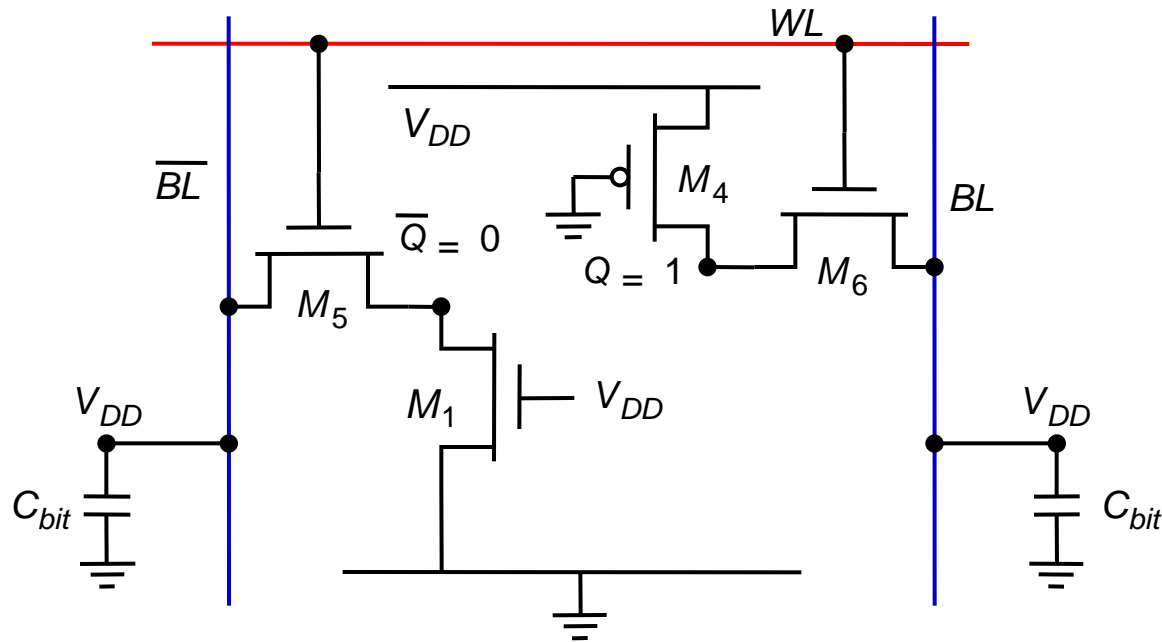**Small (1-3 transistors/cell)**
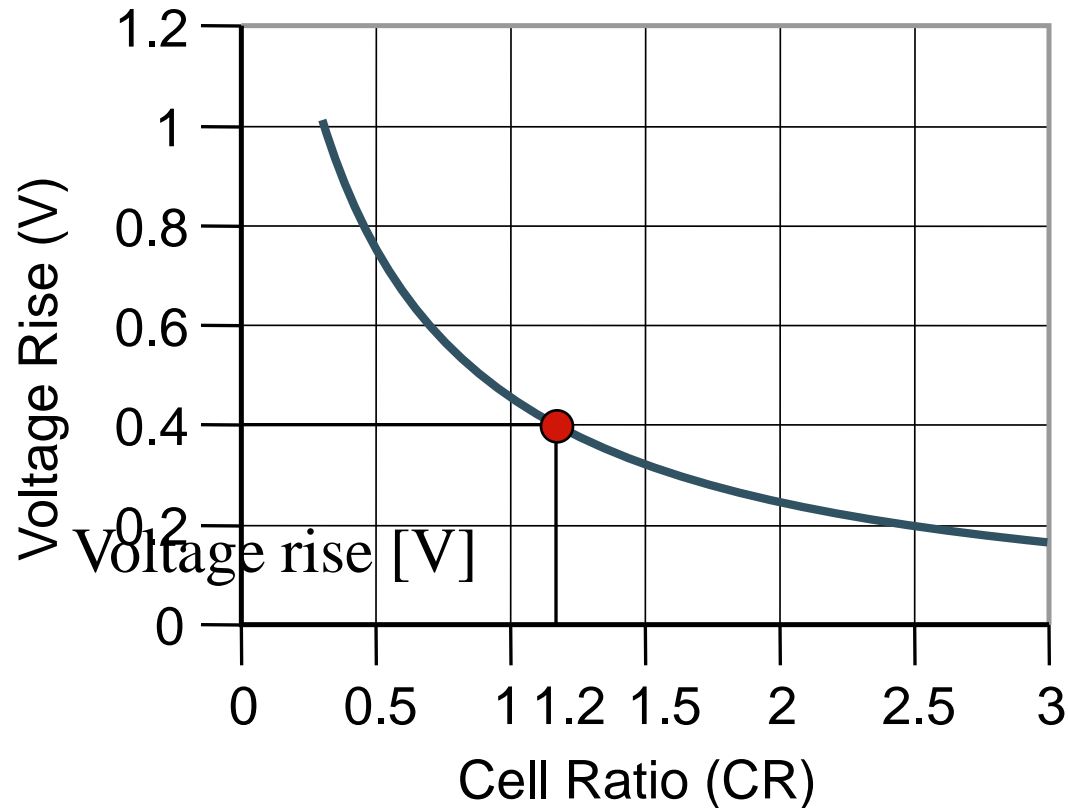**Slower**
Single Ended

6-transistor CMOS SRAM Cell

CMOS SRAM Analysis (Read)

$$k_{n,M5}\left((V_{DD} - \Delta V - V_{Tn})V_{DSATn} - \frac{V_{DSATn}^2}{2}\right) = k_{n,M1}\left((V_{DD} - V_{Tn})\Delta V - \frac{\Delta V^2}{2}\right)$$
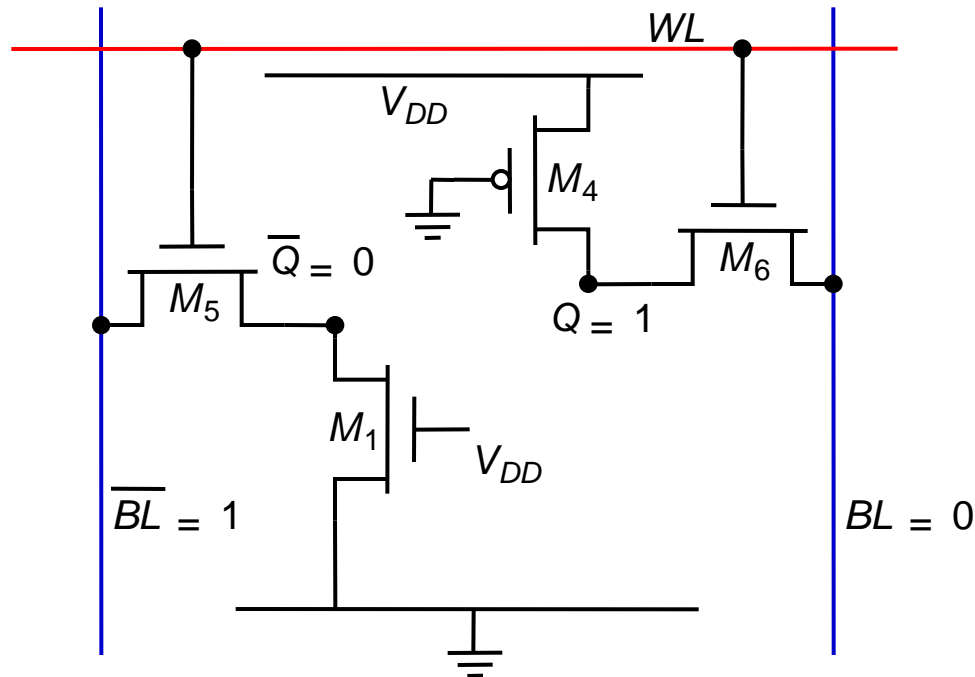
$$\Delta V = \frac{V_{DSATn} + CR(V_{DD} - V_{Tn}) - \sqrt{V_{DSATn}^2(1 + CR) + CR^2(V_{DD} - V_{Tn})^2}}{CR}$$

CMOS SRAM Analysis (Read)
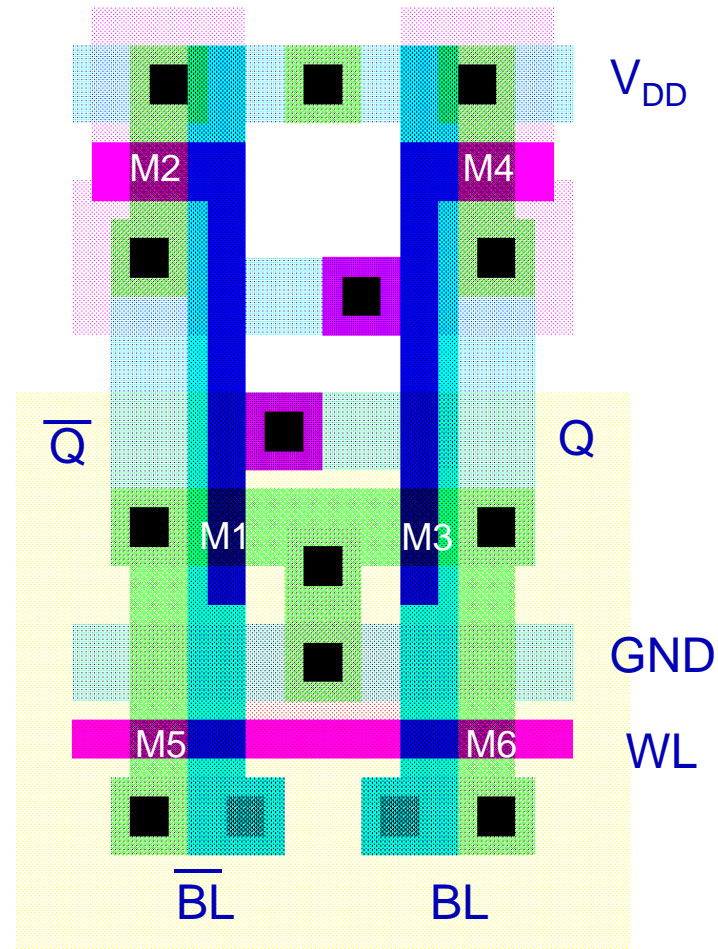

$$CR = \frac{W_1/L_1}{W_5/L_5}$$

Voltage rise [V]

CMOS SRAM Analysis (Write)

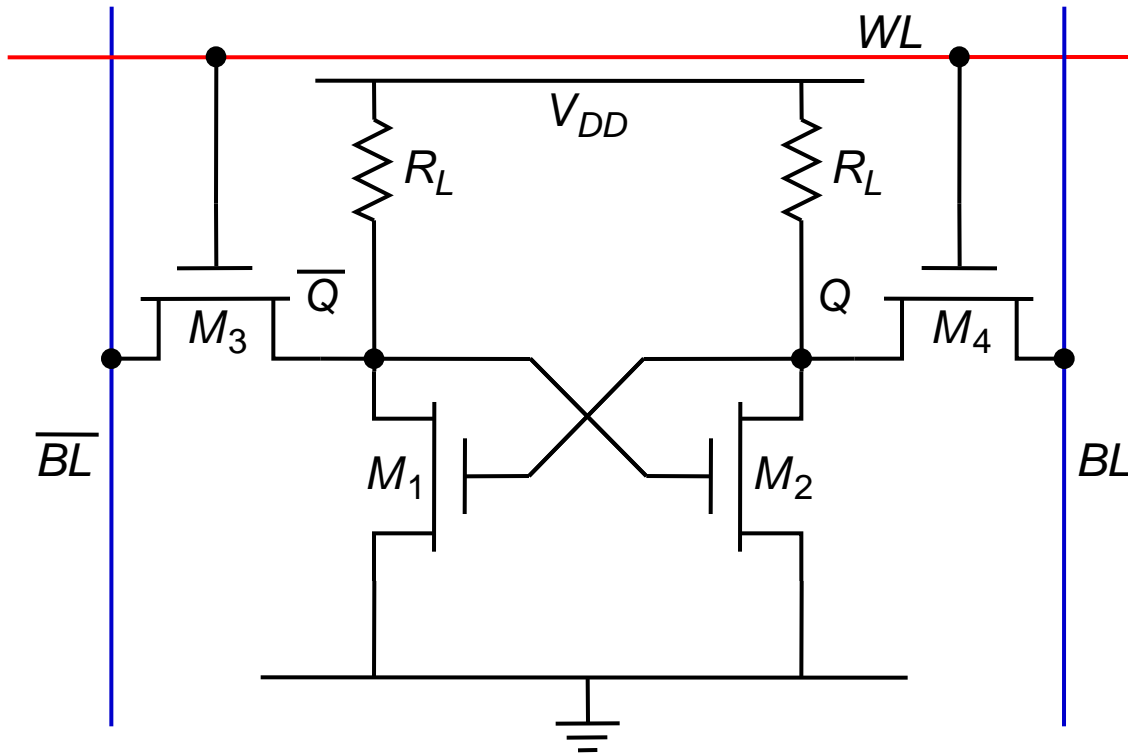
$$k_{n,M6}\left((V_{DD} - V_{Tn})V_Q - \frac{V_Q^2}{2}\right) = k_{p,M4}\left((V_{DD} - |V_{Tp}|)V_{DSATp} - \frac{V_{DSATp}^2}{2}\right)$$

$$V_Q = V_{DD} - V_{Tn} - \sqrt{(V_{DD} - V_{Tn})^2 - 2\frac{\mu_p}{\mu_n}PR\left((V_{DD} - |V_{Tp}|)V_{DSATp} - \frac{V_{DSATp}^2}{2}\right)},$$
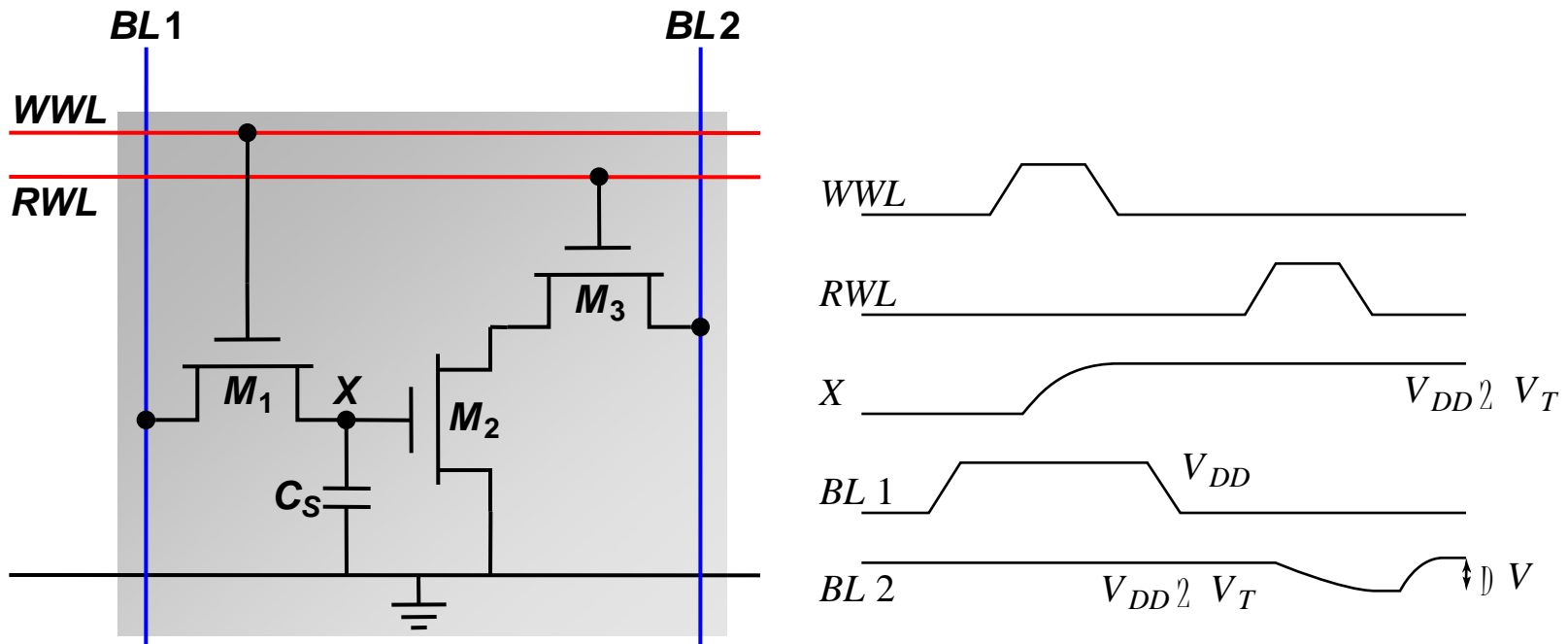
6T-SRAM — Layout

Resistance-load SRAM Cell

Static power dissipation -- Want R_L large
Bit lines precharged to V_{DD} to address t_p problem

SRAM Characteristics

Table 12-2 Comparison of CMOS SRAM cells used in 1-Mbit memory (from [Takada91])

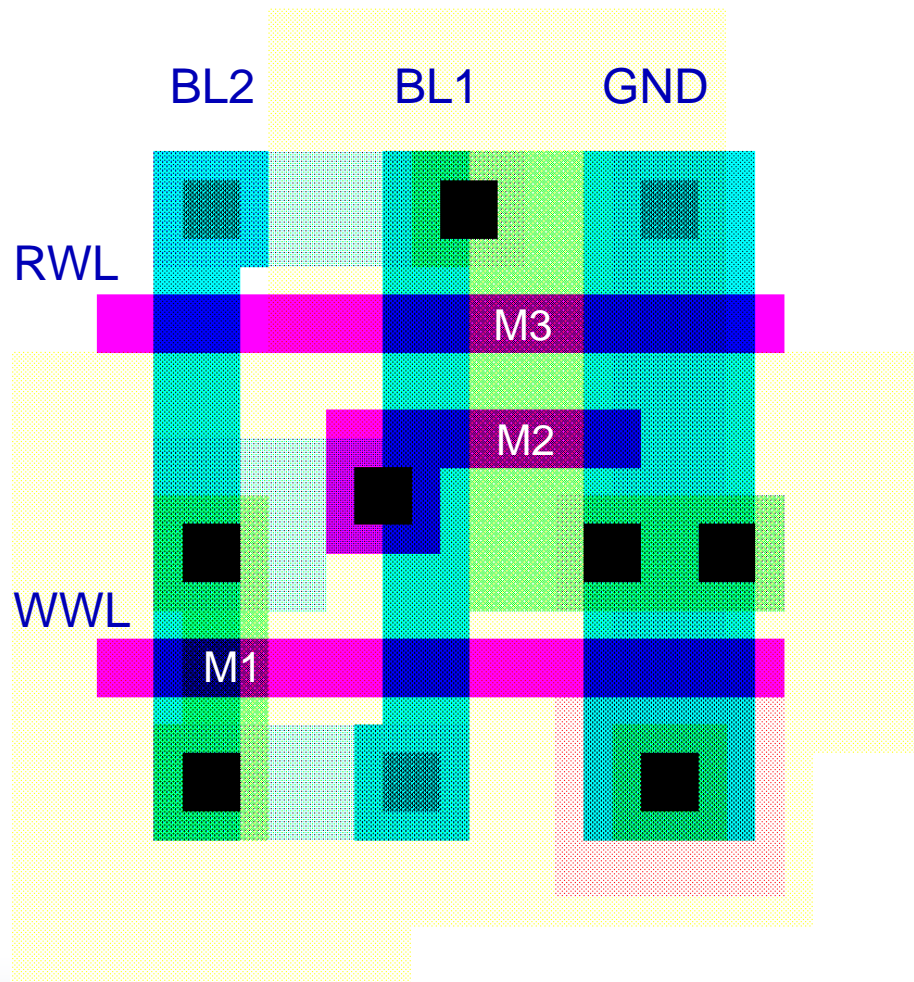|	Complementary CMOS	Resistive Load	TFT Cell
Number of transistors	6	4	4 (+2 TFT)
Cell size	58.2 μm^2 (0.7-μm rule)	40.8 μm^2 (0.7-μm rule)	41.1 μm^2 (0.8-μm rule)
Standby current (per cell)	10^{-15} A	10^{-12} A	10^{-13} A

3-Transistor DRAM Cell

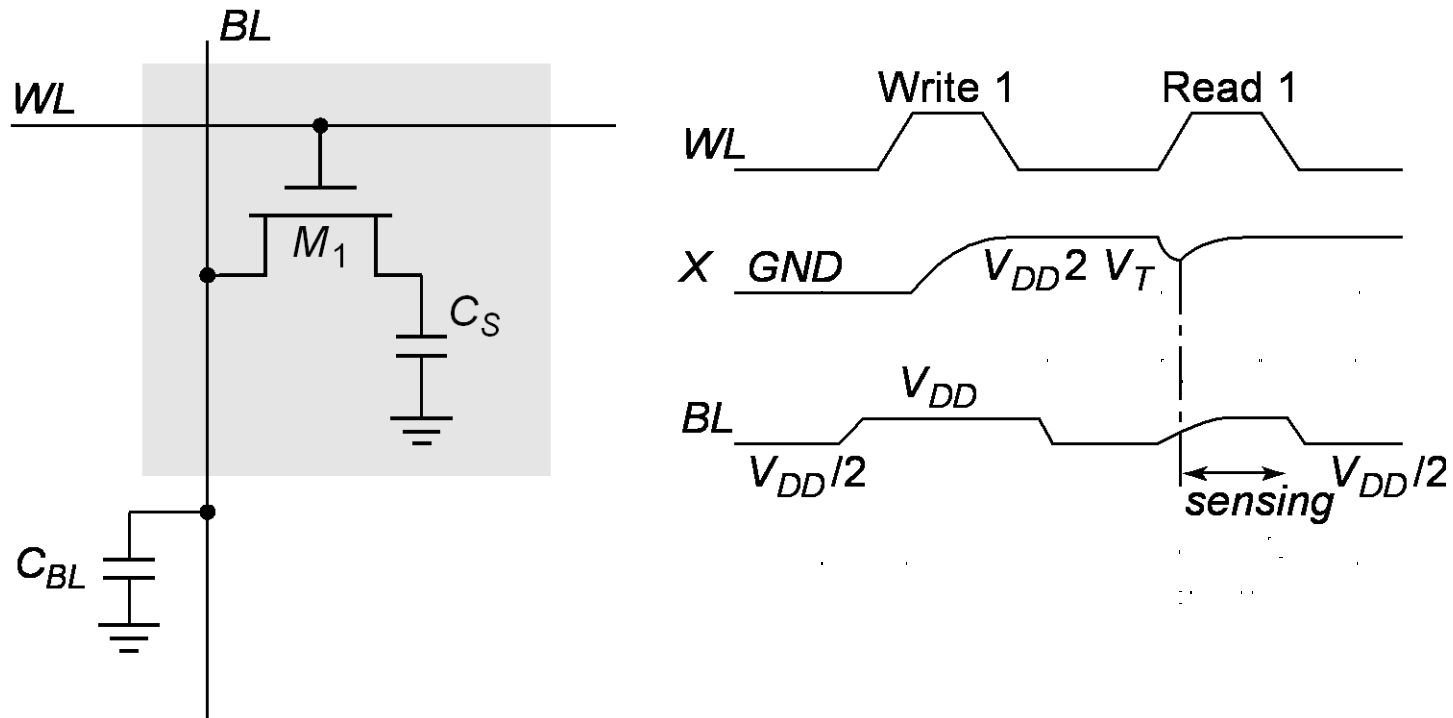**No constraints on device ratios**
Reads are non-destructive
Value stored at node X when writing a "1" = V$_{WWL}$-V$_{Tn}$

3T-DRAM — Layout

1-Transistor DRAM Cell

Write: C_S is charged or discharged by asserting WL and BL.
Read: Charge redistribution takes places between bit line and storage capacitance
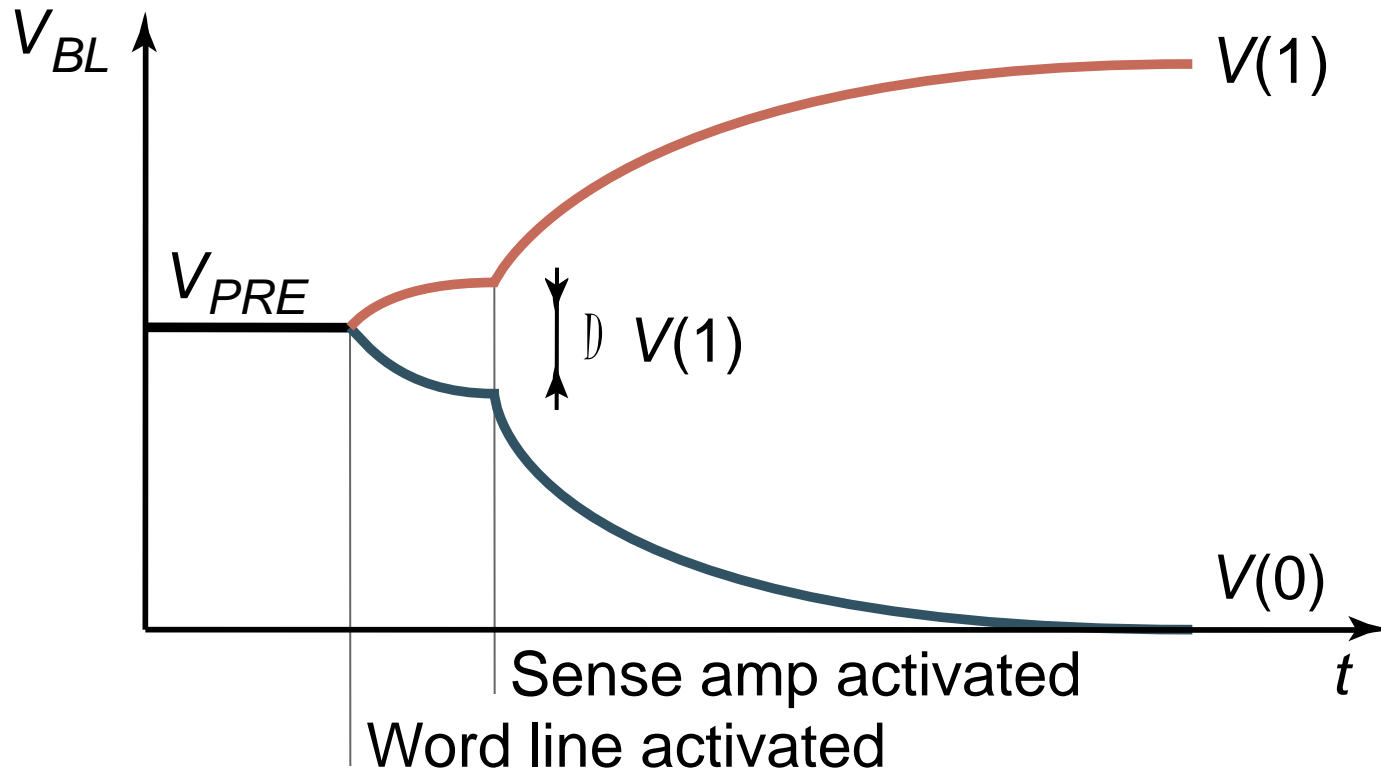
$$\Delta V = V_{BL} - V_{PRE} = V_{BIT} - V_{PRE} \frac{C_S}{C_S + C_{BL}}$$

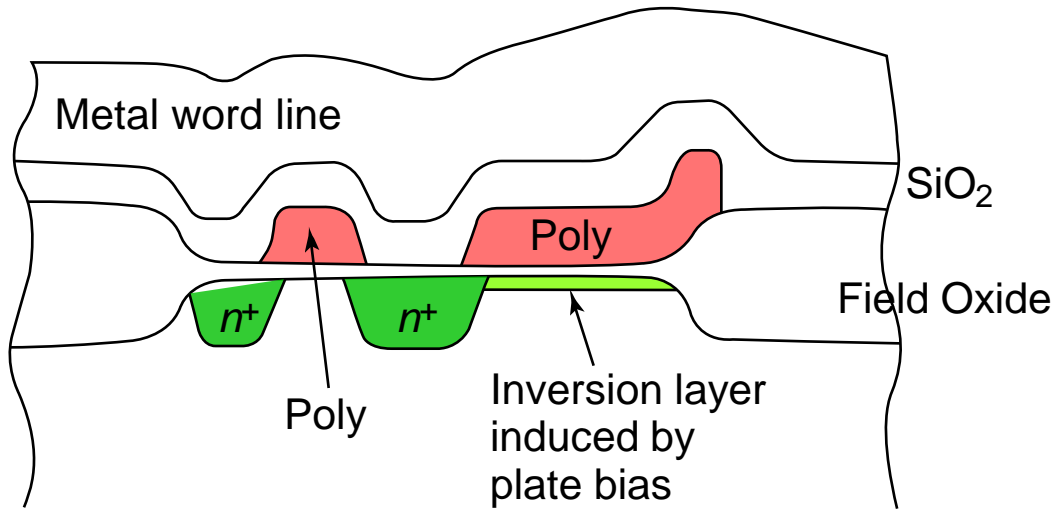**Voltage swing is small; typically around 250 mV.**

DRAM Cell Observations

❑ 1T DRAM requires a sense amplifier for each bit line, due to charge redistribution read-out.

❑ DRAM memory cells are single ended in contrast to SRAM cells.

❑ The read-out of the 1T DRAM cell is destructive; read and refresh operations are necessary for correct operation.

❑ Unlike 3T cell, 1T cell requires presence of an extra capacitance that must be explicitly included in the design.

❑ When writing a "1" into a DRAM cell, a threshold voltage is lost. This charge loss can be circumvented by bootstrapping the word lines to a higher value than V_{DD}
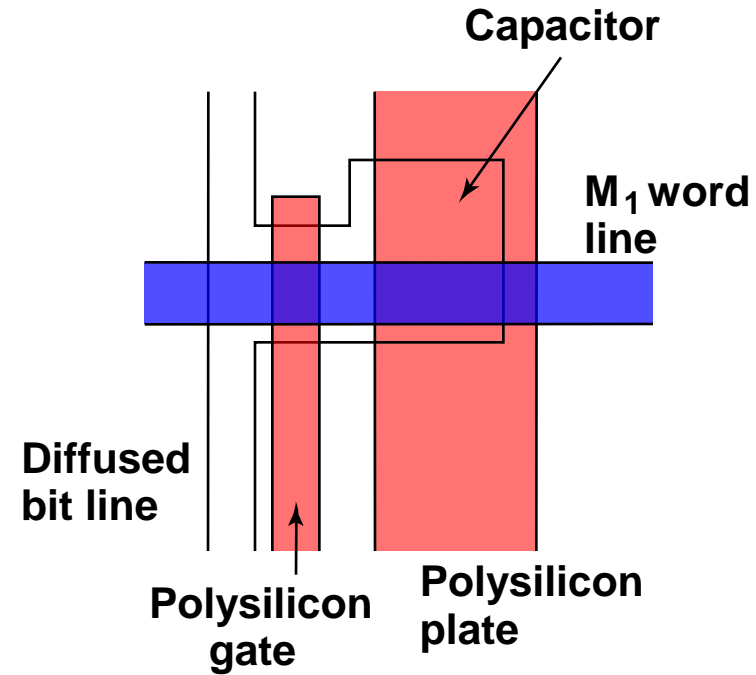
Sense Amp Operation

1-T DRAM Cell

Metal word line

SiO$_2$

Poly

Field Oxide

n^+ n^+

Poly

Inversion layer induced by plate bias

Cross-section

Capacitor

M$_1$ word line
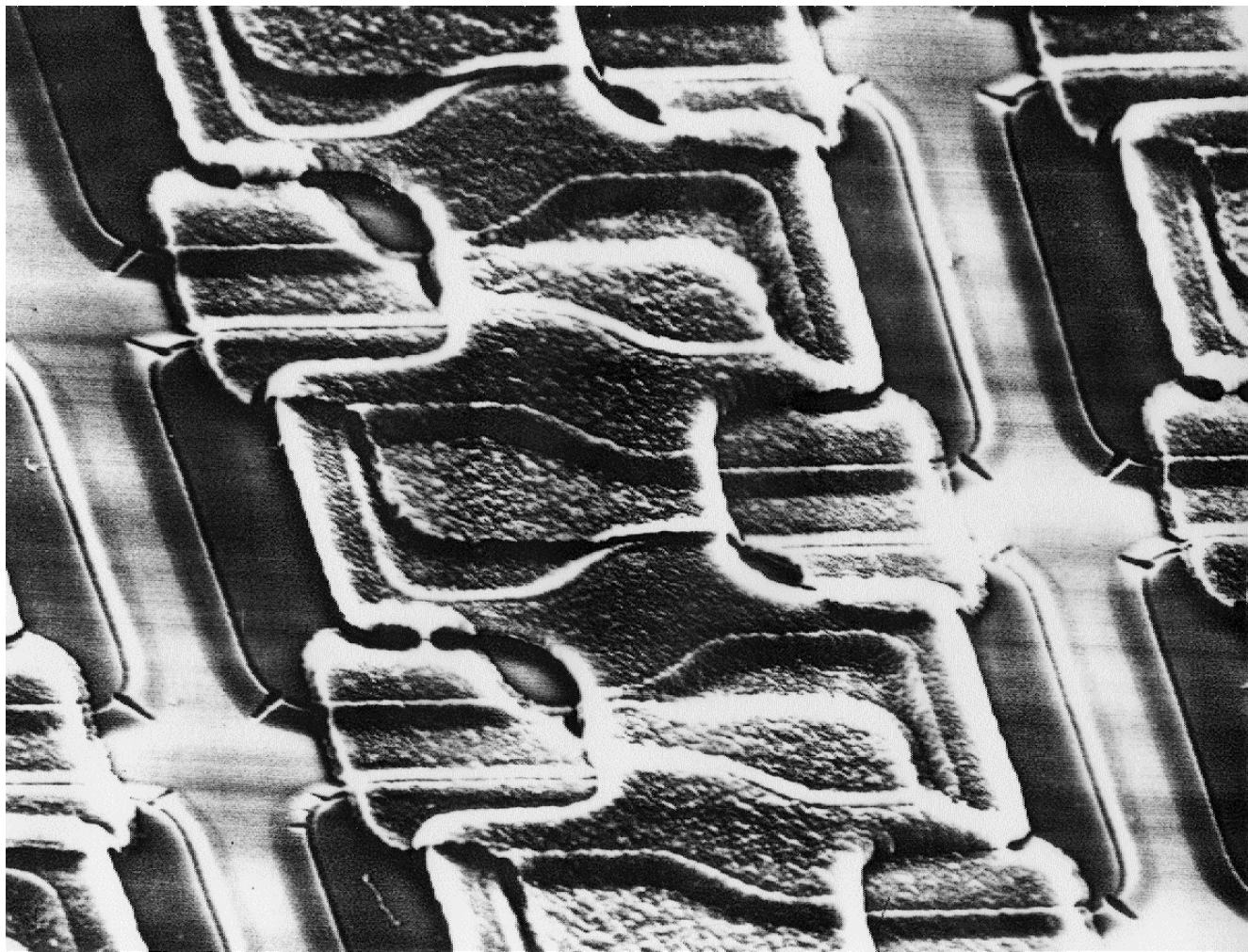
Diffused bit line

Polysilicon gate

Polysilicon plate

Layout

Uses Polysilicon-Diffusion Capacitance

Expensive in Area

SEM of poly-diffusion capacitor 1T-DRAM

Advanced 1T DRAM Cells
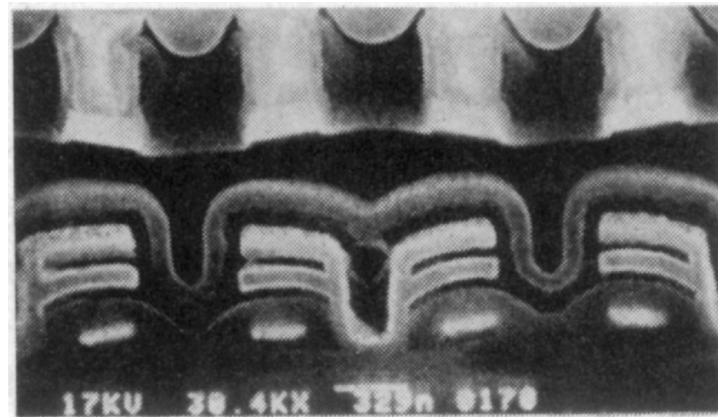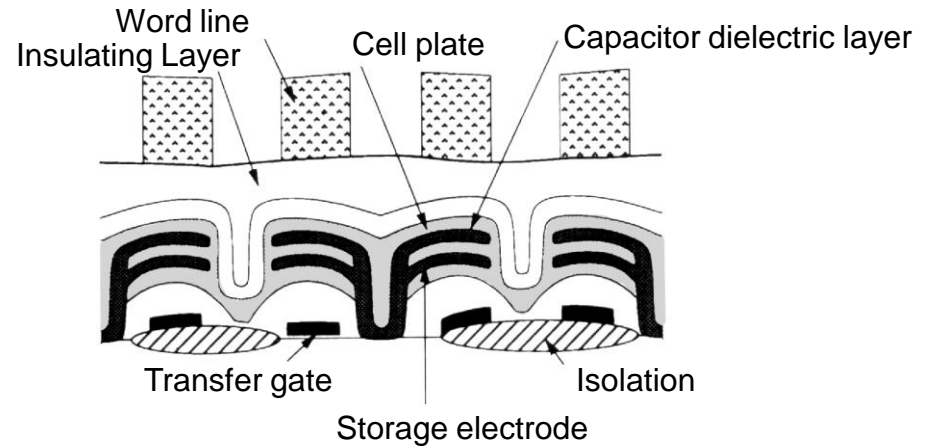

Cell Plate Si

Capacitor Insulator

Storage Node Poly

2nd Field Oxide

Refilling Poly

Si Substrate

Trench Cell

Word line
Insulating Layer

Cell plate

Capacitor dielectric layer

Transfer gate

Isolation

Storage electrode

17KV   30.4KX   329n 0170
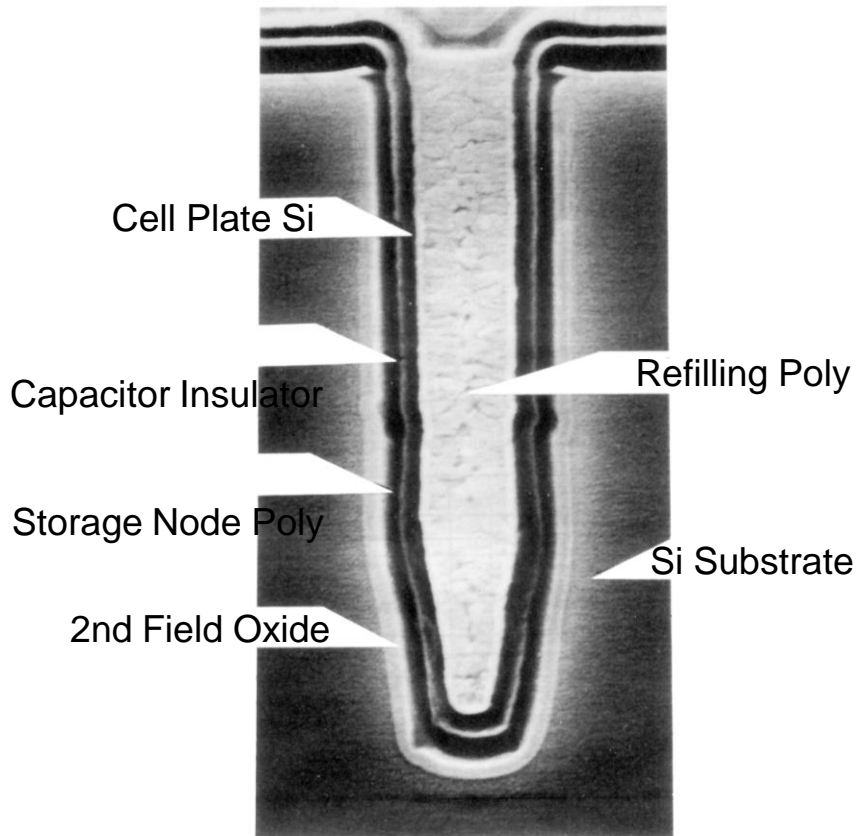
Stacked-capacitor Cell

Periphery

- ❑ **Decoders**
- ❑ **Sense Amplifiers**
- ❑ **Input/Output Buffers**
- ❑ **Control / Timing Circuitry**

Row Decoders

Collection of 2^M complex logic gates
Organized in regular and dense fashion

(N)AND Decoder

$$WL_0 = A_0 A_1 A_2 A_3 A_4 A_5 A_6 A_7 A_8 A_9$$

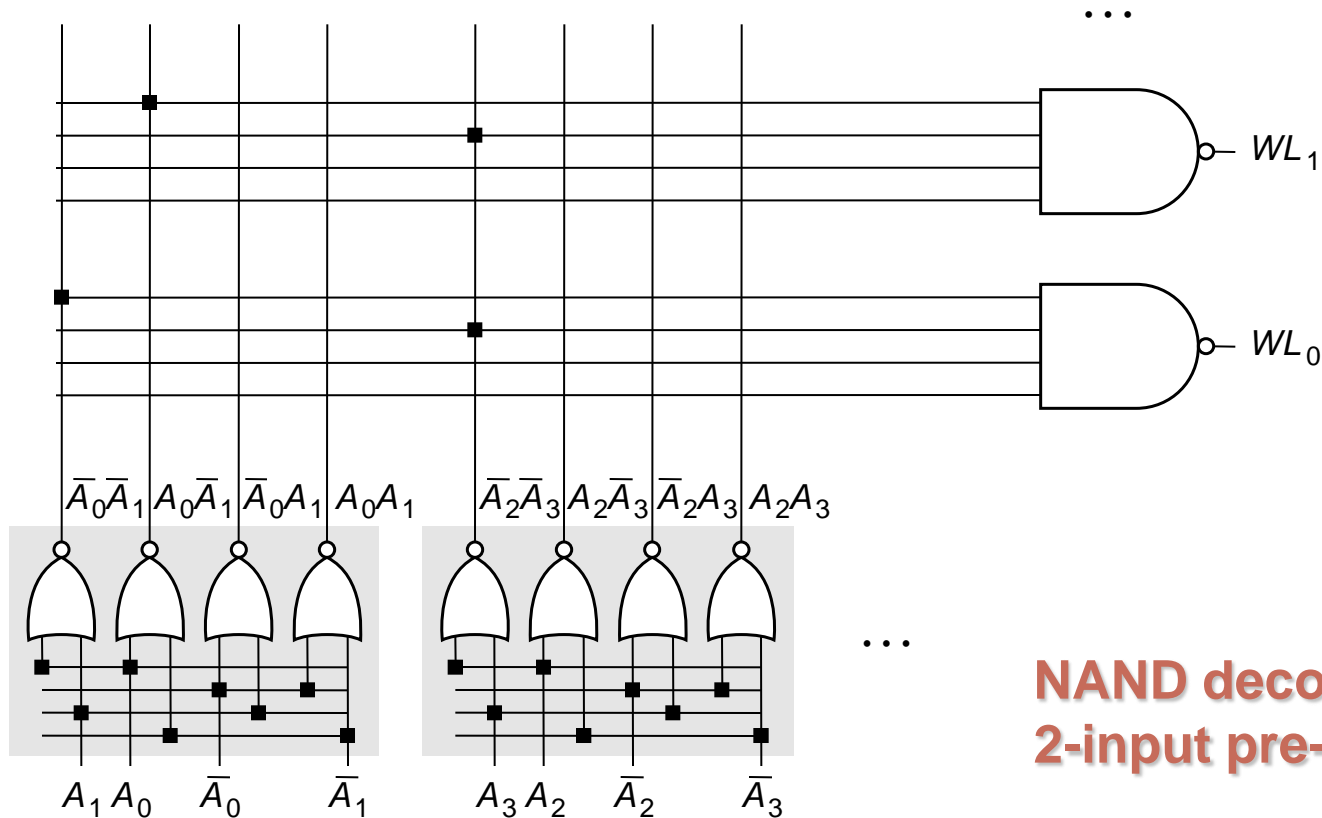$$WL_{511} = \overline{A_0} A_1 A_2 A_3 A_4 A_5 A_6 A_7 A_8 A_9$$

NOR Decoder

$$WL_0 = \overline{A_0 + A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9}$$

$$WL_{511} = \overline{A_0 + \overline{A_1} + \overline{A_2} + \overline{A_3} + \overline{A_4} + \overline{A_5} + \overline{A_6} + \overline{A_7} + \overline{A_8} + \overline{A_9}}$$
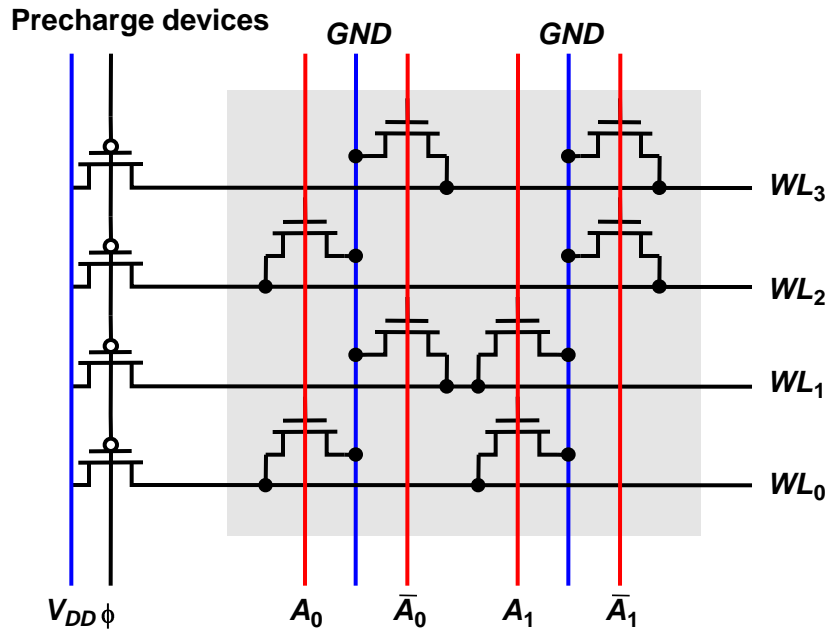
Memories

Hierarchical Decoders

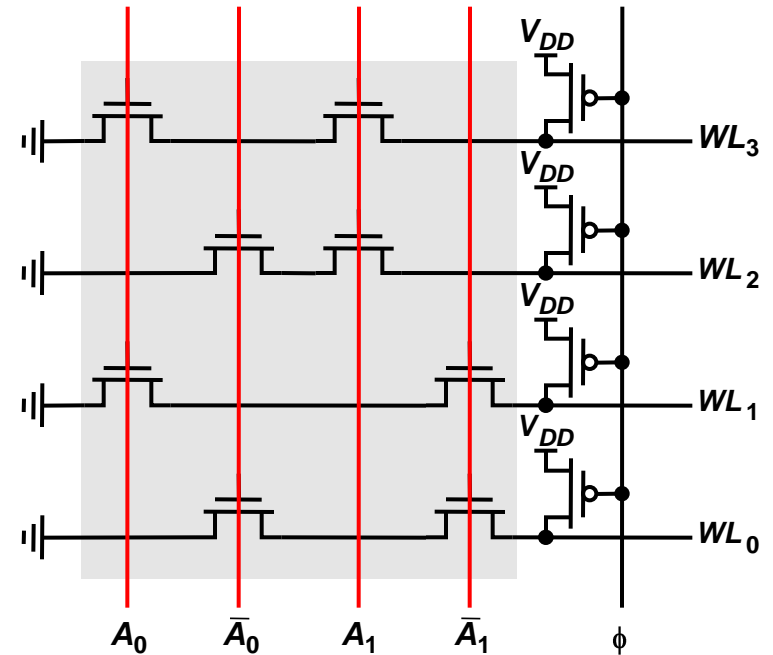## Multi-stage implementation improves performance

NAND decoder using 2-input pre-decoders
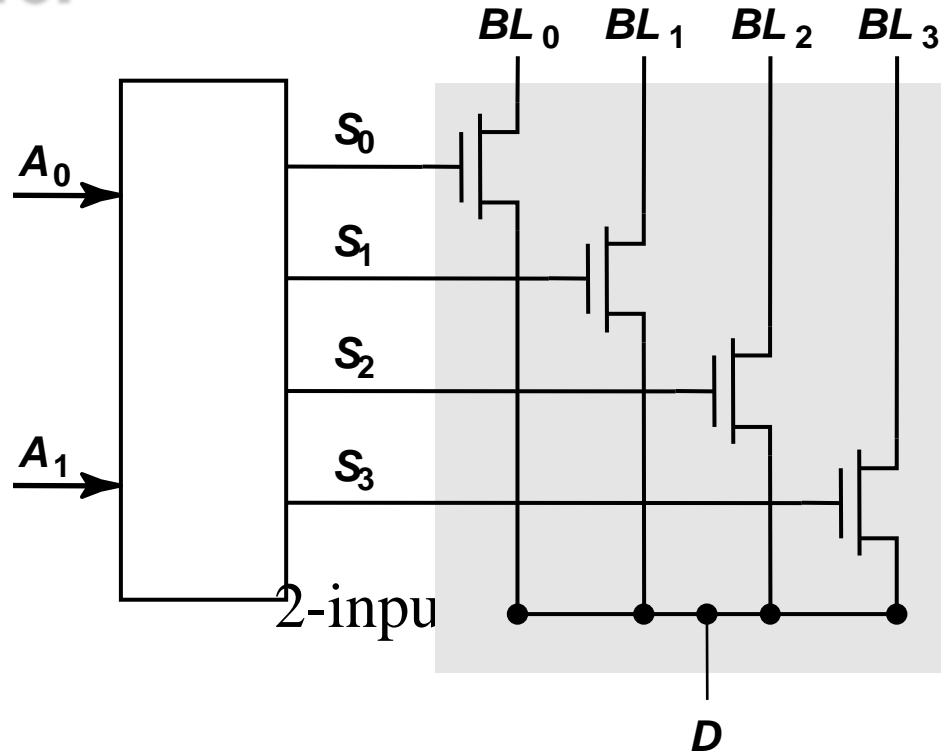
Dynamic Decoders

2-input NOR decoder

2-input NAND decoder
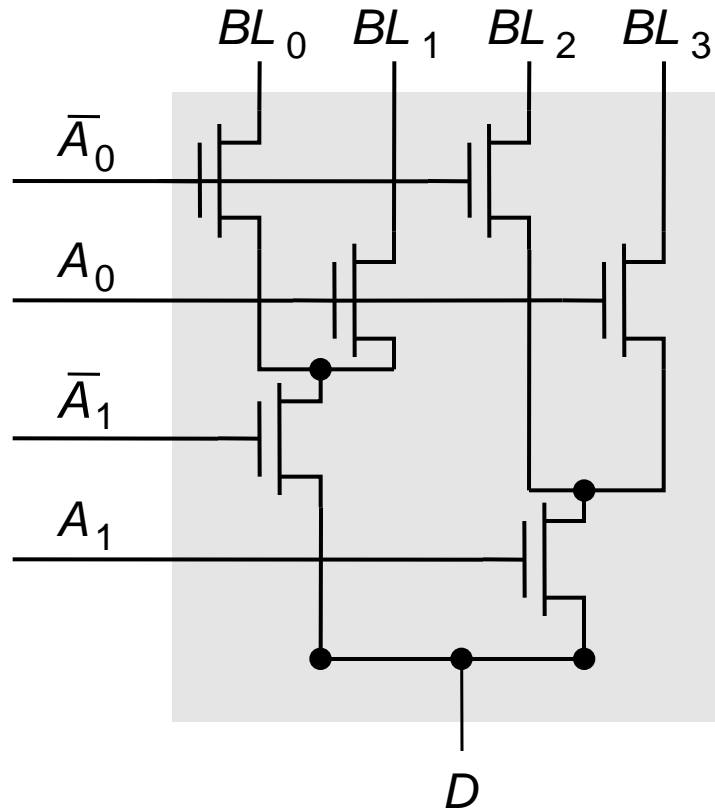
4-input pass-transistor based column decoder

Advantages: speed (t_{pd} does not add to overall memory access time)
   Only one extra transistor in signal path
Disadvantage: Large transistor count

4-to-1 tree based column decoder

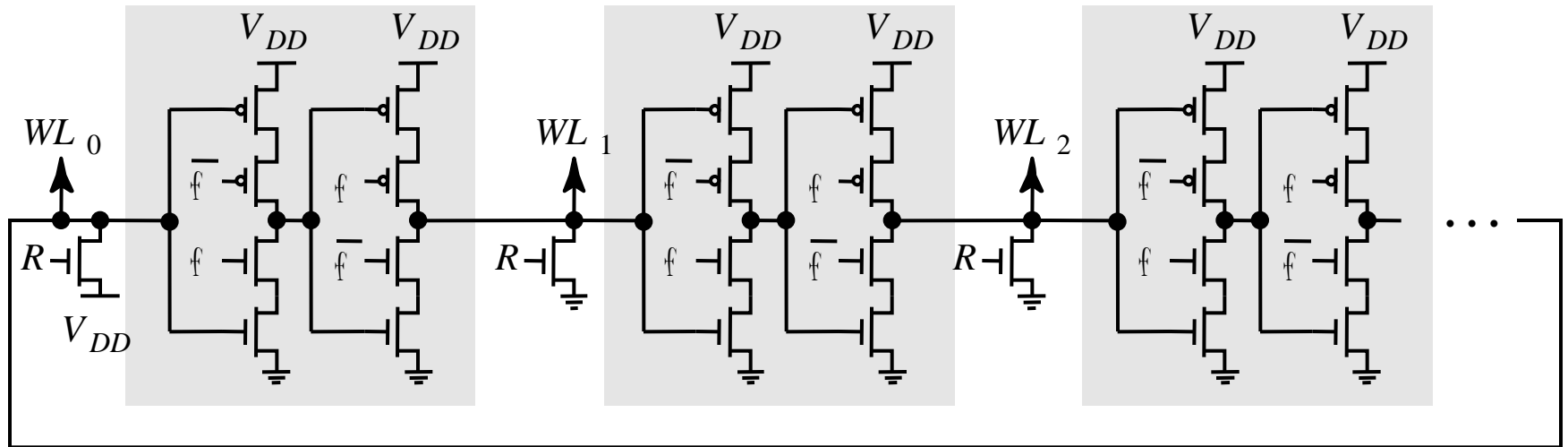**Number of devices drastically reduced**
Delay increases quadratically with # of sections; prohibitive for large decoders
Solutions: buffers
        progressive sizing
        combination of tree and pass transistor approaches

Decoder for circular shift-register
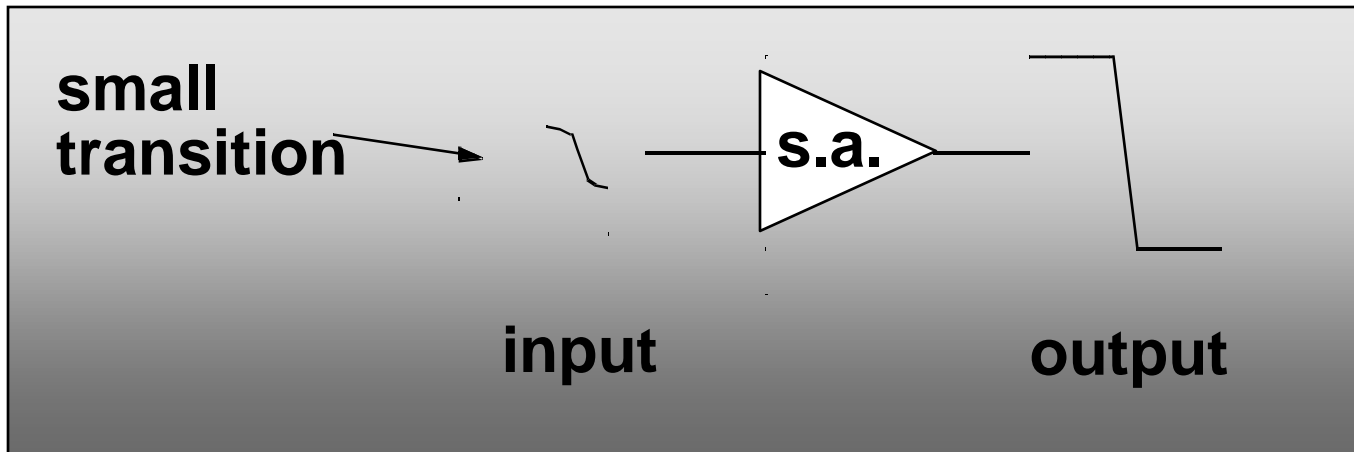
Sense Amplifiers

$$t_p = \frac{C \cdot \Delta V}{I_{av}}$$

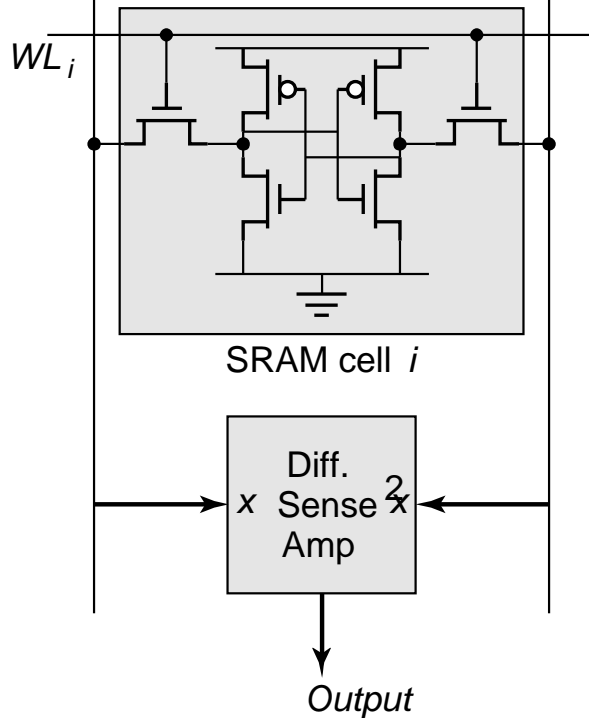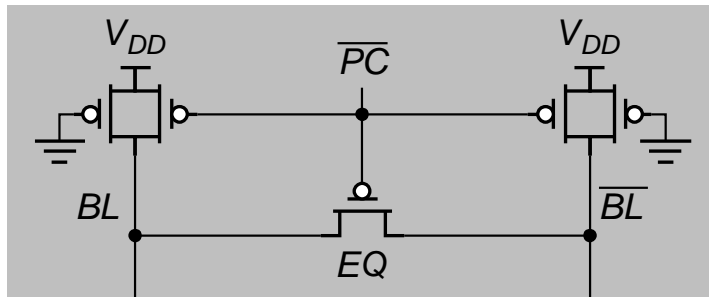**make ΔV as small as possible**

large

small

Idea: Use Sense Amplifer

small transition

s.a.

input

output

Differential Sense Amplifier
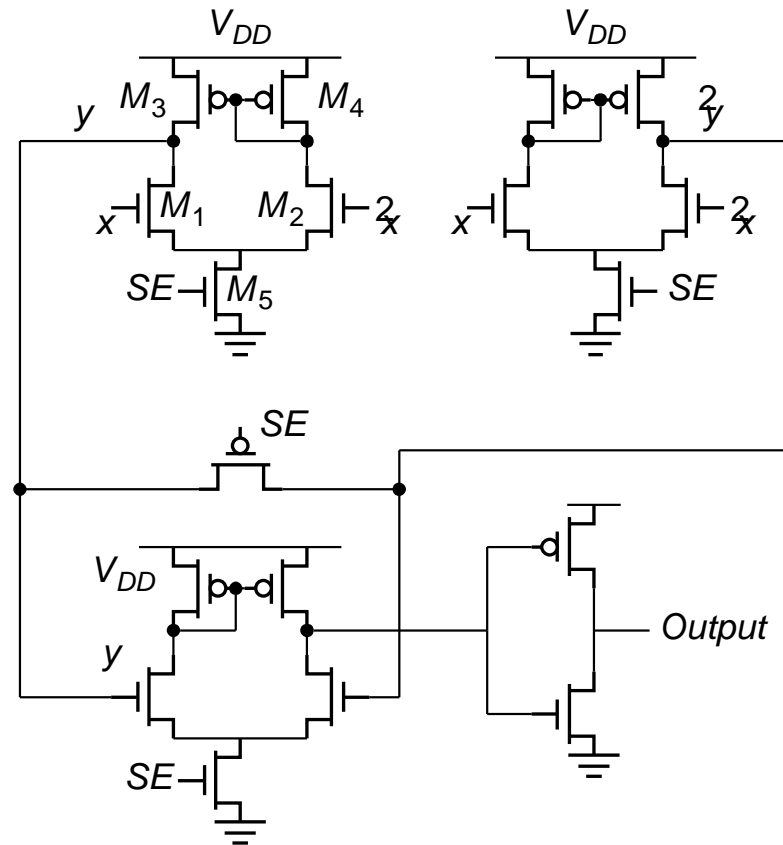

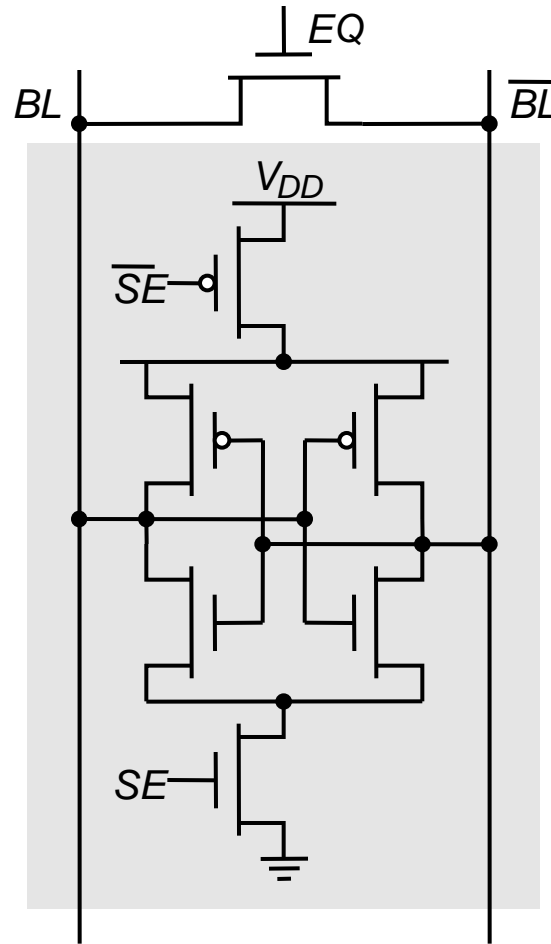**Directly applicable to SRAMs**

Differential Sensing – SRAM

(a) SRAM sensing scheme

(b) two stage differential amplifier
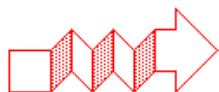
Latch-Based Sense Amplifier (DRAM)

Initialized in its meta-stable point with EQ

Once adequate voltage gap created, sense amp enabled with SE
Positive feedback quickly forces output to a stable operating point.

Reliability and Yield

- Semiconductor memories trade off noise-margin for density and performance

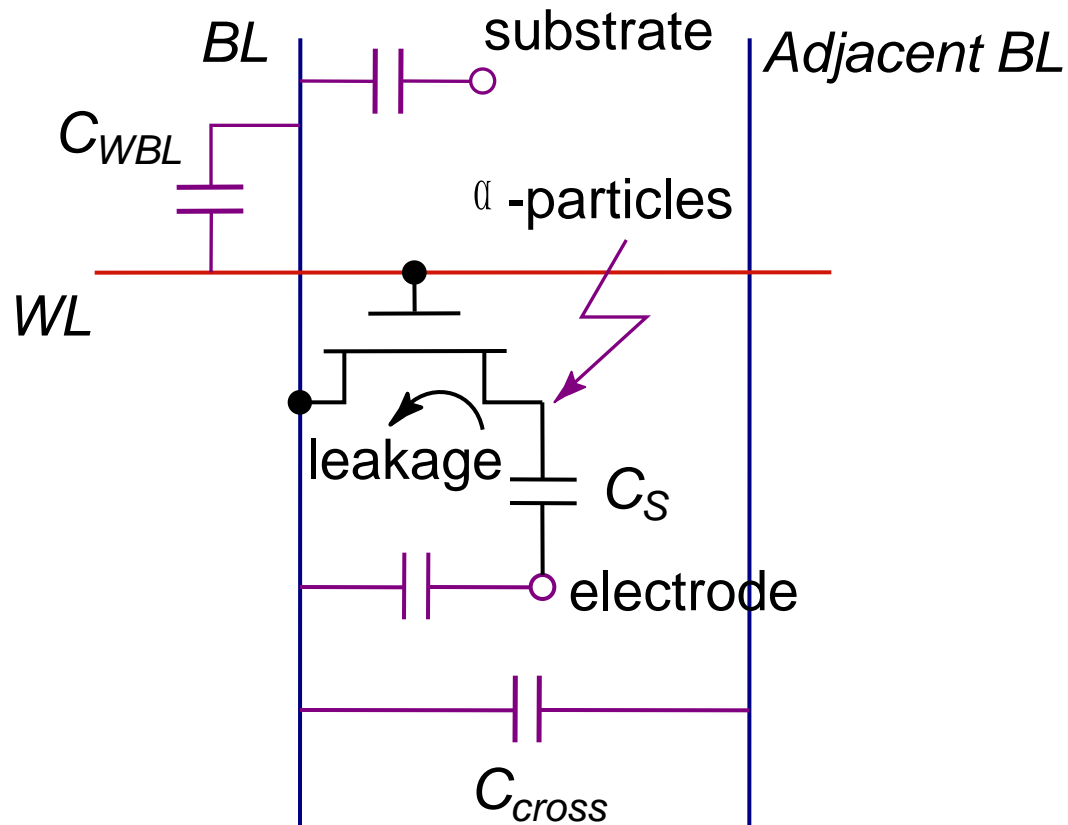Highly Sensitive to Noise (Crosstalk, Supply Noise)

- High Density and Large Die size cause Yield Problems

$$Y = 100 \frac{Number\ of\ Good\ Chips\ on\ Wafer}{Number\ of\ Chips\ on\ Wafer}$$

$$Y = \left[\frac{1 - e^{-AD}}{AD} \right]^2$$

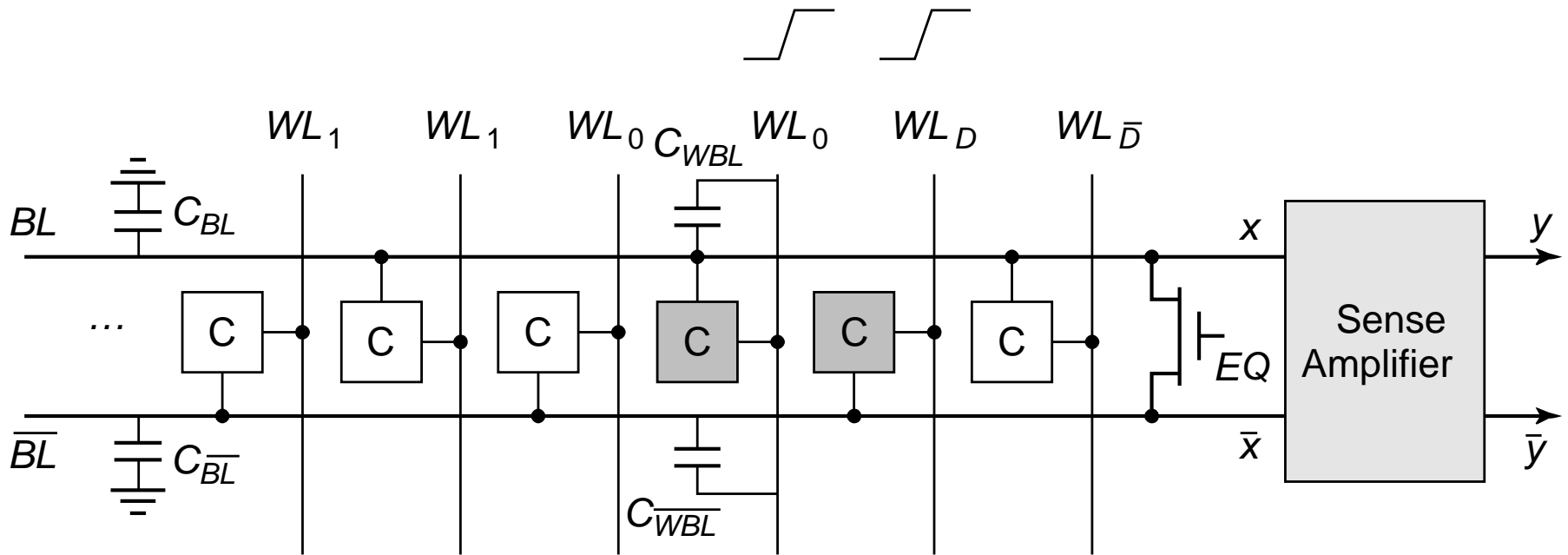**Increase Yield using Error Correction and Redundancy**
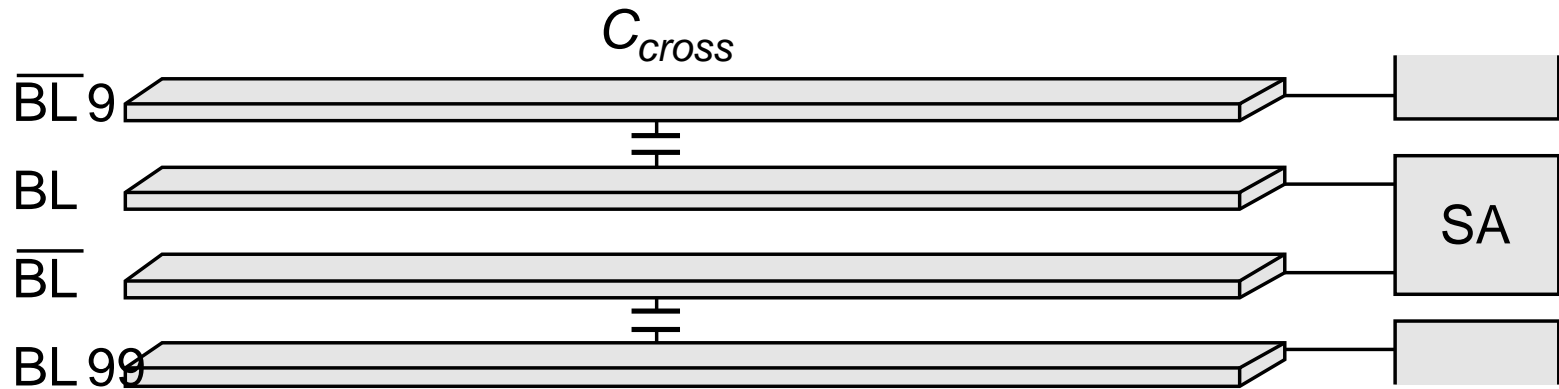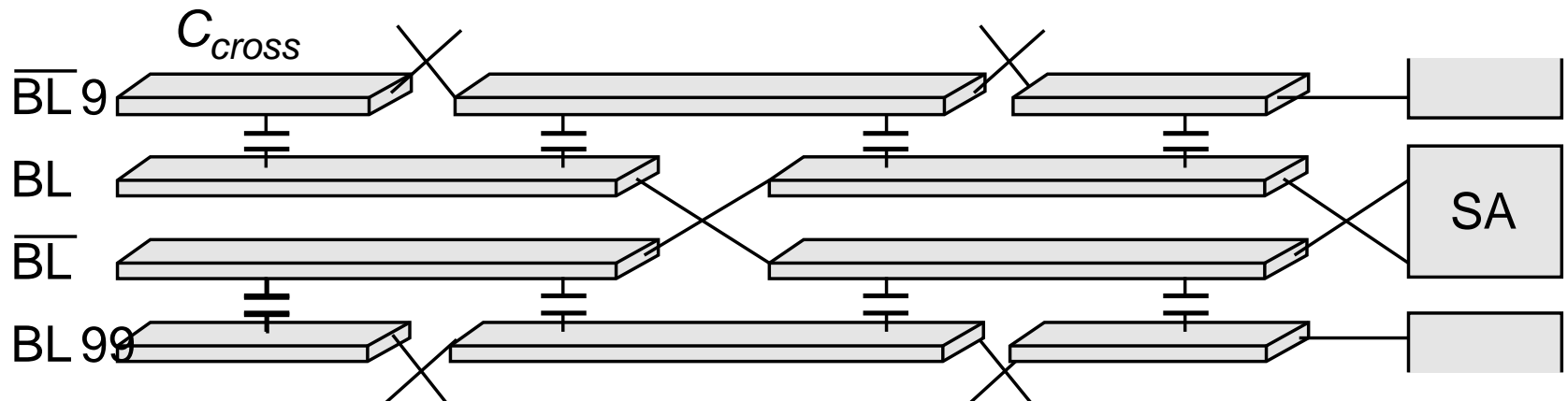
Noise Sources in 1T DRam

BL   substrate   *Adjacent BL*

C_{WBL}

α-particles

WL

leakage C_S

electrode

C_{cross}

Memories

Open Bit-line Architecture —Cross Coupling

Folded-Bitline Architecture

Transposed-Bitline Architecture

C_{cross}

$\overline{BL}$ 9

BL

$\overline{BL}$

BL 99

SA

(a) Straightforward bit-line routing

C_{cross}

$\overline{BL}$ 9

BL

$\overline{BL}$

BL 99

SA

(b) Transposed bit-line architecture

Alpha-particles (or Neutrons)

α -particle

WL

BL

V_{DD}

SiO_2

n^1
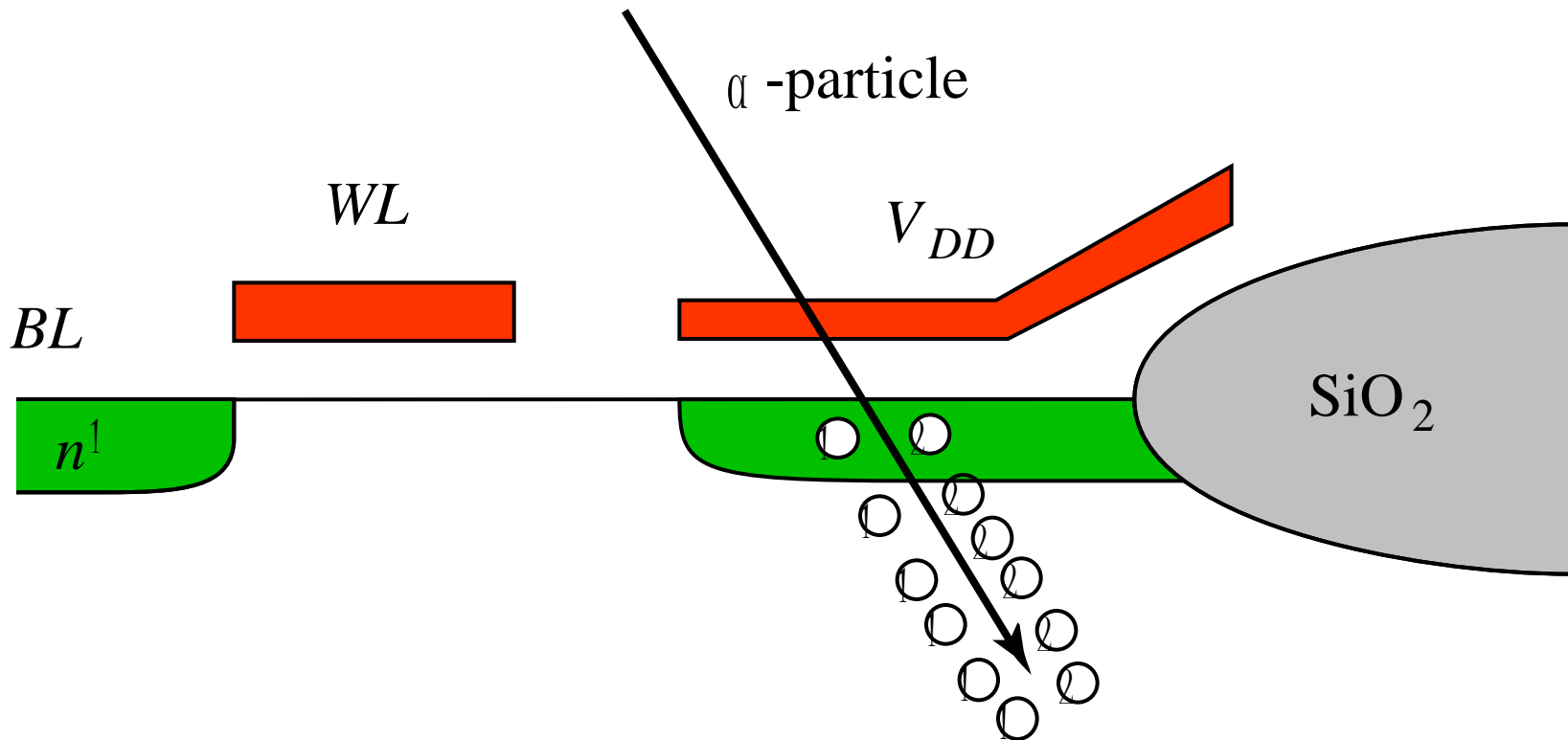
1 Particle ~ 1 Million Carriers

Redundancy

Redundant rows

Redundant columns

Memory Array

Row Decoder

Column Decoder

Row Address

Column Address

: Fuse Bank
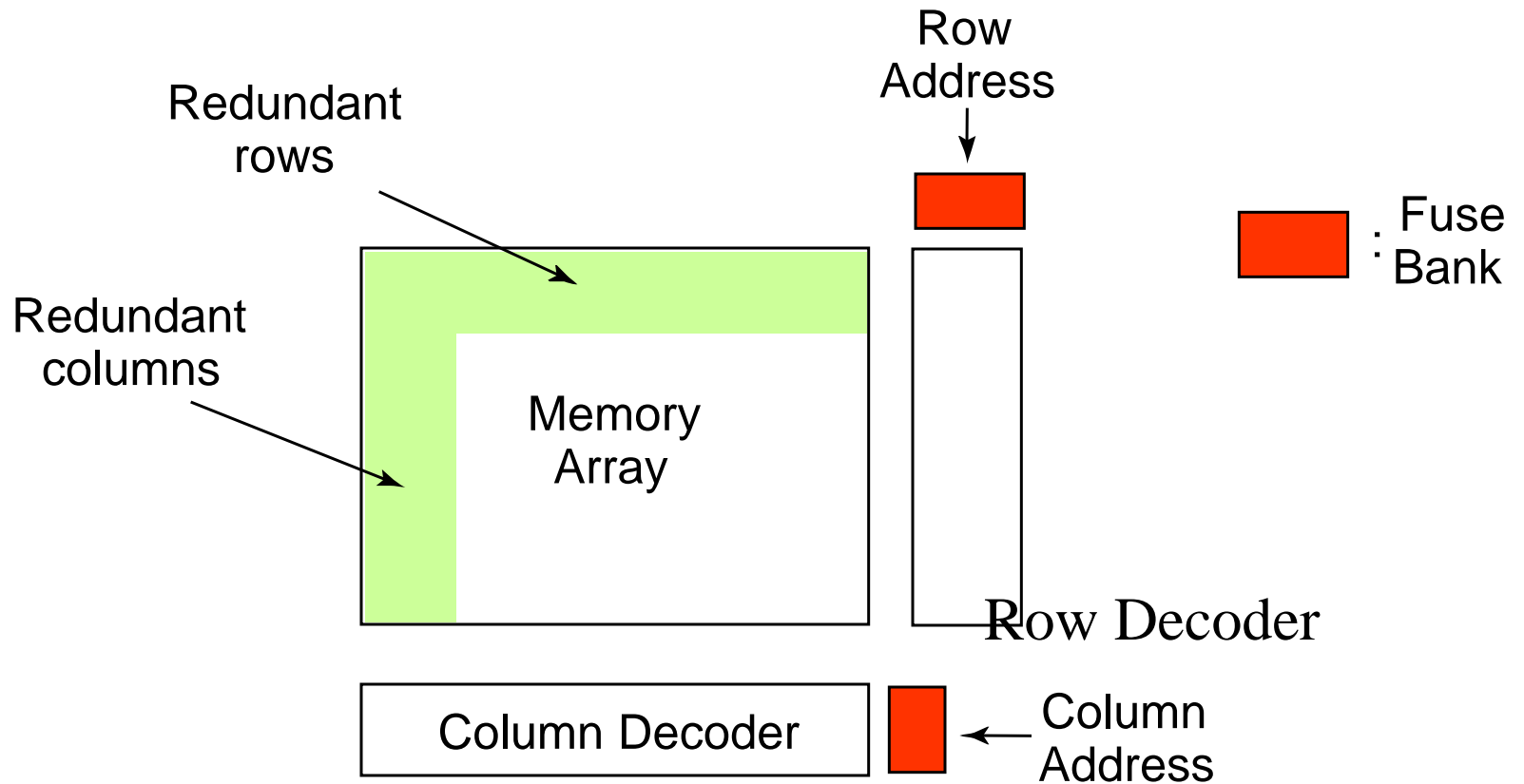
Error-Correcting Codes

Example: Hamming Codes

$$P_1 \ P_2 \ B_3 \ P_4 \ B_5 \ B_6 \ B_7$$

e.g. B3 Wrong

with

$$P_1 \oplus B_3 \oplus B_5 \oplus B_7 = 0$$ **1**

$$P_2 \oplus B_3 \oplus B_6 \oplus B_7 = 0$$ **1** *= 3*
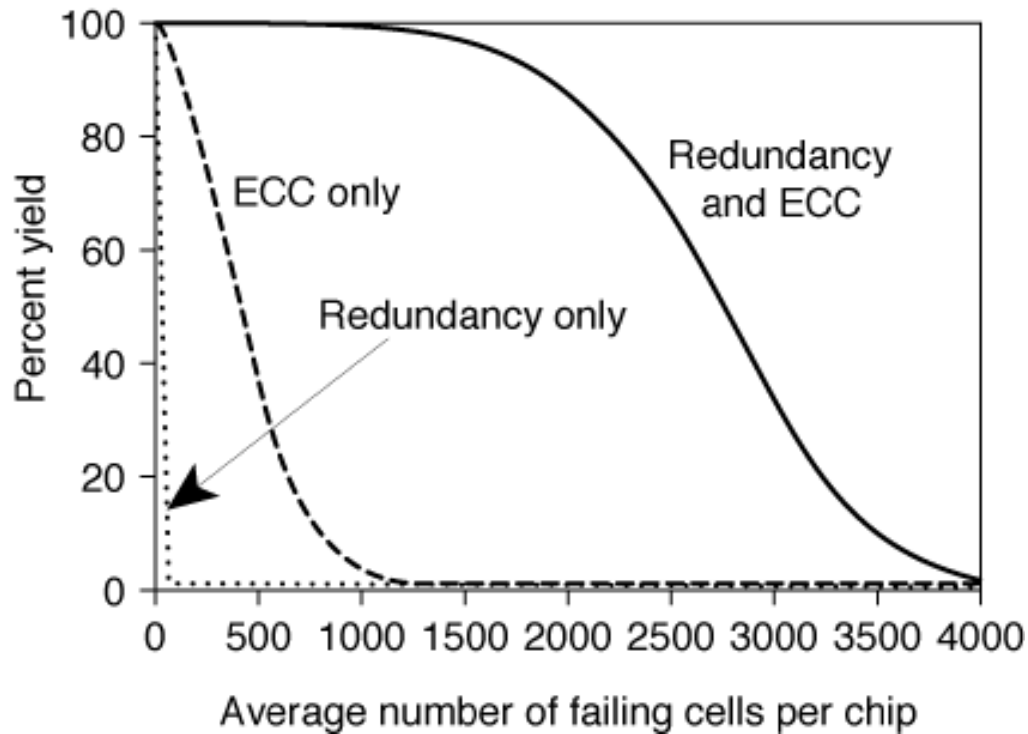
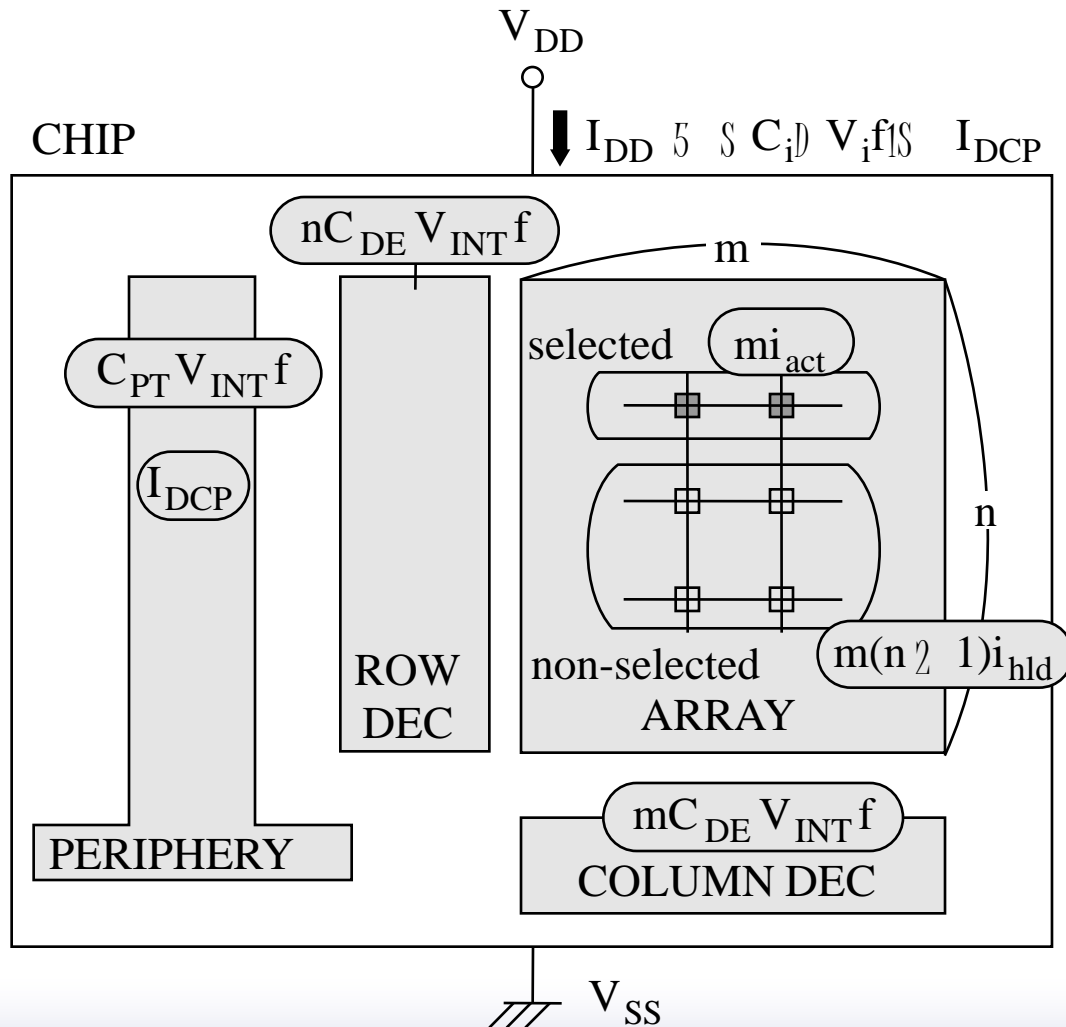$$P_4 \oplus B_5 \oplus B_6 \oplus B_7 = 0$$ **0**

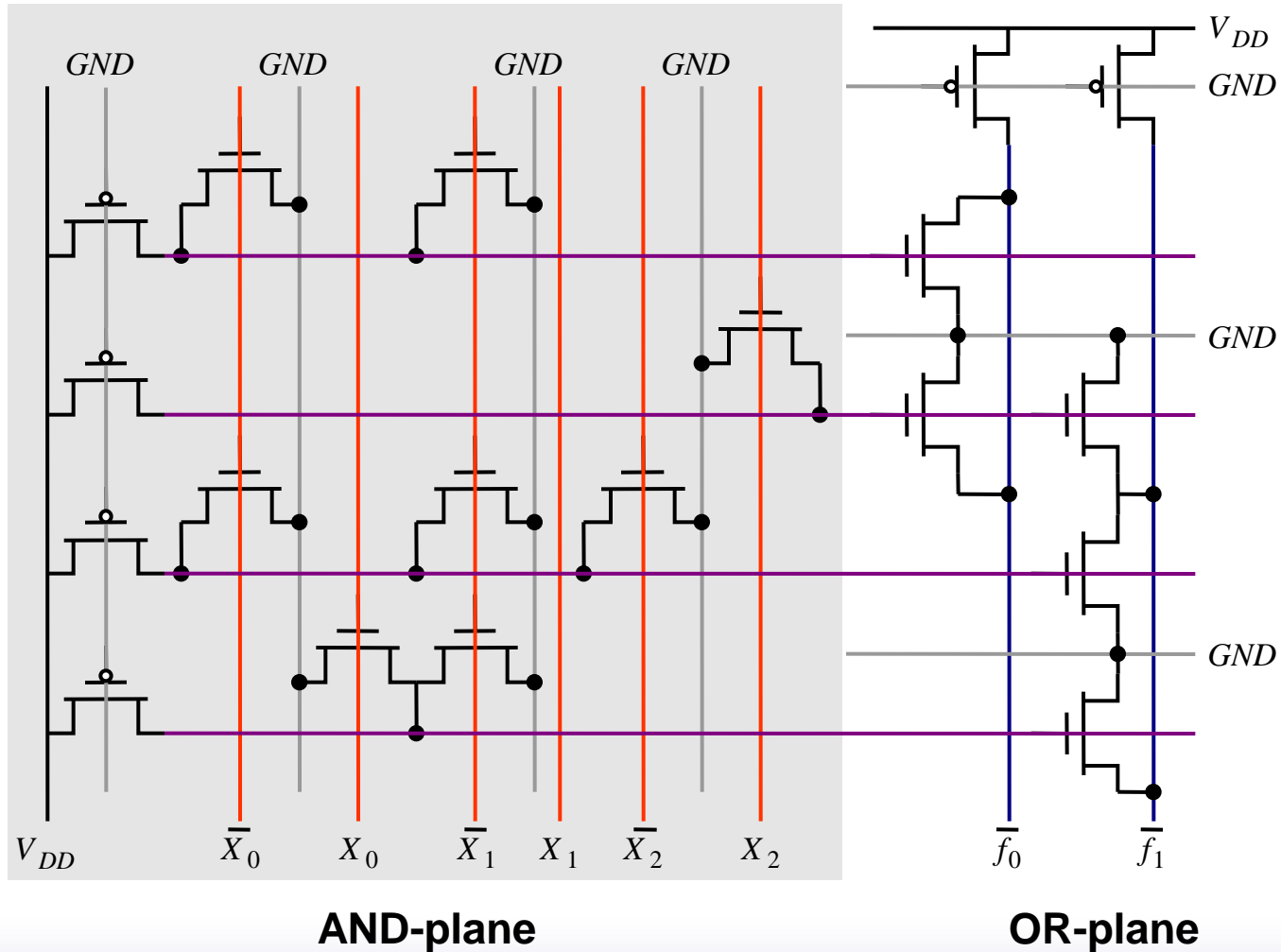# *Redundancy and Error Correction*

Sources of Power Dissipation in Memories

From [Itoh00]

Programmable Logic Array

Pseudo-NMOS PLA
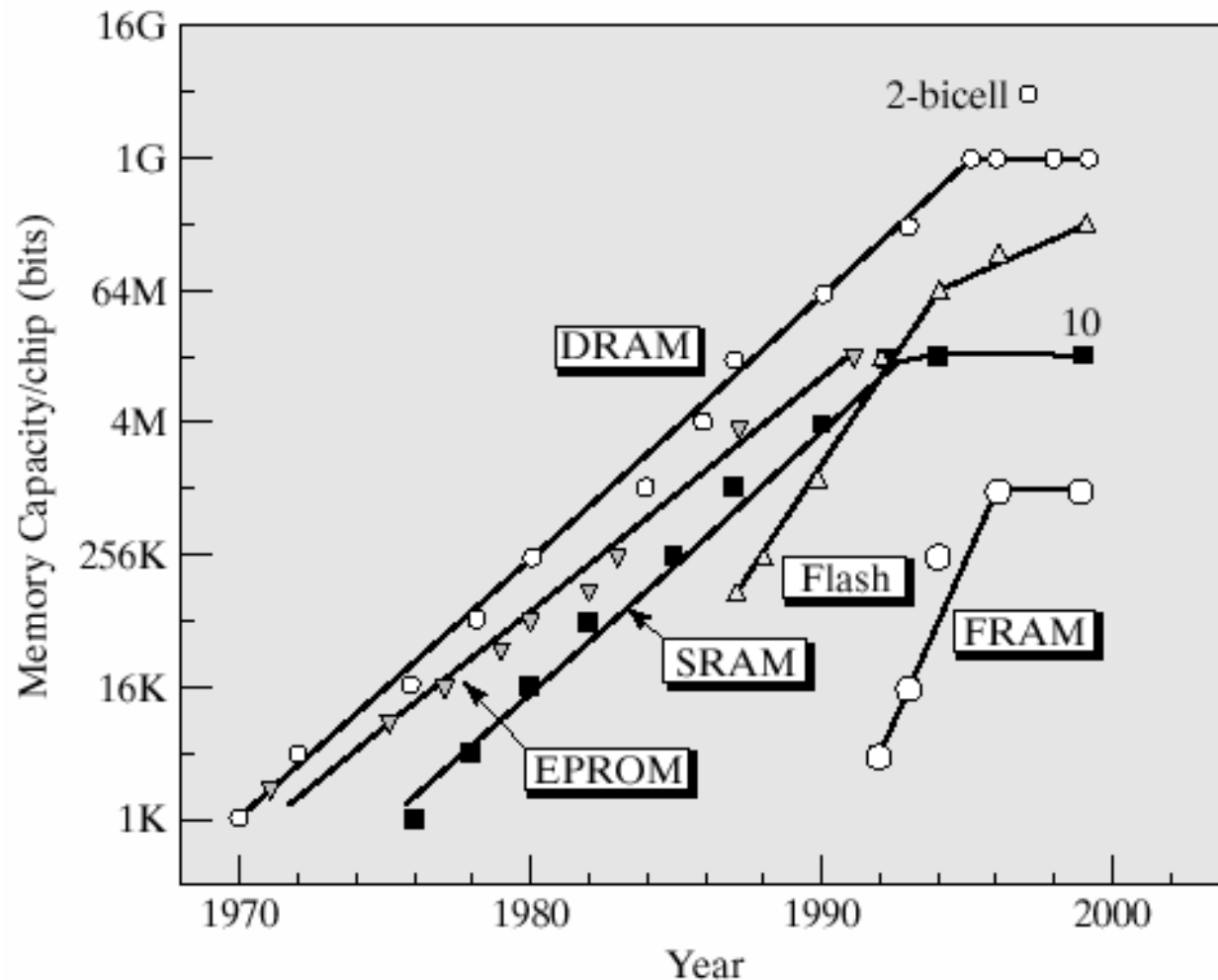

AND-plane   **OR-plane**

Dynamic PLA

AND-plane

OR-plane
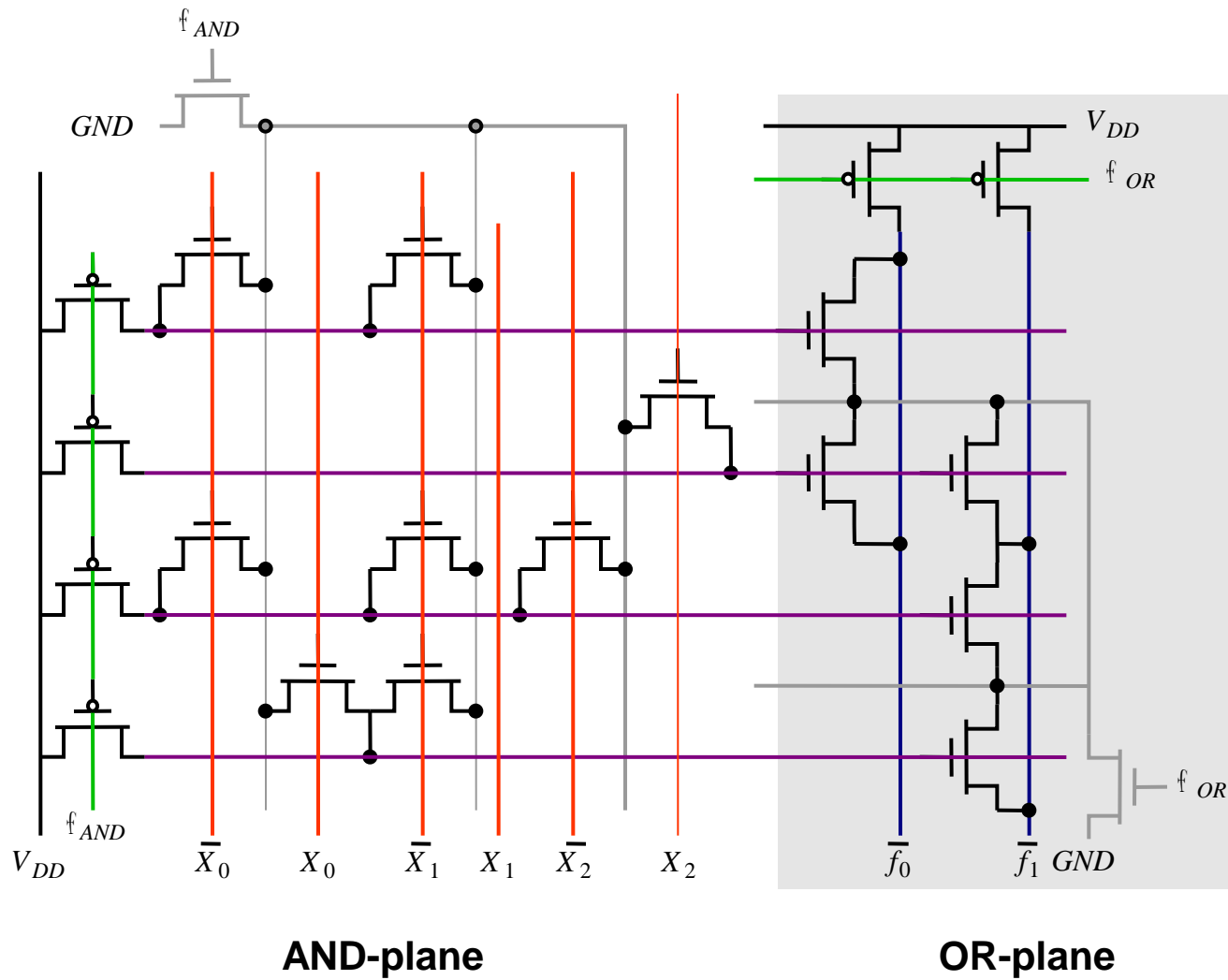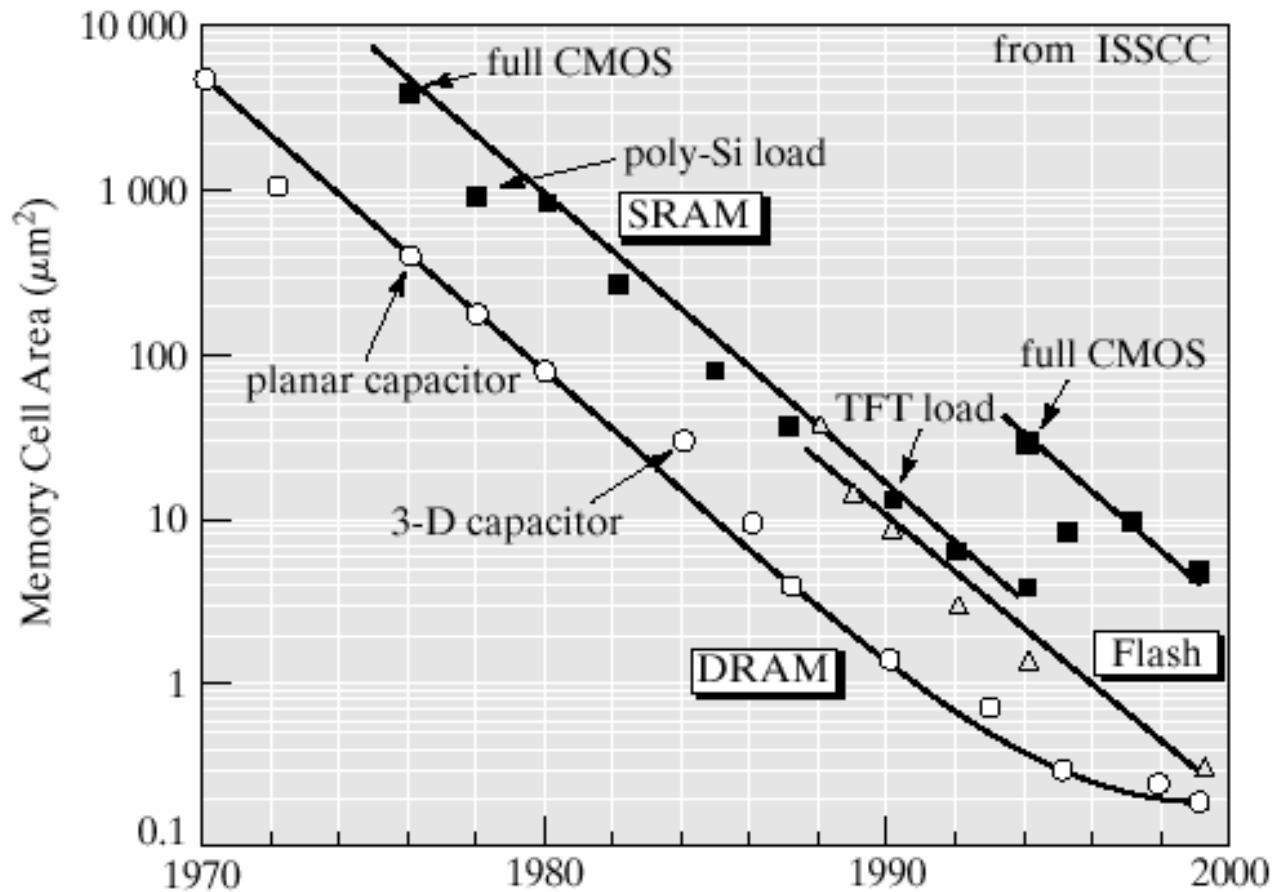
Semiconductor Memory Trends (updated)

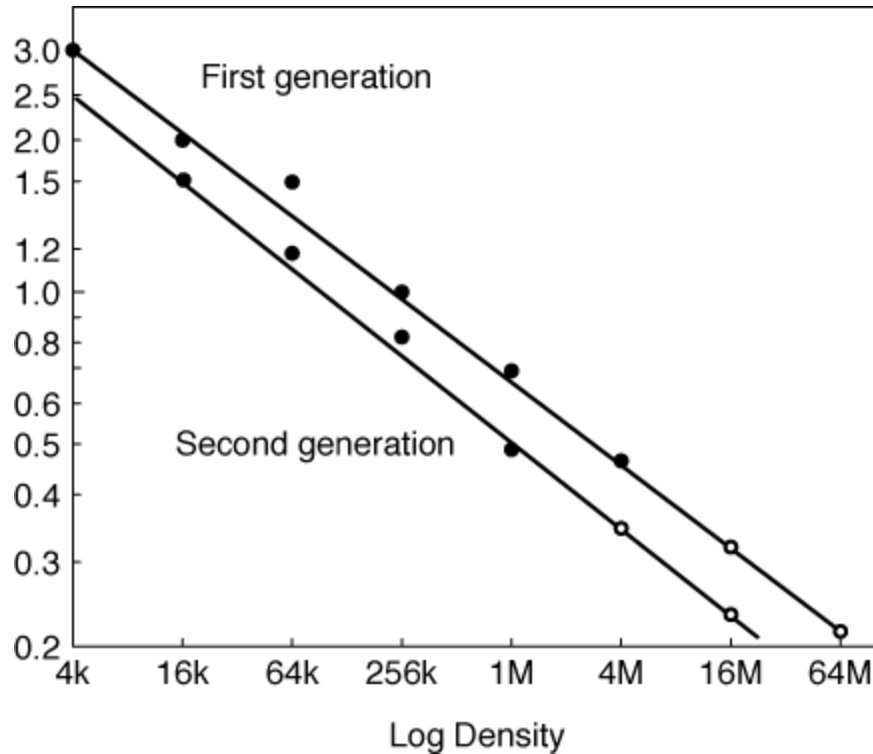**From [Itoh01]**

Trends in Memory Cell Area

From [Itoh01]

Semiconductor Memory Trends

Technology feature size for different SRAM generations