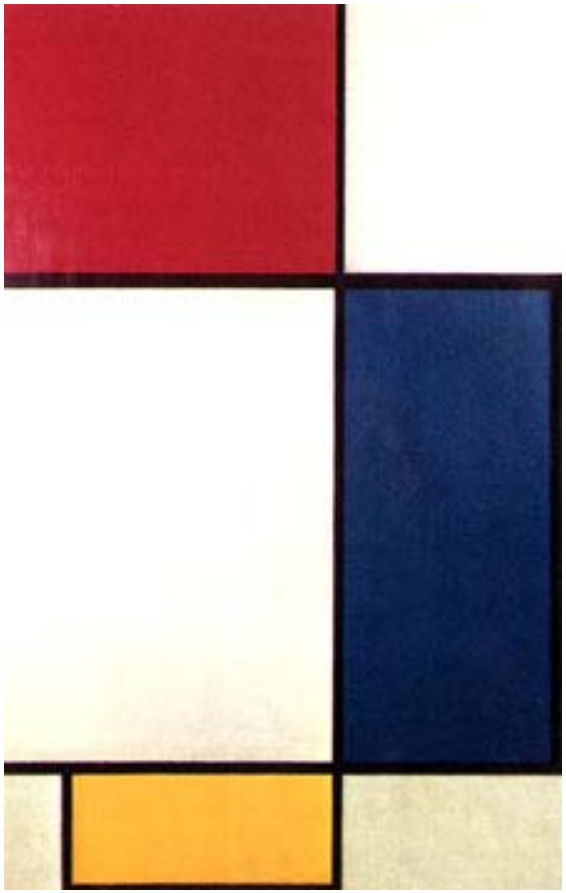
SRAM Design



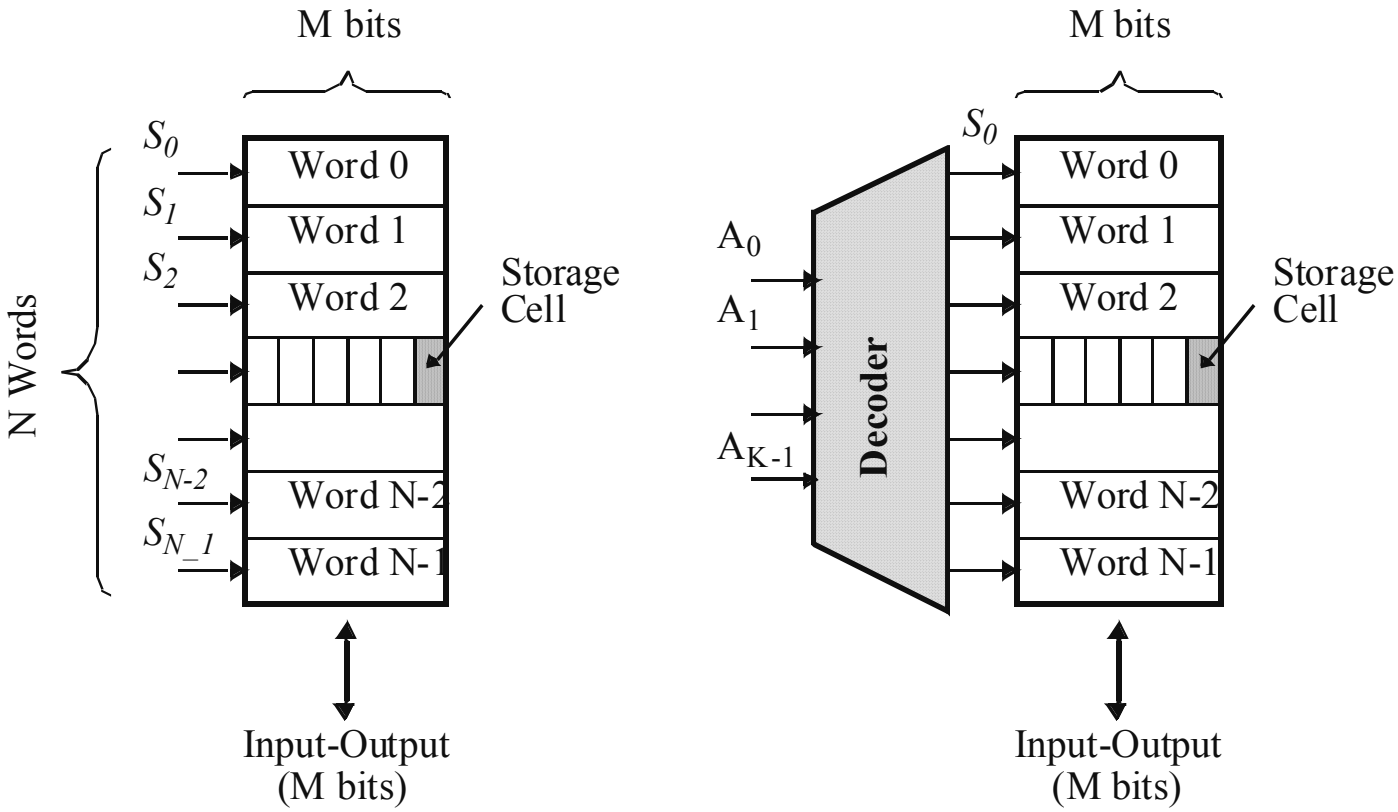
Chapter Overview

- **Memory Classification**
- **Memory Architectures**
- **The Memory Core**
- **Periphery**
- **Reliability**

Semiconductor Memory Classification

RWM		NVRWM	ROM
Random Access	**Non-Random Access**	EPROM E^2PROM FLASH	Mask-Programmed Programmable (PROM)
SRAM DRAM	FIFO LIFO Shift Register CAM		

Memory Architecture: Decoders

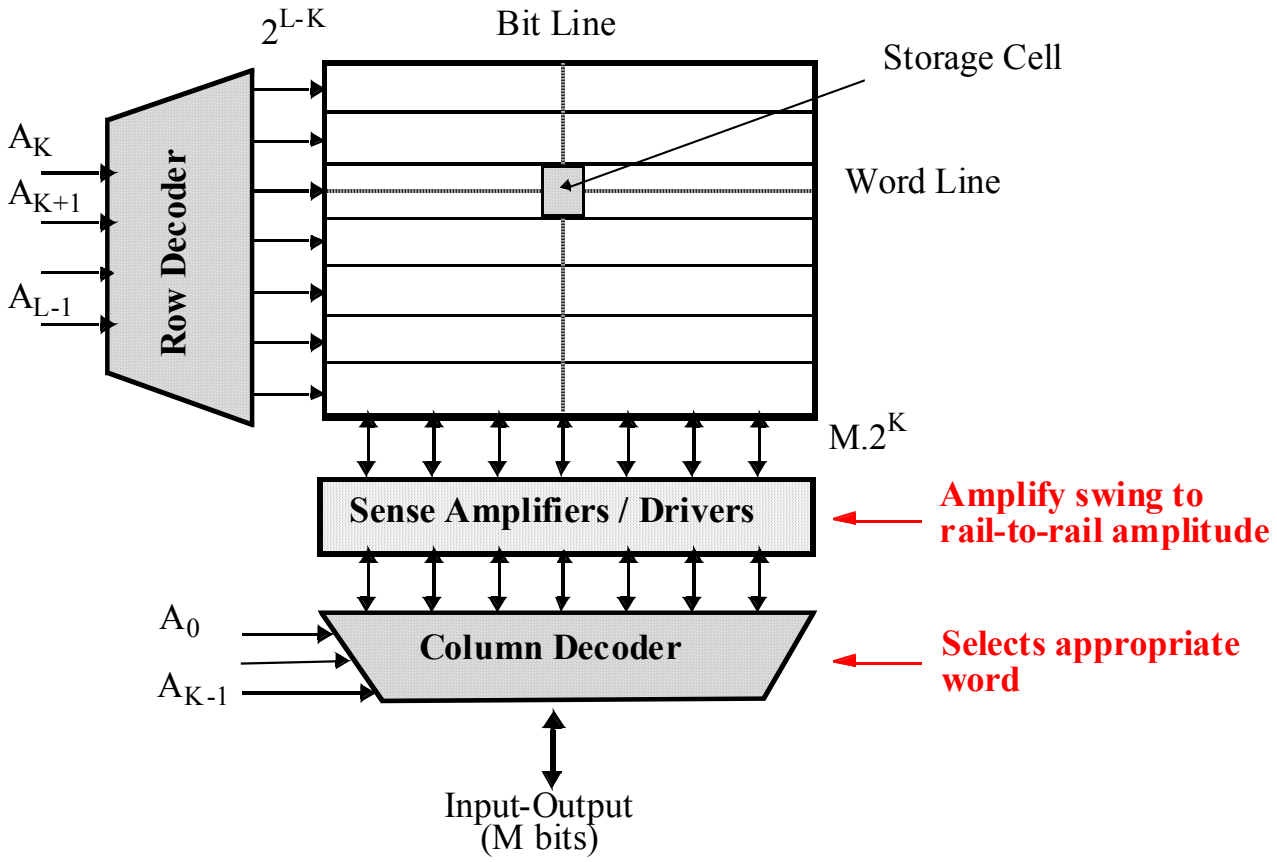


N words => N select signals
Too many select signals

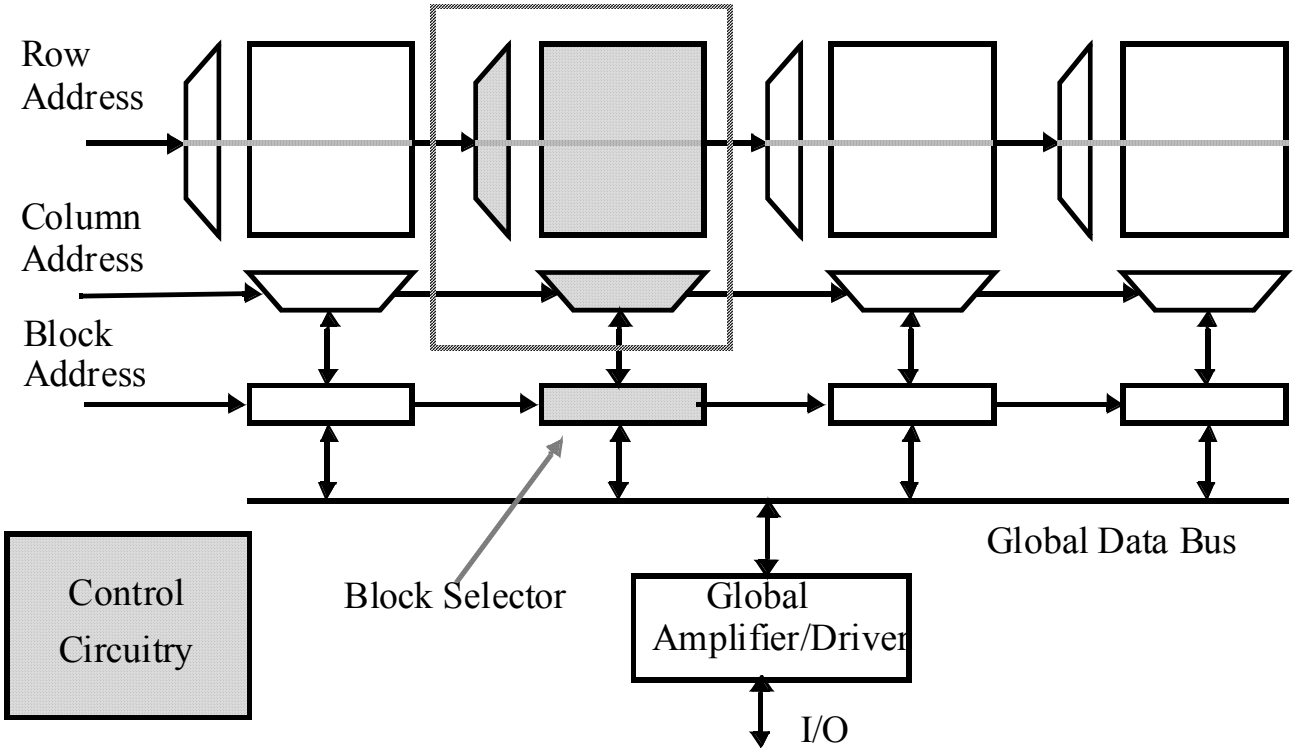
Decoder reduces # of select signals
$K = log_2 N$

Array-Structured Memory Architecture

Problem: ASPECT RATIO or HEIGHT >> WIDTH

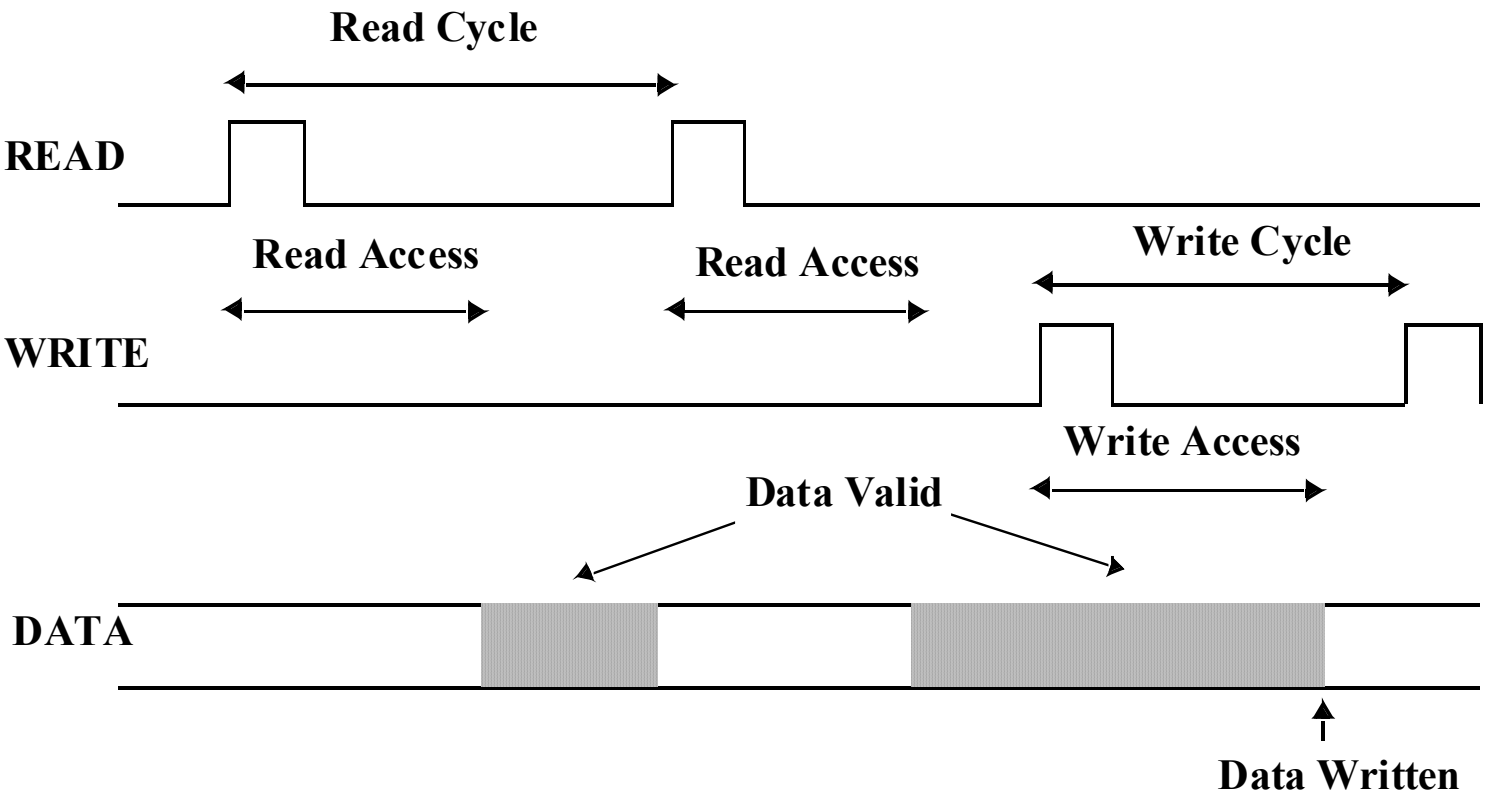


Hierarchical Memory Architecture

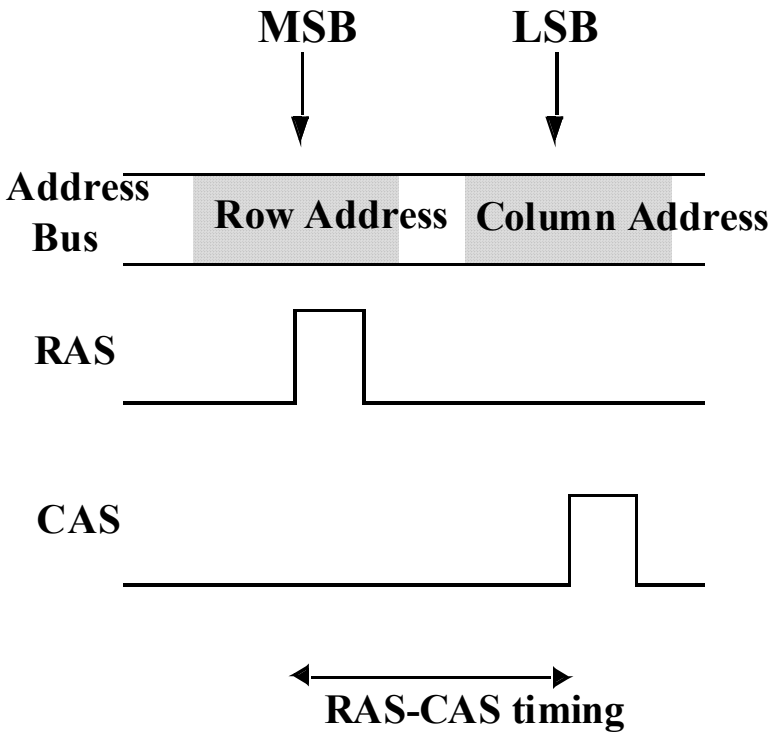


Advantages:
1. Shorter wires within blocks
2. Block address activates only 1 block => power savings

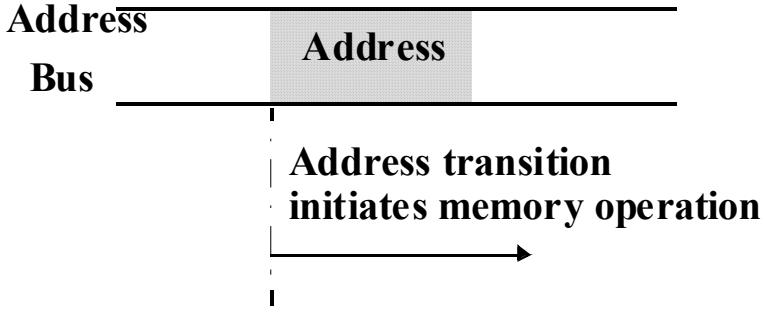
Memory Timing: Definitions



Memory Timing: Approaches



DRAM Timing
Multiplexed Adressing



SRAM Timing
Self-timed

Read-Write Memories (RAM)

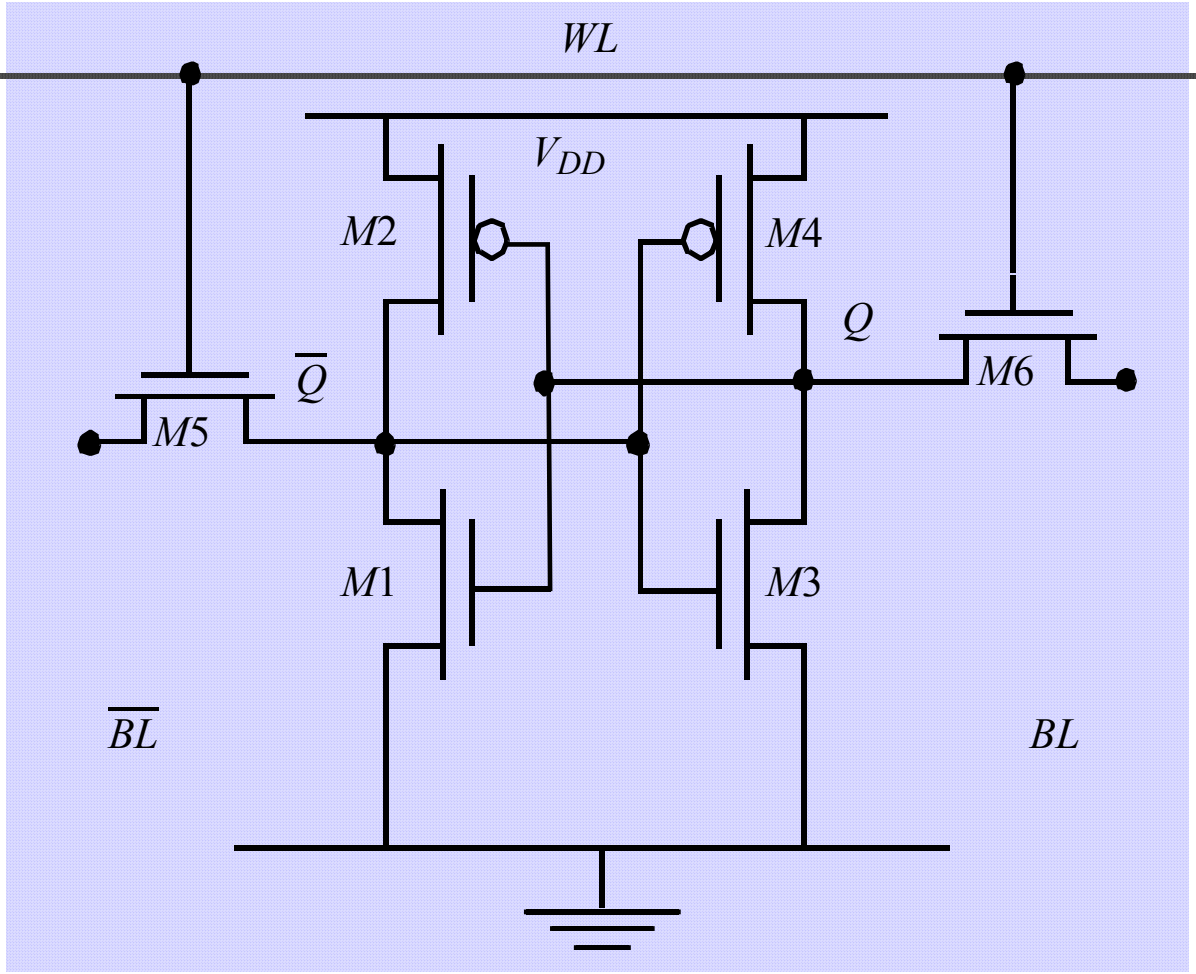
- **STATIC (SRAM)**

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

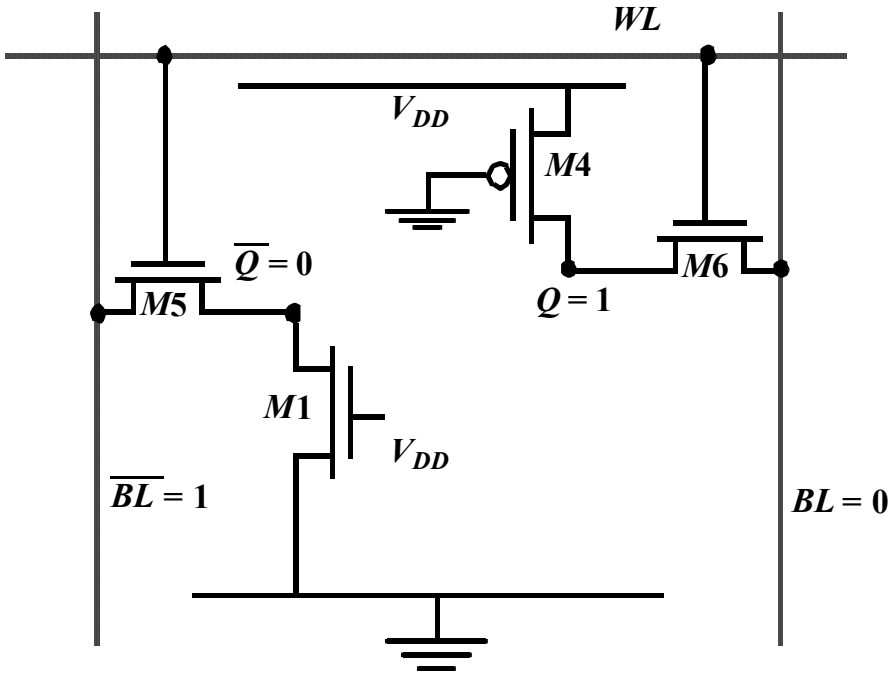
- **DYNAMIC (DRAM)**

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

6-transistor CMOS SRAM Cell



CMOS SRAM Analysis (Write)



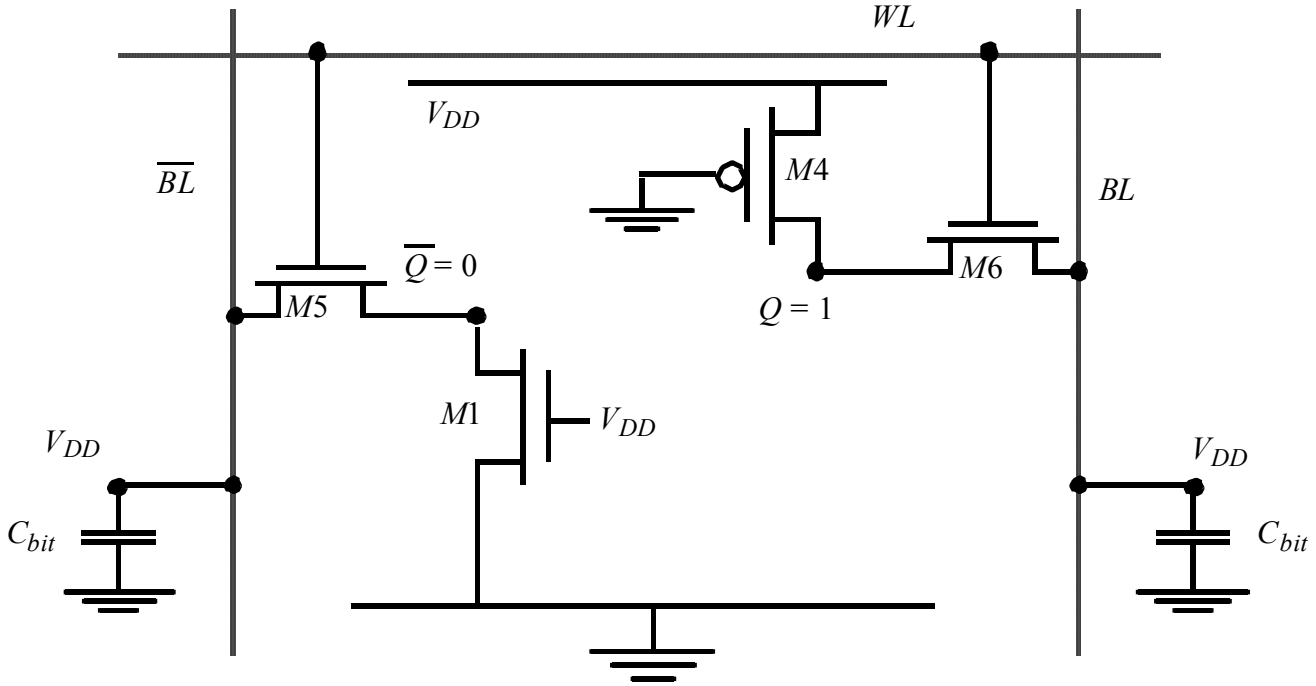
$$k_{n,\,M6}\left((V_{DD}-V_{Tn})\frac{V_{DD}}{2}-\frac{V_{DD}^2}{8}\right) = k_{p,\,M4}\left((V_{DD}-|V_{Tp}|)\frac{V_{DD}}{2}-\frac{V_{DD}^2}{8}\right)$$

$$(W/L)_{n,M6} \geq 0.33\,(W/L)_{p,M4}$$

$$\frac{k_{n,\,M5}}{2}\left(\frac{V_{DD}}{2}-V_{Tn}\left(\frac{V_{DD}}{2}\right)\right)^2 = k_{n,\,M1}\left((V_{DD}-|V_{Tn}|)\frac{V_{DD}}{2}-\frac{V_{DD}^2}{8}\right)$$

$$(W/L)_{n,M5} \geq 10\,(W/L)_{n,M1}$$

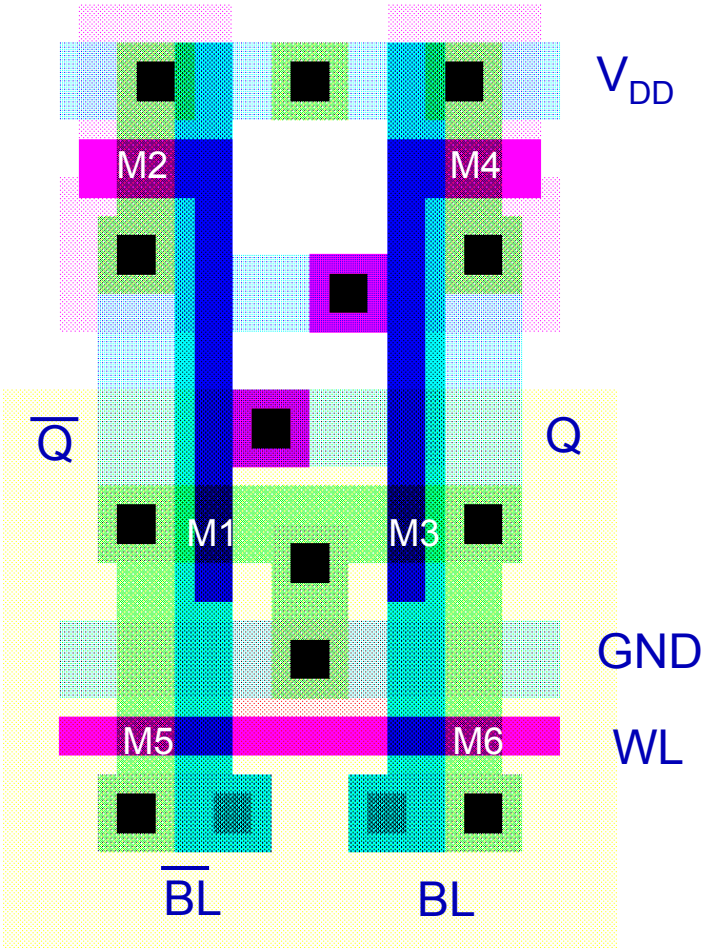
CMOS SRAM Analysis (Read)



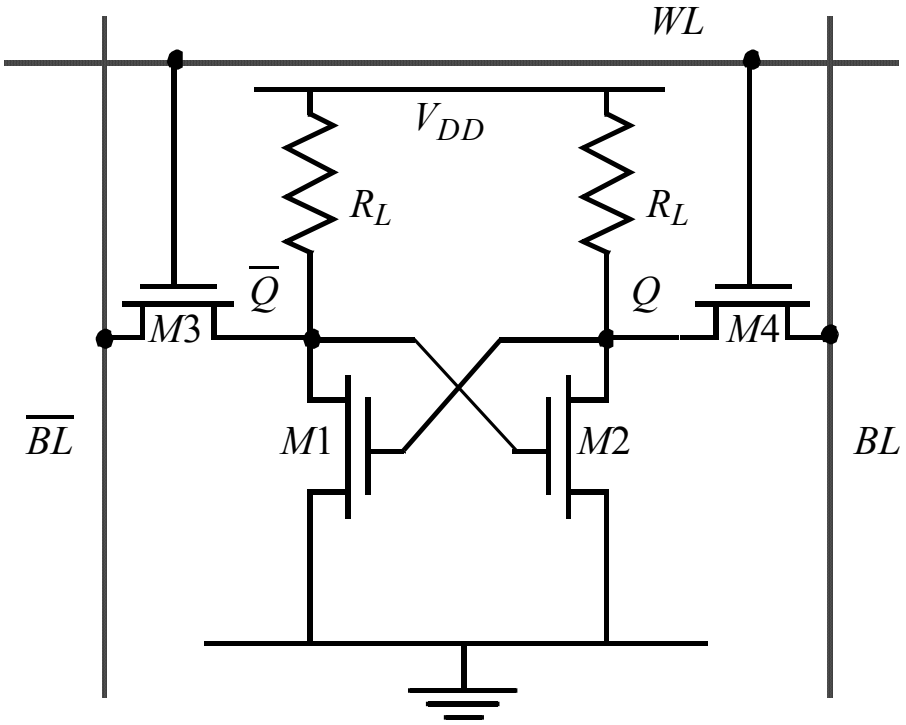
$$\frac{k_{n,\,M5}}{2}\left(V_{DD} - V_{Tn}\left(\frac{V_{DD}}{2}\right)\right)^2 = k_{n,\,M1}\left((V_{DD} - |V_{Tn}|)\frac{V_{DD}}{2} - \frac{V_{DD}^2}{8}\right)$$

$(W/L)_{n,M5} \le 10\,(W/L)_{n,M1}$ **(supercedes read constraint)**

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

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

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

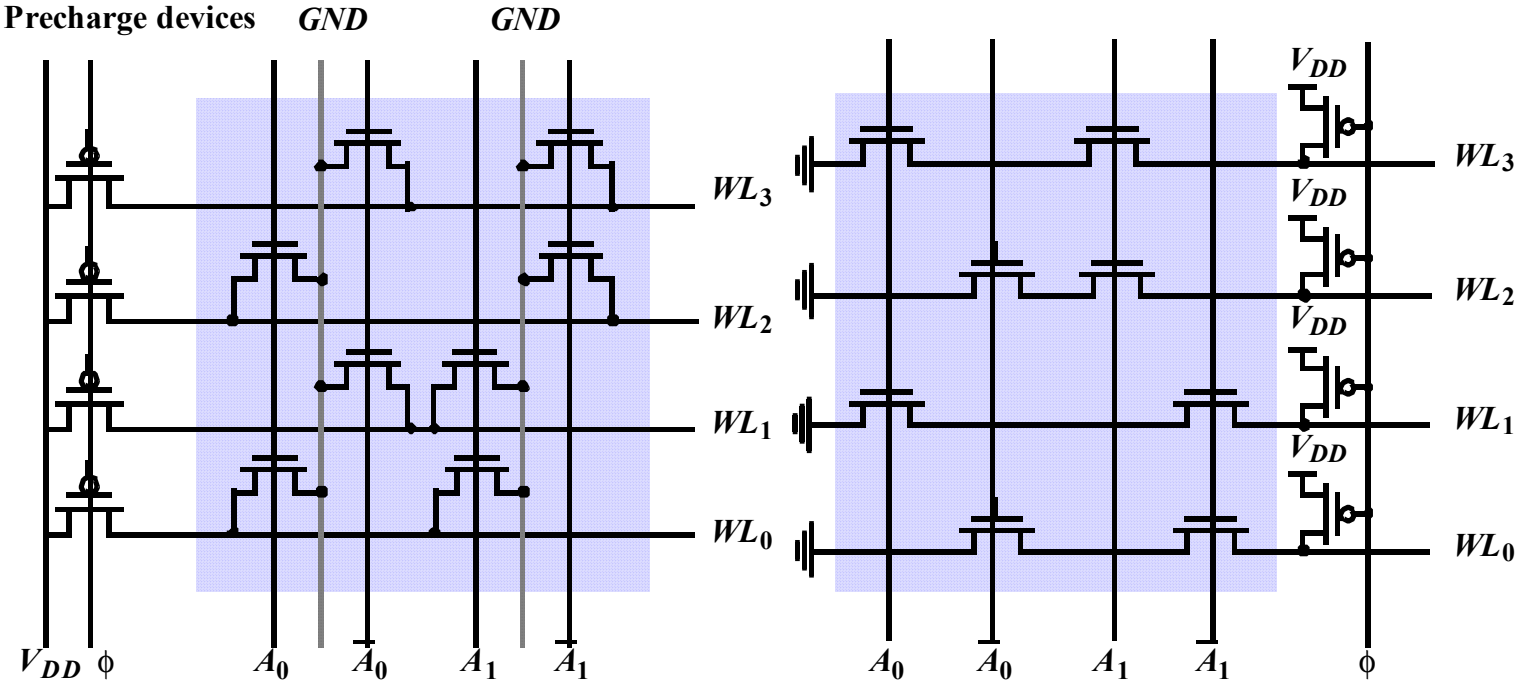
$$WL_{511} = \bar{A}_0 A_1 A_2 A_3 A_4 A_5 A_6 A_7 A_8 A_9$$

NOR Decoder

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

$$WL_{511} = \overline{A_0 + \bar{A}_1 + \bar{A}_2 + \bar{A}_3 + \bar{A}_4 + \bar{A}_5 + \bar{A}_6 + \bar{A}_7 + \bar{A}_8 + \bar{A}_9}$$

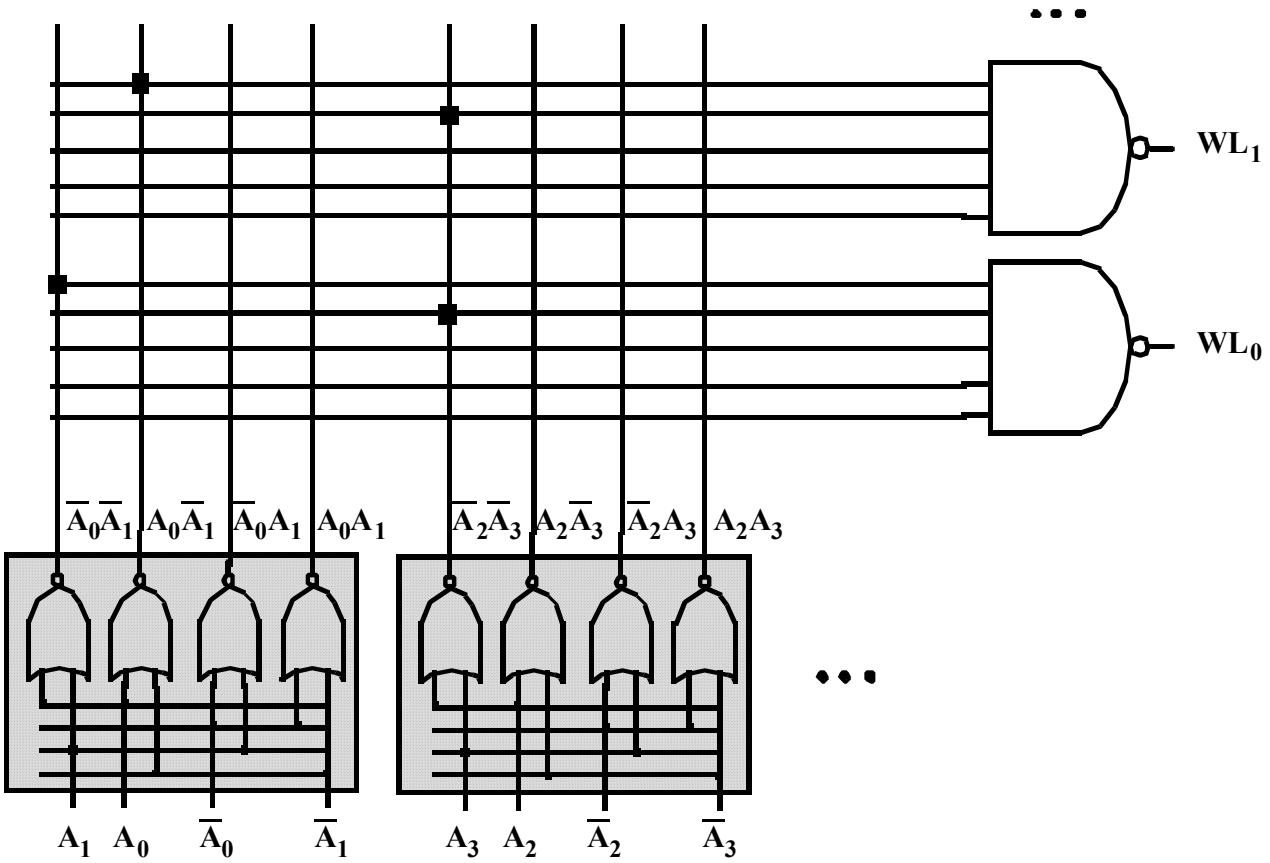
Dynamic Decoders



Dynamic 2-to-4 NOR decoder **2-to-4 MOS dynamic NAND Decoder**

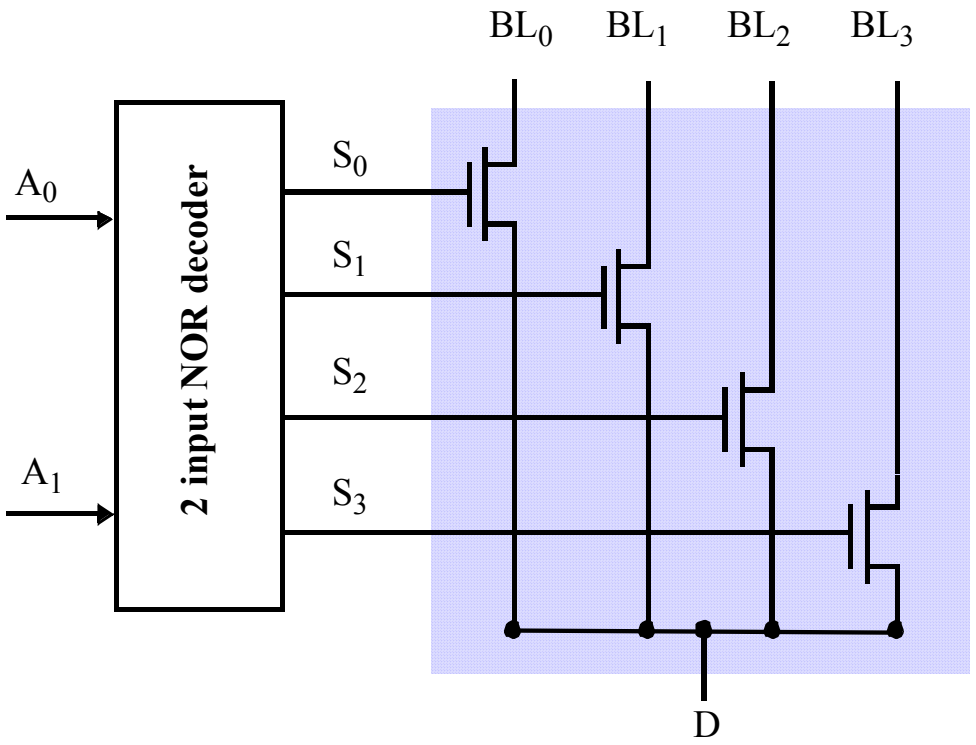
Propagation delay is primary concern

A NAND decoder using 2-input pre-decoders



Splitting decoder into two or more logic layers produces a faster and cheaper implementation

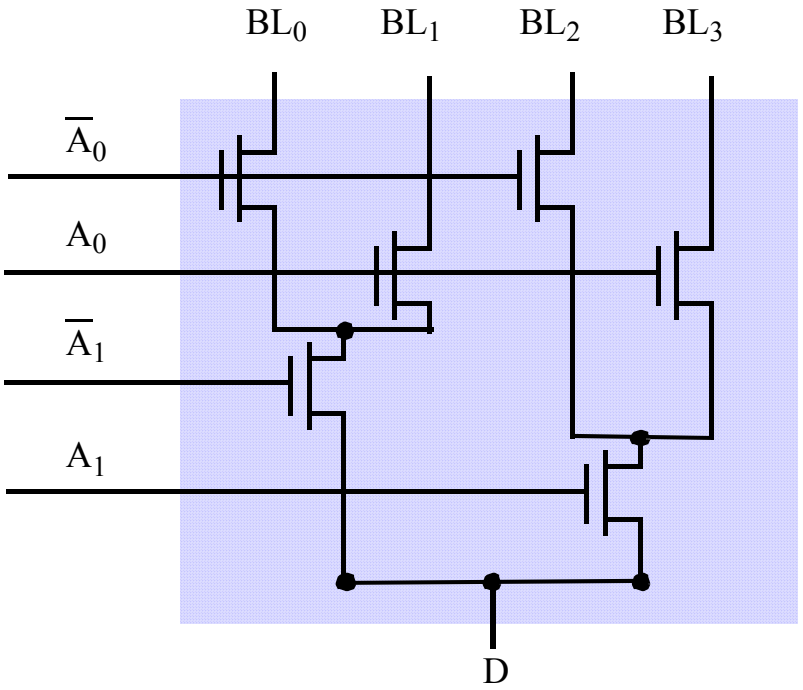
4 input pass-transistor based column decoder



lvantage: speed (t_{pd} does not add to overall memory access time)
only 1 extra transistor in signal path

sadvantage: large transistor count

4-to-1 tree based column decoder



Number of devices drastically reduced

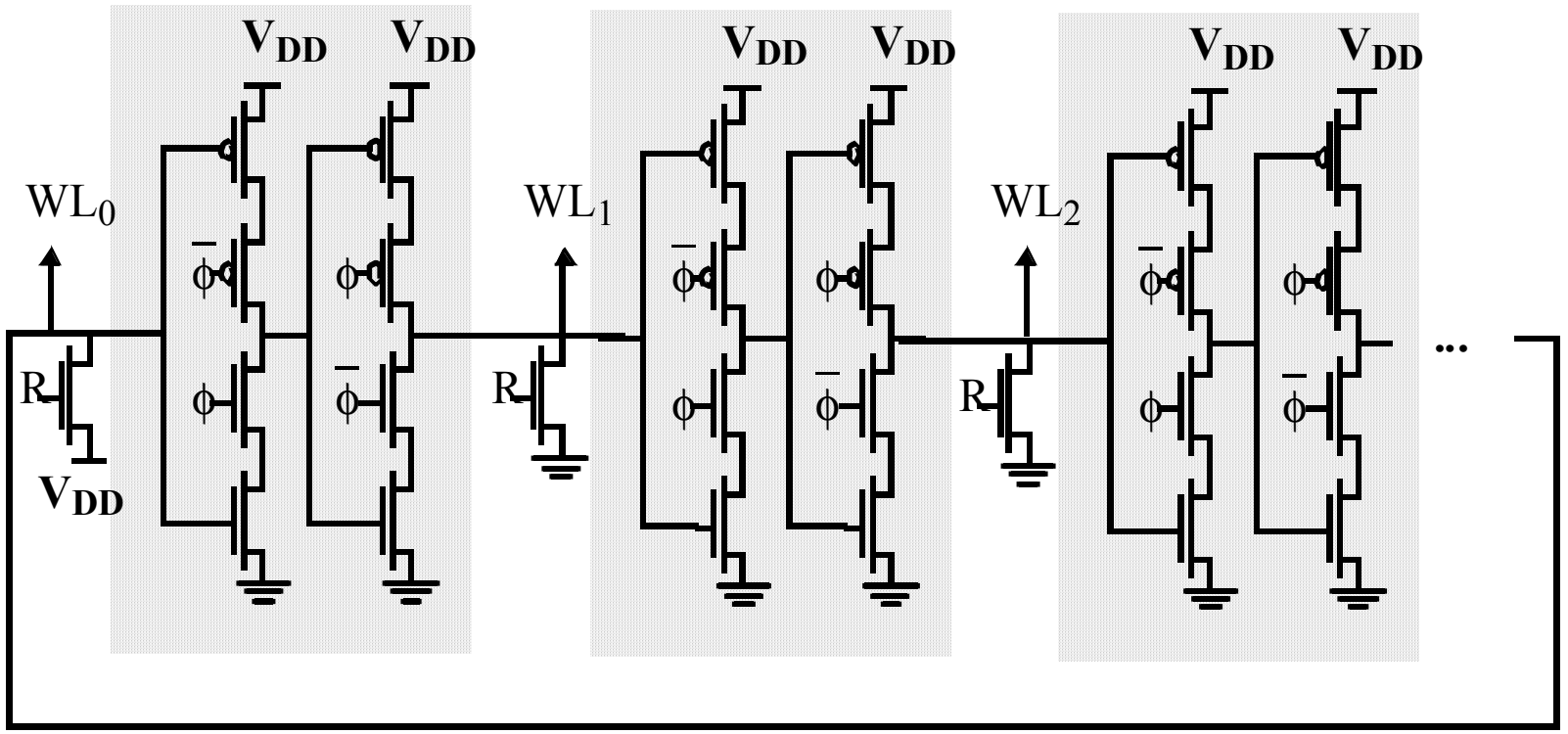
Delay increases quadratically with # of sections; prohibitive for large decoders

Solutions: **buffers**

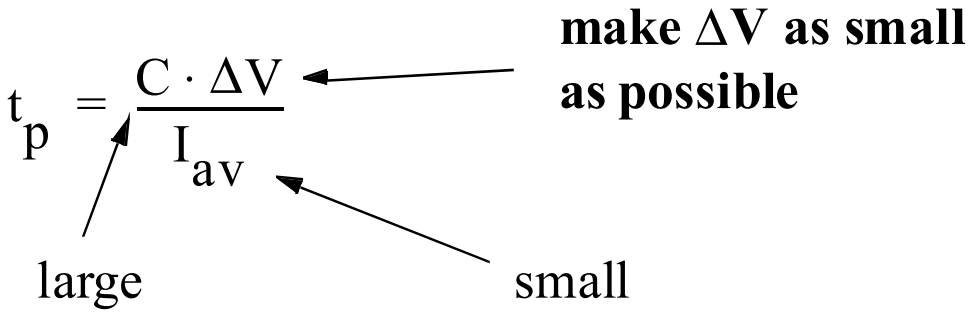
progressive sizing

combination of tree and pass transistor approaches

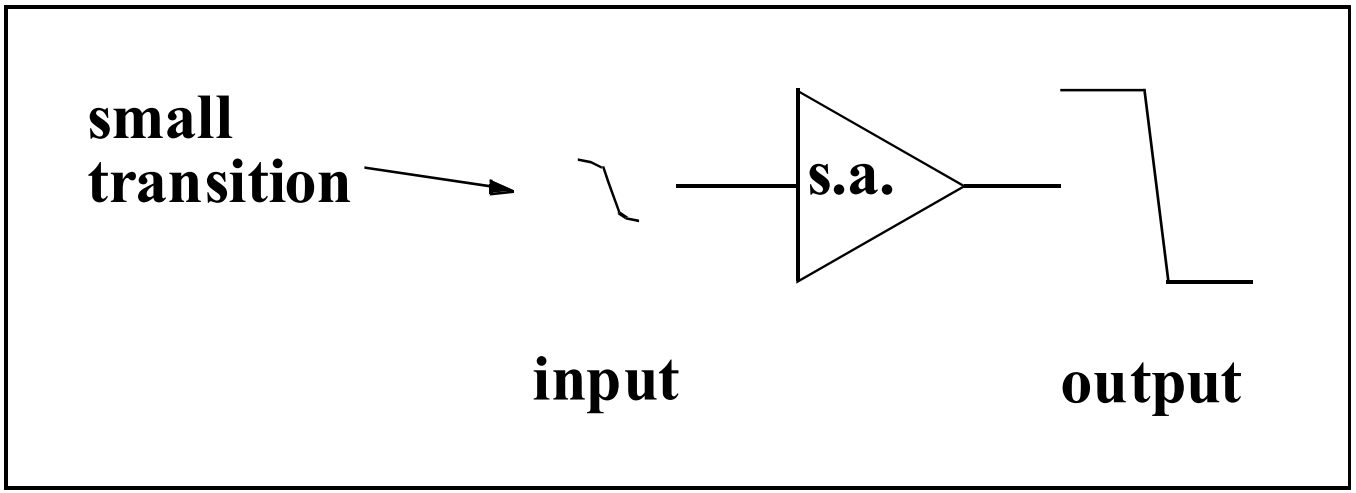
Decoder for circular shift-register



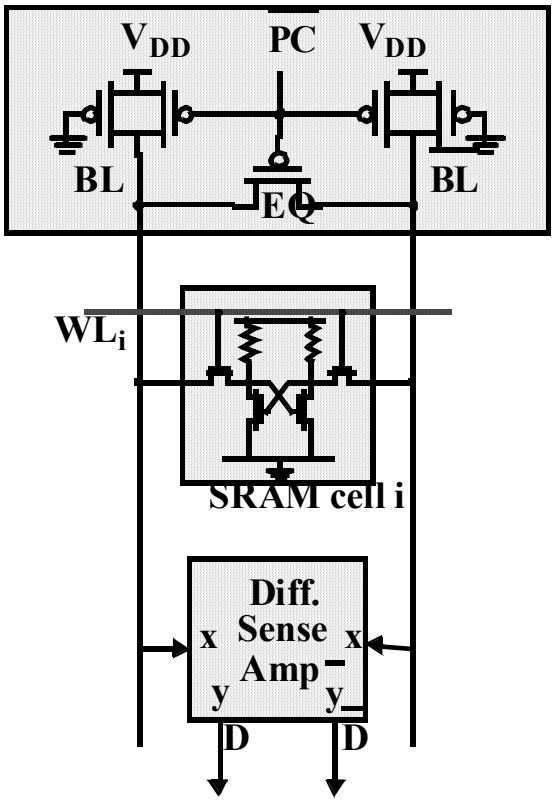
Sense Amplifiers



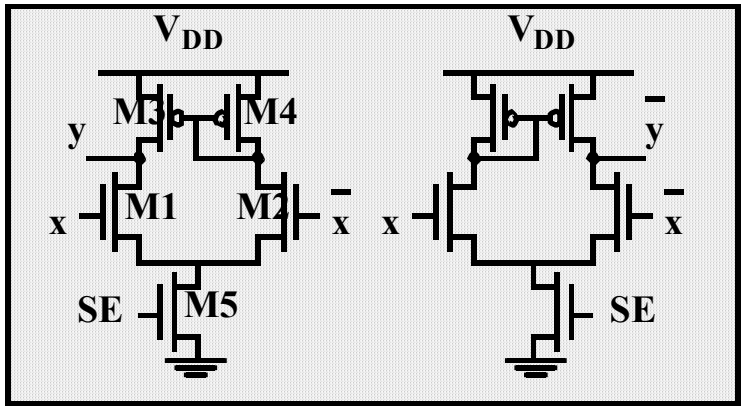
Idea: Use Sense Amplifer



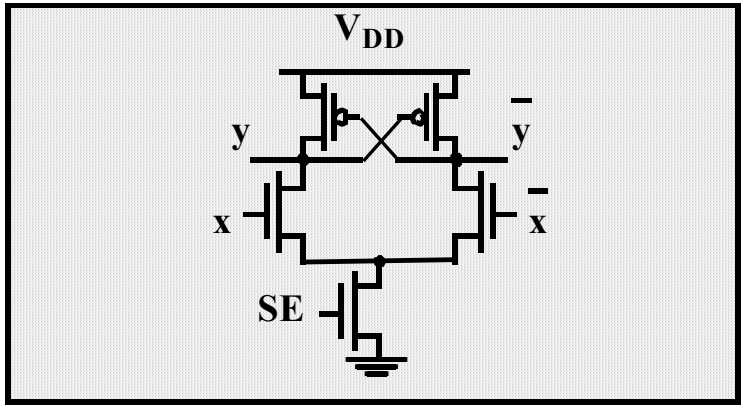
Differential Sensing - SRAM



(a) SRAM sensing scheme.

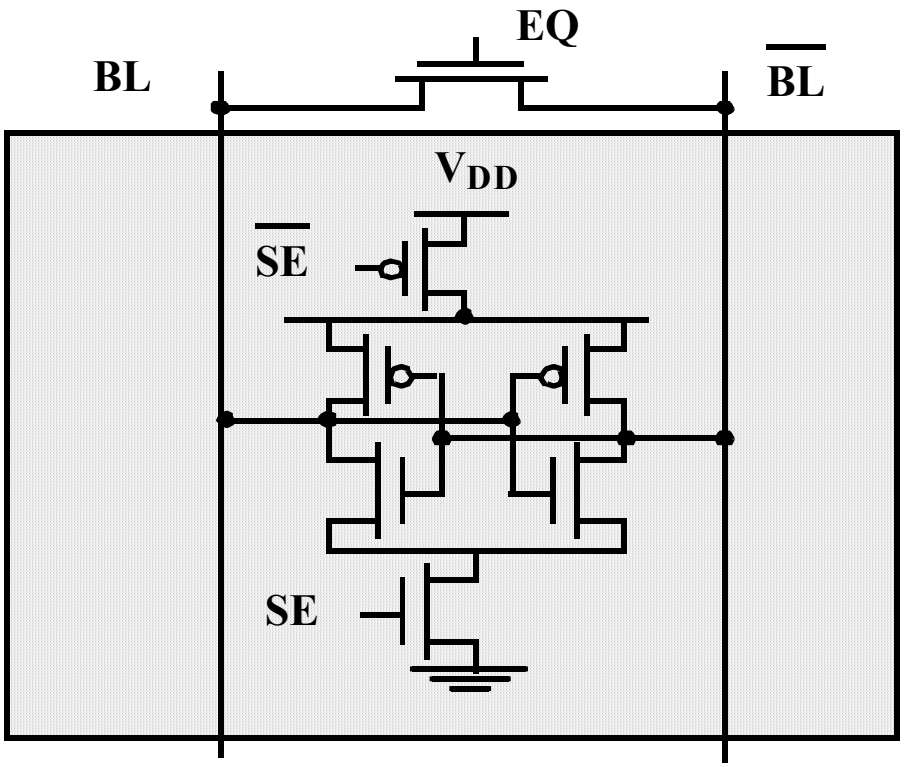


(b) Doubled-ended Current Mirror Amplifier



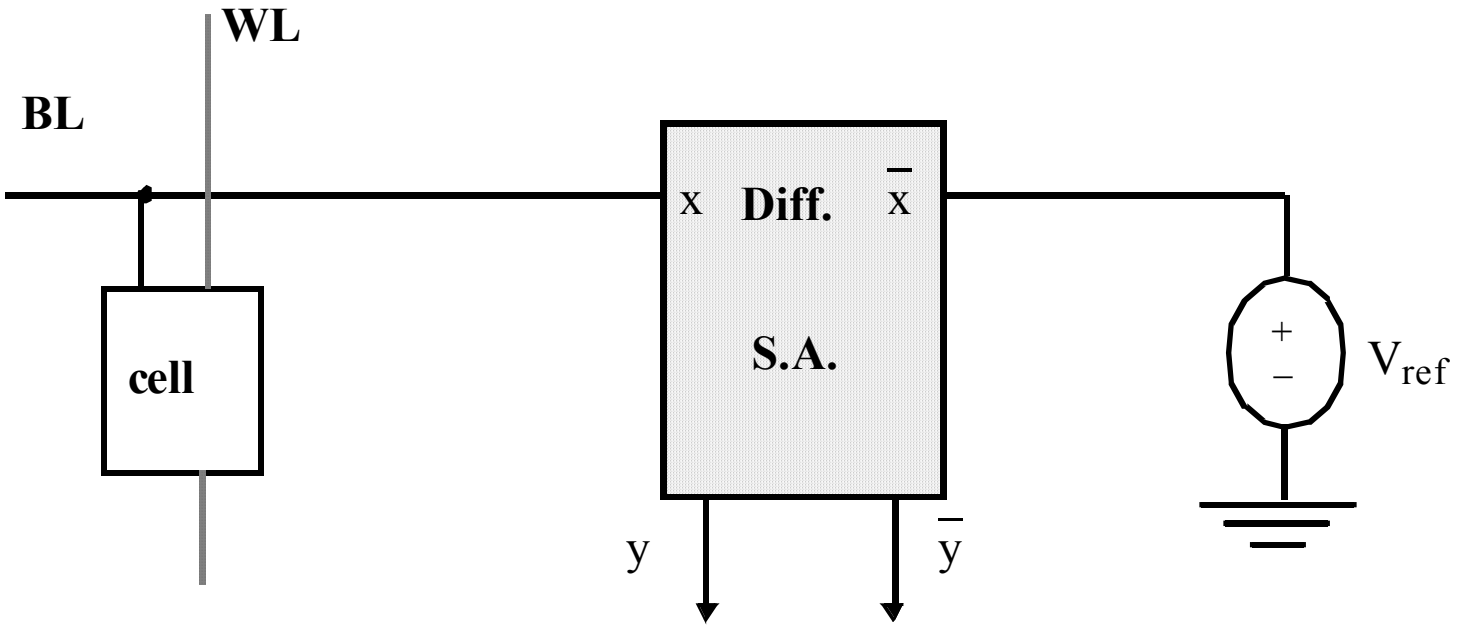
(c) Cross-Coupled Amplifier

Latch-Based Sense Amplifier



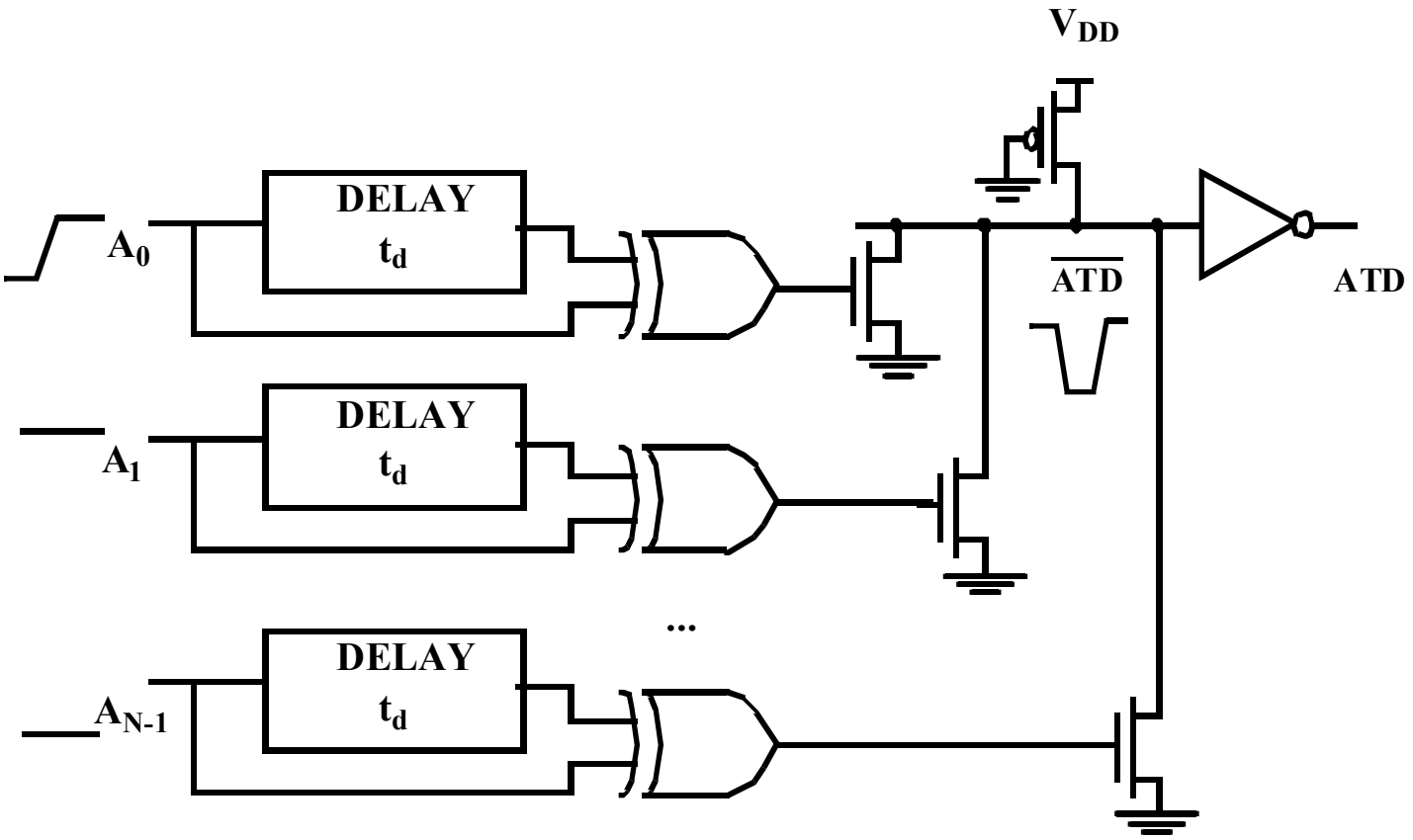
Initialized in its meta-stable point with EQ
Once adequate voltage gap created, sense amp enabled with SE
Positive feedback quickly forces output to a stable operating point.

Single-to-Differential Conversion



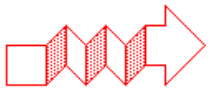
How to make good V_{ref}?

Address Transition Detection



Reliability and Yield

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



Highly Sensitive to Noise (Crosstalk, Supply Noise)

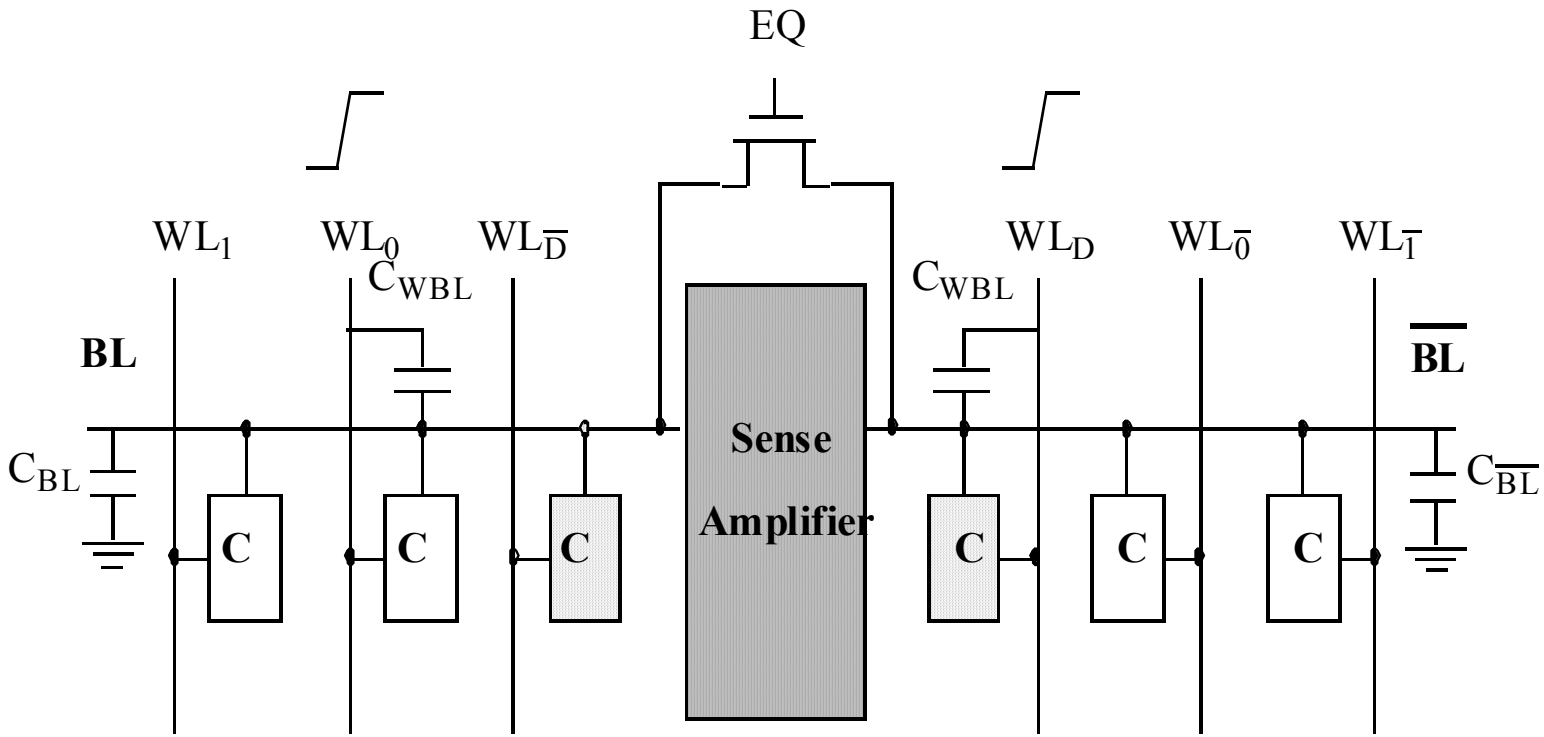
- High Density and Large Die size cause Yield Problems

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

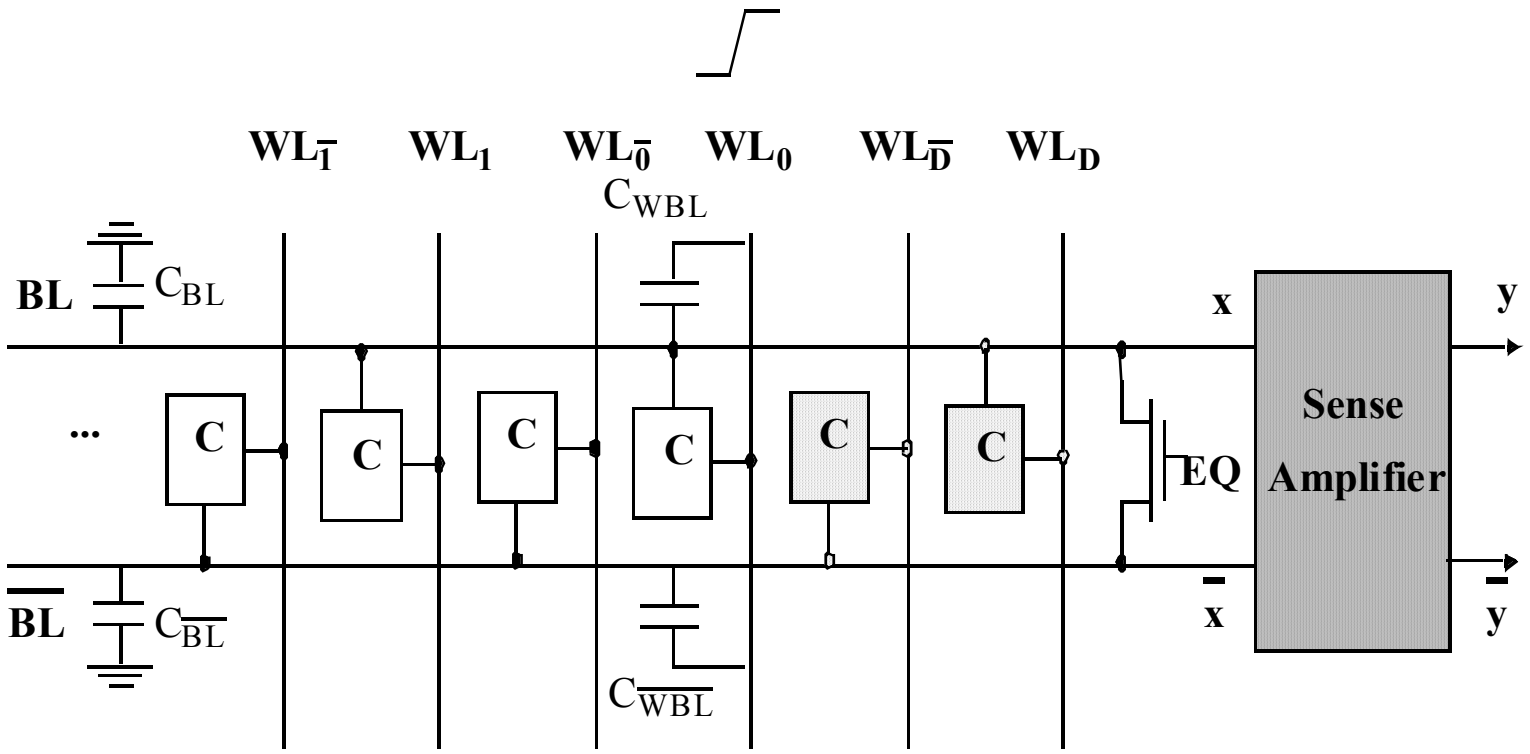
$$Y = \left[\frac{1 - e^{-AD}}{AD}\right]^2$$

Increase Yield using Error Correction and Redundancy

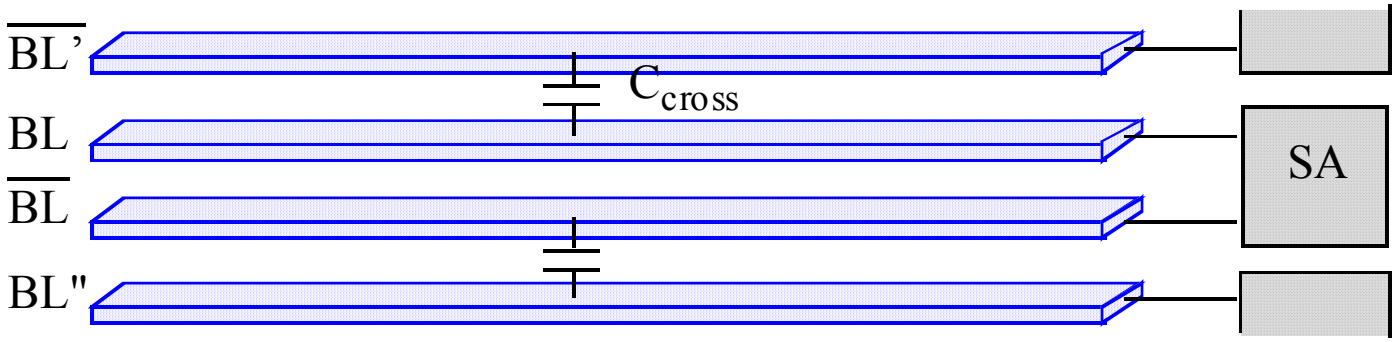
Open Bit-line Architecture —Cross Coupling



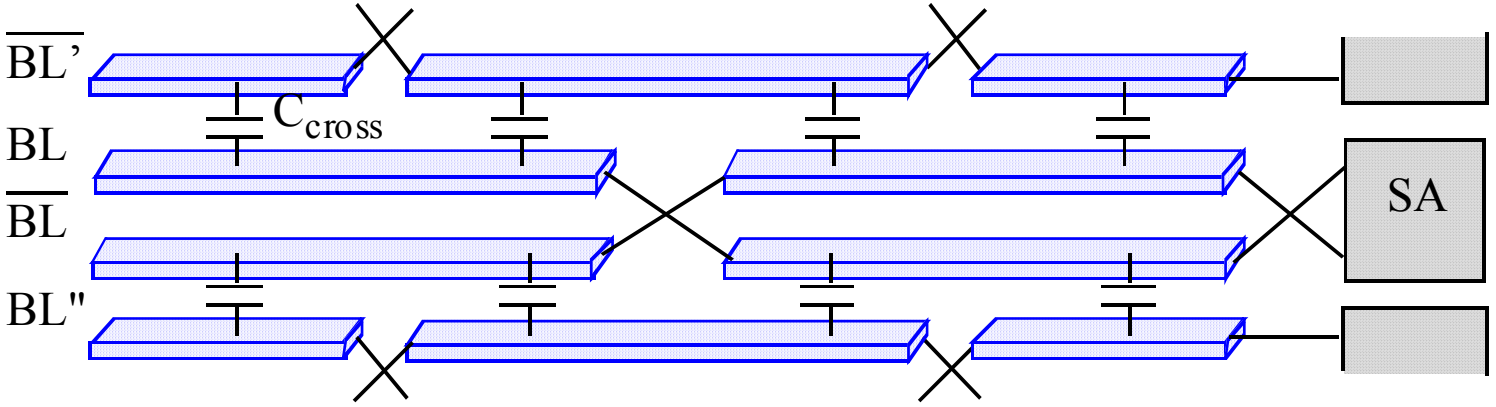
Folded-Bitline Architecture



Transposed-Bitline Architecture

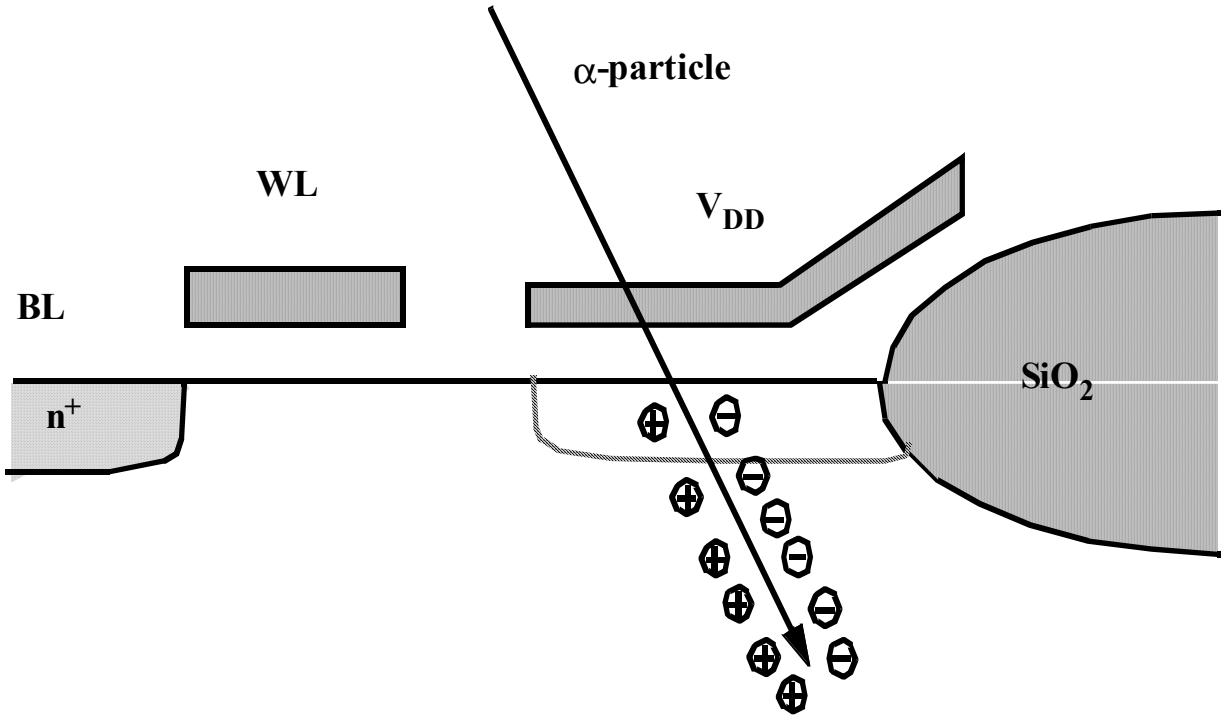


(a) Straightforward bitline routing.



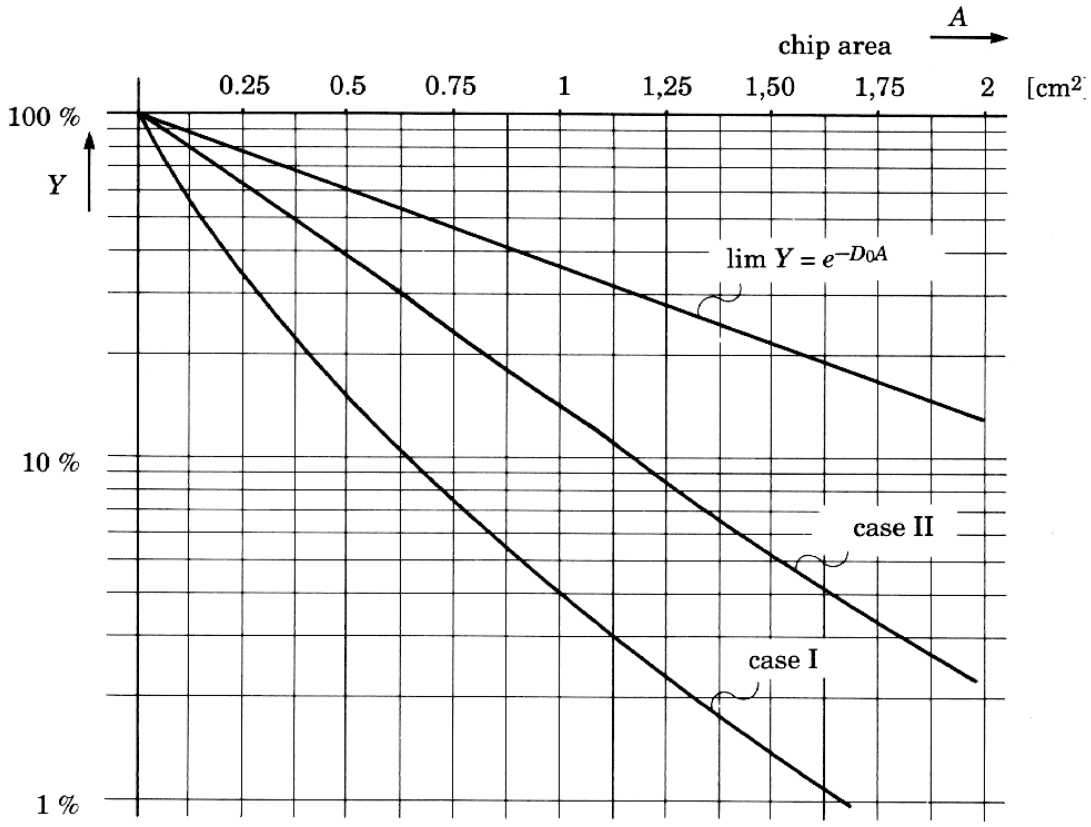
(b) Transposed bitline architecture.

Alpha-particles



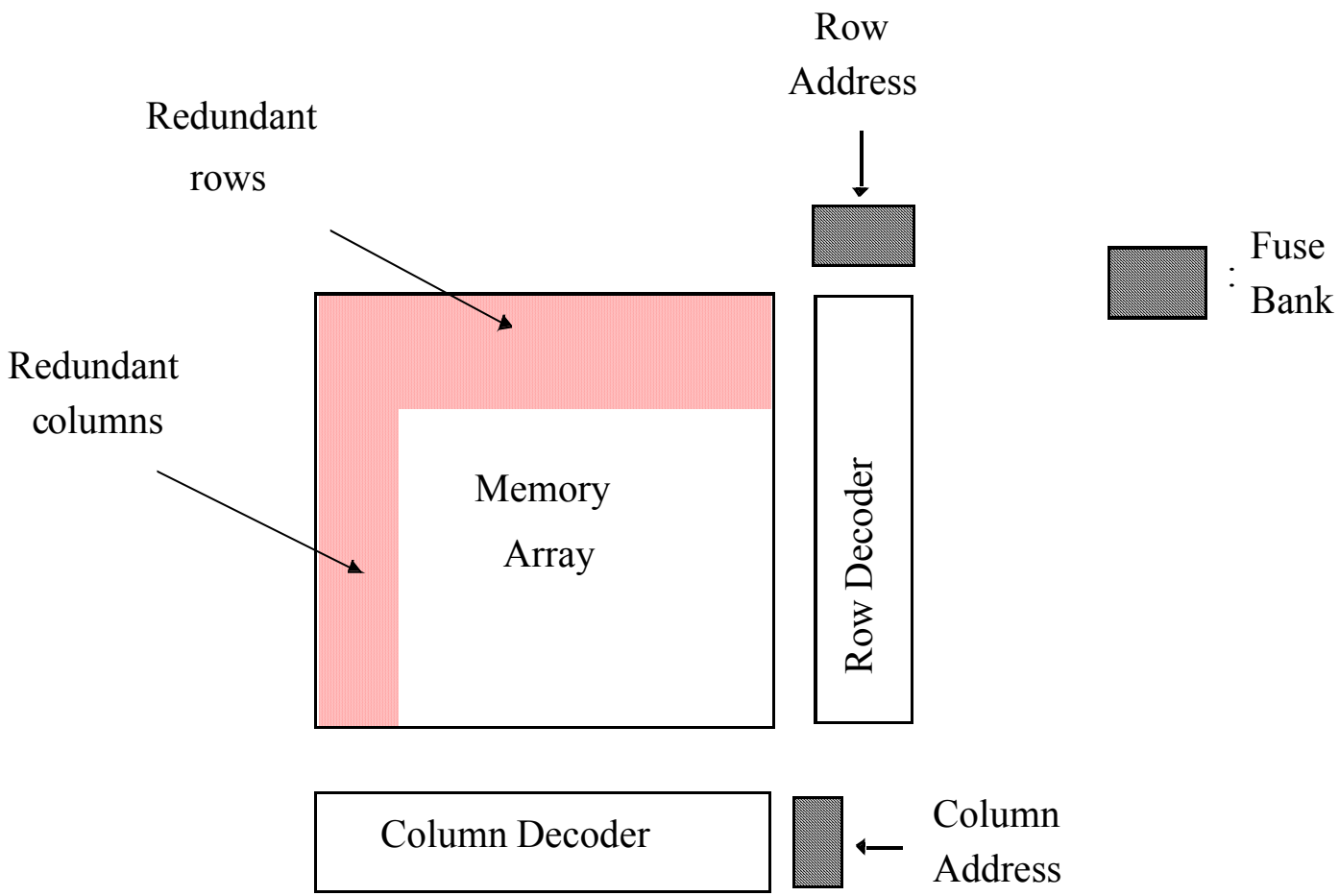
1 particle ~ 1 million carriers

Yield

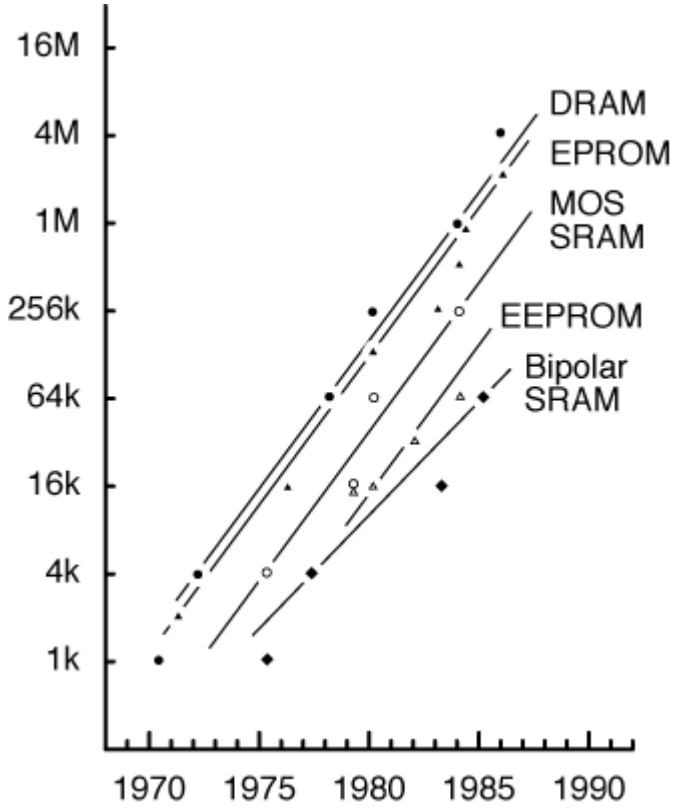


Yield curves at different stages of process maturity (from [Veendrick92])

Redundancy

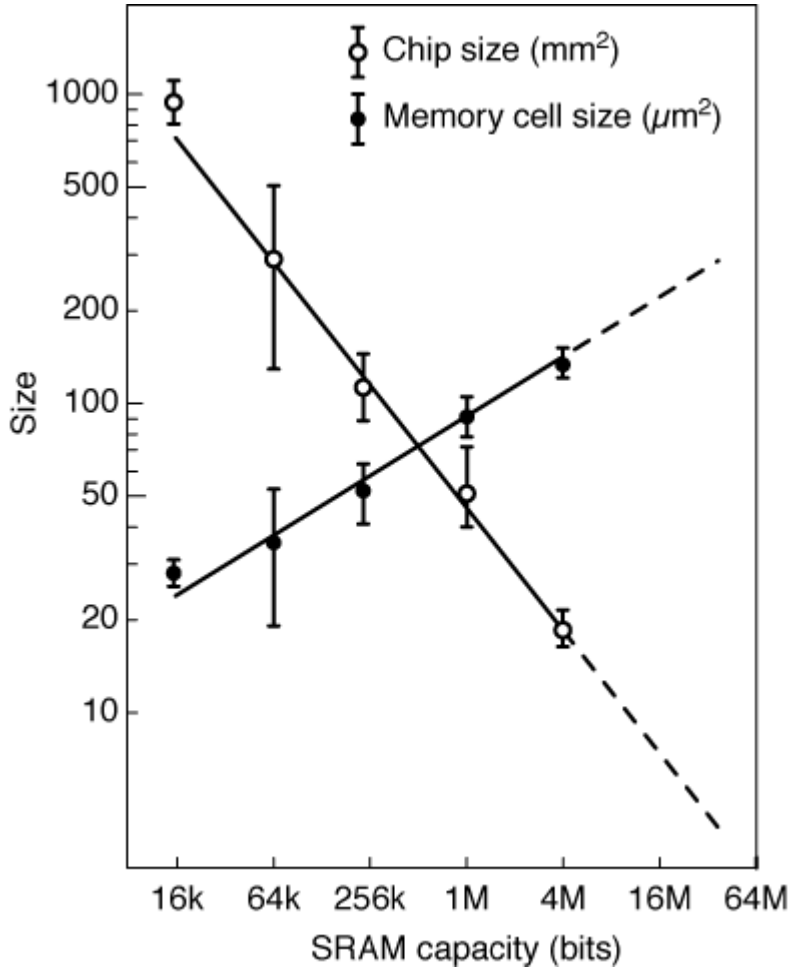

Row
Address
Redundant
rows
Redundant
columns
Memory
Array
Row Decoder
Fuse
Bank
Column Decoder
Column
Address

Semiconductor Memory Trends



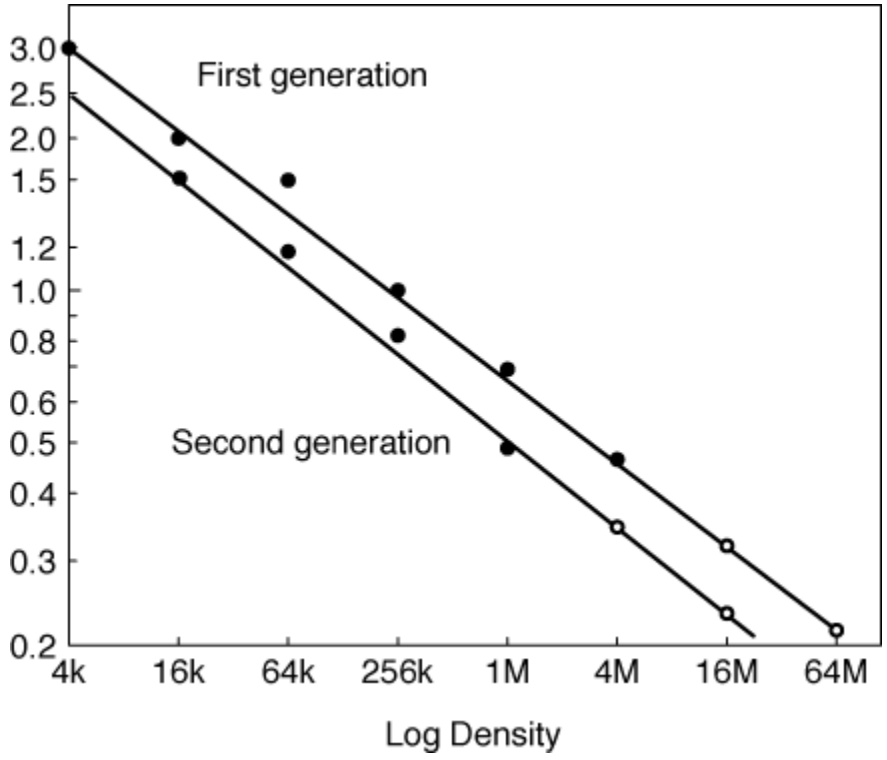
Memory Size as a function of time: x 4 every three years

Semiconductor Memory Trends



Increasing die size
factor 1.5 per generation
Combined with reducing cell size
factor 2.6 per generation

Semiconductor Memory Trends



Technology feature size for different SRAM generations