ESE3700: Circuit-Level Modeling, Design and Optimization for Digital Systems

Lec 18: April 10, 2024

Memory Overview and RAM Core

Today

- Memory
  - Classification
  - Architecture
  - RAM Core
  - Periphery (next week)
  - Serial Access Memories (next week)

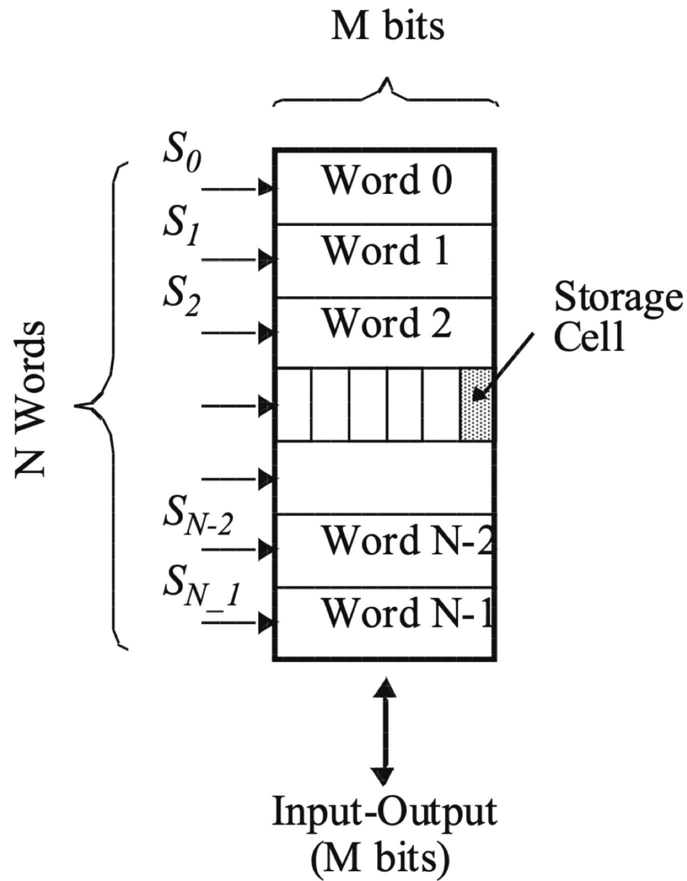
- Project 2 is on this

Memory Overview

Semiconductor Memory Classification

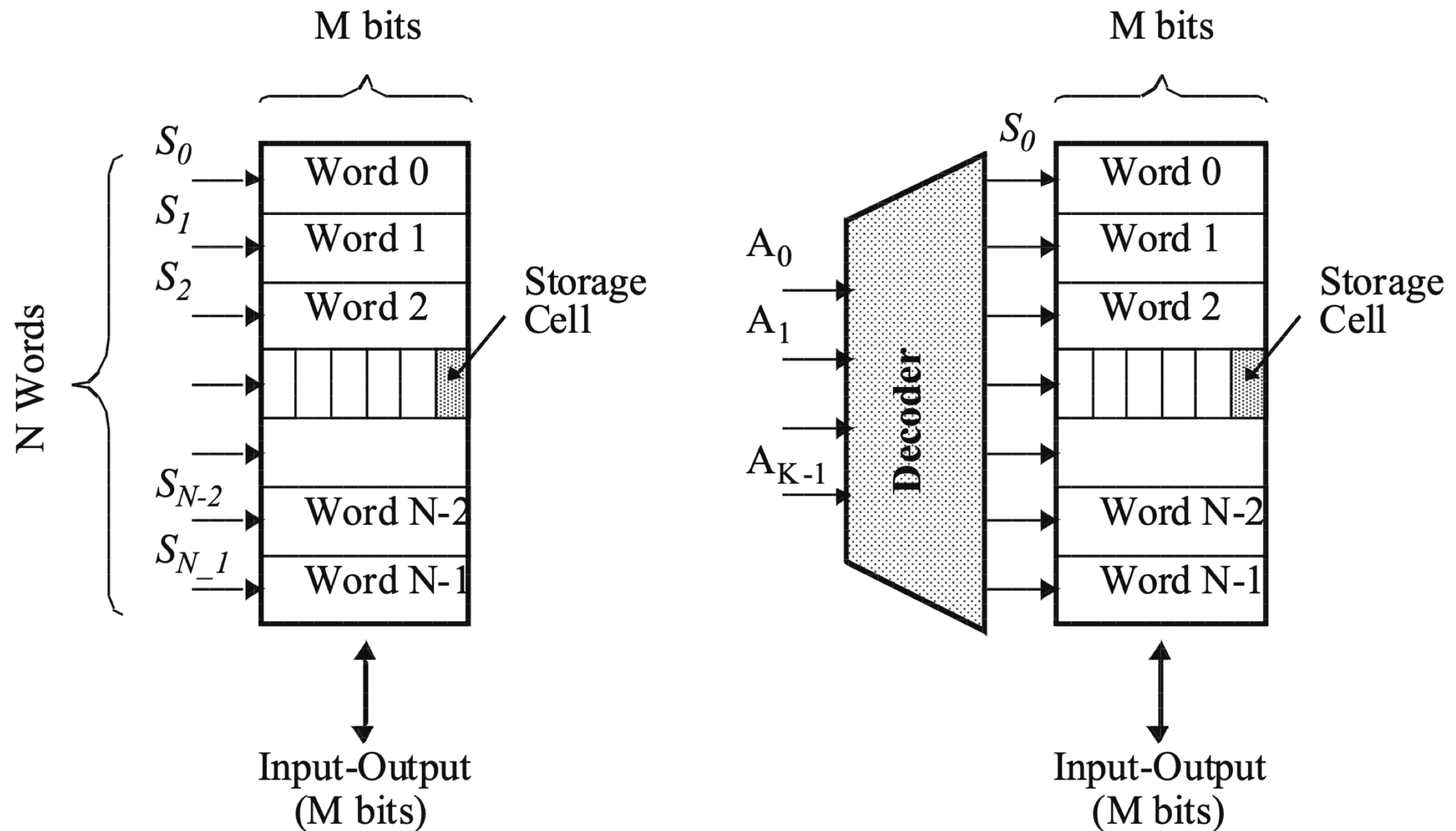
RWM		NVRWM	ROM
Random Access	**Non-Random Access**	EPROM E^2PROM FLASH	Mask-Programmed Programmable (PROM)
SRAM DRAM	FIFO LIFO Shift Register CAM		

Memory Architecture: Core



N words => N select signals
Too many select signals

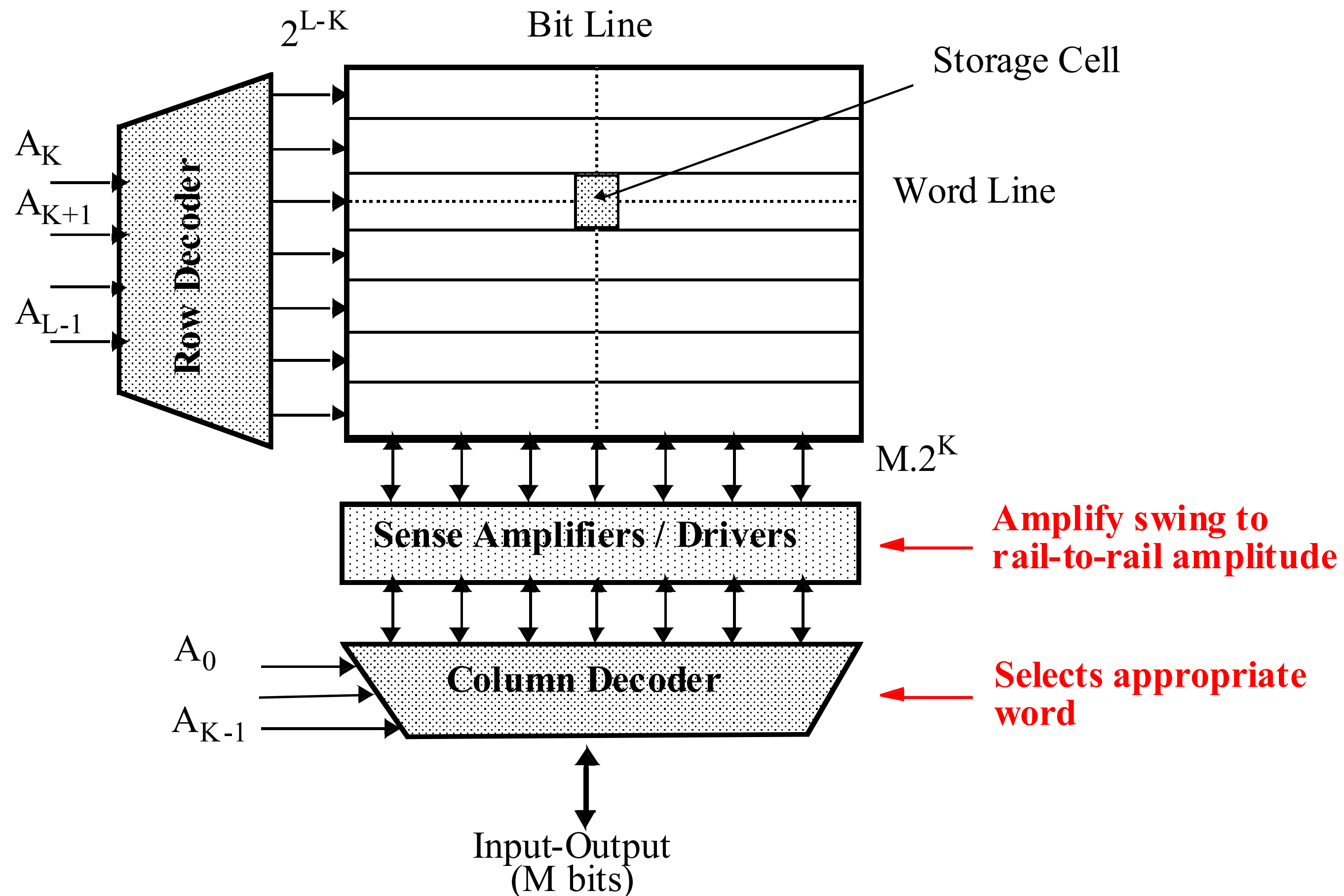
Memory Architecture: Decoders



N words => N select signals
Too many select signals

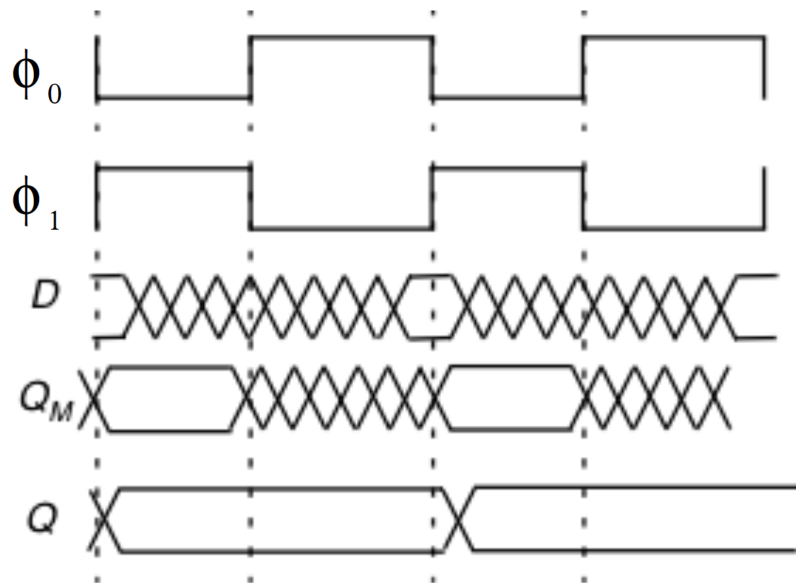
Decoder reduces # of select signals
$K = log_2 N$

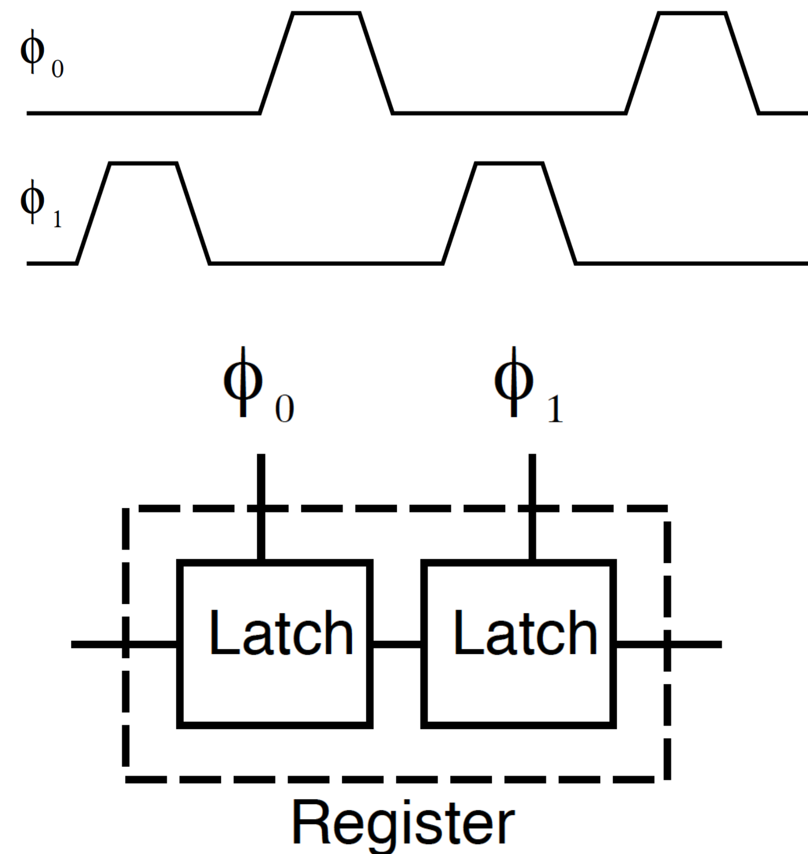
Array-Structured Memory Architecture



Latches/Register – Can Store a State

- Build register from pair of latches
- Control with non-overlapping clocks





ROM Memories

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

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

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

V_{DD}

Pull-up devices

WL[0] 1 0 1 1

GND

WL[1]

WL[2]

GND

WL[3]

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

MOS NOR ROM

V_{DD}

Pull-up devices

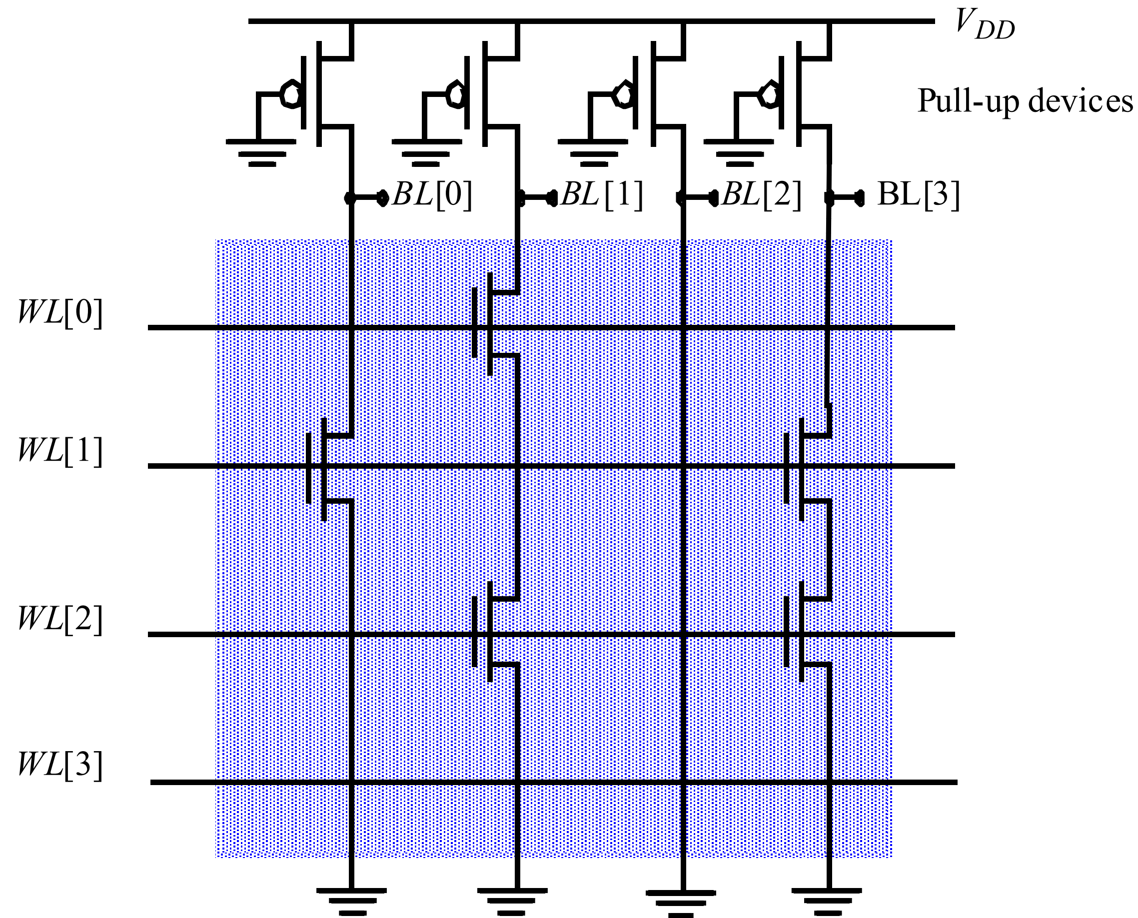
WL[0]	1	0	1	1
WL[1]	0	1	1	0
WL[2]	1	0	1	0
WL[3]	1	1	1	1

GND

GND

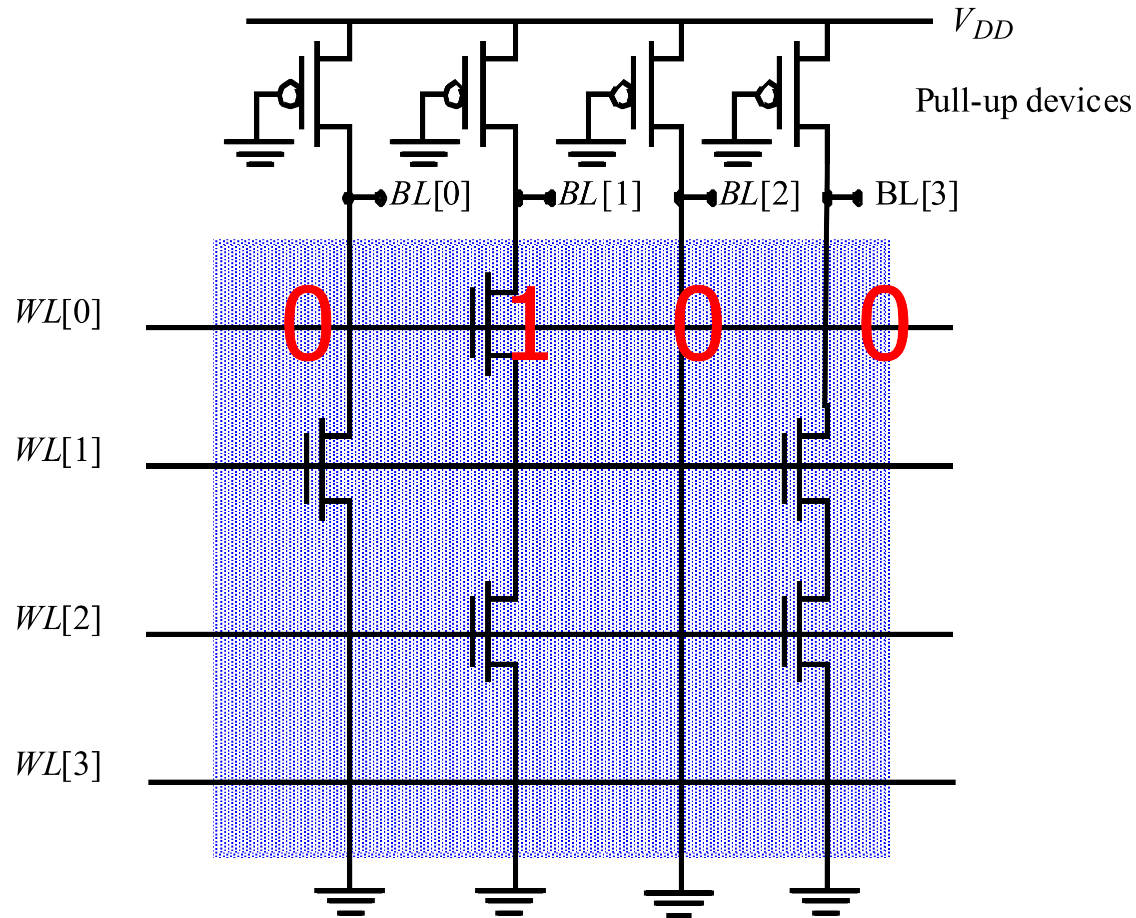
BL[0] *BL*[1] *BL*[2] *BL*[3]

MOS NAND ROM



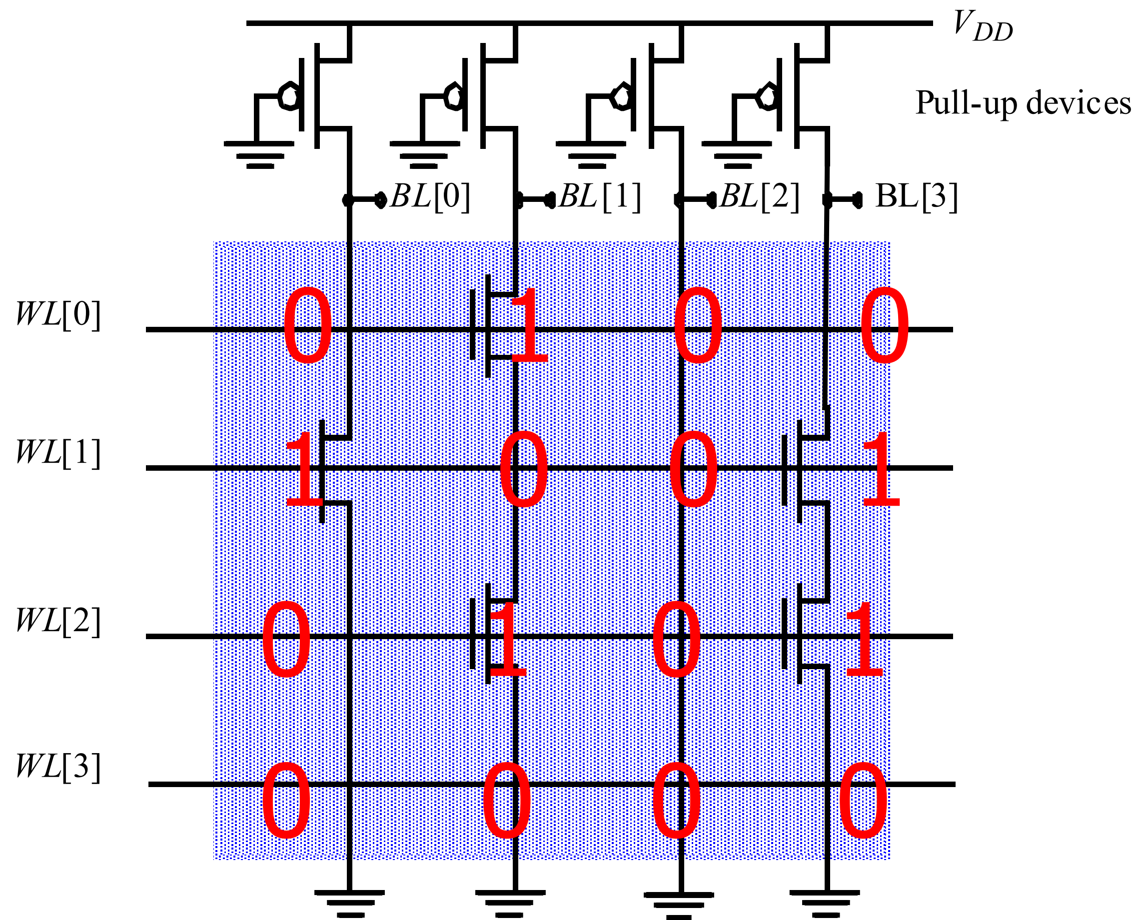
All word lines high by default with exception of selected row

MOS NAND ROM



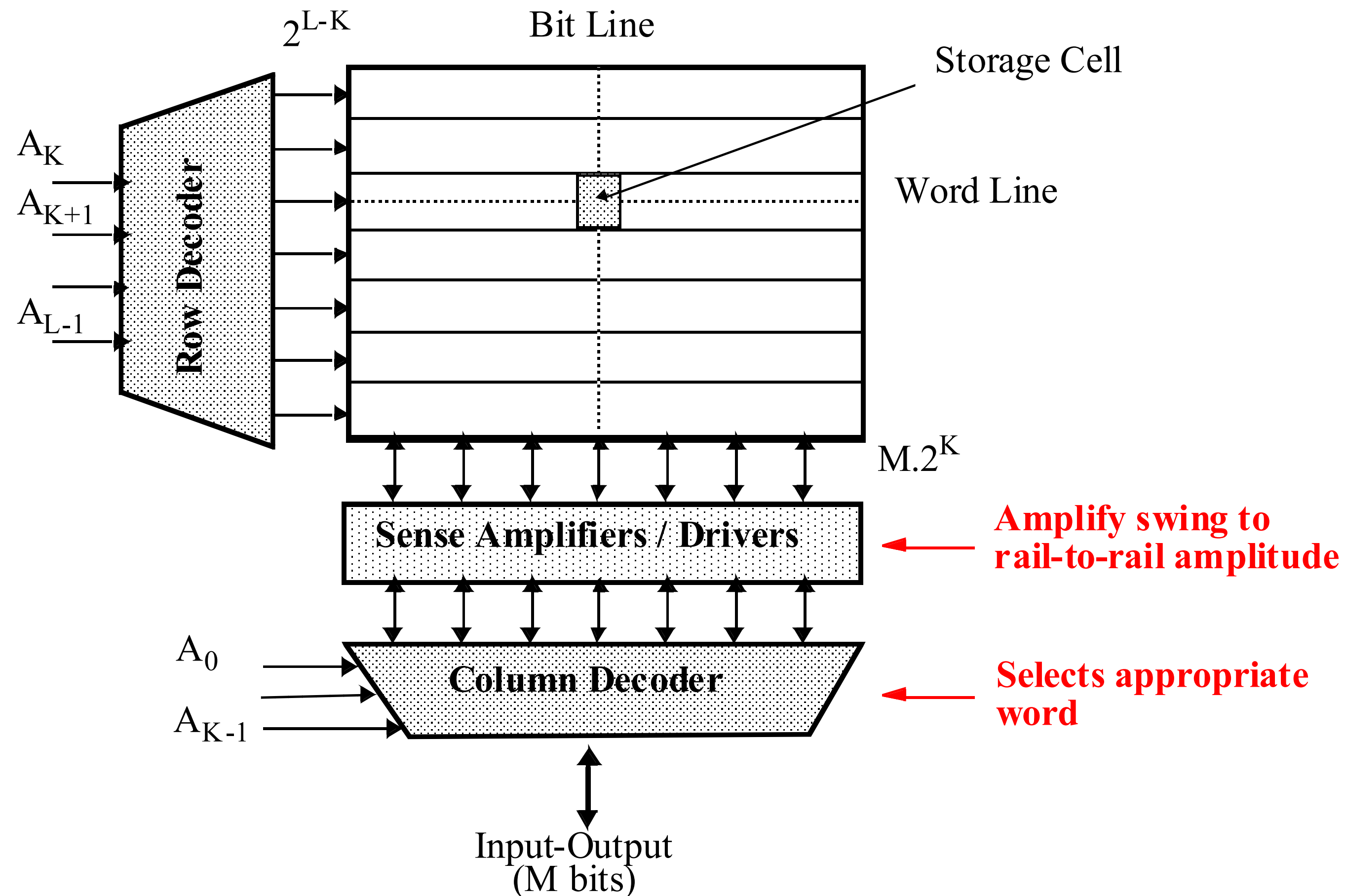
All word lines high by default with exception of selected row

MOS NAND ROM



All word lines high by default with exception of selected row

Array-Structured Memory Architecture



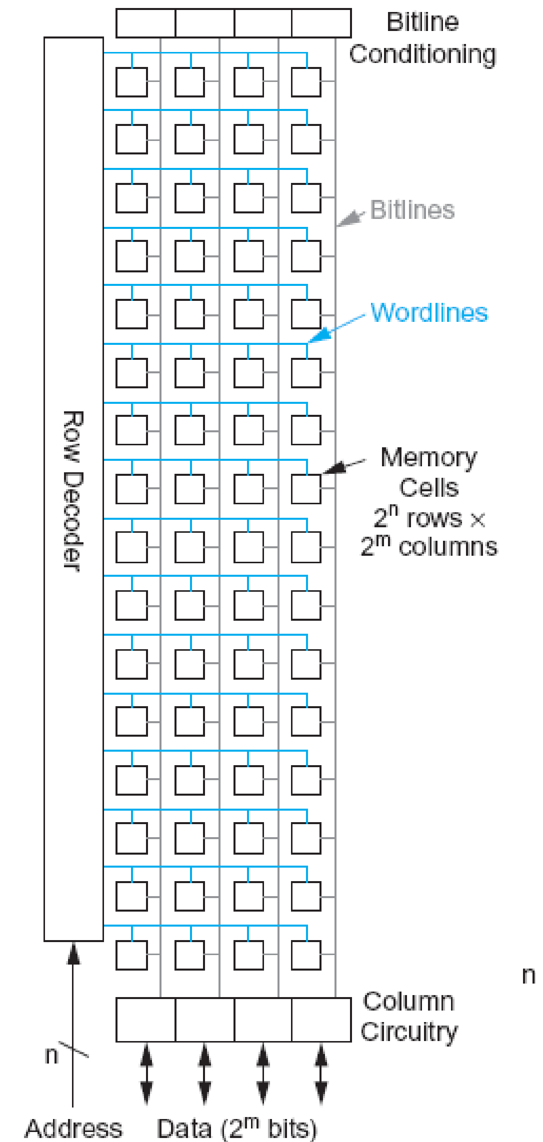
Memory Periphery

Periphery

- Decoders
- Column Circuitry
  - Bit-line Conditioning
  - Sense Amplifiers
  - Input/Output Buffers
- Control/Timing Circuitry

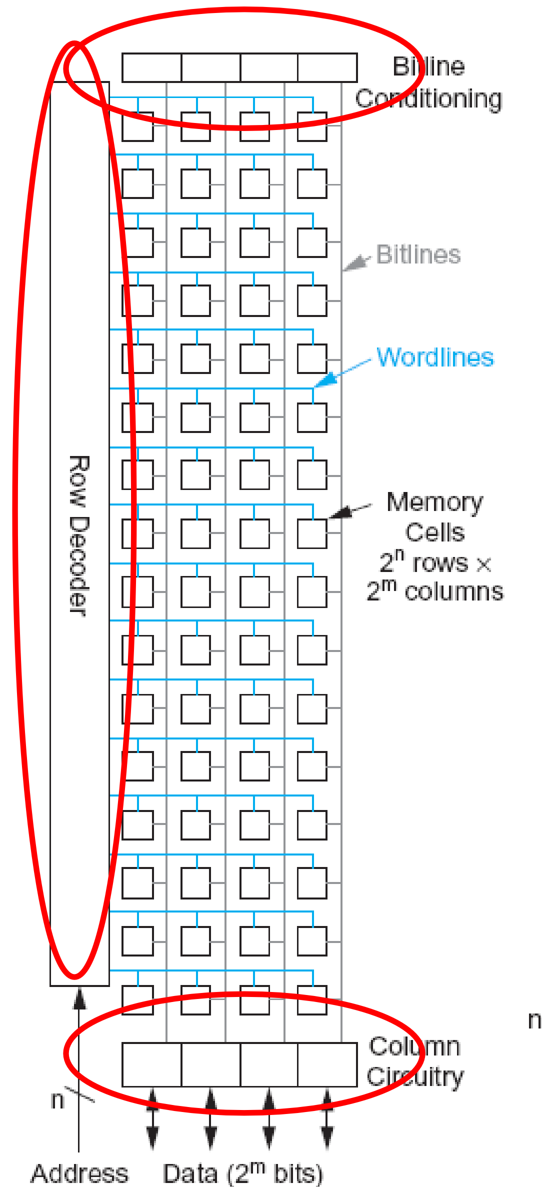
Array Architecture

- 2^n *words* of 2^m *bits* each
- Good regularity – easy to design
- Very high density if good cells are used



Array Architecture

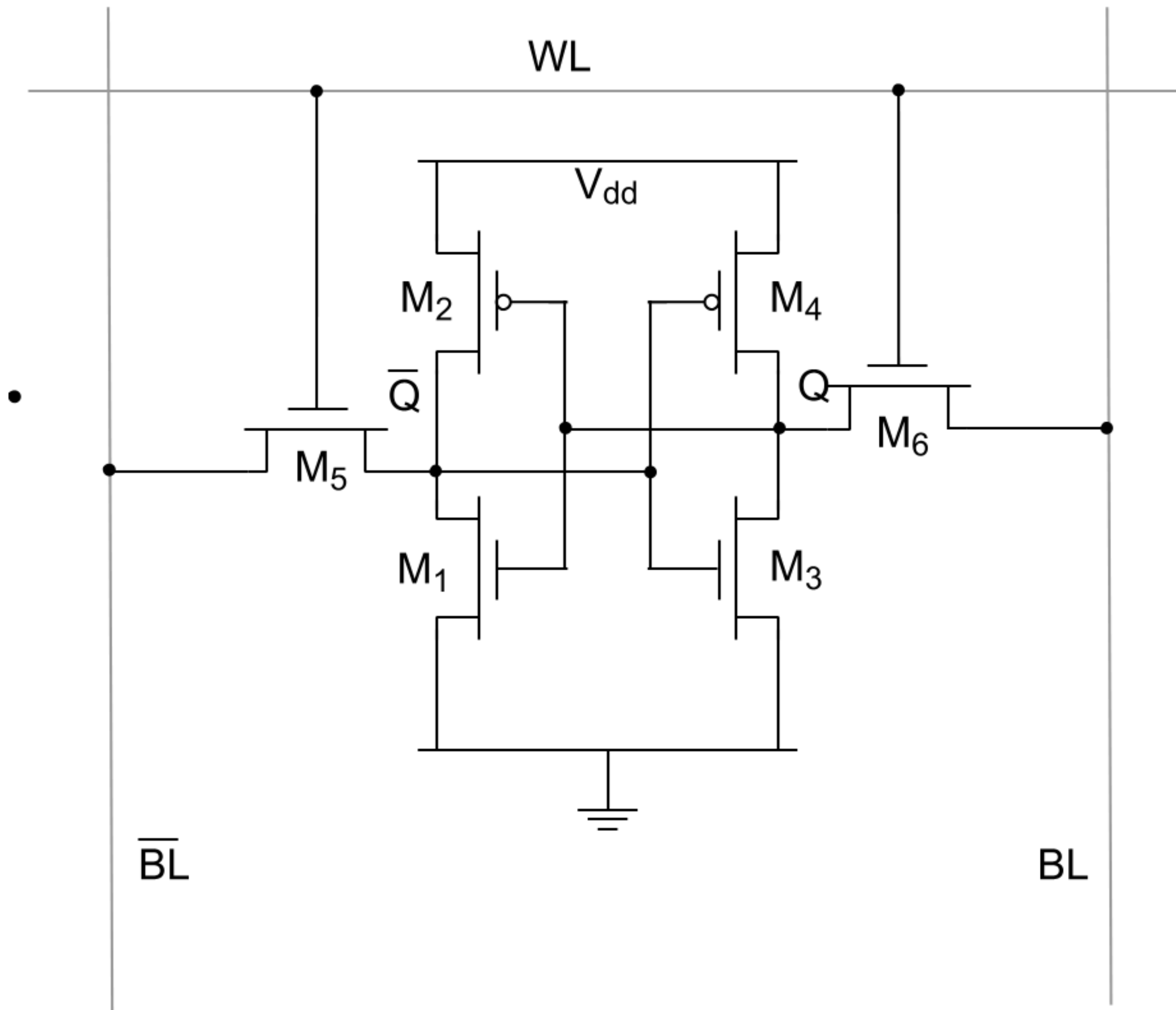
- 2^n *words* of 2^m *bits* each
- Good regularity – easy to design
- Very high density if good cells are used



6T SRAM Cell

- Cell size accounts for most of array size
  - Reduce cell size at expense of complexity
- 6T SRAM Cell
  - Used in most commercial chips
  - Data stored in cross-coupled inverters
- Read:
  - Precharge BL, BL'
  - Raise WL
- Write:
  - Drive data onto BL, BL'
  - Raise WL

6-transistor CMOS SRAM Cell



6-transistor CMOS SRAM Cell

WL

V_{dd}

M_2 M_4

$\overline{Q}$ Q

M_5 M_6

M_1 M_3

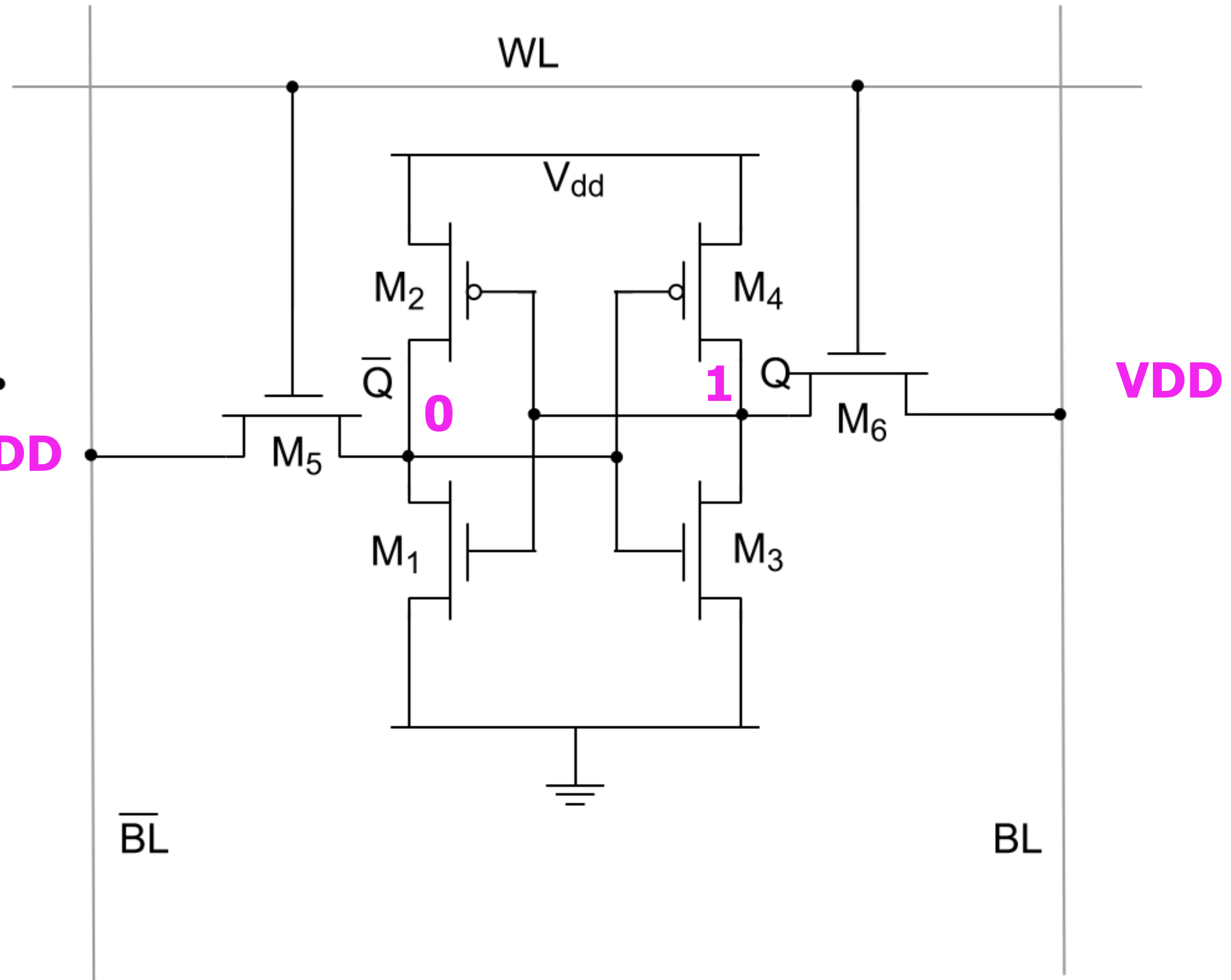
$\overline{BL}$ BL

6-transistor CMOS SRAM Cell (preclass 1)

Assume 1 is stored (Q=1)

Read Operation:

- First bitlines get precharged high (Vdd)
- Then wordline goes high (Vdd)
    - Precharge disconnected first!



CMOS SRAM Analysis (Read) (preclass 1)

VDD

WL

V_{DD}

$\overline{BL}$

M4

BL

M6

0

ΔV

1

M5

M1

V_{DD}

VDD

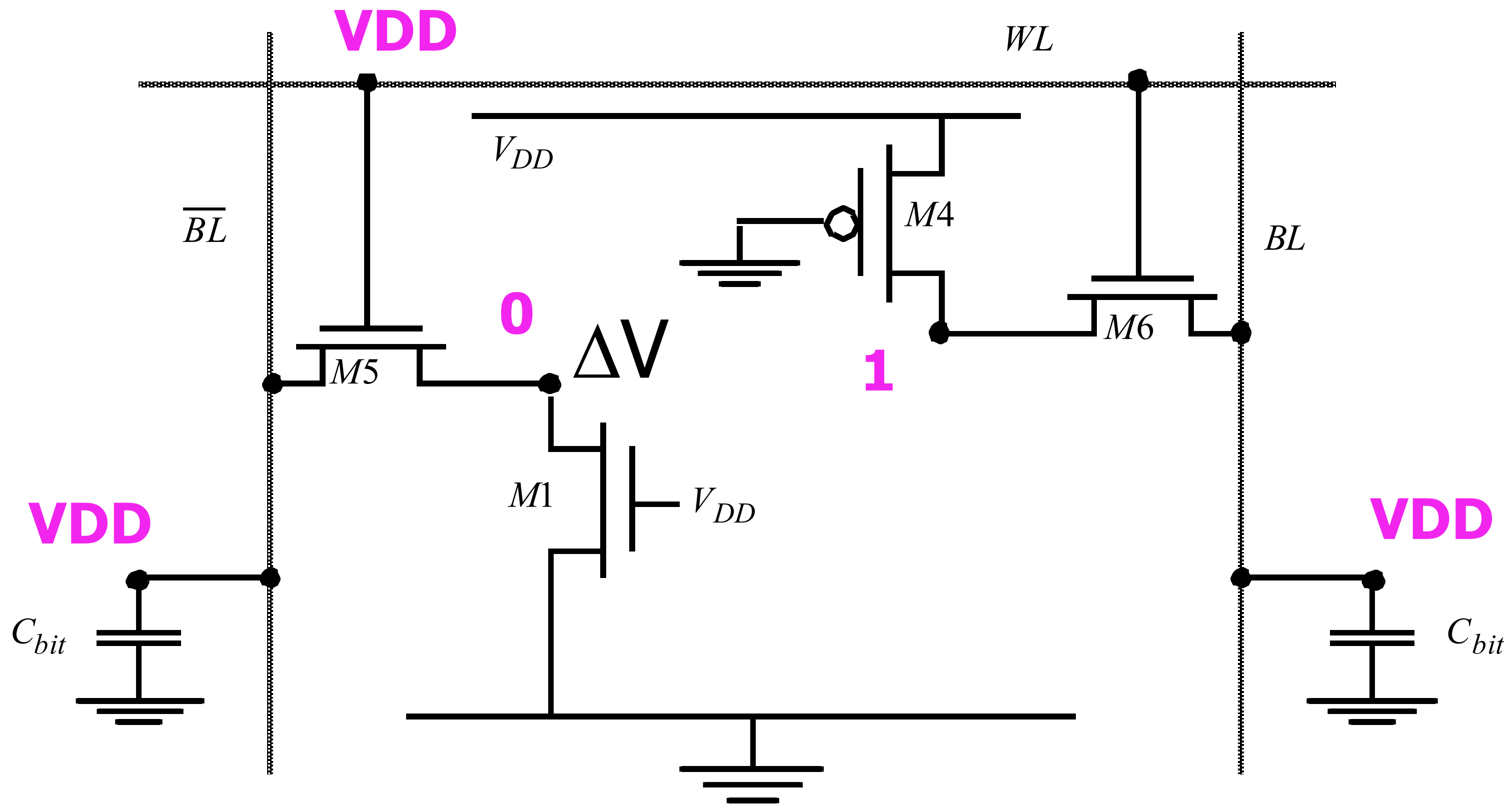
VDD

C_{bit}

C_{bit}

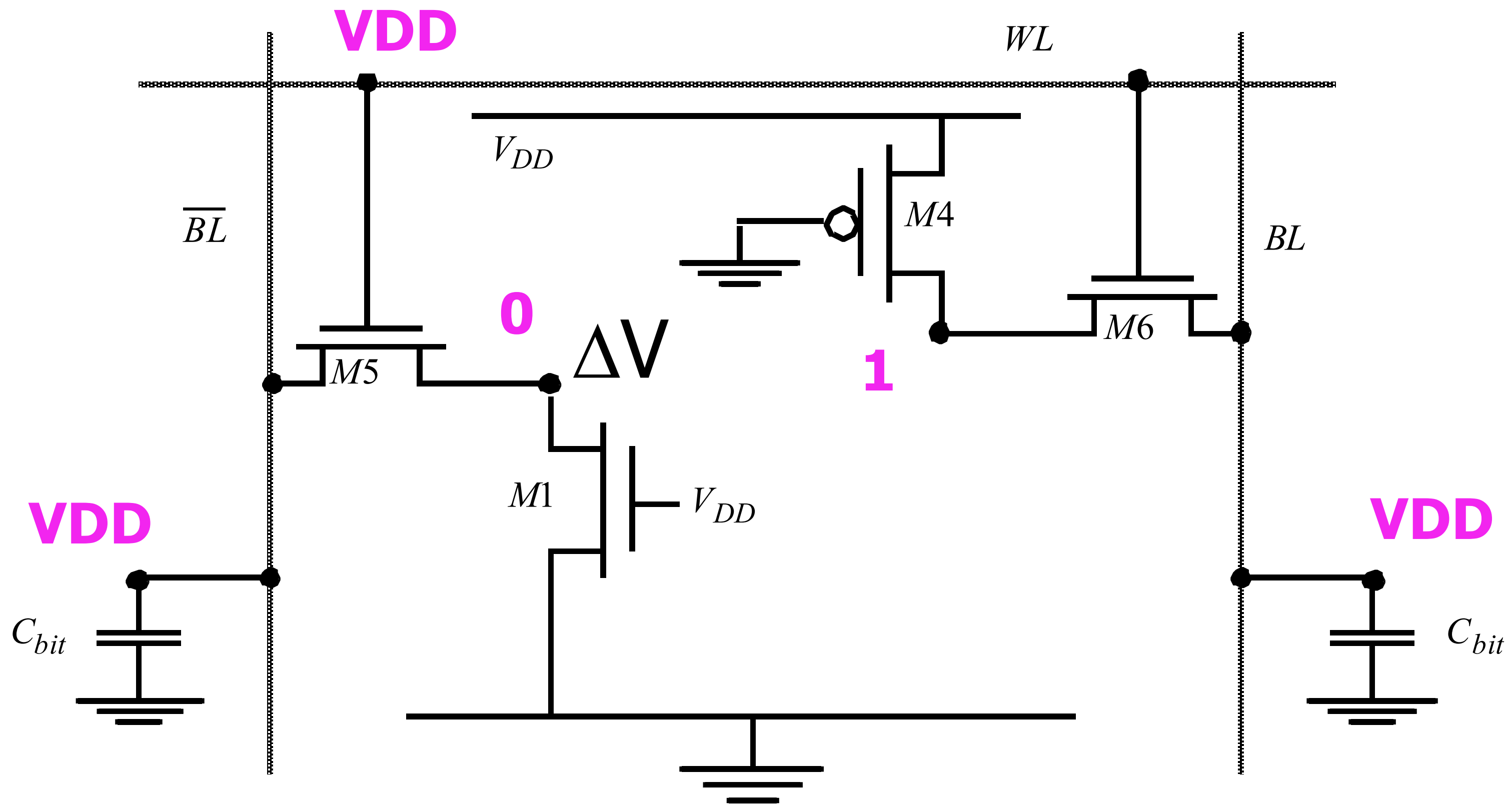
a) Which two transistors are discharging \BL to Gnd?
b) What regions of operation are the transistors in?
c) Write the KCL equation for the two transistors

CMOS SRAM Analysis (Read)



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

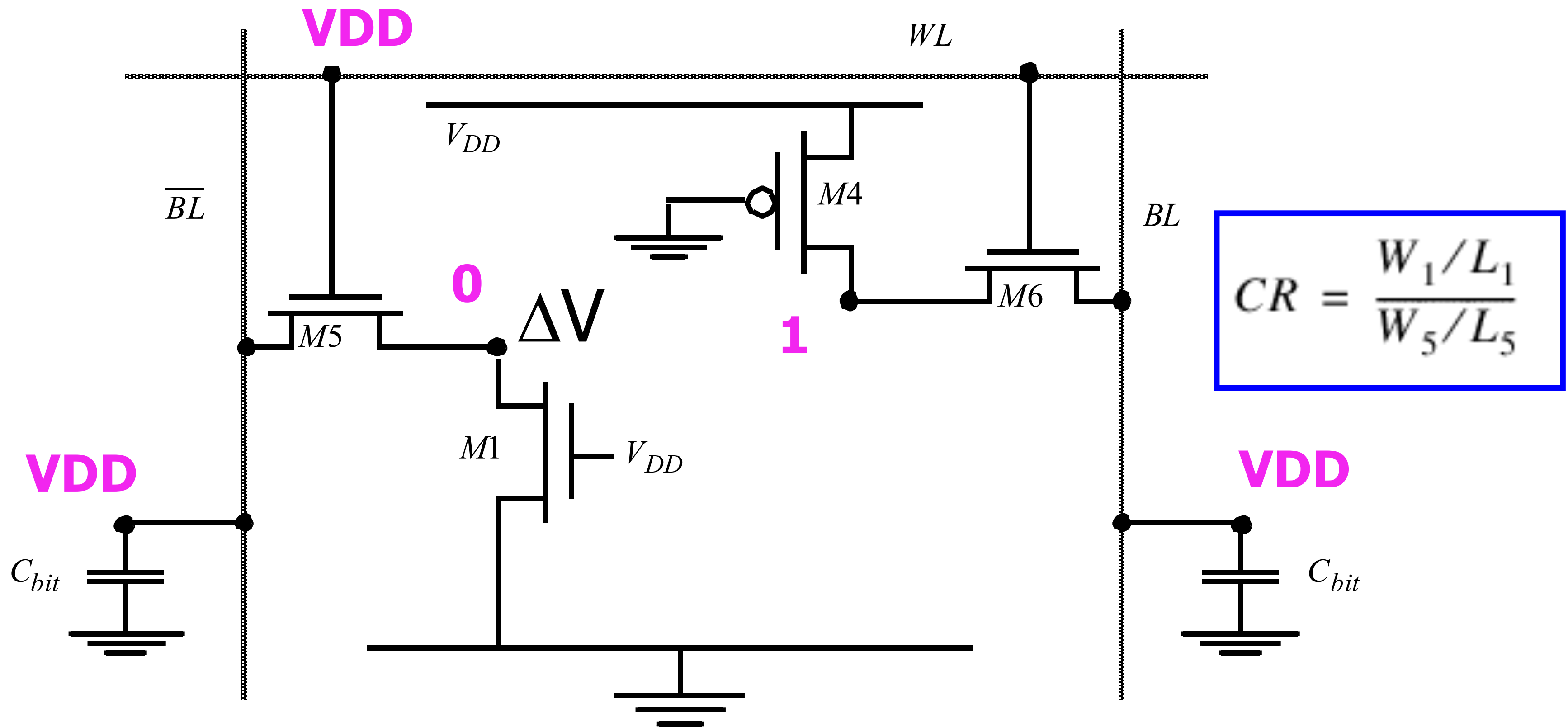
CMOS SRAM Analysis (Read)



$$\frac{k_{n,M1}}{k_{n,M5}} = \frac{\left(\frac{W}{L}\right)_1}{\left(\frac{W}{L}\right)_5} = \frac{\left(V_{DD} - \Delta V - V_{Tn}\right)^2}{(V_{DD} - V_{Tn})\Delta V - \frac{\Delta V^2}{2}}$$

Constraint on ΔV?

CMOS SRAM Analysis (Read)

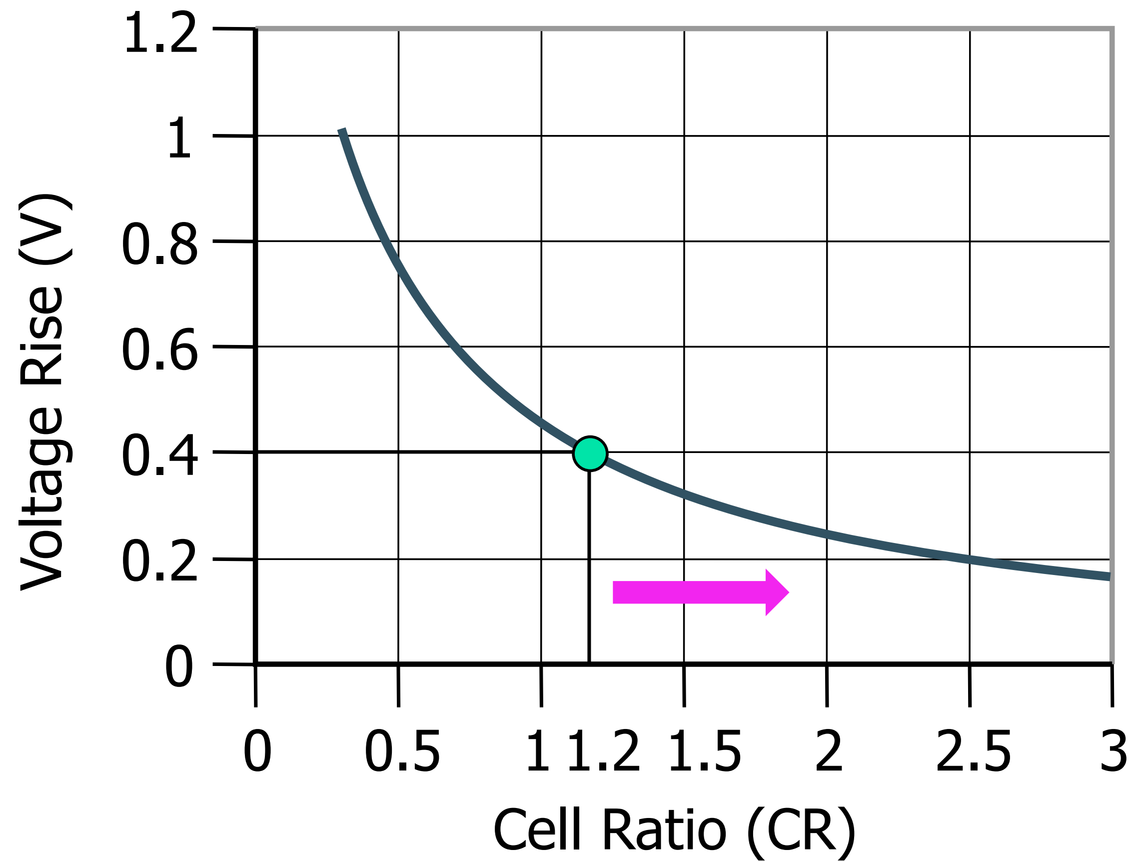


$$\frac{k_{n,M1}}{k_{n,M5}} = \frac{\left(\frac{W}{L}\right)_1}{\left(\frac{W}{L}\right)_5} = \frac{\left(V_{DD} - \Delta V - V_{Tn}\right)^2}{(V_{DD} - V_{Tn})\Delta V - \frac{\Delta V^2}{2}}$$

$\Delta V = V_{Tn}$ →

$$\frac{\left(\frac{W}{L}\right)_1}{\left(\frac{W}{L}\right)_5} = \frac{\left(V_{DD} - 2V_{Tn}\right)^2}{(V_{DD} - 1.5V_{Tn})V_{Tn}}$$

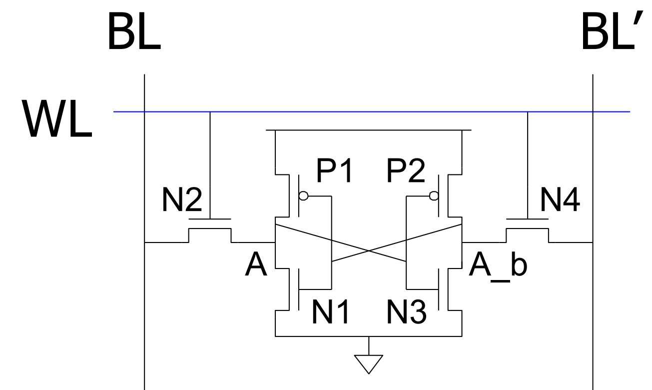
CMOS SRAM Analysis (Read)

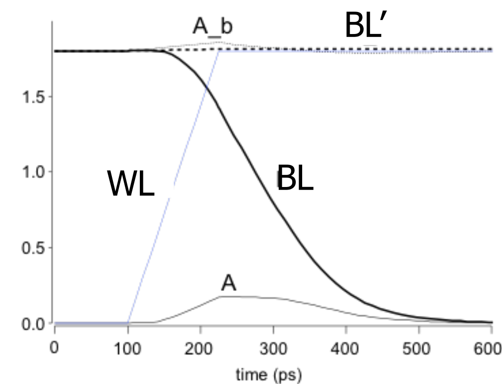


$$CR = \frac{W_1/L_1}{W_5/L_5}$$

SRAM Read

- Precharge both bitlines high
- Then turn on wordline, WL
- One of the two bitlines will be pulled down by the cell
- Ex: A = 0, A_b = 1
  - BL discharges, BL' stays high
  - But A bumps up slightly
- *Read stability*
  - A must not flip
  - N1 > N2

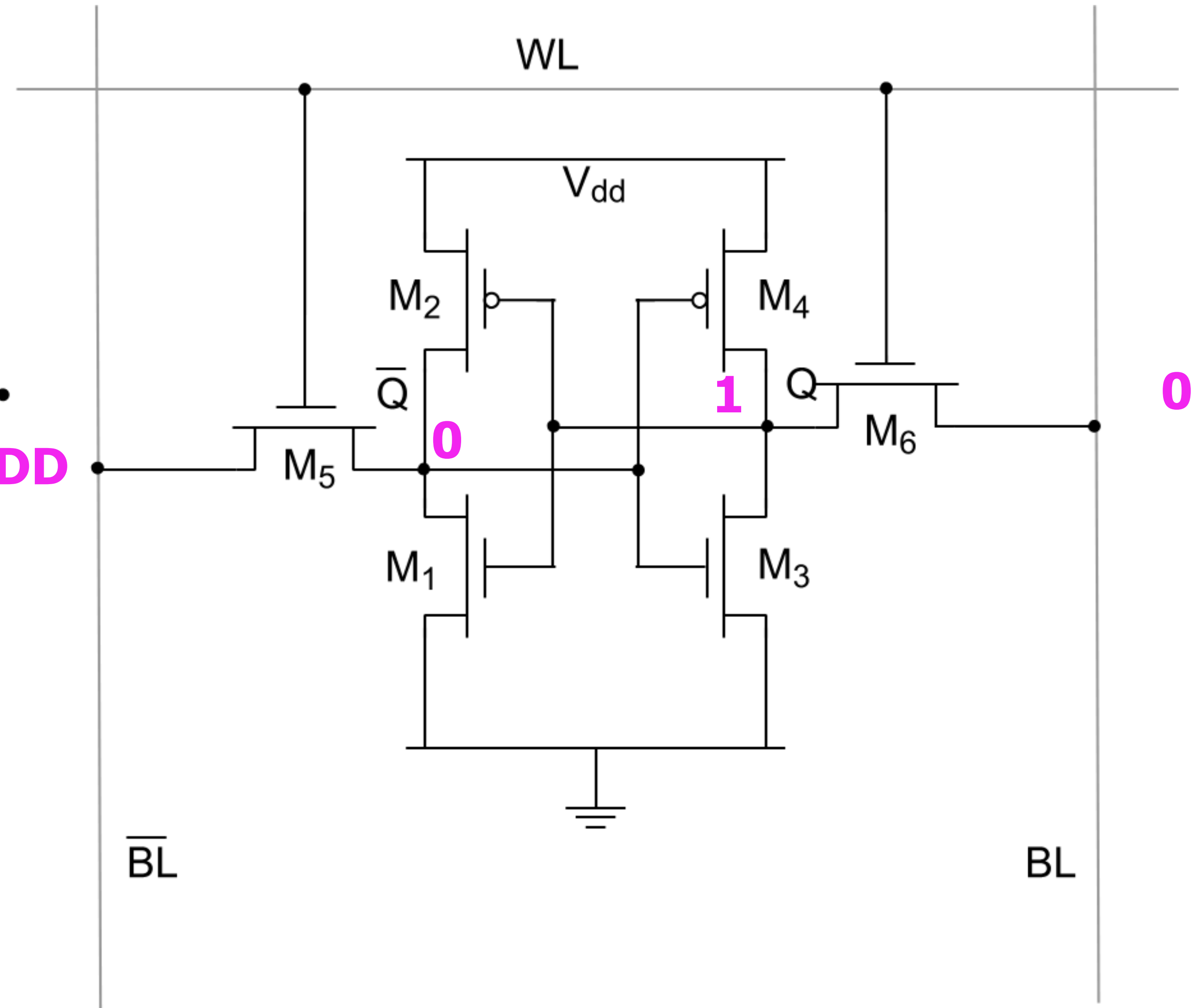




6-transistor CMOS SRAM Cell (preclass 2)

Assume 1 is stored (Q=1)

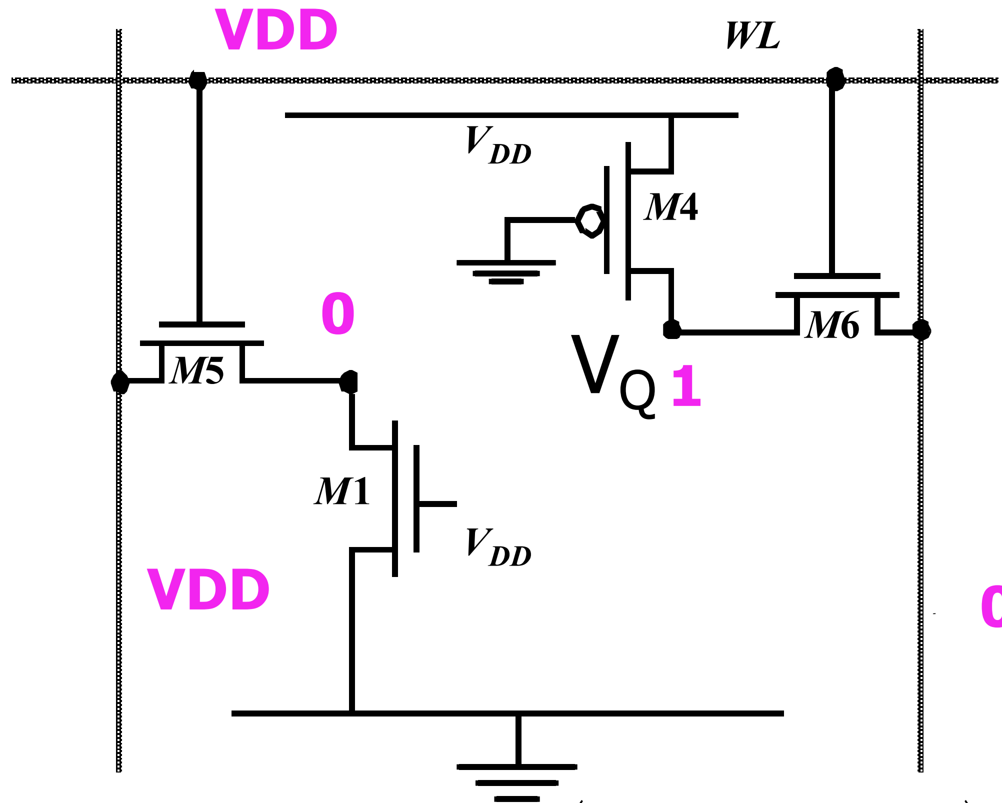
Write Operation:
- Want to write a 0
- First drive bitlines with input data
- Then wordline goes high (Vdd)
  - Still driving bitlines



CMOS SRAM Analysis (Write) (preclass 2)

a) Which transistor discharges Q to Gnd?
b) Which transistor must it overpower?
c) What regions of operation are the transistors in?
d) Write the KCL equation for the two transistors

CMOS SRAM Analysis (Write)



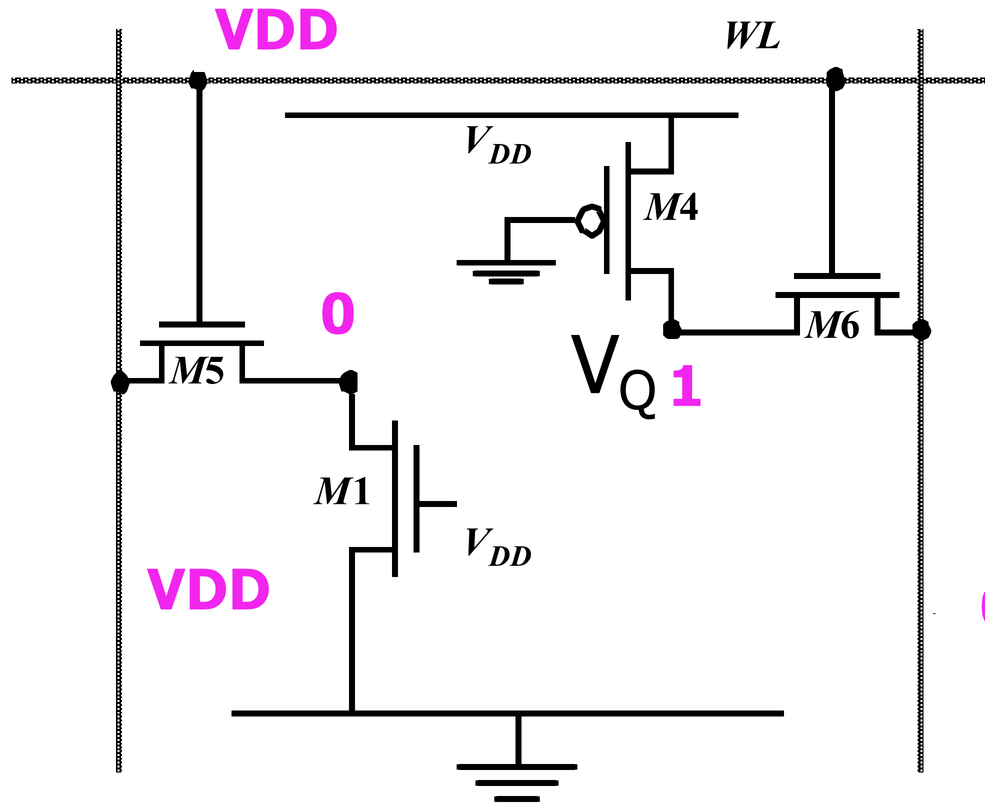
$$k_{p,M4}\left(V_{DD}-\left|V_{Tp}\right|\right)^2 = k_{n,M6}\left((V_{DD}-V_{Tn})V_Q-\frac{V_Q^2}{2}\right)$$

CMOS SRAM Analysis (Write)

VDD WL V_{DD} M4 M5 0 V_Q 1 M6 M1 V_{DD} VDD 0

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

CMOS SRAM Analysis (Write)



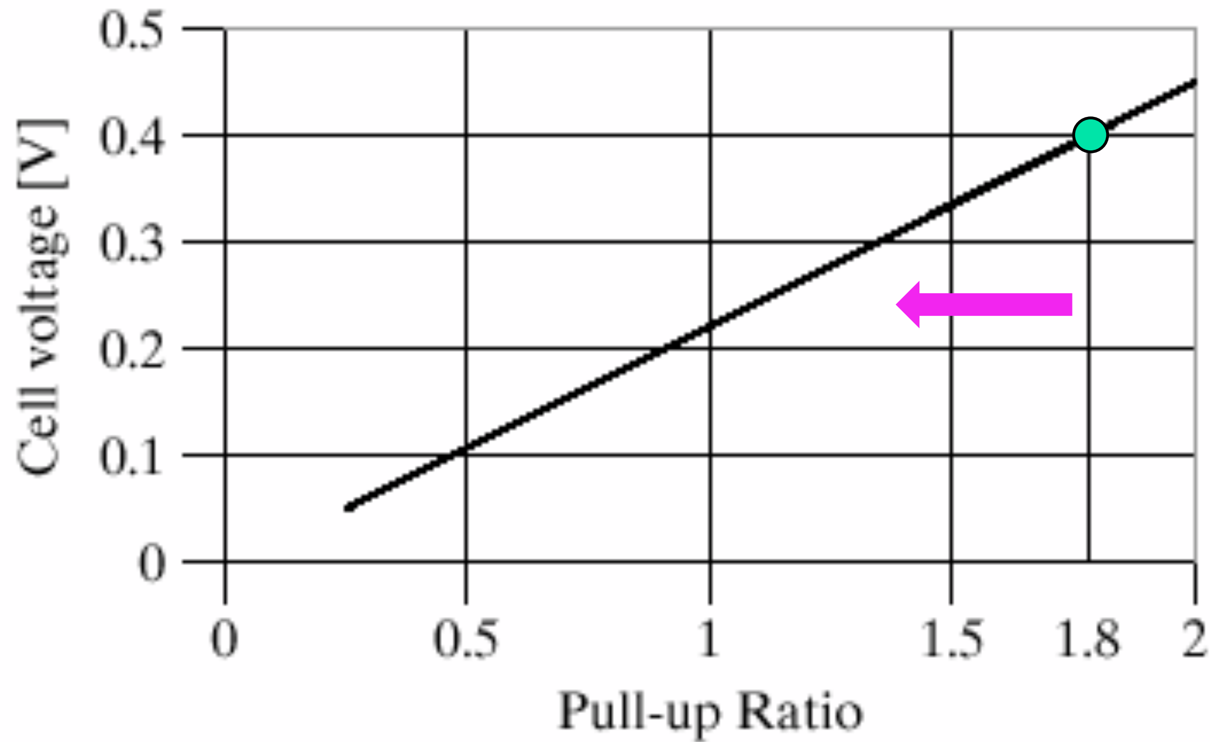
$$PR = \frac{W_4/L_4}{W_6/L_6}$$

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

$V_Q = V_{Tn}$ →

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

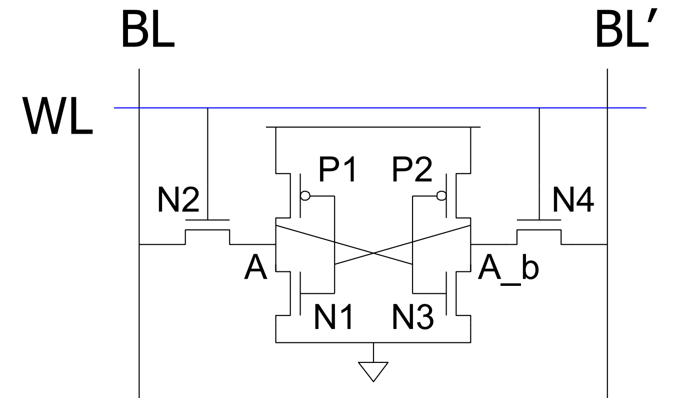
CMOS SRAM Analysis (Write)

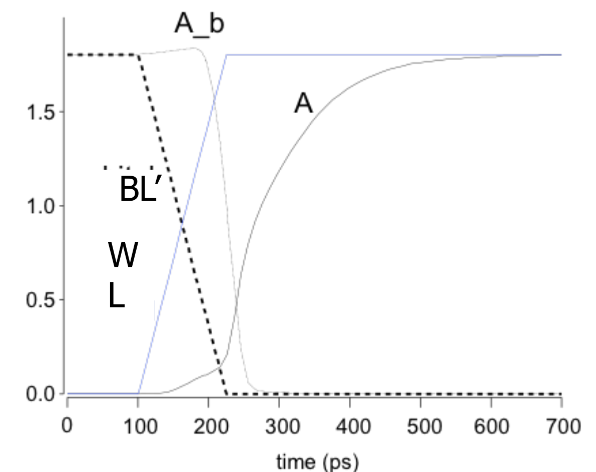


$$PR = \frac{W_4/L_4}{W_6/L_6}$$

SRAM Write

- Drive one bitline high, the other low
  - Depending on write data
- Then turn on wordline, WL
- Bitlines overpower cell with new value
- Ex: A = 0, A_b = 1, BL = 1, BL' = 0
  - Force A_b low, then A charges high
- *Writability*
  - Must overpower feedback inverter
  - N4 >> P2



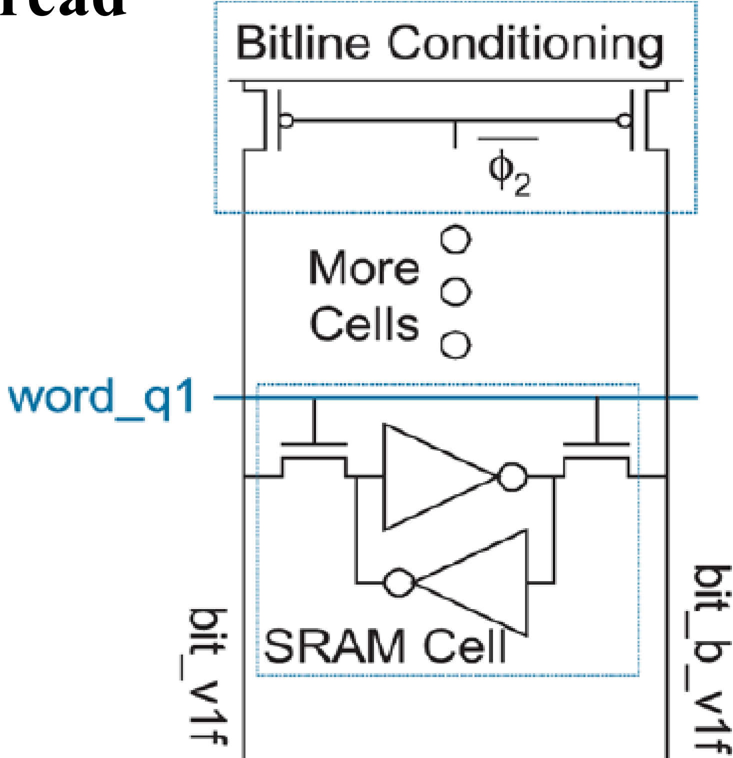


Column Circuitry

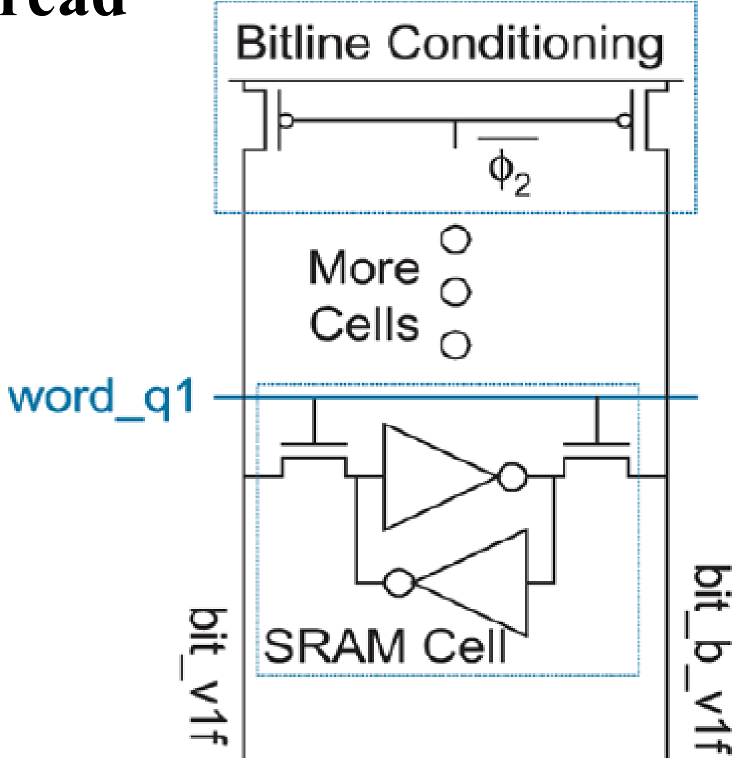
- Some circuitry is required for each column
  - Required: Bitline conditioning
    - Precharging
    - Driving input data to bitline
  - Increased speed: Sense amplifiers
  - Aspect ratio (square memory): Column multiplexing (AKA Column Decoders)

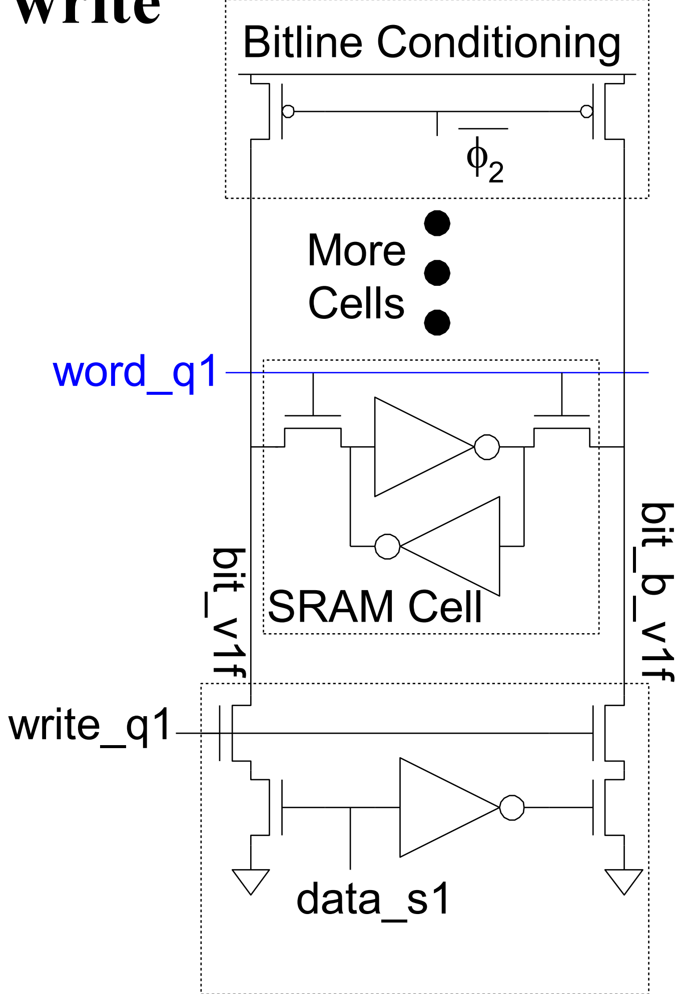
SRAM Column Example

read



SRAM Column Example





5T SRAM

WL

Waccess

Wbuf

Wrestore

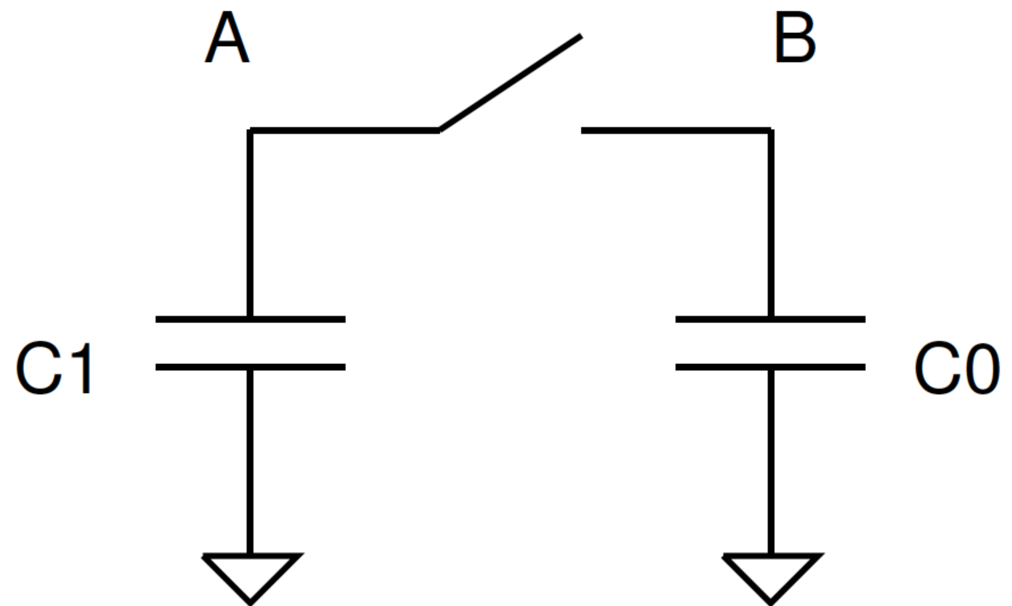
BL

Charge Sharing (Preclass 3)

Initially

- A @ 1V
- B @ 0V

❑ Q_A=1V*C1=C1



Charge Sharing (Preclass 3)

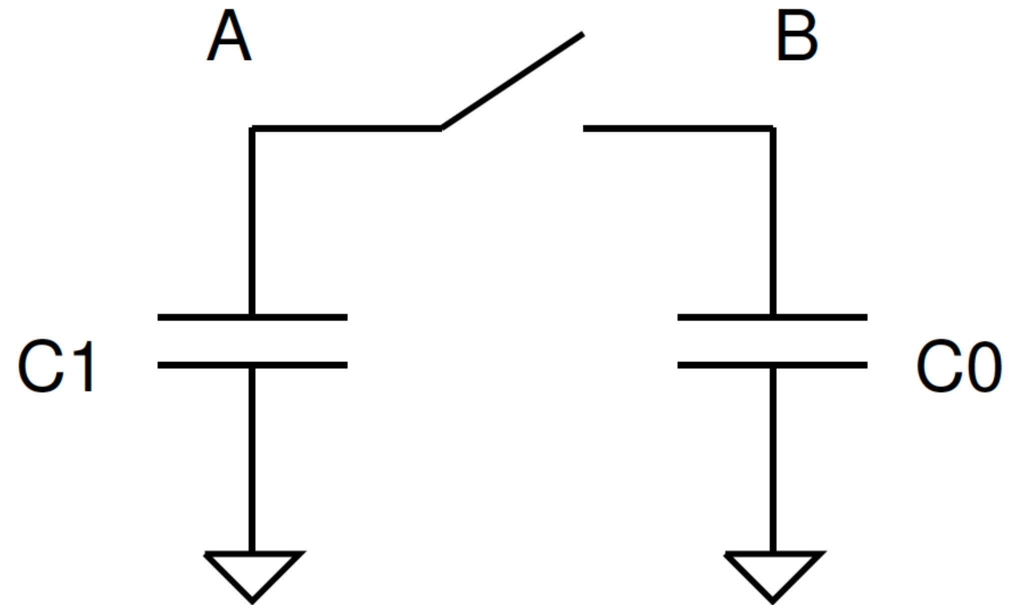
Initially

- A @ 1V
- B @ 0V

- $Q_A = 1V * C1 = C1$

Close switch

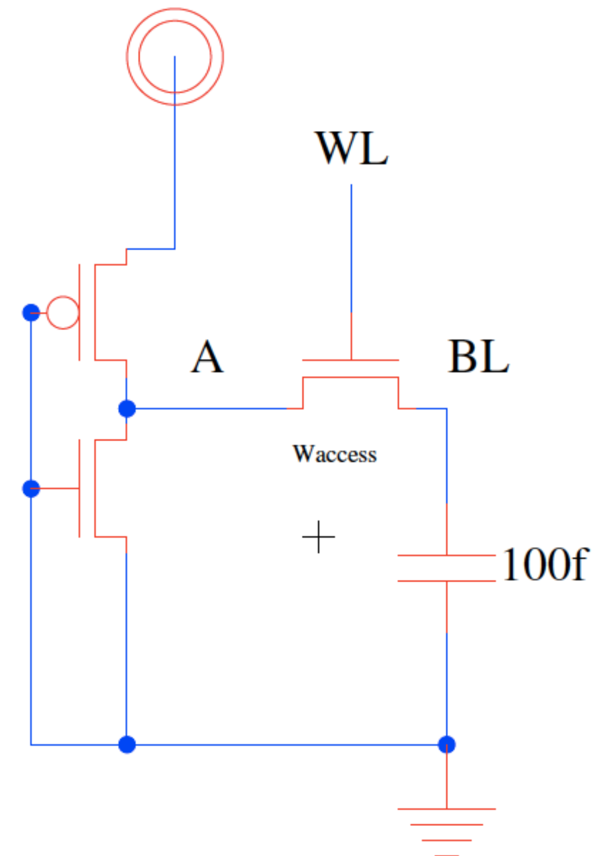
- $Q_{tot} = V_{final} * (C1 + C0)$
- Charge conservation
  - $Q_A = Q_{tot}$
- $C1 = V_{final} * (C1 + C0)$



$$V_{final} = \frac{C1}{C1 + C0}$$

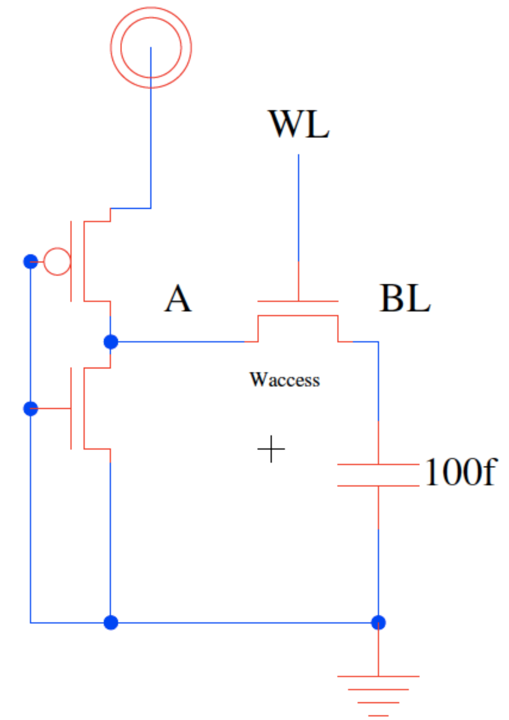
Consider (preclass 4)

- Read: What happens to voltage at A when WL turns from 0→1?
  - Assume W_{access} large
  - $W_{access} >> W_{pu}=1$
  - BL initially 0



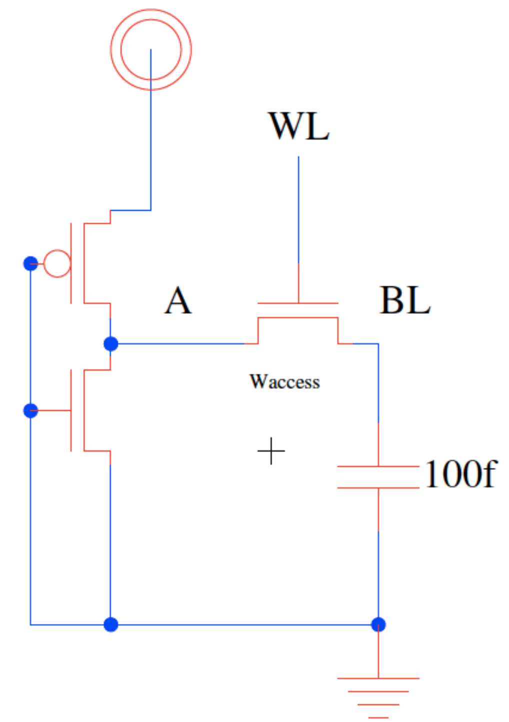
Voltage After enable Word Line

- $Q_{BL} = 0$
- $Q_A = (1V)(\gamma 2C_0+\gamma W_{access}C_0)$
- $100fF=C_{BL}>>C_A=(\gamma(2+W_{access})C_0)$

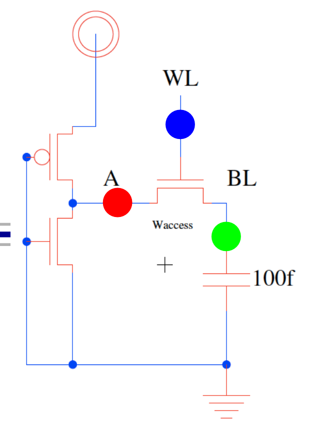


Voltage After enable Word Line

- $Q_{BL} = 0$
- $Q_A = (1V)(\gamma 2C_0 + \gamma W_{access} C_0)$
- $100fF = C_{BL} >> C_A = (\gamma(2 + W_{access})C_0)$
- After enable W_{access} (W_{access} large)
  - Total charge $Q_{BL} + Q_A$ unchanged
    - Charge conservation
  - Distributed over larger capacitance $\sim= C_{BL}$
  - $V_A = V_{BL} \sim= C_A / C_{BL}$



Simulation: W_{access}=100



Transient Response

V

1
0.8
0.6
0.4
0.2
0

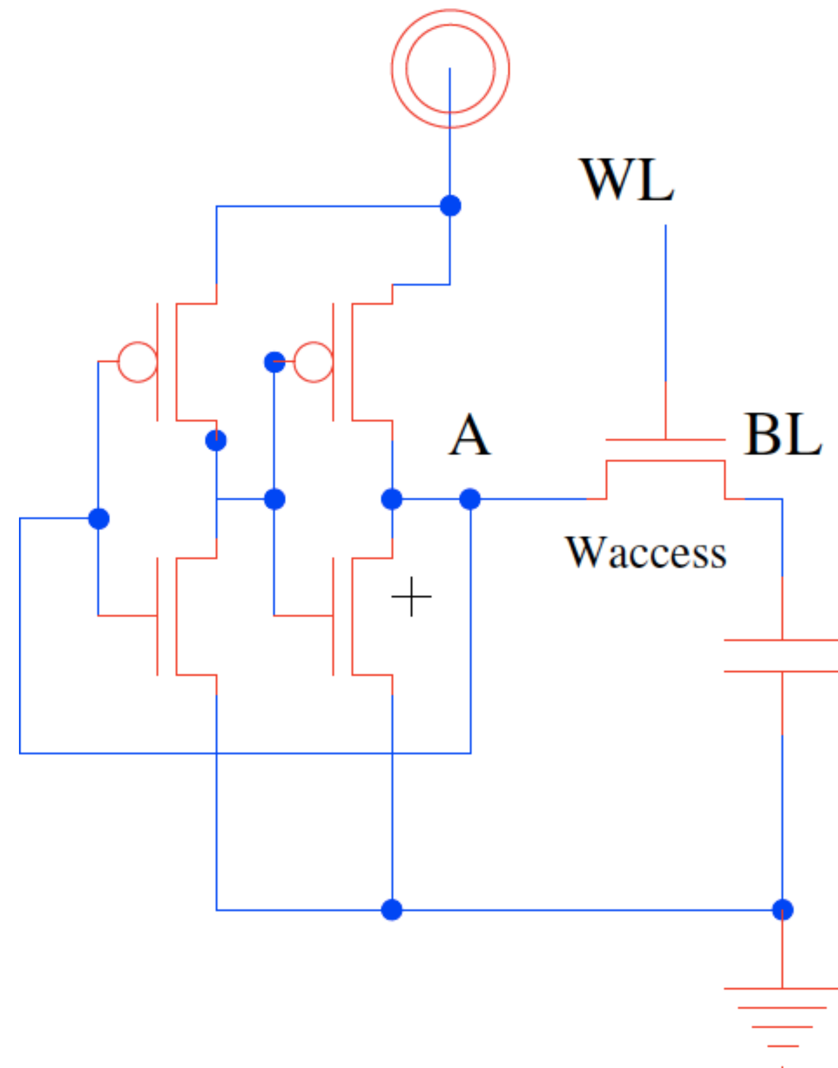
0 2e-10 4e-10 6e-10 8e-10 1e-09

s

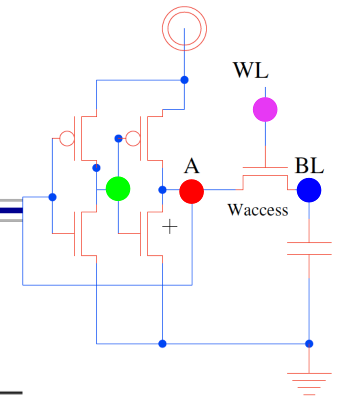
A
BL
WL

Consider (5T SRAM) (preclass 5)

- What happens to voltage at A when WL turns from 0→1?
  - Assume W_{access} large
  - A initially 1
  - BL initially 0



Simulation W_{access}=20



Transient Response

V

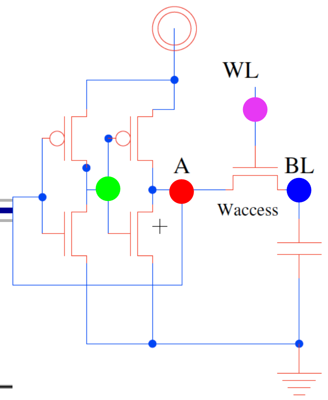
1
0.8
0.6
0.4
0.2
0

A
not A
BL
WL

1e-10 2e-10 3e-10 4e-10 5e-10 6e-10 7e-10 8e-10 9e-10 1e-09

s

Simulation W_{access}=4



Transient Response

V

1
0.8
0.6
0.4
0.2
0

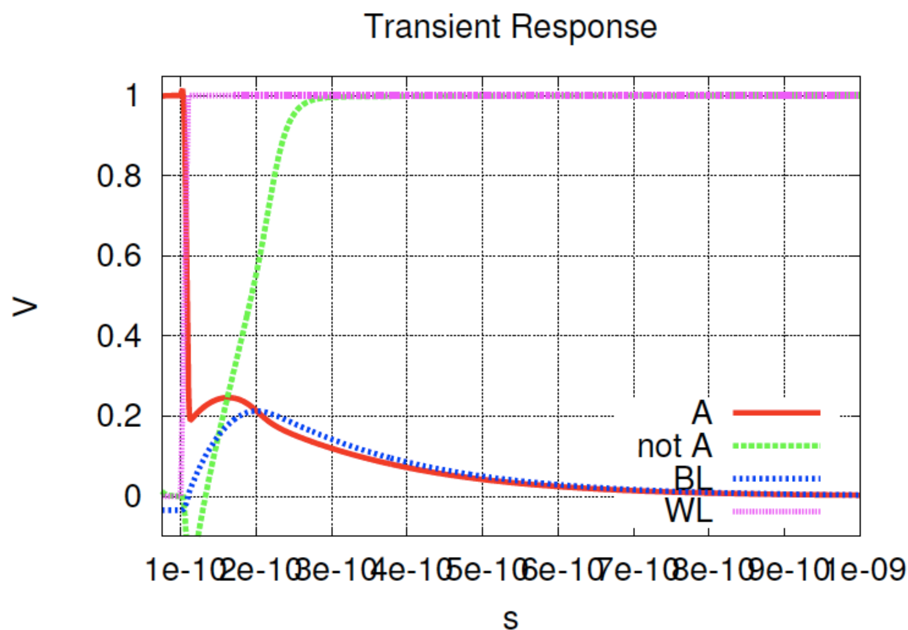
A
not A
BL
WL

1e-10 2e-10 3e-10 4e-10 5e-10 6e-10 7e-10 8e-10 9e-10 1e-09

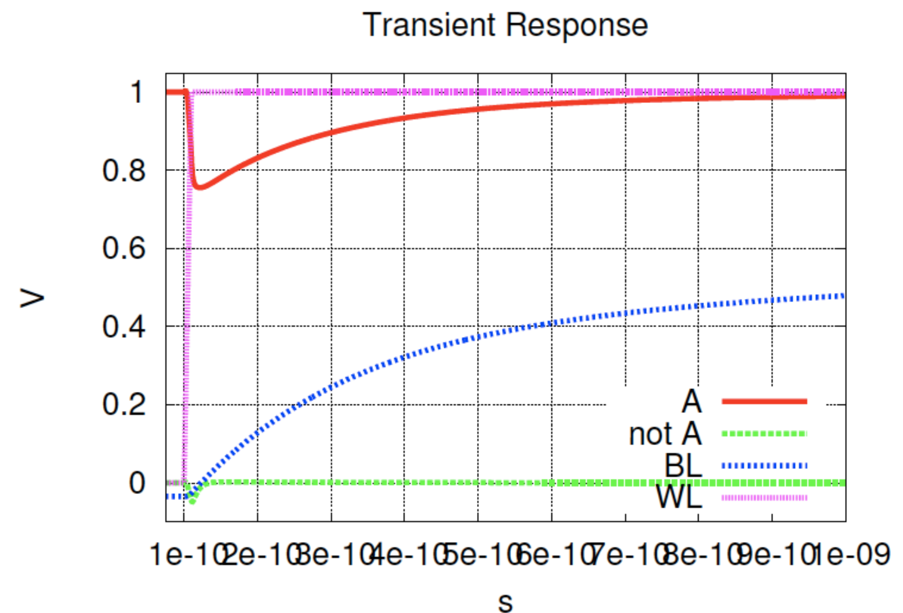
s

Charge Sharing

- **Conclude:** charge sharing can lead to read upset
  - Charge redistribution/sharing adequate to flip state of bit



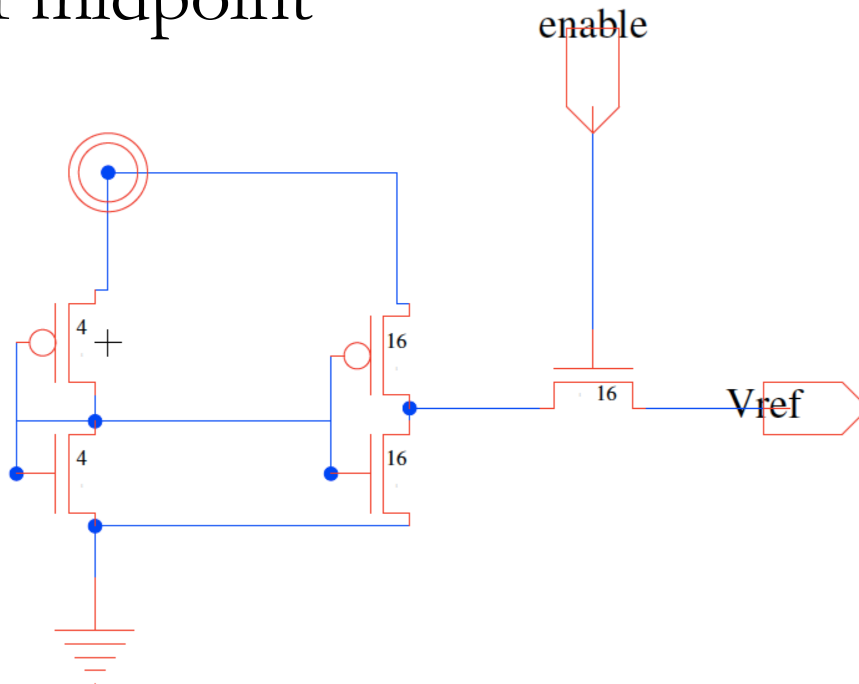
$W_{access}=20$

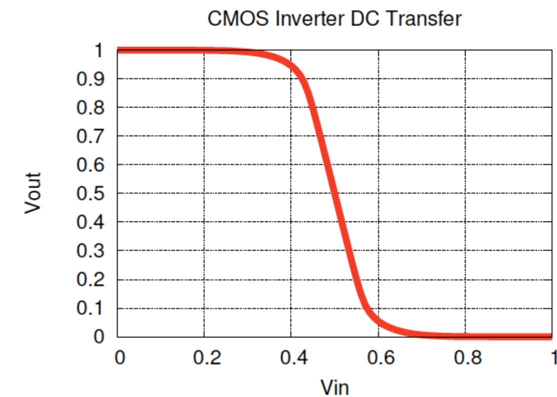


$W_{access}=4$

Charge to middle Voltage

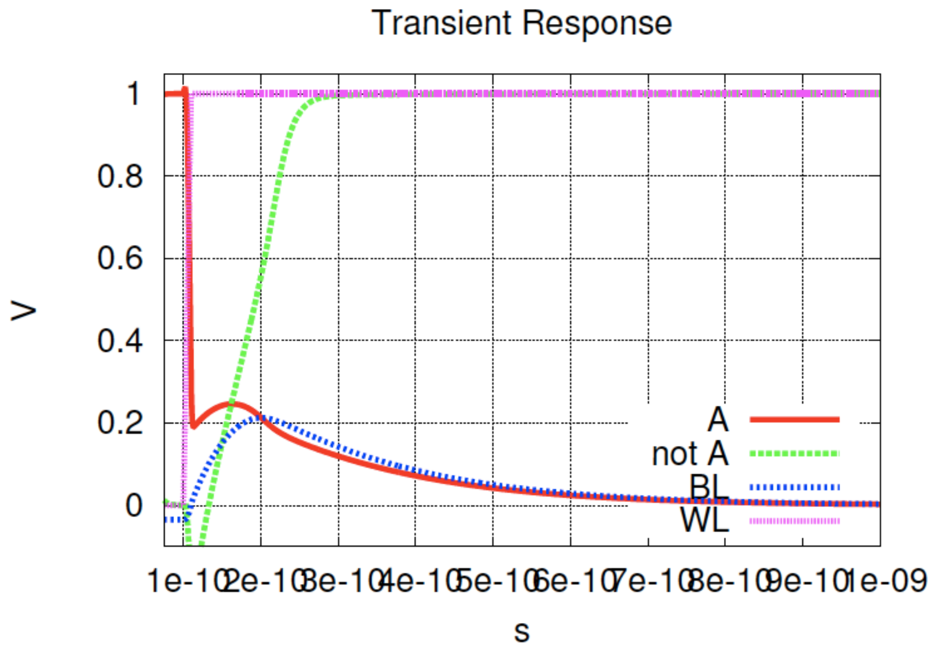
- Pre-charge bitlines to $V_{dd}/2$ before begin read operation
- Now charge sharing doesn't swing to opposite side of midpoint

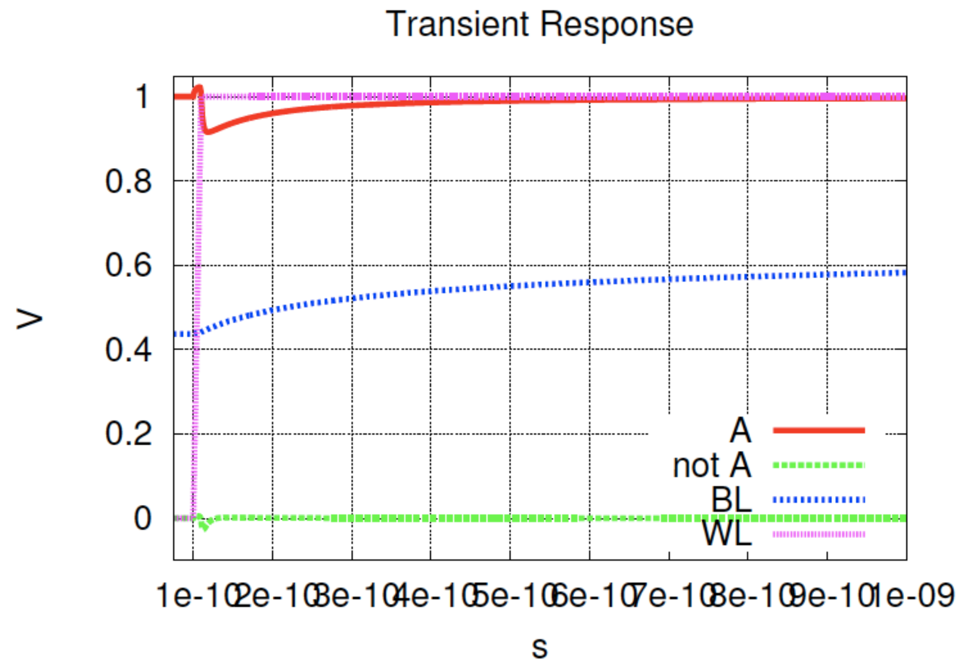




Compare

- Both W_{access}=20; vary BL precharge voltage

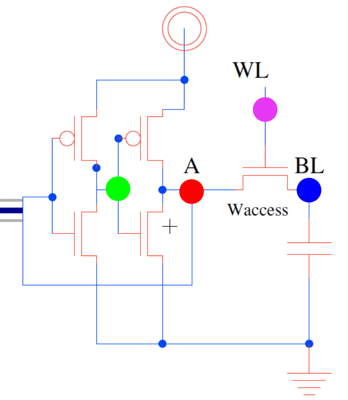




0 precharge

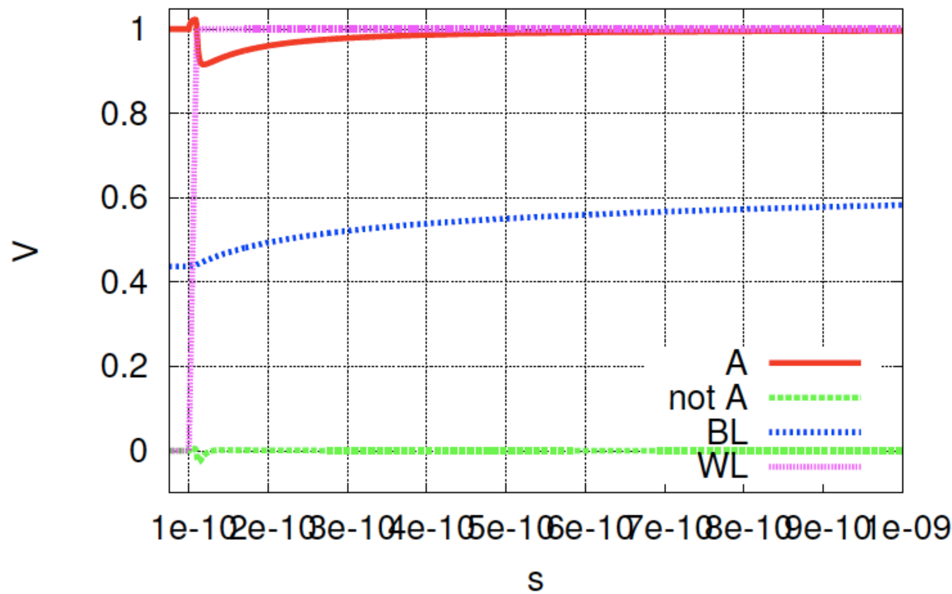
$V_{dd}/2$ precharge

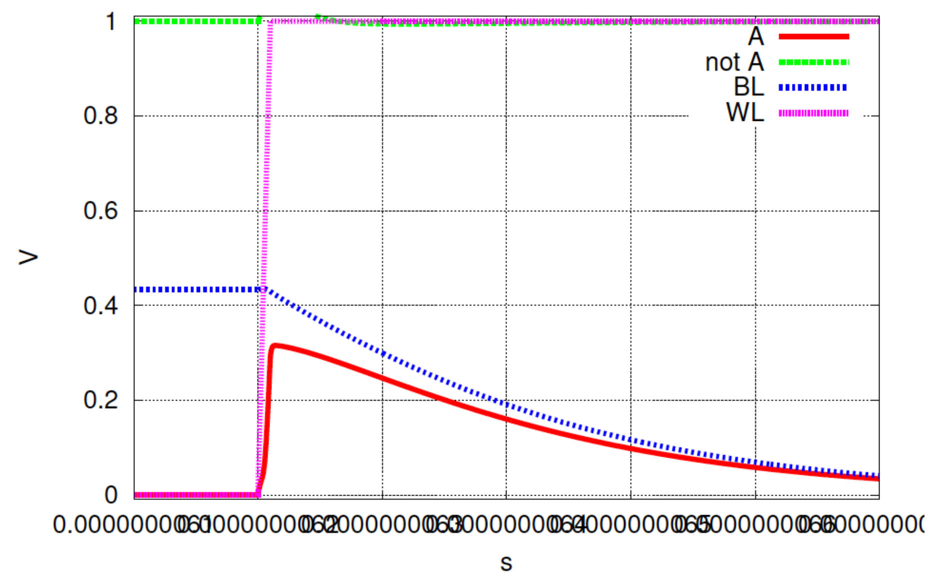
Simulation W_{access}=20 (precharge Vdd/2, reading 0)



Transient Response (reading 0)

Simulation W_{access}=20 (with precharge Vdd/2)



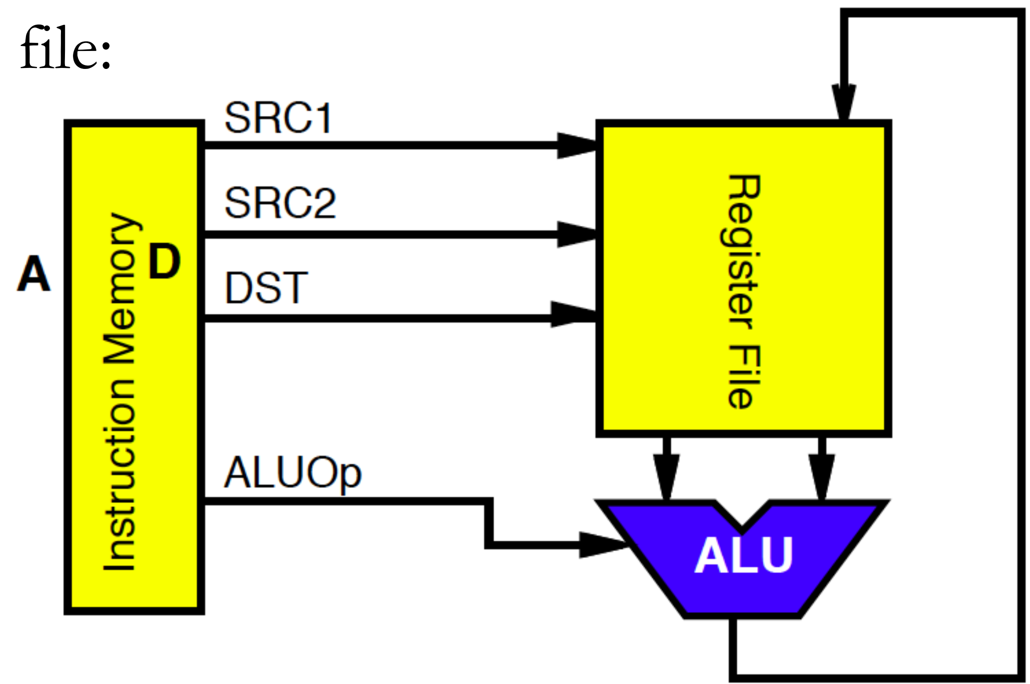


Read 1

Read 0

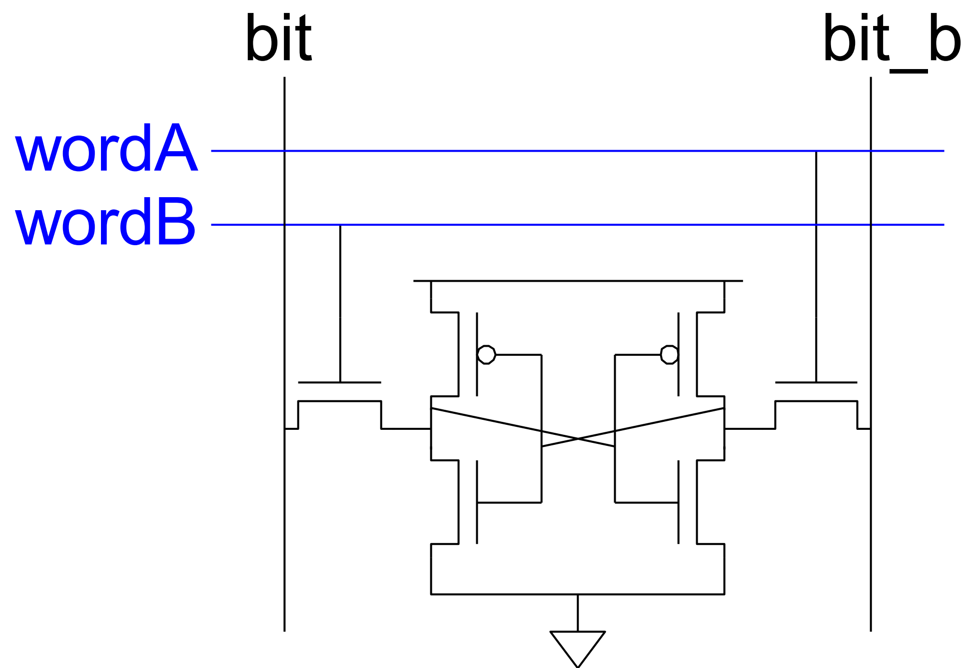
Multiple Ports

- We have considered single-ported SRAM
  - One read or one write on each cycle
- *Multiported* SRAM are needed for register files
- Examples:
  - Pipelined ALU register file:
    - add r1,r2,r3
    - R3←R1+R2
    - Requires two reads and one write



Dual-Ported SRAM

- Simple dual-ported SRAM
  - Two independent single-ended reads
  - Or one differential write



Dual-Ported SRAM

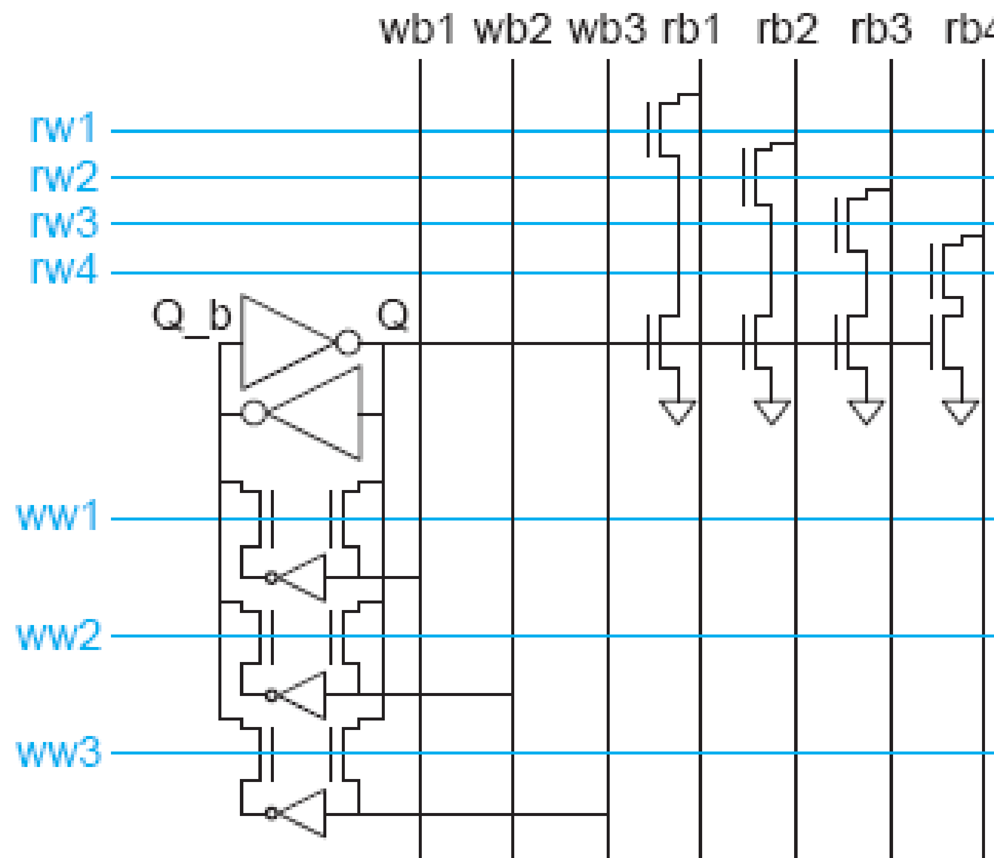
- Simple dual-ported SRAM
  - Two independent single-ended reads
  - Or one differential write

bit
bit_b
wordA
wordB

- Do two reads and one write by time multiplexing
  - Read during ph1, write during ph2

Multi-Ported SRAM

- Adding more access transistors hurts read stability
- Multiported SRAM isolates reads from state node
- Single-ended bitlines save area



Register File Cell

- Single-ended 2-read/1-write ports (Slow-write)

bitWR
bitRD1
bitRD2
I1
I3
A
N2
N1
N3
N4
I2
wordWR
wordRD1
wordRD2
wrEN

DRAM

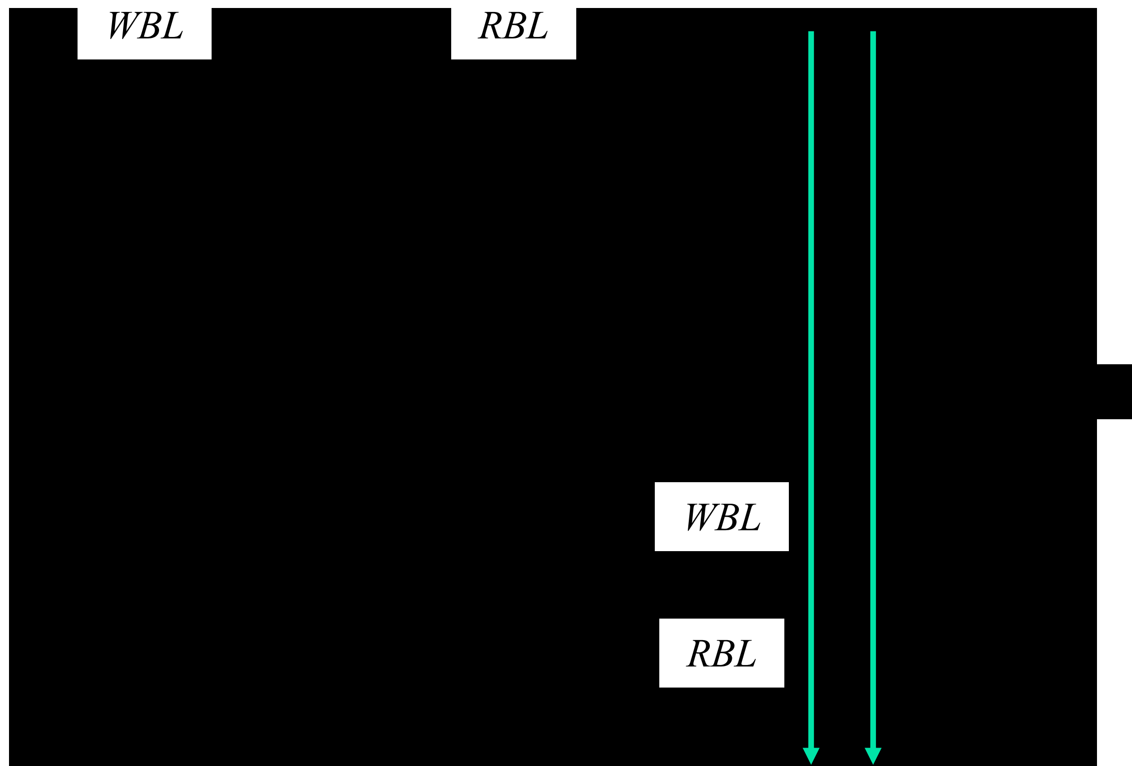
- Smaller than SRAM
- Require data refresh to compensate for leakage

3-Transistor DRAM Cell

- Cell is inverting on read operation

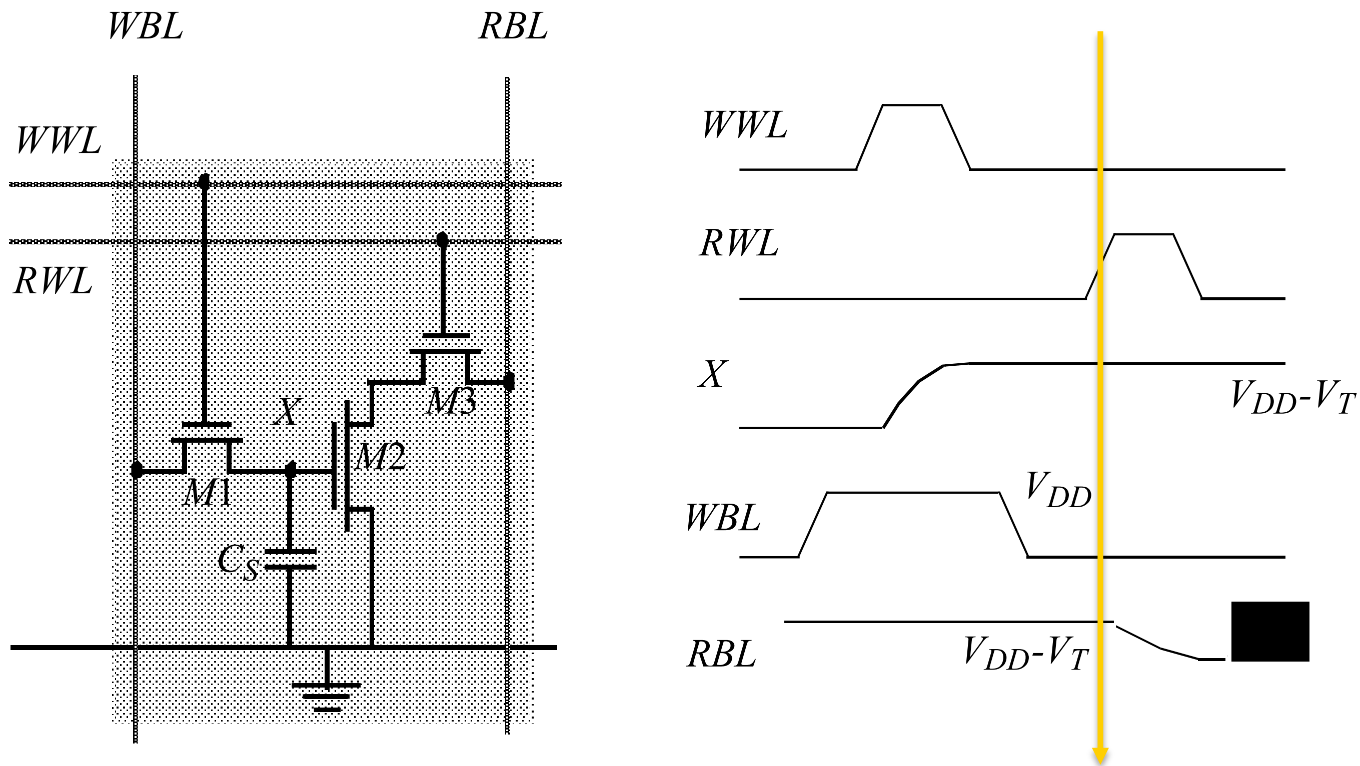
3-Transistor DRAM Cell

- Cell is inverting on read operation



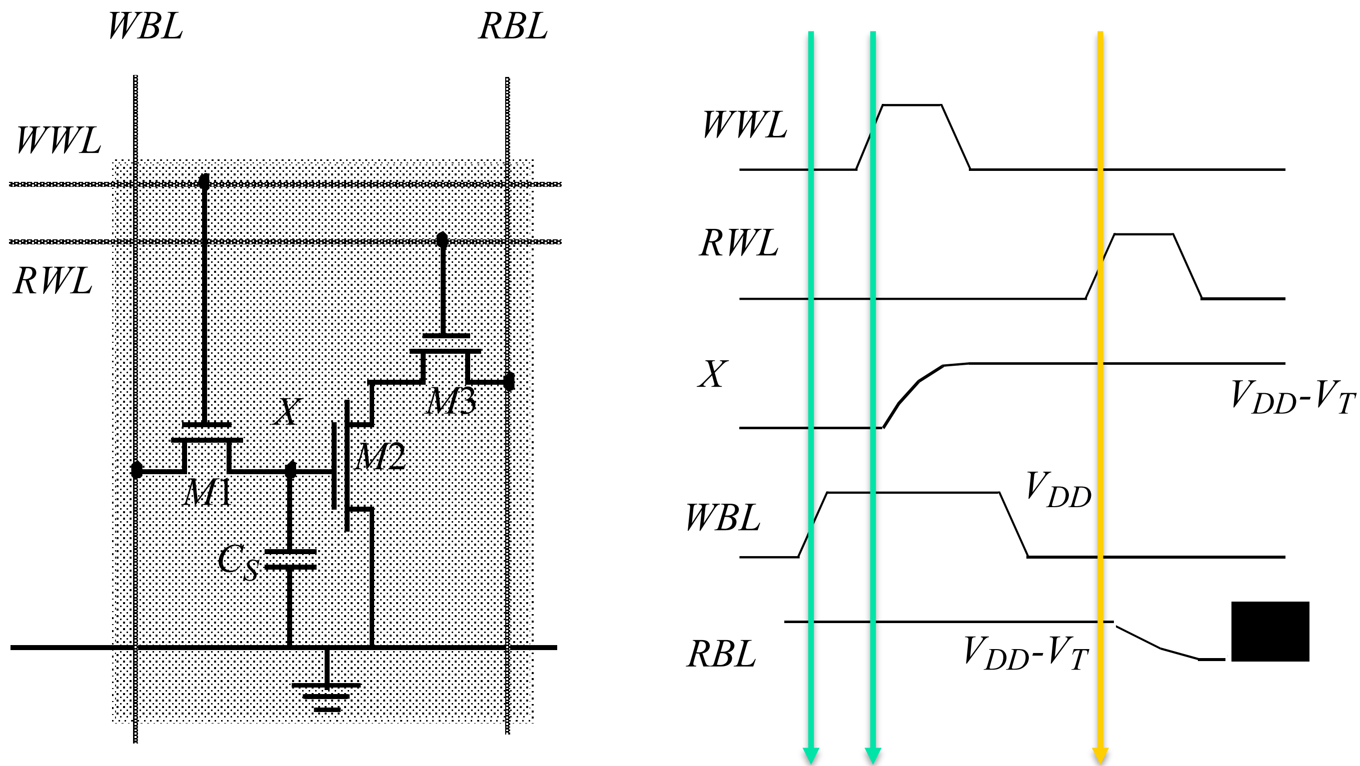
3-Transistor DRAM Cell

- Cell is inverting on read operation



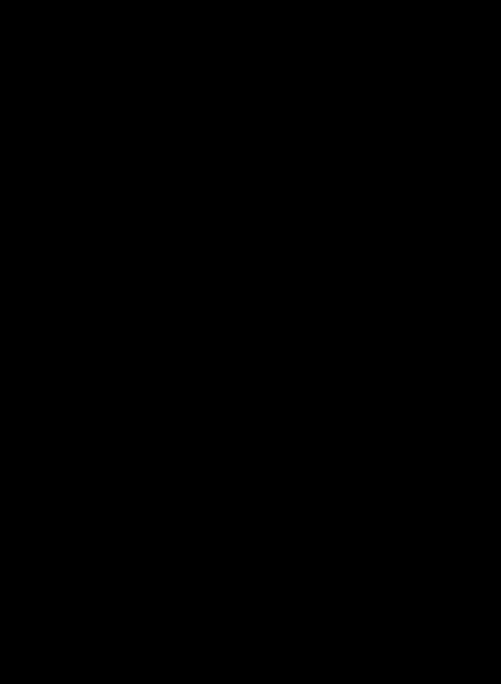
3-Transistor DRAM Cell

- Cell is inverting on read operation

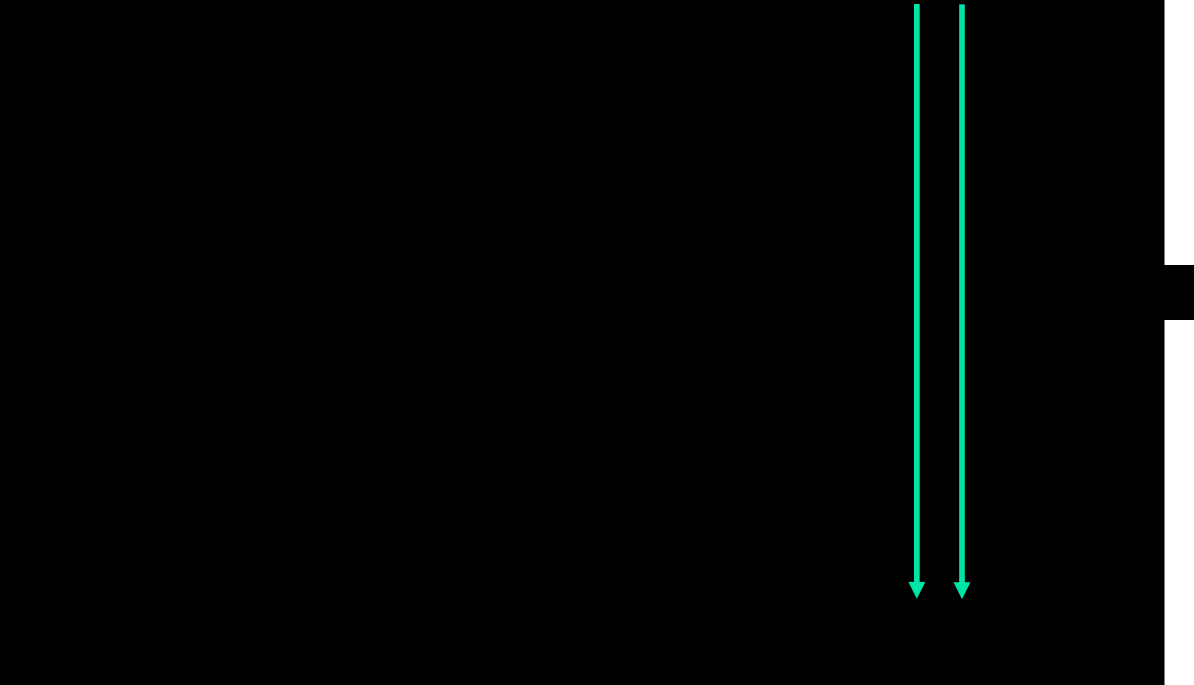


- No constraints on device ratios
- Reads are non-destructive
- Data stored has a V_T drop
  - When storing a 1, value at X = V_{WWL}-V_{Tn}

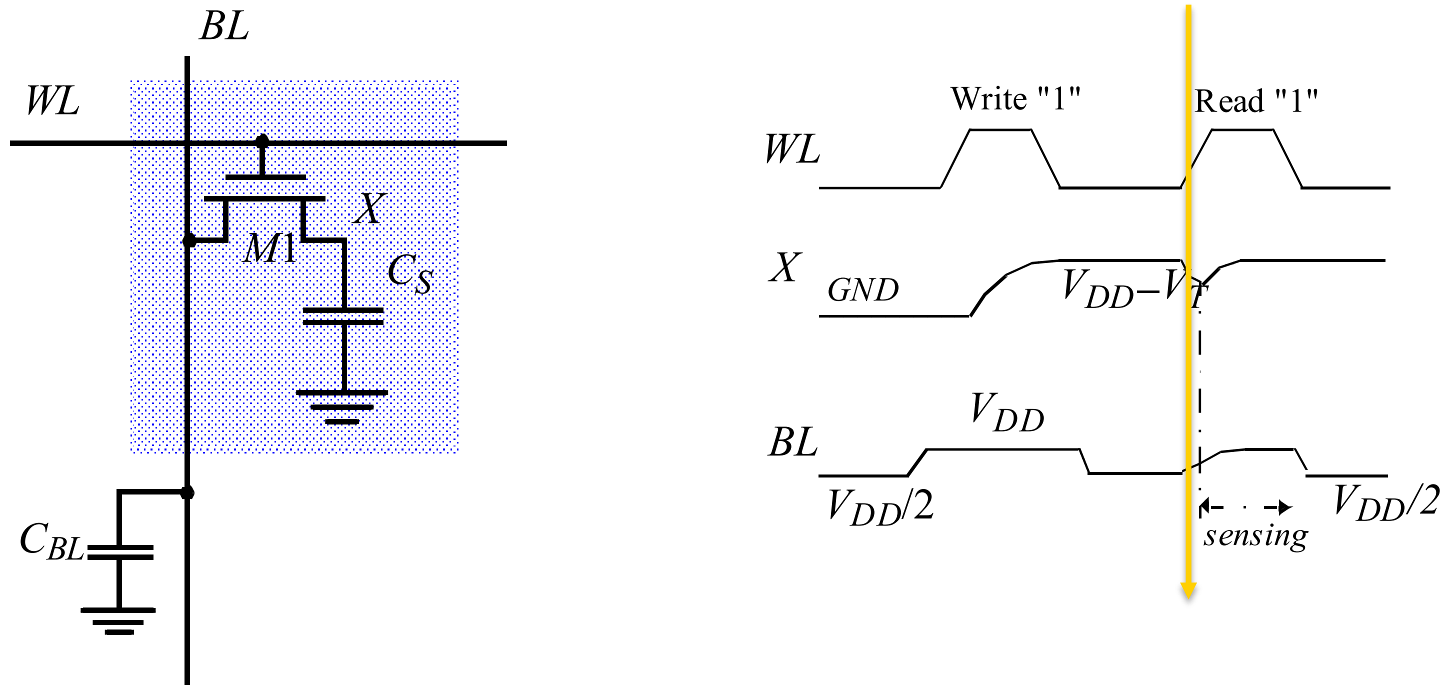
1-Transistor DRAM Cell



1-Transistor DRAM Cell



1-Transistor DRAM Cell

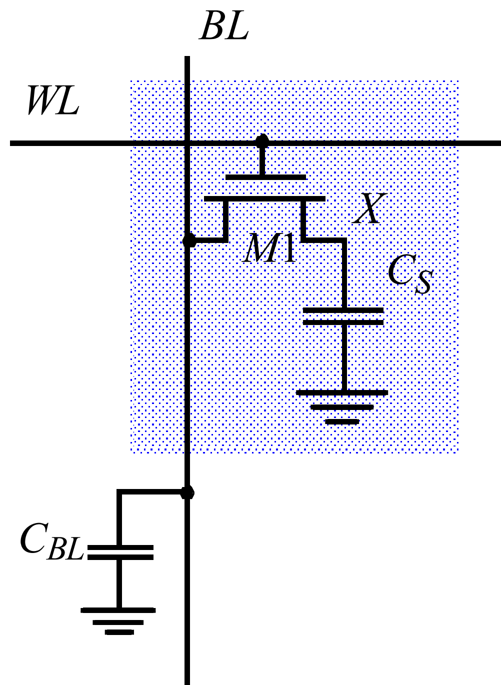


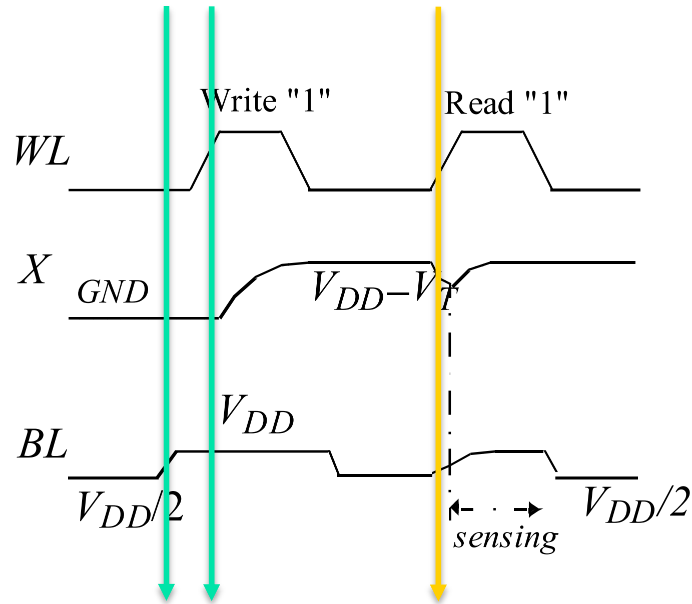
Write: C_S is charged or discharged by asserting WL and BL.
Read: Charge redistribution takes places between bit line and storage capacitance

$$\Delta V = V_{BL} - V_{PRE} = (V_{BIT} - V_{PRE})\frac{C_S}{C_S + C_{BL}}$$

Voltage swing is small; typically around 250 mV.

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 operation 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 loss can be circumvented by bootstrapping the word lines to a higher value than V_{DD}.

Big Idea

- 6T SRAM
  - Robust cell when sized carefully
- 5T SRAM
  - More sensitive to sizing than 6T SRAM
- Minimize area of repeated cell
  - 6T/5T SRAM
  - Mulitport trade off area for function
  - 1T/3T DRAM helps but slower
- Compensate with periphery (next time…)
  - Decoders
  - Bitline (column) drivers
  - Sensing/Amplification (regeneration/restoration)

Admin

- Homework 6 due tonight
- NEW Homework 7
  - Come to lab 4/22 and turn in worksheet for credit
- Project 2 out now
  - Design SRAM array
  - Work in teams of up to two
  - Milestone due F 4/19
    - Will get feedback from me by M 4/22
  - Final report due Wednesday 5/1

Acknowledgement

- Prof. André DeHon (University of Pennsylvania)
- Prof. Jing Li (University of Pennsylvania)