# EECS 427
# Lecture 16: Memory Core and Peripherals

Readings: 12.1-12.3



## Reminders

- CAD assignments
  - CAD7 is due tomorrow at noon
  - CAD8 (last one!) is due next Thursday at noon
- ECE Graduate Symposium
  - Poster session in ECE Atrium at 11-2 pm on Friday
  - Graduate students will be available to answer questions 11-12:30 pm
- HW4 (detailed proposal) is due Tuesday 11/17
  - Send your proposal (in Word or LaTeX) by email to the Zhengya by 11:59 pm
- Quiz 2 on Wednesday 11/25
  - Review in class on Monday 11/23

# Last Time

- Register timing
  - Skew and jitter
  - Used in static logic
  - Setup time, hold time, clock-to-q time
- Latch timing
  - Used in domino logic
  - Allows slack passing and time borrowing
- Master-slave register
  - Elimination of hold switch to break the feedback loop
  - Static vs. dynamic
  - Be aware of clk and clk_b overlaps
- Variants
  - $C^2MOS$ – immune to clock overlap
  - TSPC – only clk or clk_b is needed



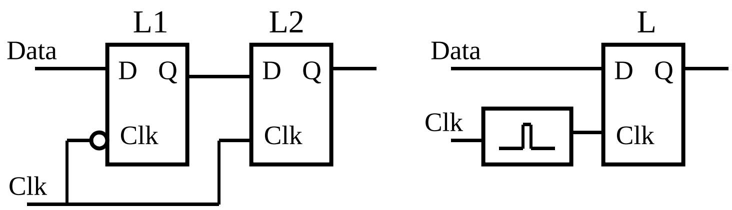
# Overview

- Pulse latches
  - Glitch generation
  - AMD K6 hybrid latch-flip-flop
- SRAM
  - Configuration
  - Sizing for read and write
  - Address decoding
  - Sense amplifier design

# Pulse-Triggered Latches

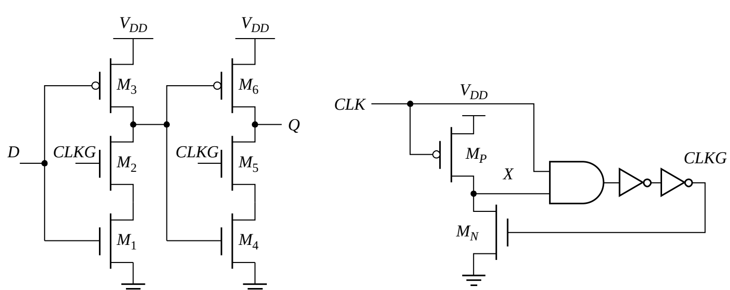
Ways to design an edge-triggered sequential cell:

Master-Slave Latches — Pulse-Triggered Latch

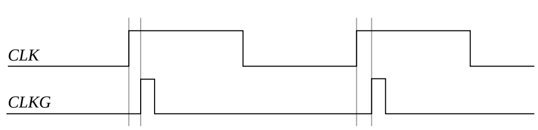





# Pulsed Latches



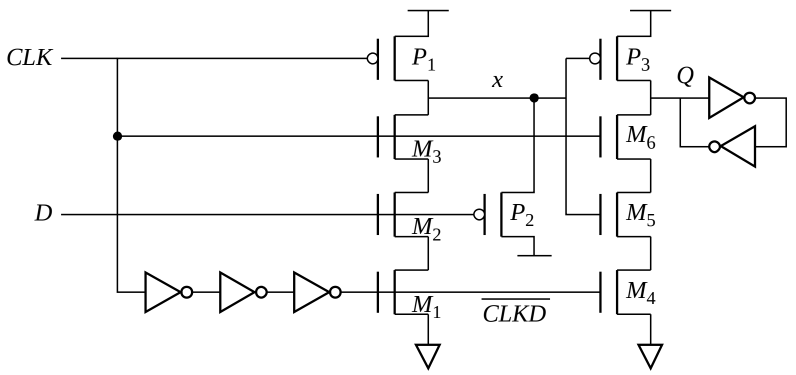

(a) register (b) glitch generation




(c) glitch clock

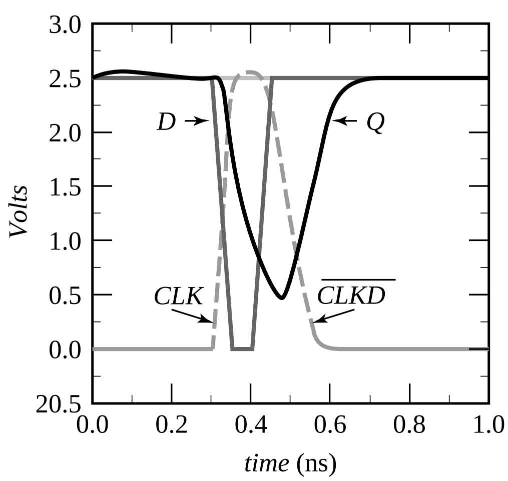
# Pulsed Latches

Hybrid Latch – Flip-flop (HLFF), AMD K-6 and K-7 :



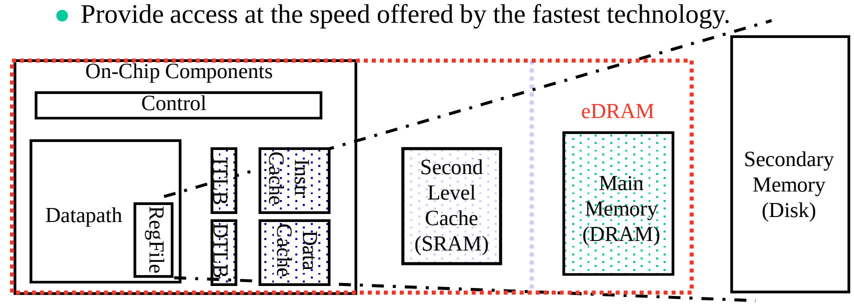



# Hybrid Latch-FF Timing



# A Typical Memory Hierarchy

- By taking advantage of the principle of locality:
  - Present the user with as much memory as is available in the cheapest technology.
  - Provide access at the speed offered by the fastest technology.



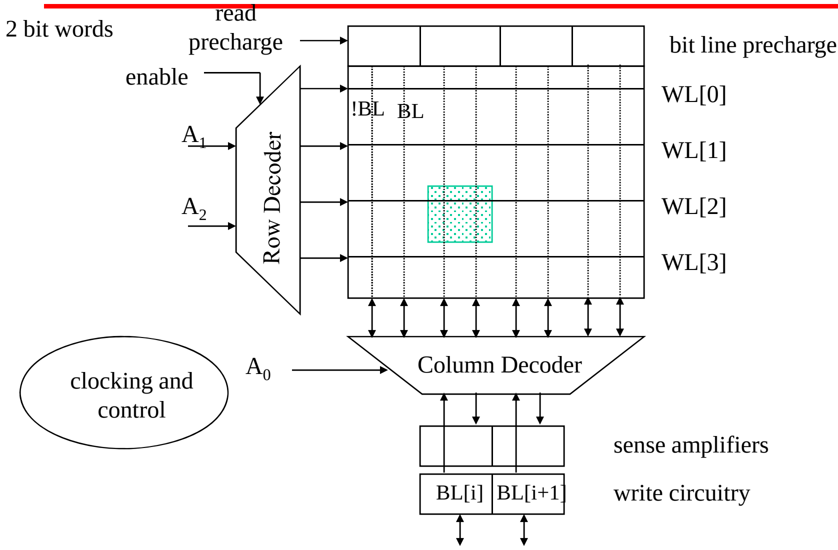

| | | | | | |
|---|---|---|---|---|---|
| **Speed (ns):** | .1's | 1's | 10's | 100's | 1,000's |
| **Size (bytes):** | 100's | K's | 10K's | M's | T's |
| **Cost:** | highest | | | | lowest |



# Read-Write Memories (RAMs)

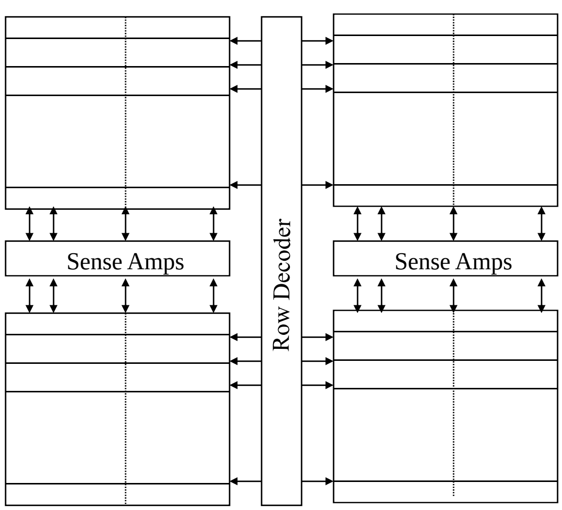
- Static – SRAM
  - data is stored as long as supply is applied
  - large cells (6 fets/cell) – so fewer bits/chip
  - fast – so used where speed is important (e.g., caches)
  - differential outputs (output BL and !BL)
  - use sense amps for performance
  - compatible with CMOS technology
- Dynamic – DRAM
  - periodic refresh required
  - small cells (1 to 3 fets/cell) – so more bits/chip
  - slower – so used for main memories
  - single ended output (output BL only)
  - need sense amps for correct operation
  - not typically compatible with CMOS technology

# 4x4 SRAM Memory

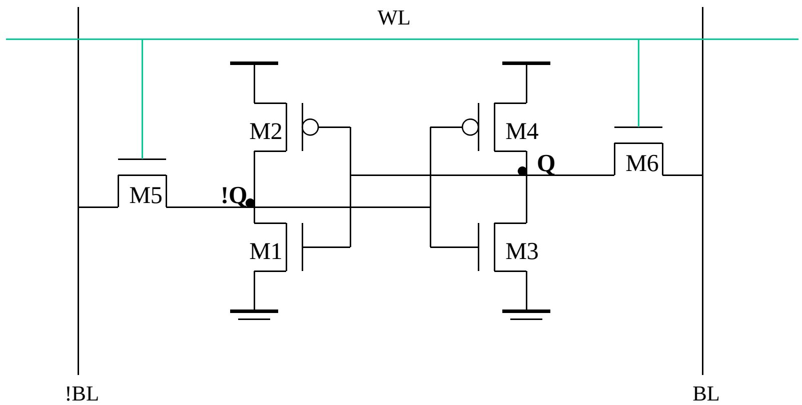





# 2D Memory Configuration



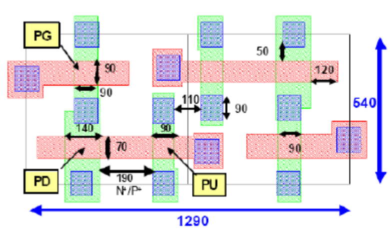
# 6-transistor SRAM Cell

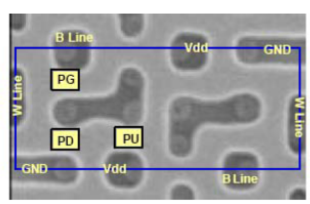





# SRAM Layout

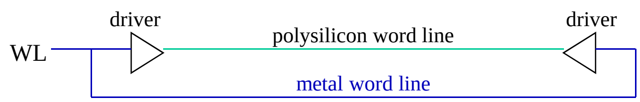
Freescale 65nm SRAM cell: $0.69um^2$



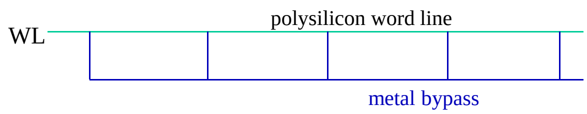



# Decreasing Word Line Delay

- Drive the word line from both sides




- Use a metal bypass



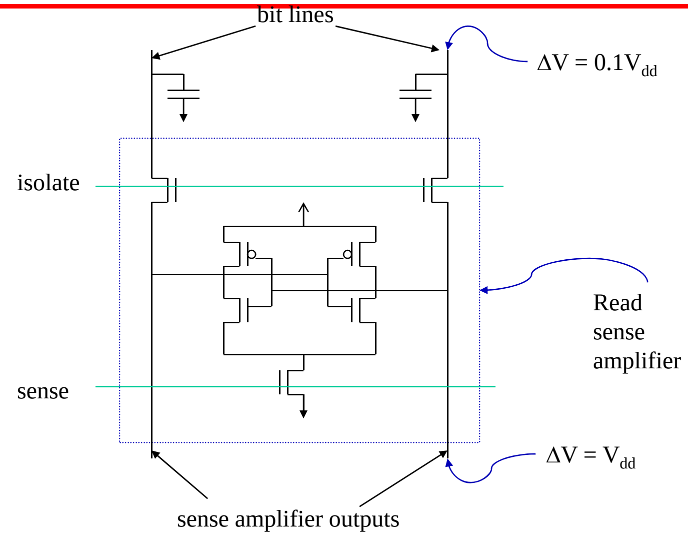

- Use silicides



# Decreasing Bit Line Delay

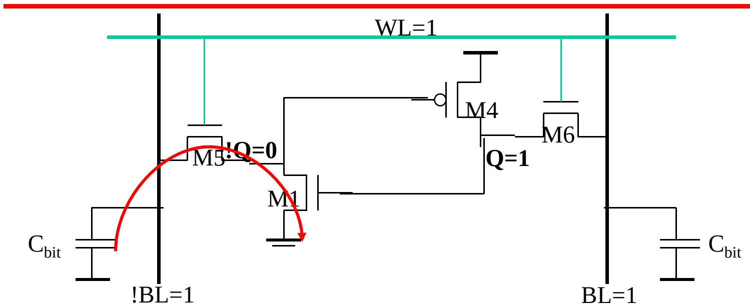
- Reduce the bit line voltage swing
  - need sense amp for each column to sense/restore signal
- Isolate memory cells from the bit lines after sensing (to prevent the cells from changing the bit line voltage further) - pulsed word line
  - generation of word line pulses very critical
    - too short - sense amp operation may fail
    - too long - power efficiency degraded (because bit line swing size depends on duration of the word line pulse)
- Isolate sense amps from bit lines after sensing (to prevent bit lines from having large voltage swings) - bit line isolation

# Bit Line Isolation



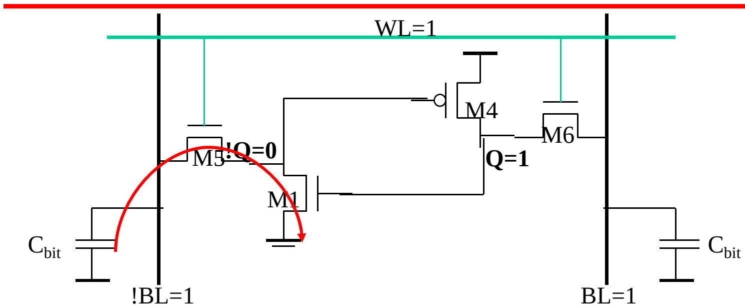



# SRAM Cell Analysis (Read)




Read-disturb (read-upset): must carefully limit the allowed voltage rise on !Q to a value that prevents the read-upset condition from occurring while simultaneously maintaining acceptable circuit speed and area constraints

# SRAM Cell Analysis (Read)

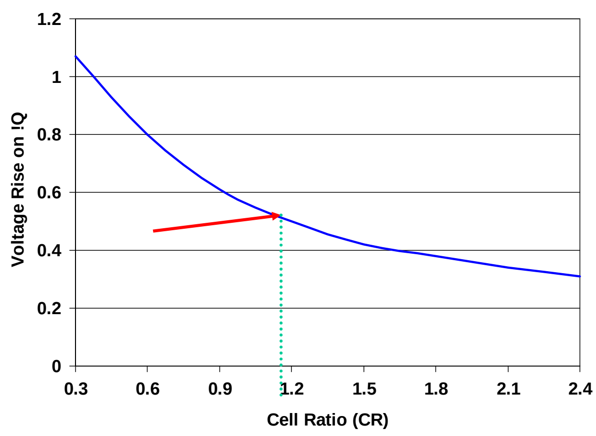



Cell Ratio (CR) = $(W_{M1}/L_{M1})/(W_{M5}/L_{M5})$

$V_{!Q} = [(V_{dd} - V_{Tn})(1 + CR \pm \sqrt{(CR(1 + CR))}]/(1 + CR)$

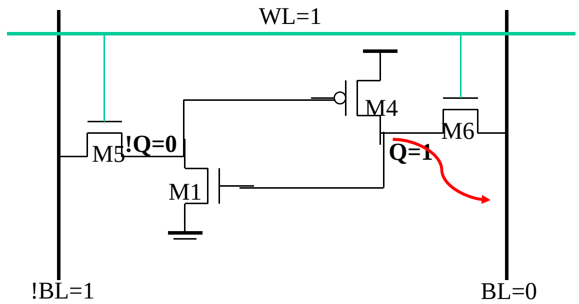
EECS 427 F09 Lecture 16 19

# Read Voltages Ratios




$V_{dd}$ = 2.5V
$V_{Tn}$ = 0.5V

# SRAM Cell Analysis (Write)

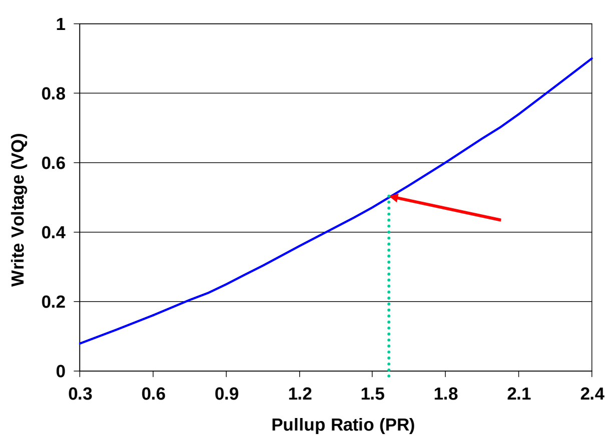



Pullup Ratio (PR) = $(W_{M4}/L_{M4})/(W_{M6}/L_{M6})$

$$V_Q = (V_{dd} - V_{Tn}) \pm \sqrt{((V_{dd} - V_{Tn})^2 - (\mu_p/\mu_n)(PR)((V_{dd} - V_{Tn} - V_{Tp})^2)}$$

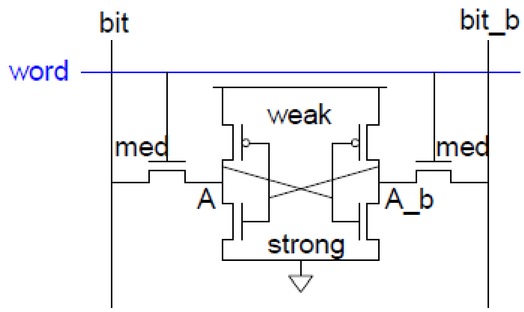


# Write Voltages Ratios



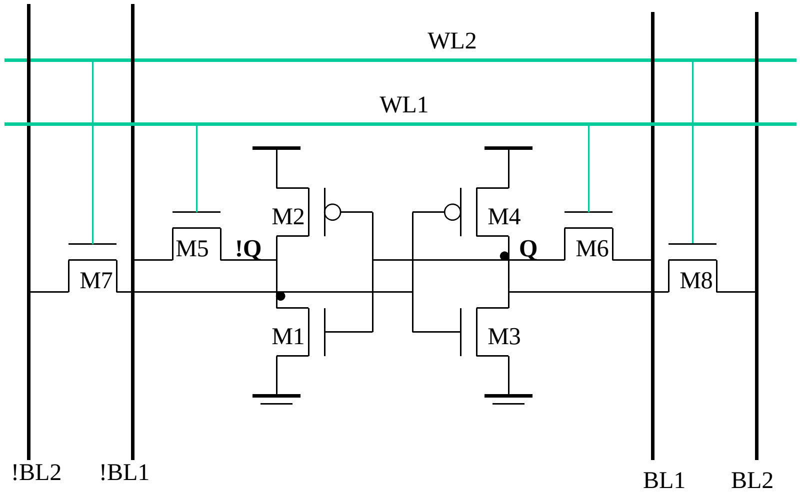

$V_{dd} = 2.5V$
$|V_{Tp}| = 0.5V$
$\mu_p/\mu_n = 0.5$

# Cell Sizing






# Multiple Read/Write Port Cell



# Row Decoders

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

(N)AND Decoder

$$WL_0 = A_0A_1A_2A_3A_4A_5A_6A_7A_8A_9$$

$$WL_{511} = \bar{A}_0A_1A_2A_3A_4A_5A_6A_7A_8A_9$$

NOR Decoder

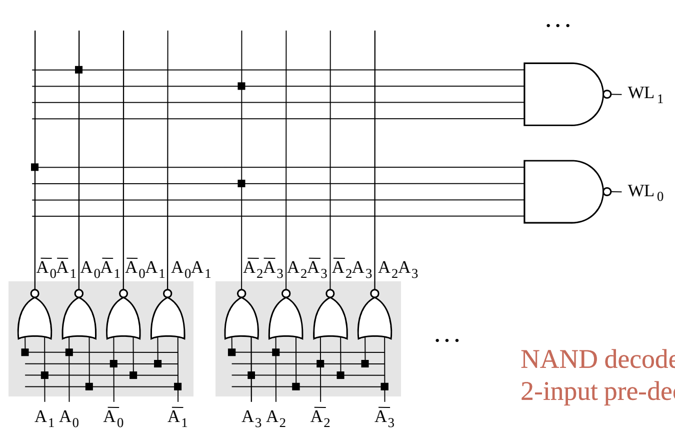
$$WL_0 = \overline{A_0+A_1+A_2+A_3+A_4+A_5+A_6+A_7+A_8+A_9}$$

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



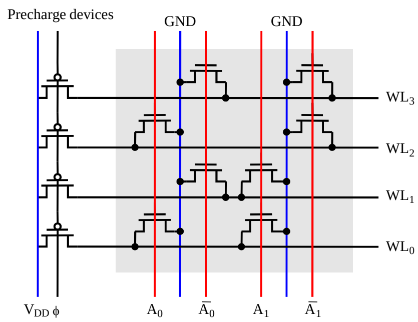
# Hierarchical Decoders

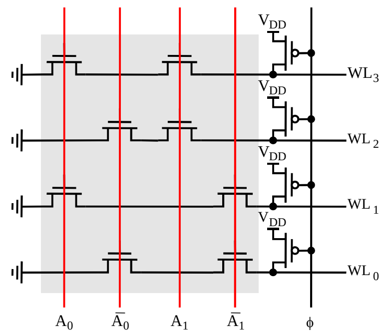
Multi-stage implementation improves performance




NAND decoder using
2-input pre-decoders

# Dynamic Decoders



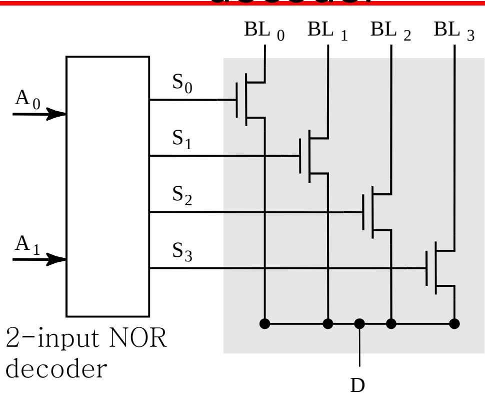




2-input NOR decoder

2-input NAND decoder



# 4-input pass-transistor based column decoder

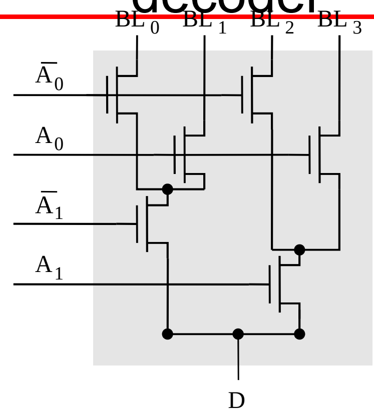



Advantages: speed ($t_{pd}$ does not add to overall memory access time)

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

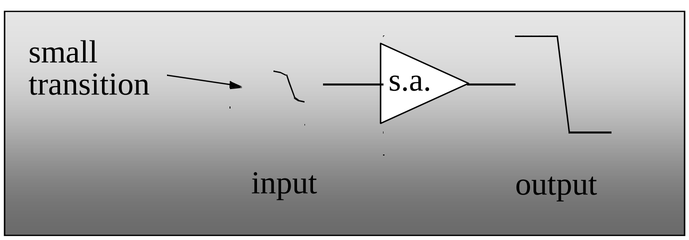


# Sense Amplifiers

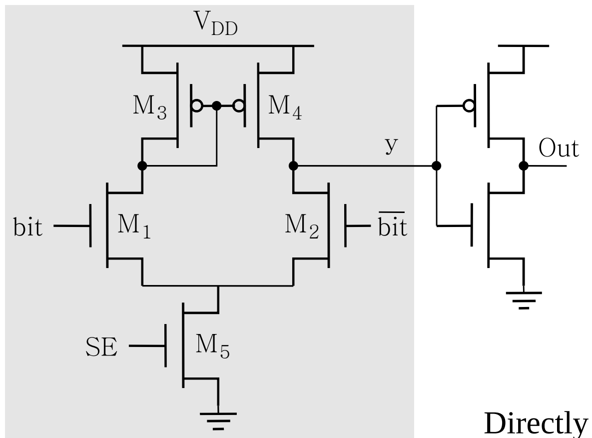
$$t_p = \frac{C \times \Delta V}{I_{av}}$$

- make $\Delta V$ as small as possible
- large ($C$)
- small ($I_{av}$)

Idea: Use Sense Amplifer



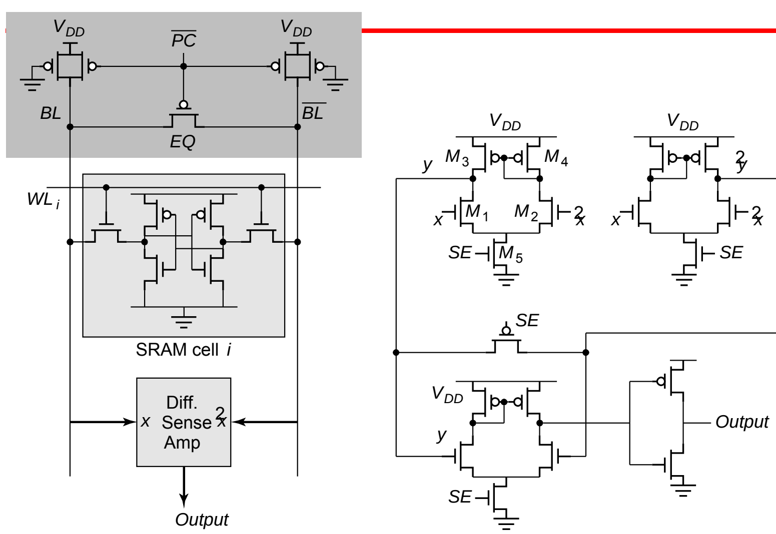
# Differential Sense Amplifier




Directly applicable to SRAMs



# Differential Sensing — SRAM




(a) SRAM sensing scheme (b) two stage differential amplifier