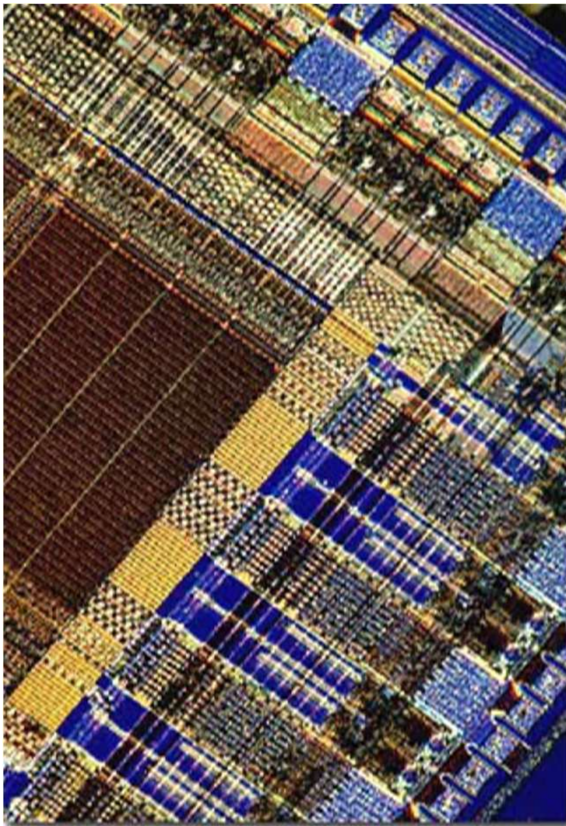
The McGraw-Hill Companies

CMOS Digital Integrated Circuits



Chapter 10
Semiconductor Memories

S.M. Kang, Y. Leblebici, and
C. Kim

Copyright © 2014 McGraw-Hill Education. Permission required for reproduction or display.

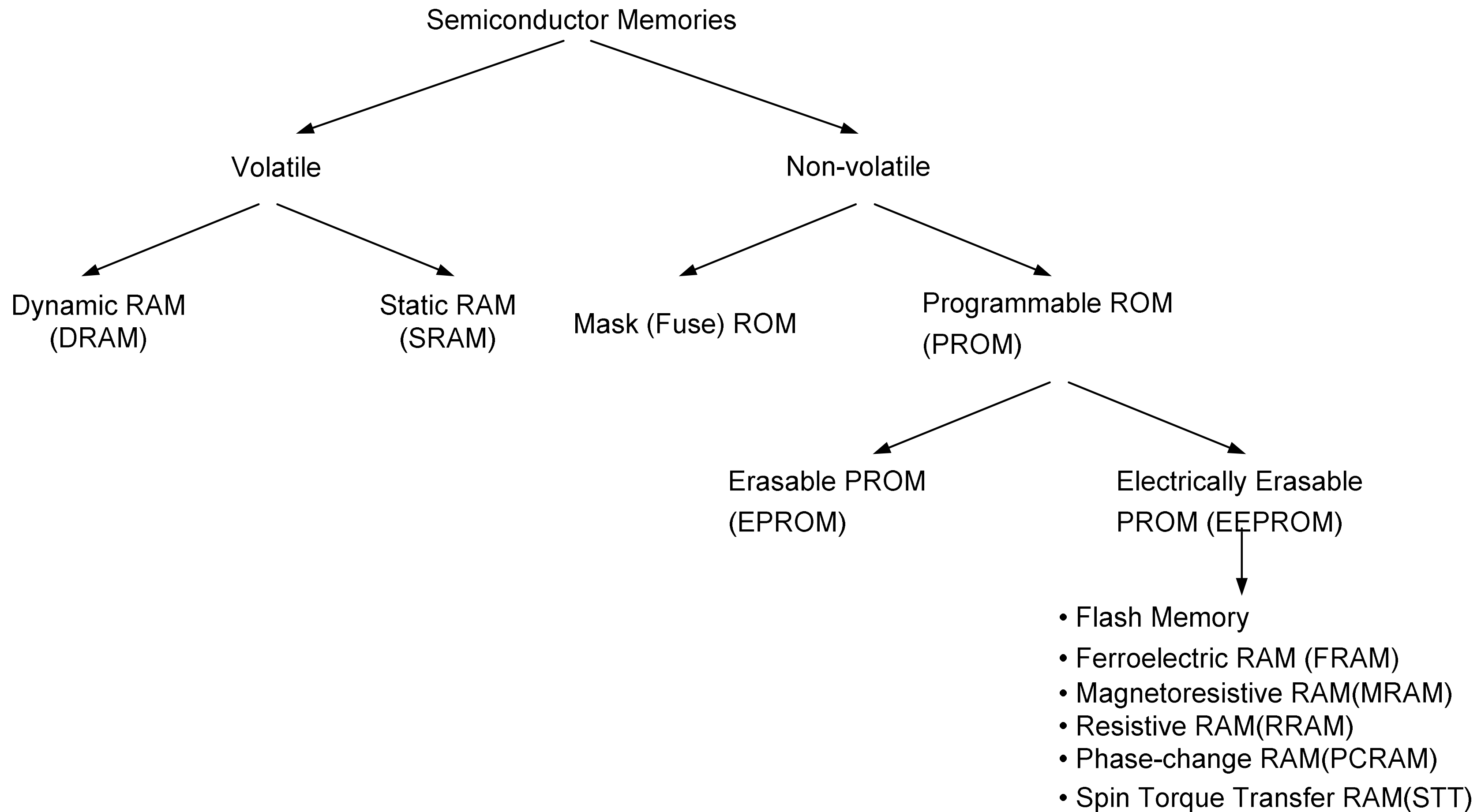
10.1 Introduction

- Design issue
  - Area efficiency → cost per bit
  - Access time → speed
  - Power consumption → low-power

© CMOS Digital Integrated Circuits – 4th Edition

10.1 Introduction

◆ Semiconductor memory types

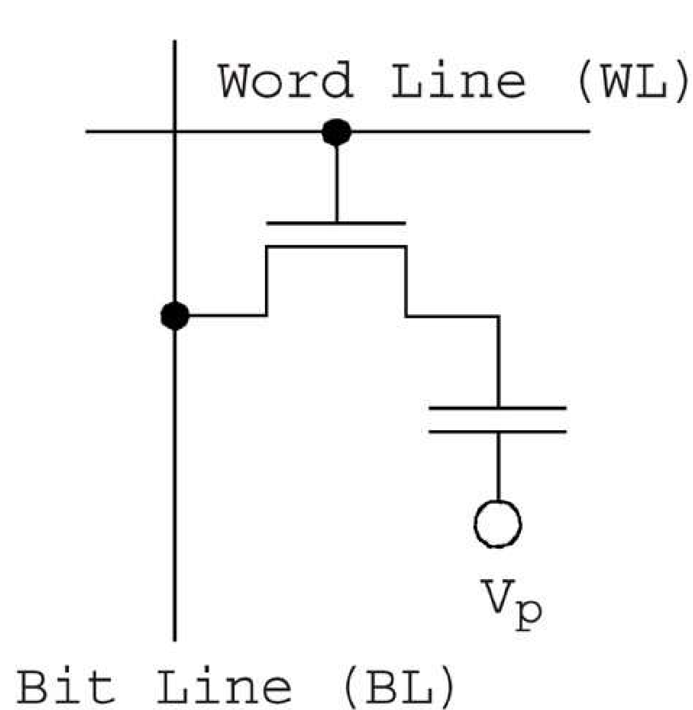


Characteristic Summary of Memory Devices

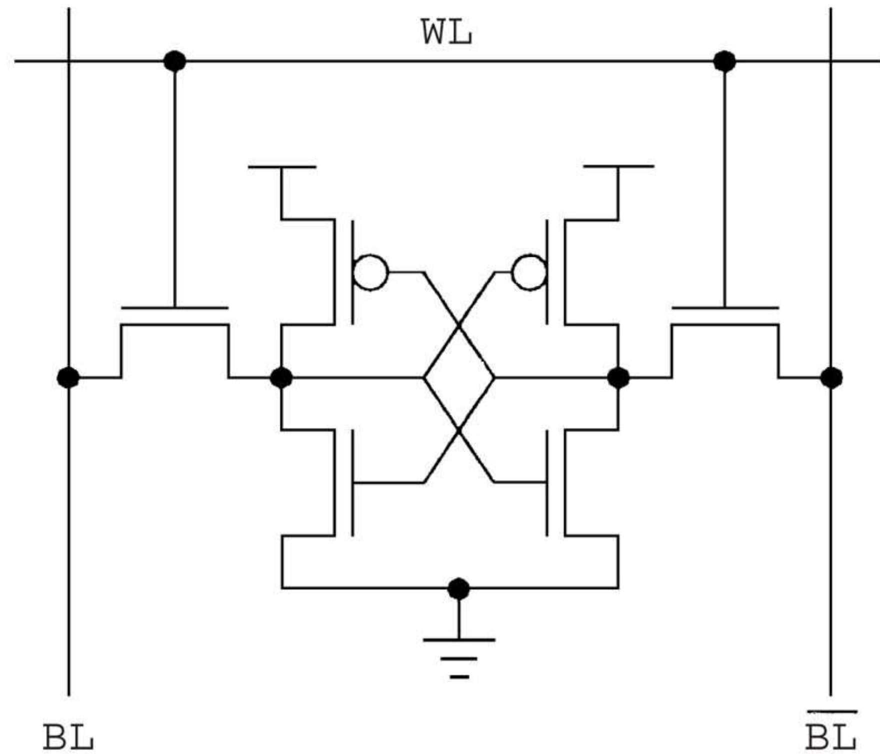
	Memory type					
	DRAM	SRAM	UV EPROM	EEPROM	Flash	FRAM
Data volatility	Yes	Yes	No	No	No	No
Data refresh operation	Required	No	No	No	No	No
Cell structure	1T-1C	6T	1T	2T	1T	1T-1C
Cell size(F^2) (F: min. feature size)	6~8	80~100			4~5(NAND) 9~10(NOR)	
Cell density	High	Low	High	Low	High	High
Power consumption	High	High/low	Low	Low	Low	High
Read speed (latency)	~50 ns	~10/70 ns	~50 ns	~50 ns	~50 ns	~100 ns
Write speed	~40 ns	~5/40 ns	~10 µs	~5 ms	~(10 µs-1 ms)	~100 ns
Endurance	High	High	High	Low	High	High
Cost	Low	High	Low	High	Low	Low
In-system writability	Yes	Yes	No	Yes	Yes	Yes
Power supply	Single	Single	Single	Multiple	Single	Single
Application example	Main memory	Cache/PDAs	Game machines	ID card	Memory card solid-state disk	Smart card, digital camera

© CMOS Digital Integrated Circuits – 4th Edition

Equivalent Circuits of Memory Cells(1)



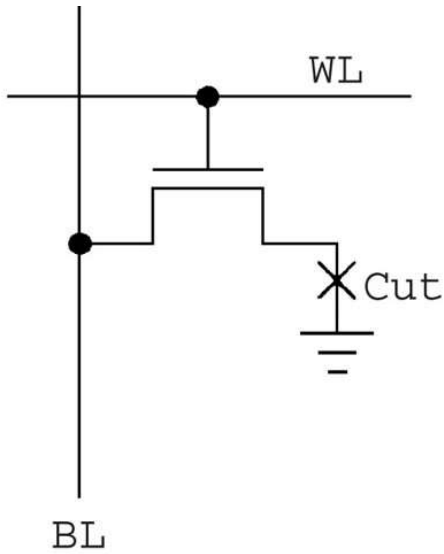
(a) DRAM

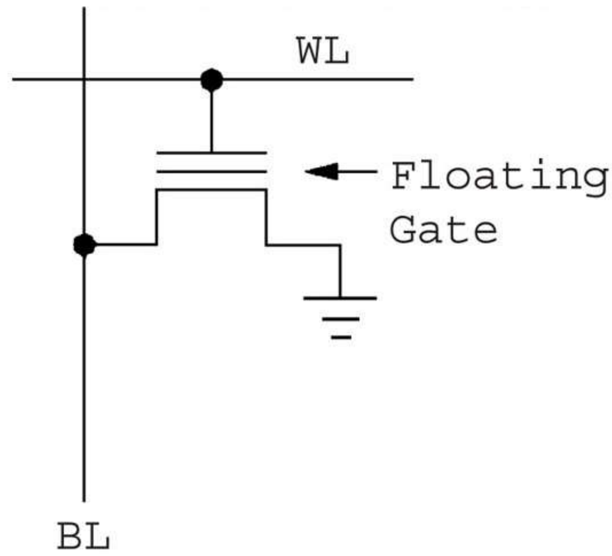


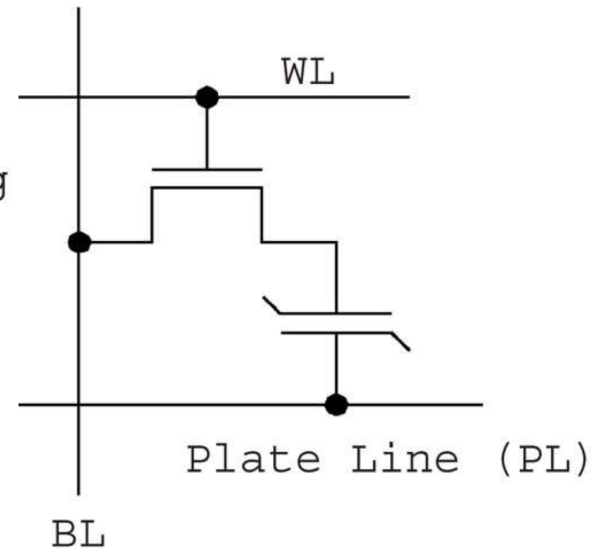
(b) SRAM

© CMOS Digital Integrated Circuits – 4th Edition

Equivalent Circuits of Memory Cells(2)







(c) Mask ROM (d) EPROM (e) FRAM

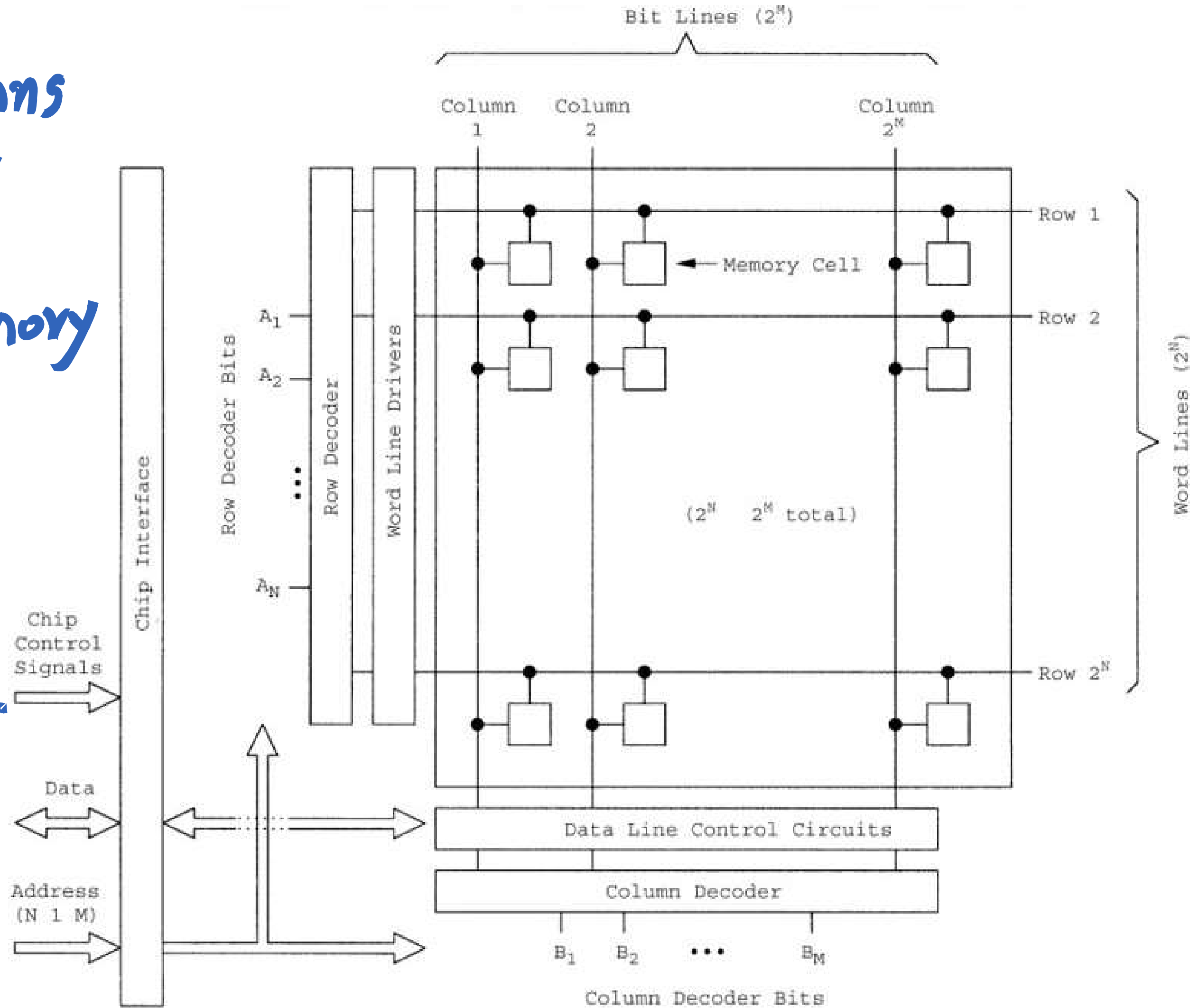
© CMOS Digital Integrated Circuits – 4th Edition

Conceptual RAM Array Organization

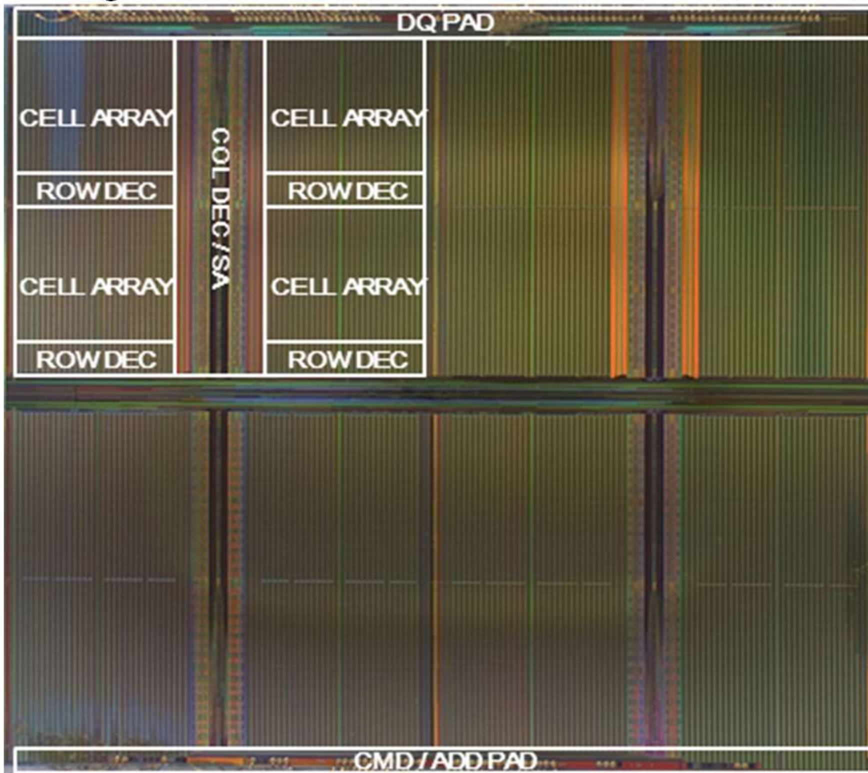
M columns
N rows

1 Kb memory
= 1024b
= 2^{10}

M = 5 6 4
N = 5 4 6



10.2 Dynamic Random Access Memory

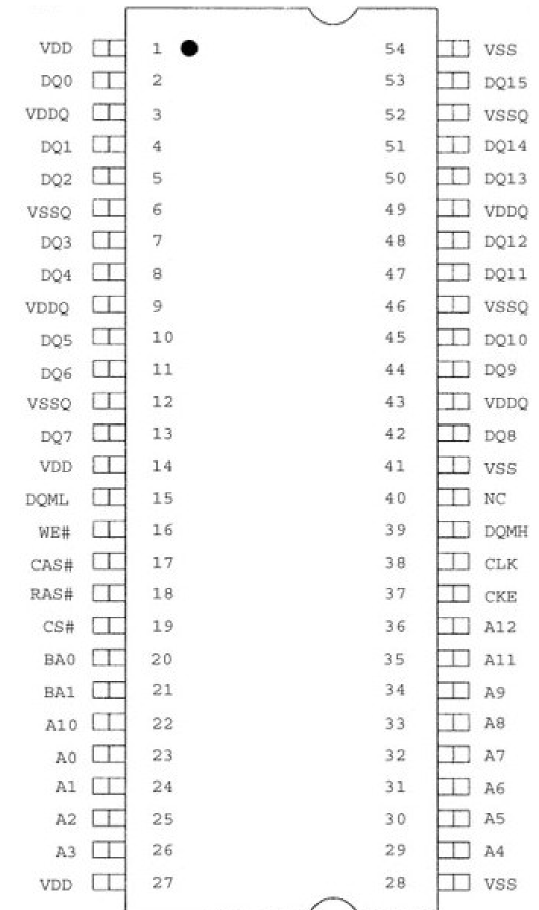


Typical configuration of DRAM chip

(1.6Gbps 4Gb 30nm LPDDR3 w/ 8 banks)

chip size ⟷ performance

→ The number of cells per word and bit lines



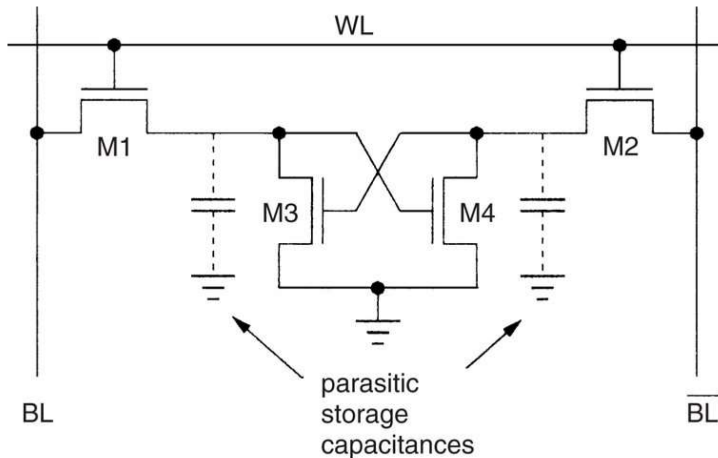
Pin assignment

© CMOS Digital Integrated Circuits – 4th Edition

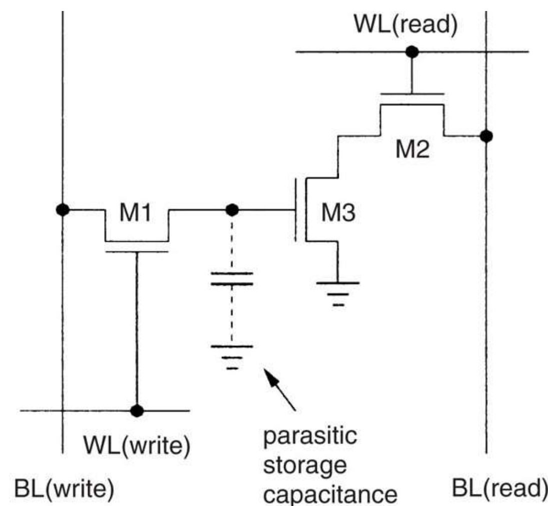
Definition and Function of DRAM Pins

Pin name	Definition	Function
CLK	Clock input	Reference system clock for the operation and data communication
CKE	Clock Enable	Control the clock input
$\overline{CS}$	Chip Select	Activate the DRAM device from a memory cluster
$\overline{RAS}$	Row Address Strobe	Latch row address and start the cell core operation
$\overline{CAS}$	Column Address Strobe	Latch column address and start the data communication operation
$\overline{WE}$	Write Enable	Activate the write operation
A0 to A14	Address input	Select a data bit
DQ0 to DQ15	Data input and output	Communicate data with external devices
DQMU/DQML	DQ Mask for Upper (Lower) Byte	Mask byte data from the operations
V_{DD}/V_{SS}	Power pins	Power for DRAM core and peripheral circuits
$V_{DD}Q/V_{SS}Q$	Power pins	Power for DQ circuits
NC	No connection	

Historical Evolution of DRAM Cell(1)

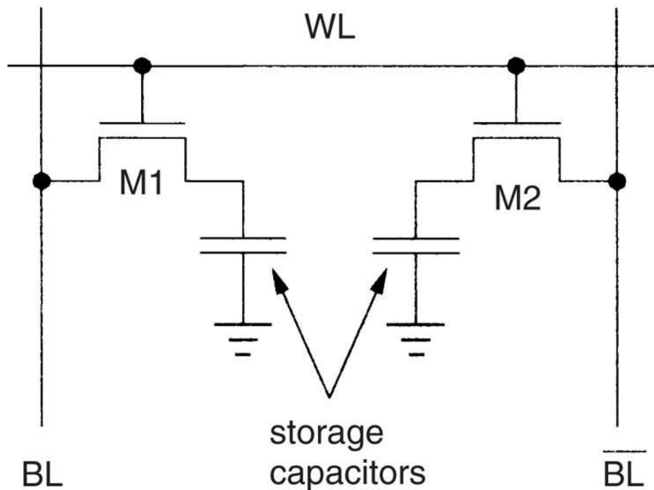


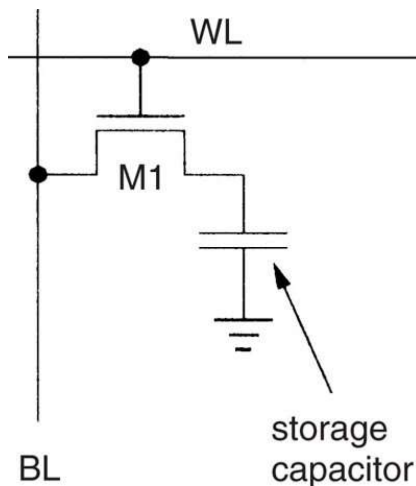
- Four-transistor DRAM cell
  - Operations are similar to SRAM cell
  - Two storage nodes
  - Periodically refresh is required
  - Non-destructive read operation



- Three-transistor DRAM cell
  - One storage node
  - One Tr. each for "read" and "write"
  - Non-destructive read operation
  - Two bit lines and two word lines (additional contacts ➡ increase area)

Historical Evolution of DRAM Cell(2)





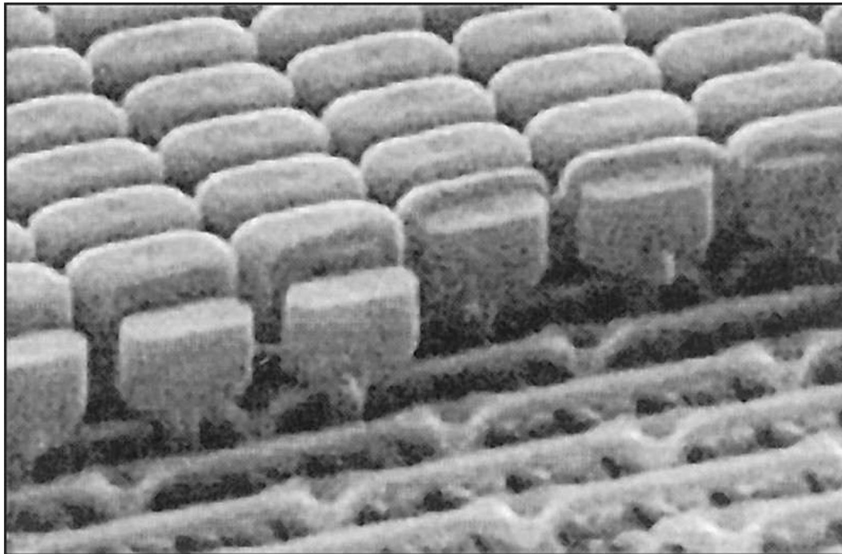
- Two-transistor DRAM cell
  - Explicit storage cap.
  - Destructive read operation (share with the bit line)
  - Two bit lines and one word line
- One-transistor DRAM cell
  - Industry-standard DRAM cell
  - Destructive read operation (share with the bit line)
  - Charge restoring operation required

© CMOS Digital Integrated Circuits – 4th Edition

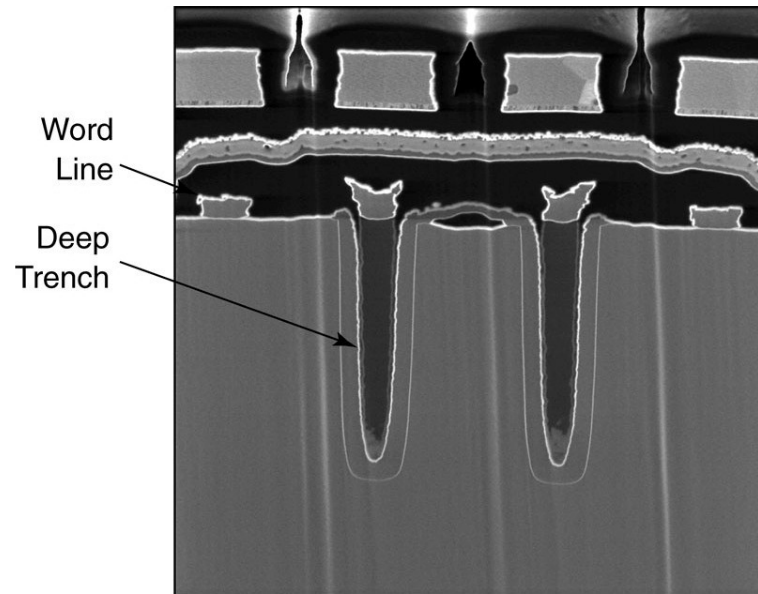
DRAM Cell Types

- With only one transistor and one capacitor
  - Smallest area of the all DRAM cells.
  - Destructive "read" operation

    → major effort : large cap. cell with minimized area



(a) DRAM cell with a stacked cap.

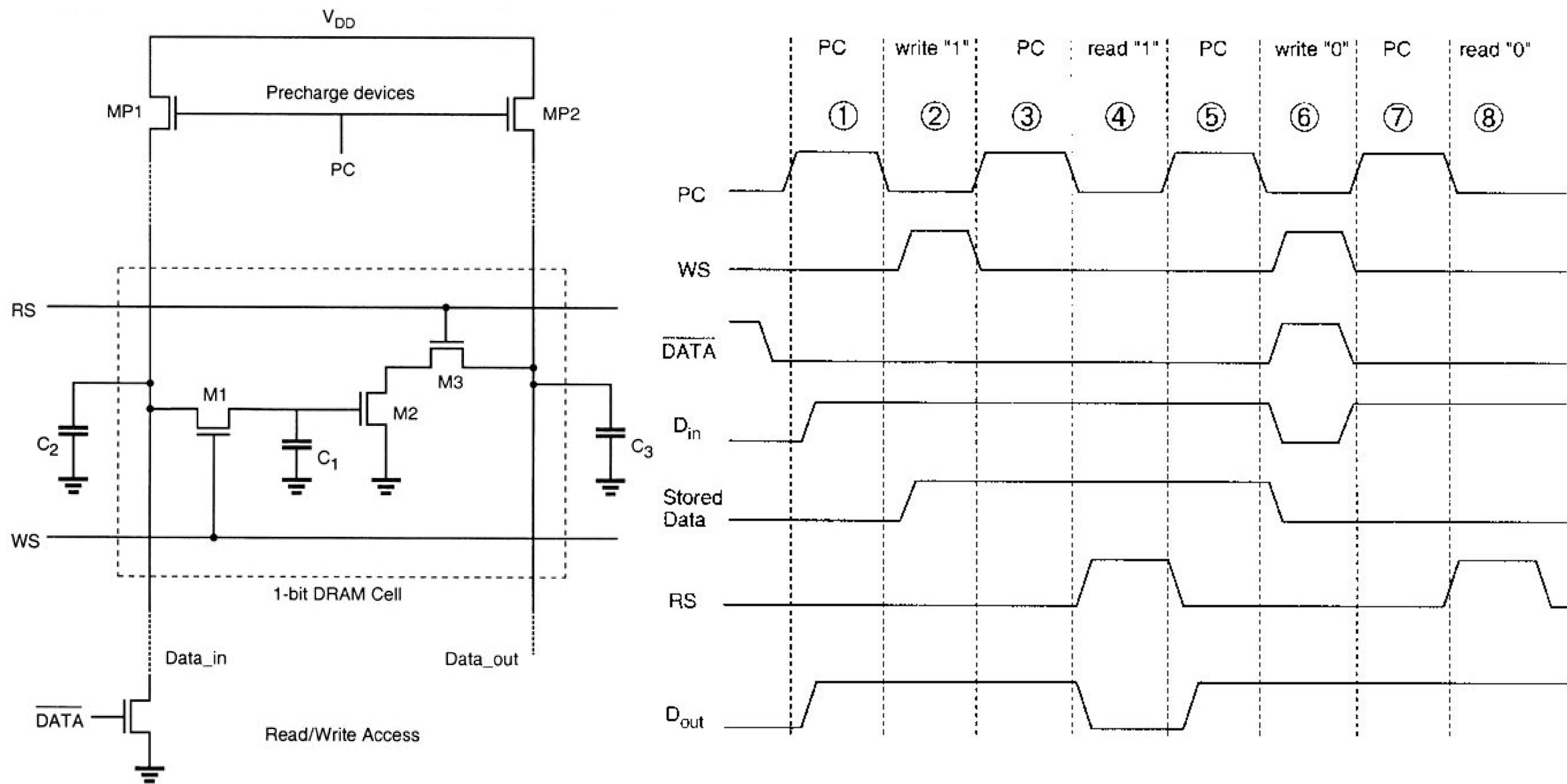


(b) DRAM cell with a trench cap.

© CMOS Digital Integrated Circuits – 4th Edition

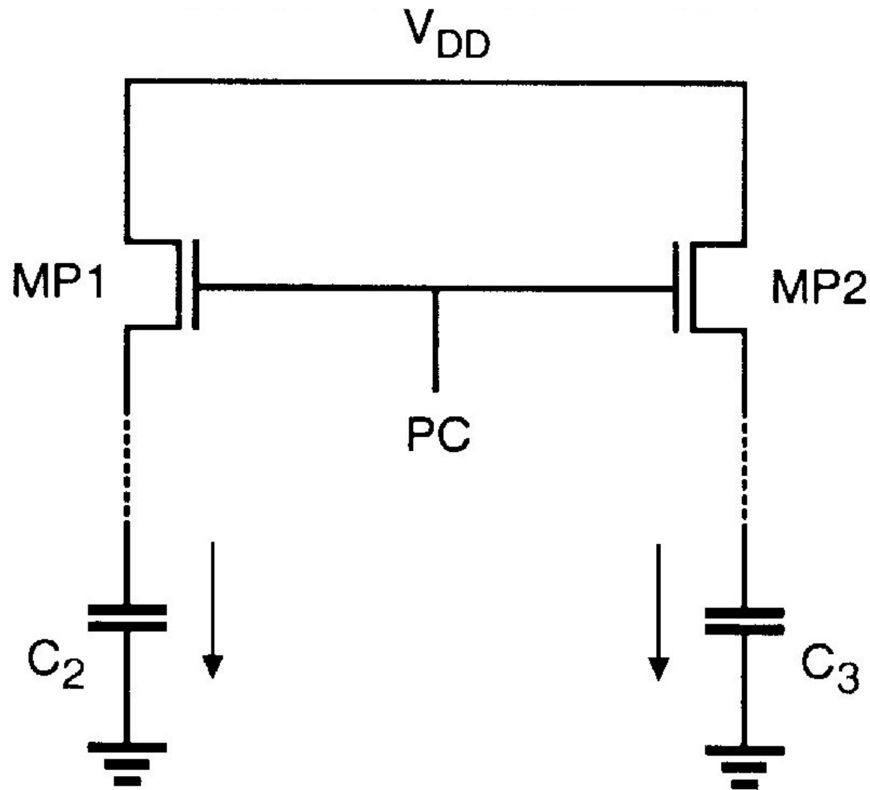
Operation of Three-Transistor DRAM Cell

- Typical 3-T DRAM cell and voltage waveforms



© CMOS Digital Integrated Circuits – 4th Edition

Precharge Events



- Base on two-phase non-overlapping clock scheme
  - Φ_1 : precharge phase
  - Φ_2 : active phase

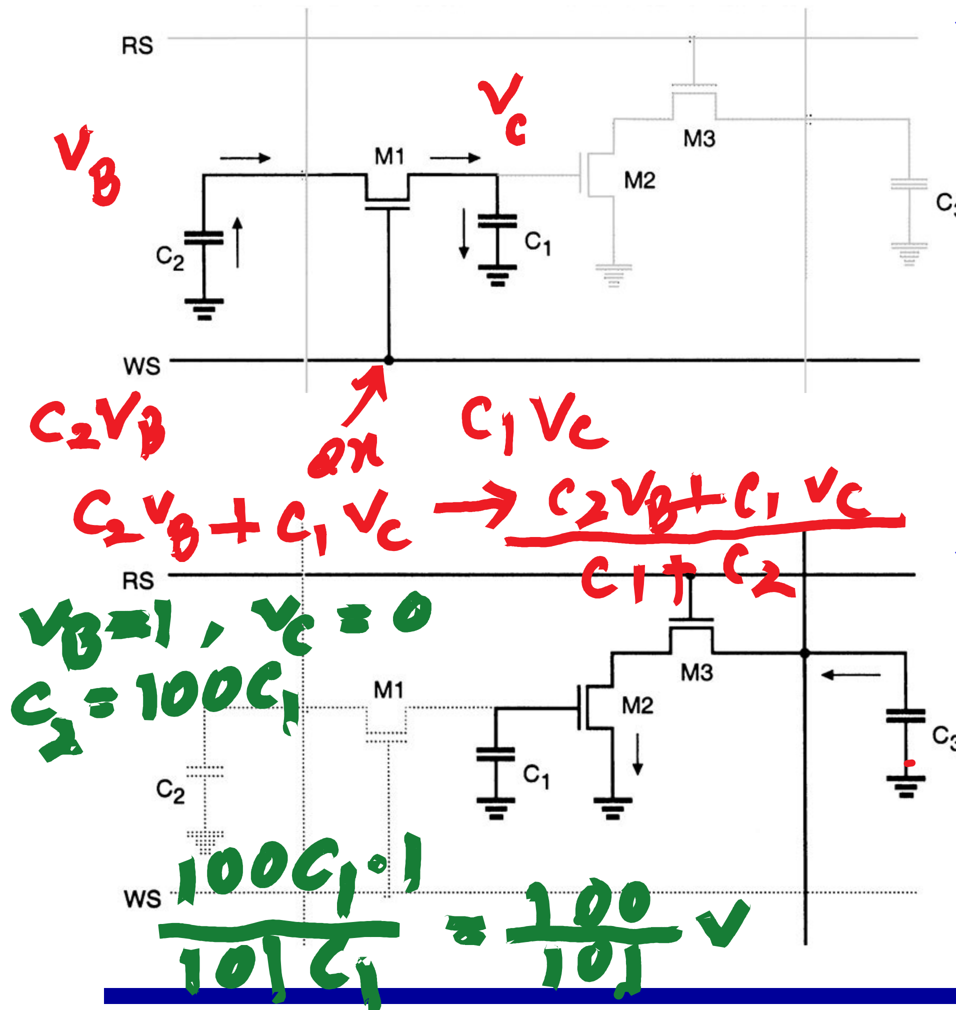
Precharge signal PC goes up

MP1 and MP2 are activated

C_2 and C_3 are charged up (Steady-state values)

© CMOS Digital Integrated Circuits – 4th Edition

Write “1” and Read “1” Operations



Write “1”

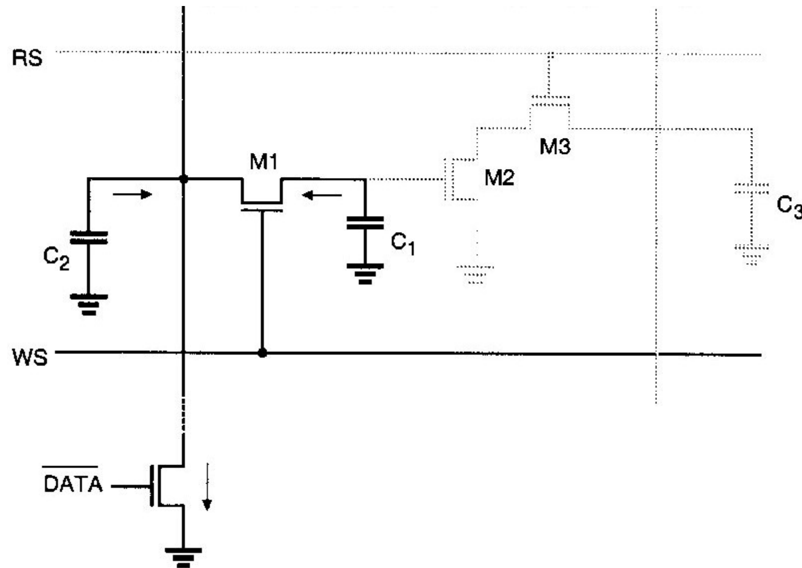
- $\overline{\text{DATA}}$ is low ➔ D_{in} remains high
- M1 turns on (WS is high)
  ➔ C_2 shared with C_1
- C_1 charge up to high (M2 is conducting)

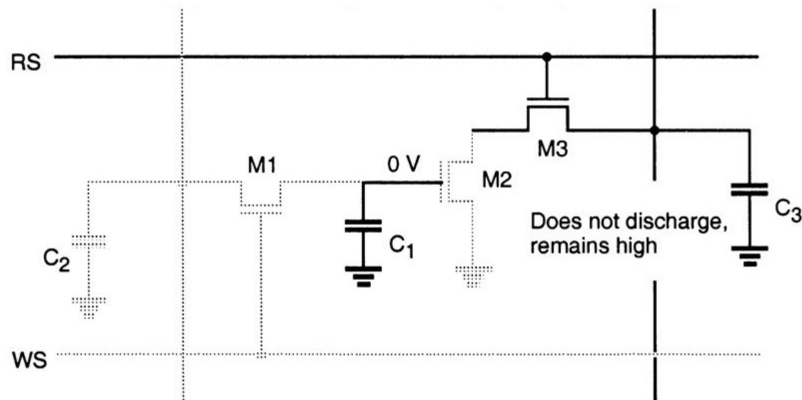
Read “1”

- RS is high ➔ M3 turns on
- M2 and M3 create conducting path from C_3 to GND
- C_3 discharges
- Non-destructive read operation

© CMOS Digital Integrated Circuits – 4th Edition

Write "0" and Read "0" Operations





Write "0"

- $\overline{\text{DATA}}$ is high → D_{in} goes low
- M1 turns on (WS is high)
- C_1 discharges (M2 turns off)

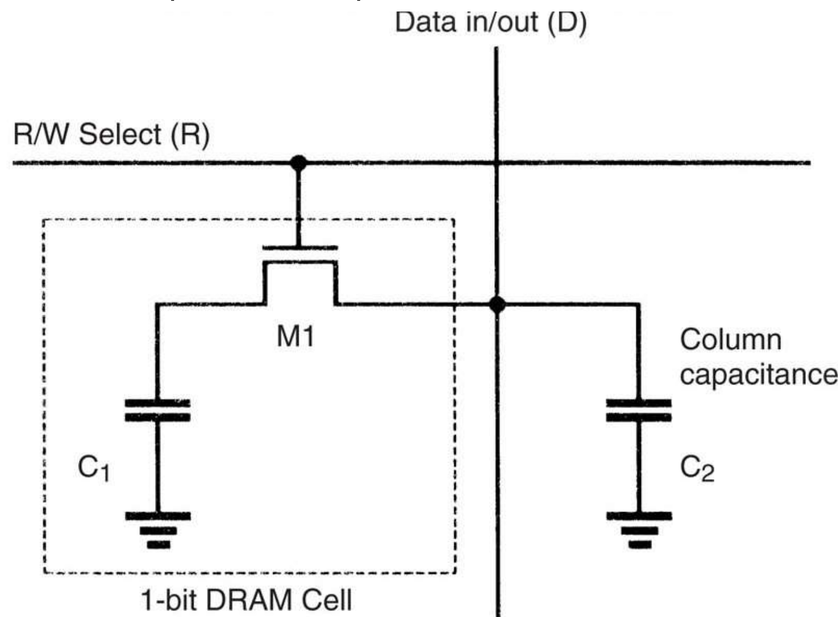
Read "0"

- RS is high → M3 turns on
- No conducting path
- C_3 does not discharge

© CMOS Digital Integrated Circuits – 4th Edition

Operation of One-Transistor DRAM Cell

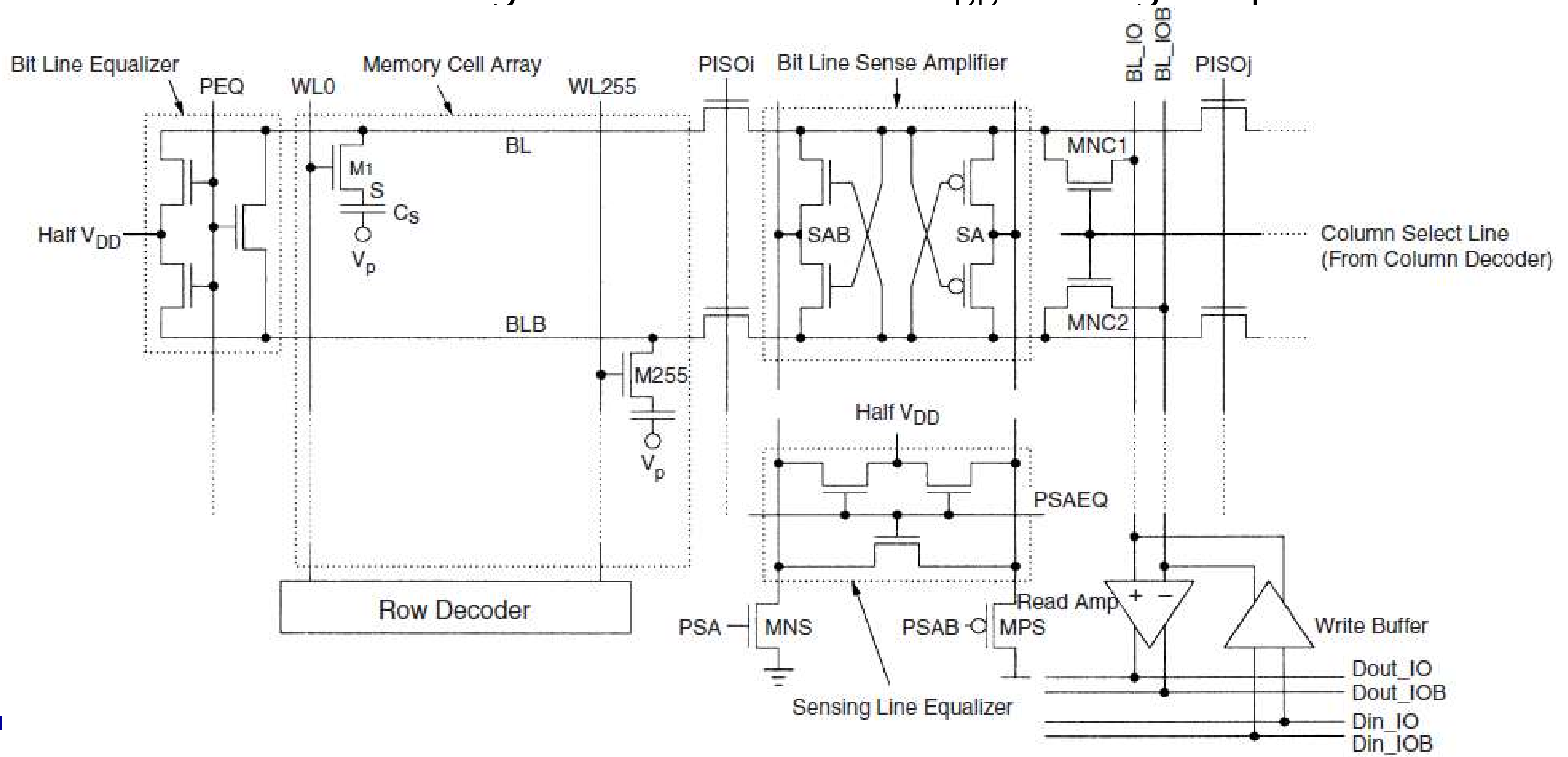
- One explicit storage cap. and one access transistor
  - The most widely used storage structure
  - Bit lines are folded and precharged to half-V_{DD}
    ➔ improve noise-immunity & reduce power consumption
  - Operation : "read", "write", "refresh"



© CMOS Digital Integrated Circuits – 4th Edition

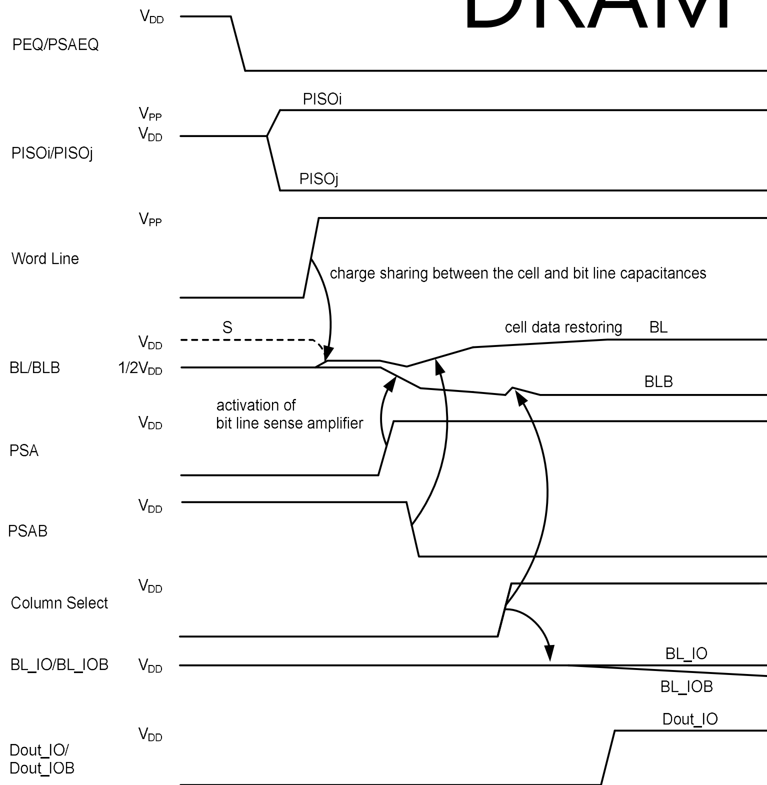
1-T DRAM Structure

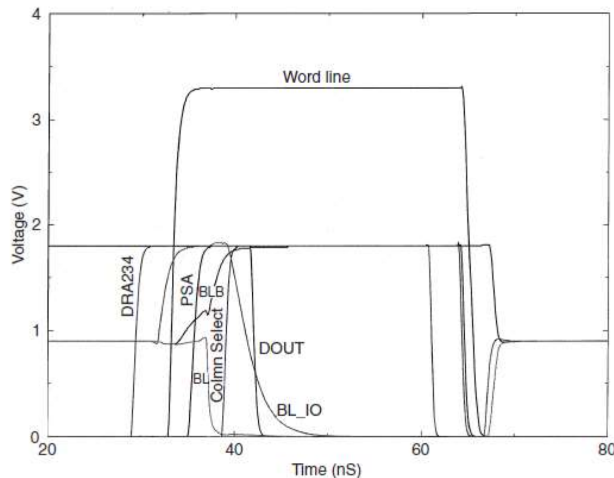
- DRAM cell array with control circuits
- Latch amplifier to sense the small signal difference
- Bit lines and sensing nodes set to half-V_{DD} through equalizer



© CMOS Digital Integrated Circuits – 4th Edition

DRAM Read Operation



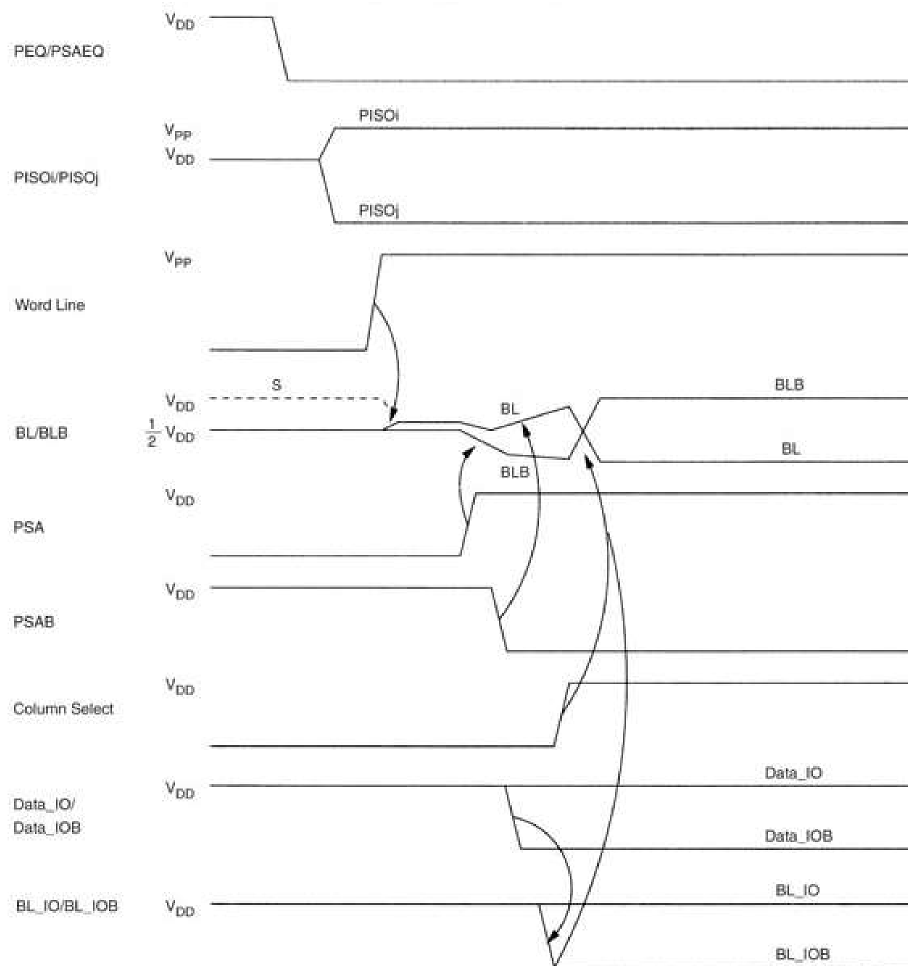


- C_S shared with C_{BL}(=initially half V_{DD})

$$\Delta V = \frac{C_S}{C_{BL} + C_S} \frac{V_{DD}}{2}$$

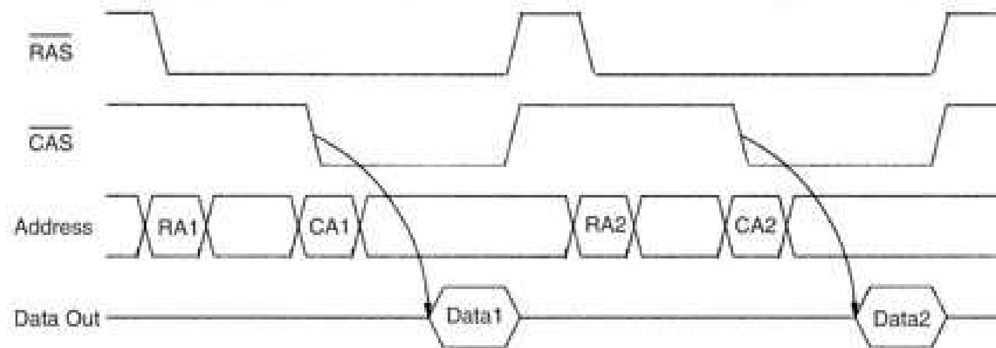
- C_S : V_{DD} ➔ ½V_{DD}+ΔV (destructive)
- BL and BLB voltage difference amplified
- BLB ➔ GND, BL ➔ V_{DD}
  storage node is recovered (restoring)
- Column switch is enabled by column decoder (BL ➔ BL_IO, BLB ➔ BL_IOB)
- Read Amp. amplifies the voltage difference
  - $V_{PP}=V_{DD}+V_{th}$ for full charge restoration
  - PSA and PSAB are sequentially activated to reduce charge injection and short circuit current

DRAM Write Operation



- Identical sequence to normal read operation
- Strong write driver (buffer) to drive BL_IO, BL_IOB line cap.
  - ➡ faster than read operation
- Column switch is selected by column decoder
- Bit line and cell data changed

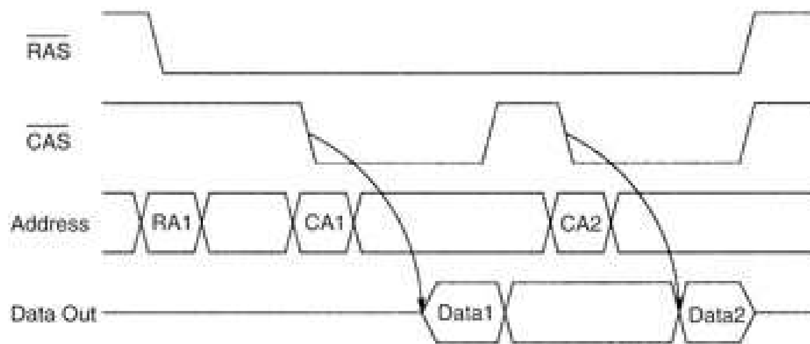
Asynchronous DRAM Mode(1)



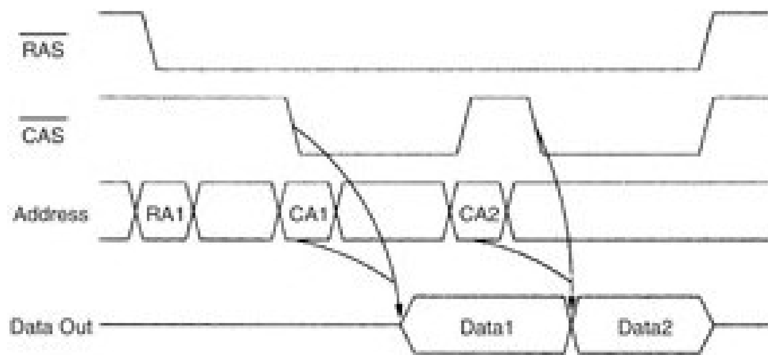
- single bit access (different row and column addresses)
- Operation uses address multiplexing scheme ($\overline{RAS}$ and $\overline{CAS}$)
  - ➡ reduce the chip package size
- $\overline{RAS}$ pull down ➡ operation start
- Falling edge of $\overline{CAS}$ ➡ data (from same word line) selected
- $\overline{RAS}$, $\overline{CAS}$ precharge before new data access
- t_{RAC} : memory read latency, time to read data from falling of $\overline{RAS}$
- Length of word line is determined by **refresh** cycle constraint

© CMOS Digital Integrated Circuits – 4th Edition

Asynchronous DRAM Mode(2)

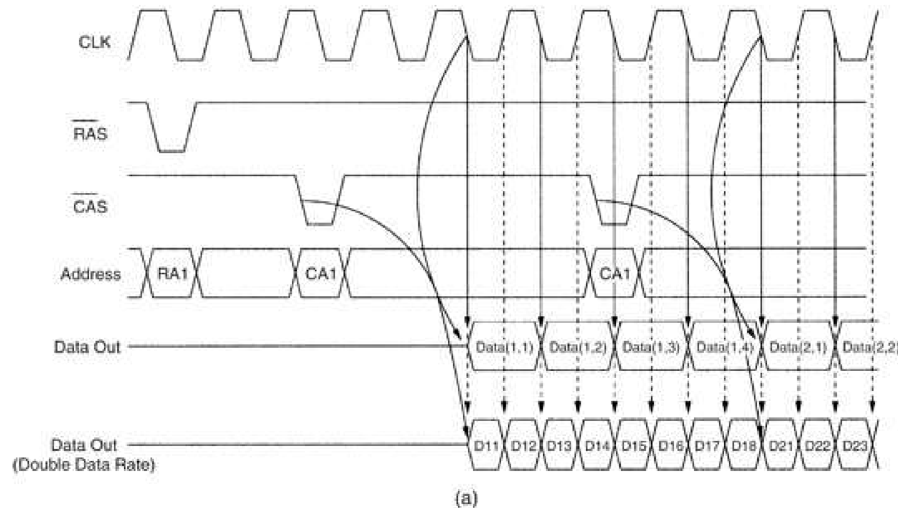


- page access
  - keep the row address
  - read cell of same row address
  - faster read operation

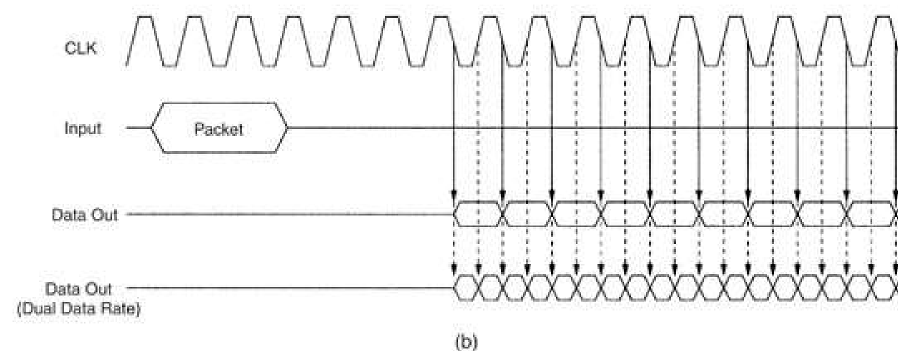


- extened data-out(EDO)
  - new column address is captured at rising edge of CAS
  - Read data maintain during precharge time
  - Fastest read opertaion

Synchronous DRAM Mode



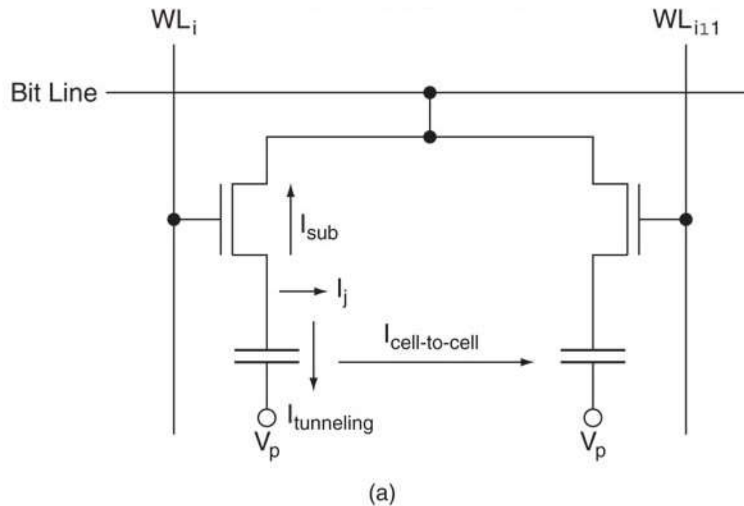
(a)



(b)

- Four bit burst read
  - Read frequency improve with use of the system clock
  - At falling edge, control signal and addresses become active
  - Pipelined based on clock to improve throughput
  - Use both of edges to improve bandwidth (Dual Data Rate)
- Serial mode read
  - Use small signal swing and clock recovery scheme to maximize the frequency
  - Send control input as packet
  - Send out data in a serial form

Leakage Currents in DRAM Cells



- Contact and bit line share by two adjacent cells

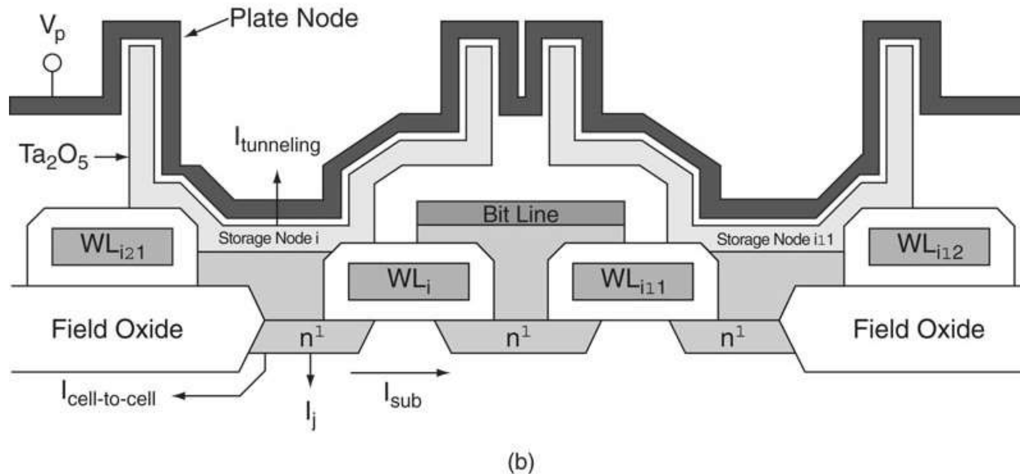
$$I_{leakage} = I_{sub} + I_{tunneling} + I_j + I_{cell-to-cell}$$

I_{sub}: leakage through cell transistor

$I_{tunneling}$: tunneling through thin dielectric

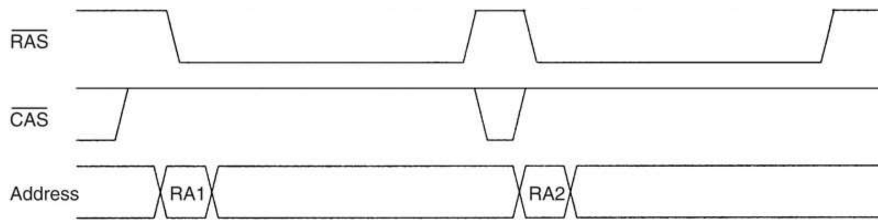
I_j: junction leakage at storage node

$I_{cell-to-cell}$: leakage across the field oxide

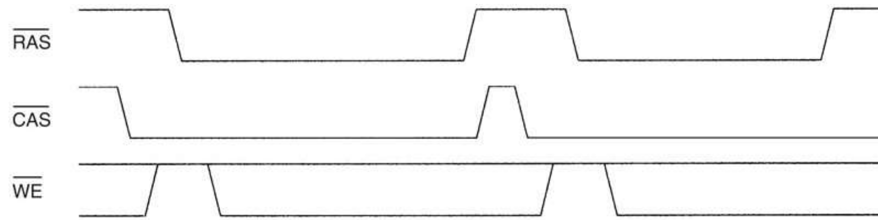


- I_{sub} depend on V_{th}
  increase V_{SB} to reduce I_{sub}
- $I_{tunneling}$ is a serious issue because thickness of dielectric is reduced to increase cell cap.

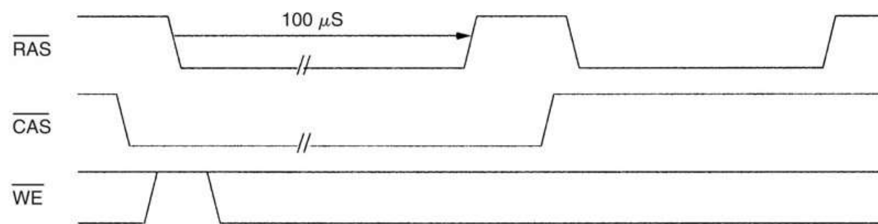
Refresh Operation



ROR refresh mode



CBR refresh mode



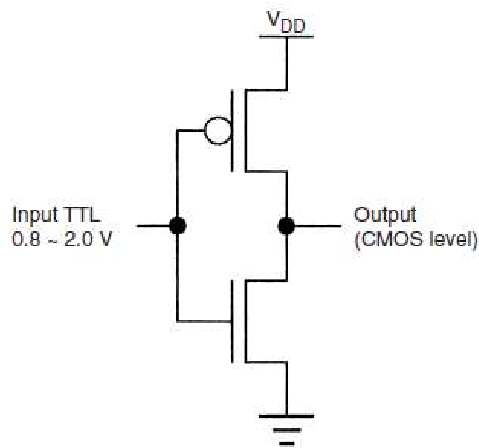
Self refresh mode

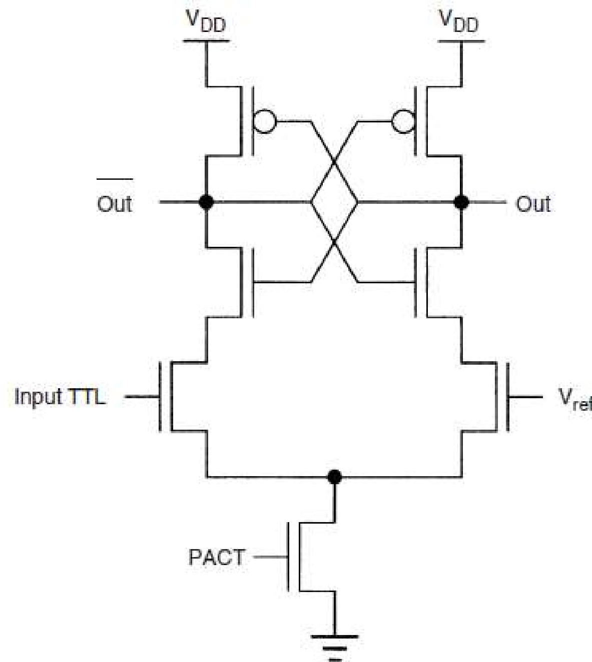
- ROR(RAS-only refresh) refresh
  - Read and restore operation
  - Does not send data out
  - similar to normal read operation
- CBR(CAS-before-RAS) refresh
  - row address generated by on-chip counter
  - performed periodically
- Self refresh
  - period set according to operating condition
  - row address and control signal generated by internal circuit

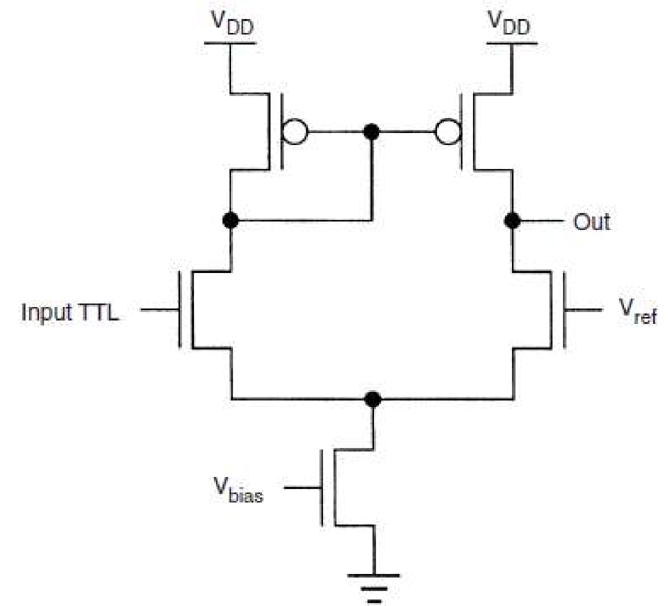
DRAM Input/Output Circuits(1)

- Logic level of system board and memory chip are different
  ➡ required to convert logic levels ➡ input/output buffers

- Input buffers







inverter type | latch type | differential amp type

© CMOS Digital Integrated Circuits – 4th Edition

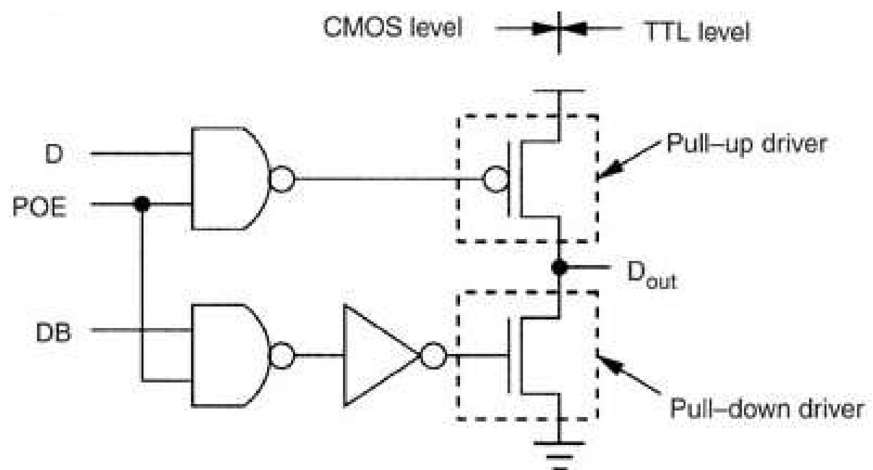
Characteristic Comparison of Input Buffers

	Buffer type		
	Inverter	Latch	Differential
Logic threshold determination (V_{IH} and V_{IL})	By W_P/W_N ratio	By V_{ref}	By V_{ref}
Speed	Slow	Fastest	Fast
Standby current	Small	Smallest	Large
Sensitivity to V_{DD} and temperature	Large	Small	Small
Noise immunity	Bad	Good	Good
Constraint	None	Precharge and activation signals needed	None

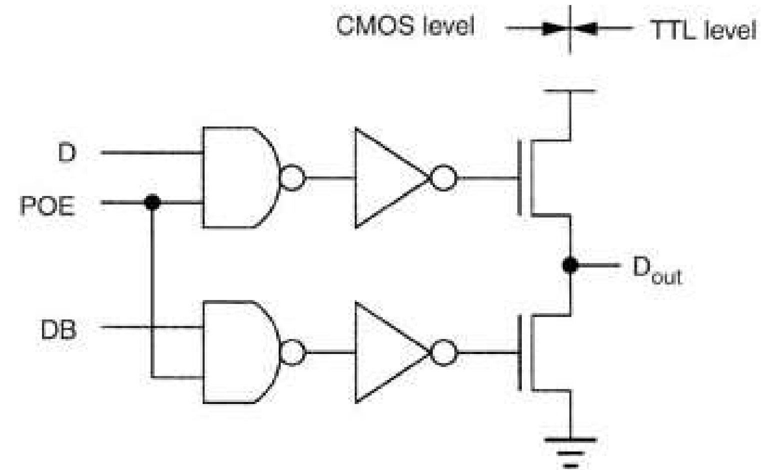
© CMOS Digital Integrated Circuits – 4th Edition

DRAM Input/Output Circuits(2)

- Memory output buffers
  - Need to drive large cap.
  - Keep a high-impedance when chip is not selected
    - → to prevent interference of output



PMOS pull-up structure

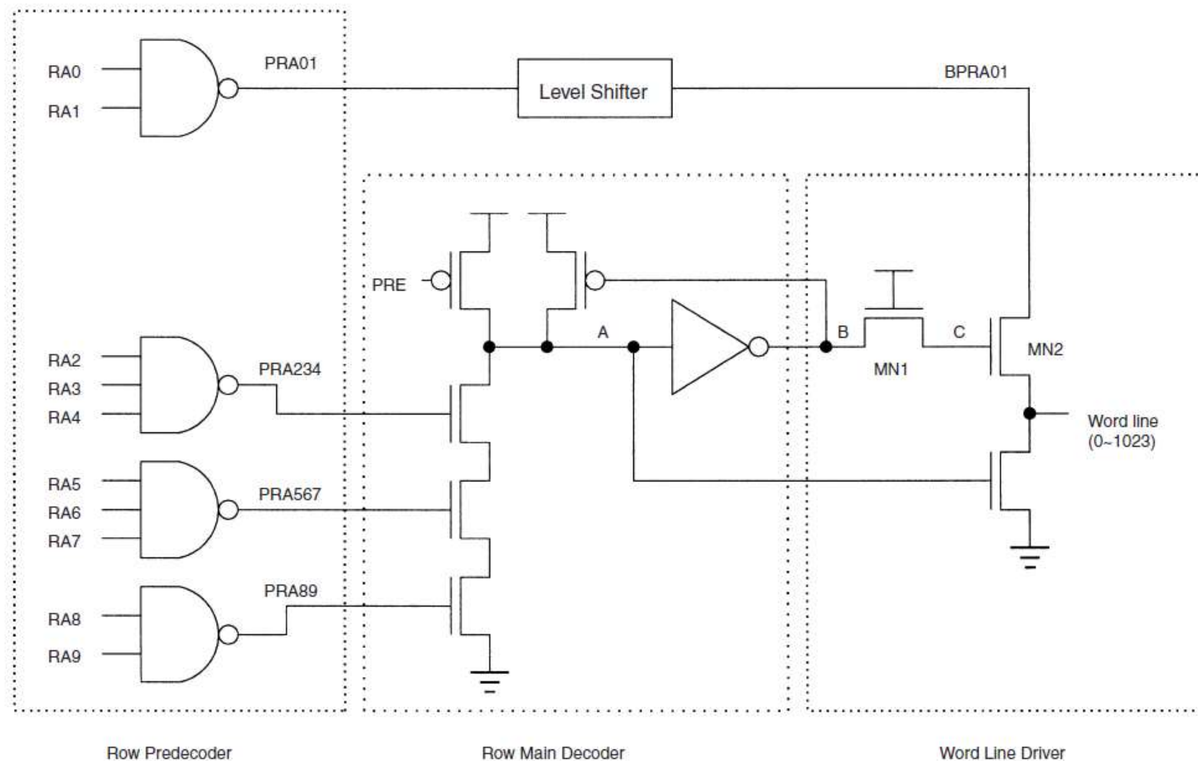


NMOS pull-up structure

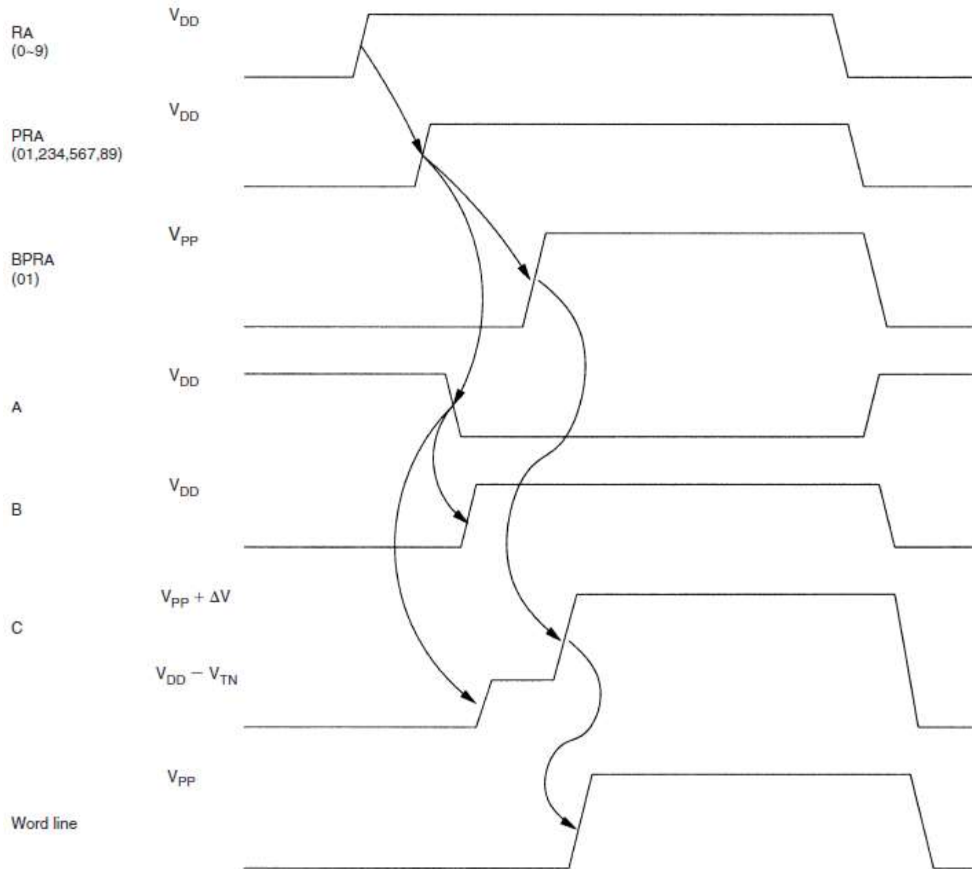
© CMOS Digital Integrated Circuits – 4th Edition

DRAM Decoder(1)

- To select cell from 2^{2M} memory array, M address bits are needed
- Practically, M transistors in series is impossible
  - ➡ decoding scheme is composed of pre and main decoder



DRAM Decoder(2)



- V_{PP} for full restoration
- Output of predecoder boosted by level shifter
- Self-bootstrapped driver for transferring to highly cap. without signal degradation

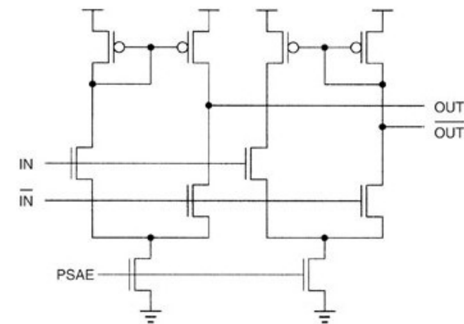
- Voltage of C when main decoder is selected

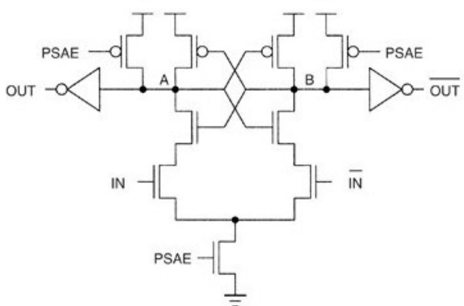
$$V_C = V_{PP} + \Delta V = V_{DD} - V_{TN} + \frac{C_{MN2}}{C_{MN2} + C_{Cparasitic}} V_{PP}$$

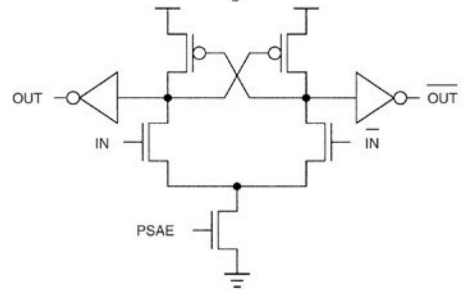
© CMOS Digital Integrated Circuits – 4th Edition

Voltage Sense Amplifiers

- To detect signal difference on data lines





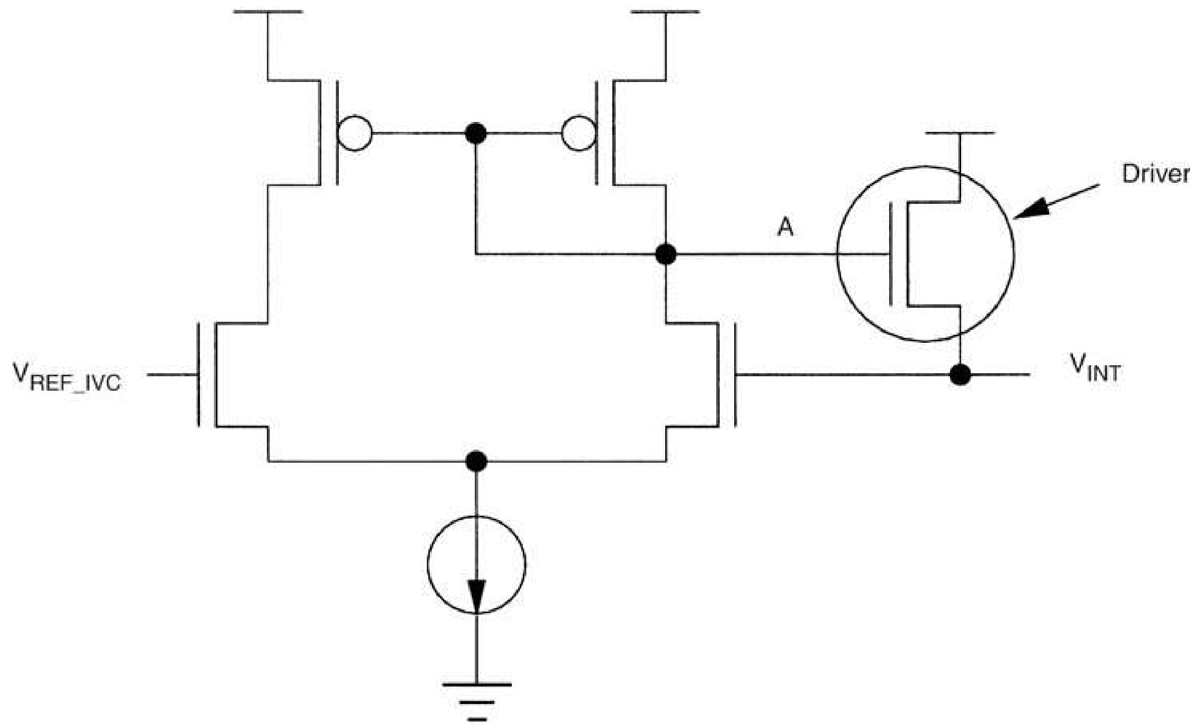


- Current-mirror differential
  - Popular and good common-mode rejection ratio
  - Large area and large power consumption
- Full CMOS latch type
  - High speed, small area and low power
  - Precharge signal required
  - operation cannot be reversed
- Semilatch type
  - Between current-mirror type and full CMOS latch type

© CMOS Digital Integrated Circuits – 4th Edition

Internal Voltage Regulator Circuit

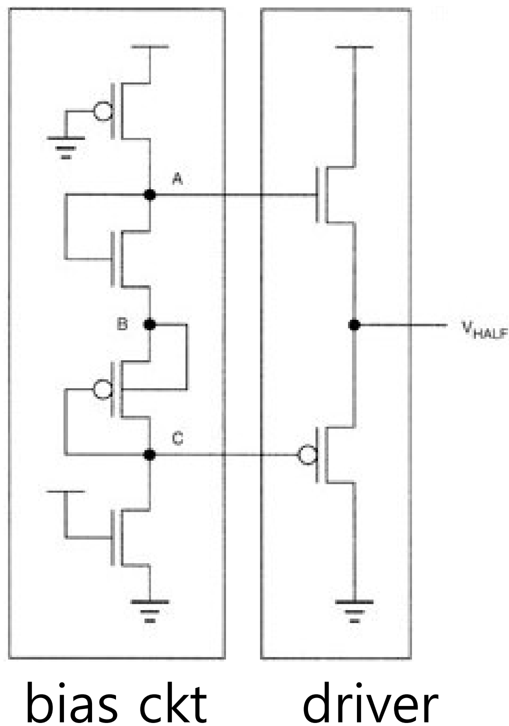
- Lowering voltage to reduce power consumption
- V_{INT}(internal voltage generator)
  - ➔ reduce operating current



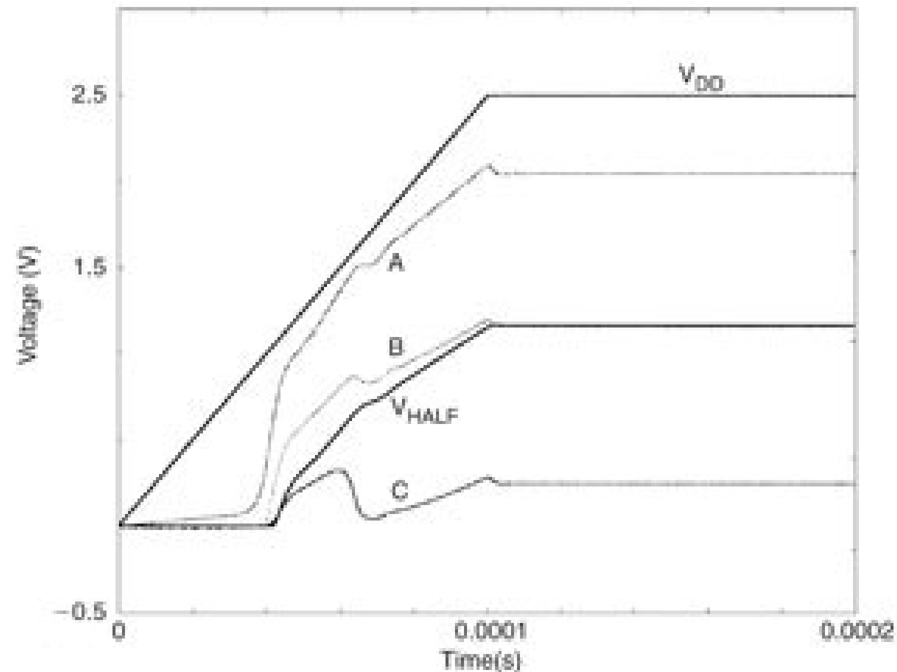
© CMOS Digital Integrated Circuits – 4th Edition

Half V_{DD} Voltage Generator

- Folded bit line structure with half V_{DD} sensing scheme
  - Improve noise immunity and low power consumption
  - Reduce electric field across thin dielectric



bias ckt driver

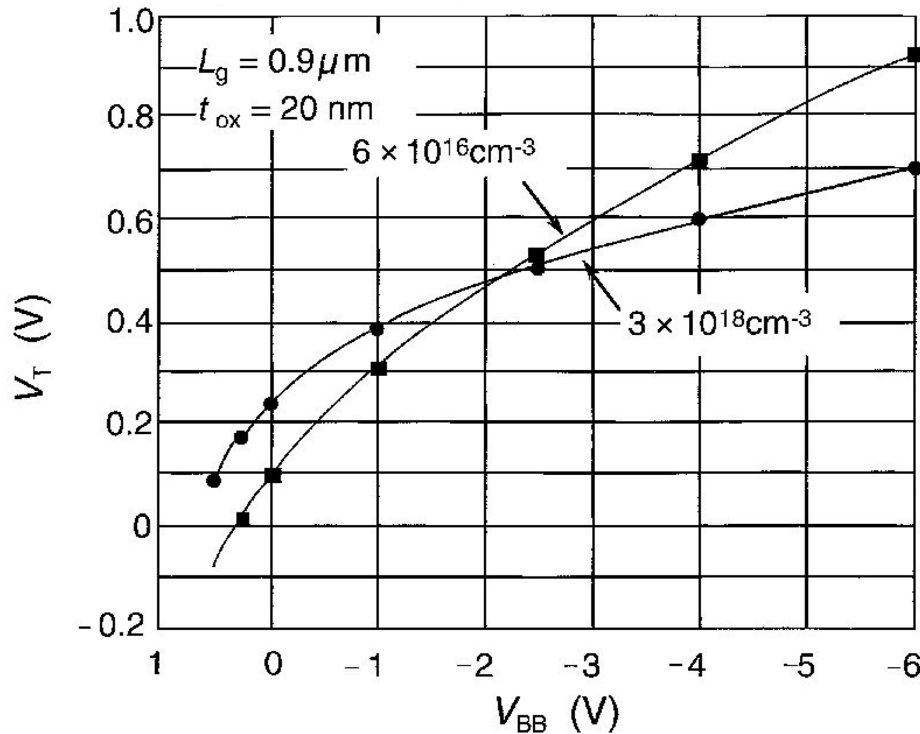


simulated output waveforms

© CMOS Digital Integrated Circuits – 4th Edition

Negative Substrate Bias Voltage Generator(1)

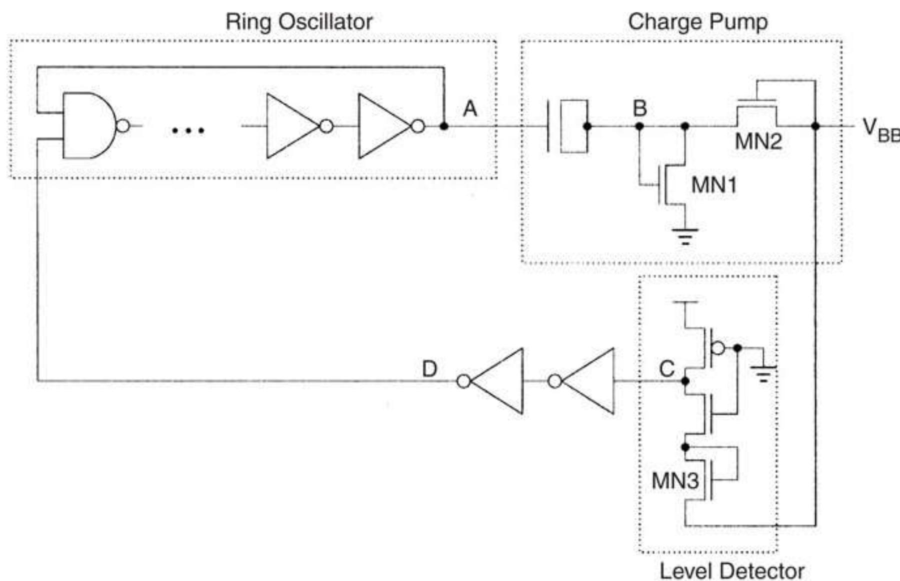
- Subthreshold current is major source of charge decay
  - Negative voltage substrate ➔ increase threshold voltage
    ➔ reduce load cap. of bit line



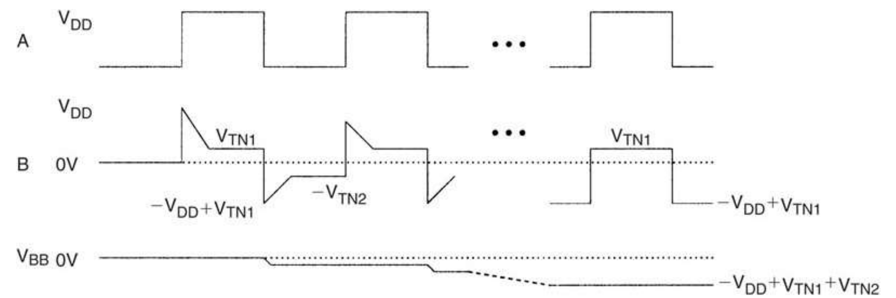
$$\frac{\Delta V_T}{\Delta V_{SB}} \propto \sqrt{V_{SB}}$$

© CMOS Digital Integrated Circuits – 4th Edition

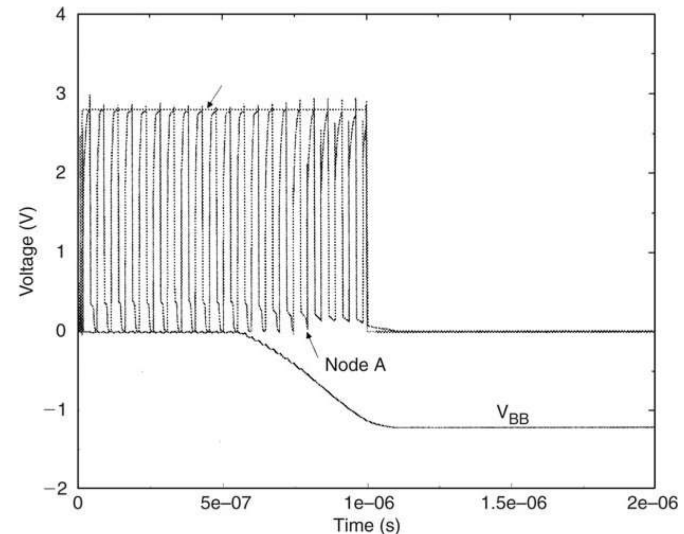
Negative Substrate Bias Voltage Generator(2)



Circuit diagram



Timing diagram

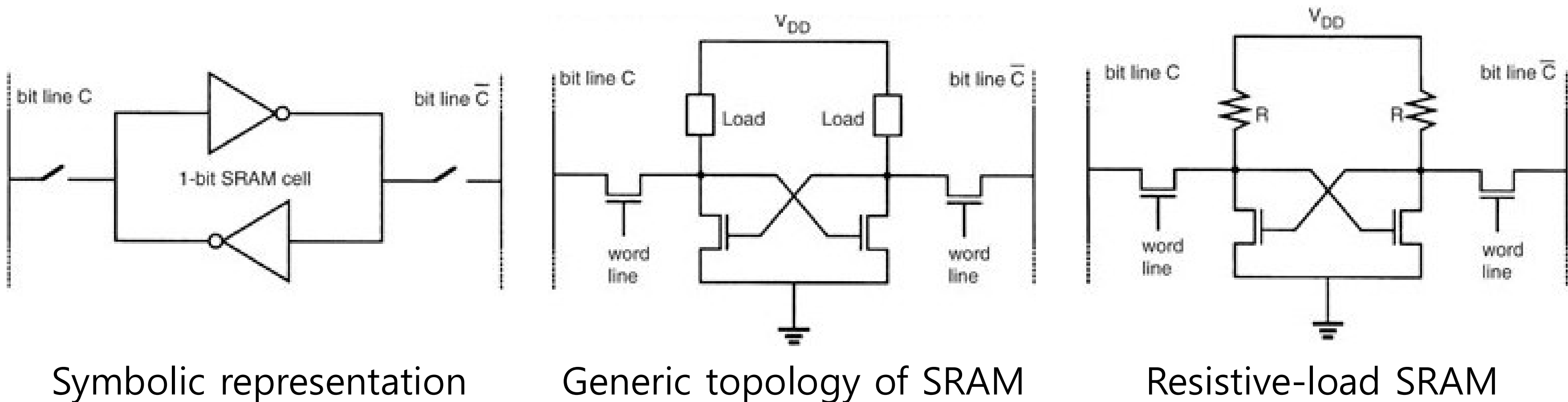


Simulated waveforms

© CMOS Digital Integrated Circuits – 4th Edition

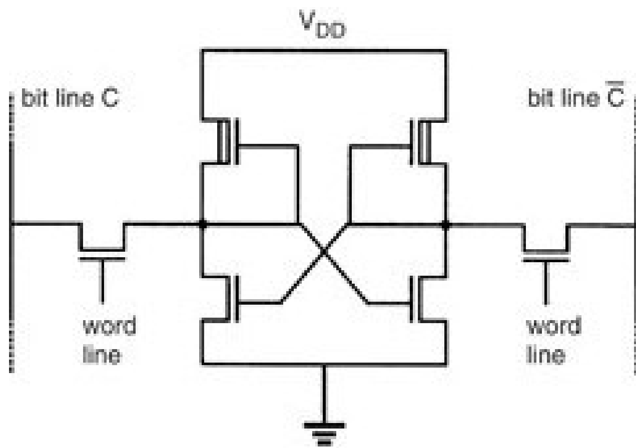
10.3 Static Random Access Memory(1)

- Stored data can be retained indefinitely
- Simple latch with two stable operating points
- Two access switches to connect 1-bit SRAM
- Poly resistor load inverter structure is more compact cell size (resistor stack on top of cell)
- Load R trade off : low power ⟷ wider noise margin, high speed

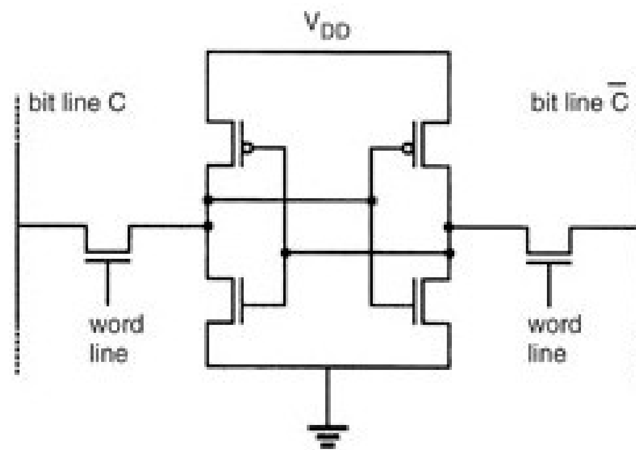


Symbolic representation　　Generic topology of SRAM　　Resistive-load SRAM

10.3 Static Random Access Memory(2)



- Depletion-load NMOS SRAM
  - Six-transistor
    (one poly and one metal layer)
  - Cell size relatively small
  - Static characteristics and noise margins better than resistive-load cell
  - Static power consumption

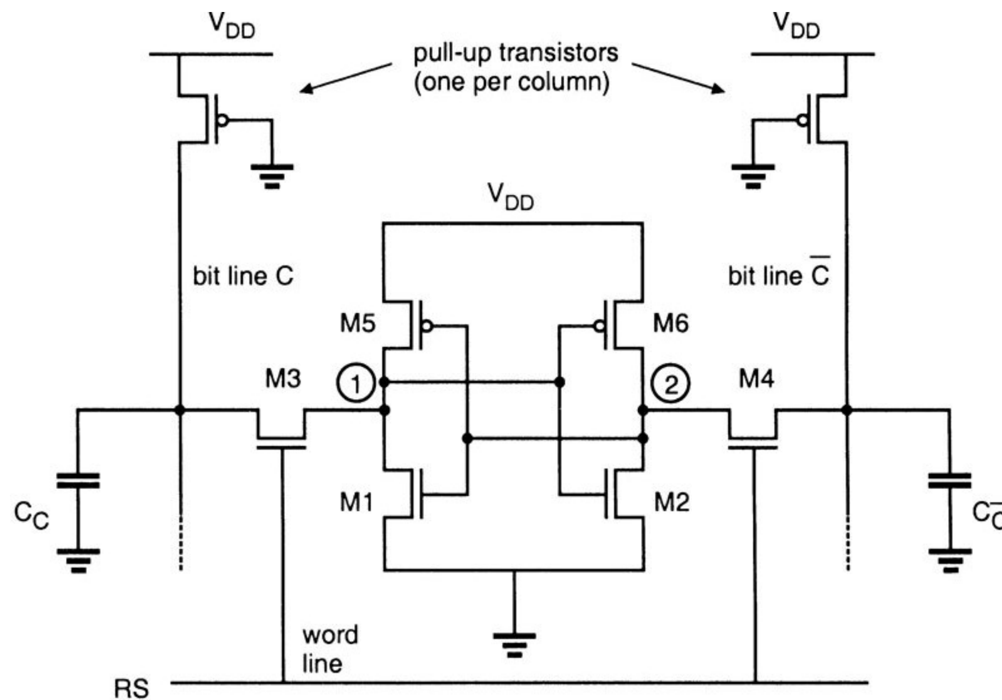


- Full CMOS SRAM
  - Most popular
  - Lowest static power
  - Superior noise margins and switching speed

© CMOS Digital Integrated Circuits – 4th Edition

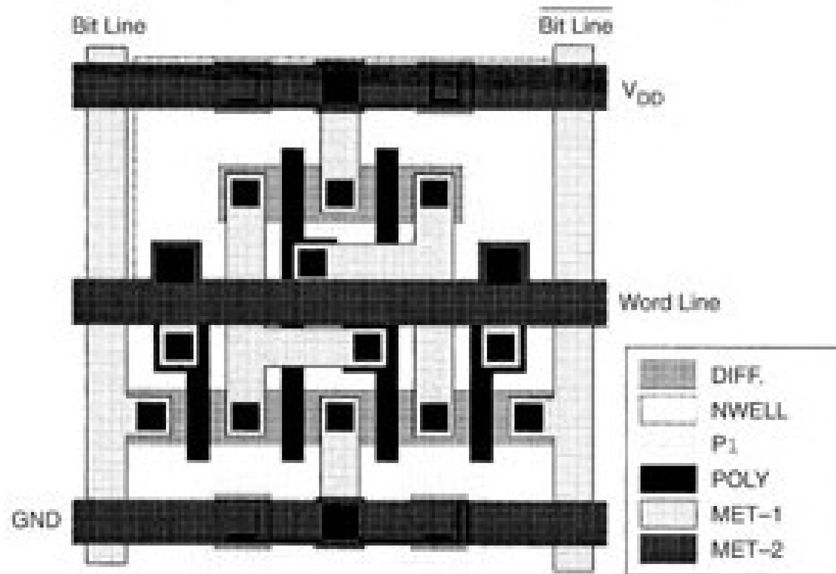
Full CMOS SRAM Cell(1)

- Very small static power dissipation (limited by leakage current)
- High noise immunity (large noise margin)
- Ability to operate at lower supply
- Disadvantage : cell area slightly larger, latch-up phenomena

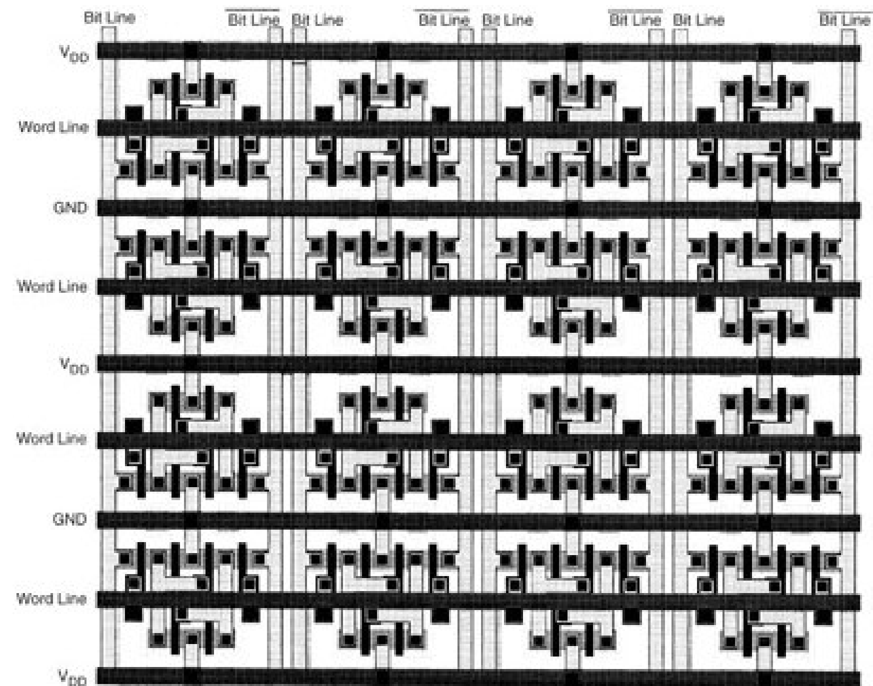


© CMOS Digital Integrated Circuits – 4th Edition

Full CMOS SRAM Cell(2)



Layout of CMOS SRAM cell

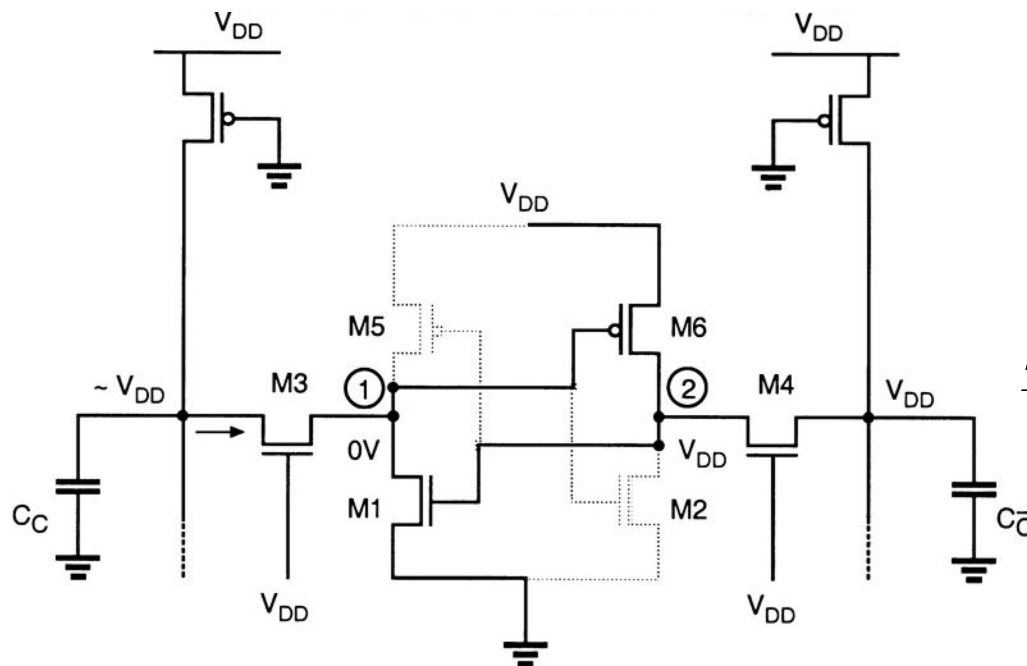


Layout of a 4-bit X 4bit SRAM array, consisting of 16 CMOS SRAM cells

© CMOS Digital Integrated Circuits – 4th Edition

CMOS SRAM Cell Design Strategy(1)

- Two basic requirements which dictate W/L ratio
  - Non-destructive data read operation
  - Modify stored data during write phase



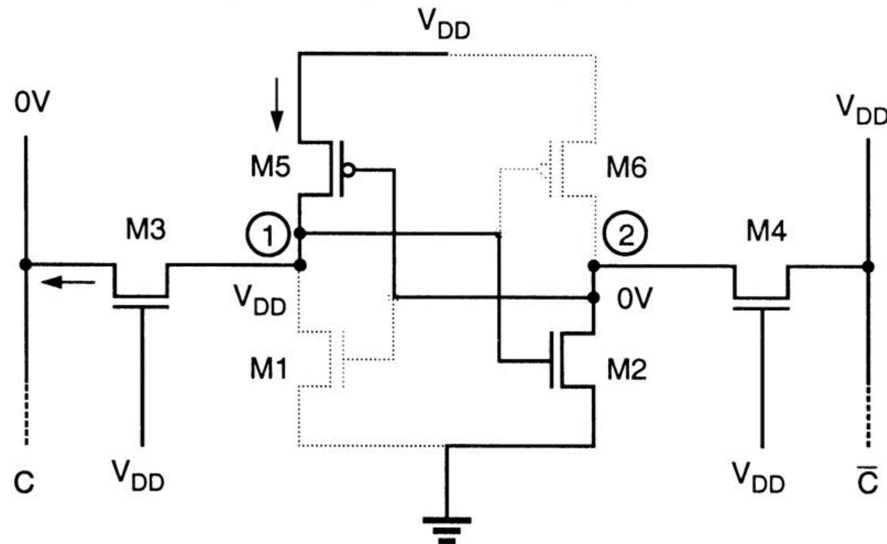
- Read Operation (0 stored)
  - M3 and M1 conduct some current
  - V_{C_C} drops slightly and V_1 increases
  - $V_{1,\max} \le V_{T,2}$ not to turn on M2
  - M3 in Saturation and M1 in linear

$$\frac{k_{n,3}}{2}(V_{DD} - V_1 - V_{T,n})^2 = \frac{k_{n,1}}{2}(2(V_{DD} - V_{T,n})V_1 - V_1^2)$$

$$\frac{k_{n,3}}{k_{n,1}} = \frac{\left(\frac{W}{L}\right)_3}{\left(\frac{W}{L}\right)_1} < \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} - 2V_{T,n})^2}$$

© CMOS Digital Integrated Circuits – 4th Edition

CMOS SRAM Cell Design Strategy(2)



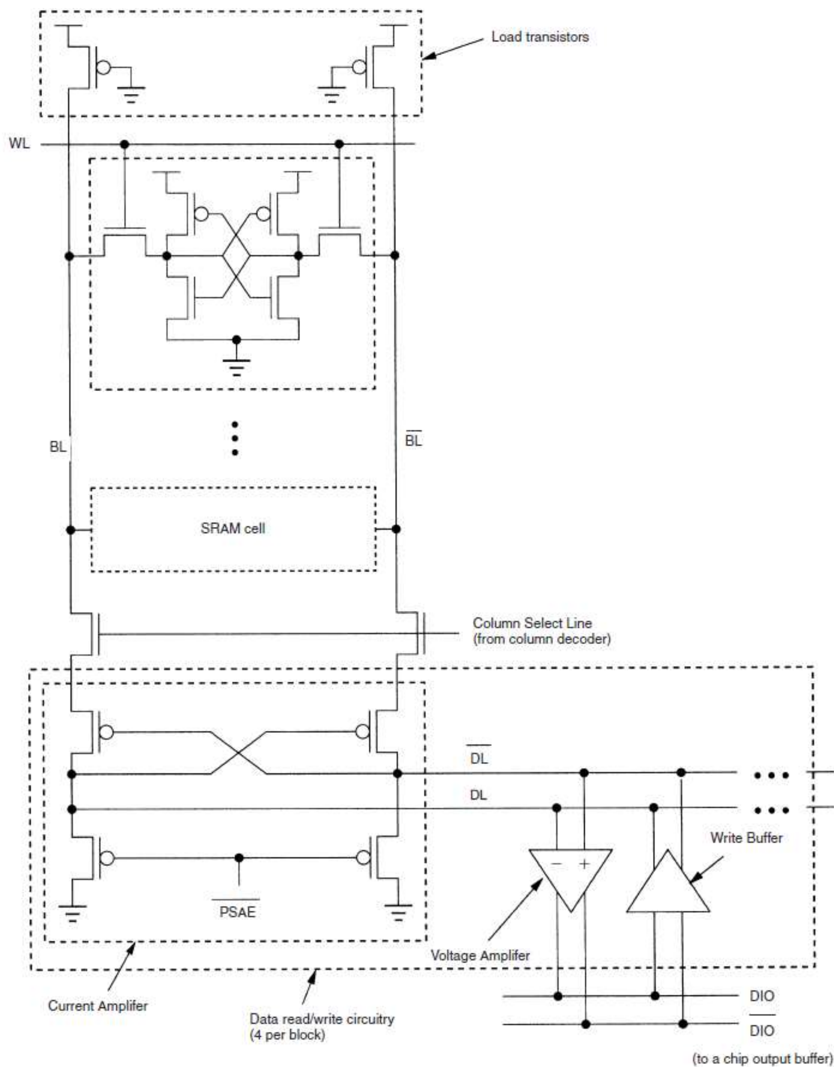
- Write 0 operation (initially, 1 was stored at node 1)
  - V_1 must be reduced below $V_{T,2}$ ➔ M2 turns off, V2 rises and V1 falls
  - When V_1=$V_{T,n}$, M3 in linear & M5 in saturation

$$\frac{k_{p,5}}{2}(0-V_{DD}-V_{T,p})^2 = \frac{k_{n,3}}{2}(2(V_{DD}-V_{T,n})V_{T,n}-V^2_{T,n})$$

$$\frac{k_{p,5}}{k_{n,3}} < \frac{2(V_{DD}-1.5V_{T,n})V_{T,n}}{(V_{DD}+V_{T,p})^2} \quad \Rightarrow \quad \frac{\left(\frac{W}{L}\right)_5}{\left(\frac{W}{L}\right)_3} < \frac{\mu_n}{\mu_p}\cdot\frac{2(V_{DD}-1.5V_{T,n})V_{T,n}}{(V_{DD}+V_{T,p})^2}$$

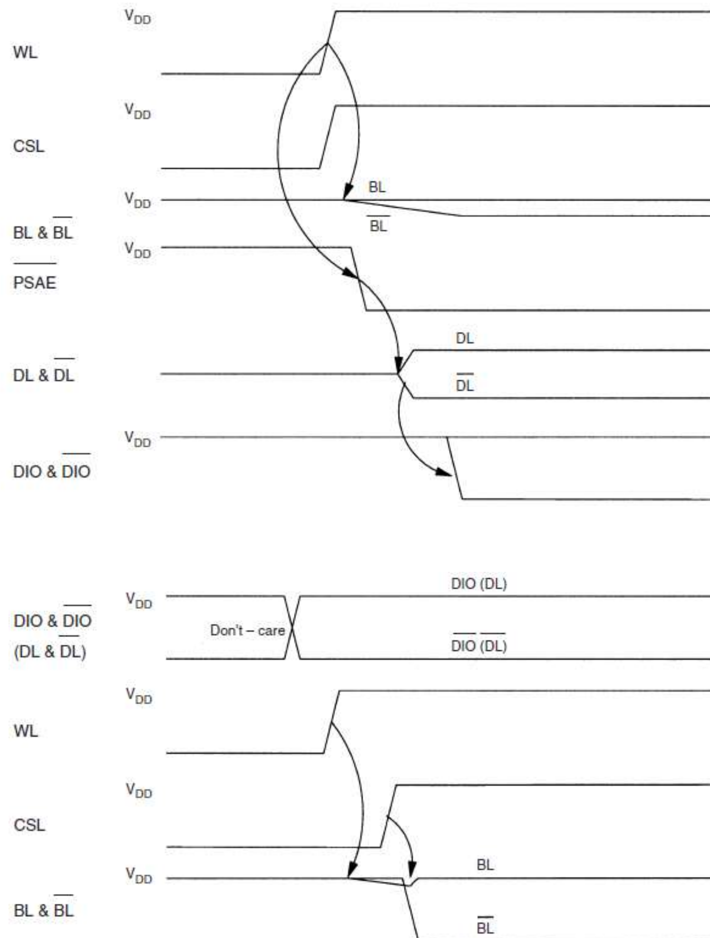
© CMOS Digital Integrated Circuits – 4th Edition

Memory Structure of SRAM



- Word line selected by row address
- Cell data kept during read operation
- Boosted voltage not required
- Address multiplexing scheme is not used (fast access time than DRAM)
- Depend on applications
  - ultra low power : load transistor turns off during read operation
  - high speed : remains on

Operation of SRAM

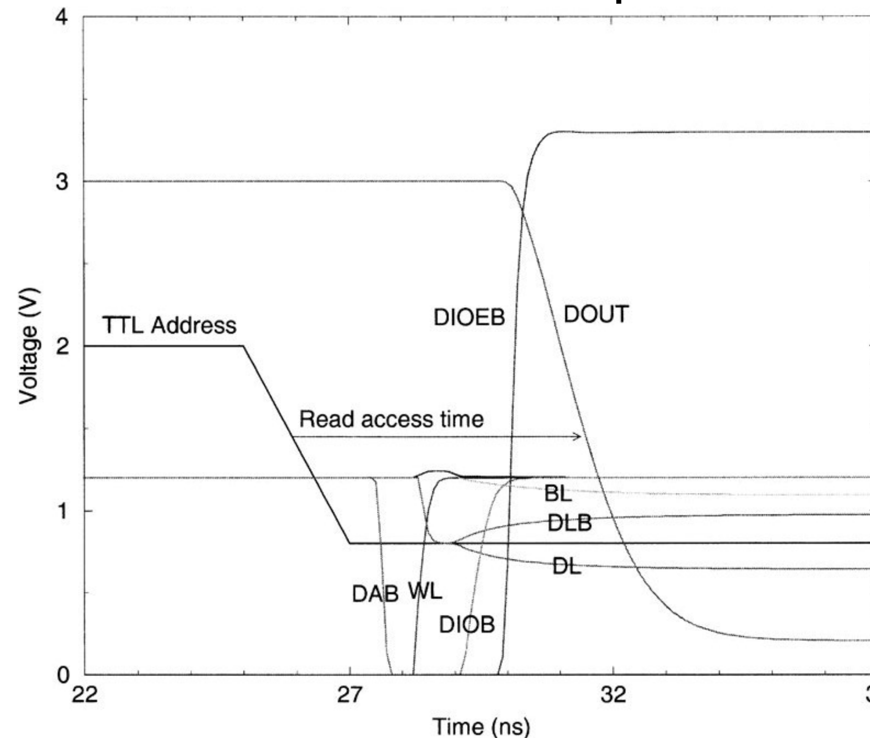


- **Read operation**
  - Word line enable
  - One bit line discharge
    (voltage change of bit line is very small)
  - Sense amp. detect the voltage difference on bit line
  - Multi-stage amp. is used to improve read speed
- **Write operation**
  - Word line selected by row address
  - Write buffer write data into cell
  - Write buffer has larger current driving capability than cell
  - Write is faster than read

© CMOS Digital Integrated Circuits – 4th Edition

SRAM Read Operation

- TTL level converts into CMOS level signal
- Internal voltage regulator to reduce power dissipation to improve reliability

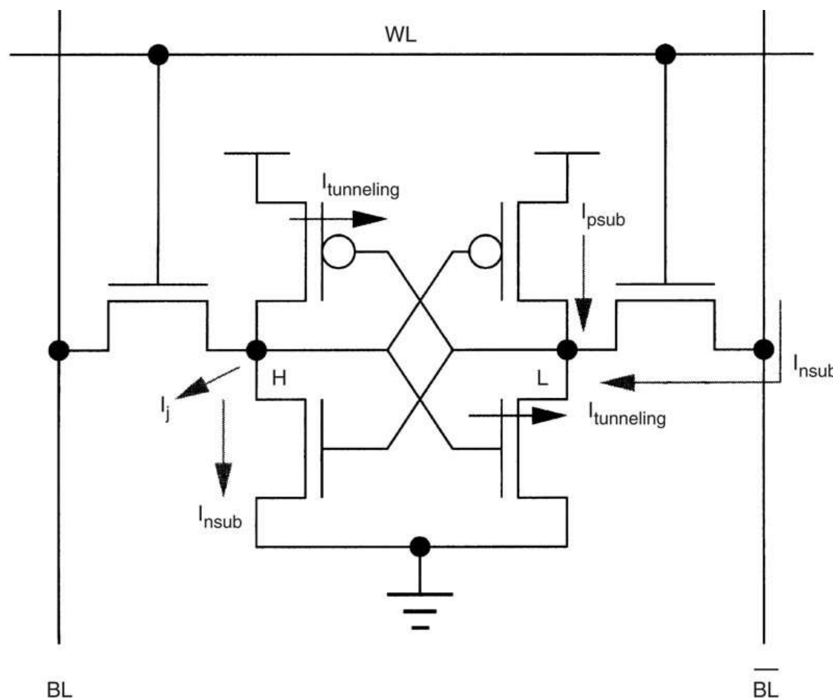


© CMOS Digital Integrated Circuits – 4th Edition

Leakage Currents in SRAM Cells

- Major portion of standby current
- Standby power is key parameter for low power design
- High threshold

  reduction of leakage ⟷ degradation of performance



I_j : junction current

  data "1" to substrate

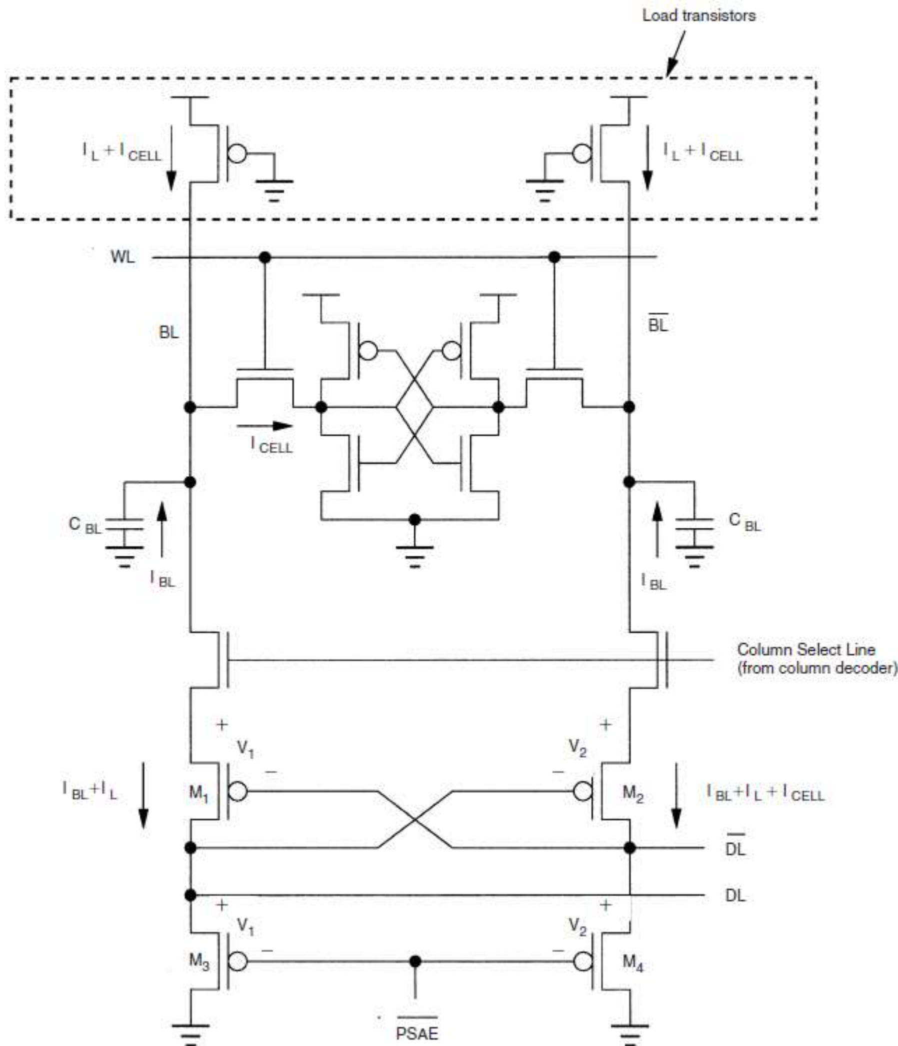
I_{nsub} and I_{psub} : subthreshold leakage

  turn off NMOS and PMOS

$I_{tunneling}$: tunneling current

  cross thin gate oxide

SRAM Read/Write Circuits



- Current-mode sense amp widely used in SRAM
  - ➔ improve signal sensing speed independent of bit line cap.
- Signal line connect to source of latch transistor
- Current difference appears on DL and $\overline{DL}$
- Open-loop gain

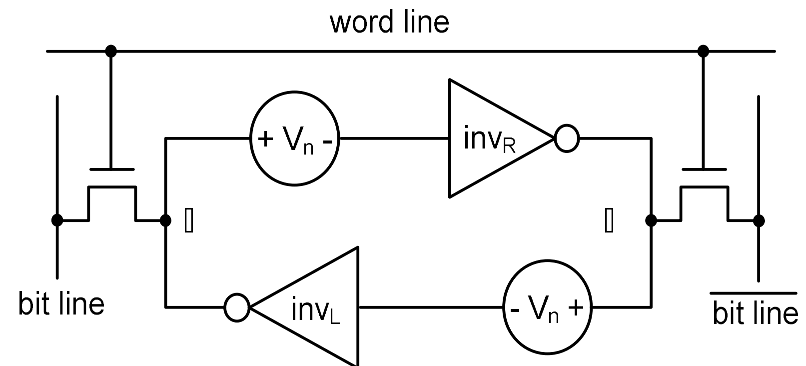
$$Gain_{open-loop} = \frac{g_m(m3) \times g_m(m4)}{g_m(m1) \times g_m(m2)}$$

- Current-mode sense amp: Drawback- larger power consumption

© CMOS Digital Integrated Circuits – 4th Edition

SRAM Cell at Low Supply Voltage

- SRAM cell susceptible to variabilities
  - Due to minimum device size to a minimize area
  - Threshold voltage variation covered in Ch. 3 plus layout induced threshold voltage variation
  - PMOS pair (M5, M6) in SRAM cell- different V_T due to NBTI
  - NMOS pair (M1, M2) in SRAM cell- different V_T due to PBTI
- Static noise margin (SNM)
  - A noise tolerant voltage before the stored data flip
  - Equivalent ckt to measure SNM
  - 6-T SRAM cell at low supply voltage degrades SNM

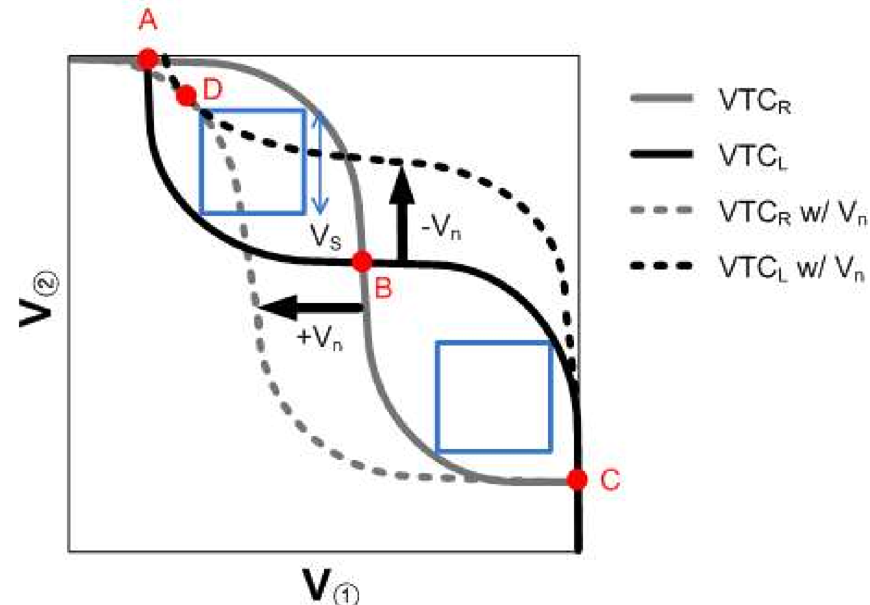


V_n: DC noise, SNM: min. DC noise which flips the state of SRAM cell during read operation

© CMOS Digital Integrated Circuits – 4th Edition

SNM Variation due to DC Noise

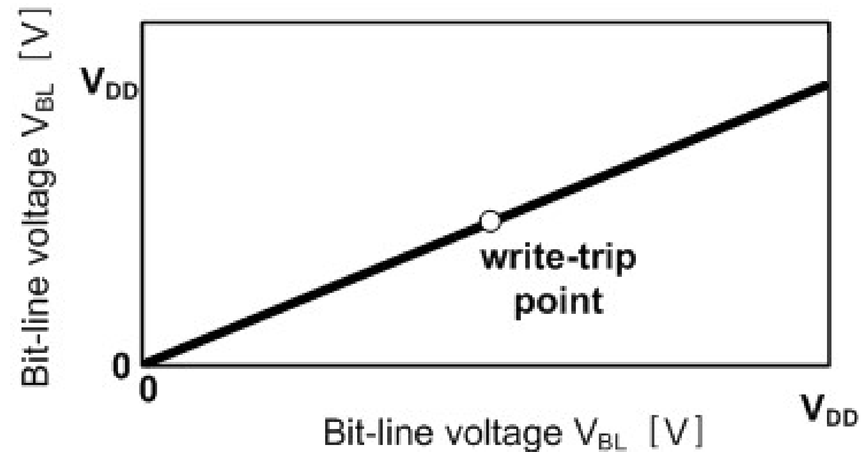
◆ How to measure SNM graphically

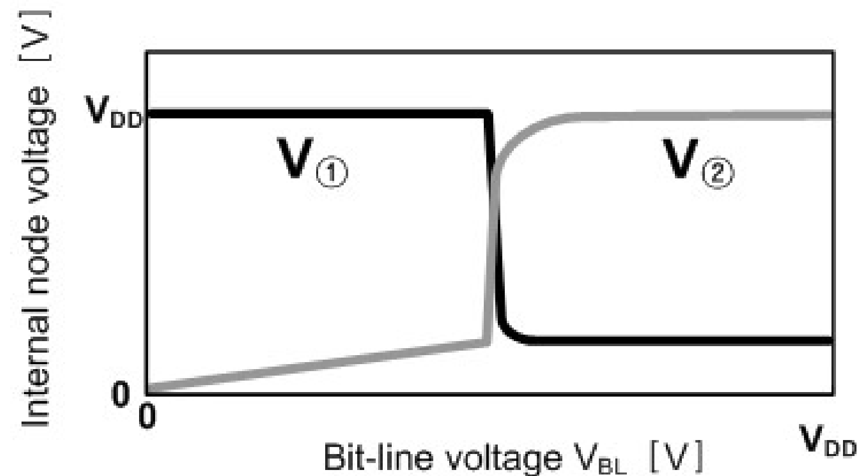


- SNM: The length of side of the smaller nested square in the two openings of butterfly curve
- Before two V_ns are fed: SNM=V_S
- After V_ns are fed: stable point A and unstable point B meets at D
- More V_ns are applied: one common point C & the stored bits are flipped

SRAM Cell Writability

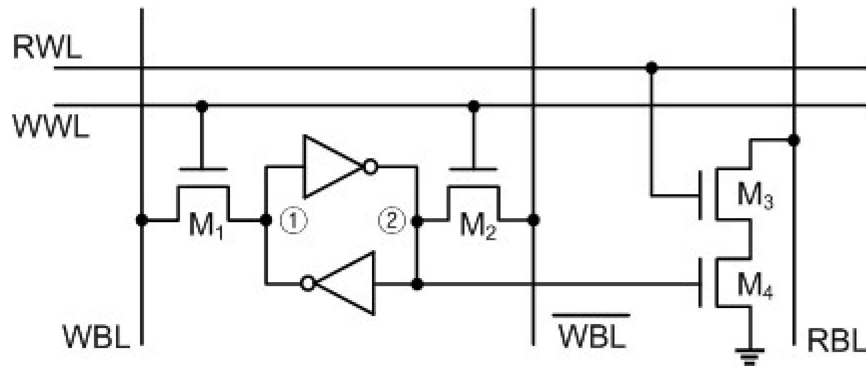
- Write-trip point
  - A metric for writability
  - Max. bit line voltage to flip the state of the SRAM cell
  - Primarily determined by the pull-up ratio of SRAM cell (Ex: $(W/L)_5/(W/L)_3$)
- Variability tolerant 6T SRAM cell
  - Trade off bw. read stability and writability
  - M_3 & M_4 ↑: SNM ↑, writability ↓

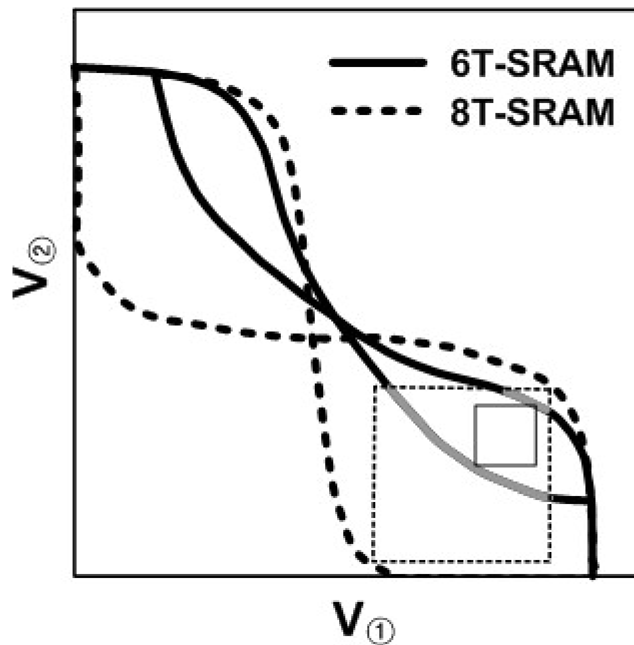




© CMOS Digital Integrated Circuits – 4th Edition

8T SRAM Cell

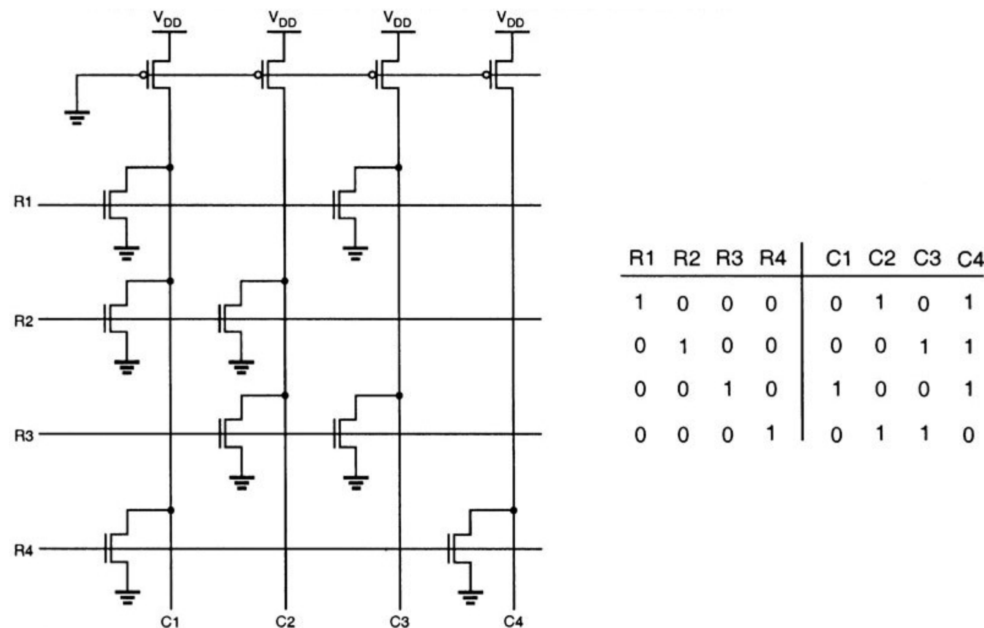




- No secondary power supplies
- Decouples the SRAM cell nodes from the bit line which enables balancing the read & write modes
- Read operation doesn't affect the stored data
- 6T cell has the worst SNM in read operation where the pass gate transistor increases the voltage at the '0' stored

10.4 Nonvolatile Memory

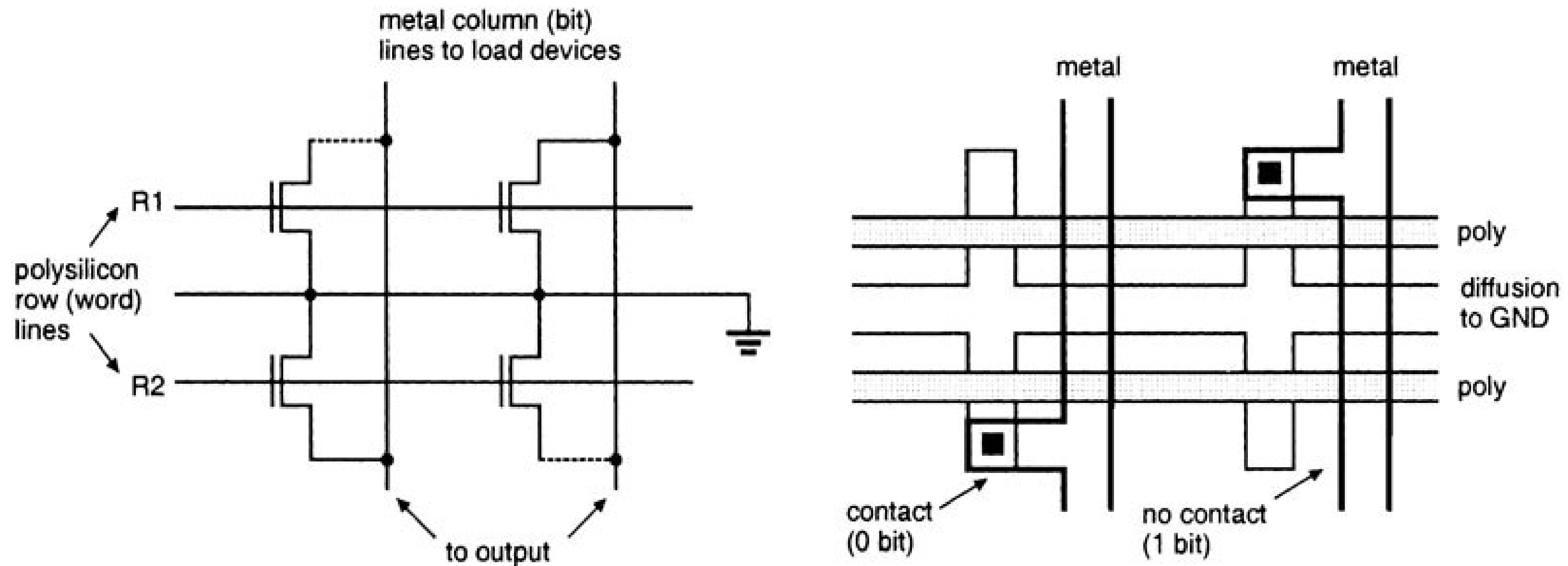
- Simple combinational Boolean network
- Only one word line selected at a time
- Active transistors exist at cross point
- Dynamic ROM
  - use periodic precharge signal to reduce static power



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

Layout of NOR ROM Array(1)

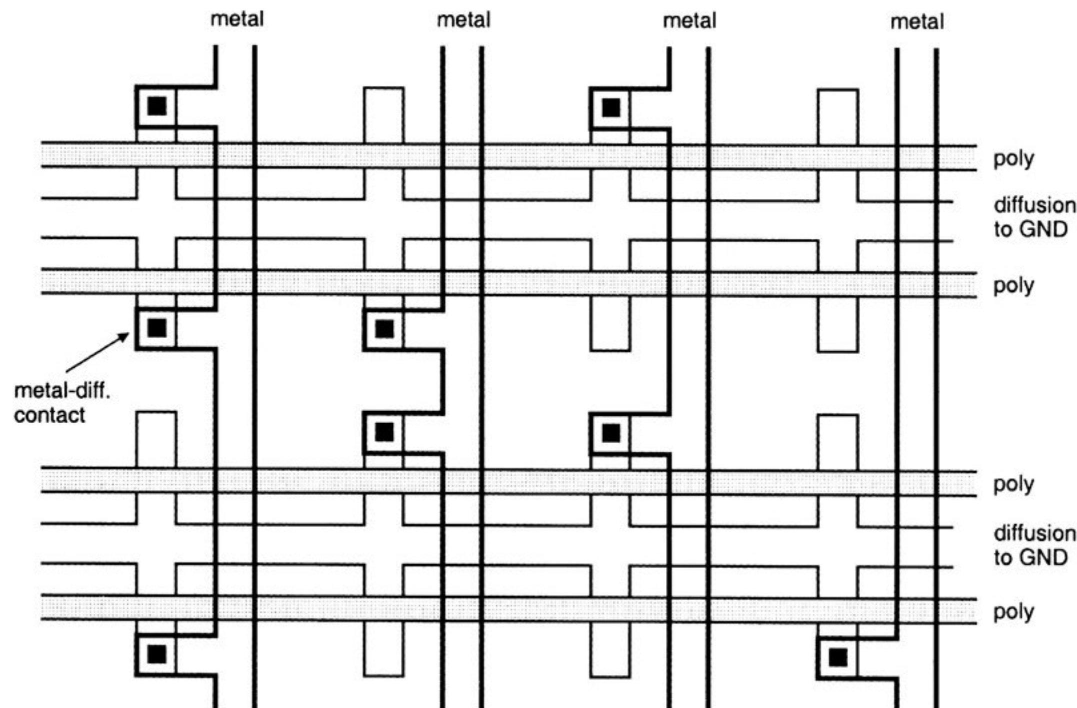
- Initially, NMOS at every row-column intersection
- '1'-bits are realized by omitting drain or source connection or gate electrode of corresponding NMOS



Layout example of a NOR ROM array

Layout of NOR ROM Array(2)

- In reality, metal column lines laid out directly on top of diffusion column to reduce horizontal dimension

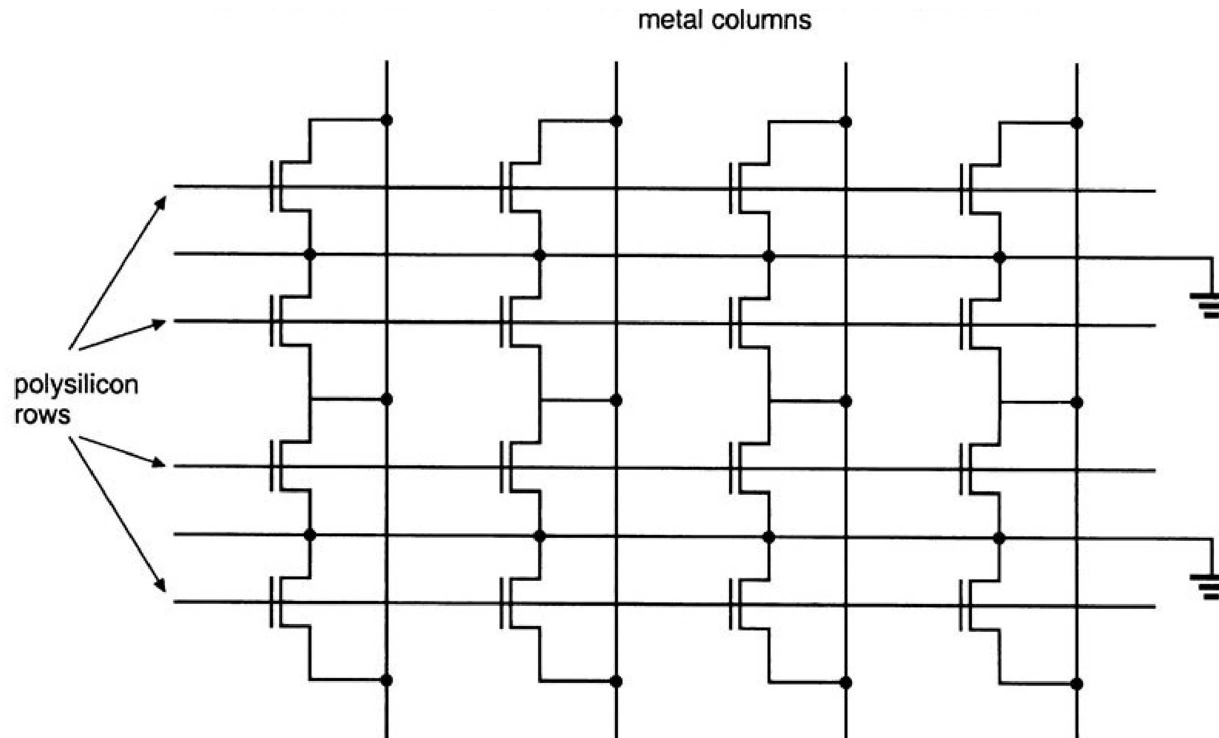


Layout of the 4-bit X 4-bit NOR ROM array (pp. 46)

© CMOS Digital Integrated Circuits – 4th Edition

Implant-mask Programmable NOR ROM

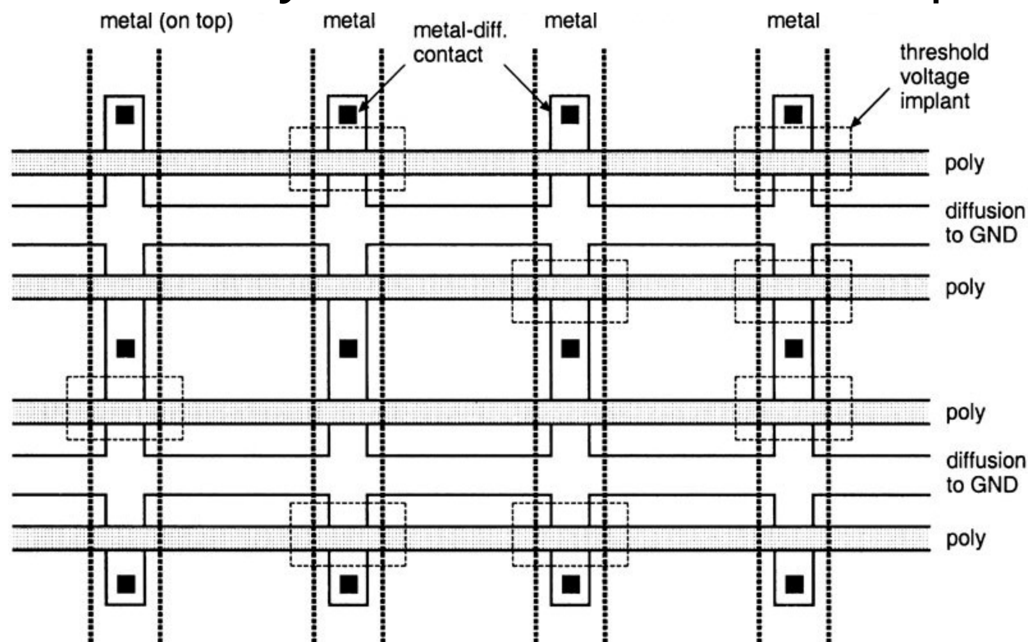
- Every two rows share a common ground connection
- Every metal to diffusion contact shared by two adjacent devices



© CMOS Digital Integrated Circuits – 4th Edition

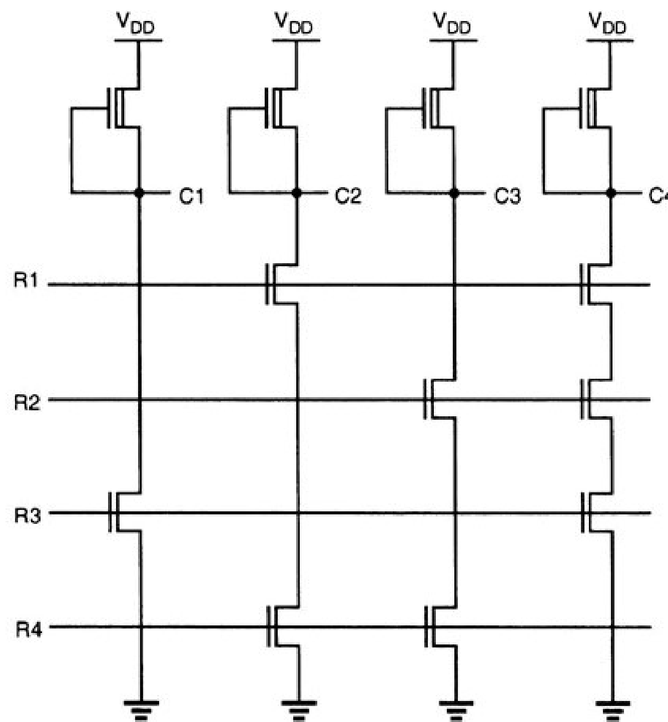
4-bit x 4-bit NOR ROM Array

- Based on implant-mask programming
- Raised threshold voltage $>V_{OH}$ ➔ "1"-bit
- Non-implanted ➔ "0"-bit
- higher core density (smaller silicon area per stored bit)



4-bit x 4-bit NAND ROM Array

- Bit line : depletion-load NAND gate
- Deactivated transistor ➔ “1”-bit
- Shorted or on transistor ➔ “0”-bit

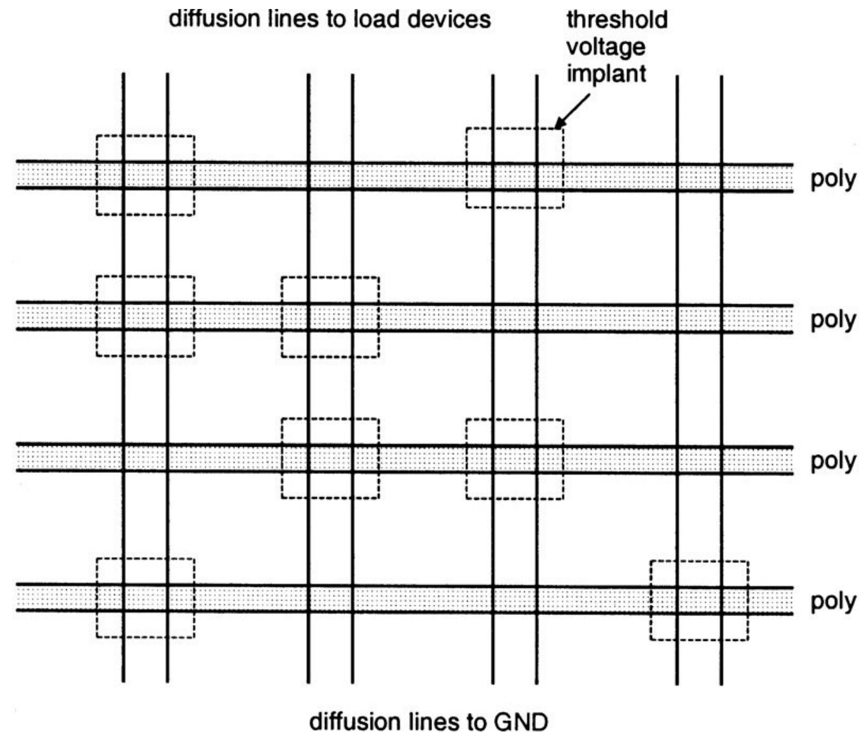


R1	R2	R3	R4	C1	C2	C3	C4
0	1	1	1	0	1	0	1
1	0	1	1	0	0	1	1
1	1	0	1	1	0	0	1
1	1	1	0	0	1	1	0

© CMOS Digital Integrated Circuits – 4th Edition

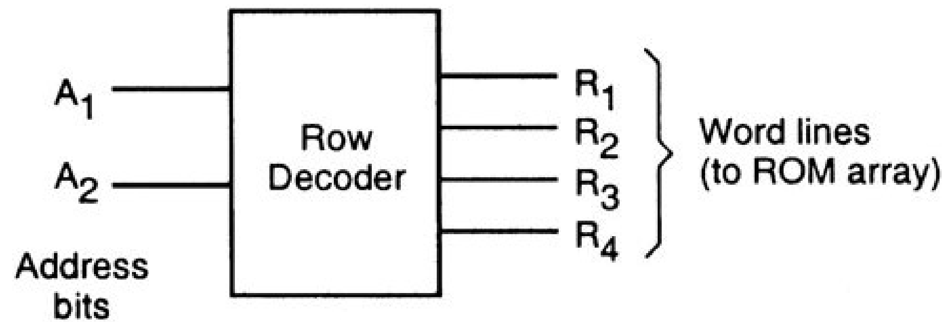
Implant-mask layout of NAND ROM

- Lowered threshold voltage < 0V ➔ "0"-bit
- Much more compact than NOR ROM
- Access time is slower than NOR ROM

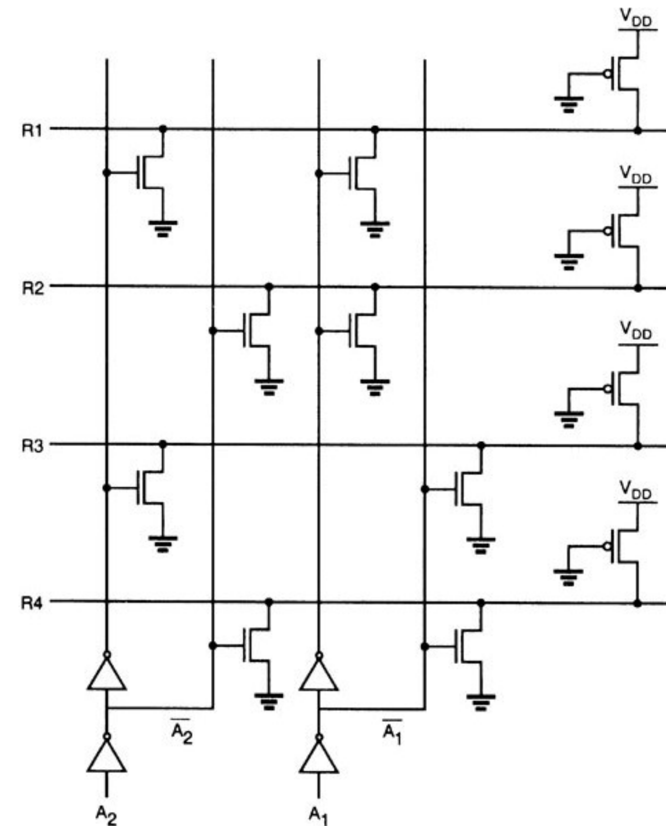


Design of Row and Column Decoders(1)

- Select a particular memory location in array
- Row address decoder example



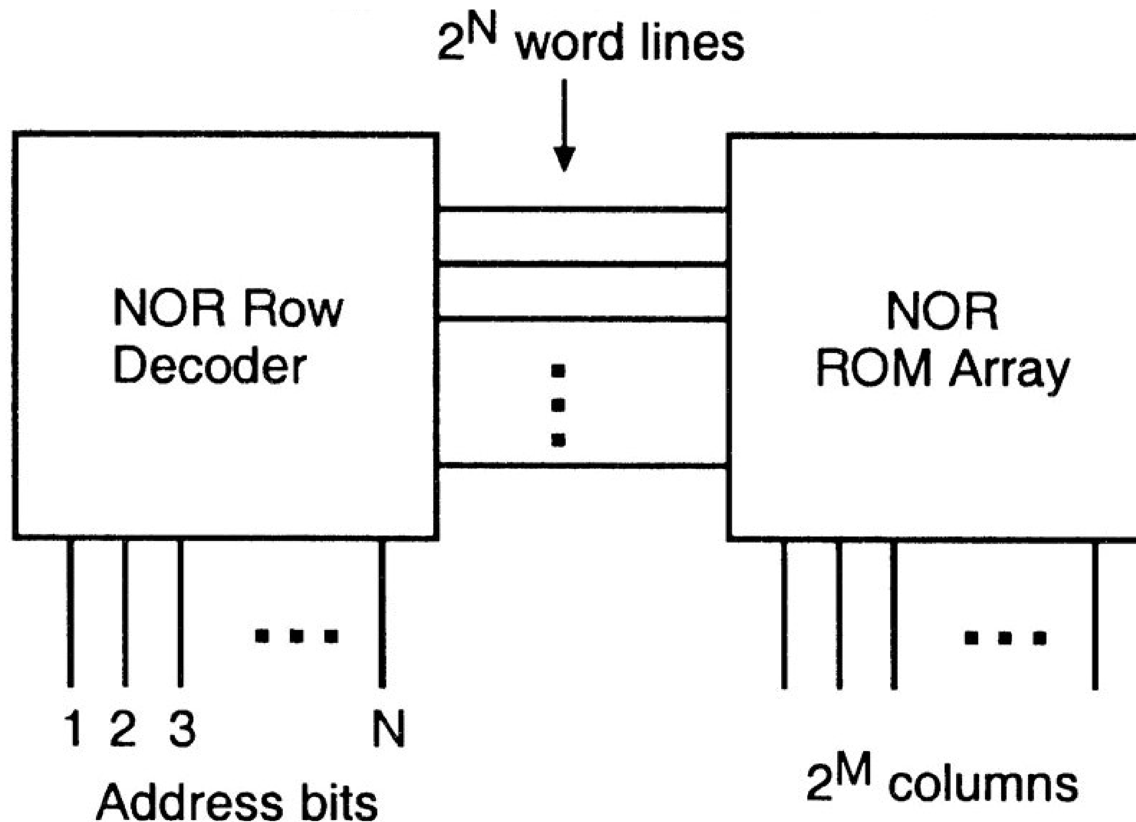
A_1	A_2	R_1	R_2	R_3	R_4
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	1



© CMOS Digital Integrated Circuits – 4th Edition

Design of Row and Column Decoders(2)

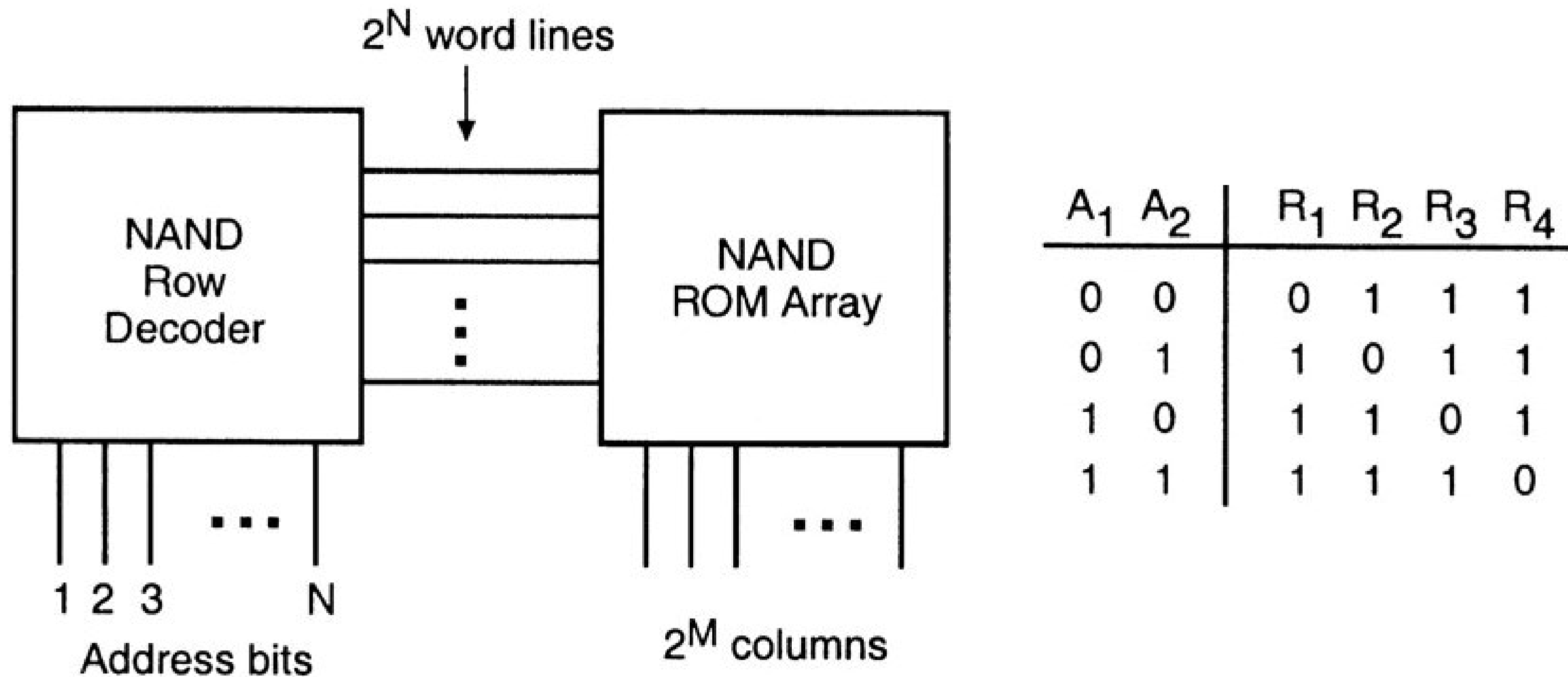
◆ ROM array and row decoder (two adjacent NOR arrays)



© CMOS Digital Integrated Circuits – 4th Edition

Row Decoder for NAND ROM

- Lower voltage for logic "0"
- Realized using same layout strategy as memory array

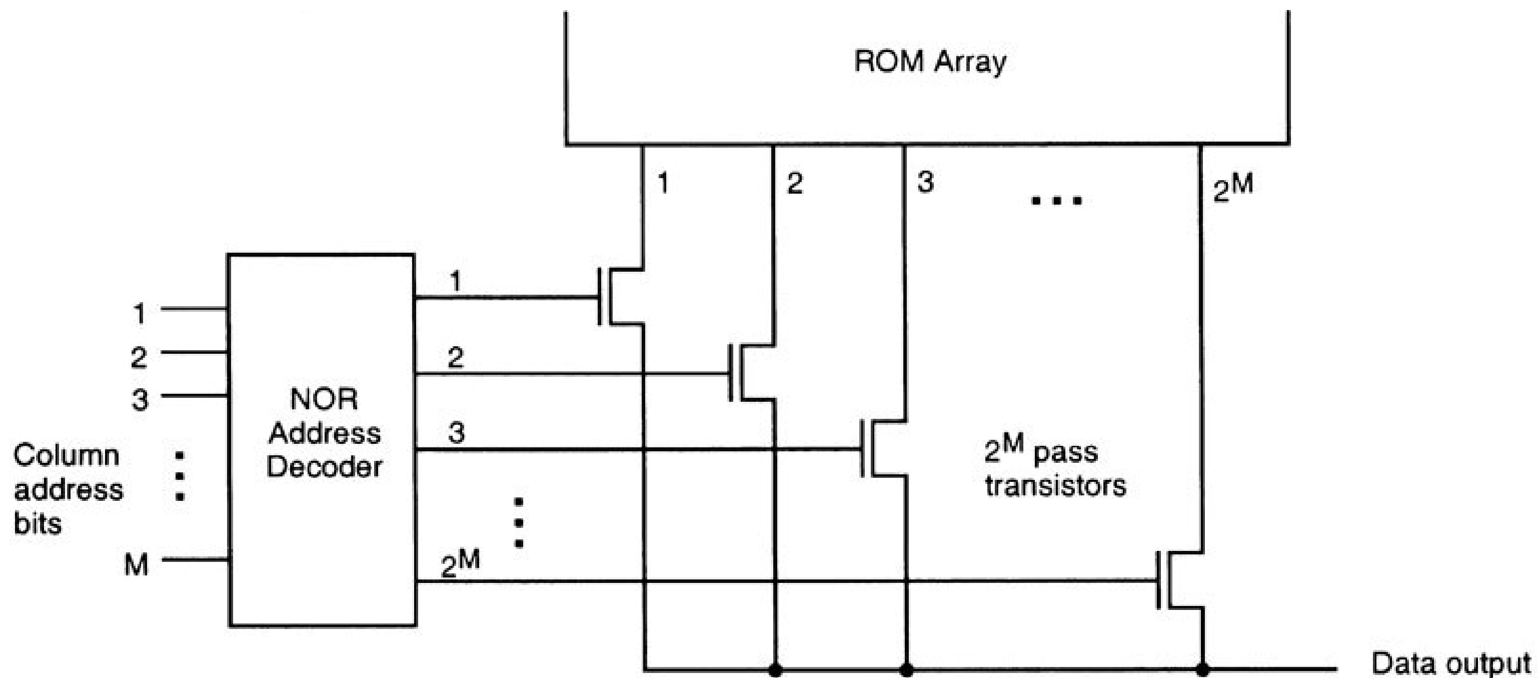


A_1	A_2	R_1	R_2	R_3	R_4
0	0	0	1	1	1
0	1	1	0	1	1
1	0	1	1	0	1
1	1	1	1	1	0

© CMOS Digital Integrated Circuits – 4th Edition

Column Decoder(1)

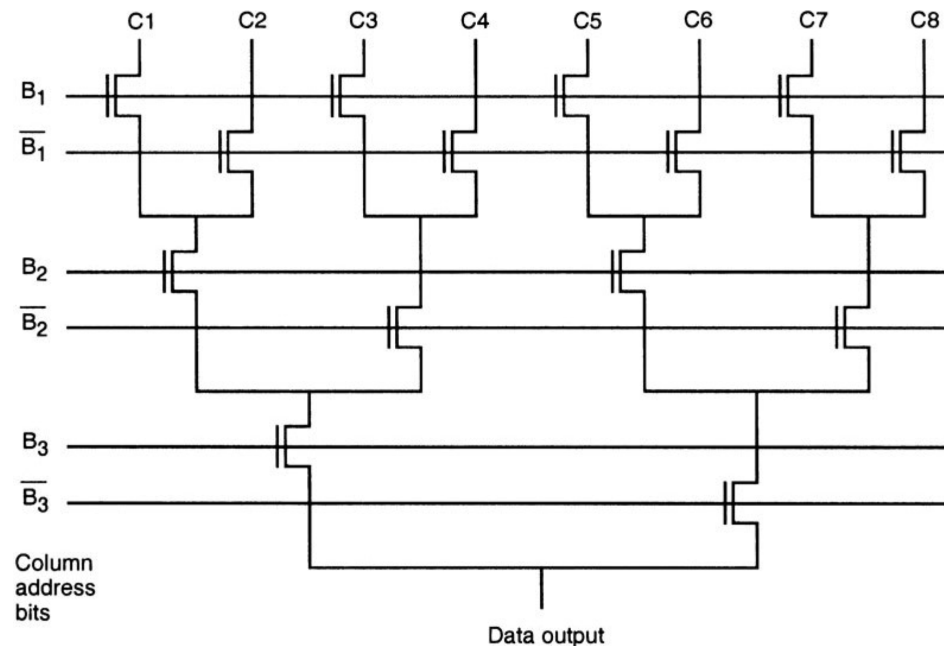
- Using NOR address decoder and NMOS pass transistor
- Only one pass transistor turned on at a time
- $2^M(M+1)$ transistors required



© CMOS Digital Integrated Circuits – 4th Edition

Column Decoder(2)

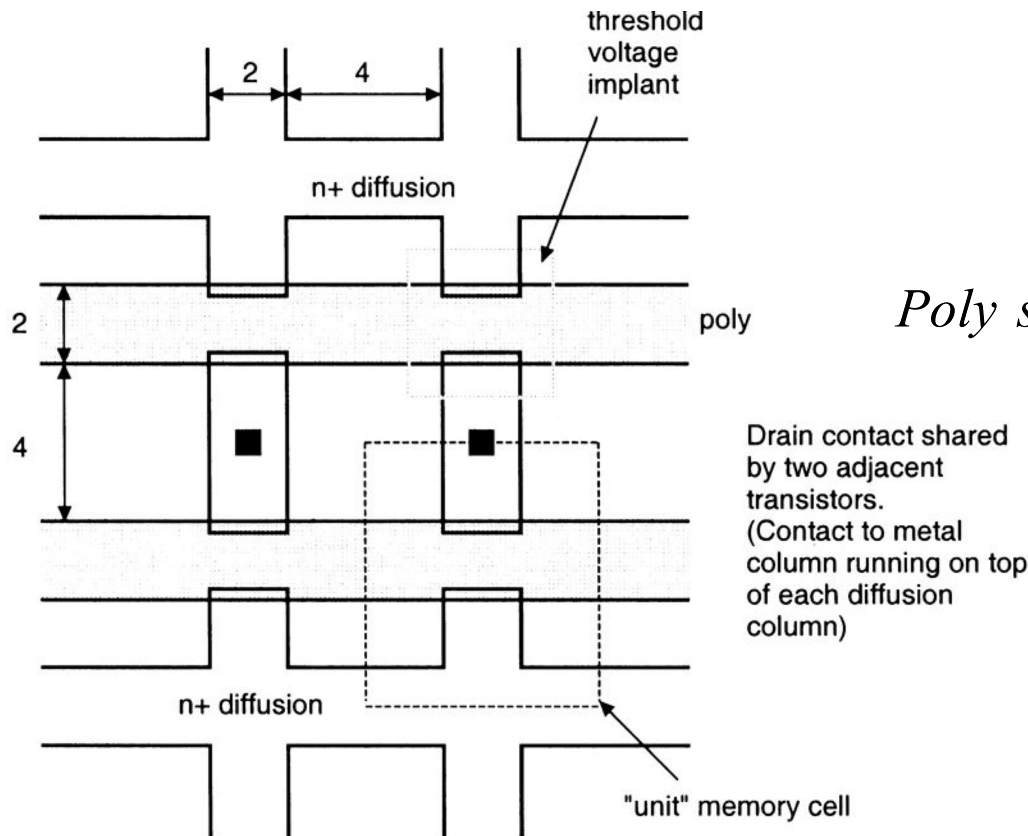
- Binary selection tree decoder
- NOR address decoder not needed
  - ➡ Reduce the number of transistors significantly
- But, long data access time



© CMOS Digital Integrated Circuits – 4th Edition

Example 10.1(1)

- Analyze the access time of a 32-kbit NOR ROM array



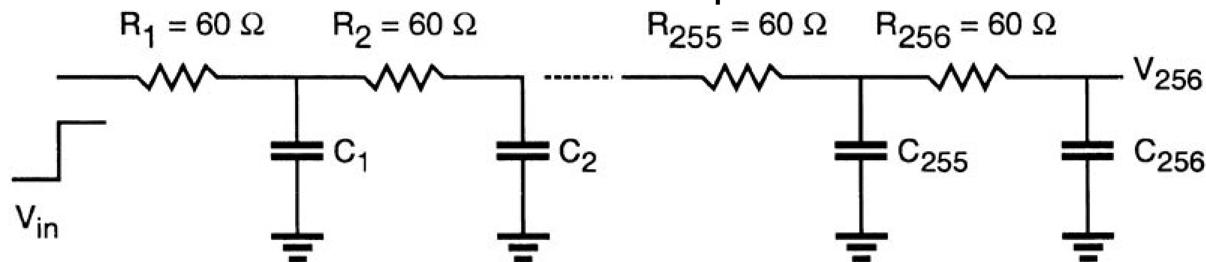
$$\mu_n C_{ox} = 20 \mu A / V^2$$

$$C_{ox} = 3.47 \mu F / cm^2$$

$$Poly\ sheet\ resistance = 20\Omega / square$$

Example 10.1(2)

- Assume 7 row address bits and 8 column address bits (128 rows and 256 columns)
- Calculate row resistance and capacitance



$$C_{row} = C_{ox} \cdot W \cdot L = 10.4\,fF\,/\,bit$$

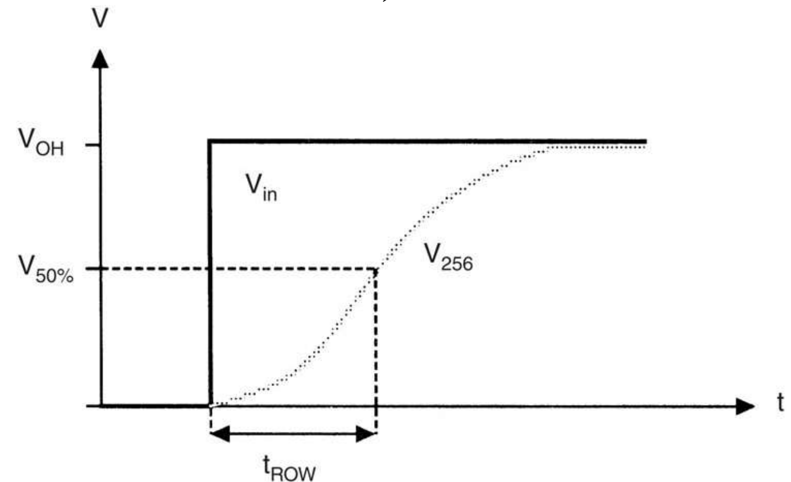
$$R_{row} = (\#\ of\ \ squares) \times (Poly\ \ sheet\ \ resistance) = 60\Omega\,/\,bit$$

- Calculate row access time

$$t_{row} \approx 0.38 \cdot R_T \cdot C_T = 15.53ns$$

$$R_T = \sum_{all\ columns} R_i = 15.36k\Omega$$

$$C_T = \sum_{all\ columns} C_i = 2.66\,pF$$



Example 10.1(3)

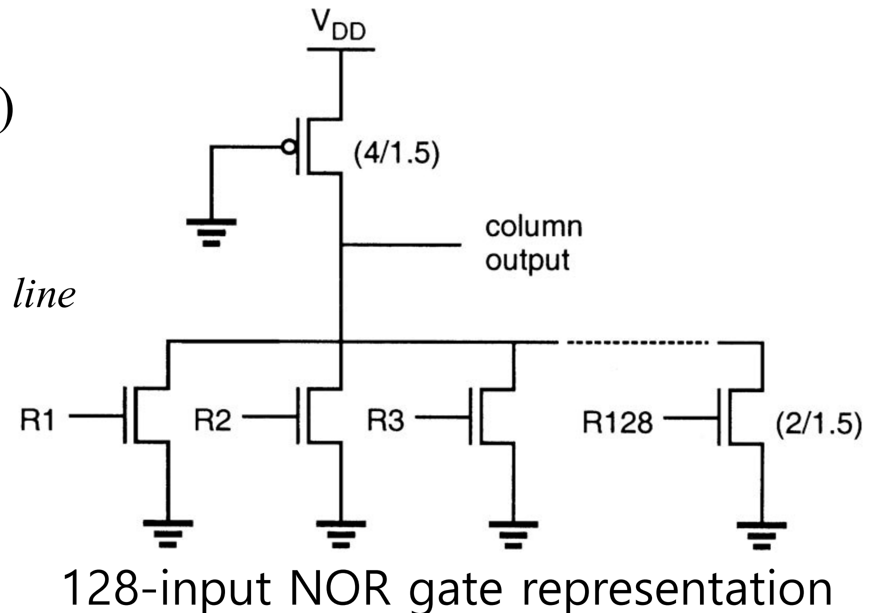
- A more accurate delay: Elmore time constant for RC ladder circuits

$$t_{row} = \sum_{k=1}^{256} R_{jk} C_k = 20.52ns \quad where \quad R_{jk} = \sum_{j=1}^{k} R_j$$

- Calculate column access time

$$C_{column} = 128 \times (C_{gd,driver} + C_{db,driver})$$
$$\approx 1.5pF$$

$$where \quad C_{gd,driver} + C_{db,driver} = 0.0118pF / word \ line$$

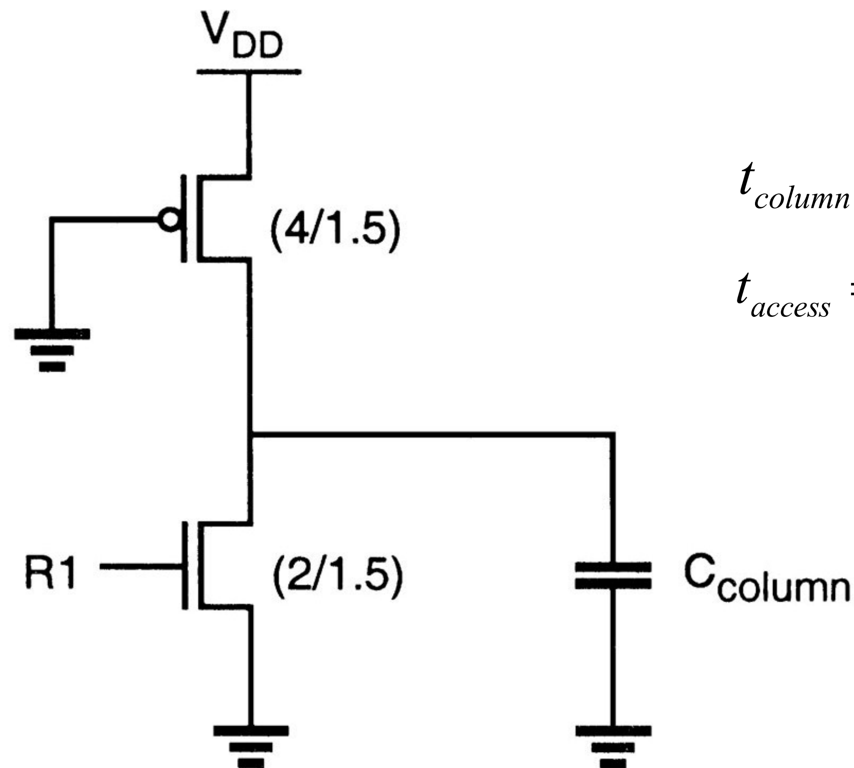


128-input NOR gate representation

© CMOS Digital Integrated Circuits – 4th Edition

Example 10.1(4)

- To calculate column access time, consider the worst-case signal propagation delay τ_{PHL} for below inverter



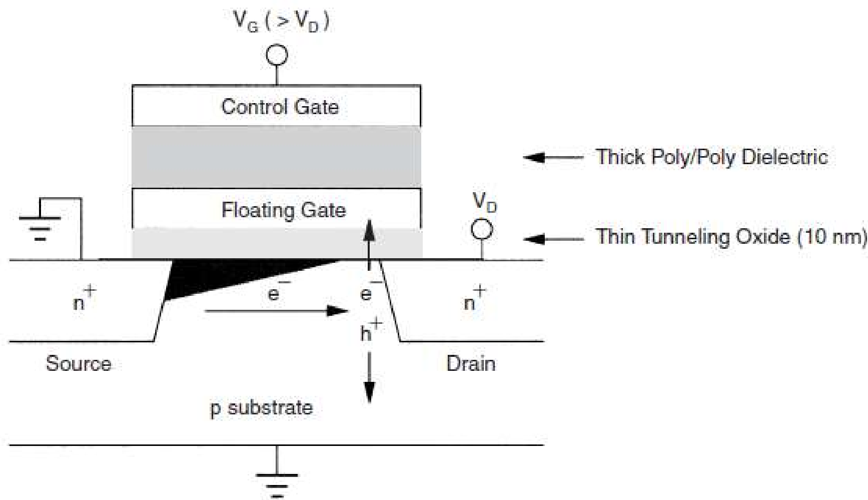
$$t_{column} = 18ns$$ (using eq. 6.18: t_{PHL})

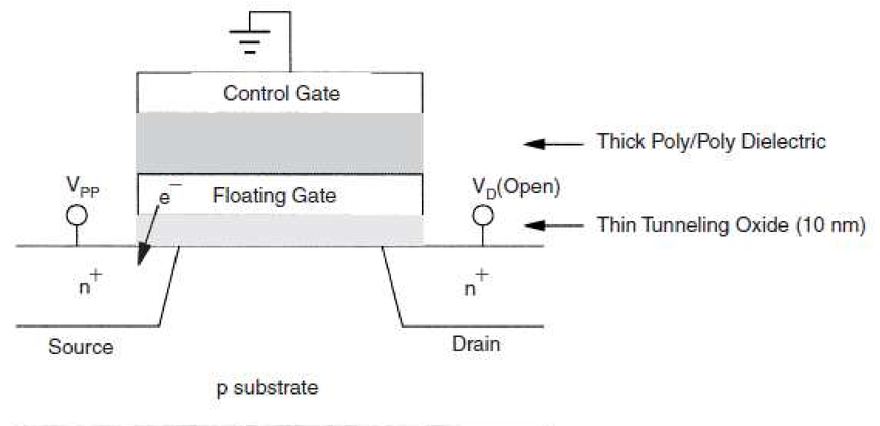
$$t_{access} = t_{row} + t_{column} = 38.5ns$$

© CMOS Digital Integrated Circuits – 4th Edition

10.5 Flash Memory

- One transistor with floating gate
- Memory cell can have two states (two threshold)
- Electron accumulated at the floating gate → higher threshold → "1" state
- Electron removed from the floating gate → lower threshold → "0" state



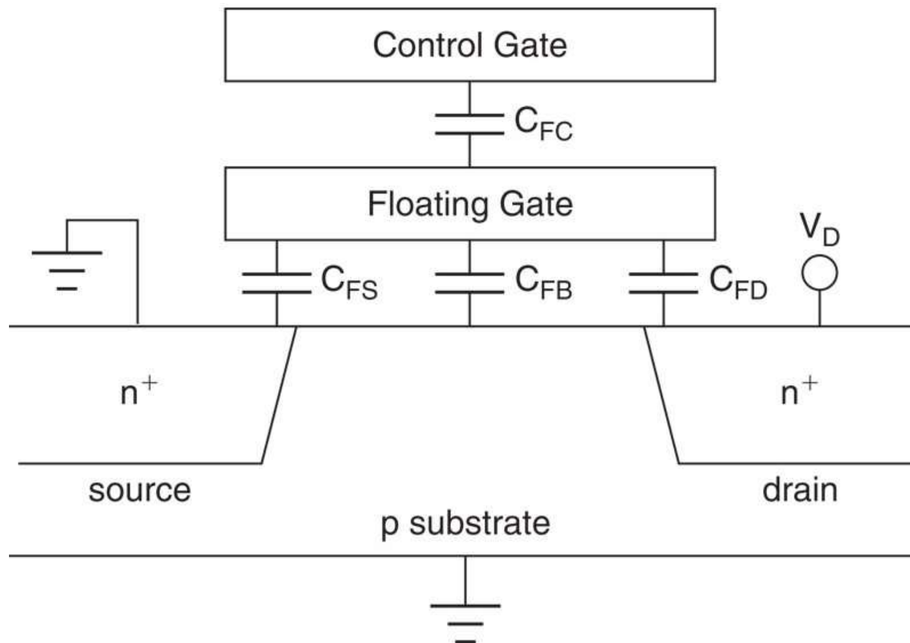


Hot electron injection mechanism (Data programming)

Fowler-Nordheim tunneling mechanism (Data erasing)

© CMOS Digital Integrated Circuits – 4th Edition

Equivalent Capacitive-Coupling Circuit



- V_{FG} by capacitive coupling after V_{CG} & V_D applied

$$V_{FG} = \frac{Q_{FG}}{C_{total}} + \frac{C_{FC}}{C_{total}} V_{CG} + \frac{C_{FD}}{C_{total}} V_D$$

$$C_{total} = C_{FC} + C_{FS} + C_{FB} + C_{FD}$$

- min. V_{CG} to turn on the control gate transistor

$$V_T(CG) = \frac{C_{total}}{C_{FC}} V_T(FG) - \frac{Q_{FG}}{C_{FC}} - \frac{C_{FD}}{C_{FC}} V_D$$

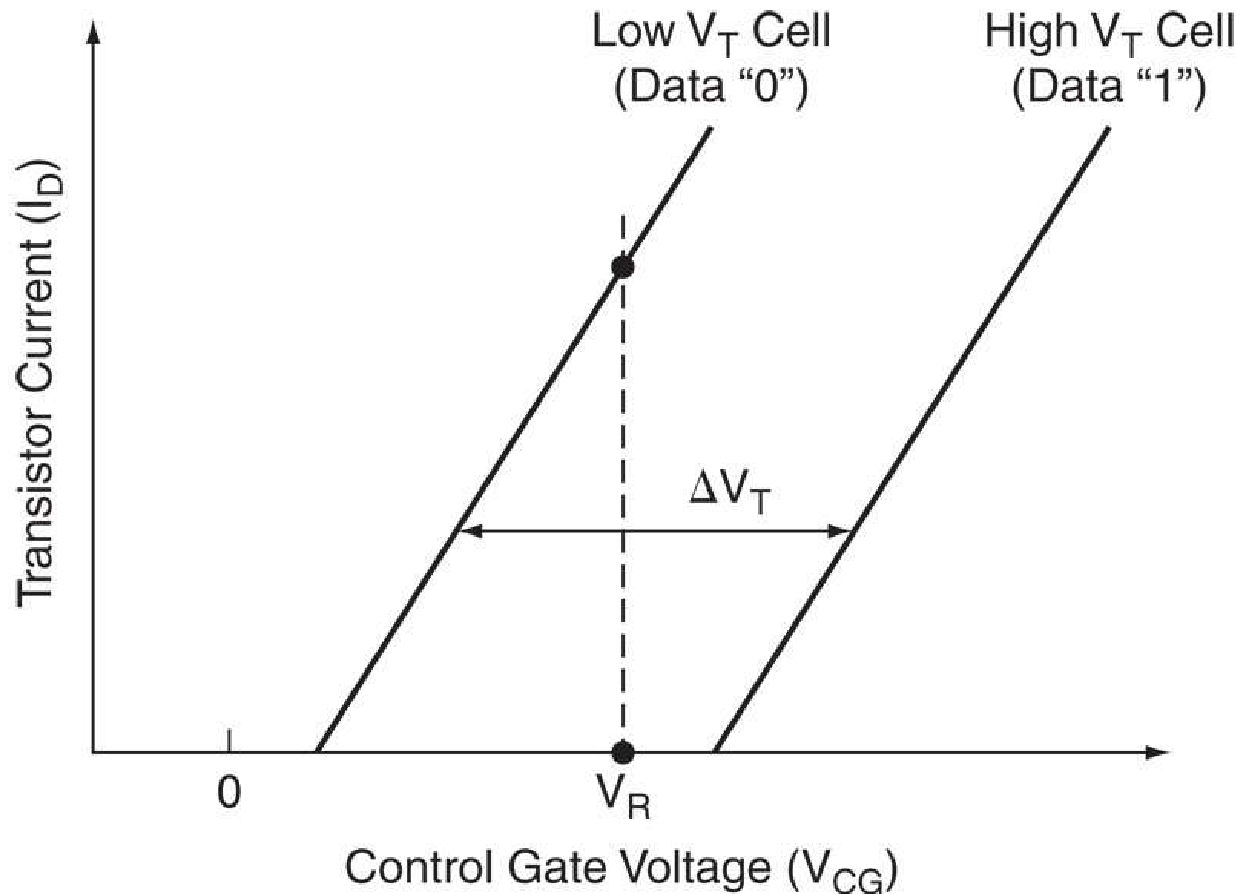
$$\Delta V_T(CG) = -\frac{\Delta Q_{FG}}{C_{FC}}$$

Q_{FC} : charge stored at floating gate
C_{total} : total cap.
C_{FC} : cap. between floating and control gate
C_{FS}, C_{FB} and C_{FD} : cap. between floating gate and source, bulk and drain
V_{CG} and V_D : voltage at control gate and drain
$V_T(FG)$: threshold voltage to turn on the floating gate transistor

© CMOS Digital Integrated Circuits – 4th Edition

I-V Characteristic of Flash Memory

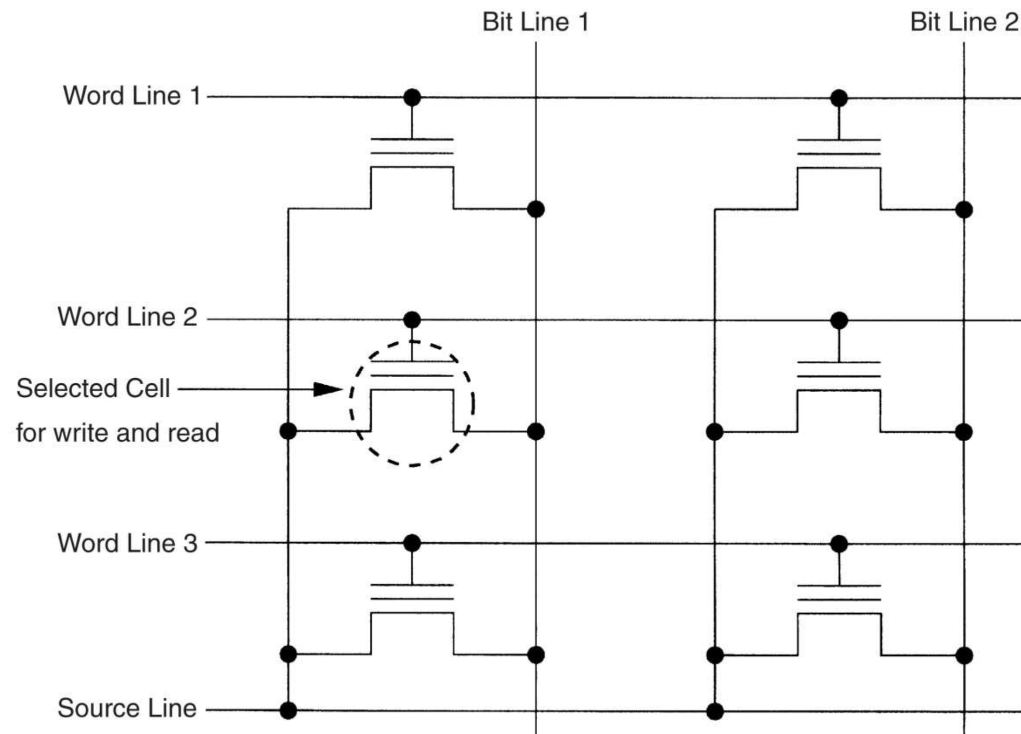
- Low and high threshold voltages for control gate voltage



© CMOS Digital Integrated Circuits – 4th Edition

NOR Flash Memory Cell

- Bias conditions and configuration of NOR Cells



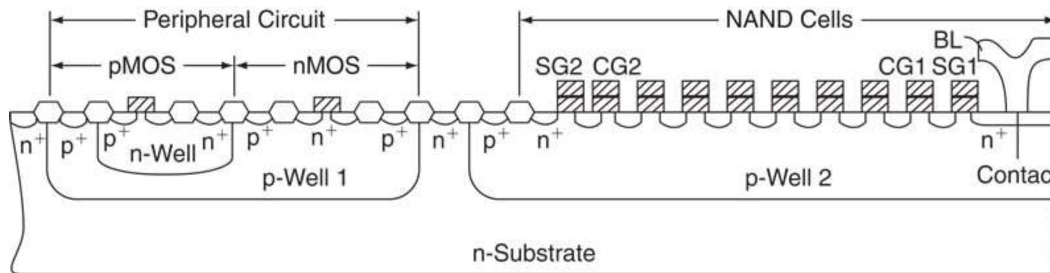
- F-N tunneling mechanism for erase operation
- Hot-electron injection mechanism for programming operation

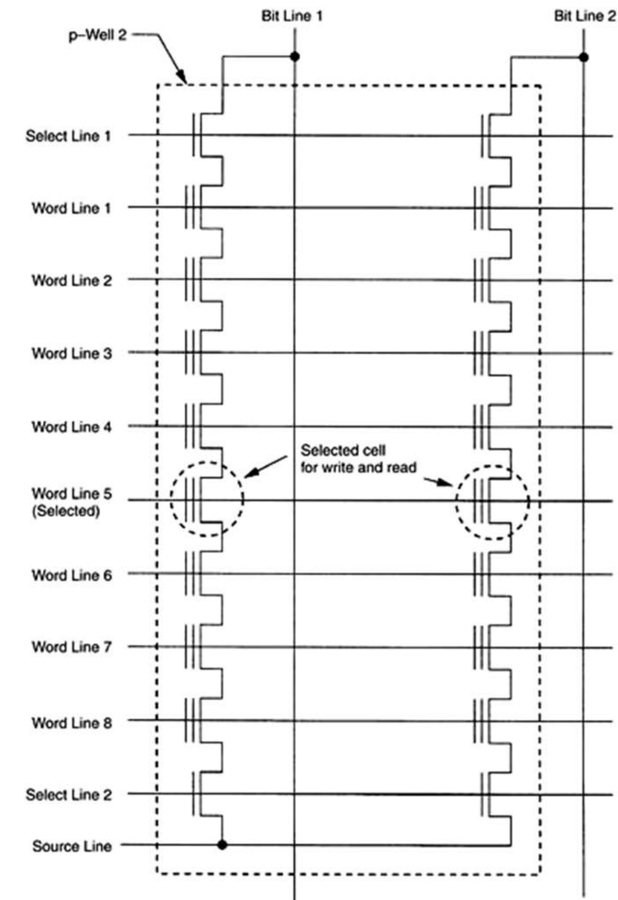
Bias Conditions of NOR Cell

Signal	Operation		
	Erase	Programming	Read
Bit line 1	Open	6V	1V
Bit line 2	Open	0V	0V
Source line	12V	0V	0V
Word line 1	0V	0V	0V
Word line 2	0V	12V	5V
Word line 3	0V	0V	0V

NAND Flash Memory Cell

- Cross-section view and configuration of NAND cells





- F-N tunneling mechanism for erase
- F-N tunneling mechanism for program
- Slower programming and read speed but smaller area than NOR cell structure

Bias Conditions of NAND Cell

Signal	Operation		
	Erase	Programming	Read
Bit line 1	Open	0V	1V
Bit line 2	Open	0V	1V
Select line 1	Open	5V	5V
Word line 1	0V	10V	5V
Word line 2	0V	10V	5V
Word line 3	0V	10V	5V
Word line 4	0V	10V	5V
Word line 5	0V	20V	0V
Word line 6	0V	10V	5V
Word line 7	0V	10V	5V
Word line 8	0V	10V	5V
Select line 2	Open	0V	5V
Source line	Open	0V	0V
p-well 2	20V	0V	0V
n-sub	20V	0V	0V

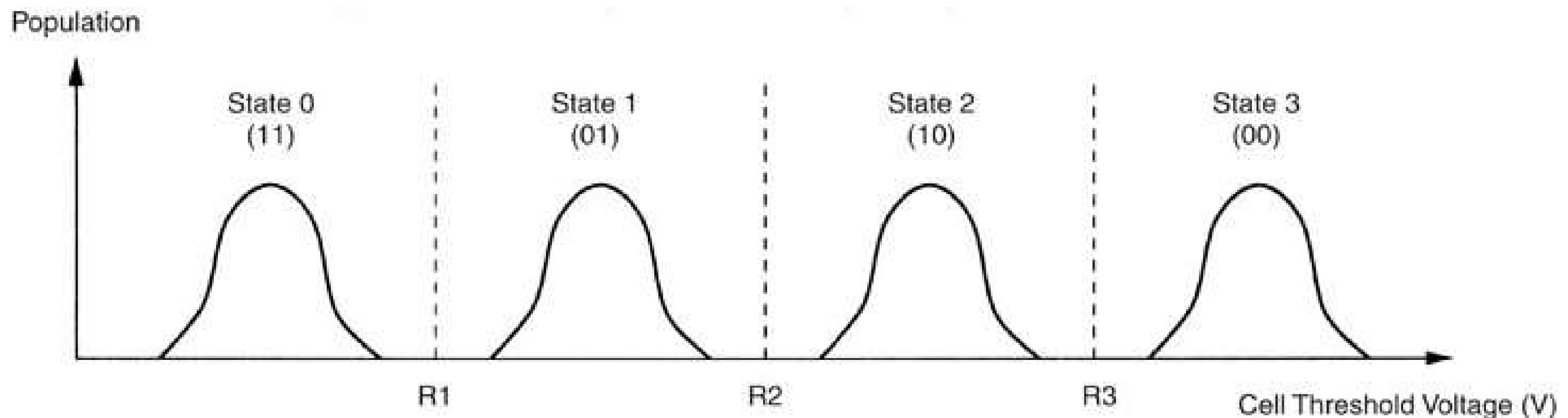
© CMOS Digital Integrated Circuits – 4th Edition

Comparison between NOR and NAND

	NOR	NAND
Erase method	F-N tunneling	F-N tunneling
Programming method	Hot electron injection	F-N tunneling
Erase speed	Slow	Fast
Program speed	Fast	Slow
Read speed	Fast	Slow
Cell size	Large	Small
Scalability	Difficult	Easy
Application	Embedded system	Mass storage

Multilevel Cell Concept

- Effective memory density can be improved
- Possible state number limited by
  - Available charge range
  - Accuracy of programming and read operations
  - Disturbance of state over time



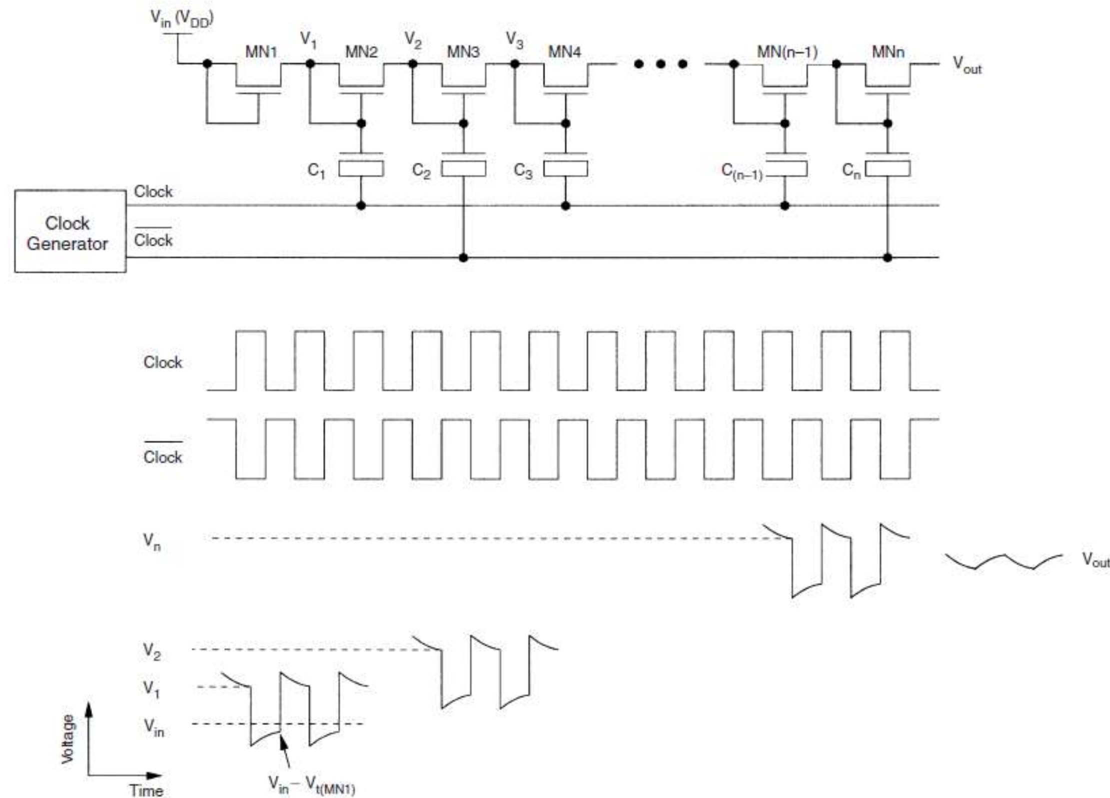
Threshold voltage distribution of 2bits/cell storage

© CMOS Digital Integrated Circuits – 4th Edition

Flash Memory Circuit

- On-chip charge pump used to generate programming voltage
- Chain of diode and cap. to charge or discharge each half cycle

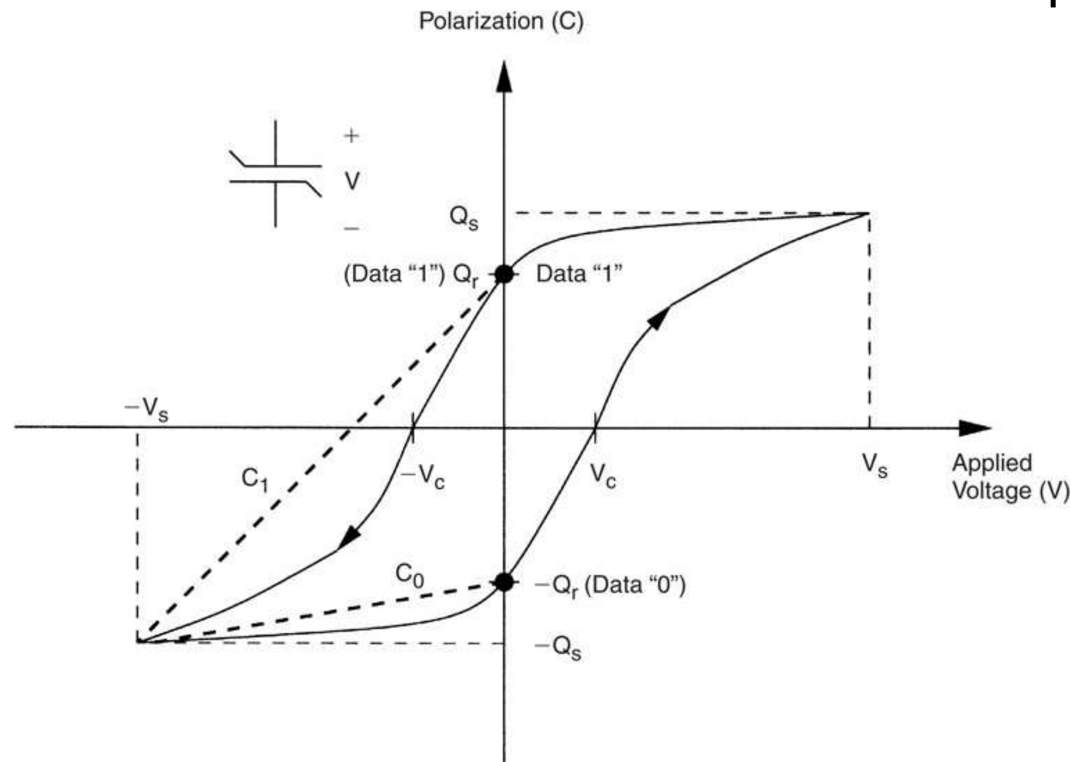
$$V_{out} = V_{in} + (\gamma V_{DD} - V_T(MN1)) + \cdots + (\gamma V_{DD} - V_T(MNn))$$



© CMOS Digital Integrated Circuits – 4th Edition

10.6 Ferroelectric Random Access Memory

- Hysteresis characteristic of a ferroelectric cap.

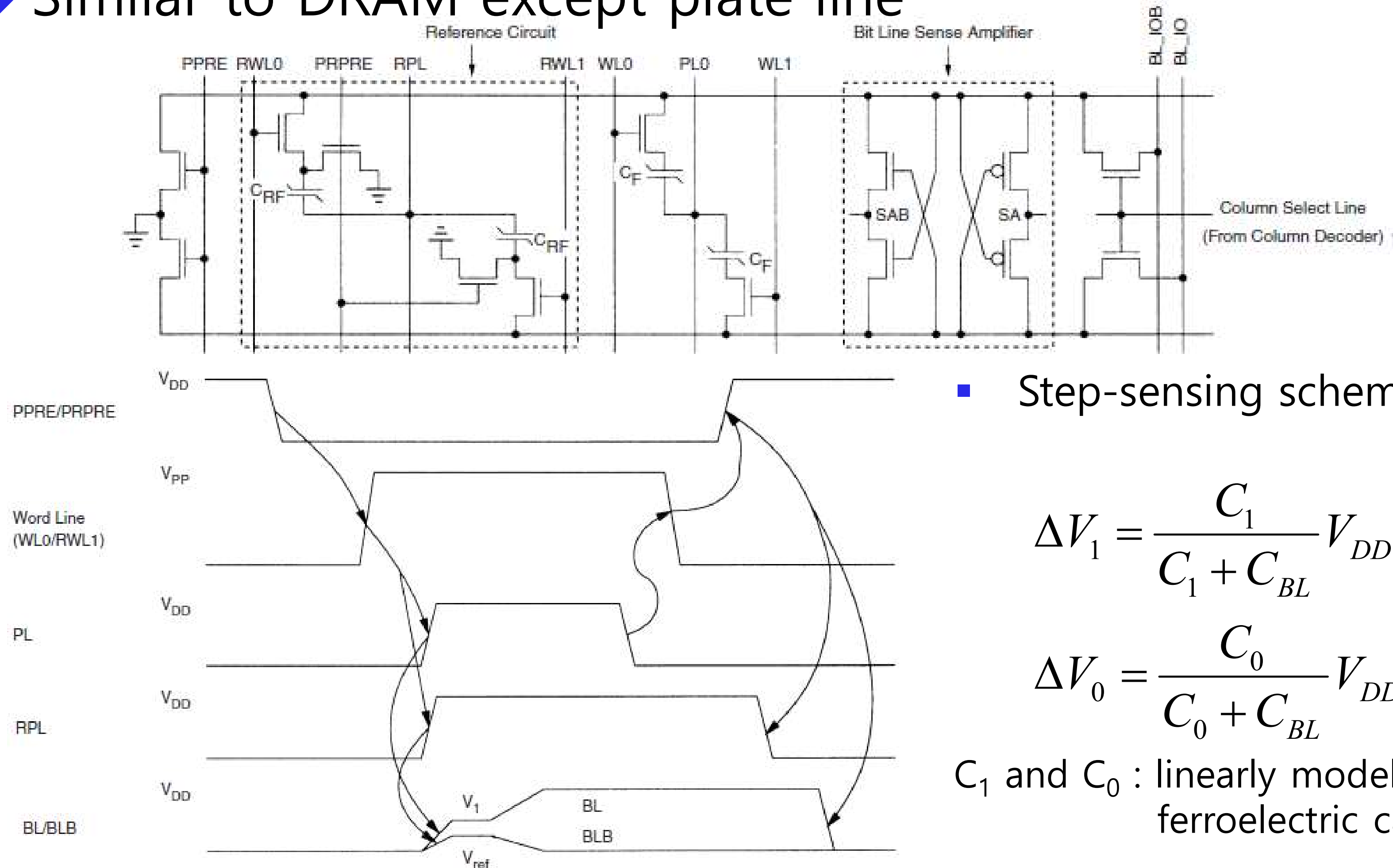


- Total charge varies as function of applied voltage

© CMOS Digital Integrated Circuits – 4th Edition

Structure and Operation of FRAM

- Similar to DRAM except plate line



- Step-sensing scheme

$$\Delta V_1 = \frac{C_1}{C_1 + C_{BL}} V_{DD}$$

$$\Delta V_0 = \frac{C_0}{C_0 + C_{BL}} V_{DD}$$

C_1 and C_0 : linearly modeled ferroelectric cap

© CMOS Digital Integrated Circuits – 4th Edition

Problems of FRAM

- Step-sensing scheme cause reliability issues
  - Pulse sensing scheme also used with read speed penalty
- Fatigue
  - Capacitance charge gradually degraded with repeated use
- Imprint
  - Ferroelectric cap tends to stay at one state preferably when state maintained for a long time