CG2027 Transistor-Level Digital Circuits

Handout #5: Memory

National University of Singapore
Kelvin Fong

NUS National University of Singapore

Lecture Overview

In this lecture, you will learn about

- Memory
  - SRAM
  - DRAM
  - Flash
  - Future memories

Overview

- Memory classification
- Memory architecture
  - Core / Peripheral
- Memory in the market
  - Flash memory (NAND / NOR)
  - SRAM
  - DRAM (EDO, DDR, DDR2, DDR3 / ~ GDDR5)
  - Future memory

MOS Memory Classification

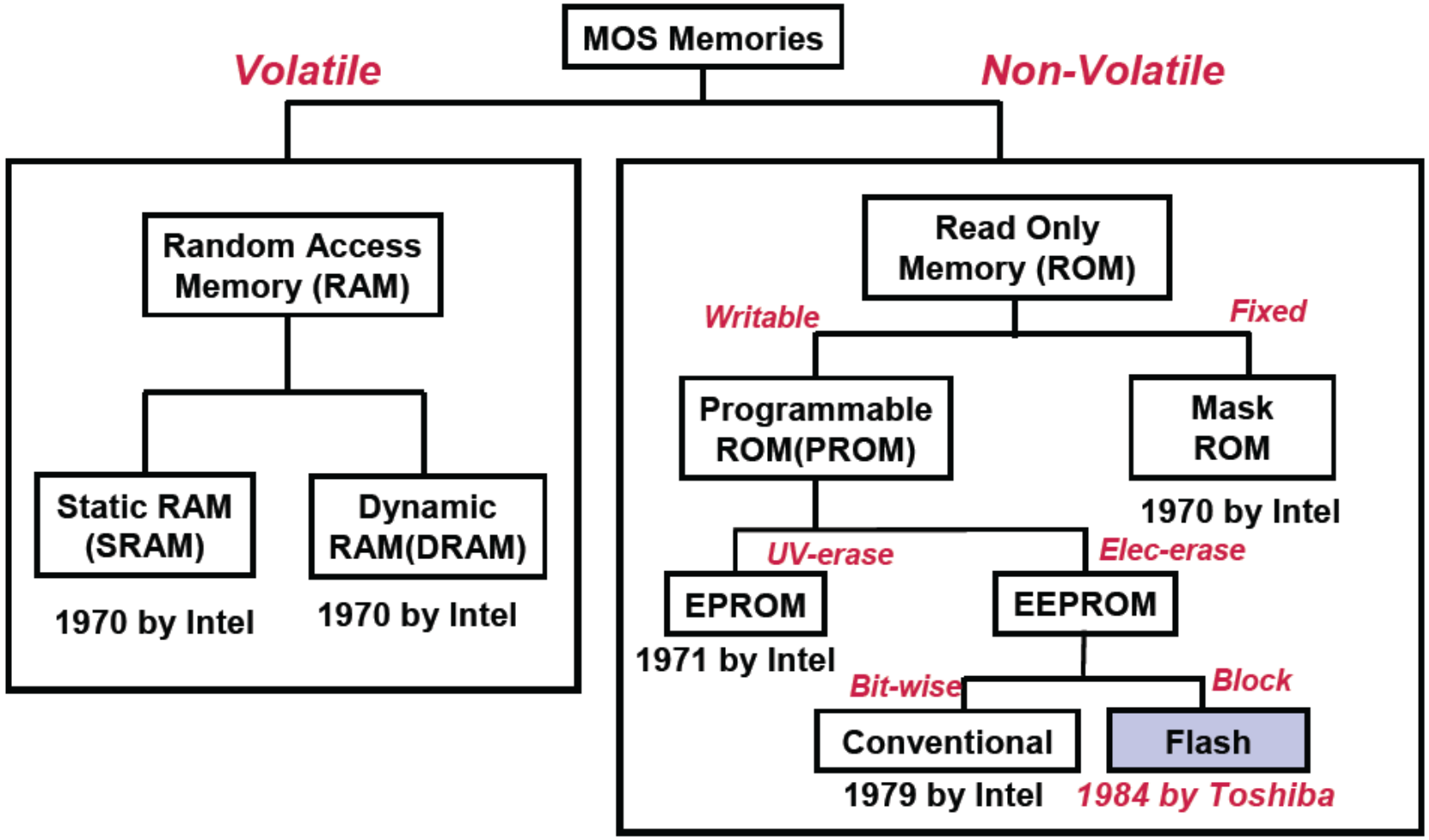


Figure-1 4Q17 Global Branded DRAM Revenue Ranking Unit: Million USD

Ranking	Company	Revenue			Market Share	
		4Q17	3Q17	QoQ	4Q17	3Q17
1	Samsung	10,066	8,790	14.5%	46.0%	45.8%
2	SK Hynix	6,291	5,514	14.1%	28.7%	28.7%
3	Micron Group	4,562	4,023	13.4%	20.8%	21.0%
4	Nanya	558	439	26.9%	2.5%	2.3%
5	Winbond	173	177	-2.2%	0.8%	0.9%
6	Powerchip	104	103	0.6%	0.5%	0.5%
	Others	144	135	7.0%	0.7%	0.7%
	Total	21,898	19,181	14.2%	100.0%	100.0%

Note 1: 3Q17 USD$1: KRW$1,131; US$1: TWD$30.25
Note 2: 3Q17 USD$1: KRW$1,105; US$1: TWD$30.1
Source: DRAMeXchange, Feb., 2018

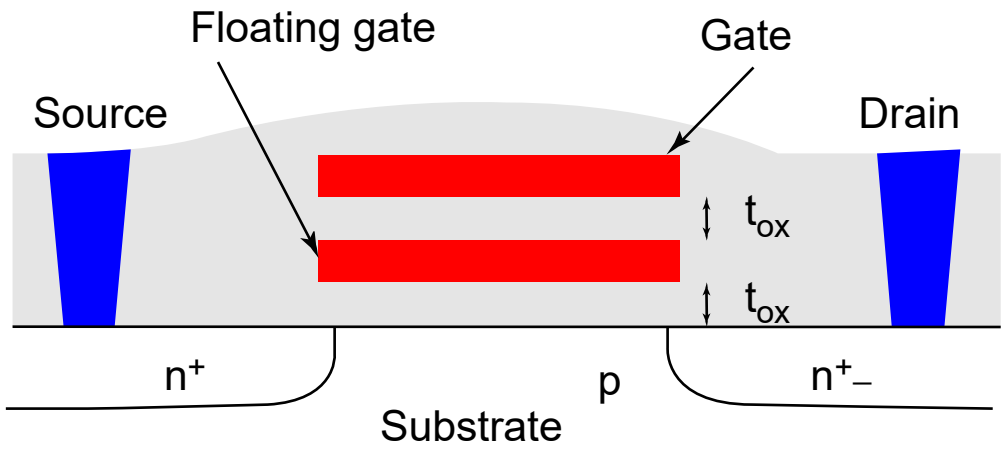
Table: 4Q17 Revenue Ranking of Branded NAND Flash Makers

Company	Revenue (US$M)		Market Share(%)	
	4Q17	QoQ (%)	4Q17	3Q17
Samsung	6,169.6	9.8%	38.0%	37.0%
Toshiba	2,779.7	1.4%	17.1%	18.0%
WDC	2,616.8	3.7%	16.1%	16.6%
Micron	1,866.0	1.5%	11.5%	12.1%
SK Hynix	1,797.5	19.5%	11.1%	9.9%
Intel	889.0	-0.2%	5.5%	5.9%
Others	116.1	30.9%	0.7%	0.6%
Total	16,234.6	6.8%	--	--

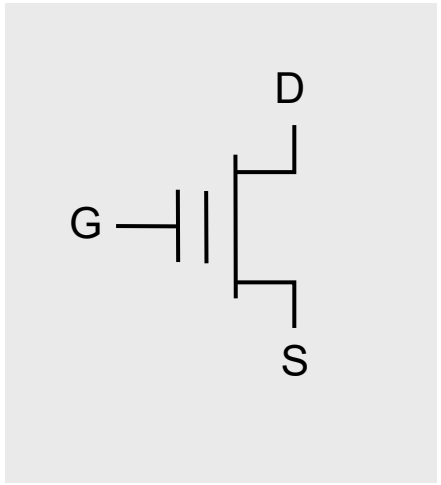
Note 1. 4Q17 USD/JPY= 1:112.9; USD/KRW=1:1,105.0
Note 2. 3Q17 USD/JPY= 1:111.0; USD/KRW=1:1,131.2
Note 3. Non-captive, license and royalty are excluded from Western Digital's revenue calculation.
Source: DRAMeXchange, Feb., 2018

FLASH MEMORY

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

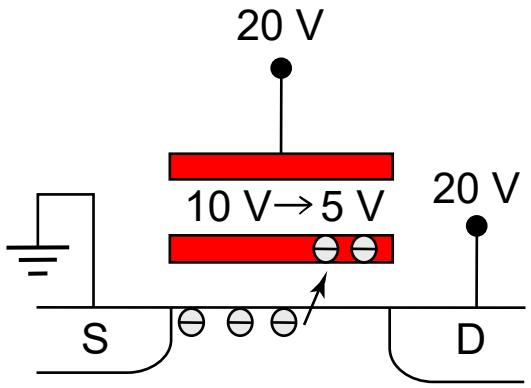


Device cross-section

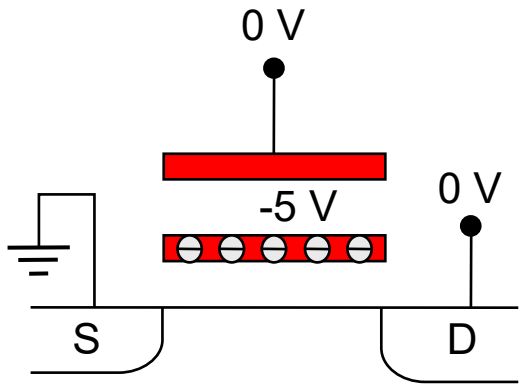


Schematic symbol

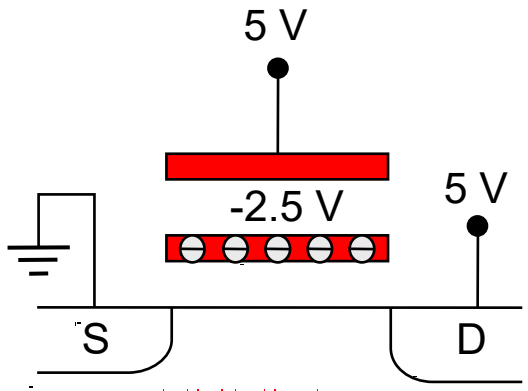
Floating-Gate Transistor Programming



Avalanche injection Channel Hot Electron (CHE)

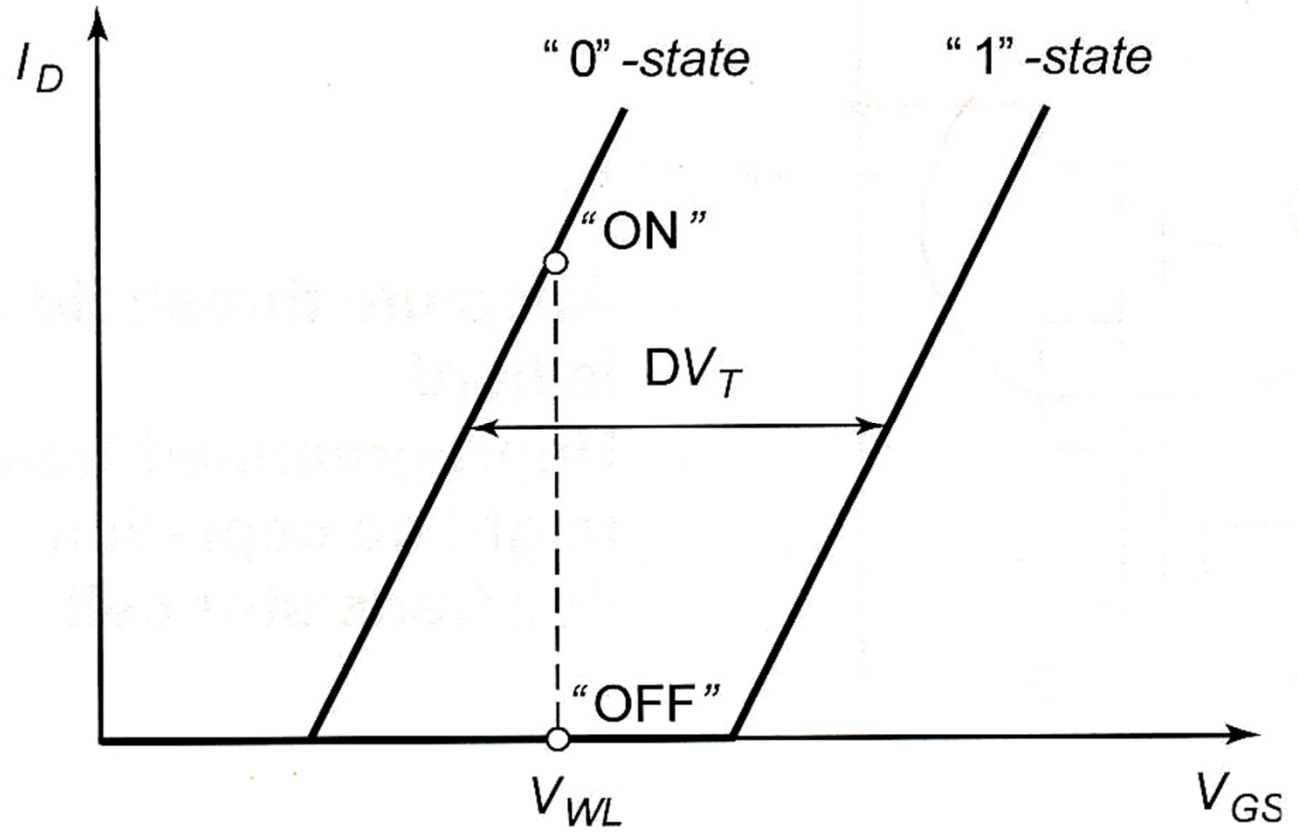


Removing programming voltage leaves charge trapped

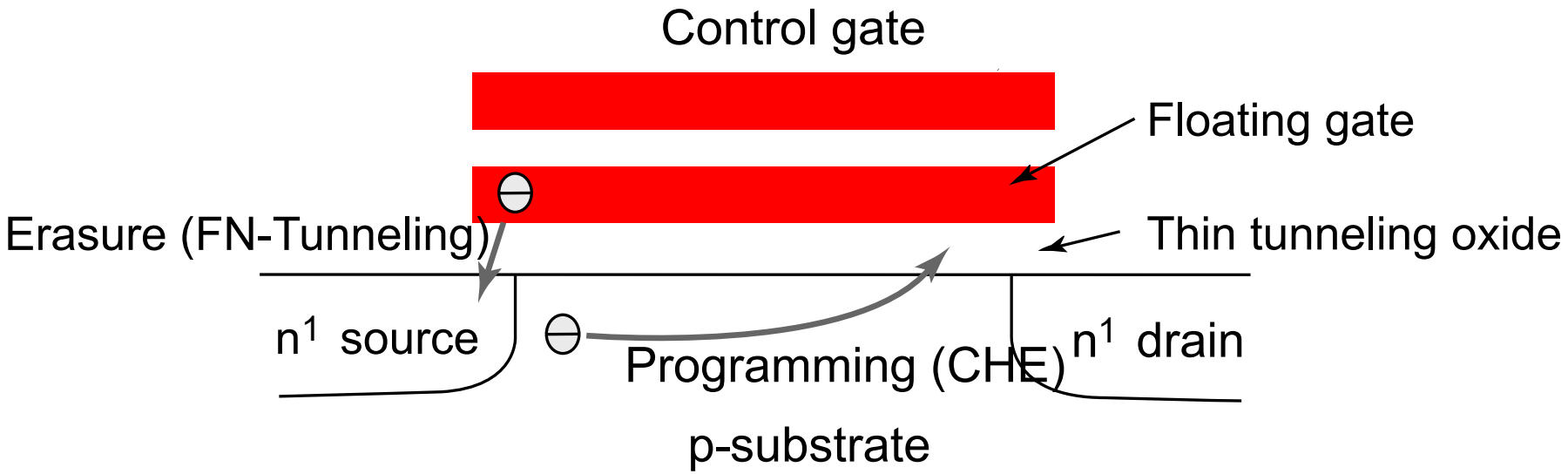


Programming results in higher V_T.

A “Programmable-Threshold” Transistor



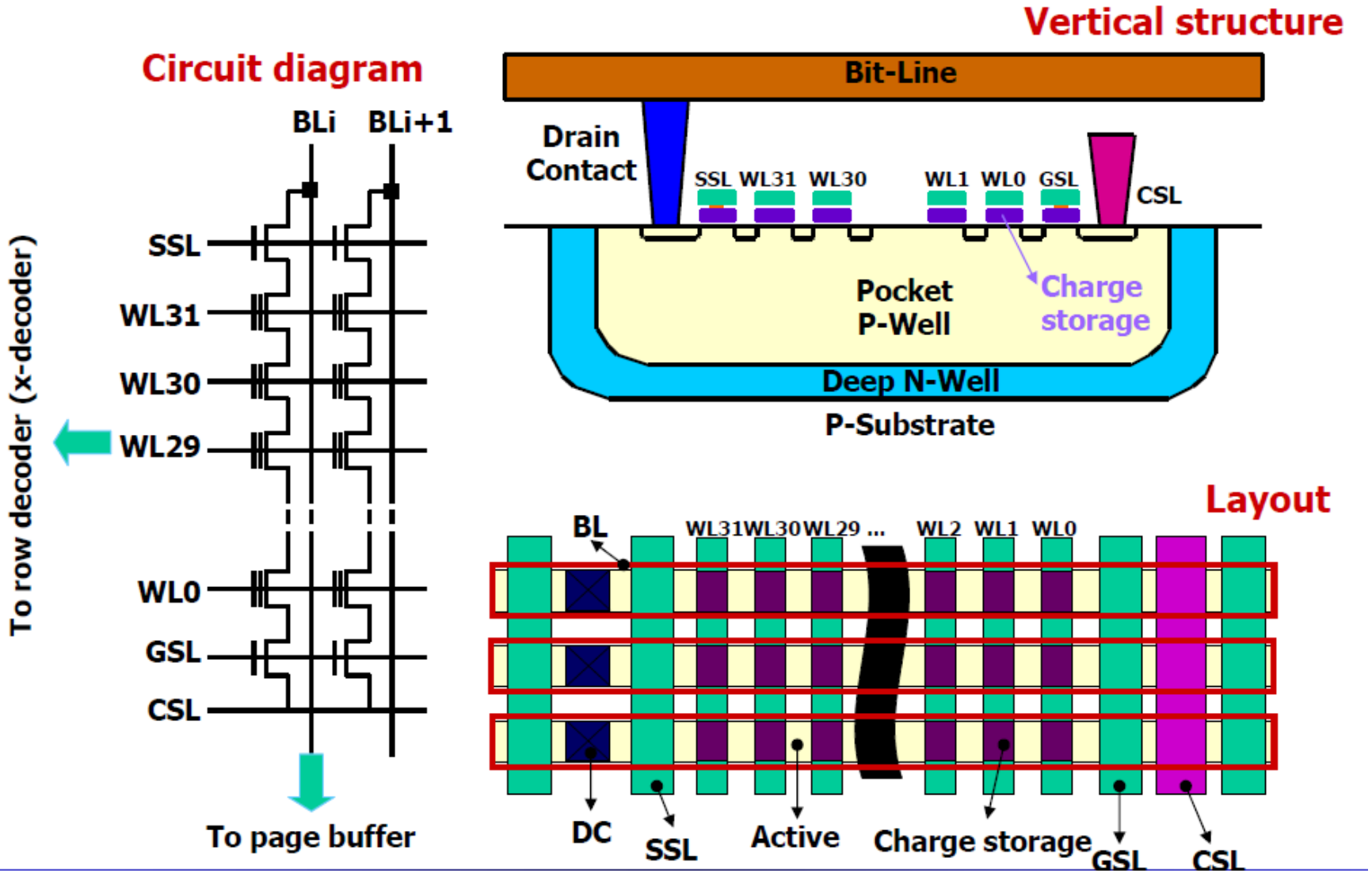
Flash EEPROM



Classical case of Channel Hot Electrons (CHE) and FN-Tunneling
Many other options …

Q) How do we store a new data into a pre-programmed Floating-Gate Transistor?

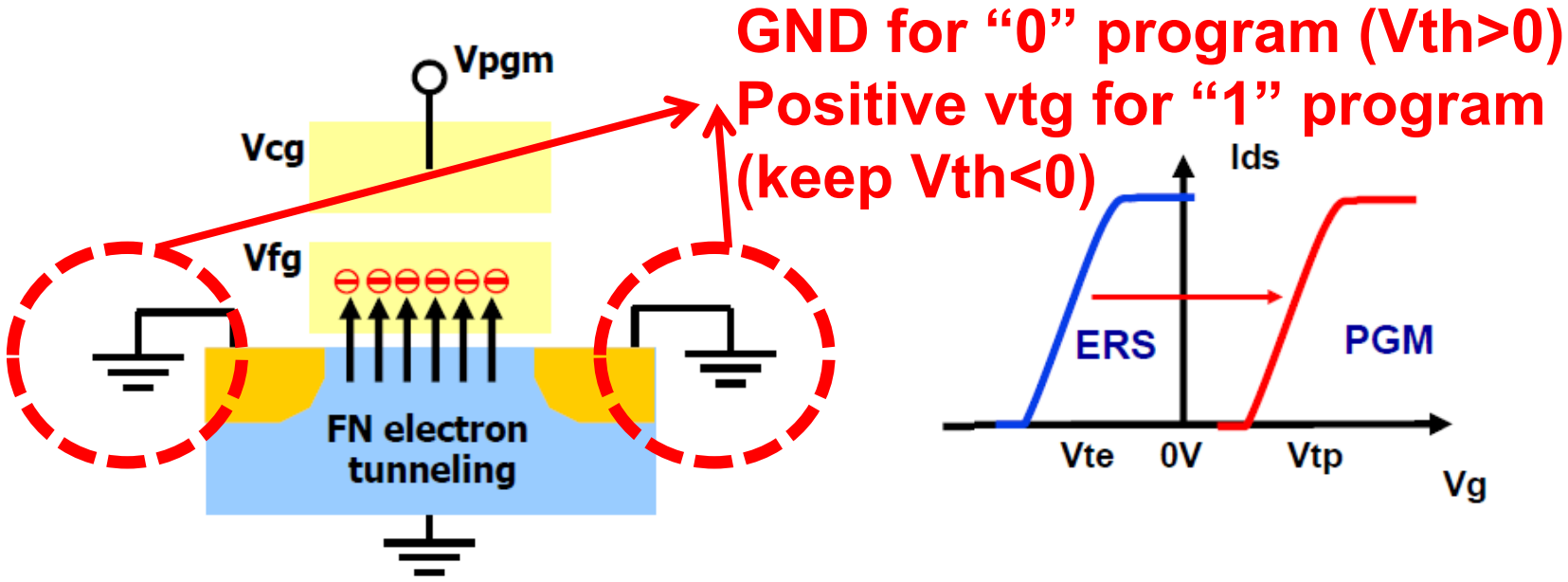
NAND Flash Memory



Courtesy Samsung

NAND Flash: Write Operation

- Inject electrons to the floating gate ➔ Programs Vth

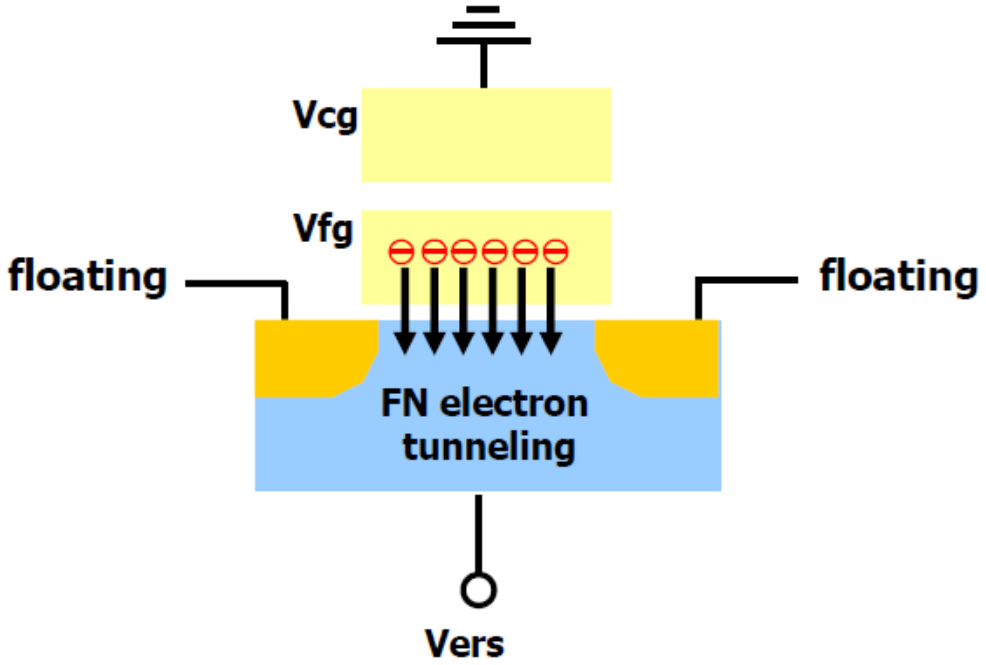


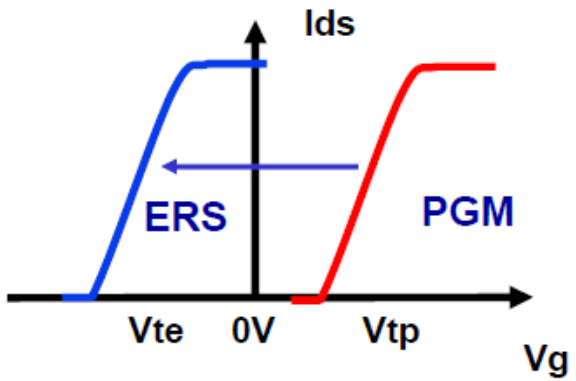
(Note) All cells within a "page" is programmed simultaneously!

NAND Flash: Erase Operation

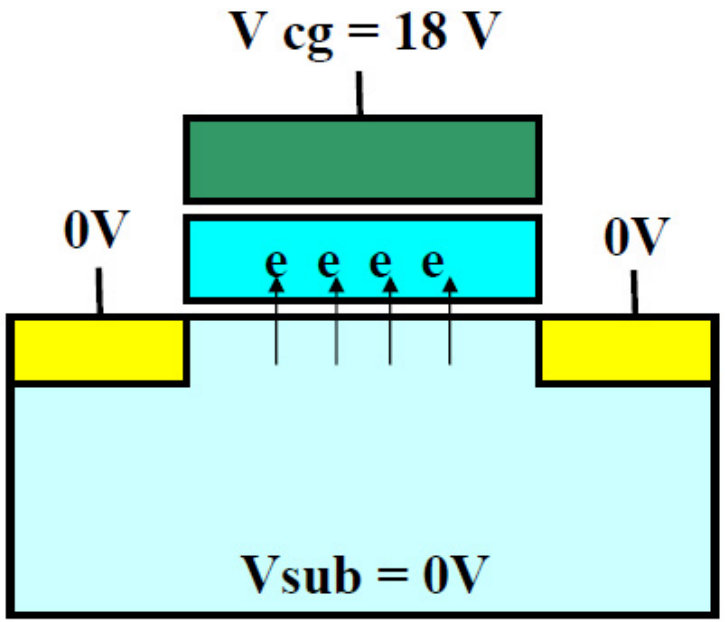
- Remove electrons from the floating gate

  ➔ Lowers Vth



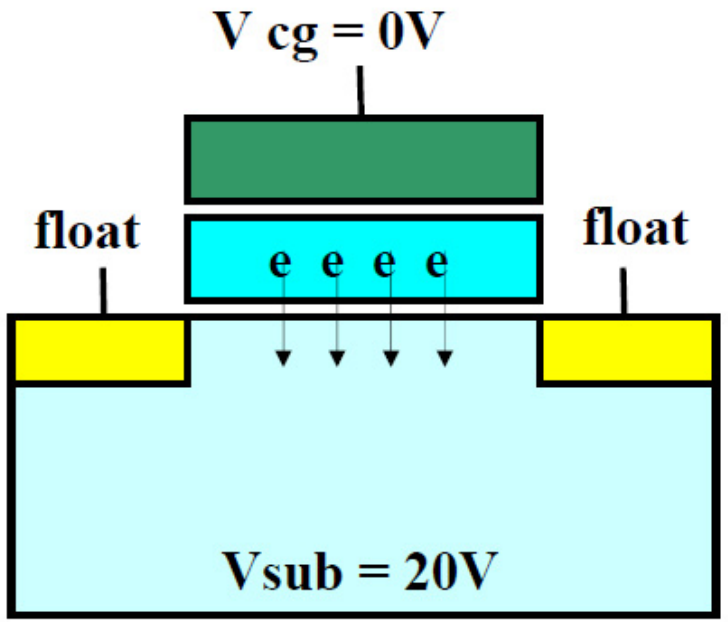


Summary: NAND Program/Erase



Program
F-N Tunneling

Off cell
(Solid-0)

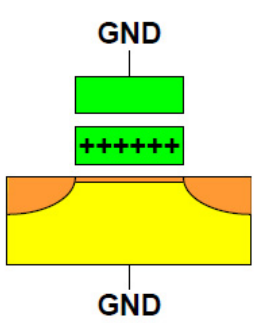


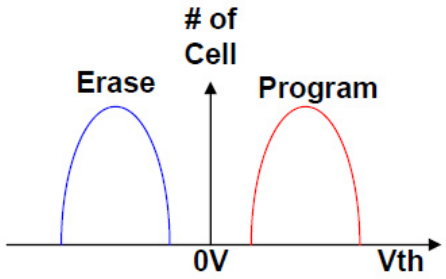
Erase
F-N Tunneling

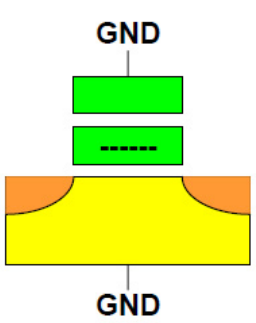
On cell
(Solid-1)

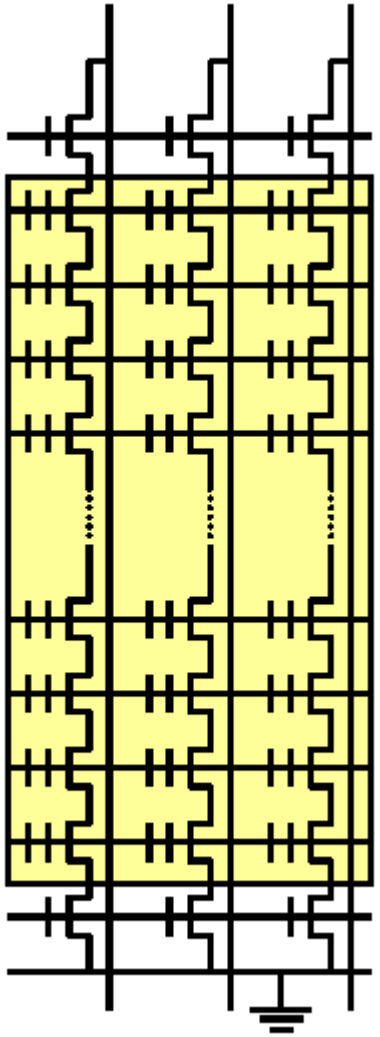
Note: NAND Erase

- Erase by “Block” (e.g., 1Mb)
  - Reverse biased body (high V)
  - All WL set to 0V
- Program: Vth >0
- Erase: Vth < 0

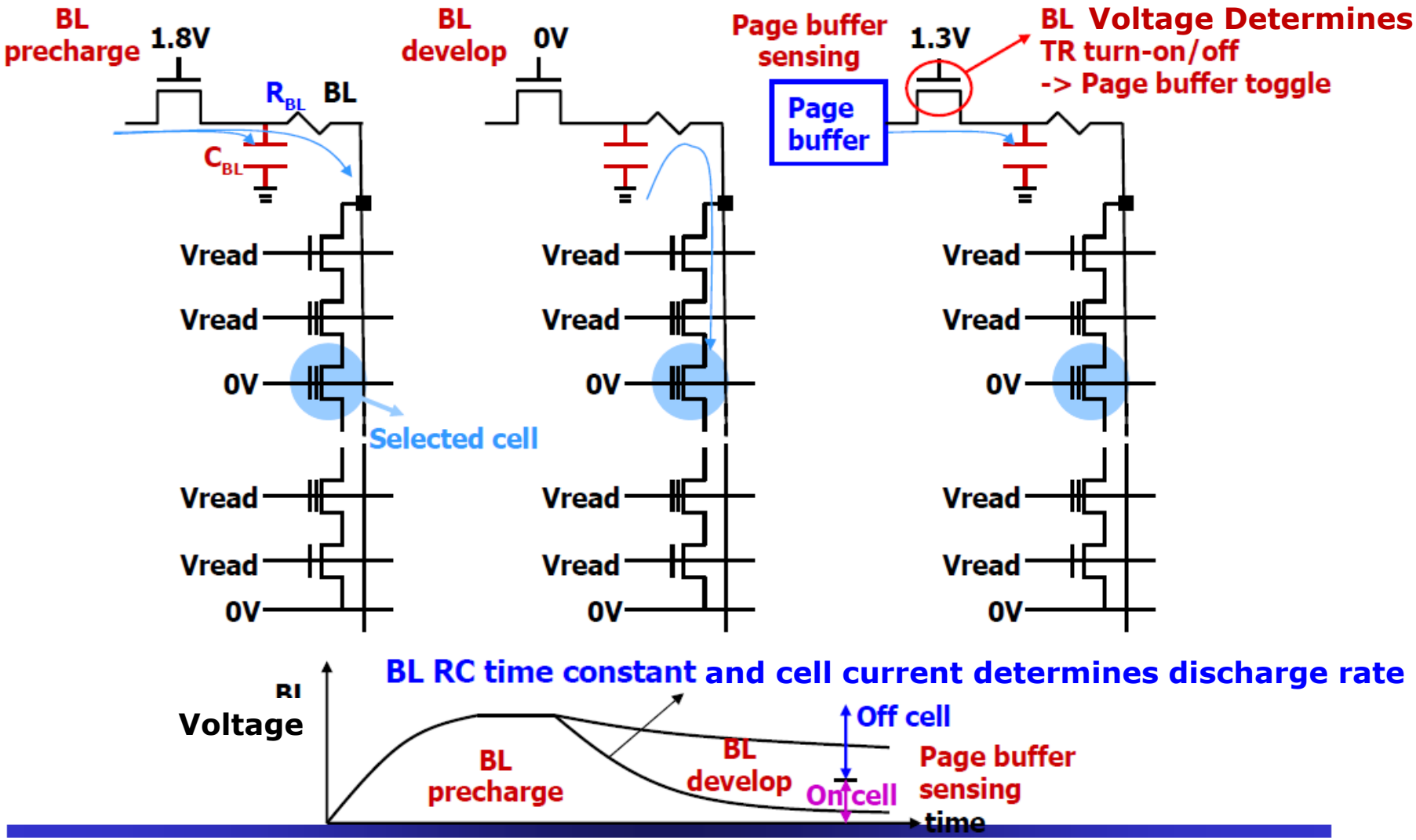






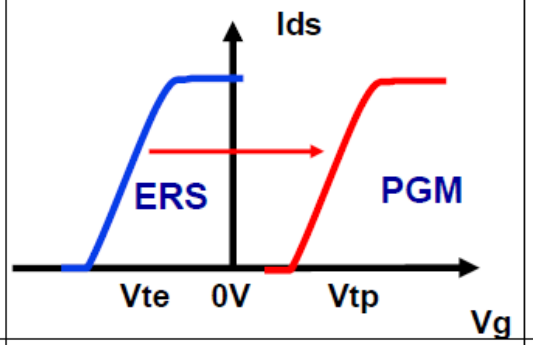
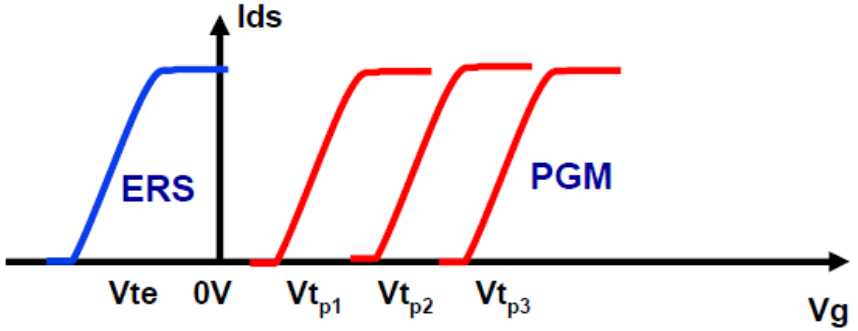
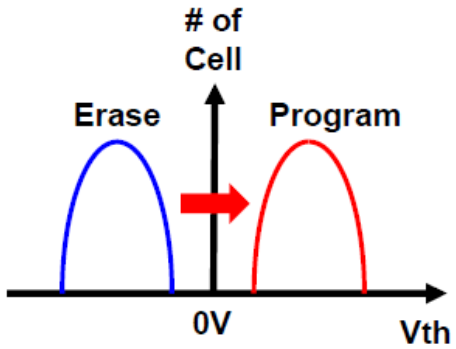
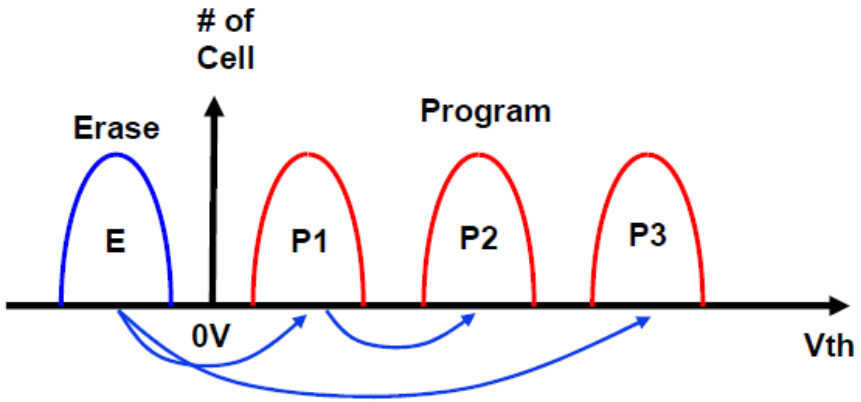


NAND Read Sequence



Courtesy Samsung

SLC vs. MLC

	SLC (Single Level Cell)	MLC (Multi Level Cell)
Vg-Id		
Vth Dist.		
Note	1 bit per cell	Multiple bits per cell

SRAM / DRAM

Read-Write Memories (RAM)

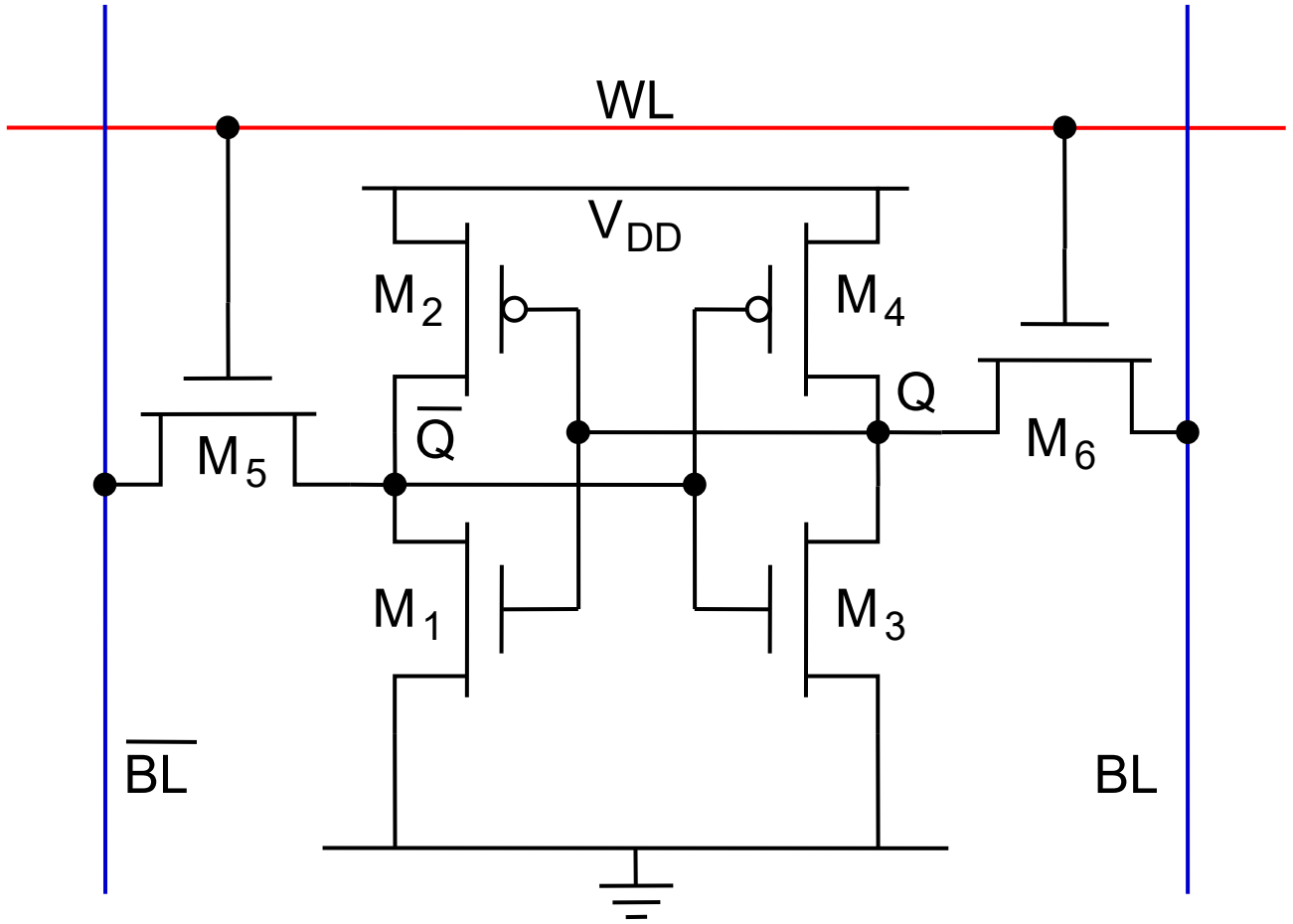
- STATIC (SRAM)

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

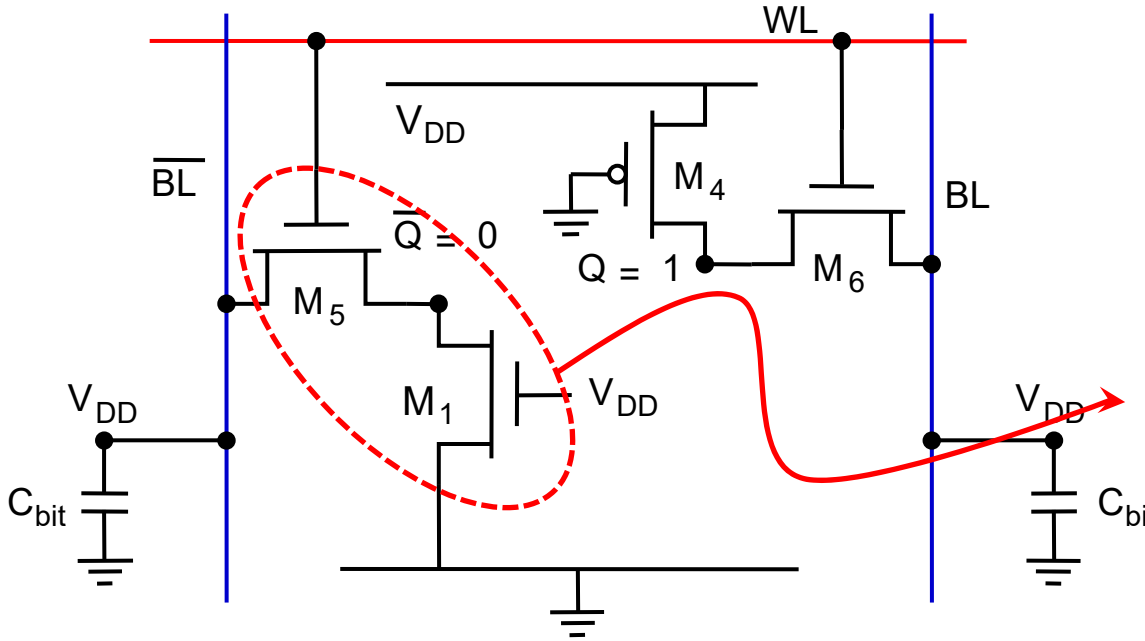
- DYNAMIC (DRAM)

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

6-transistor CMOS SRAM Cell



CMOS SRAM Analysis (Read)



Assume V_{BL}, $V_{/BL}$ precharged to VDD

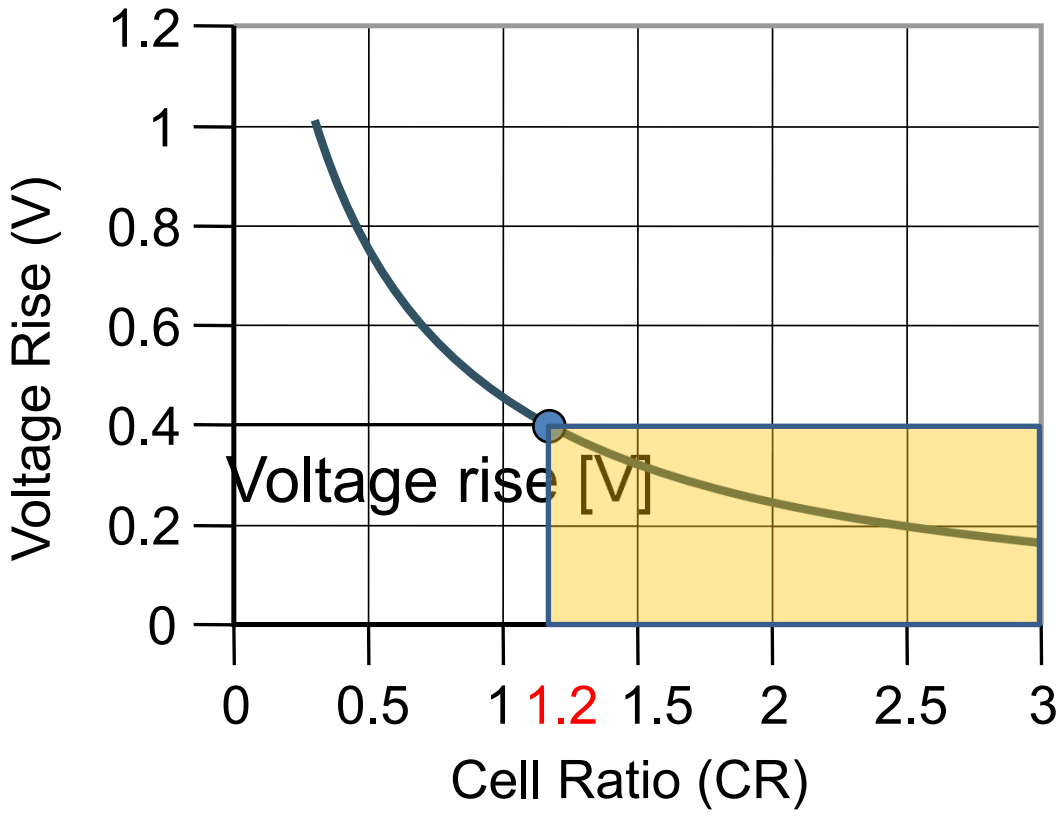
$V_{/Q} < V_{th,NMOS,M3}$ to prevent accidental pull down of V_Q

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

$$\Delta V=\frac{V_{DSATn}+CR(V_{DD}-V_{Tn})-\sqrt{V_{DSATn}^2(1+CR)+CR^2(V_{DD}-V_{Tn})^2}}{CR}$$

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

CMOS SRAM Analysis (Read)

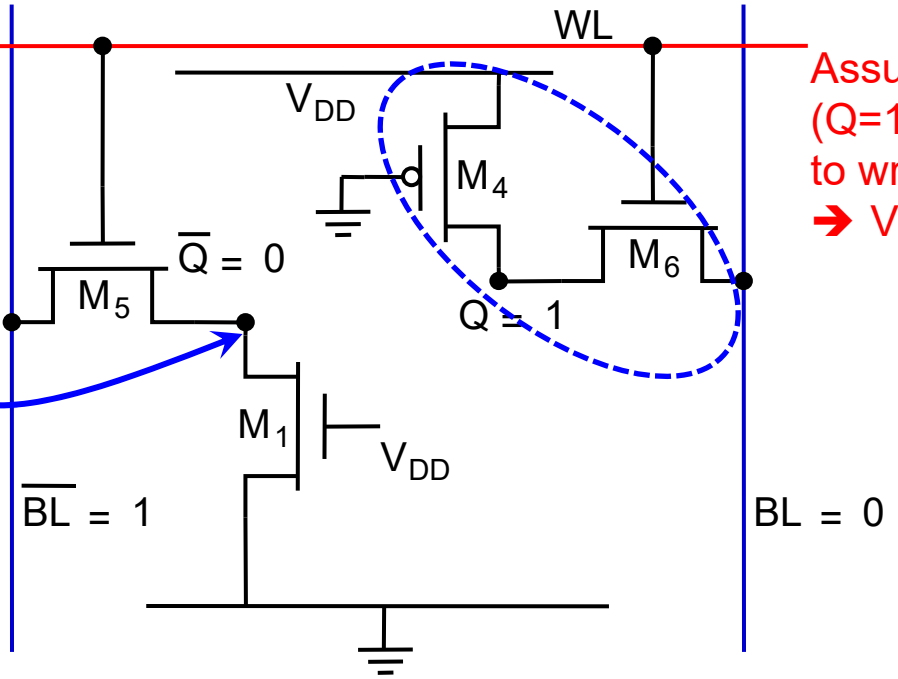


$$CR = \frac{W_1/L_1}{W_5/L_5}$$

CR > 1.2 to ensure read stability
(otherwise, bit flip occurs while read operation)

CMOS SRAM Analysis (Write)

From read analysis, we know $V_{/Q}$<0.4(V)
➔ Write of "0" should be initiated by M4/M6 pairs.



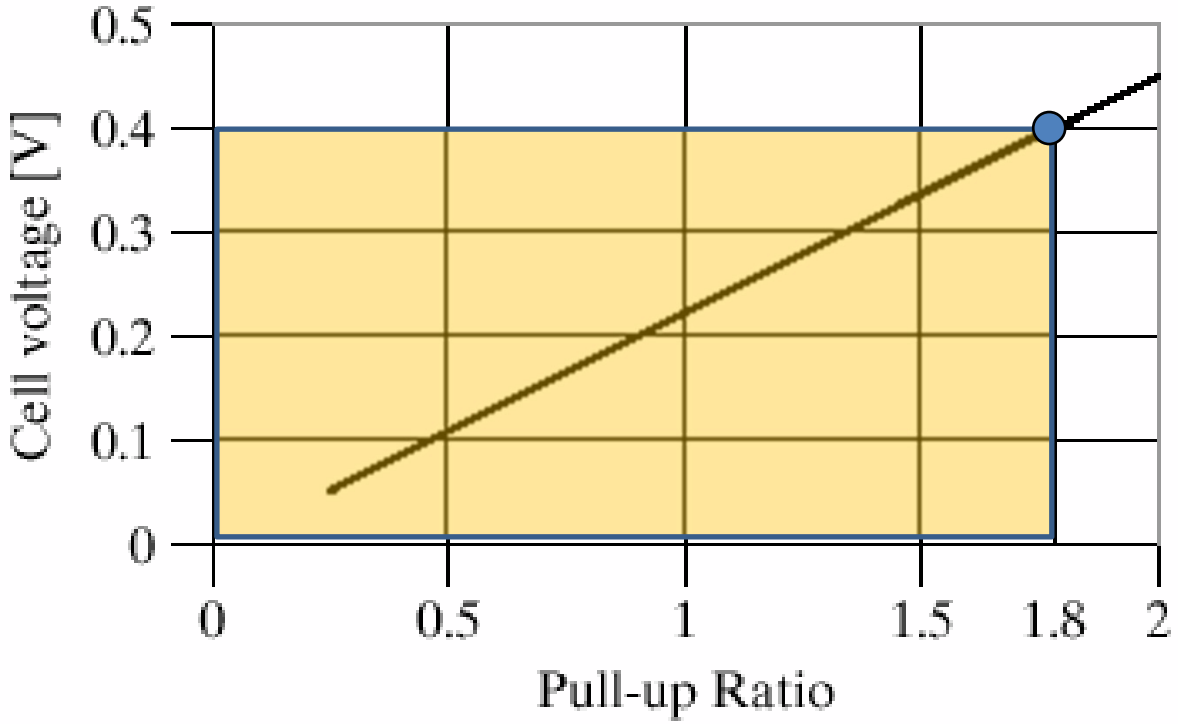
Assume 1 is stored (Q=1) and we want to write "0".
➔ V_{BL},=0, $V_{/BL}$=1

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

$$V_Q=V_{DD}-V_{Tn}-\sqrt{(V_{DD}-V_{Tn})^2-2\frac{\mu_p}{\mu_n}PR\left((V_{DD}-|V_{Tp}|)V_{DSATp}-\frac{V_{DSATp}^2}{2}\right)},\qquad PR=\frac{W_4/L_4}{W_6/L_6}$$

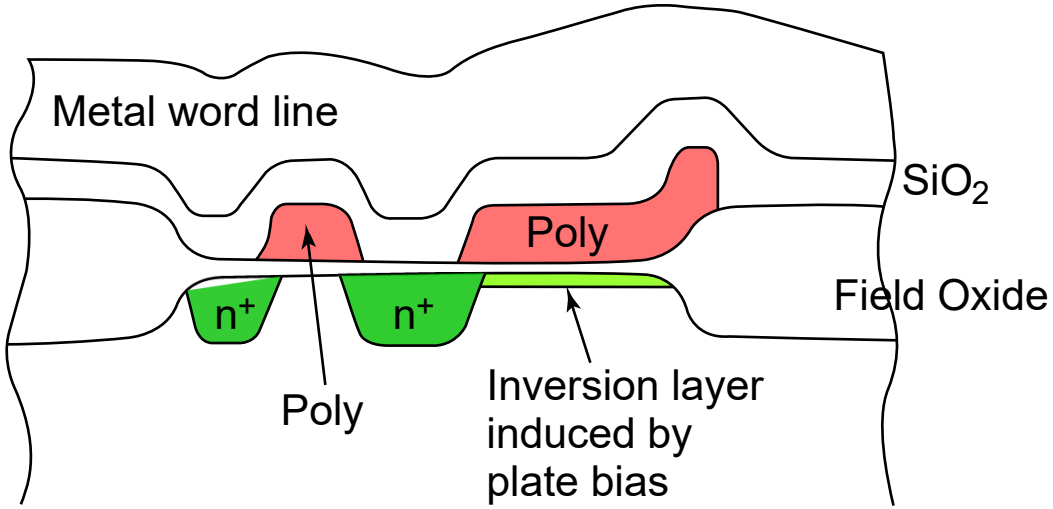
➔V_Q should be lower than $V_{th,M1}$
(otherwise, M1 is kept on, and 0 cannot be written to Q)

CMOS SRAM Analysis (Write)

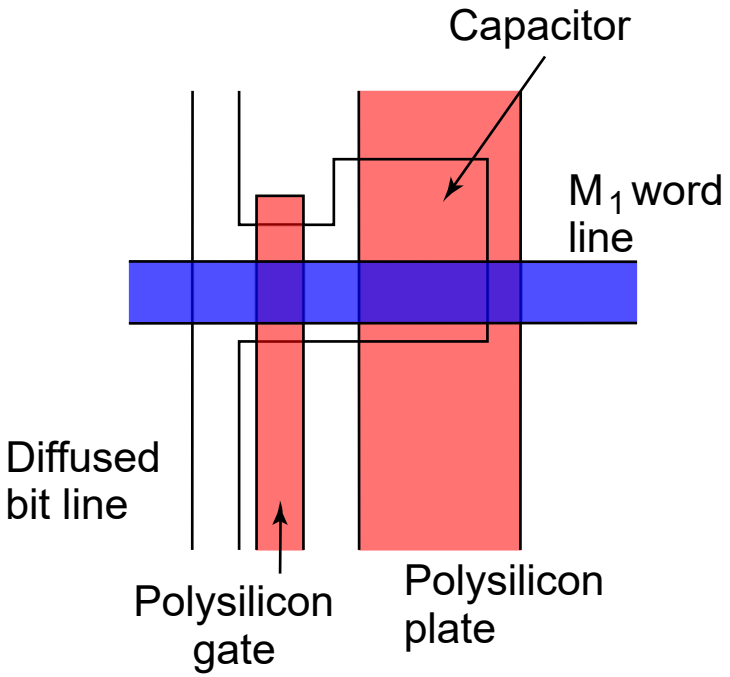


PR < 1.8 to ensure write of “0”
(otherwise, M1 stays on and we cannot write “0” into Q)

1-T DRAM Cell



Cross-section

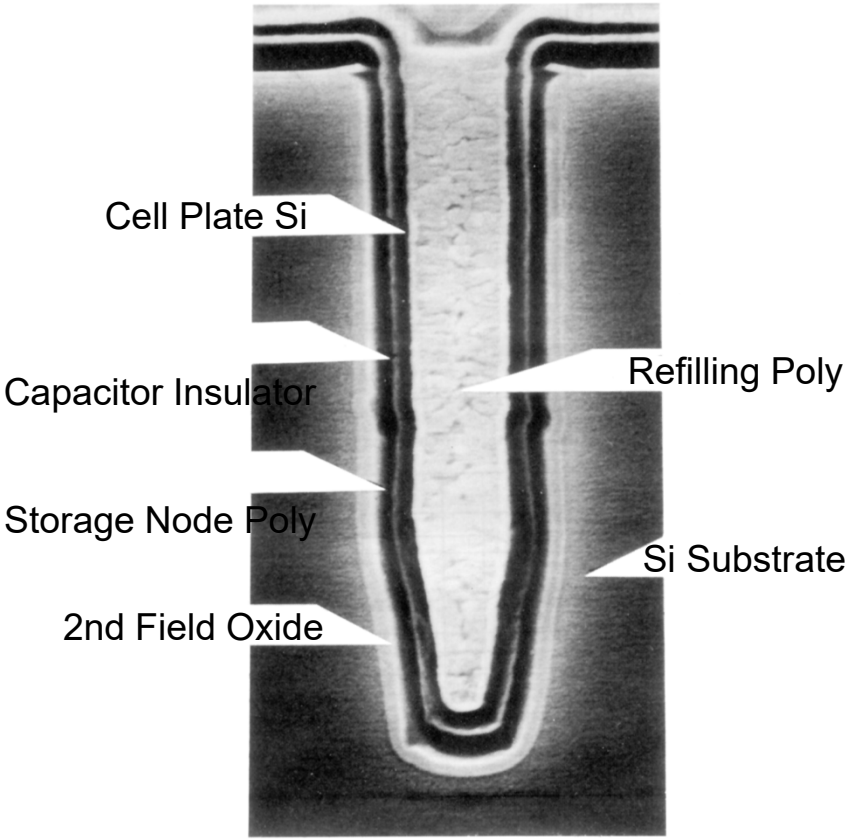


Layout

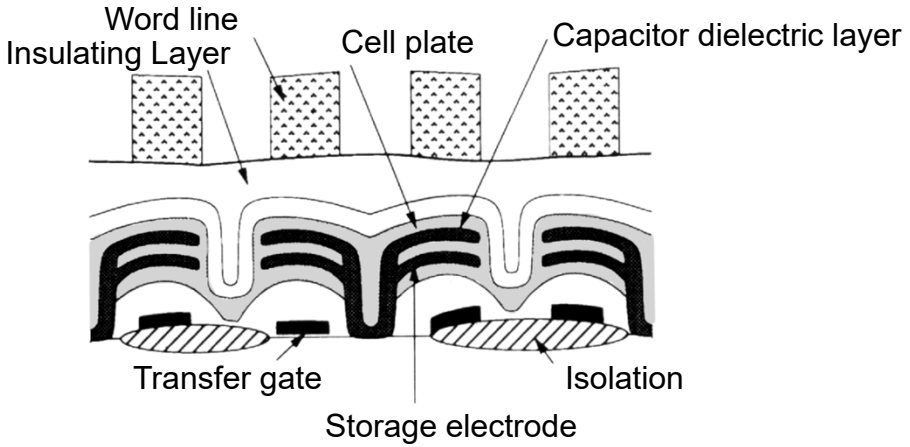
Uses Polysilicon-Diffusion Capacitance

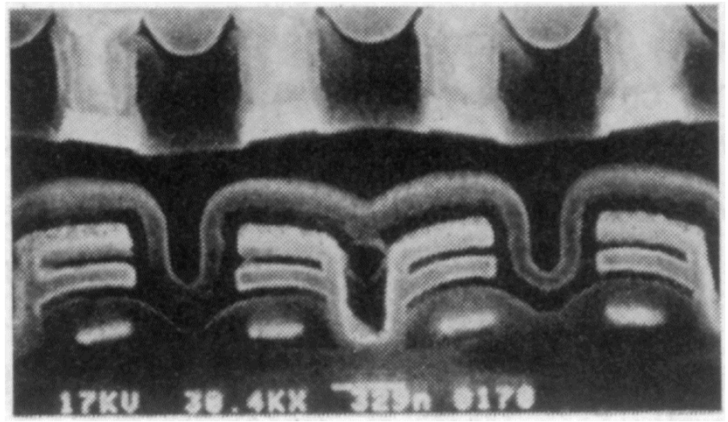
Expensive in Area

Advanced 1T DRAM Cells



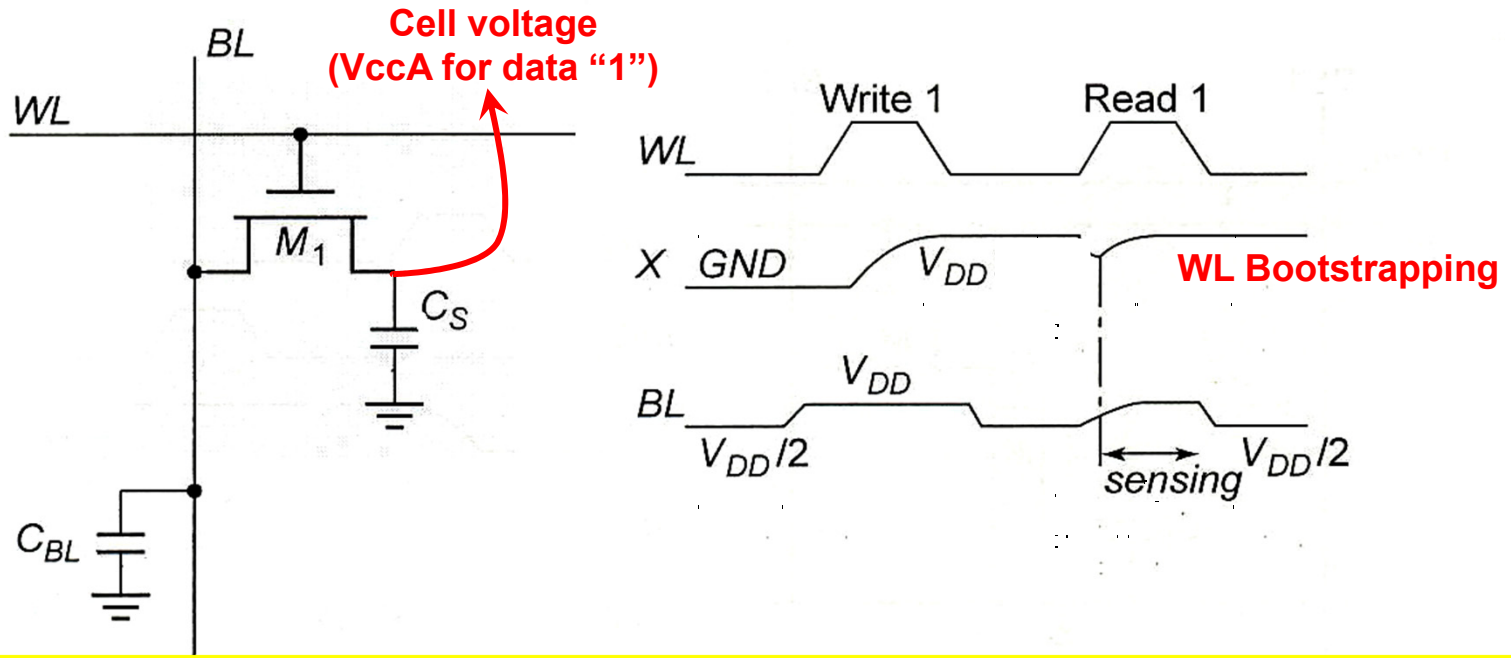
Trench Cell





Stacked-capacitor Cell

1-Transistor DRAM Cell

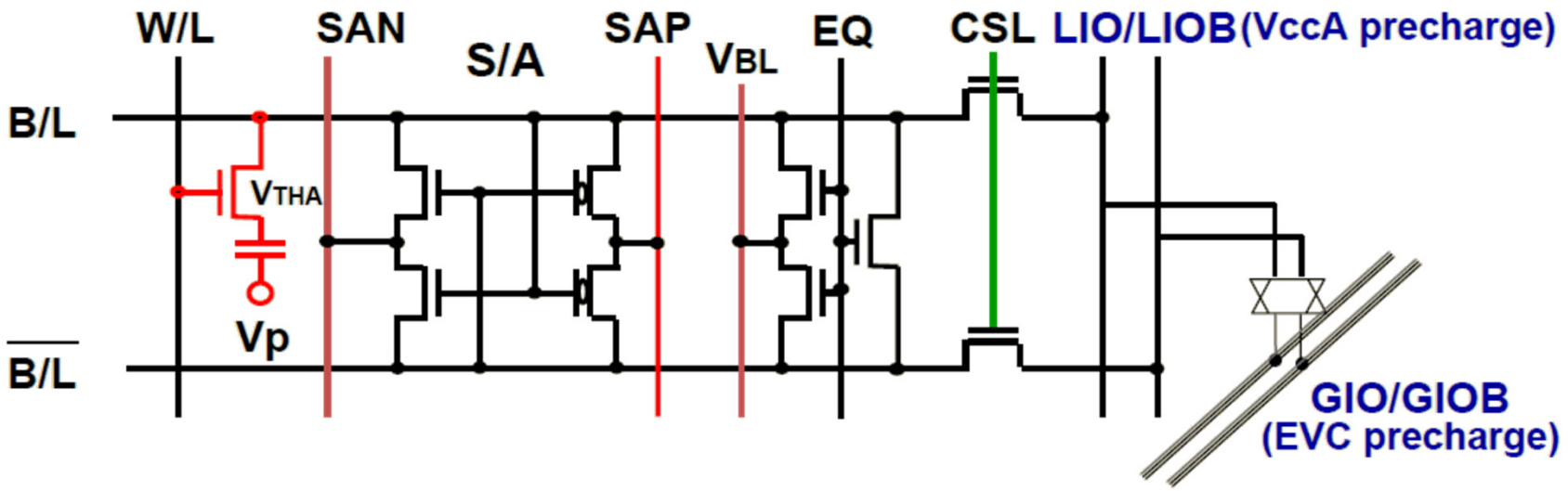


If VCs was initially charged up to VccA (data "1"),

$$V_{Final}\,(C_{BL}+C_S) = C_{BL}VccA/2 + C_SVccA$$

$$\rightarrow \Delta V = V_{Final} - 1/2VccA = \frac{VccA/2}{1+CBL/CS}$$

DRAM Capacitance Ratio



Parasitic Capacitance Example

	Capacitance	Ratio
Cell, Cs	25 fF	1
Bit Line, Cb	100 fF	4
LIO Line	200 fF	8
GIO Line	1.5 pF	60

Charge Sharing

Stored Charge : $VccA \cdot Cs$

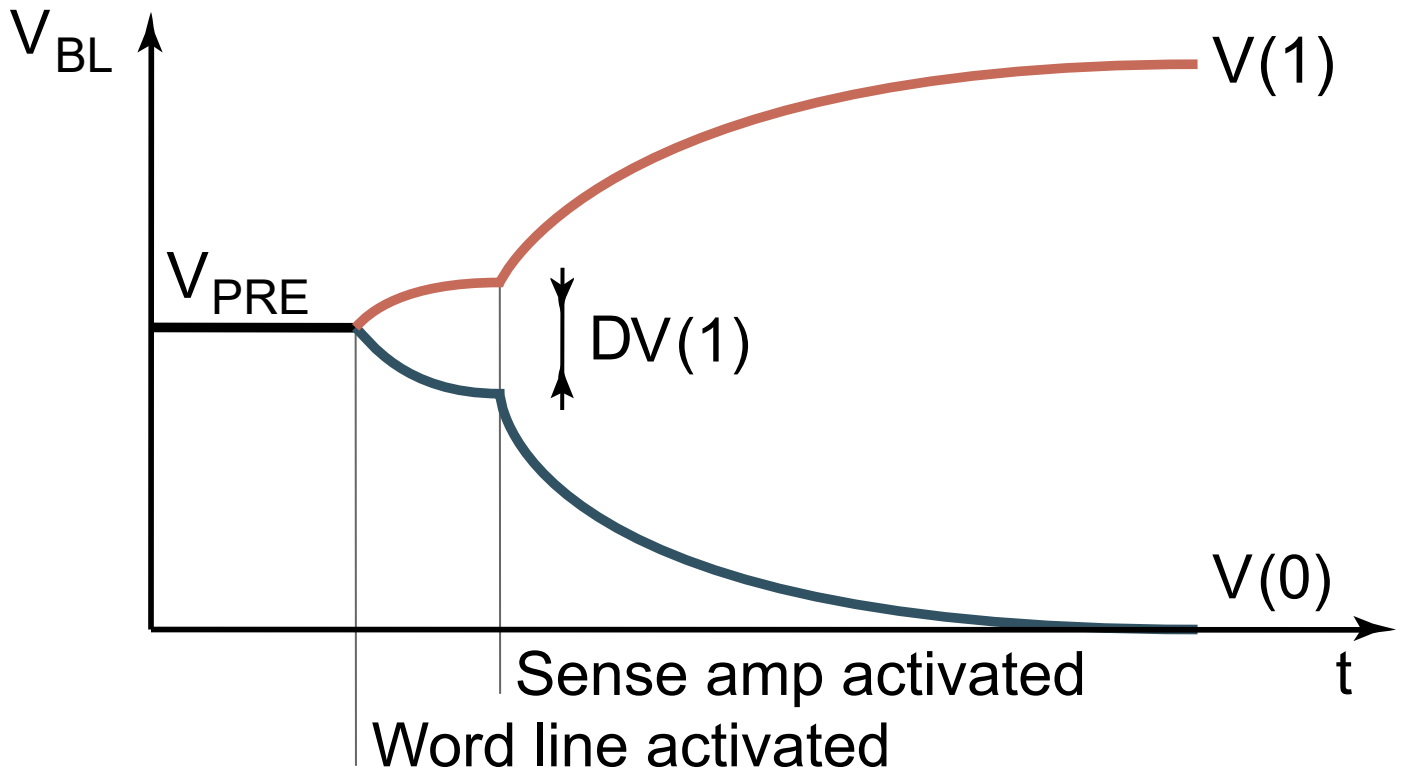
$$\Delta VBL : VccA/2 \cdot \frac{Cs}{Cb + Cs}$$

Example, $\Delta VBL = 150mV$
for $VccA = 1.5V$

DRAM Cell Observations

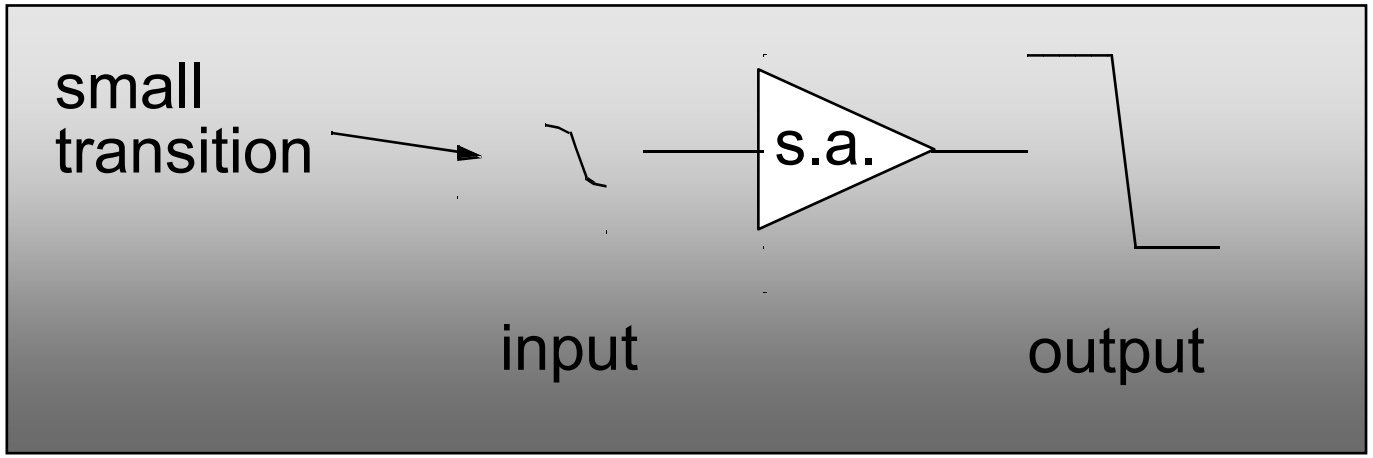
- 1T DRAM requires a sense amplifier for each bit line, due to charge redistribution read-out.
- DRAM memory cells are single ended in contrast to SRAM cells.
- The read-out of the 1T DRAM cell is destructive; read and refresh operations are necessary for correct operation.
- When writing a “1” into a DRAM cell, a threshold voltage is lost. This charge loss can be circumvented by bootstrapping the word lines to a higher value than V_{DD}

Sense Amp Operation

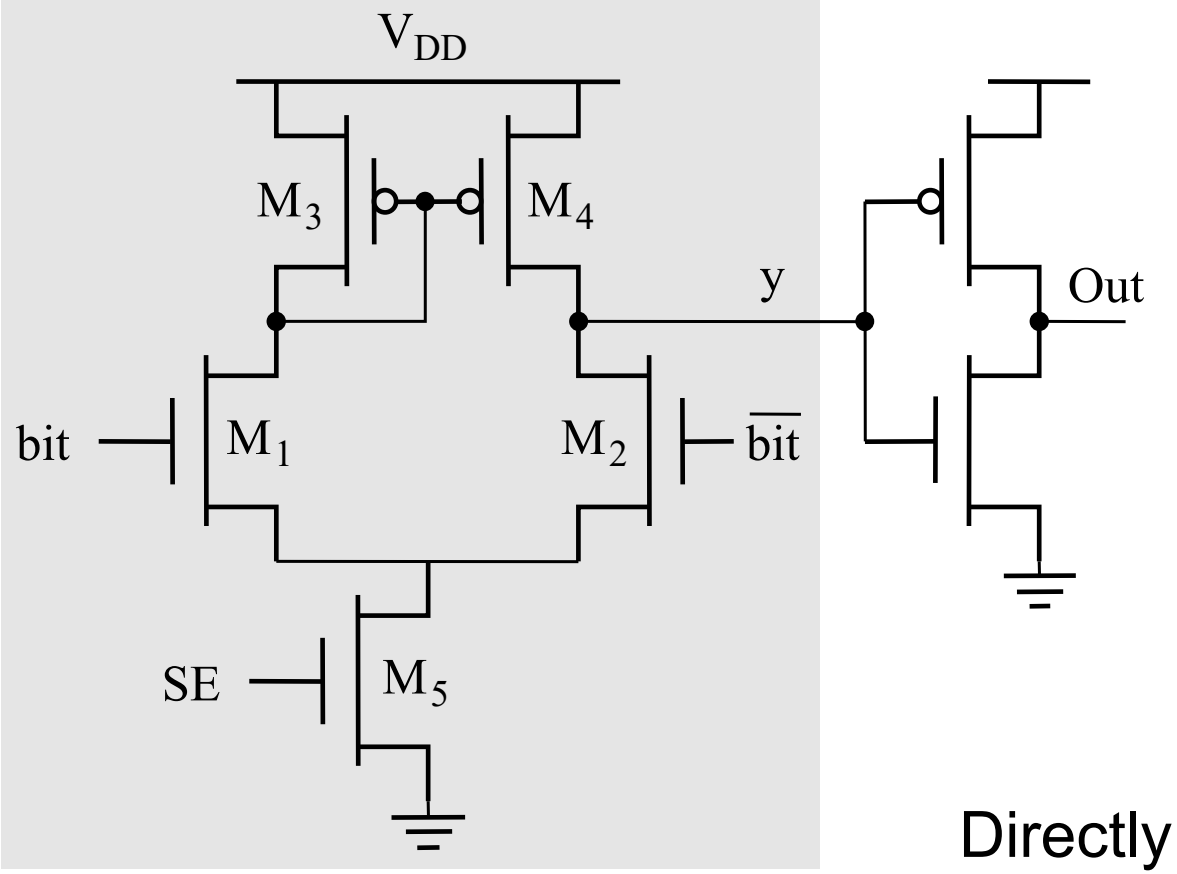


Sense Amplifiers

Idea: Use Sense Amplifer

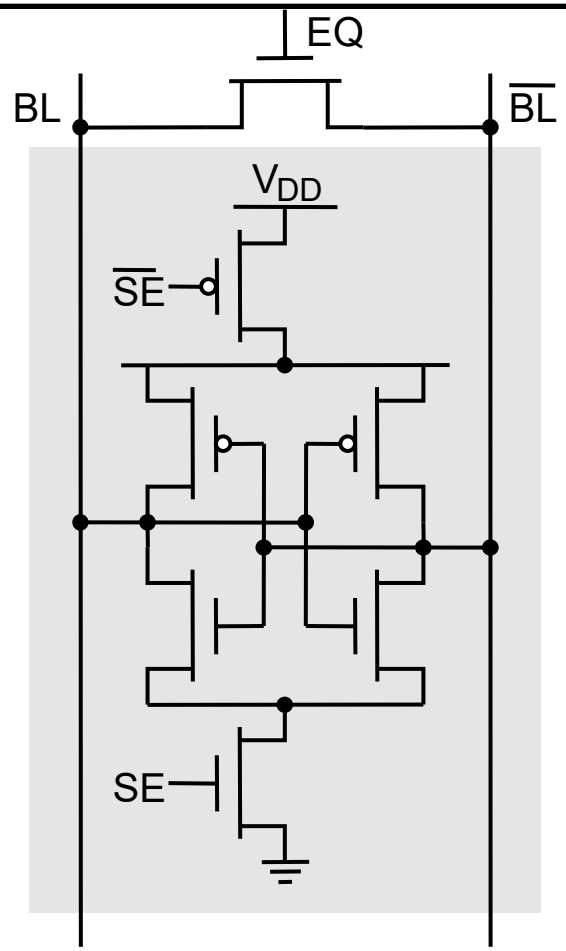


Differential Sense Amplifier



Directly applicable to SRAMs

Latch-Based Sense Amplifier (DRAM)



Initialized in its meta-stable point with EQ

Once adequate voltage gap created, sense amp enabled with SE

Positive feedback quickly forces output to a stable operating point.

FUTURE MEMORIES

Ideal Memory System

- Supply all (any) data that processor needs
  - ➔Capacity, Cost = small cell/die size, scalability
- In a moment
  - ➔Latency / Bandwidth
- With low power consumption and good reliability
  - ➔Non-volatility

Bandwidth vs. Latency

- Latency (=Execution Time, Response Time)
  - The time between the start and the completion of an event.
- Bandwidth (=Throughput)
  - The total amount of work done in a given time

➔ The terms “latency” and “bandwidth” are “the terms of choice” when discussing a high performance memory system

Comparison of Si Memories

	DRAM	SRAM	NAND	NOR	FRAM	PRAM
Cell size	2	6	1	4	5	2
Latency	2	1	6	5	2	4
Data rate	2	1	4	5	2	6
Low Vcc	2	1	6	5	2	3
Non-volatility	6	6	1	1	1	1
Endurance	1	1	6	6	2	3
Tech. migration	1	6	1	4	6	1
High Density	2	6	1	3	6	2
Market size	1	4	2	3	5	N.A

1 ← Best Worst → 6

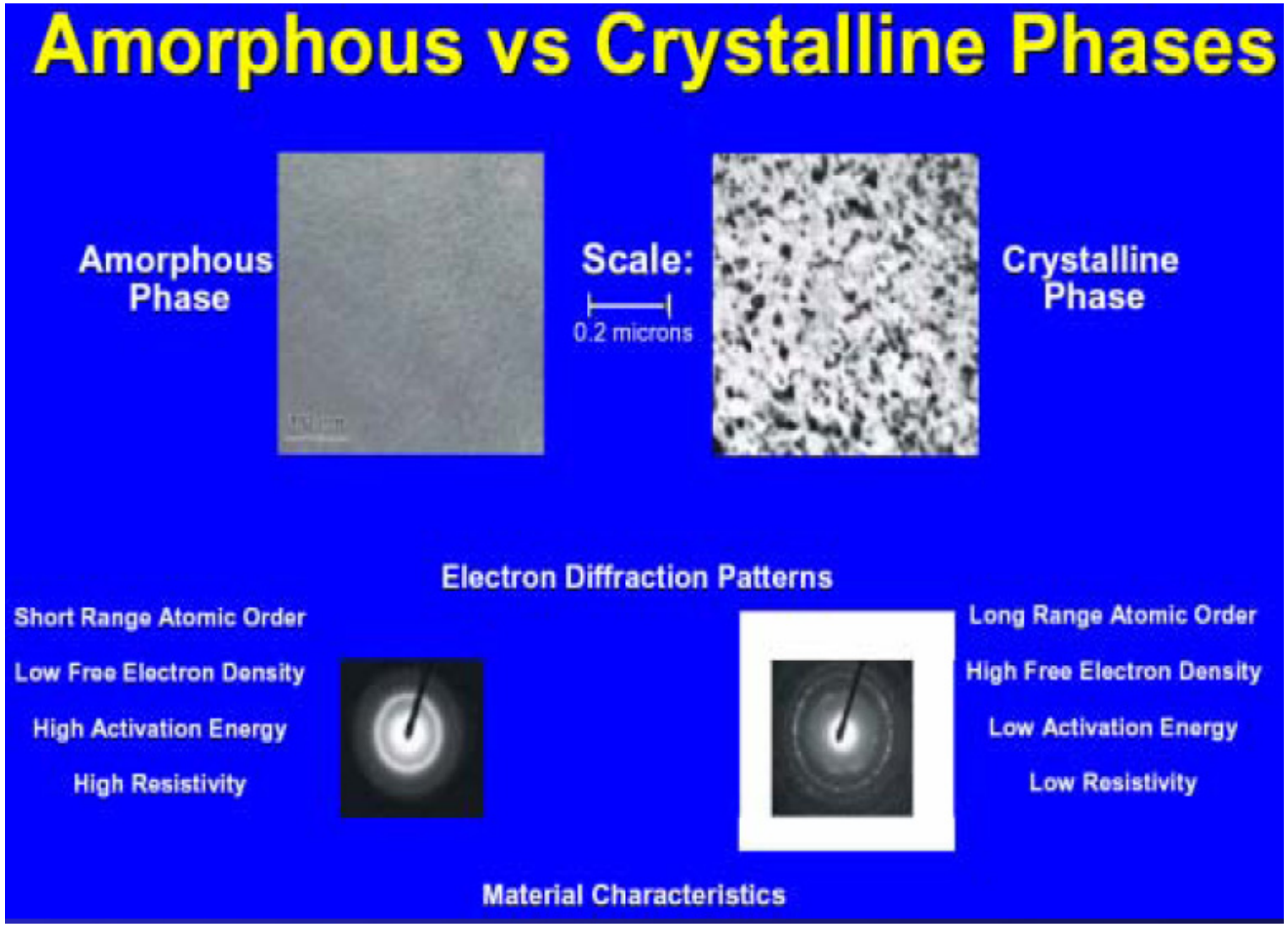
Courtesy Samsung

Chalcogenide Material

- The general class of switching media in CD-RW and DVD-RW
  - In high volume production and low cost
- Laser beam energy is used to control the switching between crystalline and amorphous phases
  - Higher energy ➔ amorphous
  - Medium energy ➔ crystalline
- Low energy laser beam to read

Courtesy Intel

Phase-Change RAM (PRAM)

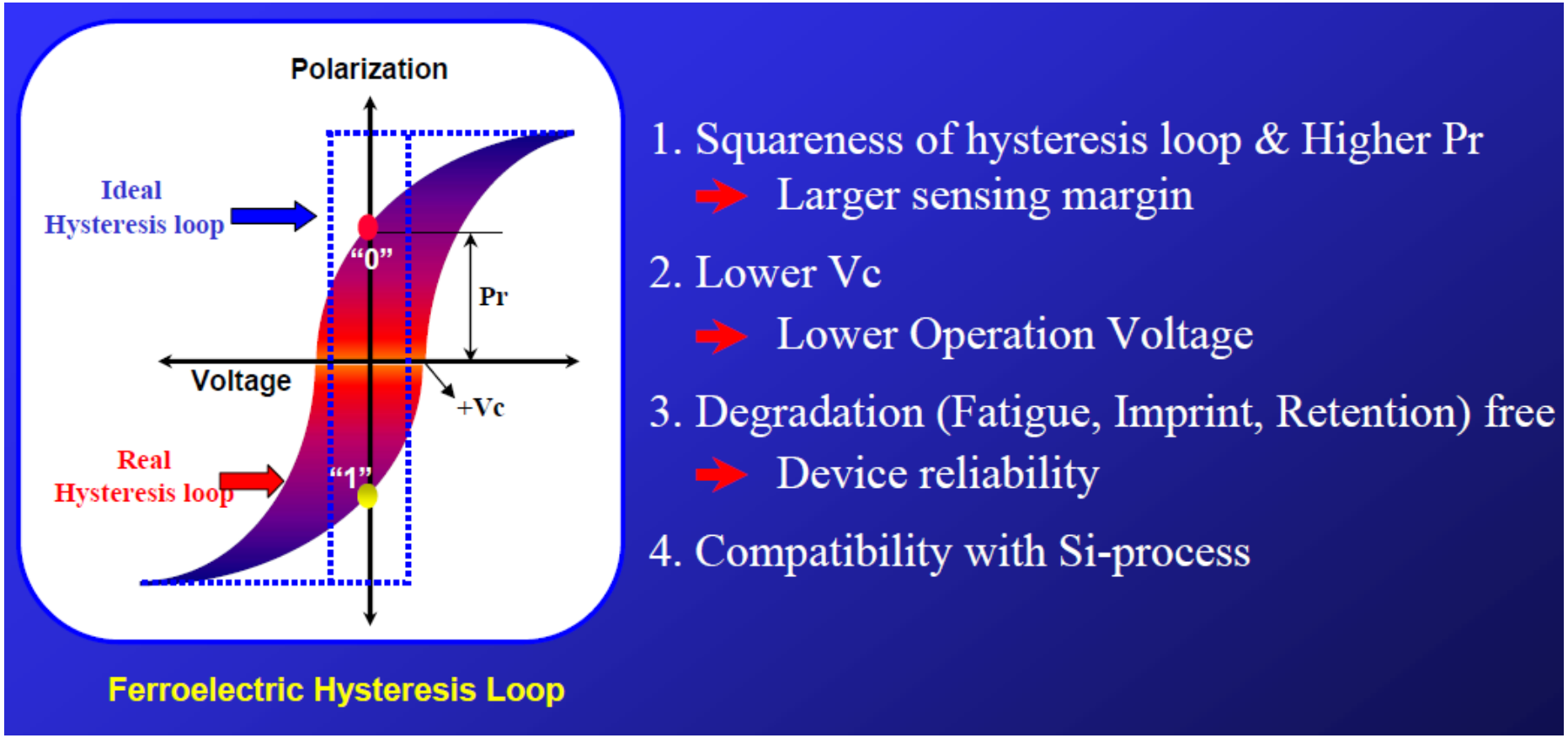


Courtesy Intel

PRAM Properties

- **Strong points and General properties**
  - Signal Sensing Margin : R_{reset} / R_{set} ~ 10^3
  - Programming Speed : ≤ 300 nsec
  - Endurance : > 10^9 (Read), 10^7 ~ 10^9 (Write)
  - Retention Characteristics : 10 years @110°C ~ 130°C
- **Technical Challenges**
  - Programming Current : 0.6 mA ~ 1mA
  - Set Resistance : 1kΩ ~ 10kΩ
  - Degradation due to Integration : Heat, Contamination

Ferroelectric RAM (FRAM)

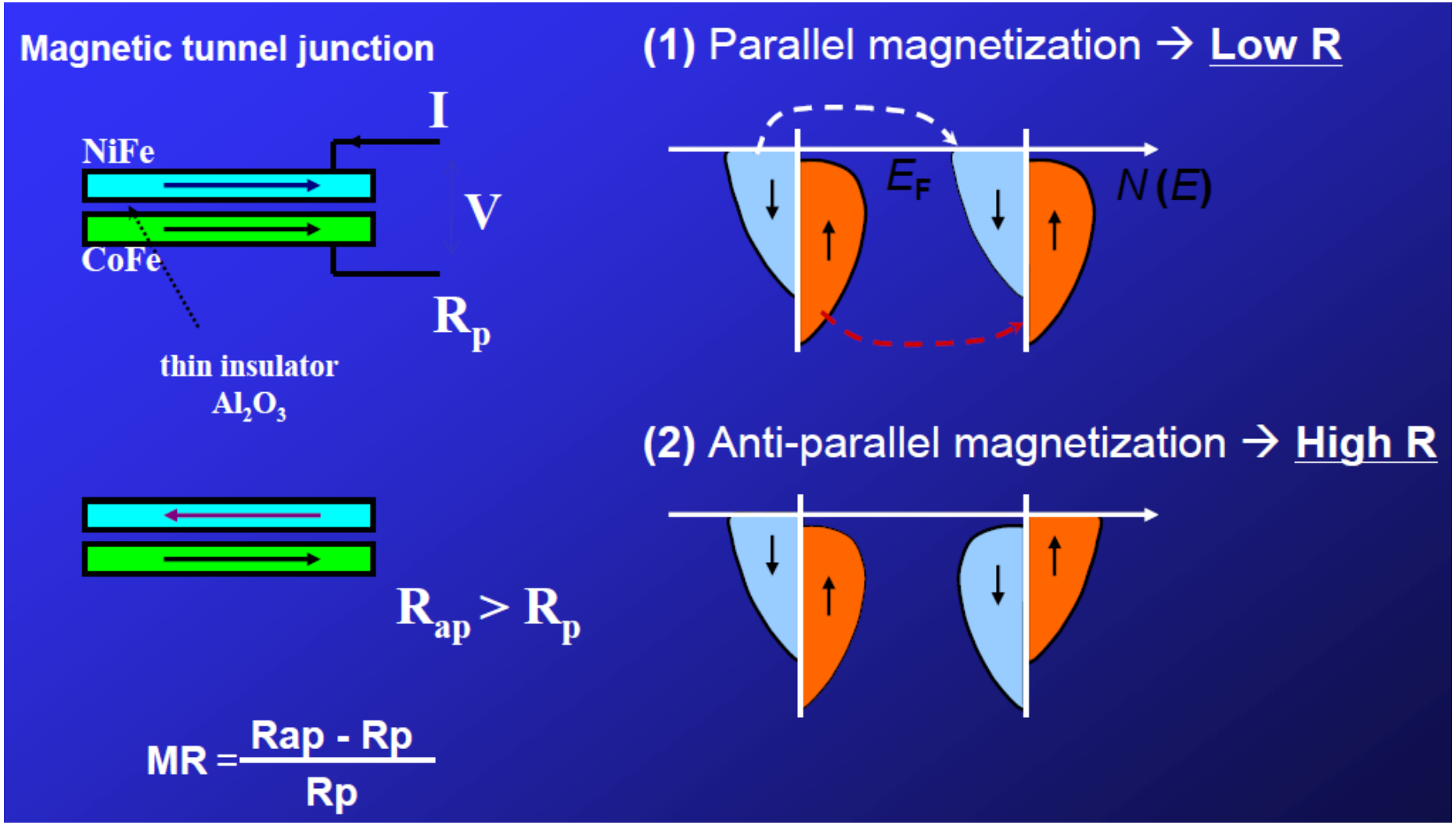


1. Squareness of hysteresis loop & Higher Pr
   - → Larger sensing margin
2. Lower Vc
   - → Lower Operation Voltage
3. Degradation (Fatigue, Imprint, Retention) free
   - → Device reliability
4. Compatibility with Si-process

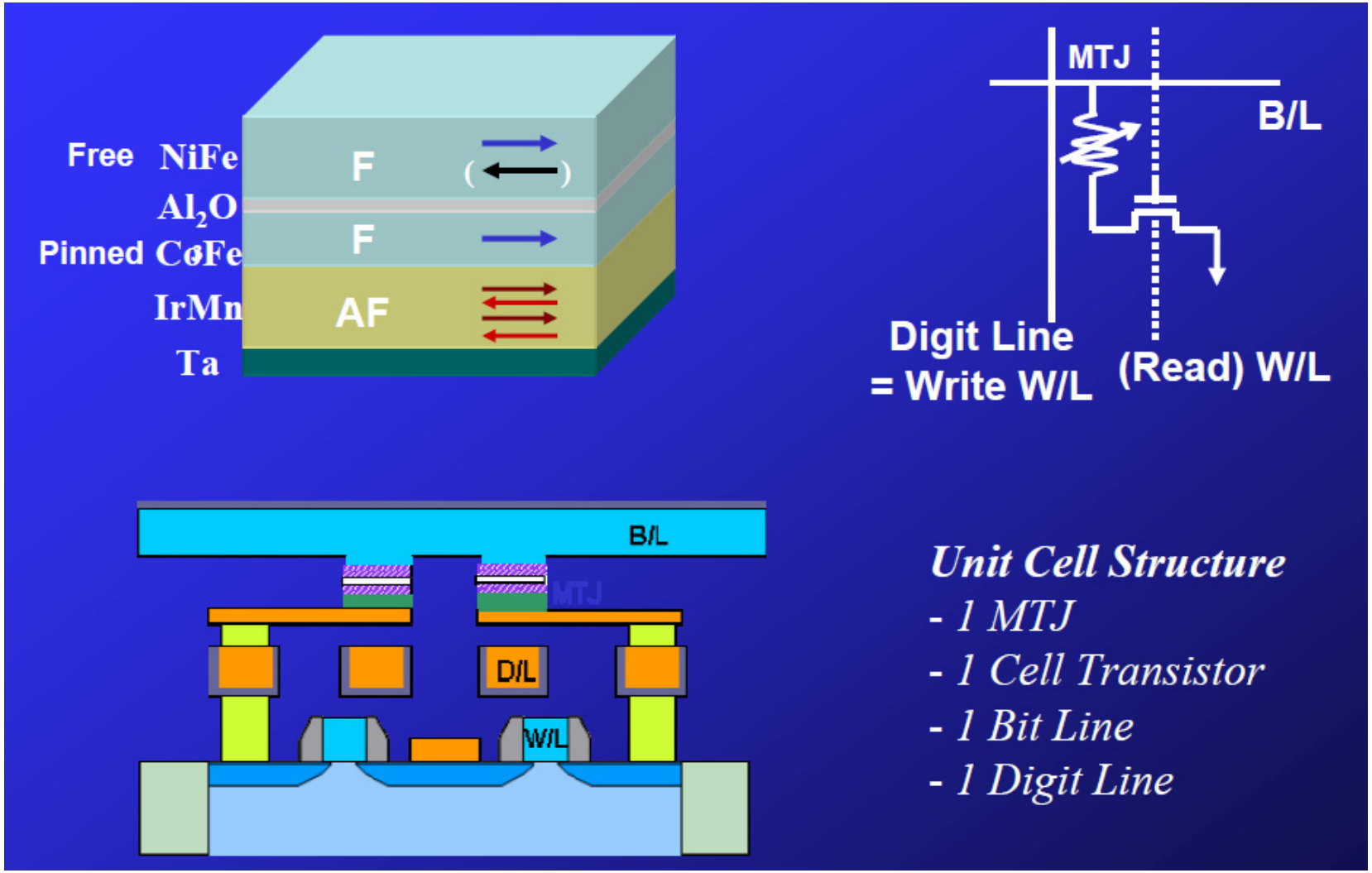
FRAM Properties

- **Strong points and General properties**
  - Programming Speed : ≤ 100 nsec
  - Endurance : > 10^{11} (Write / Read)
  - Retention Characteristics : > 10 years @85°C
- **Technical Challenges**
  - Thin Ferroelectric Film : ~ 70nm
  - Low Voltage Operation : ≤ 1.2 V
  - High Etching Slope : ~ 82°
  - Degradation Free Integration
  - 3-Dim. Conformal Deposition : Not Yet

Magnetic RAM (MRAM)



MRAM cell structure



Comparison of NVM

Parameter	NAND Flash	NOR Flash	FRAM	MRAM	PRAM
Random access	no	yes	yes	yes	yes
Cell Size (Achieved)	$4F^2$	$10F^2$	$15F^2$	$>40F^2$	$\sim 10F^2$(Bi) $\sim 30F^2$ (MOS)
Cell Size (Achievable)	$4F^2$	$10F^2$	$15F^2$	$30F^2$	$6F^2$
Scalability Limit	Good Coupling	Good Drain Disturb.	Bad Capacitor	Worse Writing Current	Good Reset Current
Write cycles	10^5	10^5	$>10^{12}$	$>10^{12}$	$>10^{12}$
Write Speed (Pgm.) (Erase)	200us/Page 1ms/Block	10us/Byte 1s/Sect.(64kB)	< 100ns N.A.	<100ns N.A.	~100ns N.A.

- PRAM (density)
- MRAM & FRAM (nearly ideal except for density)