



ES - 626 Research Presentation

Floating Gate Flash Memories

Instructor:
Prof. Nihar R Mohapatra

Presented By:
Rutu Patel
Kailash Prasad
PhD Electrical Engineering
IIT Gandhinagar



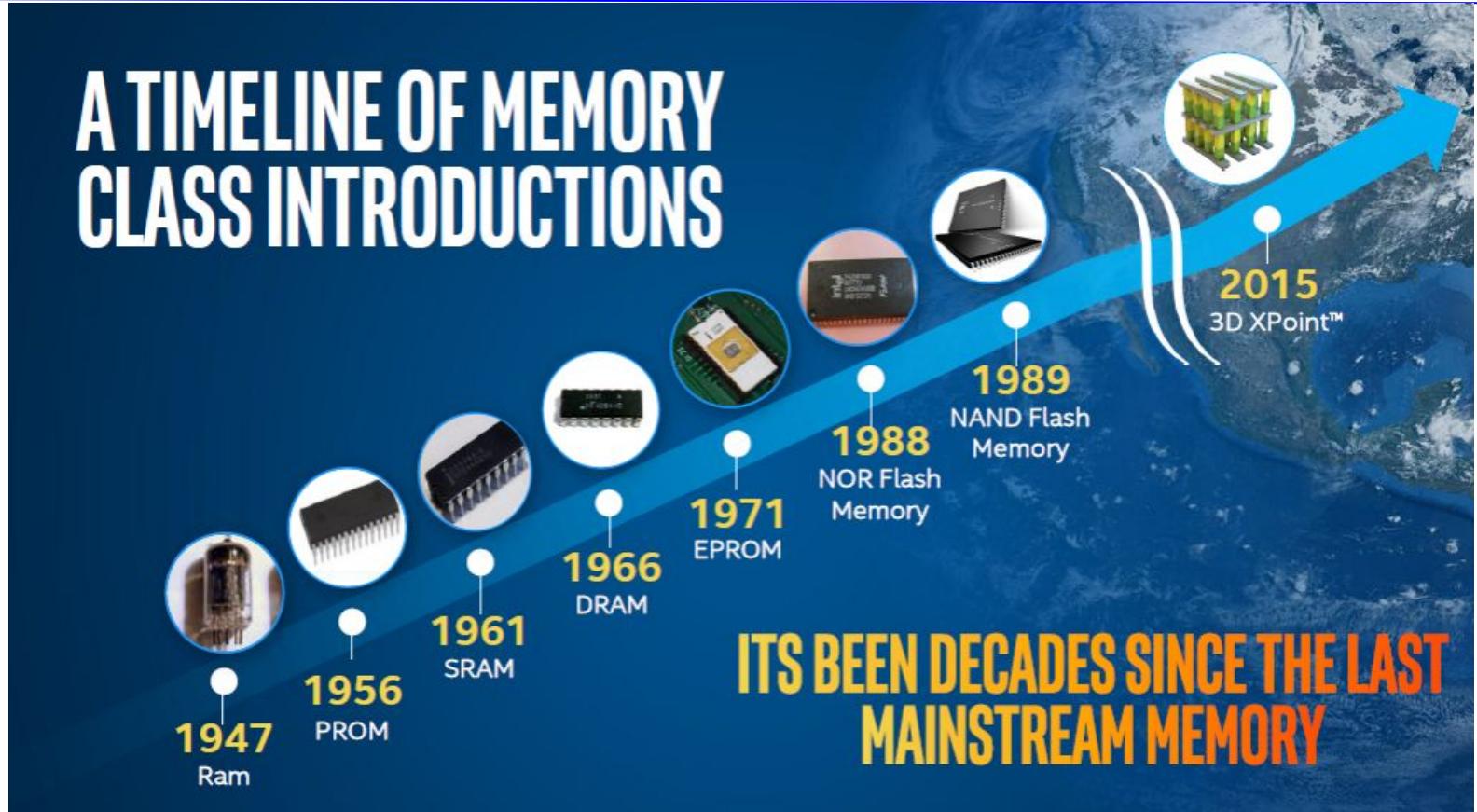
Outline :

- Why Flash Memories?
 - History of Flash Memories
 - NOR and NAND flash memories
- NAND flash devices
 - read/ write (program) / erase operations
- Process Technology
- Multi Level Cell
- Scaling and Reliability issues in NAND flash.
- Flash Memories: Present and Future



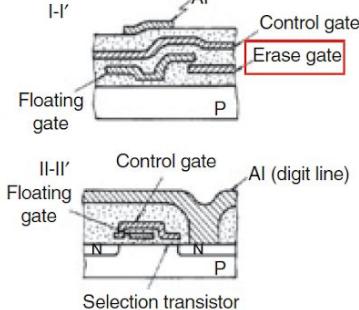
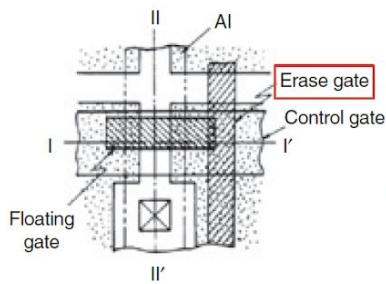
Why NAND Flash Memories?

History of Flash memory :



Floating Gate Flash memory :

- 1984 IEDM; 1st paper on **Flash Memory**
- 1987 IEDM; 1st paper on **NAND Flash**

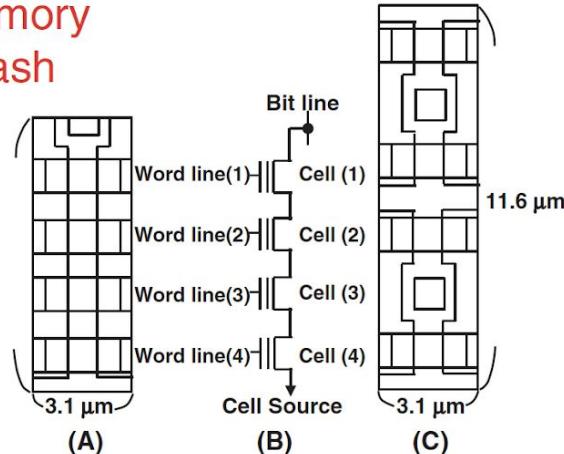


F. Masuoka et al., IEDM 1984

- Erase Gate ; Electron ejection from FG
- All cells can be erased at same time
- Fast Access

NOR Flash → Code Storage

(a)



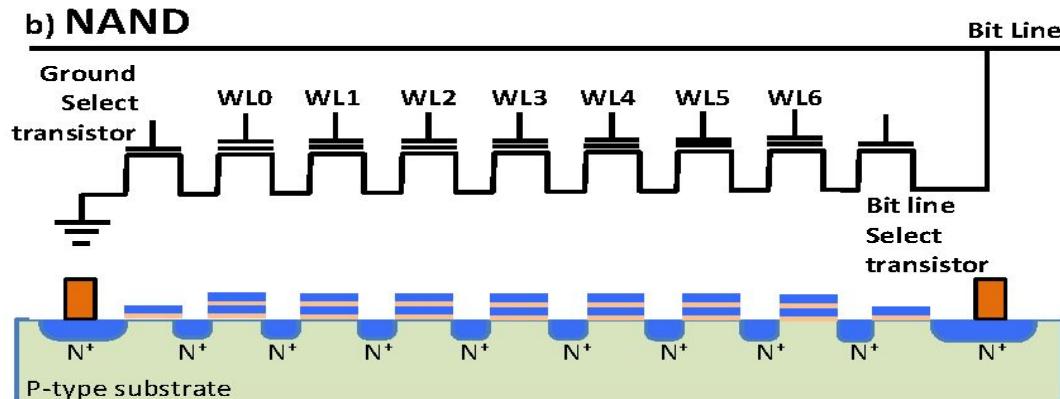
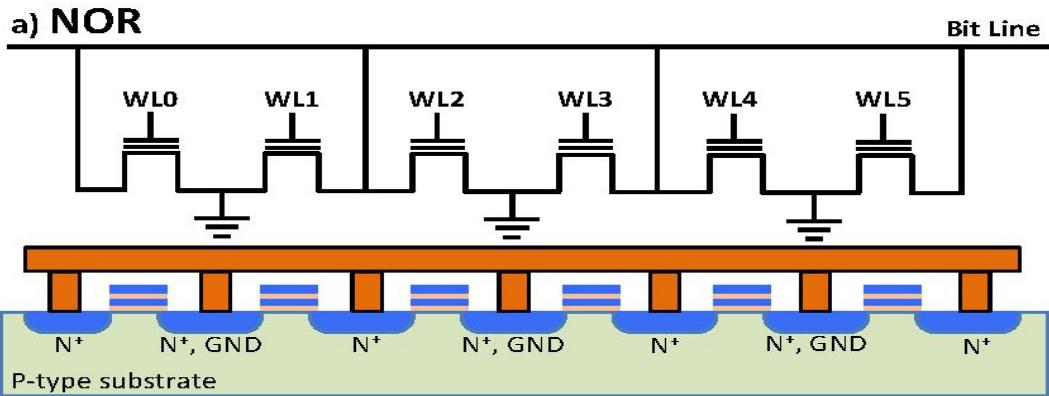
F. Masuoka et al., IEDM 1987

- Small Cell Size
- 1 Contact/2 cells → 1 con/64 cells
- Slow access

NAND Flash → File Storage

(b)

NAND vs NOR Flash Array :



Where Flash is being used

NAND

Mass Storage



Mobile Phone (Storage)

Tablet PC

Solid-State Disk

Memory Cards

NOR

Code Memory



BIOS/Networking

Telecommunications

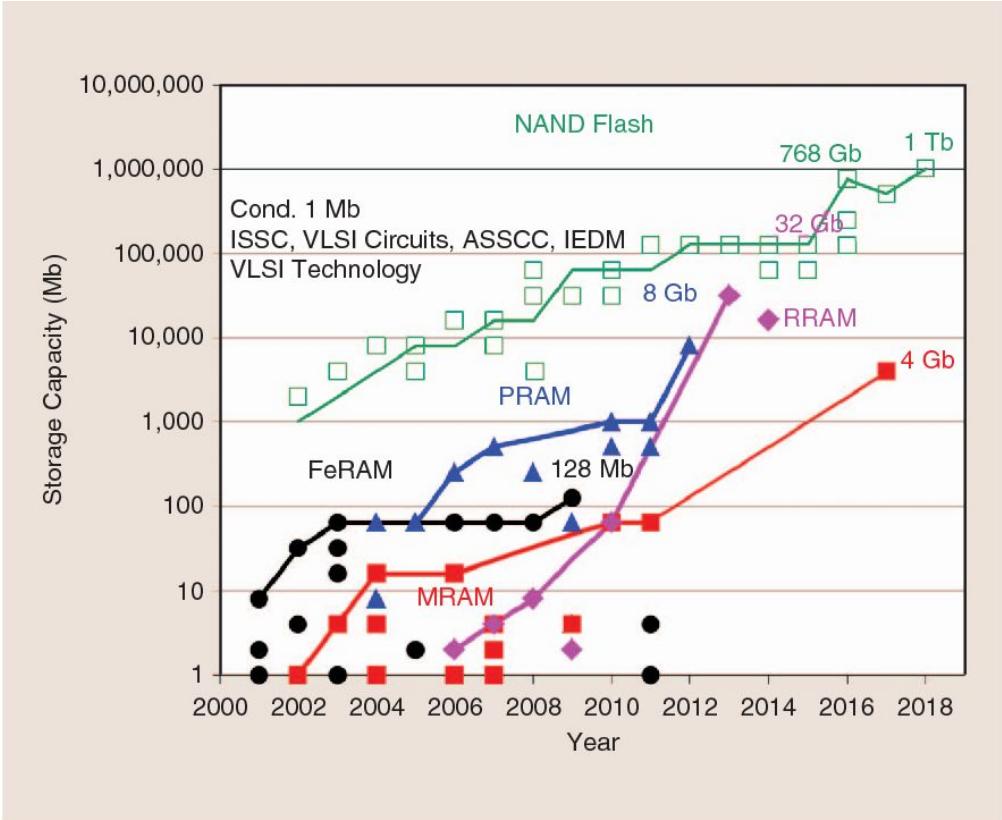
Mobile Phone (Code)

POS / PDA / PCA

- **Low Cost** and High Density
- **Page Mode Program**

- **Fast Random Access**
- **Fast Read Speed**

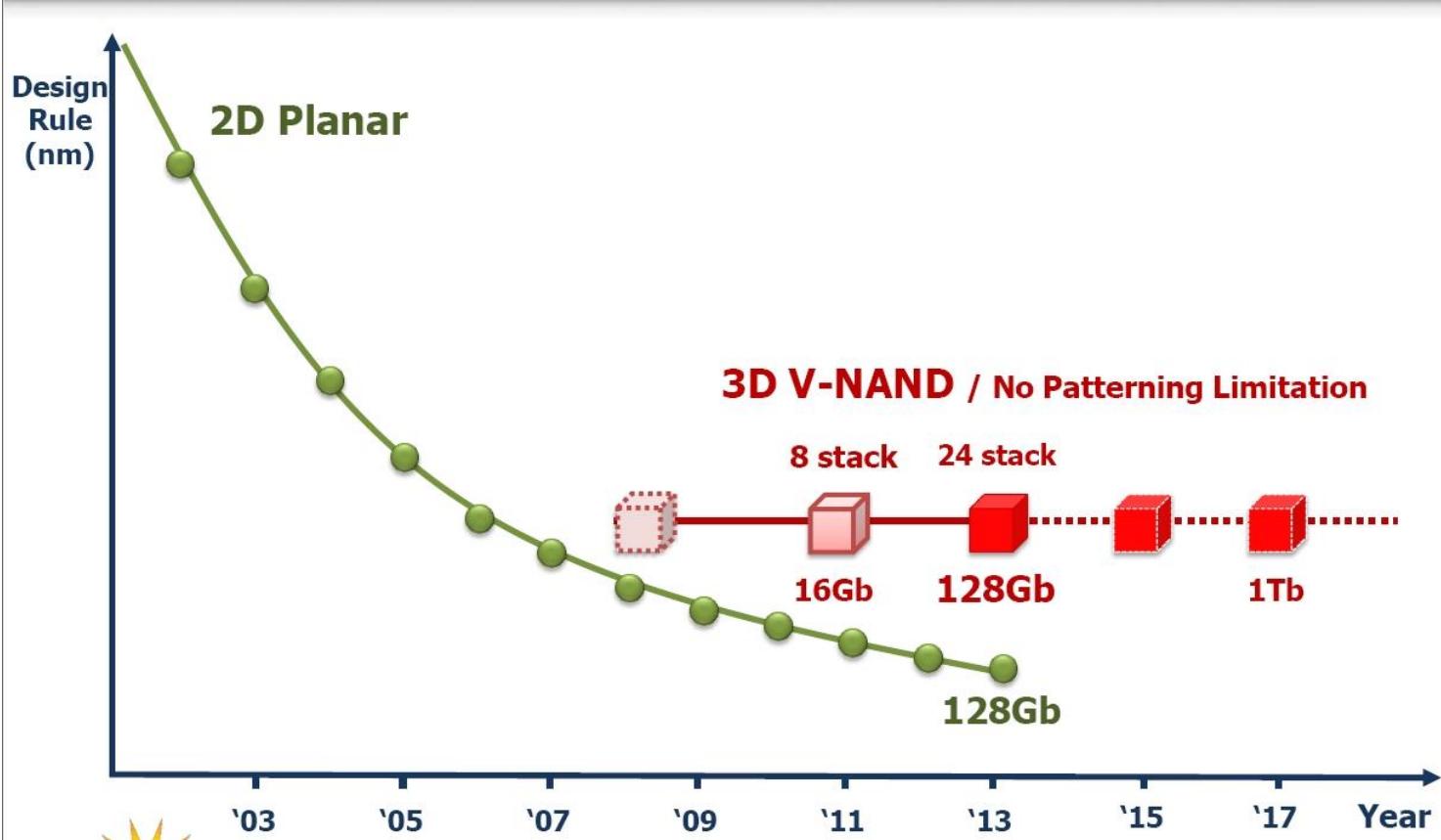
Flash Memory Scaling Trends



Hwang's Law:
Double Capacity of
Chip every Year

FIGURE 30: Memory-capacity trends for emerging NVMs. ASSCC: IEEE Asian Solid-State

Flash Memory Technology





Flash in News

Samsung NAND flash bit supply growth to drop below 30% in 2020

Siu Han, Taipei; Jessie Shen, DIGITIMES



Friday 22 November 2019



...

Samsung Electronics' NAND flash bit supply growth is expected to drop below 30% and even this year, according to industry sources.

\$62.27 Bn NAND Flash Memory Market - Global Forecasts from 2019 to 2024

RESEARCHANDMARKETS
THE WORLD'S LARGEST MARKET RESEARCH STORE

NEWS PROVIDED BY
[Research and Markets](#) →
Sep 03, 2019, 12:15 ET

SHARE THIS ARTICLE



SUPPLY CHAIN

Highlights of the day: Samsung cautious about expanding NAND flash output

DIGITIMES staff



Friday 22 November 2019



...

A flash memory made out of graphene and molybdenite

EPFL scientists have combined two materials with advantageous electronic properties – graphene and molybdenite – into a flash memory prototype that is promising in terms of performance, size, flexibility and energy consumption.



NAND Flash Memory Device

Requirement of NAND Flash Memory :

➤ Low Bit Cost

Small cell size & Scalability → Self-Aligned STI (SA-STI)
Multi-bit cell (MLC)

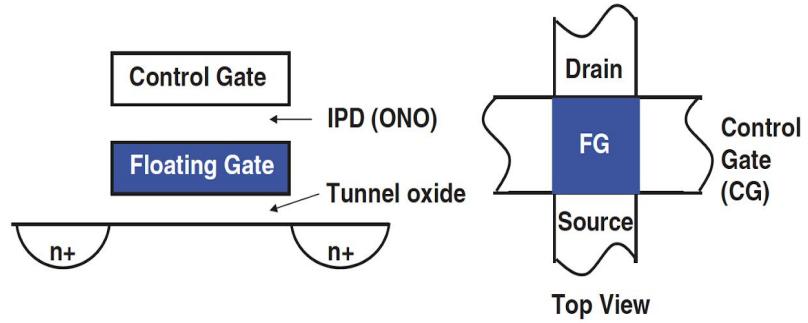
➤ High-Speed Program

Parallel (Low power) Program → Page program
Bit-by-bit verify
 V_{pgm} step up (ISPP)

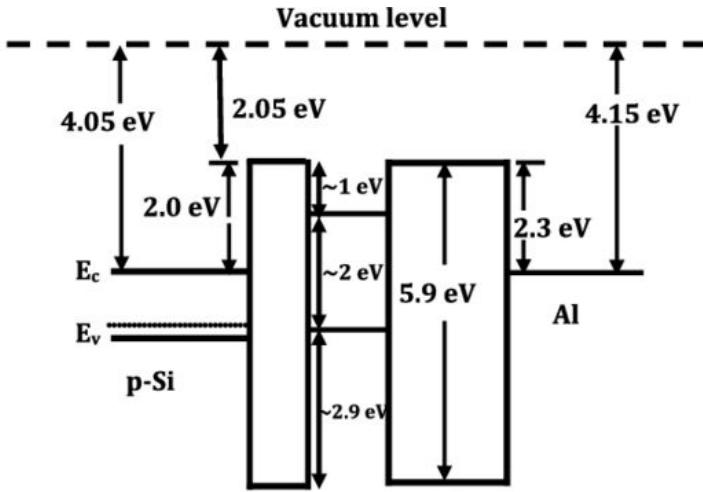
➤ High Reliability

Less degradation on tunnel oxide
→ Uniform P/E scheme

Floating gate transistor as memory cell :

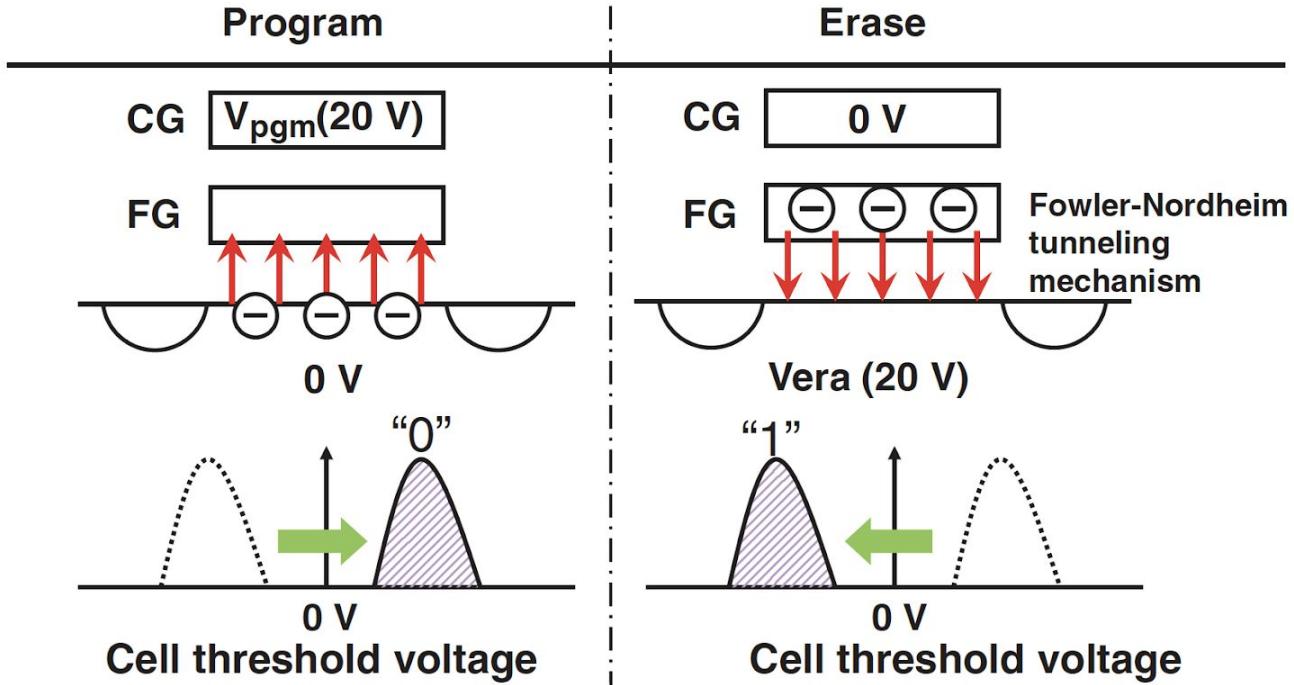


a) Floating gate transistor structure

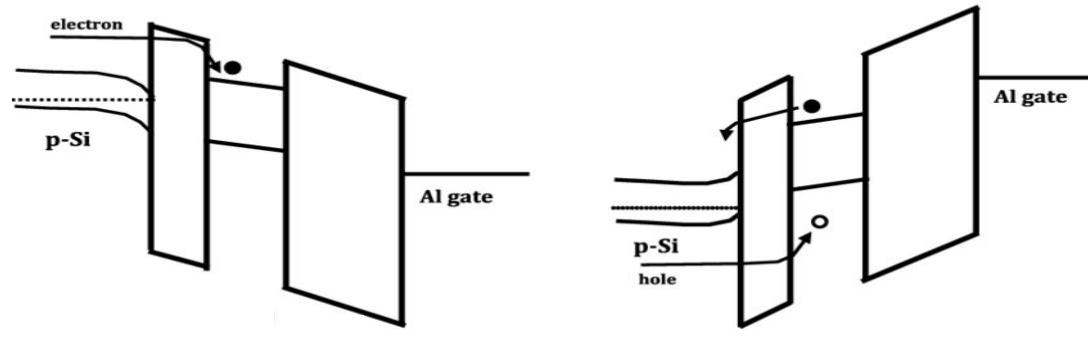
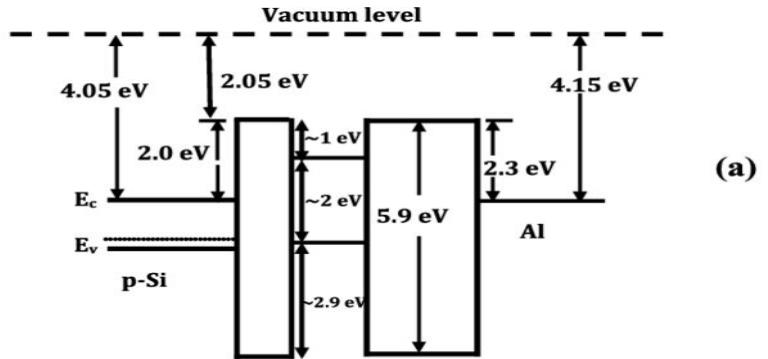


b) Energy band diagram for floating gate transistor

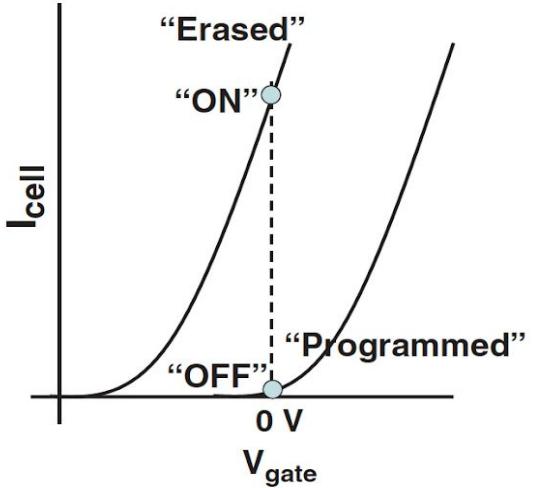
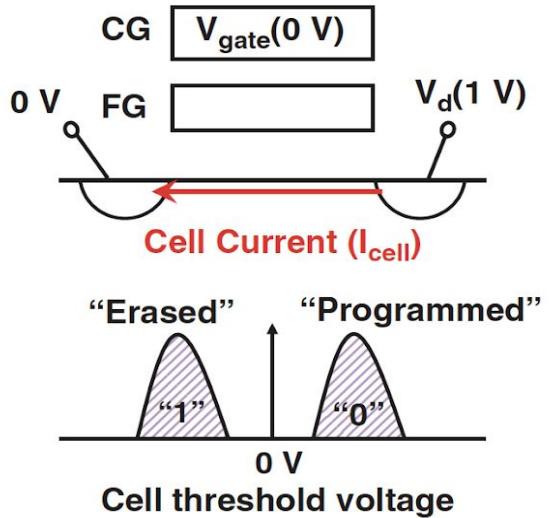
Erase and Program operations :



Energy Band Diagram for Erase and Program :



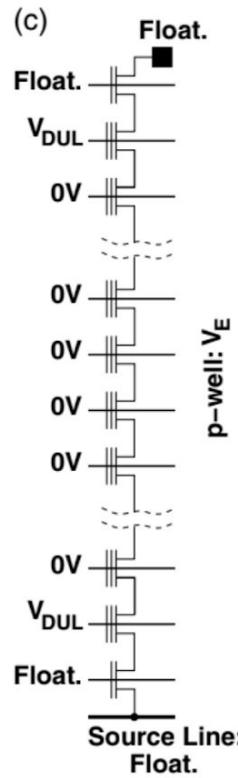
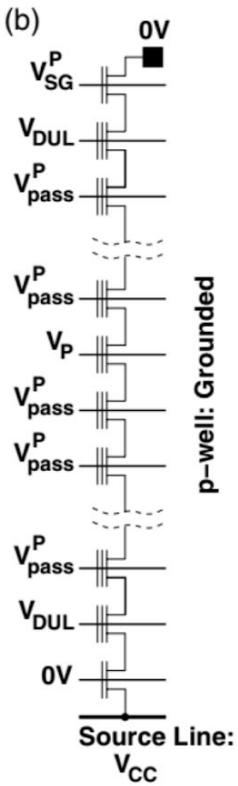
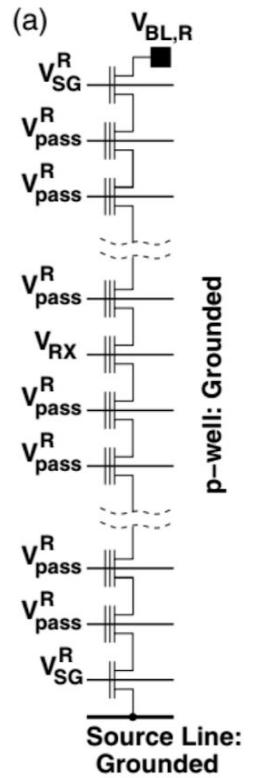
Read operation:



Principle of read of single cell

- Erased: +ve charge in FG. I_{cell} “ON”.
- Programmed: -ve charge in FG. I_{cell} “OFF”.

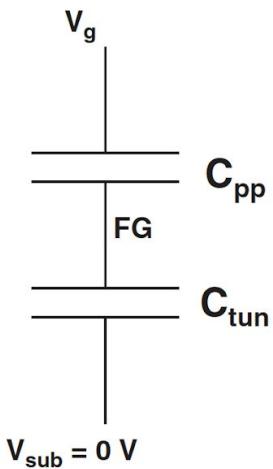
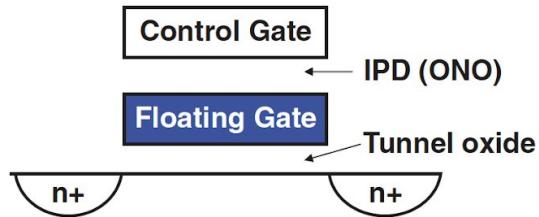
Operations in NAND array:



Voltages applied to the NAND String :

- (a) read a selected memory cell in the string by applying V_{RX} to WL
- (b) program a selected memory cell in the string by applying V_P to WL
- (c) Erase the entire nand block by applying V_{SG}^R and V_{SG}^P are the voltages applied to the gates of the biased select transistors during read and program, respectively.

Capacitive coupling of the floating gate:



$$|V_{tun}|_{write} = V_g \cdot K_w$$

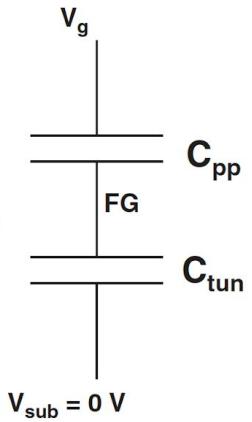
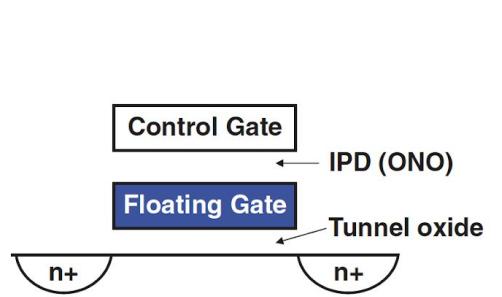
$$K_w = \frac{C_{pp}}{C_{pp} + C_{tun}}$$

$$|V_{tun}|_{erase} = V_{well} \cdot K_e$$

$$K_e = 1 - \frac{C_{tun}}{C_{pp} + C_{tun}}$$

V_{tun} is the voltage across tunneling oxide when Q_{fg} is 0.

Capacitive coupling of the floating gate:



$$|V_{tun}|_{write} = V_g * K_w + \frac{Q_{fg}}{C_{pp} + C_{tun}}$$

$$|V_{tun}|_{erase} = V_{well} * K_e - \frac{Q_{fg}}{C_{pp} + C_{tun}}$$

$$\Delta Vt = - \frac{Q_{fg}}{C_{pp}}$$

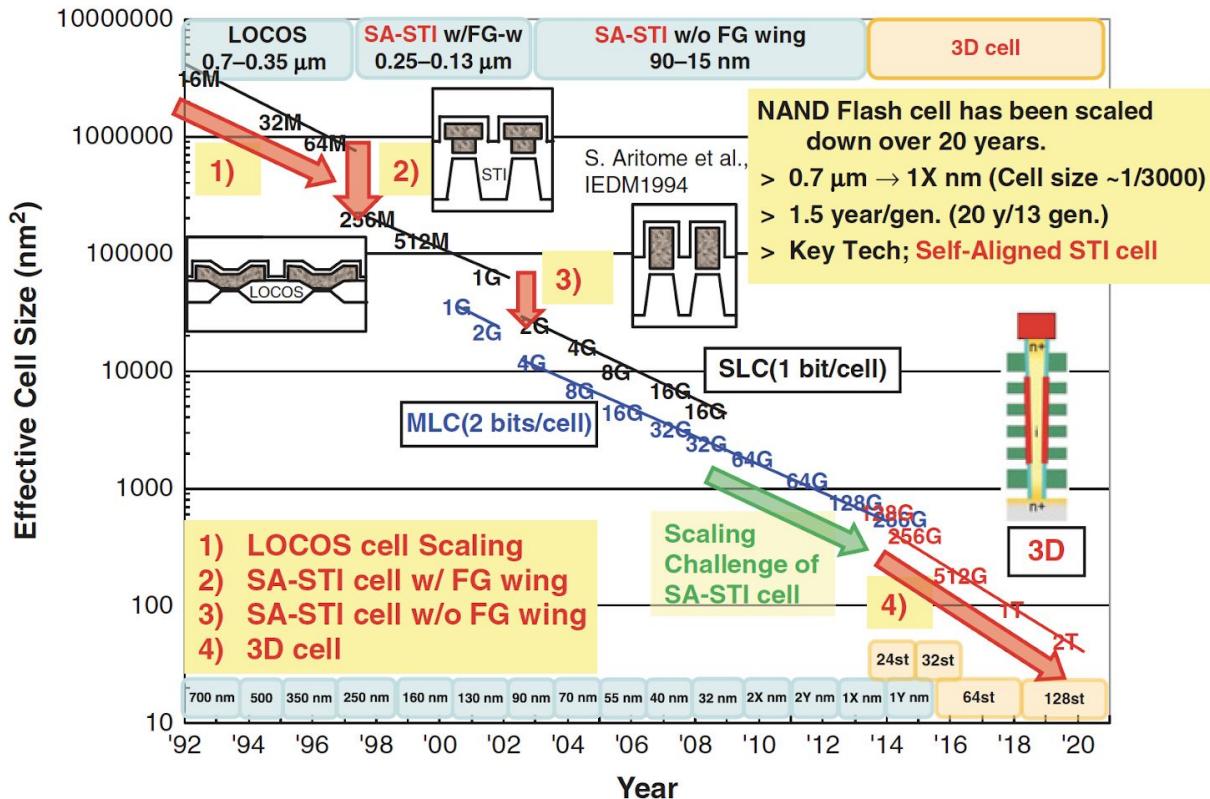
$$V_{tw} = V_{ti} - \frac{Q_{fg}}{C_{pp}} = V_{ti} + V_g * \left(1 - \frac{V_{tun}}{K_w * V_g} \right)$$

$$V_{te} = V_{ti} - \frac{Q_{fg}}{C_{pp}} = V_{ti} - V_{well} * \left(\left(\frac{K_e}{K_w} \right) - \frac{V_{tun}}{K_w * V_{well}} \right)$$

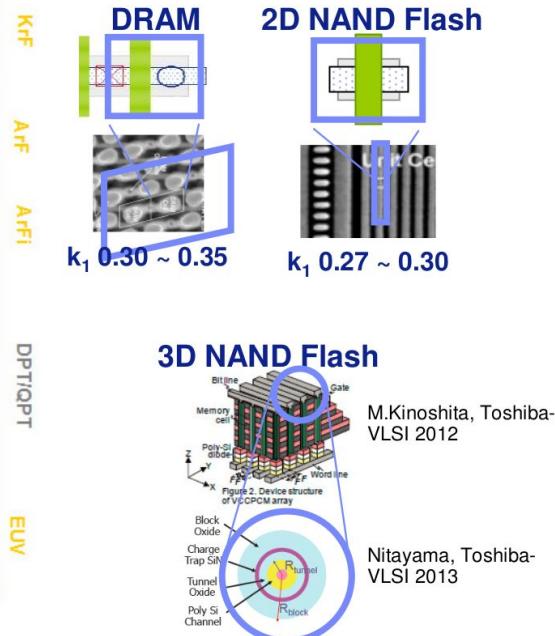
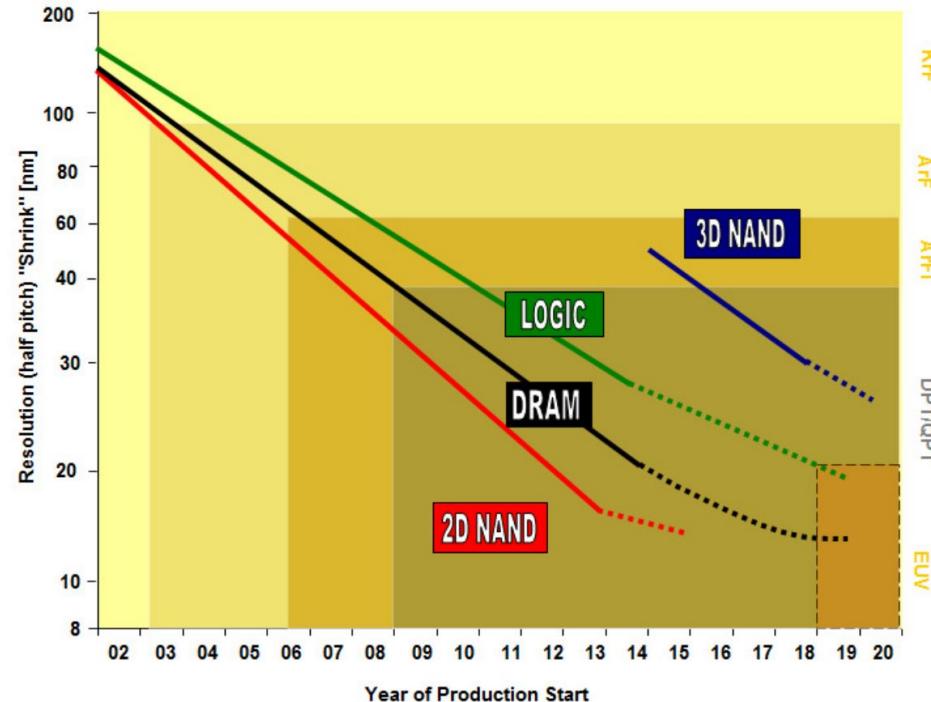


Process Technology

NAND Flash memory Technology Road Map :

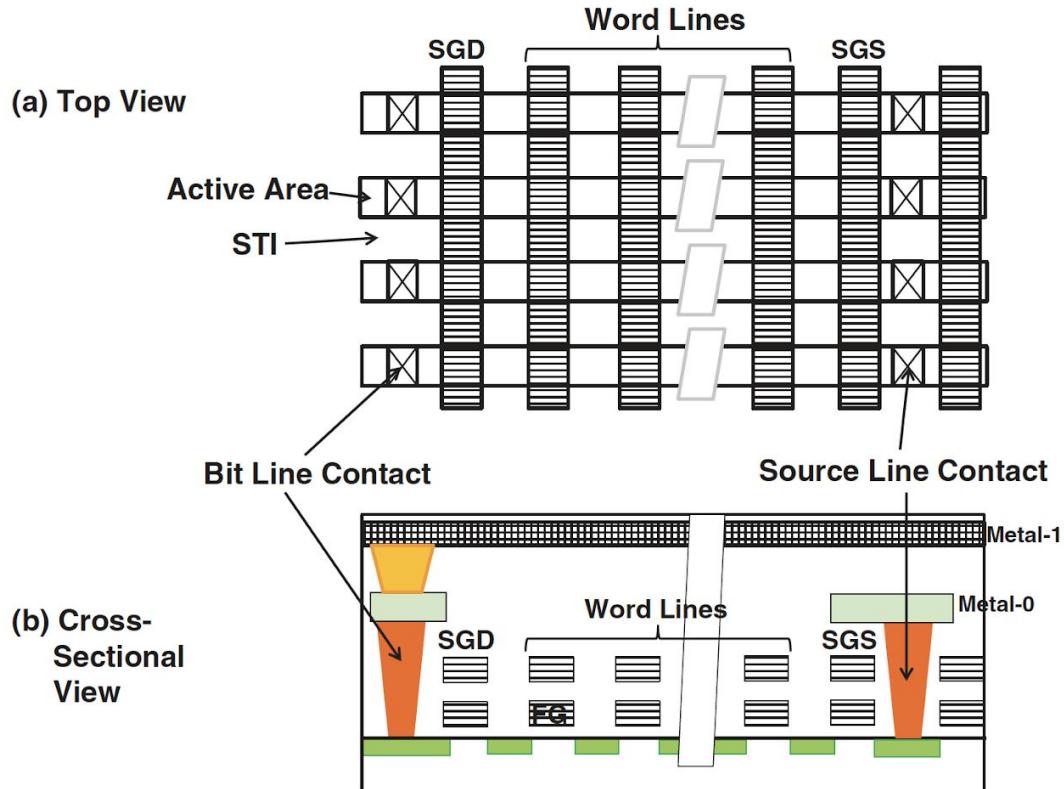


Flash Memory : The Driver of Lithography

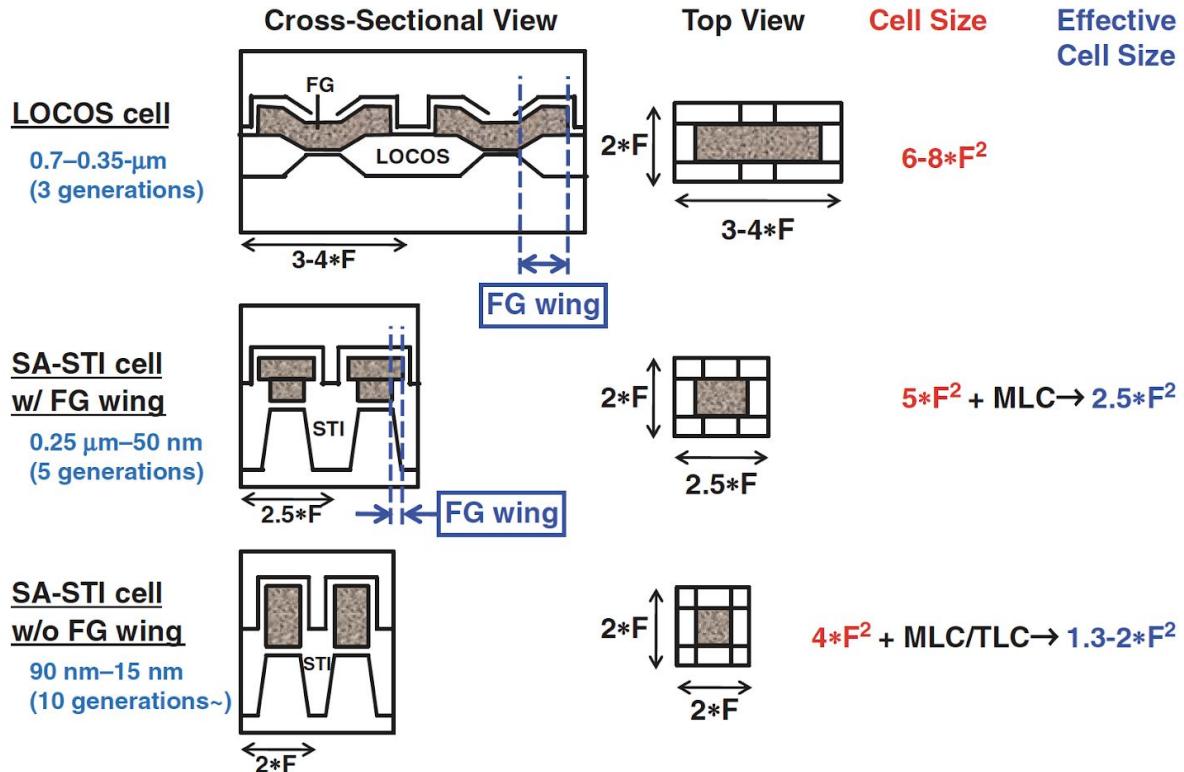


**Due to simple layout and Explosion in Demand
NAND Flash became the driver of Lithography**

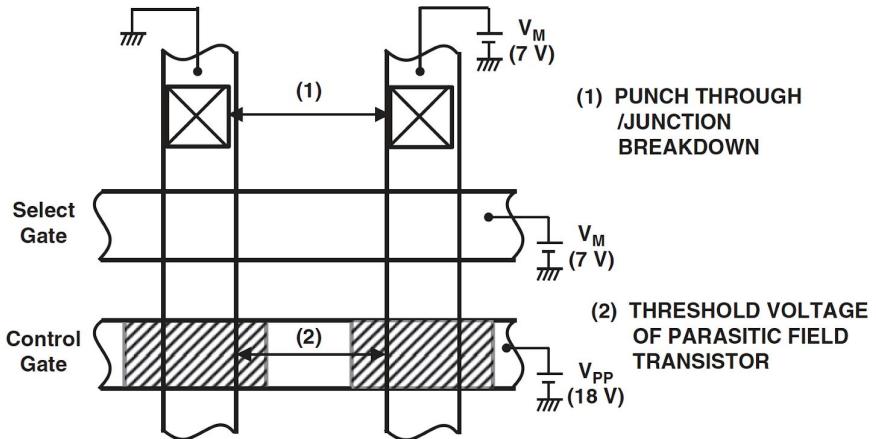
Layout of NAND memory cells Array :



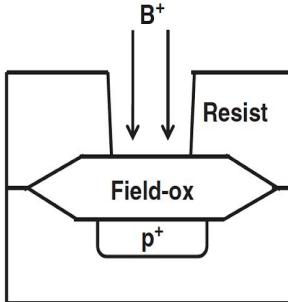
Structure scaling of NAND Flash Memory :



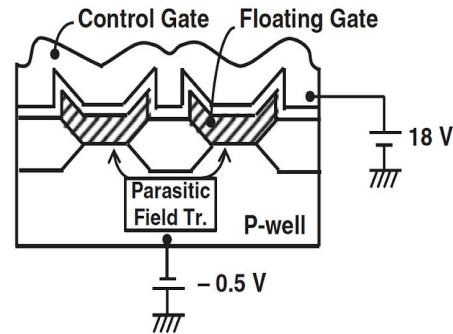
Scaling in LOCOS cell :



1) FTI (Field-Through
Implantation) PROCESS



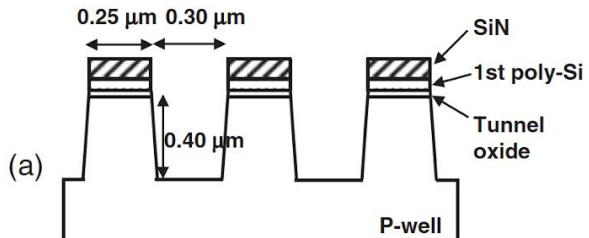
2) P-WELL NEGATIVE BIAS
DURING WRITE OPERATION



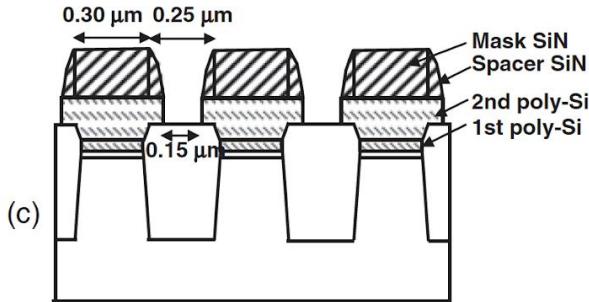
Issue

Solution

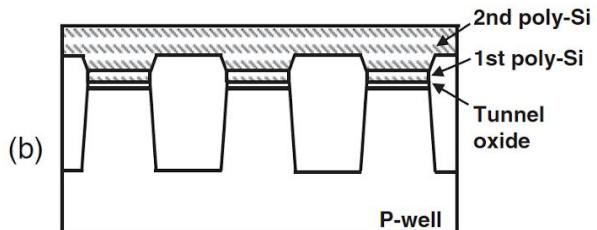
Self-Aligned STI cell with FG wing (Fabrication) :



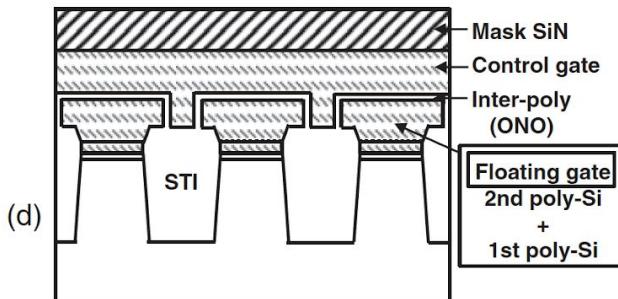
(a) Trench etching.



(c) Floating gate formation by SiN spacer process.

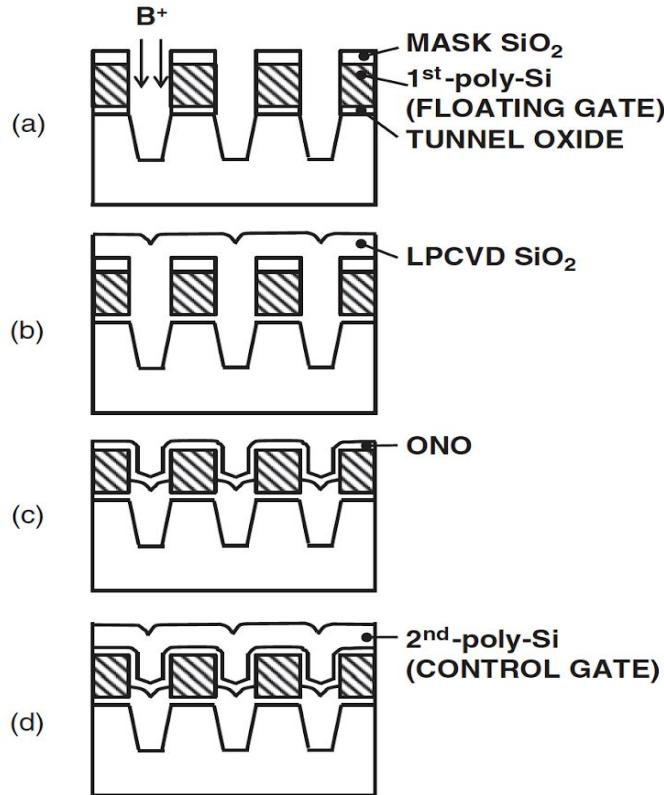


(b) LPCVD SiO_2 fill-in and planarization by CMP, second poly-Si gate deposited.



(d) ONO and the control-gate formation.

Self-Aligned STI cell without FG wing (Fabrication)

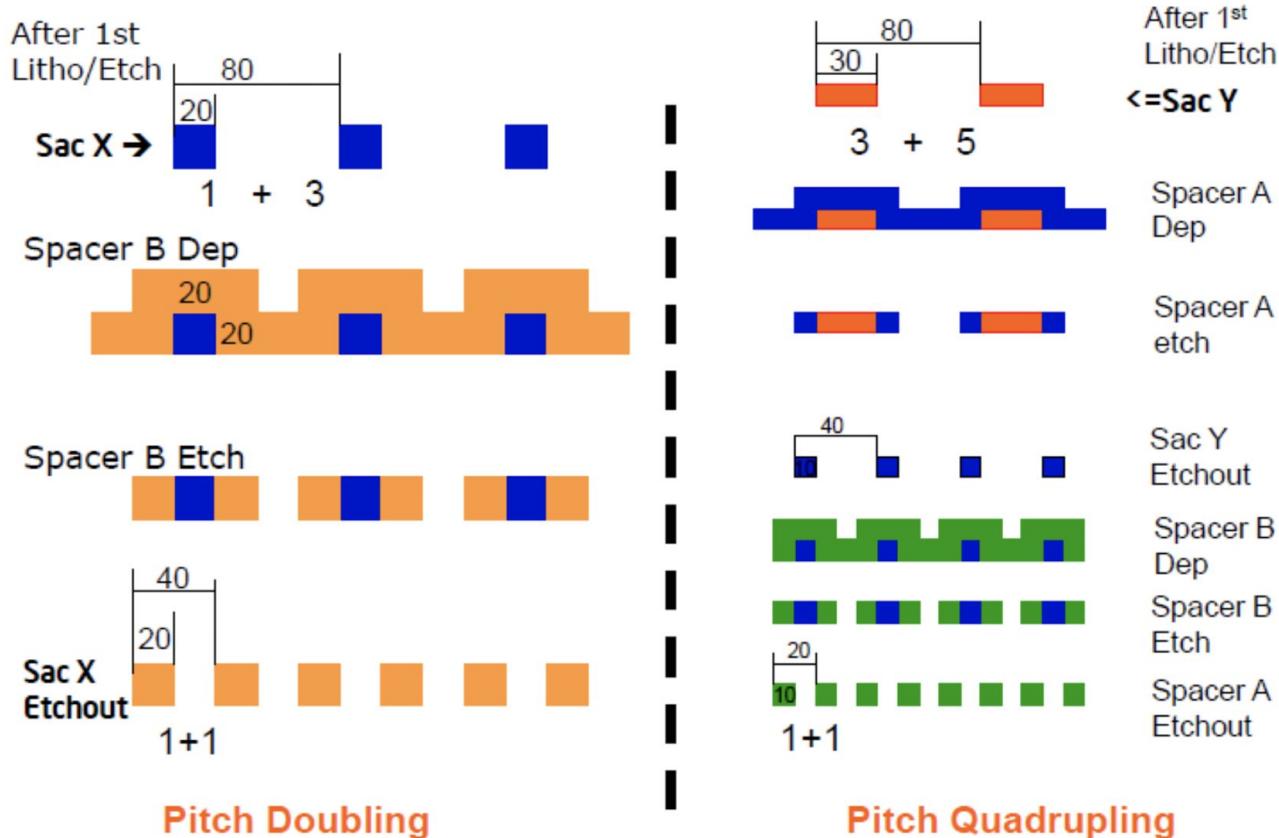


The process sequence of the SA-STI process without FG wing.

- (a) Trench etching, B+ implantation.
- (b) LP-CVD SiO₂ fill-in.
- (c) Oxide etch-back and ONO formation.
- (d) Control-gate formation.

The floating-gate and STI patterning are carried out by the same mask, so the number of fabrication steps for the SA-STI process can be decreased.

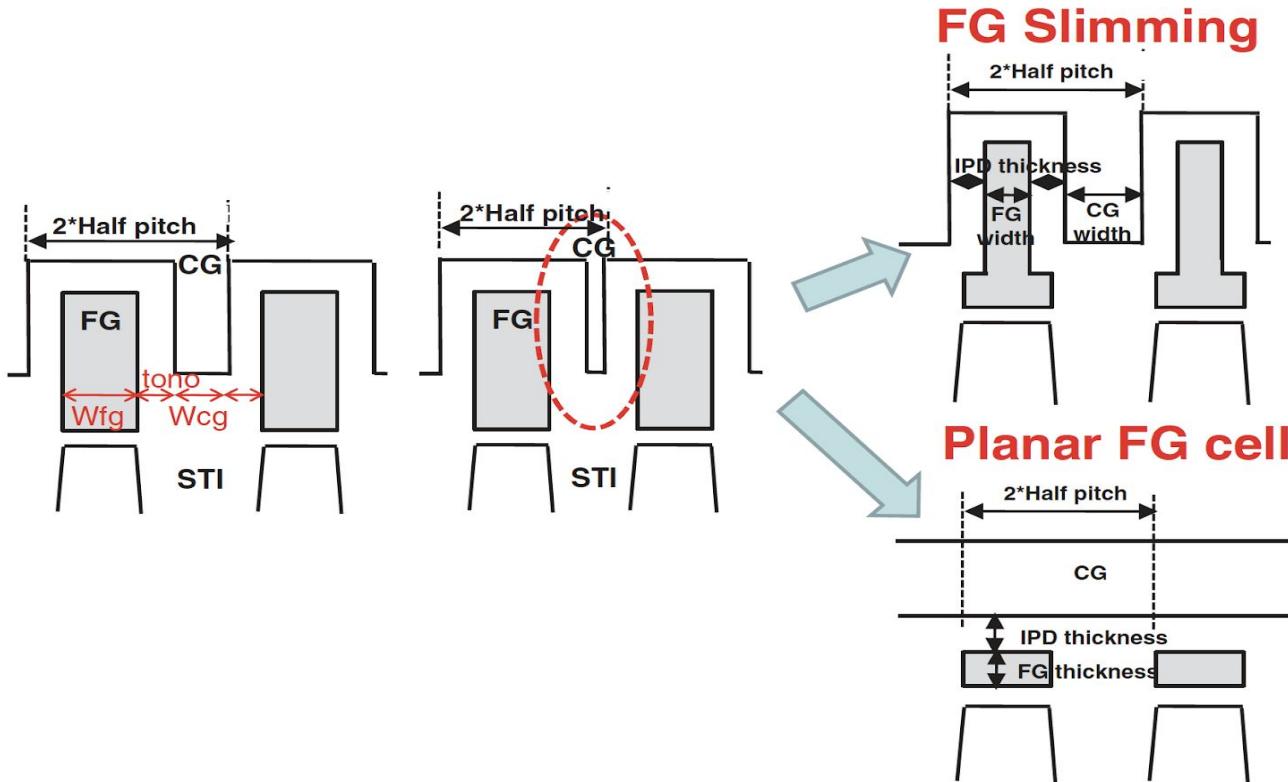
Double patterning and Quadruple patterning :



Dimensional Scaling of NAND Flash :

Generation	2X	2Y	1X	1Y	1Z	0X	Scaling factor
BL half-pitch (nm)	27	23.0	19.5	16.6	14.1	12.0	×0.85 assumption
WL half-pitch (nm), Gate length L	26	20.8	16.6	13.3	10.6	8.5	×0.8 assumption
Channel W (nm)	20	18.0	16.2	14.6	13.1	11.8	×0.9 assumption
ONO thickness (nm)	12	11.4	10.8	10.3	9.8	9.3	×0.95 assumption

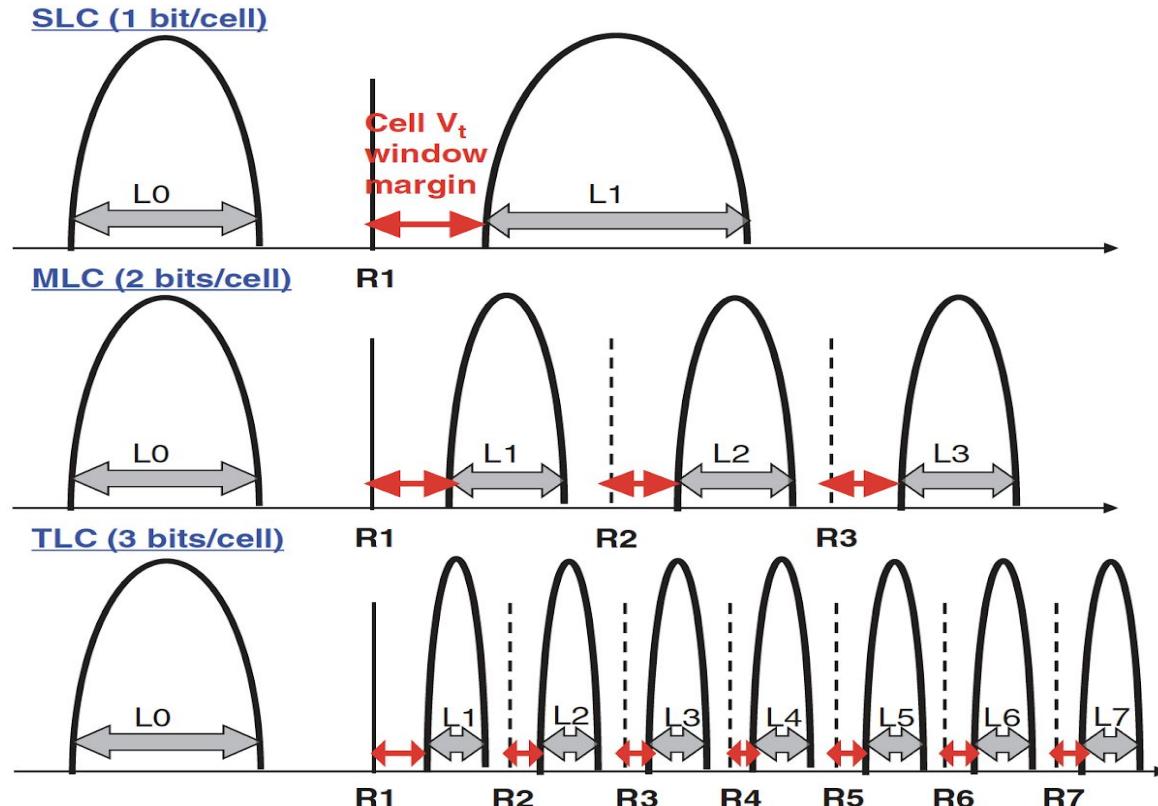
Planar FG cell with high k dielectric:



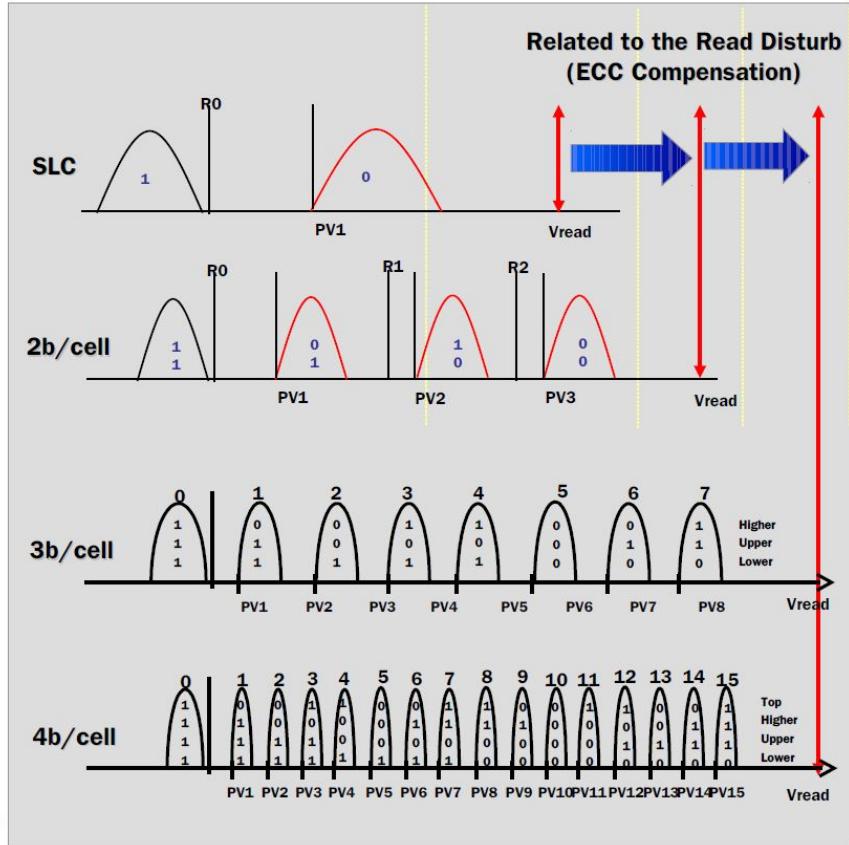


Multi Level Cell

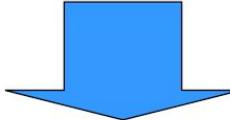
Multi-level (Multi Vt) cells (MLC) :



Multi Level Cell - MLC



More cell states in the limited voltage range



Requires tight control of each cell distribution



Performance and Reliability concerns

MLC vs SLC

ATTRIBUTE	SLC	pseudoSLC	MLC	TLC
Prevalent Process Geometry	2X - 4X nm	1x nm	1x nm	1x nm
Typical P/E Cycle Rating	◆◆◆◆◆ 60 - 100K	◆◆◆◆◆ ~20K	◆◆ 3K	◆ 500
Typical Data Retention	◆◆◆◆◆	◆◆◆◆◆	◆◆◆◆◆	◆◆◆◆◆
<10% of P/E cycles	10 years	5 years	5 years	1 - 4 years
100% of P/E cycles	1 year	1 year	1 year	0.5 - 1 year
t _{prog} typical (μs)	400 - 600	650 - 900	1300	1650
Typical Mono-die Density	1Gb - 32Gb	16Gb - 64Gb (in pSLC mode)	32Gb - 128Gb	128Gb - 256Gb
Typical Page/Block Size	2KB - 8KB Pages 128KB - 1MB Blocks	8KB - 16KB Pages 2MB - 4MB Blocks	8KB - 16KB Pages 2MB - 4MB Blocks	16KB Pages 4MB Blocks
Typical Life Cycle	5 years	18 months	18 months	12 - 18 months
Typical Cost per GB	◆◆◆◆◆	◆◆◆	◆◆	◆



NAND flash memory technologies evolve to support QLC

QLC NAND pushes the storage capacity of flash devices higher, but devices can wear out more quickly than other NAND types because more data can be written to the same cell.



By **Brien Posey**

Published: 01 Mar 2018

Innovation around NAND flash memory technologies has centered on improving 3D NAND and increasing the number of bits that can be stored in each flash memory cell.



Sponsored News

Delivering datacentre performance and cost savings with Intel Optane technology
—Intel



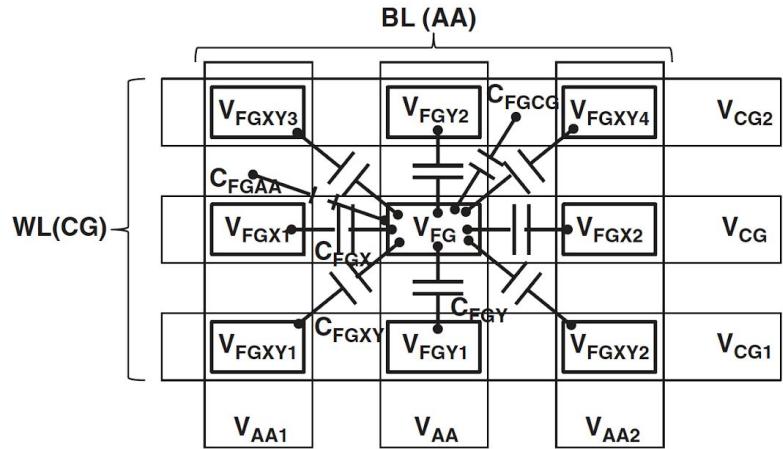
Scaling and Reliability Issues



Scaling and Reliability Issues

- Floating gate capacitive coupling interference:
- Stress Induced Leakage Current (SILC)
- Endurance in MLC or TLC
- Few Electrons Limitations

Floating gate capacitive coupling interference:



$$V_{FG} = \frac{C_{IPD}}{C_{TUN} + C_{IPD}} V_{CG} = \frac{C_{IPD}}{C_{Total}} V_{CG} = CR * V_{CG}$$

C_{FGX} ; Capacitance with x-direction FG (BL-BL)

C_{FGY} ; Capacitance with y-direction FG (WL-WL)

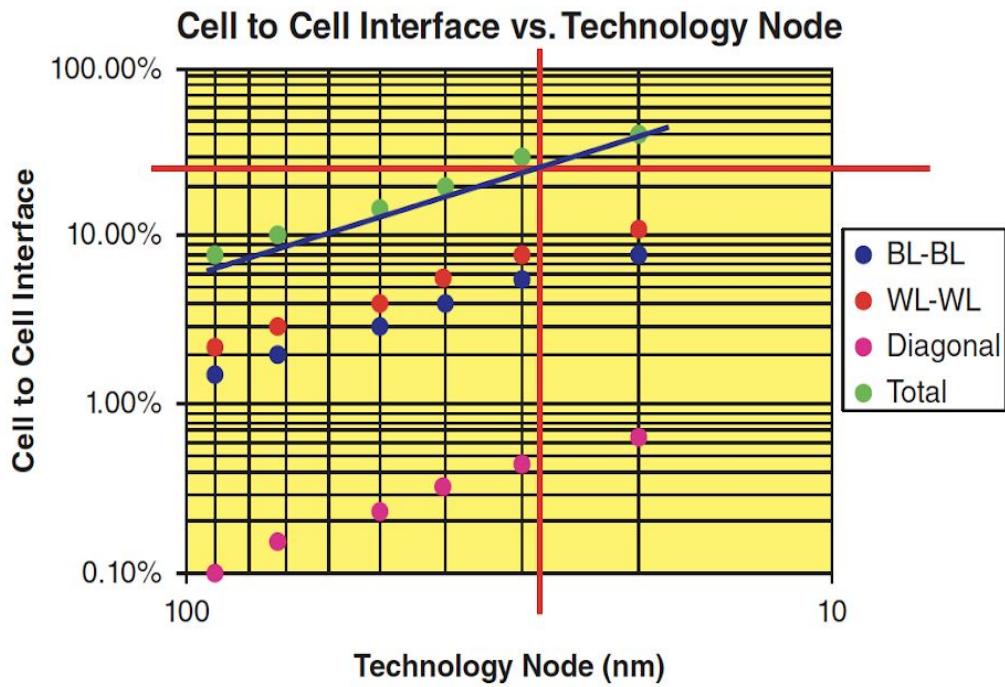
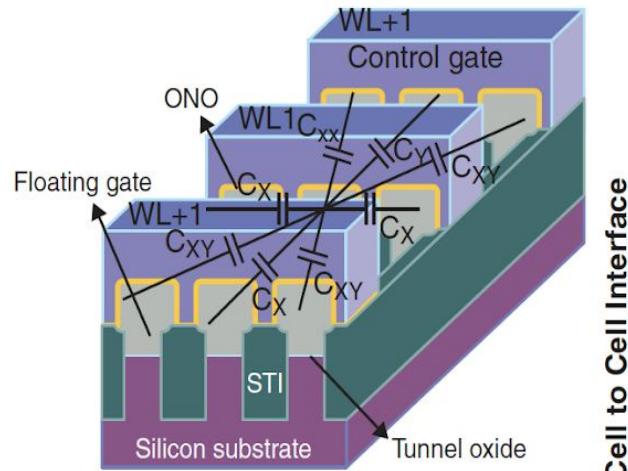
C_{FGXY} ; Capacitance with diagonal-direction FG

C_{FGCG} ; Capacitance with neighbor Control Gate

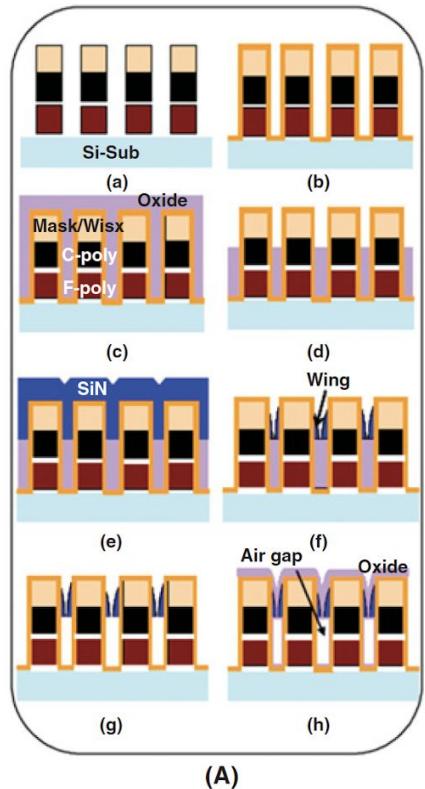
C_{FGAA} ; Capacitance with neighbor Active Area

$$V_{FG} = \frac{C_{IPD}V_{CG} + C_{PCX}(V_{PCX1} + V_{PCX2}) + C_{PCY}(V_{PCY1} + V_{PCY2}) + C_{PCXY}(V_{PCXY1} + V_{PCXY2} + V_{PCXY3} + V_{PCXY4}) + C_{PGCG}(V_{PGC1} + V_{PGC2}) + C_{FG(A)A}(V_{(A)A1} + V_{(A)A2})}{C_{TUN} + C_{IPD} + 2C_{FAX} + 2C_{FAY} + 4C_{FOXY} + 2C_{FOCG} + 2C_{FG(A)A}}$$

Floating gate capacitive coupling interference:



Air gap to reduce FG coupling interference:



The process flow of the air gap:

1. the gate patterning and the barrier
2. silicon dioxide deposition (150 \AA),
3. the barrier SiN deposition (200 \AA)
4. thick-oxide deposition (1000 \AA),
5. thick oxide is removed by dry etch
6. SiN deposition (150 \AA),
7. the wing is formed,
8. the thick oxide inside gate to gate space is removed by wet etch.
9. the air gap is formed.

Stress Induced Leakage Current (SILC)

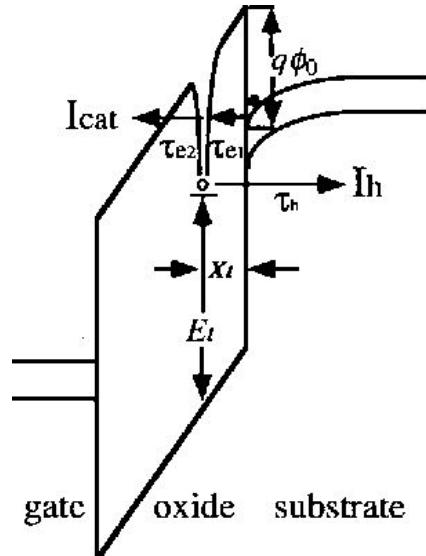
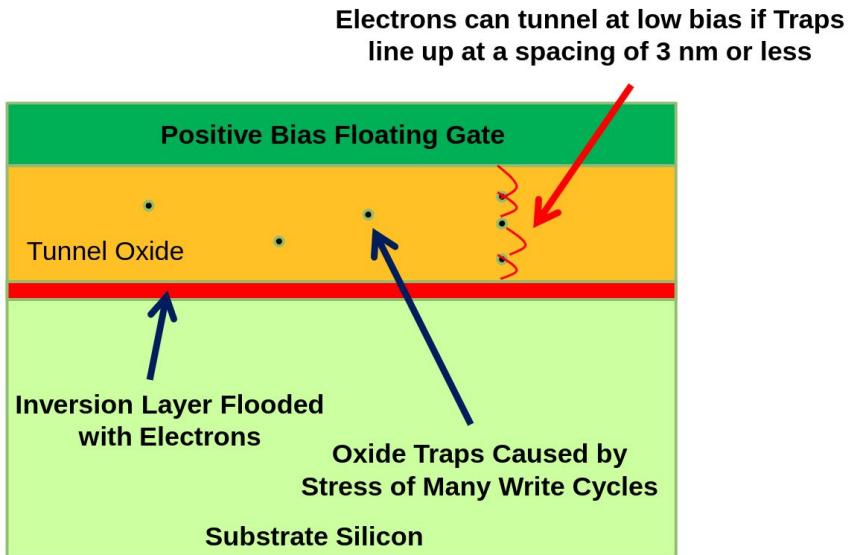
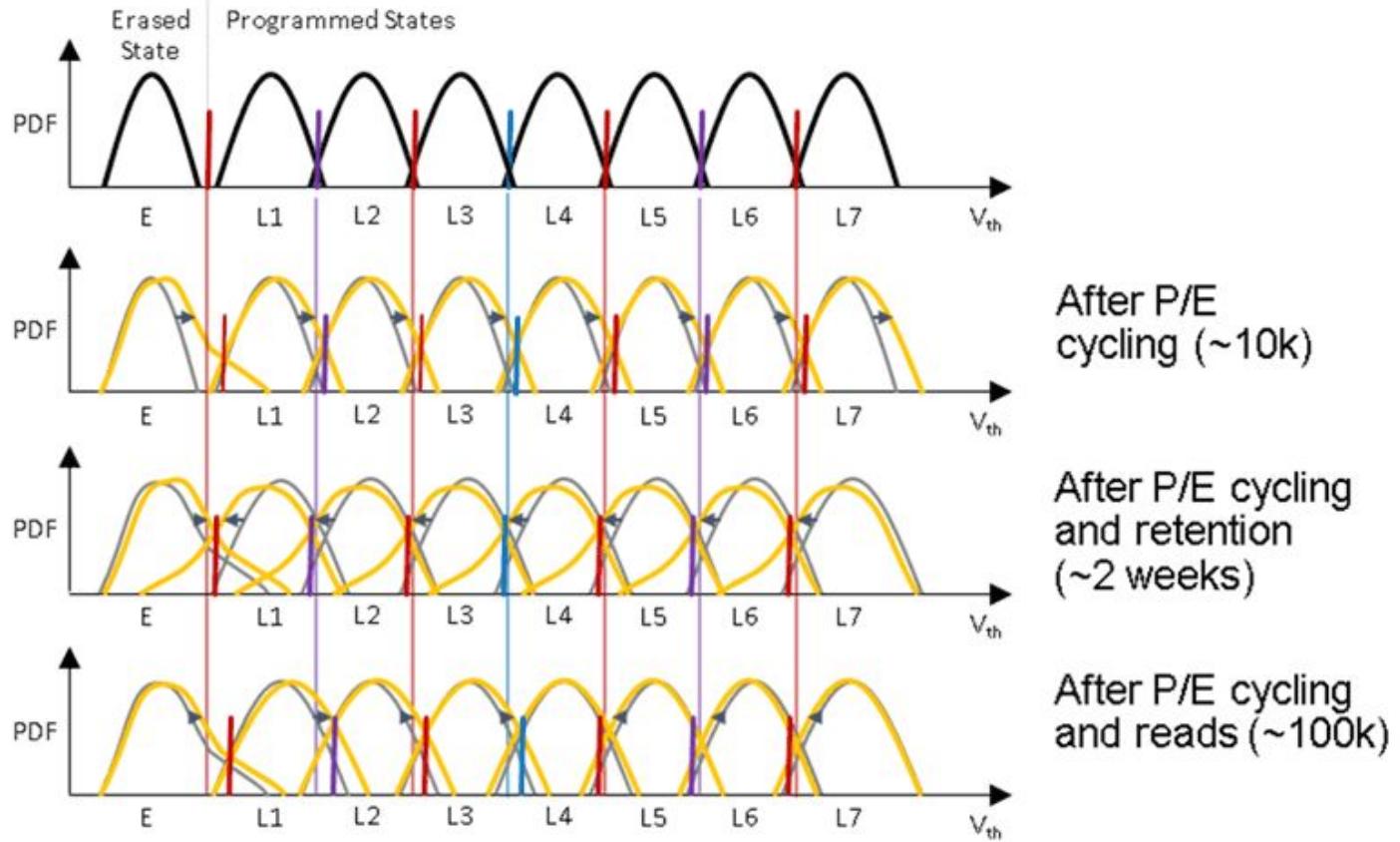
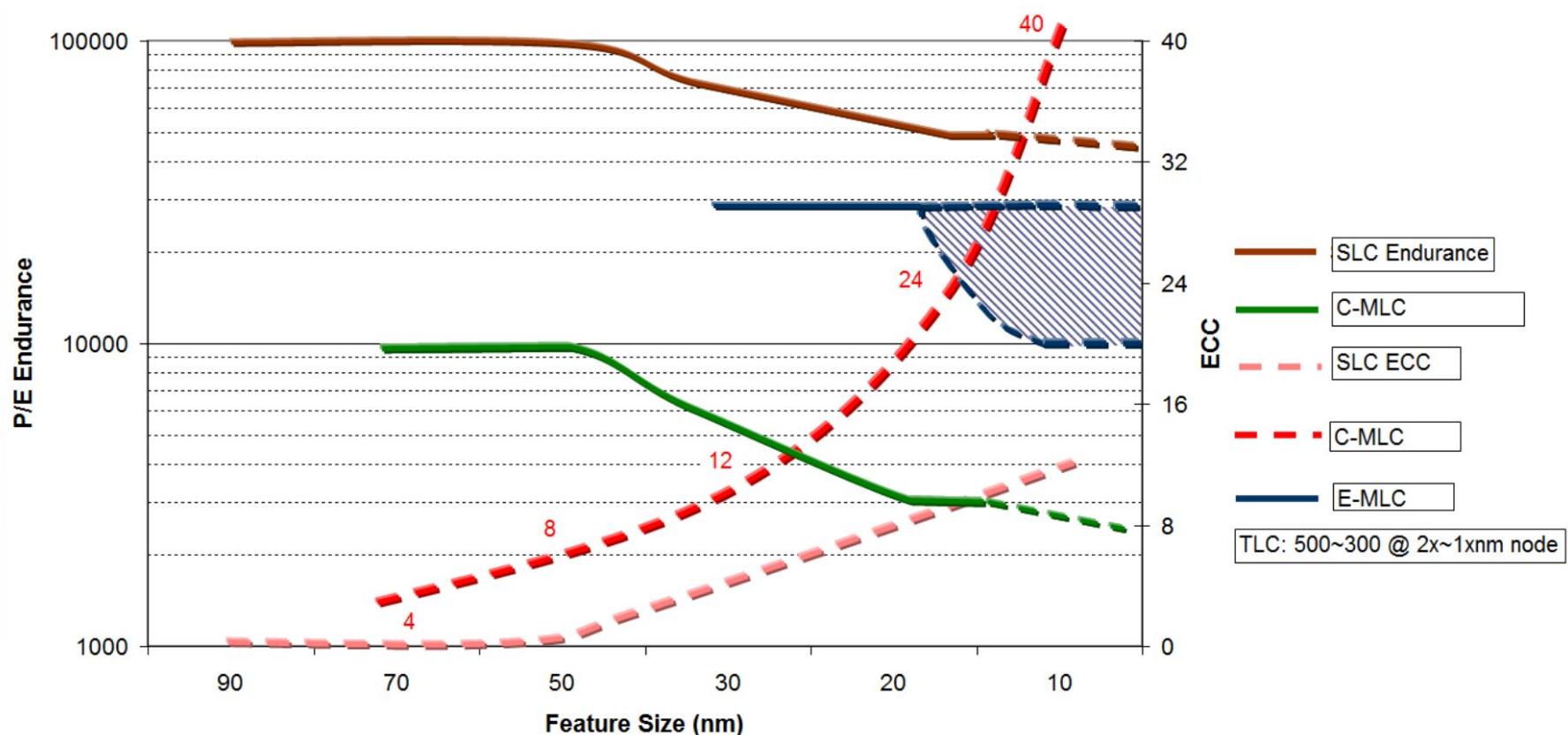


Chart courtesy of Microsemi Corp.

Endurance in MLC or TLC

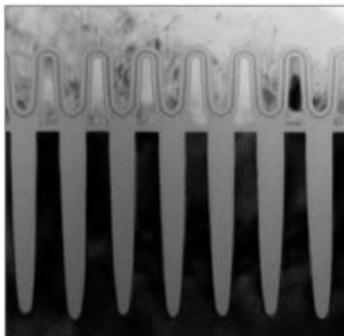
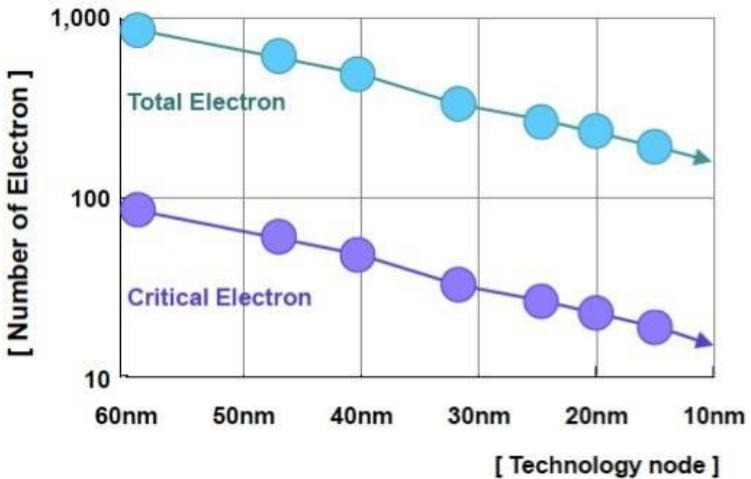


ECC to Improve Endurance



Few Electrons Limitations

FG Limitation : Number of Electrons



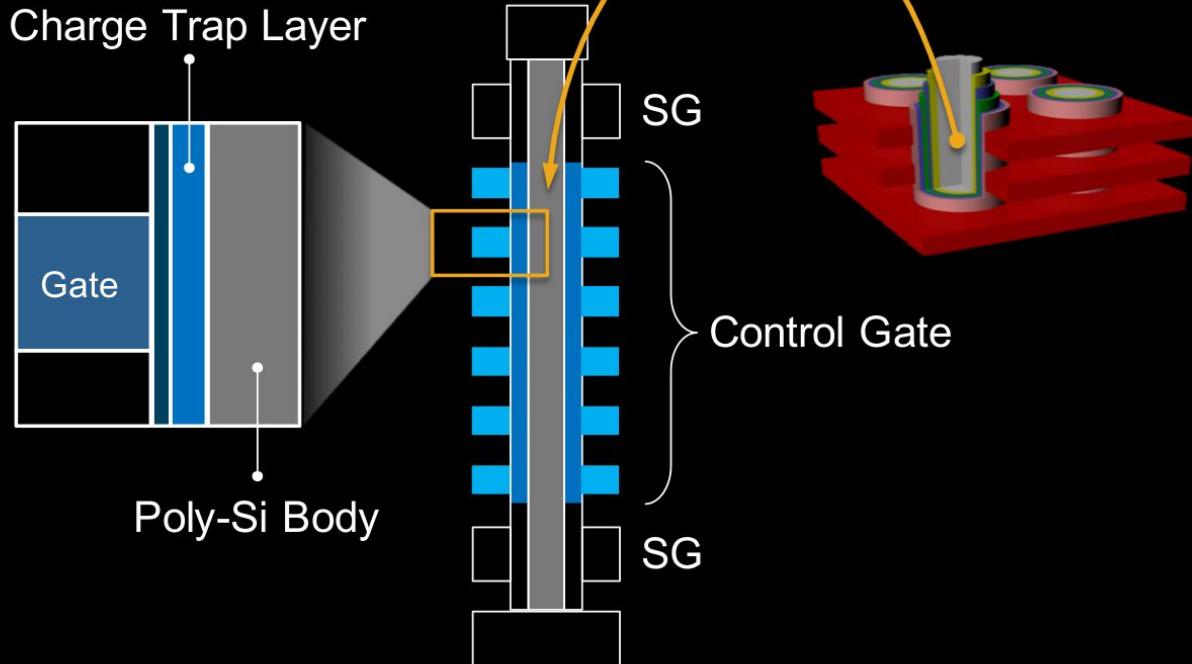
How to Manage 10 electrons in sub-1xnm design rule?



Flash Memories: Present and Future

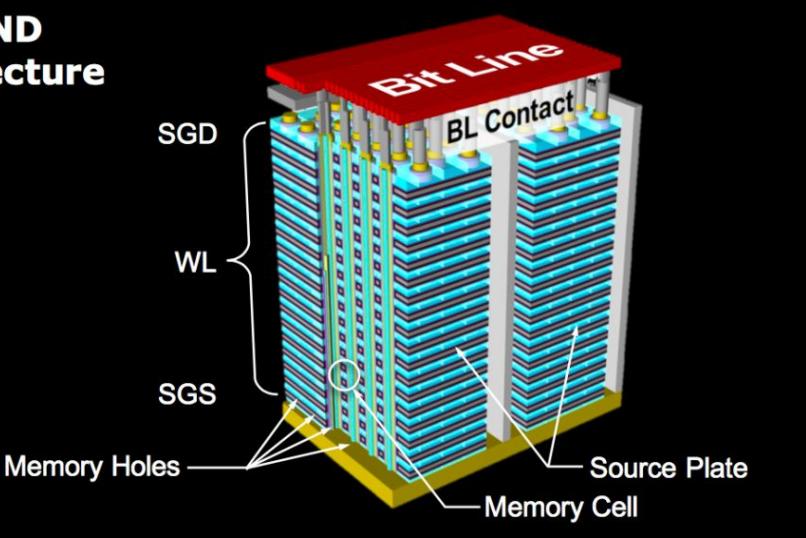
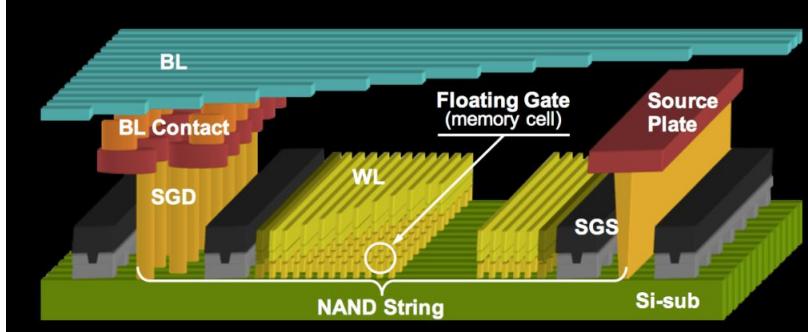
3D NAND Cell

3D NAND Cell

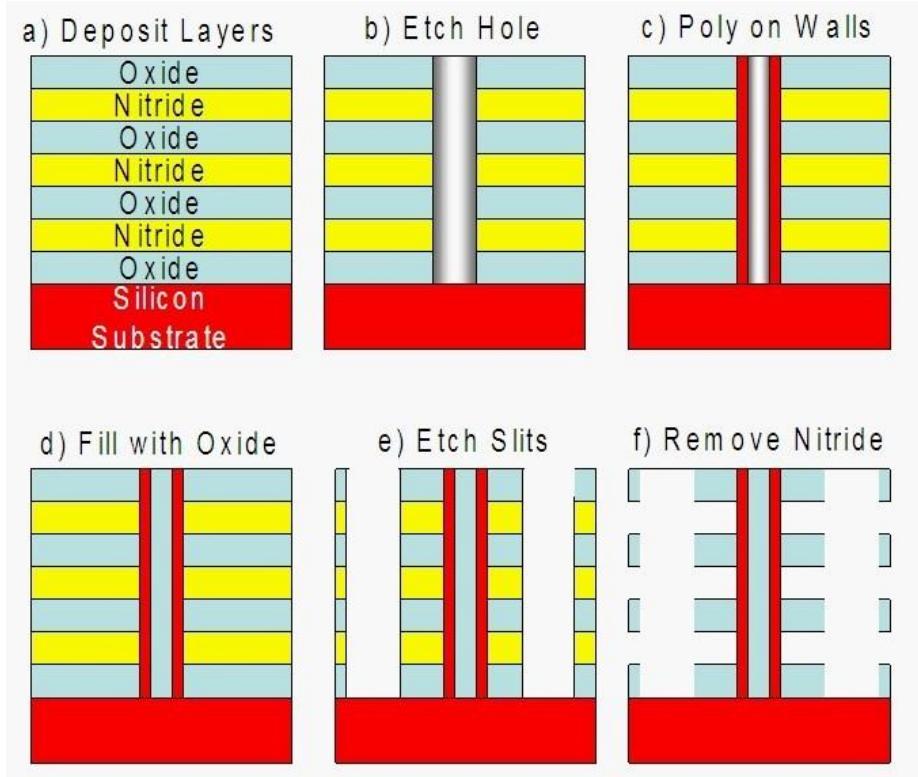


3D NAND

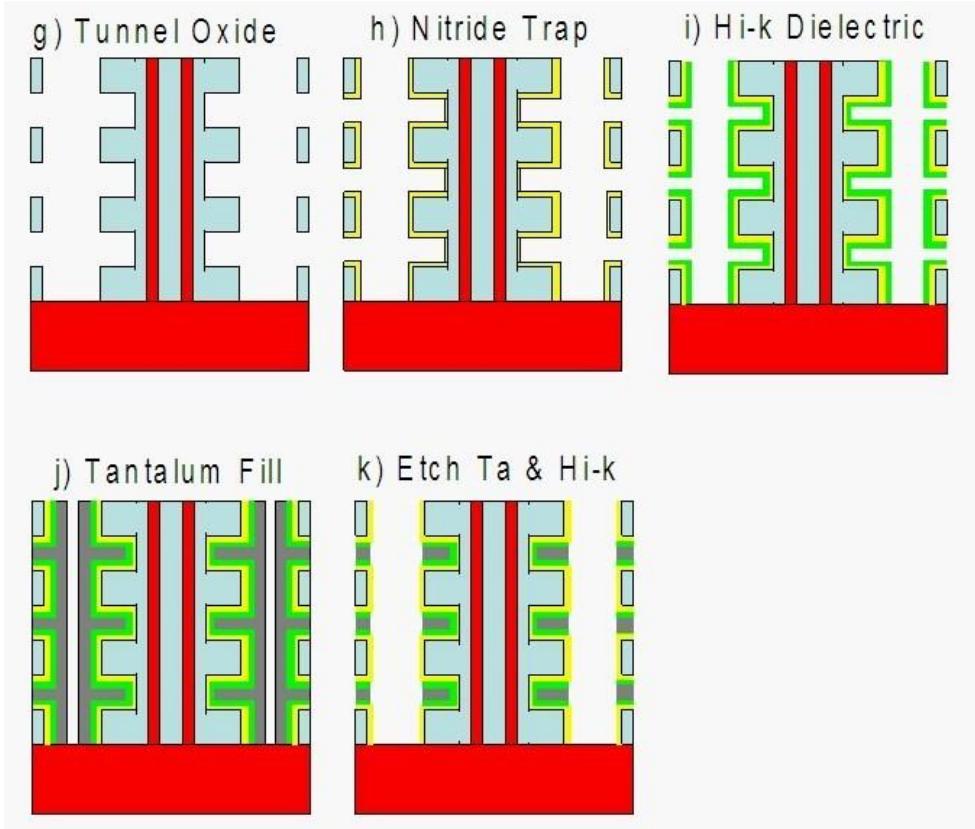
3D NAND Architecture



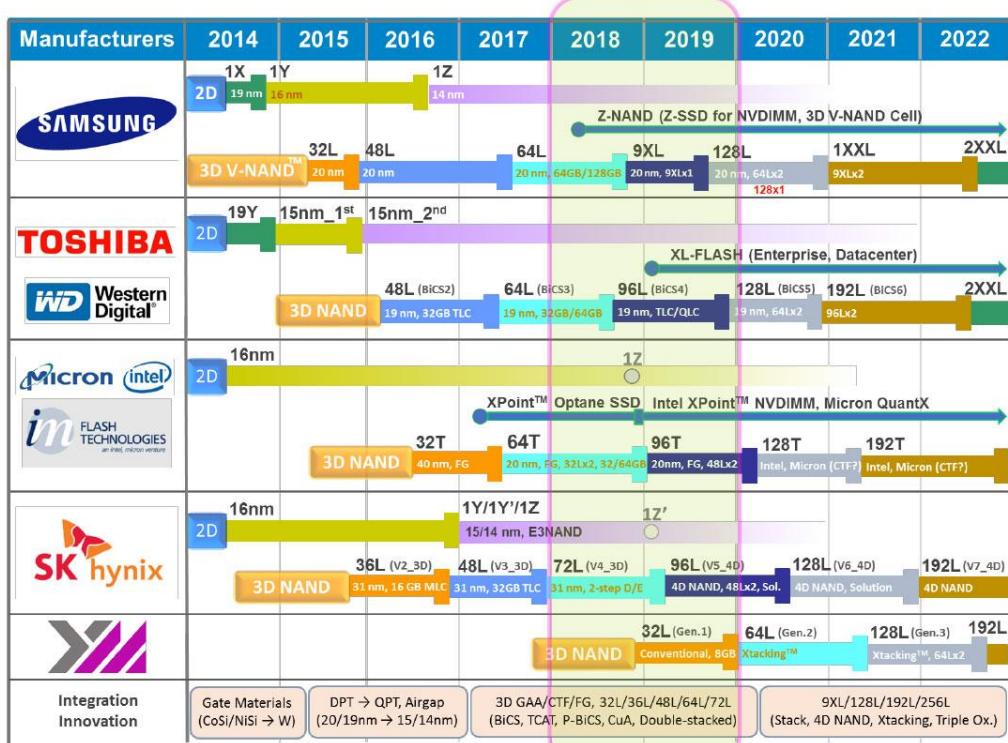
Basic Fabrication Steps



Basic Fabrication Steps



Flash Memory in Recent Times



Tech
Insights

TechInsights memory technology update from IEDM'18

Type of Memories						
	Emerging Memories			Established Memories		
	FeRAM (or FRAM)	MRAM	ReRAM (or RRAM)	PCRAM (or PRAM, PCM)	DRAM	Flash NAND
Nonvolatile	YES	YES	YES	YES	NO	YES
Endurance	High (10^{11})	High (10^{15})	Medium (10^6)	Medium (10^6)	High (10^{15})	Low (10^3)
2012 latest technological node produced (nm)	130 nm	130 nm	R&D	45 nm	30 nm	20 nm
Cell Size (cell size in μm^2)	Large (15-20)	Large/Medium (6-40)	Medium (6-12)	Medium (6-12)	Small (6-10)	Very small (4)
Write speed	Medium (100ns)	High (10 ns)	Medium (75 ns)	Medium (75 ns)	High (10ns)	Low (10 000 ns)
Power Consumption	Low	High/Low	Low	Low	Low	Very High
Cost (\$/Gb)	High (\$ 10 000/Gb)	High (\$ 1000 – 100 /Gb)	R&D	Medium (few \$/ Gb)	Low (\$1/Gb)	Very Low (\$ 0.1/Gb)

Other References :

1. Masuoka, Fujio, et al. "New ultra high density EPROM and flash EEPROM with NAND structure cell." 1987 International Electron Devices Meeting. IEEE, 1987.
2. Masuoka, Fujio, et al. "A new flash E 2 PROM cell using triple polysilicon technology." 1984 International Electron Devices Meeting. IEEE, 1984.
3. Koh, Yohwan. "NAND flash scaling beyond 20nm." 2009 IEEE International Memory Workshop. IEEE, 2009.
4. Lai, Stefan K. "Flash memories: Successes and challenges." IBM Journal of Research and Development 52.4.5 (2008): 529-535.
5. Li, Yan, and Khandker N. Quader. "NAND flash memory: Challenges and opportunities." Computer 46.8 (2013): 23-29.
6. Compagnoni, Christian Monzio, et al. "Reviewing the evolution of the NAND Flash technology." Proceedings of the IEEE 105.9 (2017): 1609-1633.
7. Lee, Seok-Hee. "Technology scaling challenges and opportunities of memory devices." 2016 IEEE International Electron Devices Meeting (IEDM). IEEE, 2016.



Thank You
