THE FLOATING – GATE MEMORY
From Concept to Flash Memory to
THE FOURTH INDUSTRIAL REVOLUTION

S. M. SZE

National Chiao Tung University

Hsinchu, Taiwan

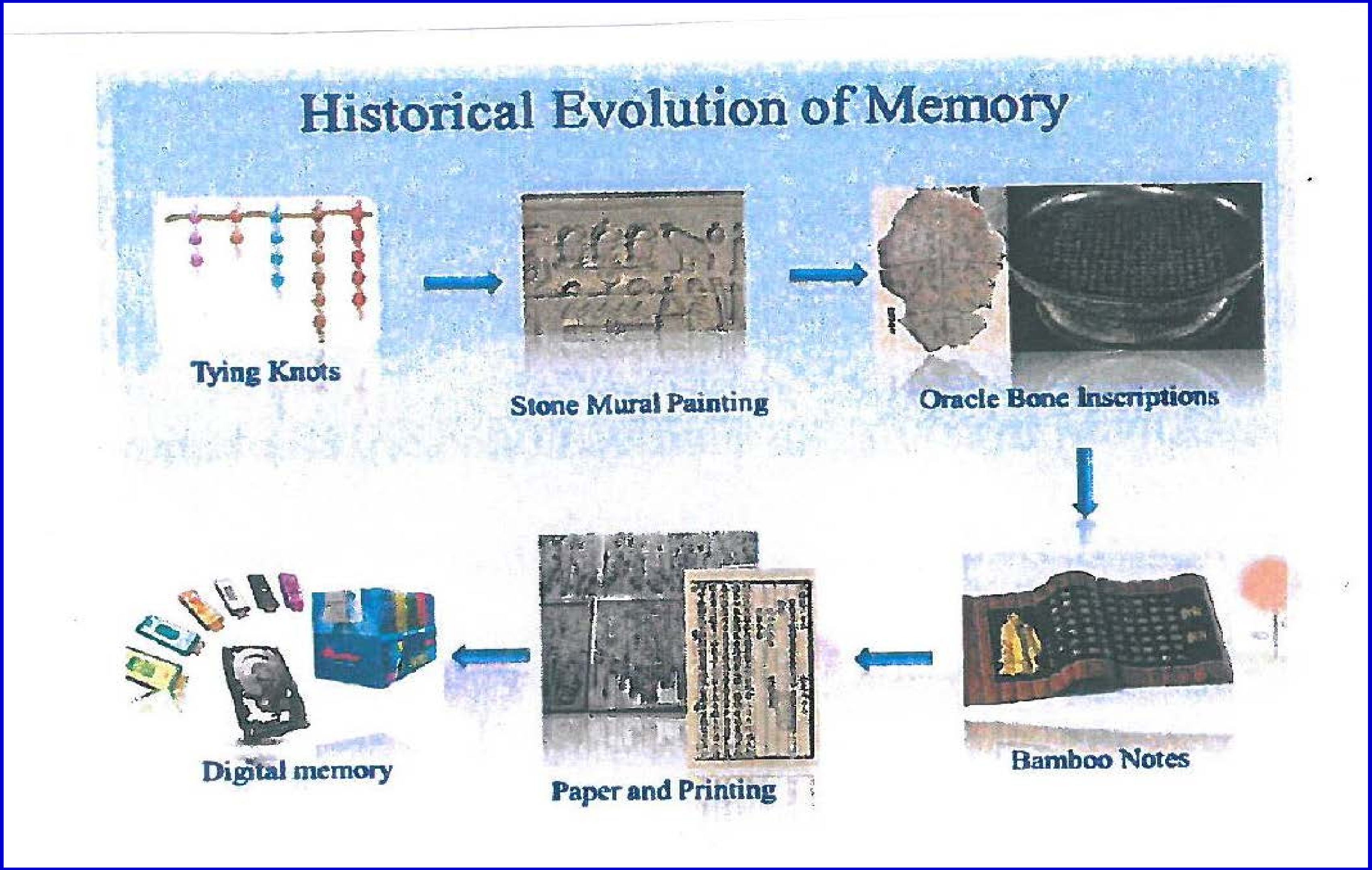
OUTLINE

- INTRODUCTION
- THE FLOATING-GATE MEMORY EFFECT
- HISTORICAL DEVELOPMENT
- APPLICATIONS
- SCALING CHALLENGES
- CONCLUSION

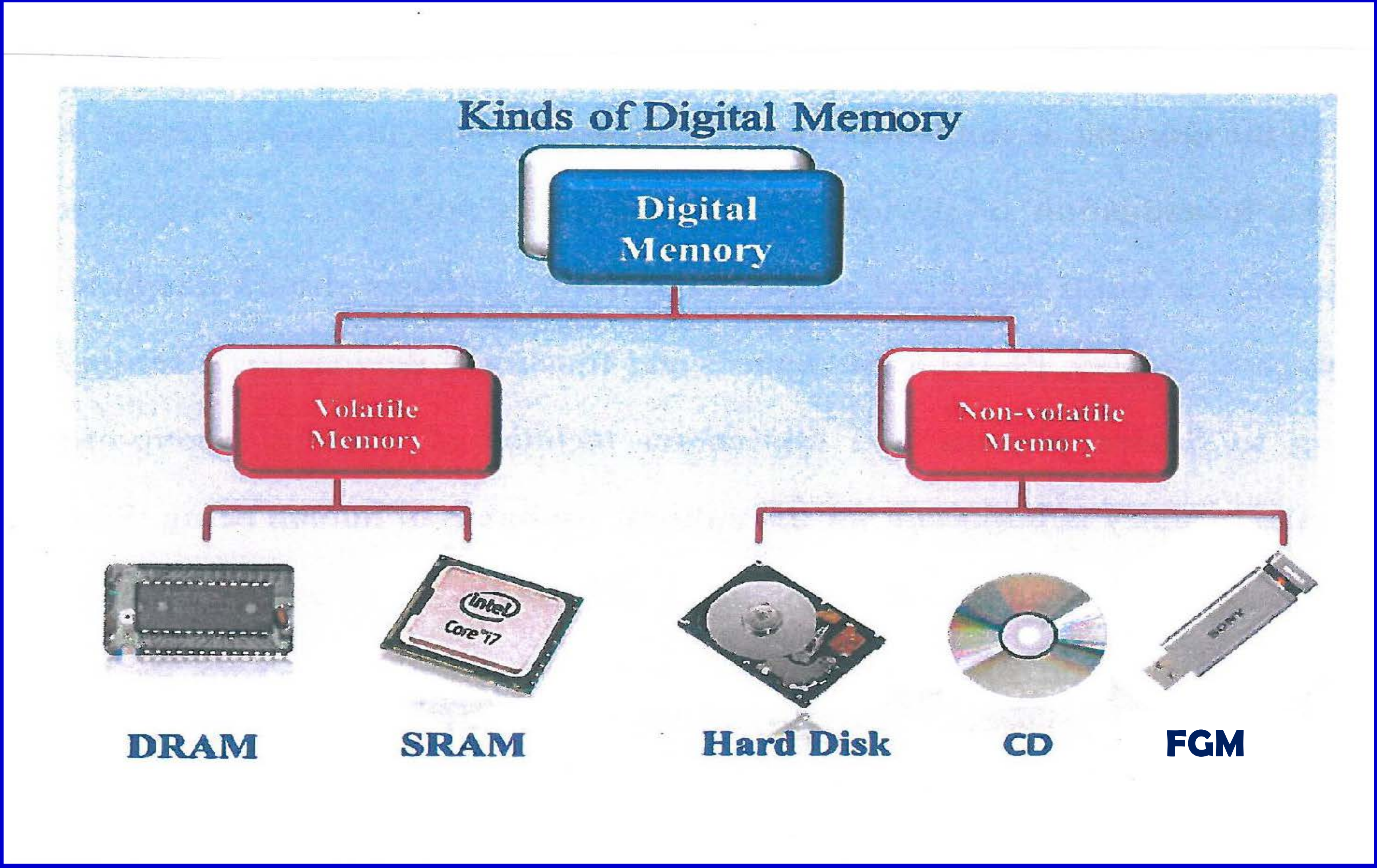
FOUR INDUSTRIAL REVOLUTIONS

- First (1769-1830)： *Steam Engine* — to power coal mines, textile mills, railroads, ships, etc
- Second (1876-1914)： *Electricity* — to enable the invention of lightbulbs, telephone, phonograph, motion picture, etc
- Third (1947-2000)： *Transistor* — to lead to computer, microchip, internet, etc
- Fourth (2000-)： *Floating-gate memory* — to empower technology for global connectivity, artificial intelligence, IoT, big data, robotics, etc

HISTORICAL EVOLUTION OF MEMORY



CLASSIFICATION OF DIGITAL MEMORY



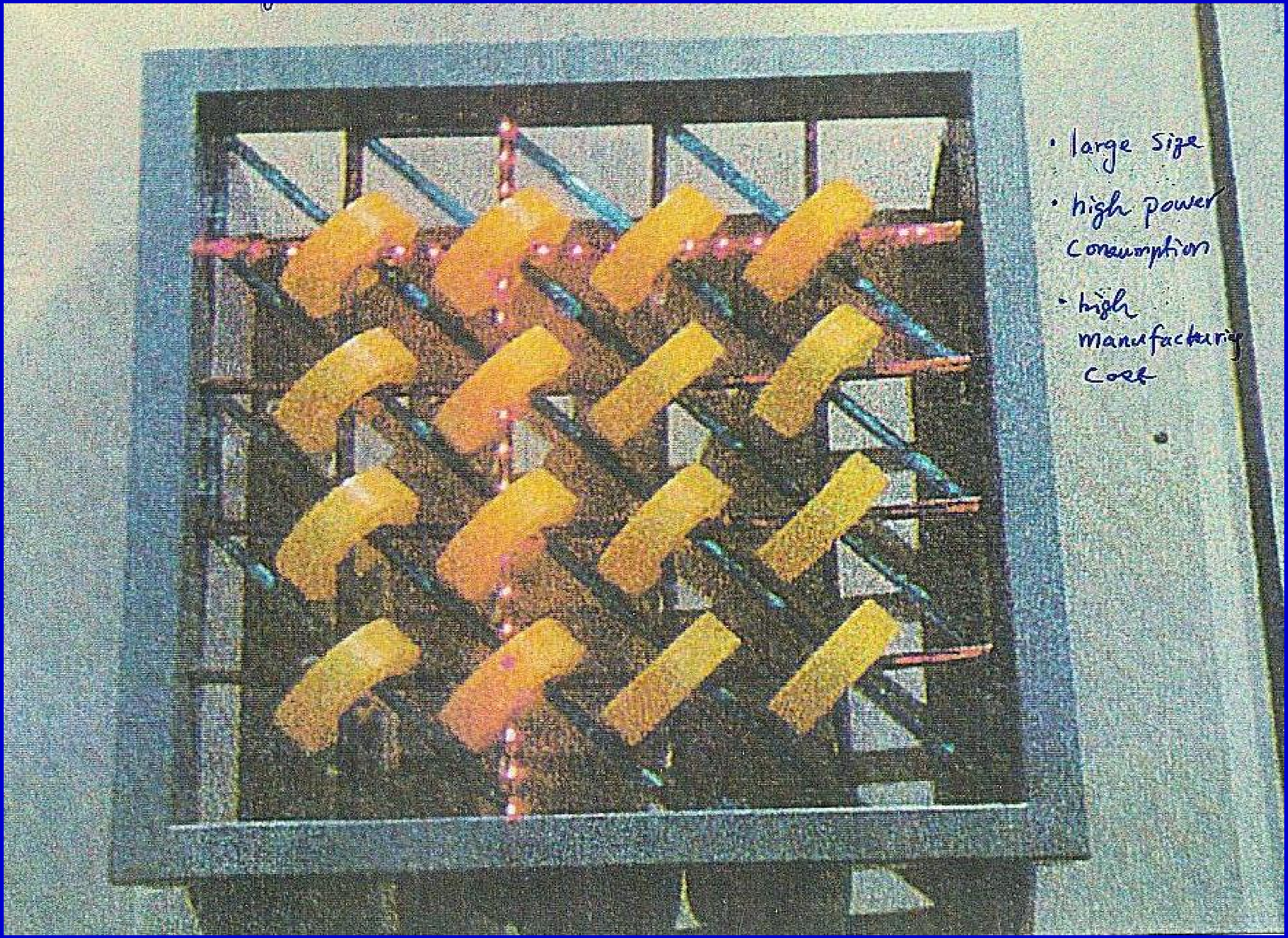
DRAWBACKS OF HDD AND OPTICAL DISK

- Low density
- Slow access time
- Large form factor
- High power consumption
- Non-compatibility with CMOS technology

- **THE FLOATING-GATE MEMORY (FGM) EFFECT**

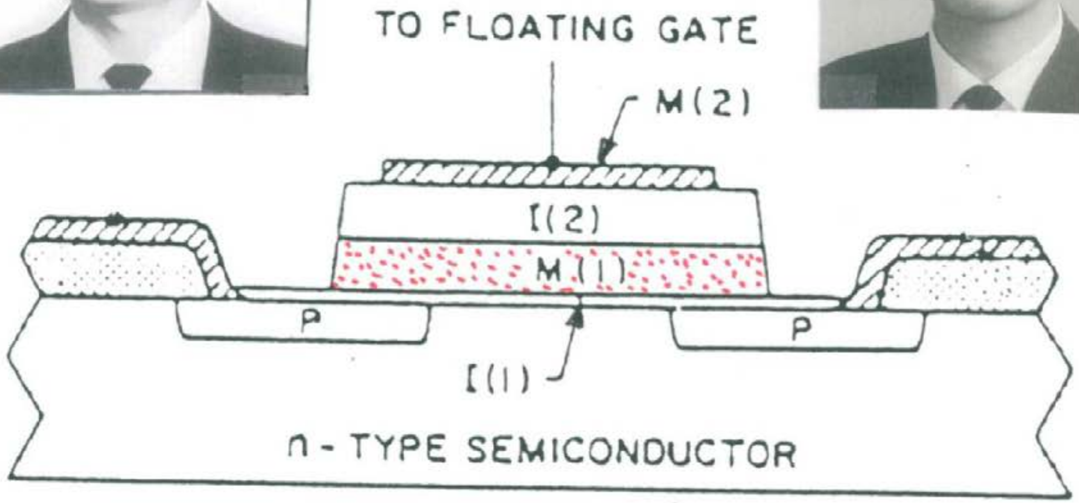
MAGNETIC CORE MEMORY INVENTED BY Dr. A. WANG (1949)

— The non-volatile memory for mainframe computers and communication equipment in the 1950s to 1970s



DISCOVERERS OF THE FLOATING-GATE MEMORY EFFECT

D. Kahng and S. M. Sze in May 1967



(1967)

FIRST PAPER ON FLOATING-GATE MEMORY EFFECT

(by D. Kahng and S. M. Sze in May 1967)

THE BELL SYSTEM

TECHNICAL JOURNAL

VOLUME XLVI | JULY–AUGUST 1967 | NUMBER 6

A Floating Gate and Its Application to Memory Devices

By D. KAHNG and S. M. SZE

(Manuscript received May 16, 1967)

A structure has been proposed and fabricated in which semi-permanent charge storage is possible. A floating gate is placed a small distance from an electron source. When an appropriately high field is applied through an outer gate, the floating gate charges up. The charges are stored even after the removal of the charging field due to much lower back transport probability. Stored-charge density of the order of $10^{12}/cm^2$ has been achieved and detected by a structure similar to an metal-insulator-semiconductor (MIS) field effect transistor. Such a device functions as a bistable memory with nondestructive read-out features. The memory holding time observed was longer than one hour. These preliminary results are in fair agreement with a simple analysis.

It has been recognized for some time that a field-effect device, such as that described by Shockley and Pearson,[1] can be made bistable utilizing switchable permanent displacement charges on ferroelectric material.[2] Subsequent studies of ferroelectric material have revealed,[3] however, that the inherent speed capability of a device incorporating a ferroelectric material is limited by domain motion, whose highest speed is limited by the acoustic velocity. In the absence of highly ordered, near-ideal thin film ferroelectric material, the speed capability of a bistable device, therefore, is in the microsecond range at best.[4] In addition, many ferroelectric materials suffer from irreversible mechanical disorder after many cycles of polarization switching,[2] rendering some uncertainty on the long term device reliability aspect.

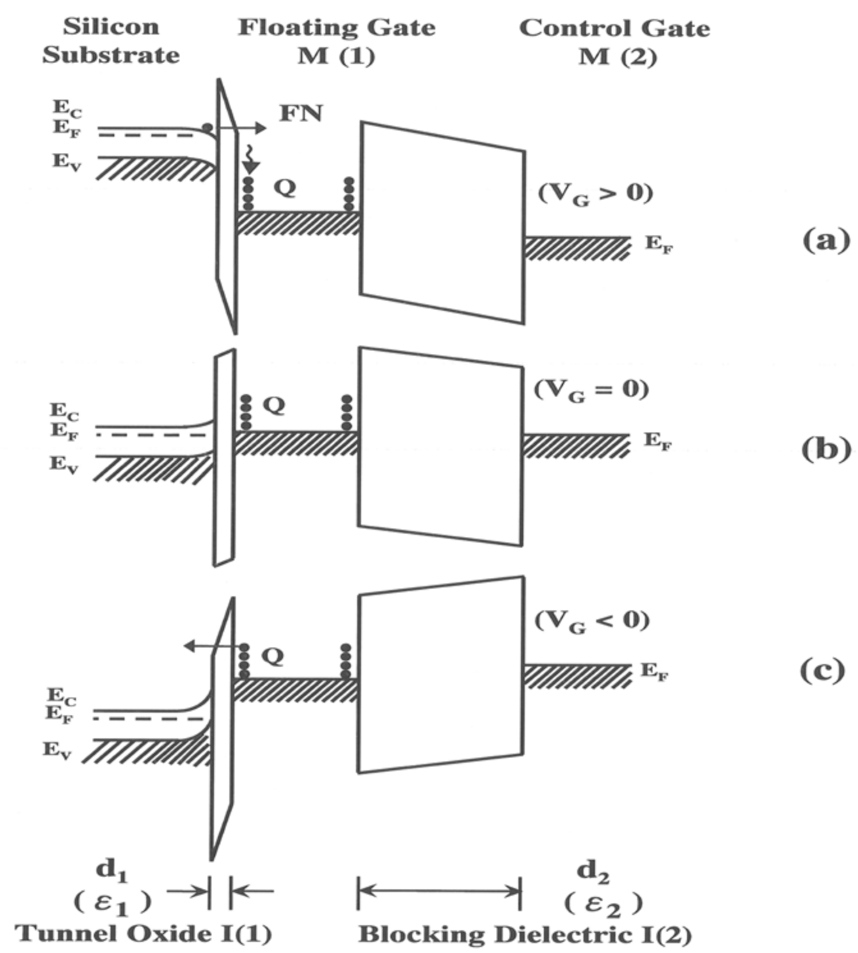
FIRST PAPER ON FGM EFFECT (1967)

- **In the mid 1960s magnetic core memory was the dominant nonvolatile memory for computers. Such memory was considered not possible in a semiconductor because of its *short lifetime* (~1ms)**

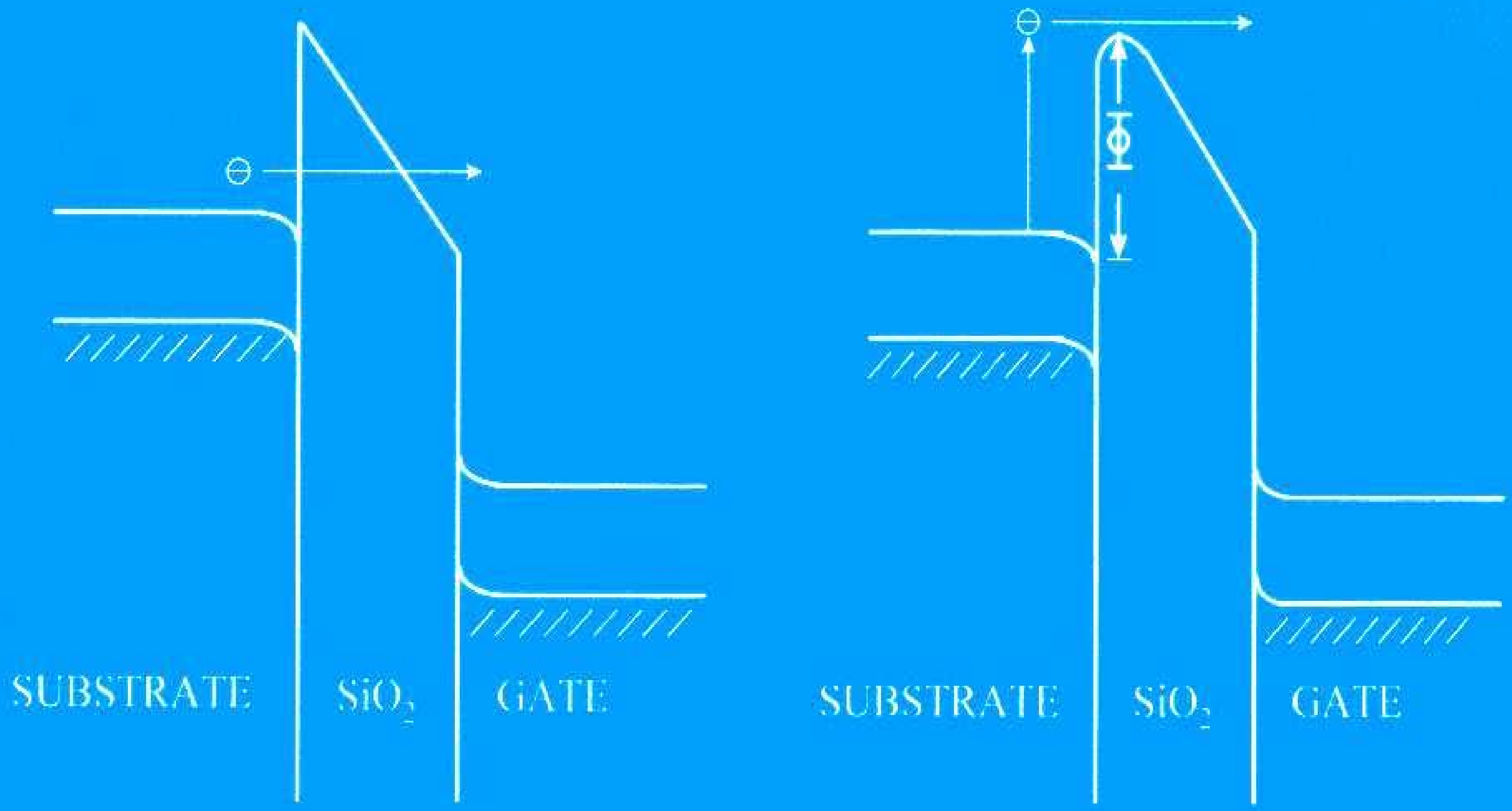
- **The possibility of nonvolatile storage in semiconductors device was recognized for the *first time* in the 1967 paper, and an experimental EEPROM was demonstrated (>1 hr storage time)**

- **The paper introduced not only the *basic concept of* nonvolatility in semiconductor, but also the *floating-gate structure* which has been the dominant technology for nonvolatile information storage**

BAND DIAGRAM OF FGM OPERATIONS (Programming, Storage, and Erase)



CHARGE INJECTION MECHANISMS



(a) FOWLER-NORDHEIM TUNNELING (b) HOT-ELECTRON INJECTION

CURRENT DENSITY EQUATIONS

For Fowler-Nordheim Tunneling

$$J = C_1 \mathrm{E}^2 \exp\left(- \mathrm{E}_0 / \mathrm{E} \right)$$

$$\mathrm{E} = \frac{V_G}{d_1 + d_2 \left(\in_1 / \in_2 \right)} - \frac{Q}{\in_1 + \in_2 \left(d_1 / d_2 \right)}$$

$$Q = \int_0^t J dt' = \text{stored charge}$$

For Hot-Electron Injection

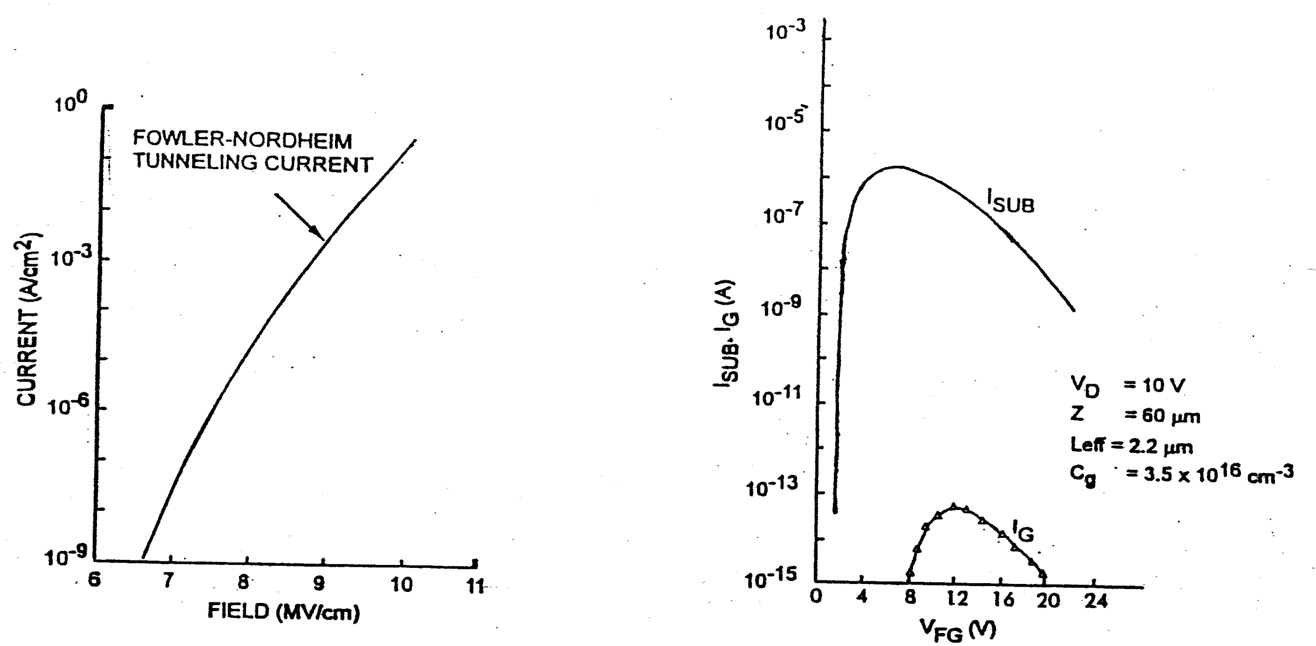
$$J = C_2 \mathrm{E}_m^2 \exp\left(- \Phi / \lambda \mathrm{E}_m \right)$$

E_m = Maximum Channel Lateral Field

λ = Inelestic Scattering Length

Φ = Barrier Height between Si and SiO_2

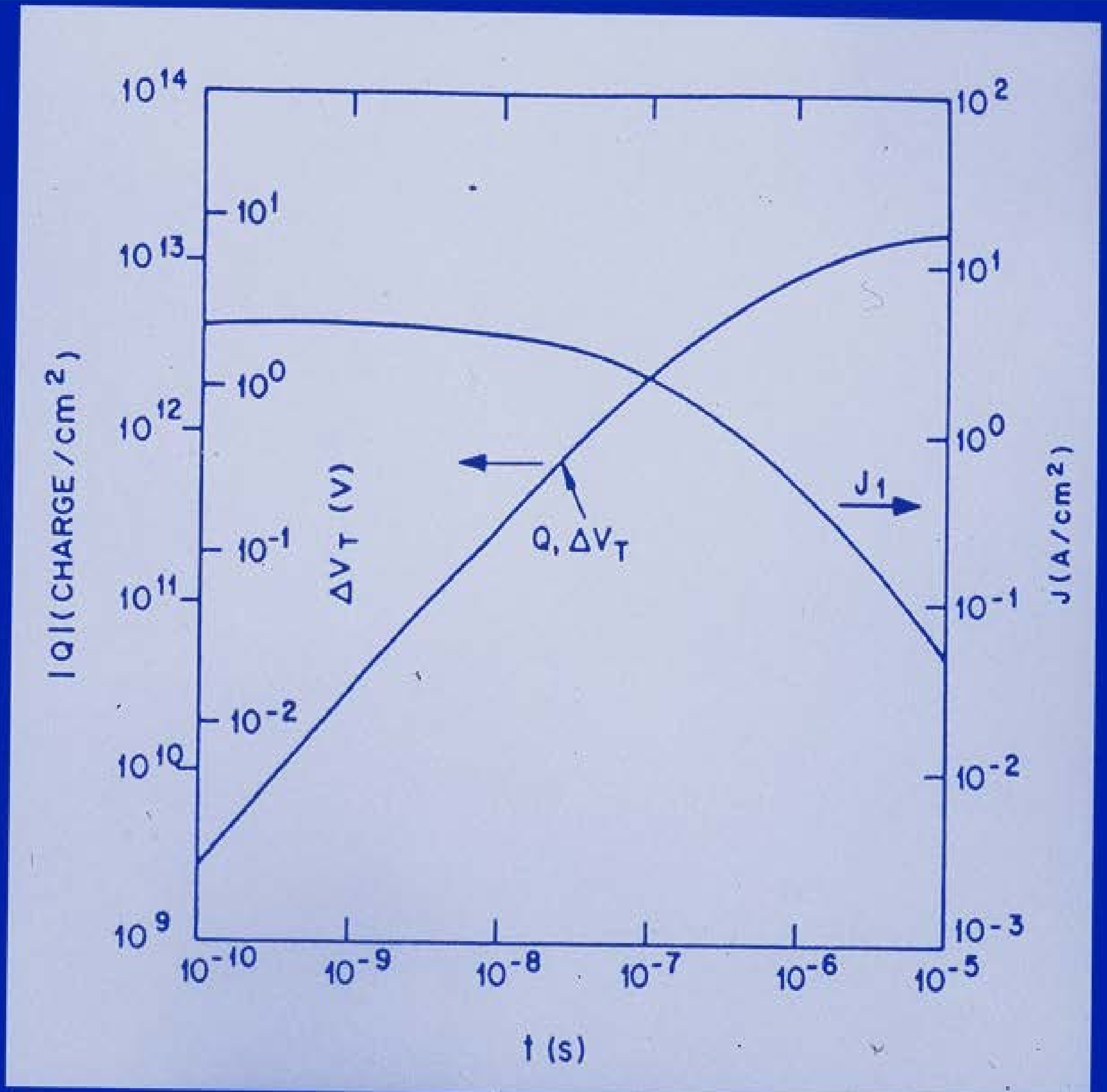
GATE CURRENT VERSUS GATE VOLTAGE



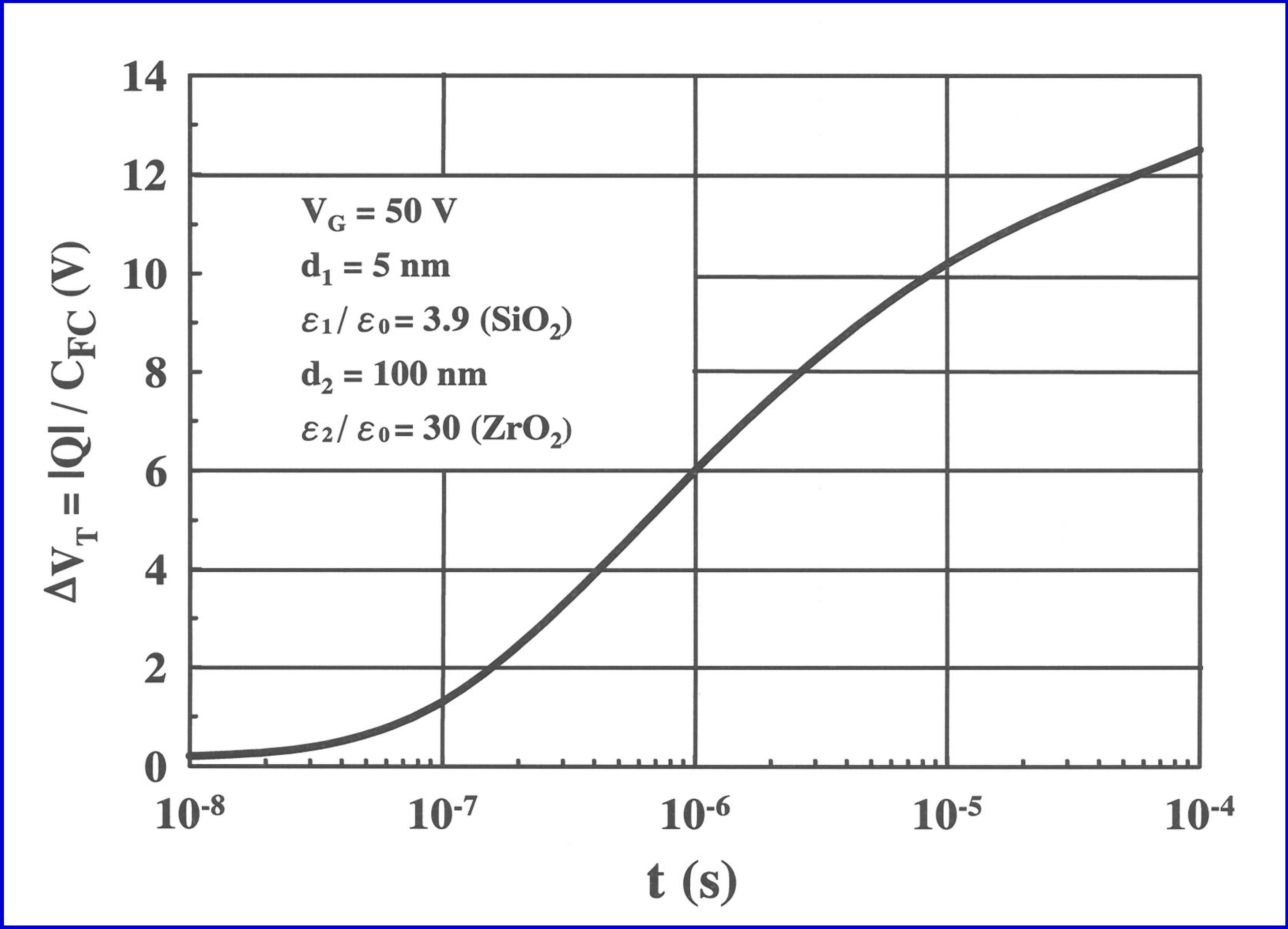
(a) Fowler-Nordhem Tunneling

(b) Hot-Electron Injection

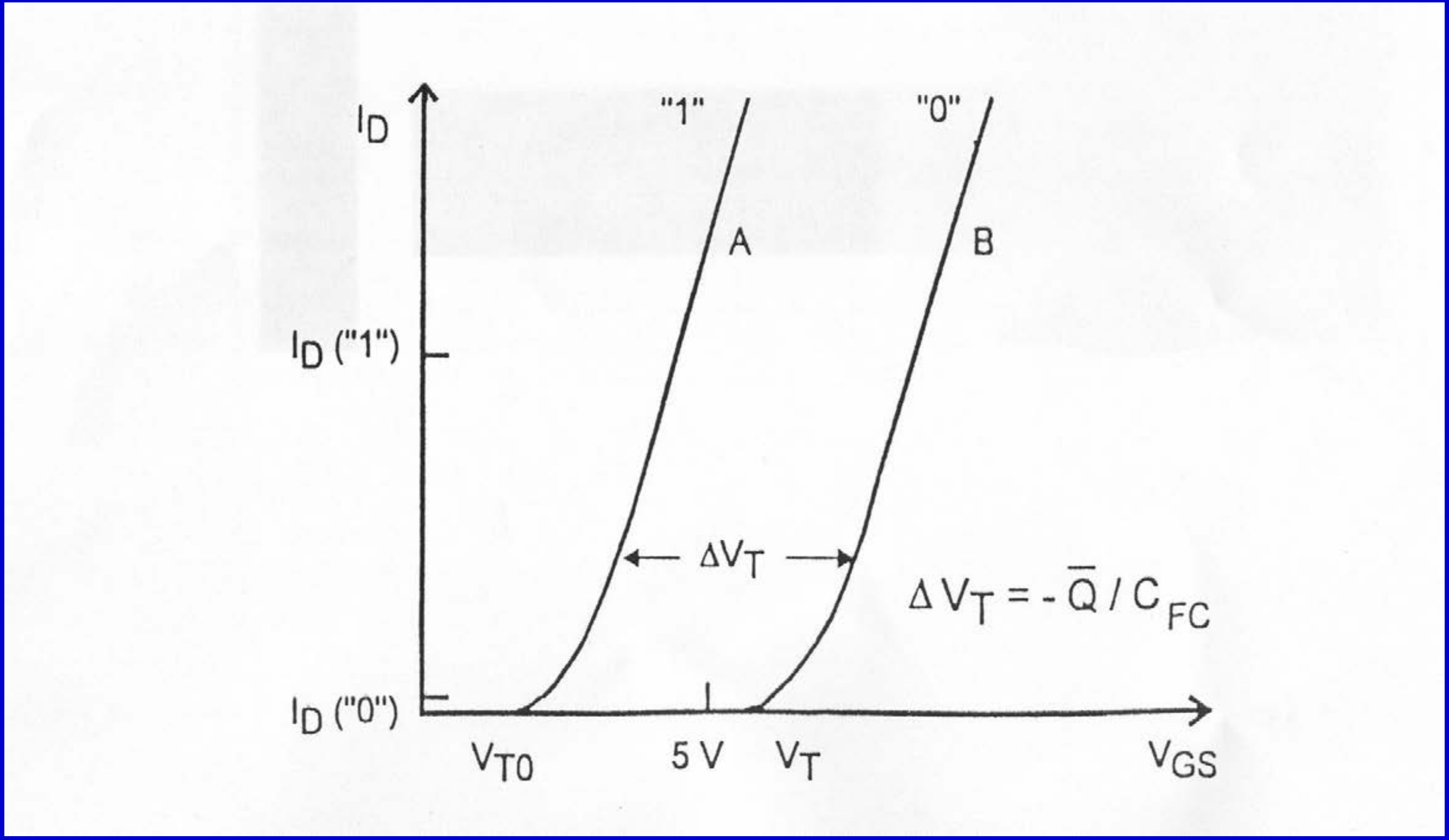
PROGRAMMING CURRENT AND STORED CHARGE



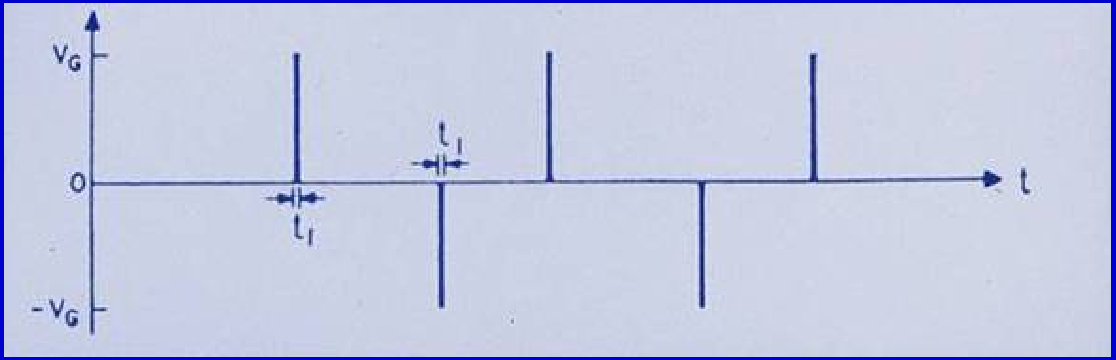
THRESHOLD VOLTAGE SHIFT VERSUS TIME



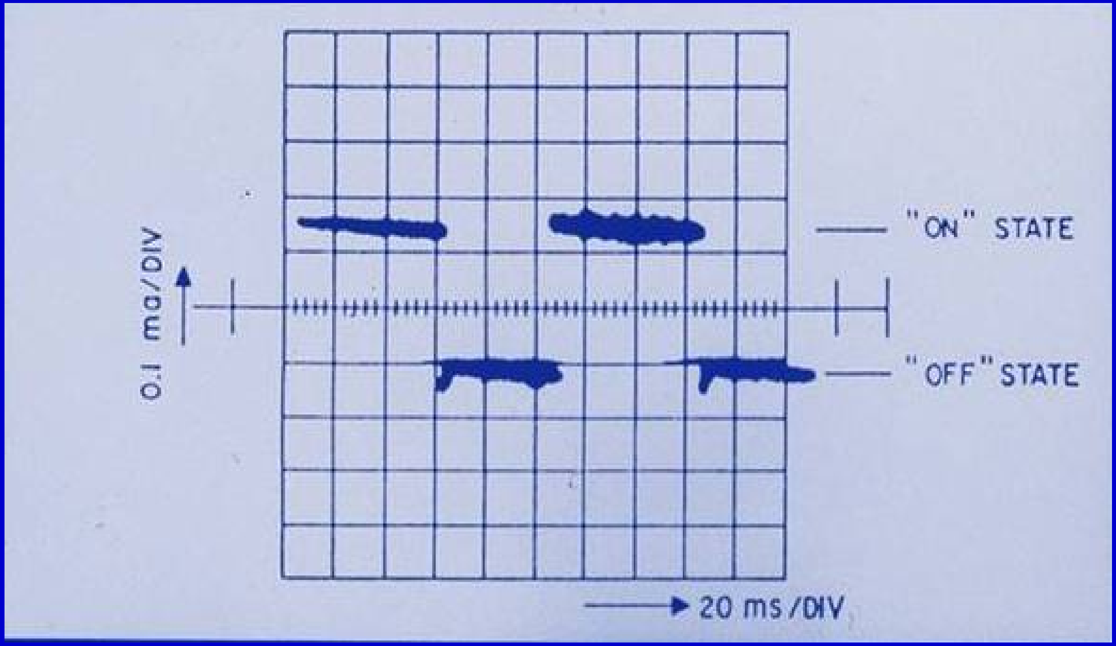
THRESHOLD VOLTAGE SHIFT DUE TO CHARGE STORAGE IN THE FLOATING GATE



FIRST EXPERIMENTAL RESULT OF FGM OPERATION IN 1967
(The storage time was one hour)

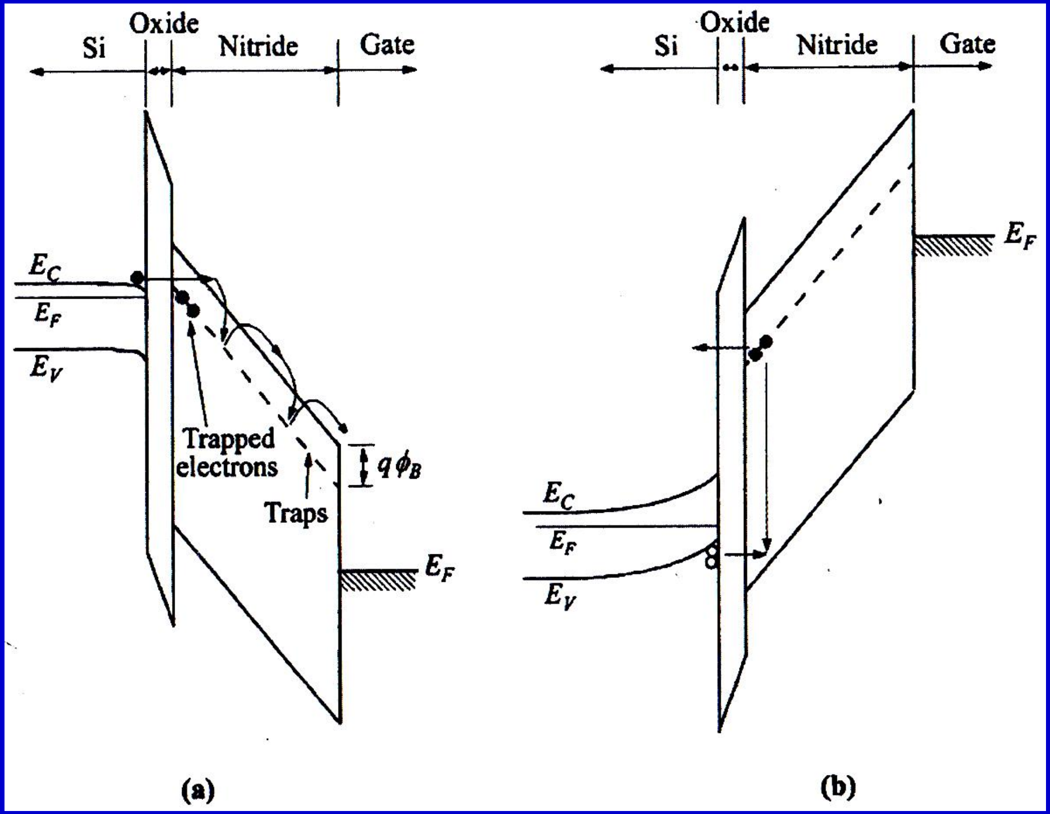


(a) Applied gate pulse voltage



(b) Source-drain current
For $V_G = \pm 50$ V, $t_1 = 0.5$ μs

FLOATING–TRAP OR CHARGE–TRAPPING MEMORY
(A Limiting Case of Floating-Gate Memory)



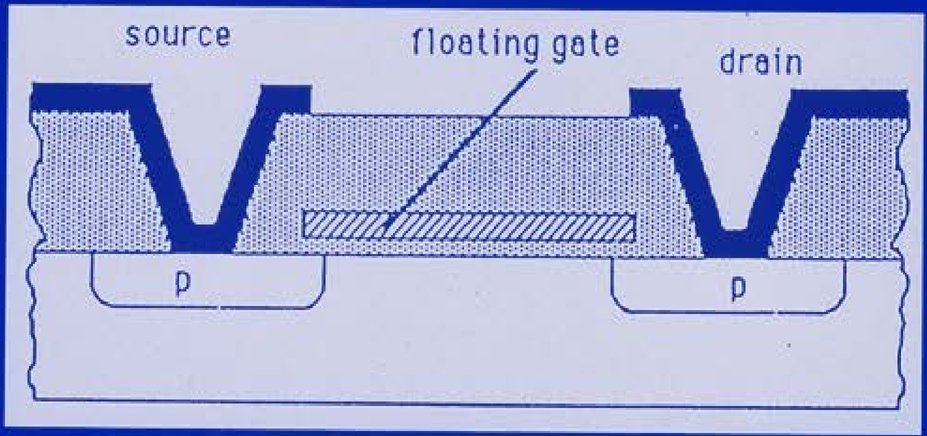
- Charge-trapping memory (CTM)： when the thickness of the floating gate approaches zero and the charges are stored in the upper insulator
- CTM is based on the same MOS structure and the same charge-storage principle as FGM

- **HISTORICAL DEVELOPMENT**

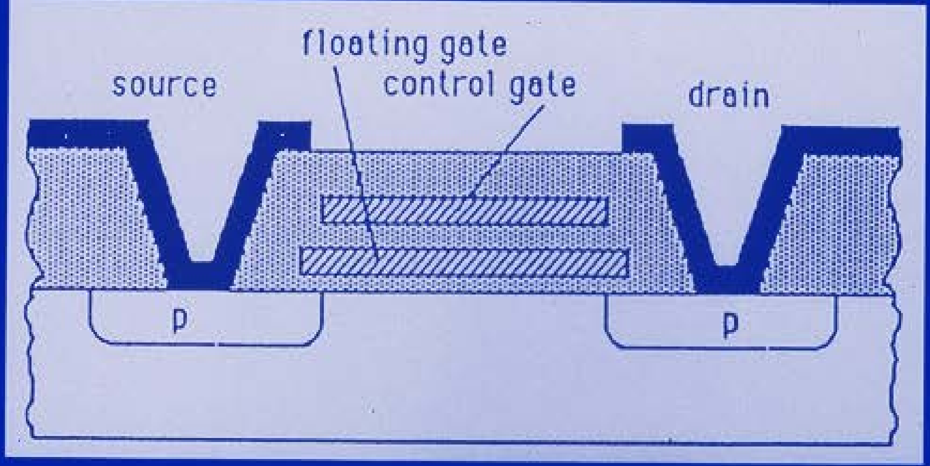
HISTORY OF FGM

YEAR	DEVICE	DISCOVERER (S)/INVENTOR(S)	ORGANIZATION
1967	Floating-Gate Concept	Kahng and Sze	Bell Labs
1971	EPROM-FAMOS	Frohman-Bentchkowsky	Intel
1976	EEPROM-SAMOS	Iizuka et al.	Toshiba
1984	Flash Memory	Masuoka et al.	Toshiba
1987	NAND Flash	Masuoka et al.	Toshiba
1995	Multilevel Cell	Bauer et al.	Intel
2007	Multi-layer Integration (BiCS)	Fukuzumi et al.	Toshiba
2014	128Gb 2b/cell 2D NAND Flash	Helm et al.	Micron
2016	768Gb 3b/cell 3D NAND Flash	Tanaka et al.	Micron Intel
2018	1Tb 4b/cell 3D NAND Flash	Lee et al.	Samsung

TWO EARLY VERSIONS OF FLOATING GATE MEMORY

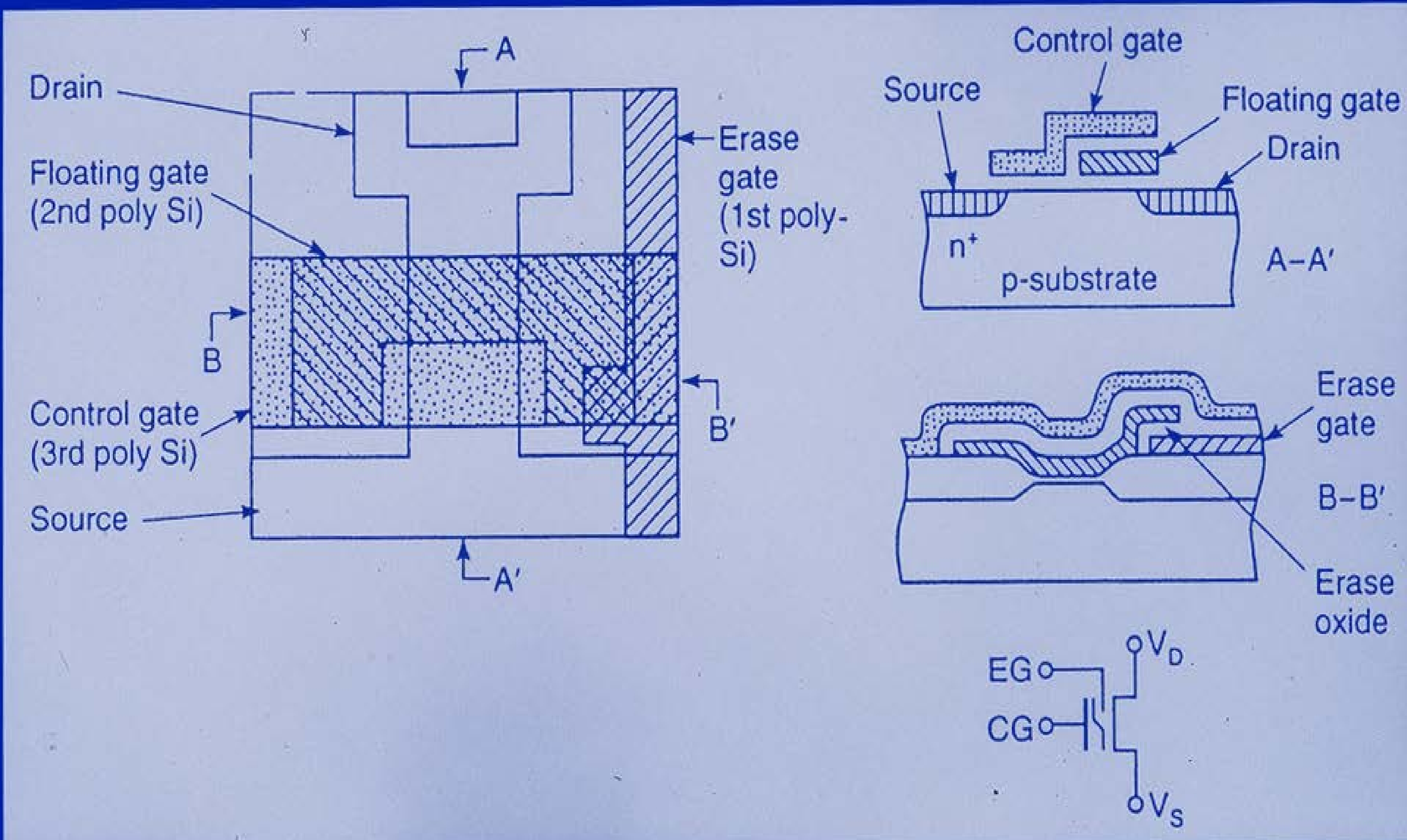


In 1971
FAMOS (Floating Gate Avalanche Injection MOS)

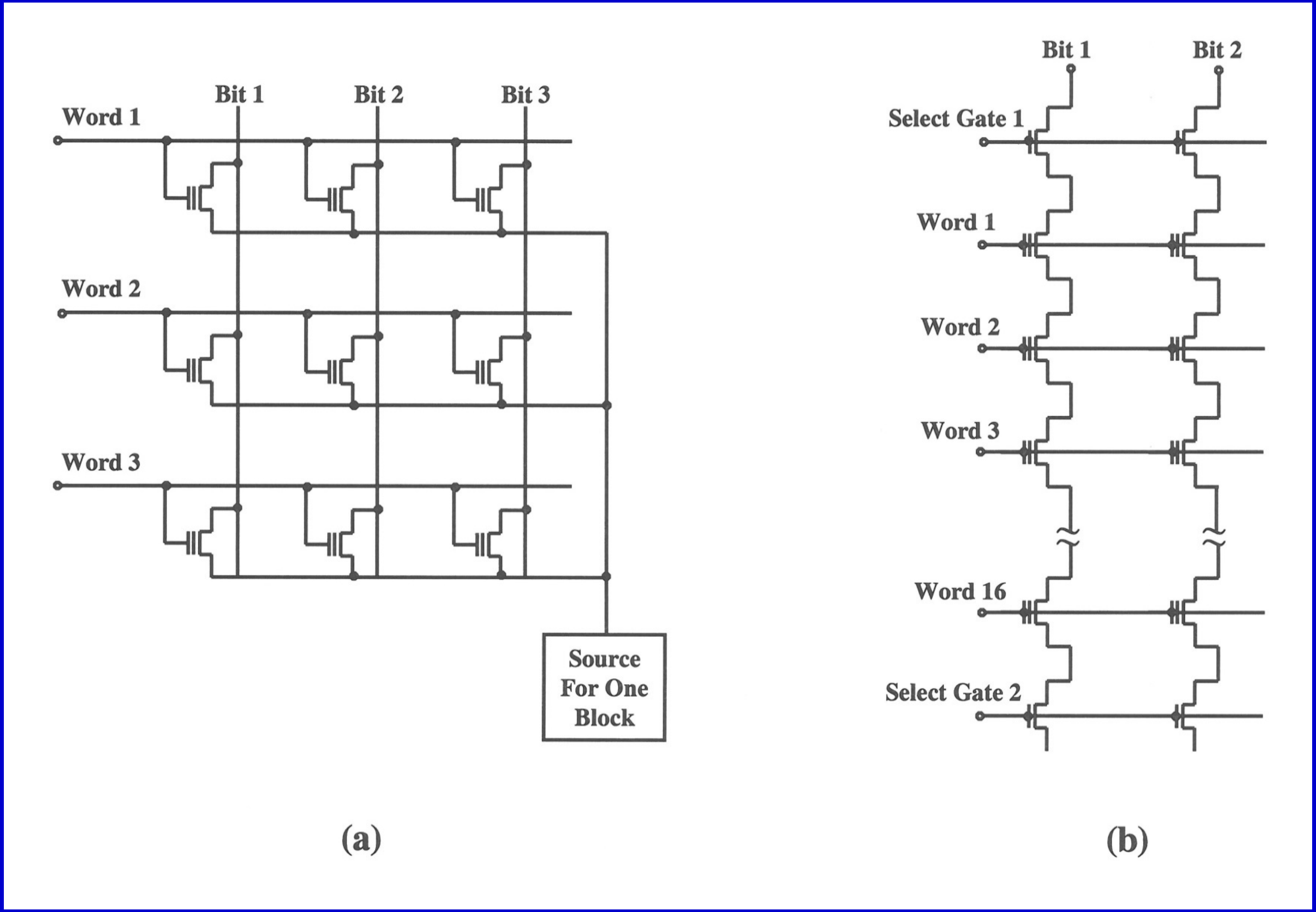


In 1976
SAMOS (Stacked Gate Avalanche Injection MOS)

TOP AND CROSS-SECTIONAL VIEW OF A FLASH MEMORY (1984)



NOR AND NAND ARCHITECTURES



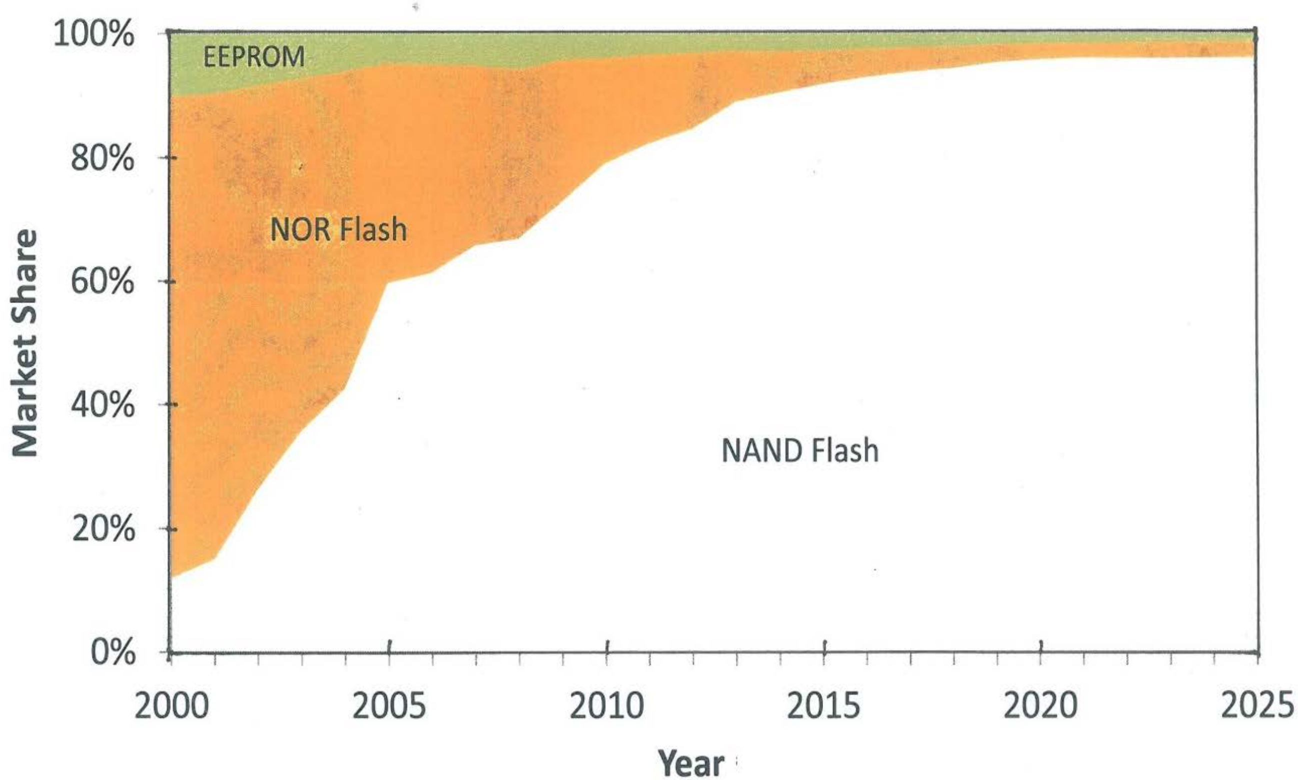
CLASSIFICATION OF FGM

- EPROM (Electrically Programmable Read Only Memory)

- EEPROM (Electrically Erasable Programmable Read Only Memory)

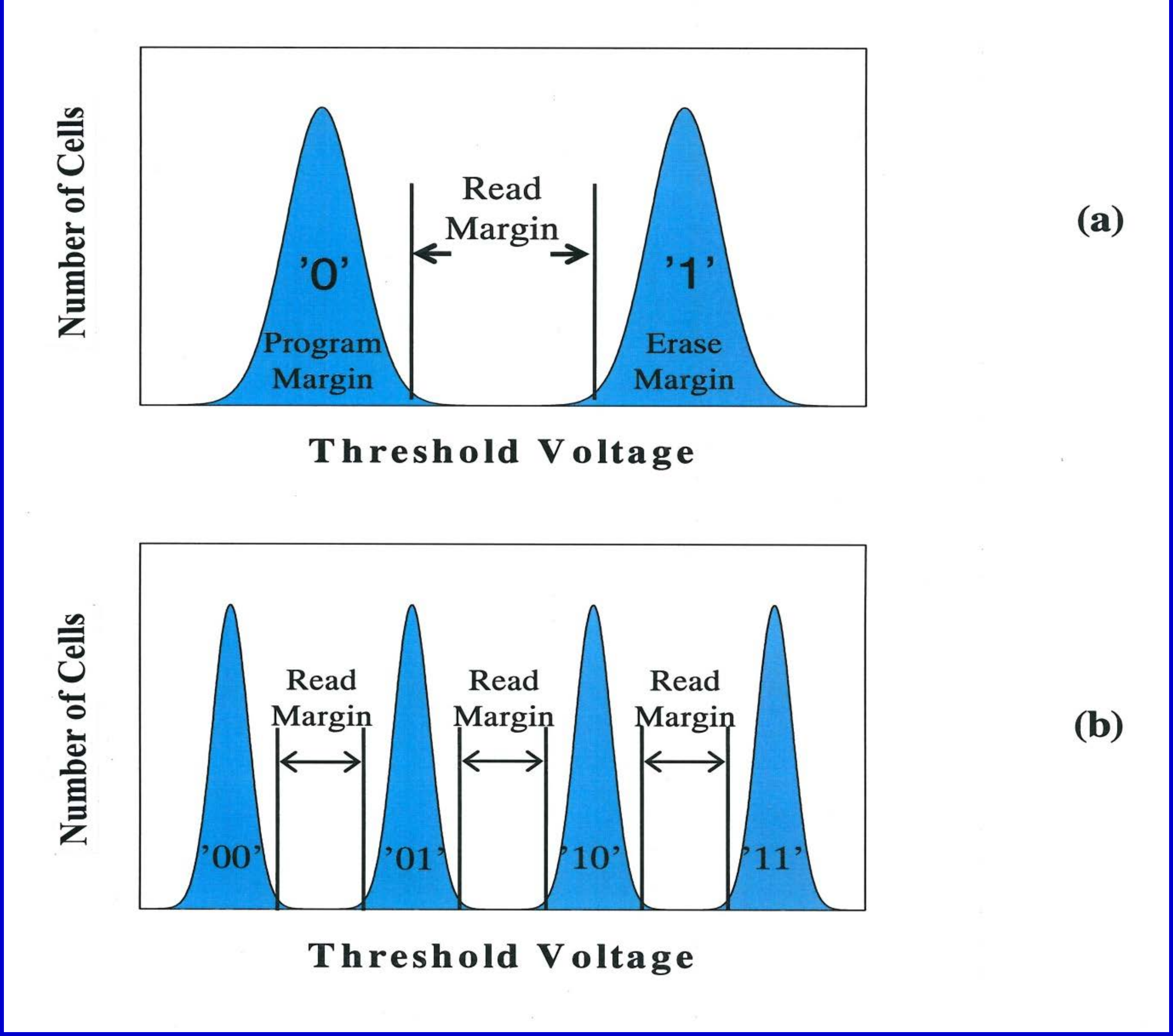
- Flash Memory
  - NOR Flash (with a basic unit of one memory device)
  - NAND Flash (with a basic unit of 16 or more memory devices)

MARKET SHARE OF THREE NVSM PRODUCTS



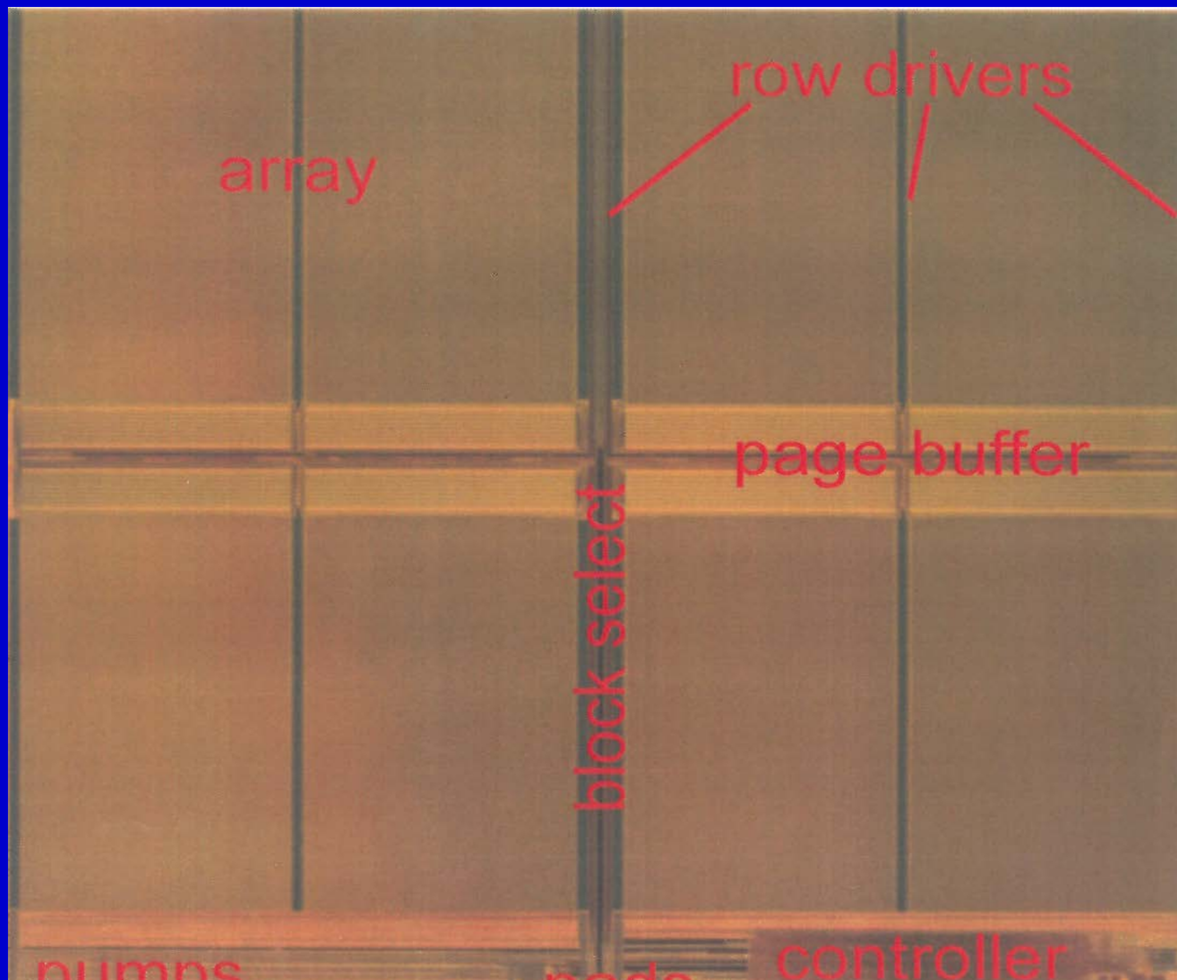
Source: Gartner, adjusted by Macronix

SINGLE-LEVEL AND MULTI-LEVEL CELLS

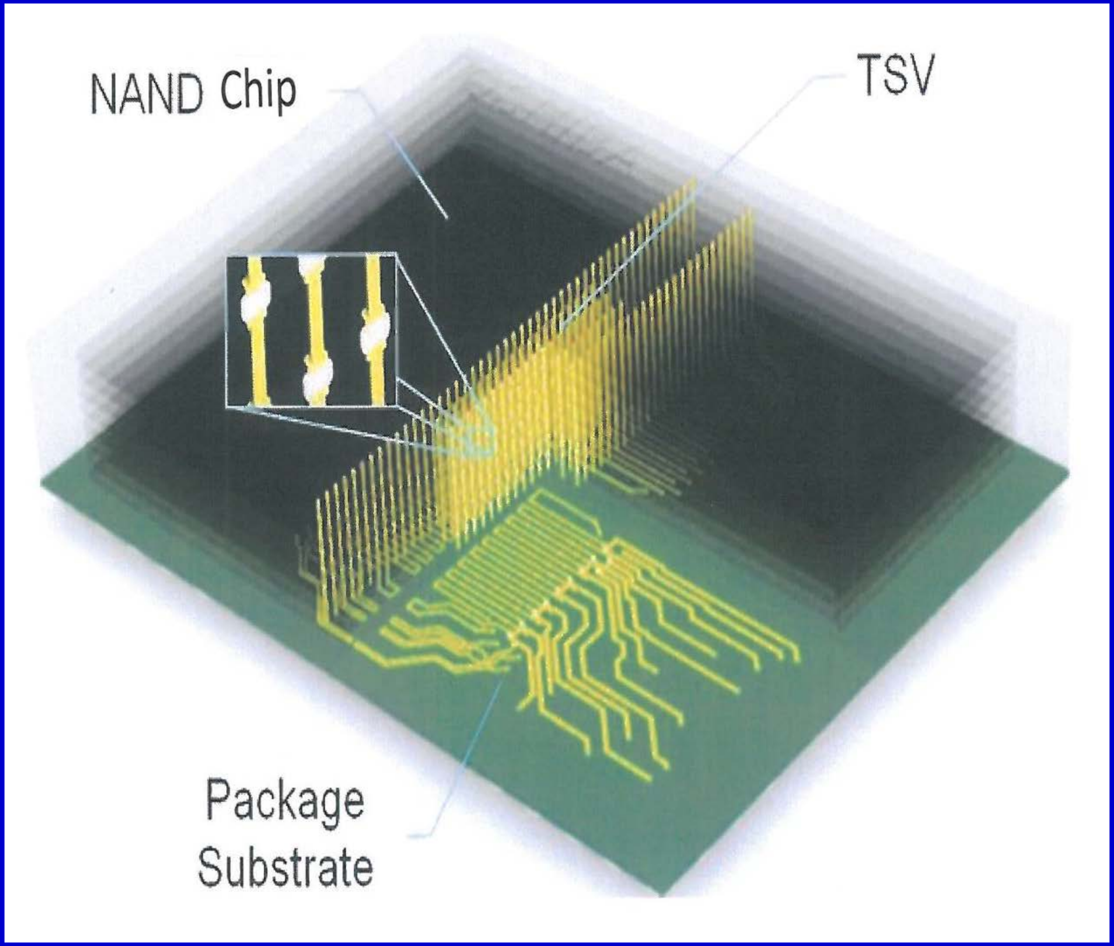


128 Gb 2b/cell 2D NAND FLASH MEMORY

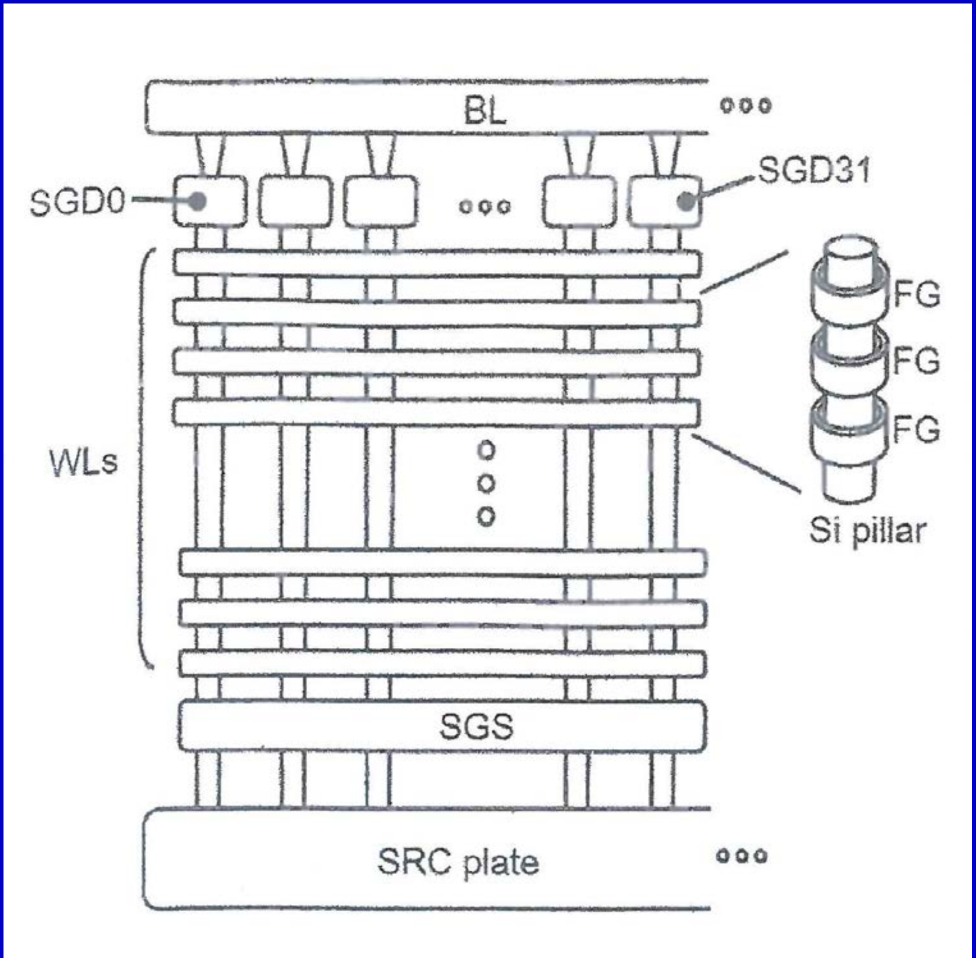
(1.8 V   16 nm   173 mm^2 0.0013μm^2/cell)



3D NAND FLASH USING MULTI-CHIP STACKING WITH THROUGH SILICON VIA (TSV)

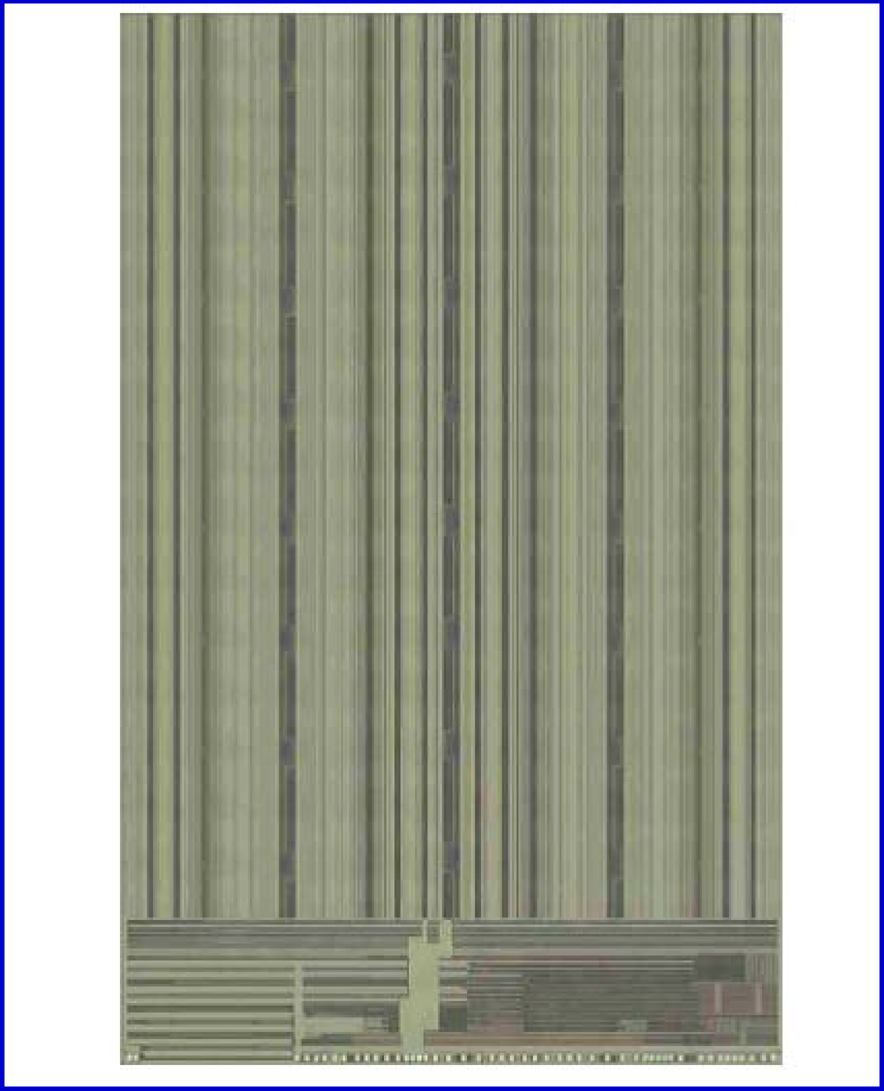


CROSS-SECTIONAL VIEW OF 3D NAND FLASH IN MULTI-LAYER INTEGRATION



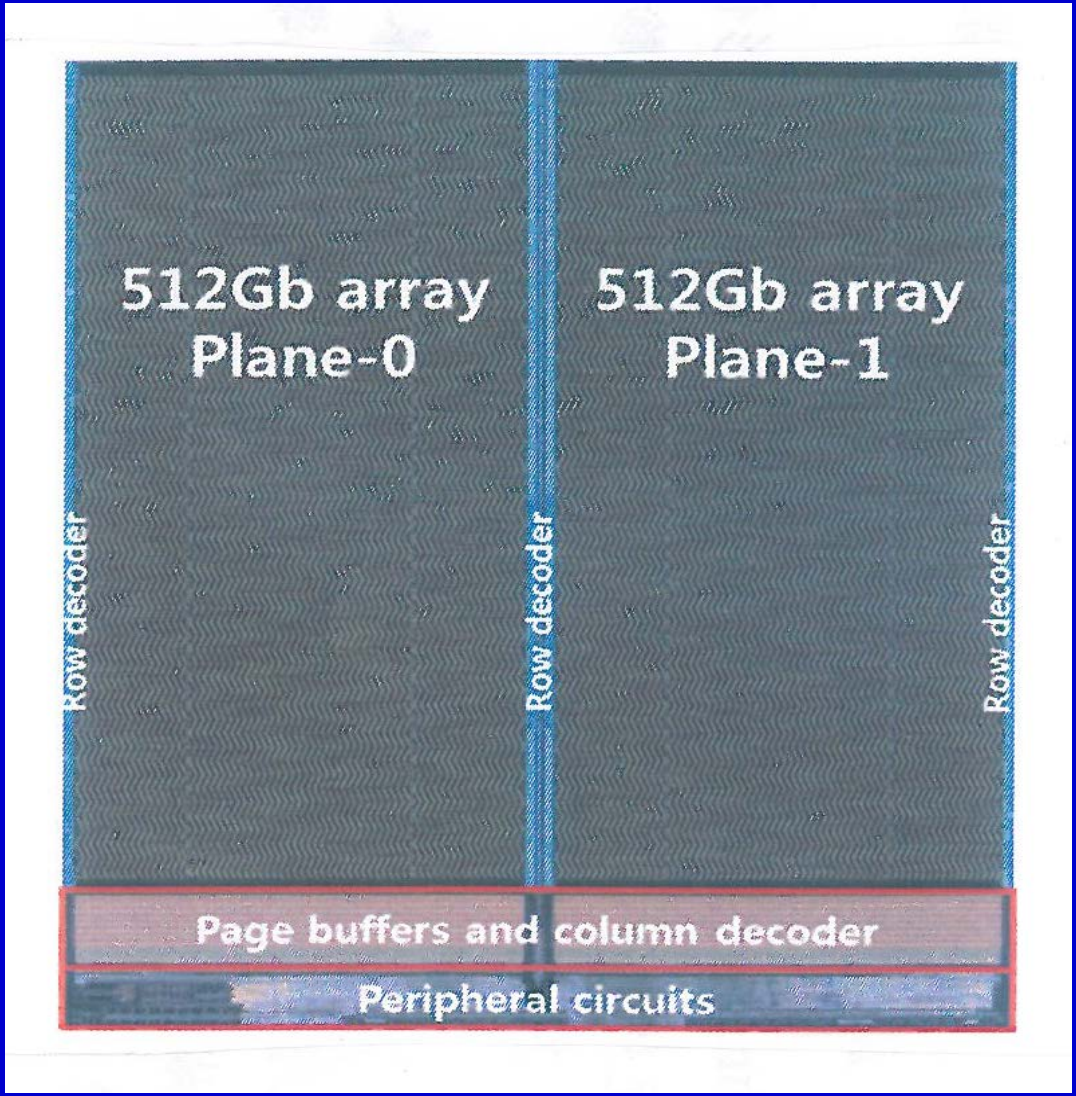
768 Gb 3b/cell 3D NAND FLASH MEMORY

(2.7V–3.6v   179 mm^2 0.0002 μm^2/cell)

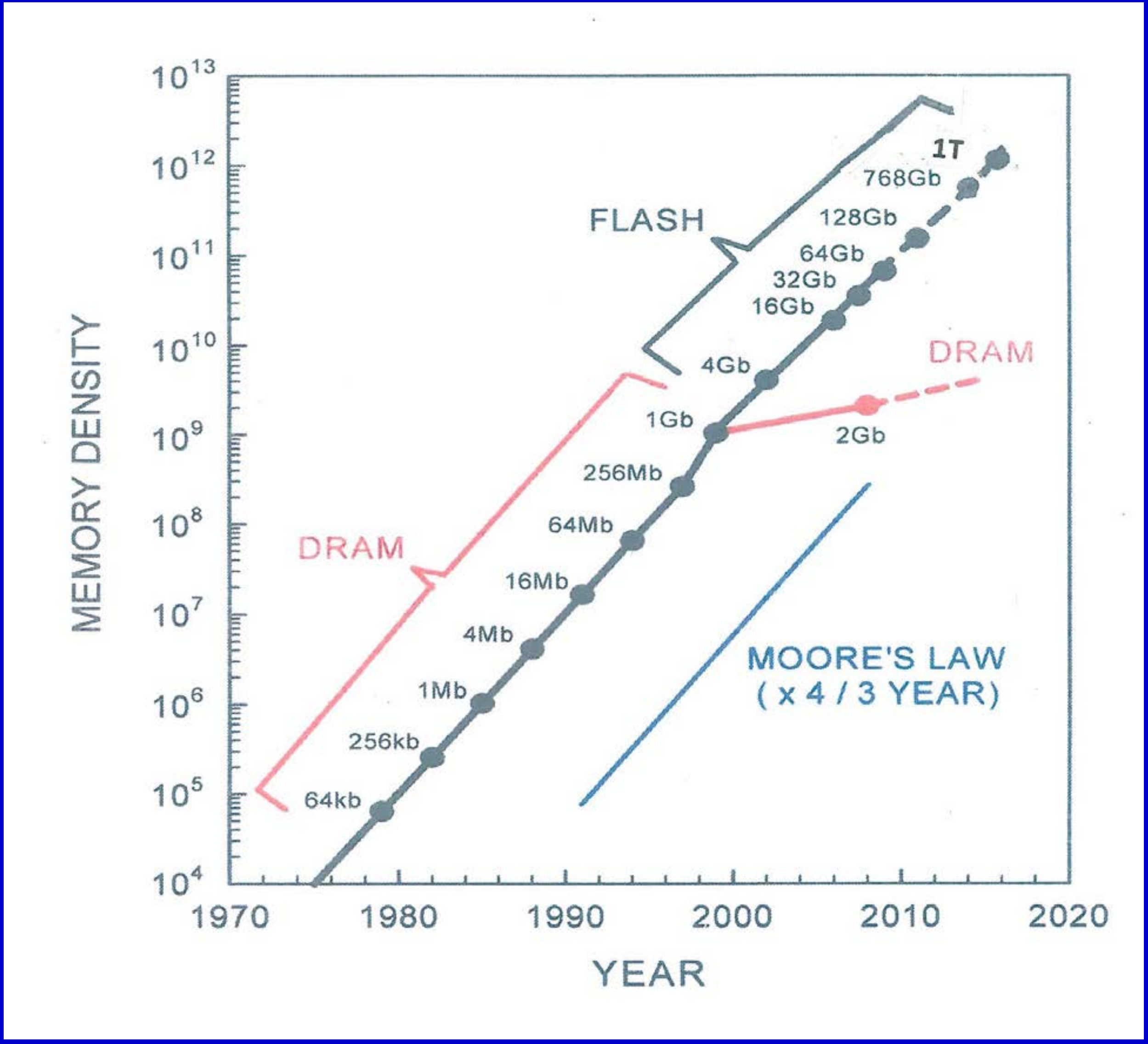


1Tb 4b/cell 3D NAND FLASH MEMORY

(1.2V 182 mm^2 0.00018 μm^2/cell)



MOORE'S LAW



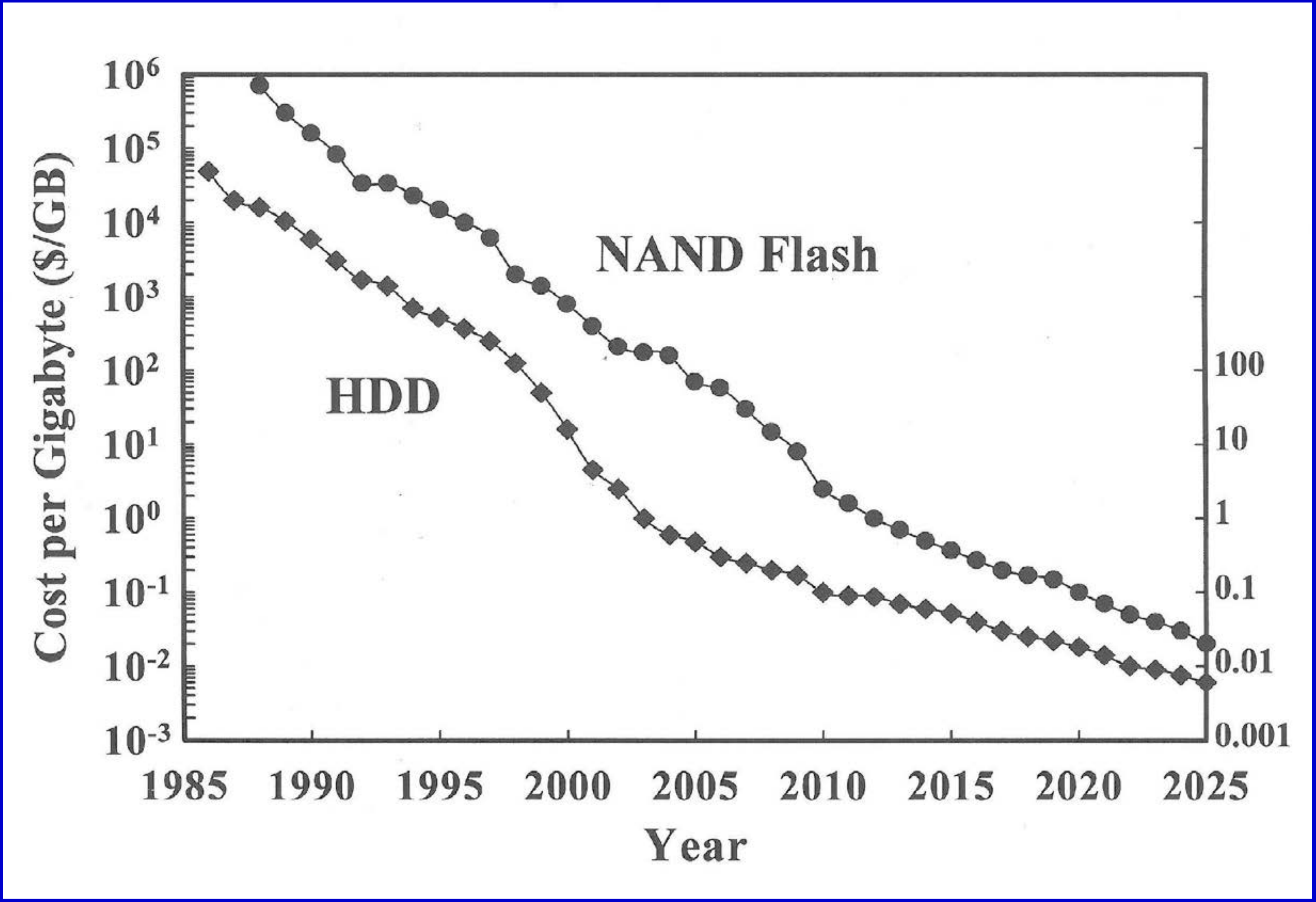
MEMORY DENSITY
YEAR
FLASH
DRAM
DRAM
MOORE'S LAW
(x 4 / 3 YEAR)
64kb
256kb
1Mb
4Mb
16Mb
64Mb
256Mb
1Gb
4Gb
16Gb
32Gb
64Gb
128Gb
768Gb
1T
2Gb

Drs. G. Moore & S. M. Sze

UNIQUE FEATURES OF FG MEMORY

- Nonvolatility — long term storage (10-100 years) is maintained after the source of power is discontinued
- High bit density — FGM (NAND Flash) has the smallest area/cell among all digital memories
- Low power consumption — FGM consumes the least power/cell among all digital memories (10% of HDD, 1% of DRAM)
- Compatibility with CMOS technology

COST PER GIGABYTE OF HDD AND NAND FLASH



- **APPLICATIONS**

APPLICATIONS OF FG MEMORY

- Communication
  - ·Over 5 billion FGM-based cellular phone subscribers in the world
  - ·Bluetooth, pagers, network systems, etc.
- Computing
  - ·Billions of FGM-based tablet computers, USB memory sticks
  - ·Artificial Intelligence
  - ·Big Data
  - ·Cloud Computing
  - ·Internet of Things
  - ·Robotics
- Consumer Electronics
  - ·Billions of digital TVs, DVD players, MP3 music players
  - ·Digital cameras, digital camcorders, electronic books
  - ·Smart IC cards
  - ·Bar-code readers

APPLICATIONS OF FG MEMORY (Continued)

- Energy Conservation

  ·FGM-based microcontroller units (NCUs) for buildings will save 50% electricity in 2030

  ·Smart sensors, smart grids, and smart systems to minimize energy consumption

  ·FGM-based brushless motors can increase energy efficiency of household appliances by 30% or more

- Health Care

  ·Wearable medical devices

  ·Implantable systems—pacemakers, defibrillators

  ·Biomedical systems-on-a-chip

  ·Micro-surgery (deVinci robot)

- Transportation

  ·50-100 FGM-based MCUs are installed in every modern car

  ·Next generation automobiles—autonomous self-driving cars

  ·FGM-based systems in every modern aircraft and marine vessel to improve performance and travel safety

JOBS NOT TO BE REPLACED BY AI

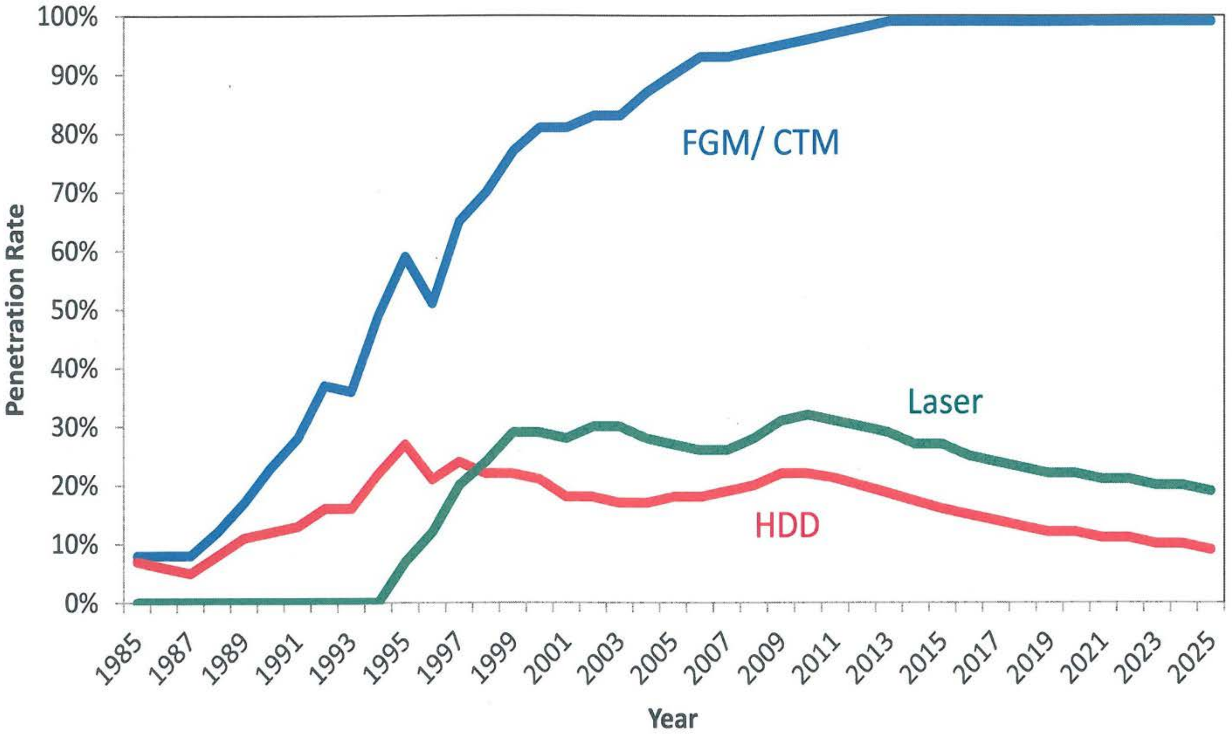
- Creative Jobs： Scientists, Novelists, Artists, etc
- Complex, Strategic Jobs： Executives, Diplomats, Economists, etc
- Empathetic and Compassionate Jobs：Teachers, Doctors, Nannies, etc

It is extremely unlikely for AI to have conceptual learning common sense, self-awareness, and emotions such as love, feelings, empathy and solidarity.

60 TB SOLID-STATE DRIVE (60 x 10^{12} Bytes)

(to store 10 million copies of the Bible, 5 million songs, or 3000 movies)

PENETRATION RATE IN ELECTRONICS INDUSTRY

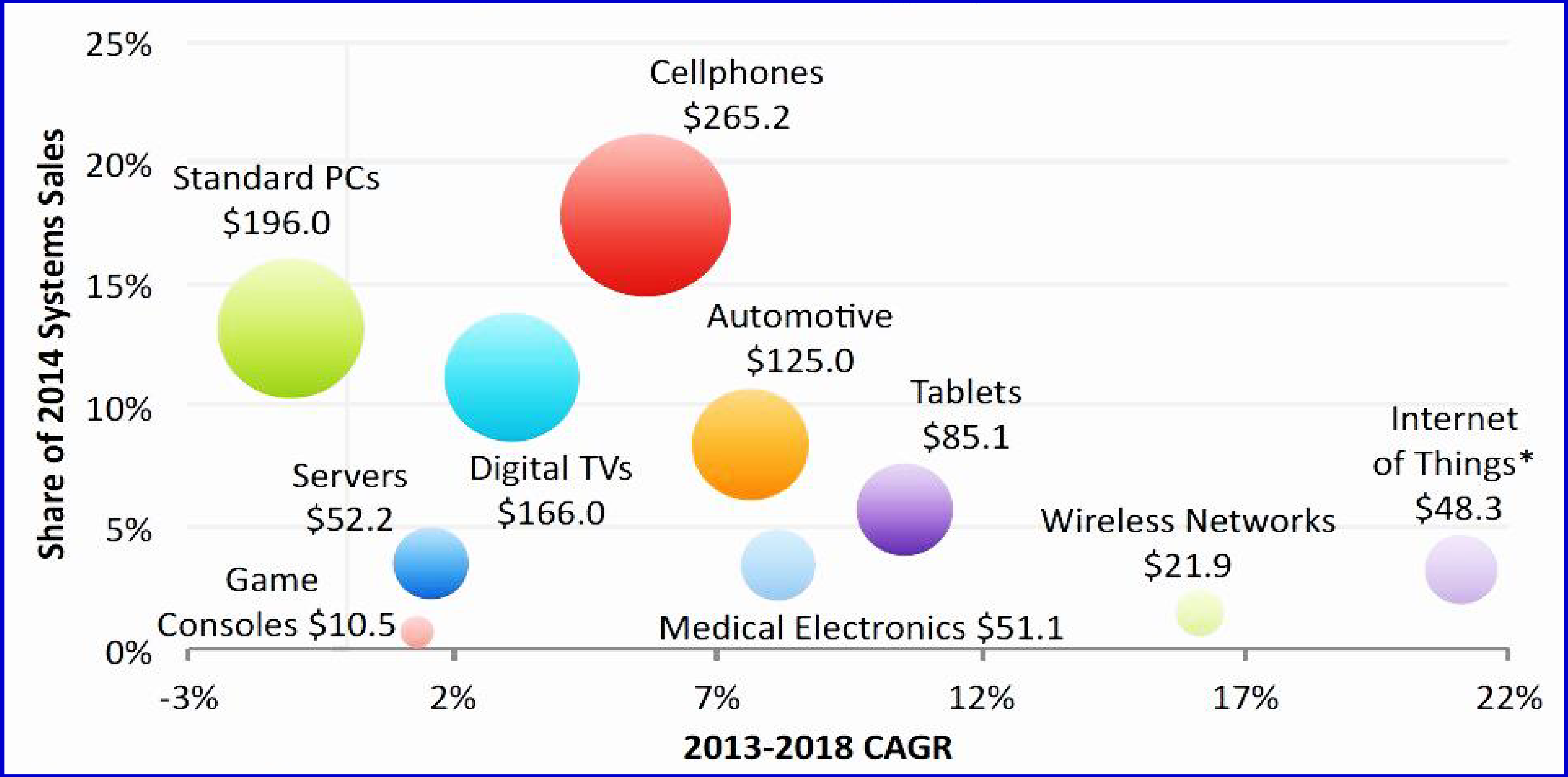


Source: The Information Network
estimated by Macronix

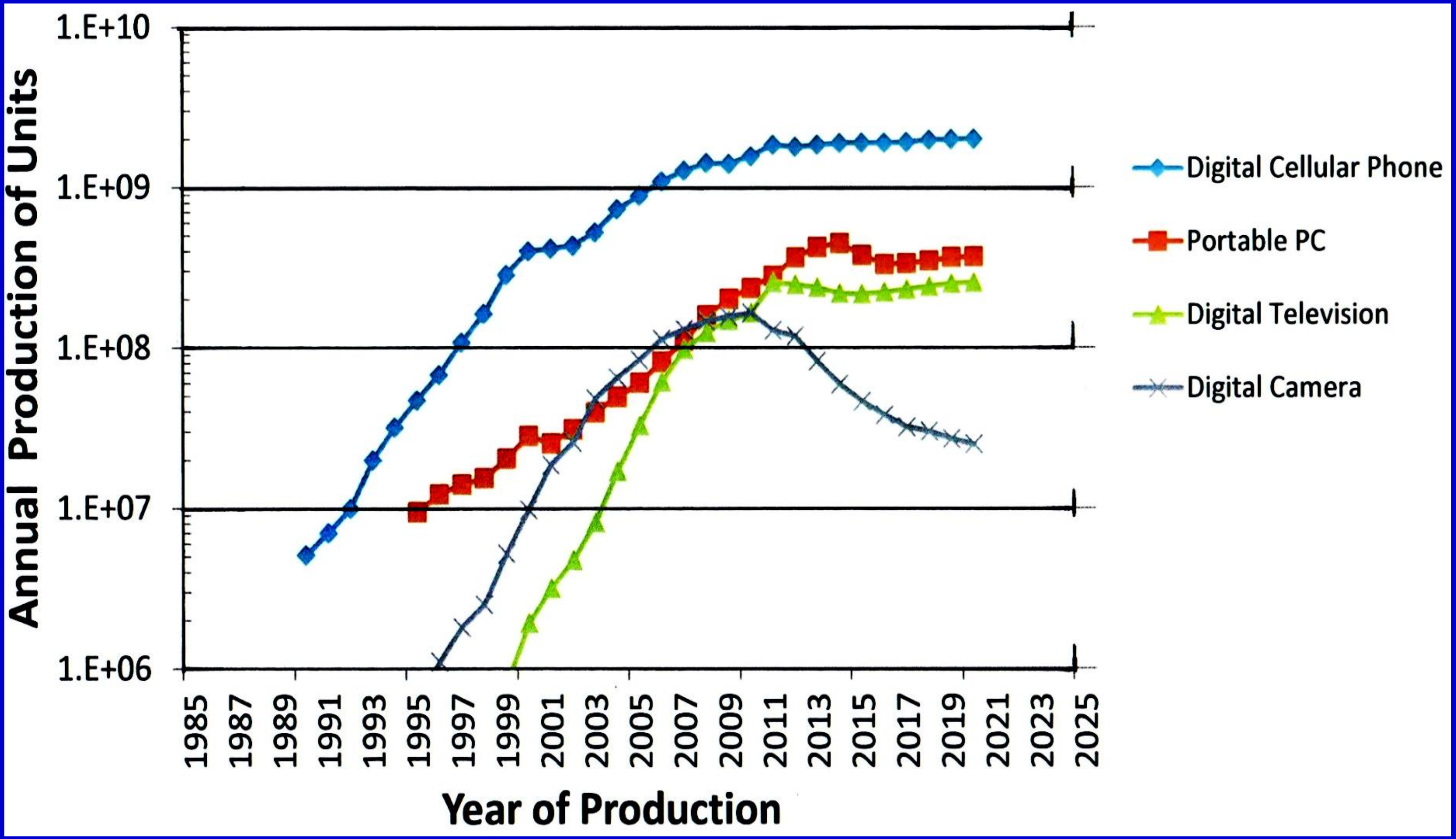
TEN FGM-BASED SYSTEMS

- 2000 USB Flash drive
- 2003 Digital TV
- 2006 Cloud computing
- 2007 Smart phone (e.g., iphone)
- 2010 3D printer
- 2011 Big data
- 2011 Ultrabook computer
- 2013 Internet of things
- 2016 Virtual reality player
- 2018 Advanced AI systems

TOP TEN FGM APPLICATIONS IN ELECTRONICS INDUSTRY
(Sales Volumes and Growth Rates)



ANNUAL PRODUCTIONS OF SELECTED DIGITAL ELECTRONIC PRODUCTS

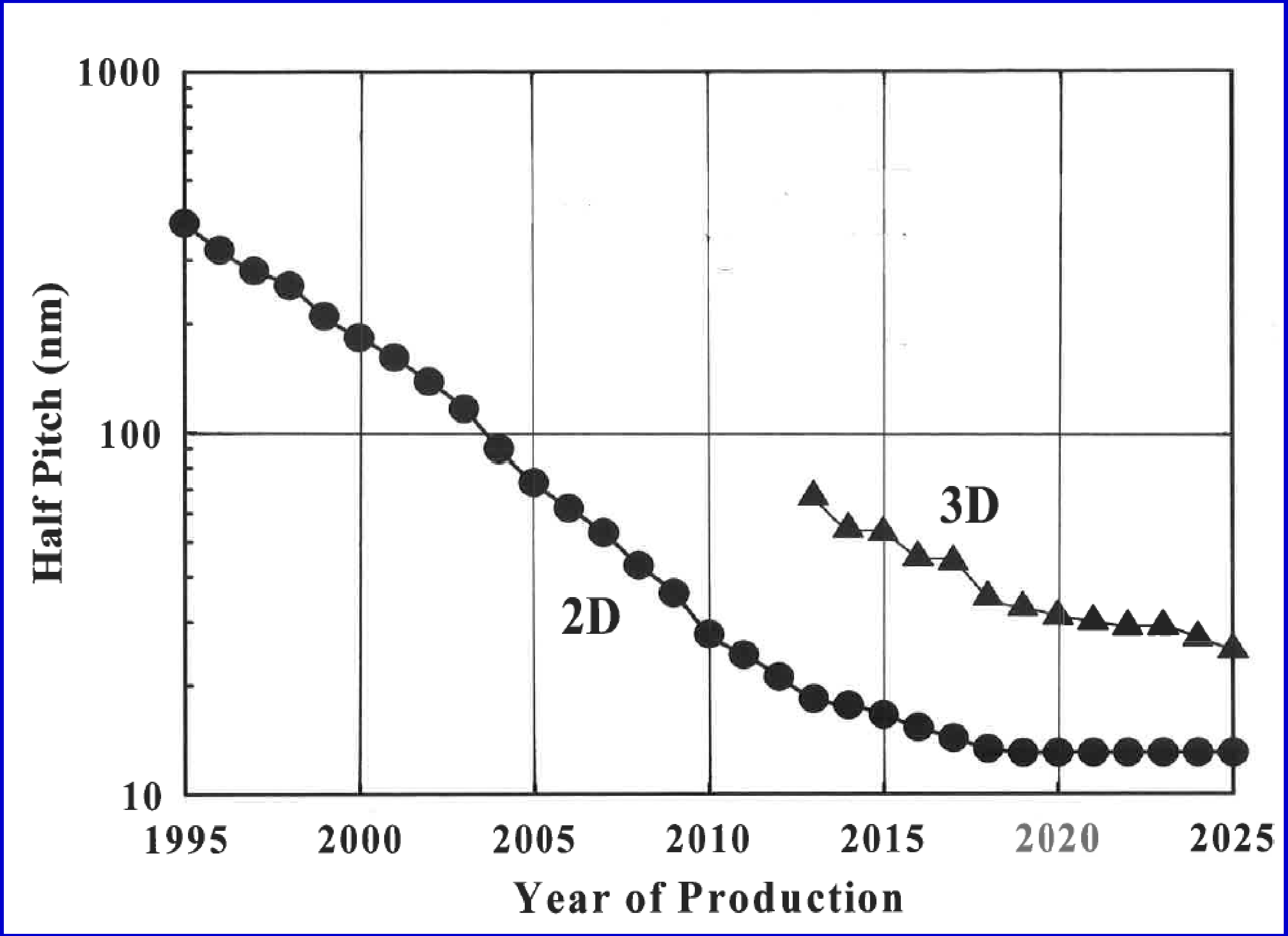


- **SCALING CHALLENGES**

SCALING CHALLENGES

- Retention of 10 ~ 20 years
- Endurance of program/erase cycles
- Interference of neighboring cells
- Reduction of coupling ratio
- Reduction of number of electrons
- Dielectric leakage
- Variability of doping and line edge roughness
- Random telegraph noise

HALF PITCH FOR 2D AND 3D NONVOLATILE MEMORIES



DOMINANT CELL TYPE (2019- 2025)

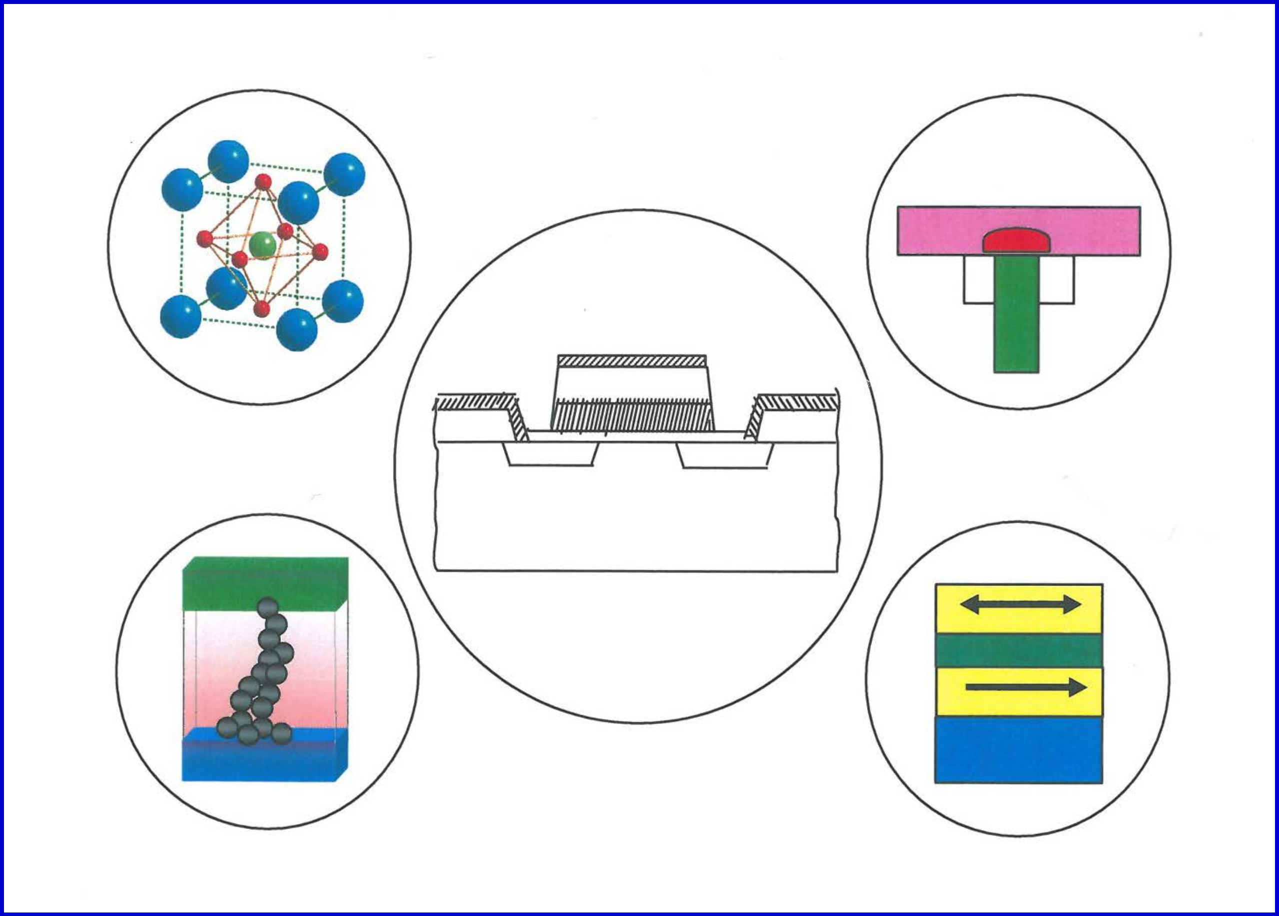
- EEPROM： Floating-gate memory
- NOR Flash : Floating-gate memory
- 2D NAND Flash : Floating-gate memory
- 3D Super-high density NAND Flash : charge-trapping memory (For ≧516 Gb, CTM has higher scalability : no larger 3D pillar due to the presence of the floating gate)

EMERGING NONVOLATILE MEMORIES

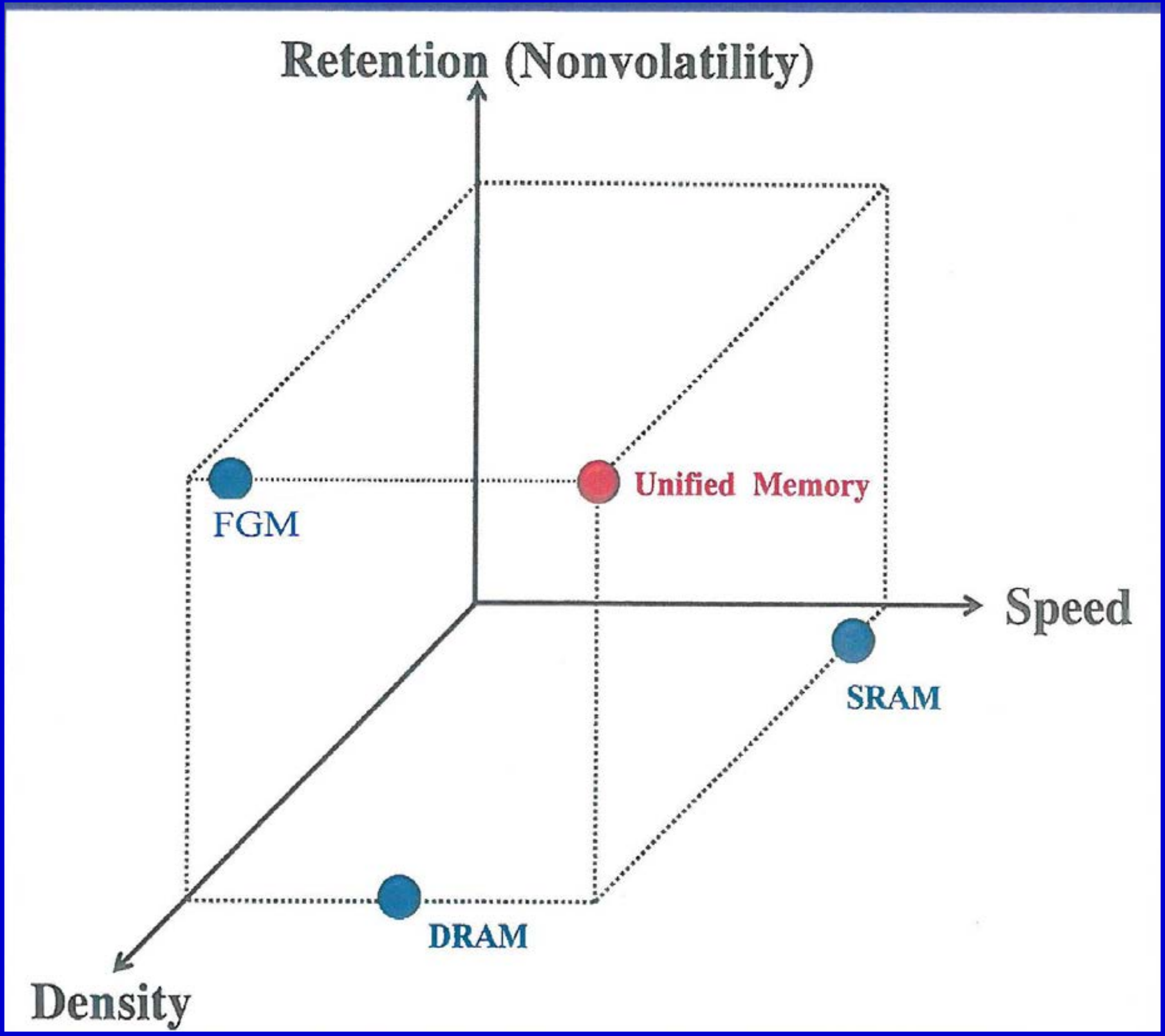
- FeRAM (Ferroelectric Random Access Memory) : based on remanent polarization in Perovskite materials
- PCRAM (Phase-Change RAM) : based on reversible phase convertion between the amorphous and crystalline state of a chalcogenide glass which is accomplished by heating and cooling of the glass
- RRAM (Resistance RAM) : based on change in resistance with applied electric field, e.g., in lead zirconium titanate and tantalum pentoxide
- STT- MRAM (Spin-torque - transfer Magnetic RAM) : based on spin polarization that can be used to control the magnetic orientation of layers in an MRAM cell

EMERGING NONVOLATILE MEMORIES

- FeRAM
- PCRAM
- RRAM
- STT-MRAM



THE UNIFIED MEMORY WITH HIGH SPEED, HIGH DENSITY, AND NONVOLATILITY



CONCLUSION

- The floating-gate memory(FGM) is a ground-breaking digital memory that supersedes virtually all prior long-term information-storage technologies such as the hard-disk drive, optical disk, and magnetic tape

- The floating-gate memory has enabled the development of nearly all modern electronic systems such as the cellular phone , tablet computer, GPS and digital TV, and brought *unprecedented benefi*t to humankind

- Many innovations are being made to overcome the scaling challenges, and a *Unified Memory* may be developed to enhance system performances

- Since year 2000, FGM has been the *prime technology driver* of the *Fourth Industrial Revolution* which is reshaping our world and leading hopefully to more empowerment ,creativity and sustainability