University of Chinese Academy of Sciences

Memory IPs and Applications

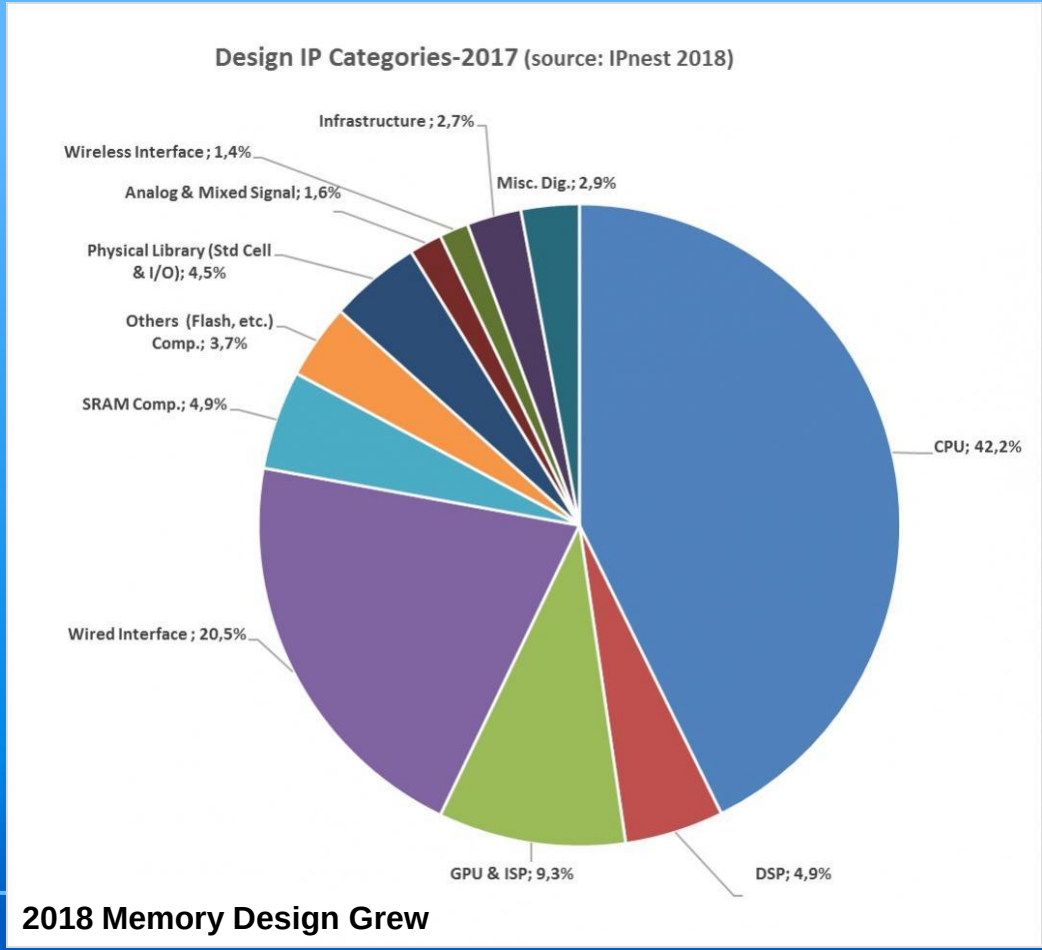
SRAM, DRAM and Flash

Chun-Zhang Chen, Ph.D.

June 28 - July 2, 2021

中国科学院大学2021年夏季

Market of Memory Design



2018 Memory Design Grew

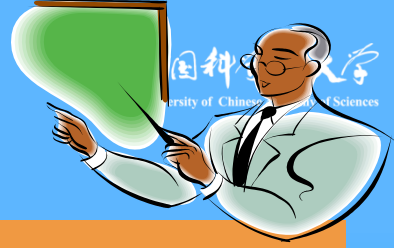
Types of Memory (IP) Design

- Memory: ROM, RAM
  - Volatile Memory (VM)
  - Non-Volatile Memory (NVM)
- ROM (OTP/MTP):
  - MTP: PROM, EPROM, EEPROM (E^2PROM, E2PROM)
- RAM:
  - **SRAM, DRAM (DDR SDRAM)**
- Flash (NVRAM): NAND and NOR
  - NAND (USB, SSD), 3D NAND, Xpoint
  - NOR

Digital Standard Cells and Memory Cells

- Combinational (or non-regenerative) logic
  - AND, OR, BUF; NAND, NOR, INV
  - Decoder/Encoder; Mux; XOR/XNOR; Add/Sub/ALU
  - PUN (pull-up network), PDN (pull-down network)
- ***Sequential*** (or regenerative) logic
  - Latches/FFs; registers (D/FF), counters, oscillators,
  - State-Machine; **Memory**
    - **Memory:** ROM**, RAM (SRAM, DRAM)**

Memory IP Designs & Applications



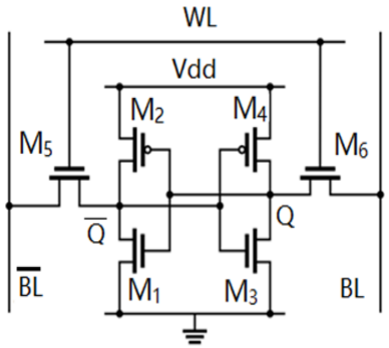
- SRAM and Its Application
- DRAM and Its Application
- Flash Memory and Its Application
- New Memory and HBM
- Discussion

CPU Memories: SRAM and Cache

- **Level 1 caching**
  - **L1 cache**, **primary cache**, **internal cache**, or **system cache**
  - The L1 cache stores the most critical files that need to be executed and is the first thing the processor looks when performing an instruction.
- **Level 2 cache**, **L2 cache**, **secondary cache**, or **external cache**
- **L3 cache** is cache memory located on the die of the CPU. (See Intel Core i7-3960X processor die)

SRAM Cell Sizes

SRAM 6T-Cell



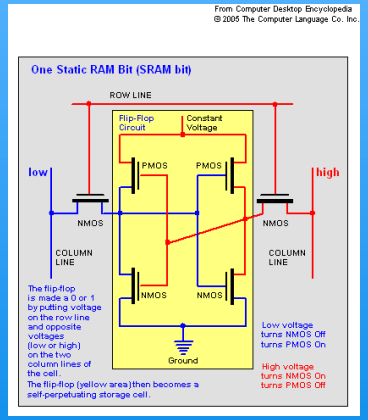
1T-SRAM Cell sizes (μm²/bit or mm²/Mbit)

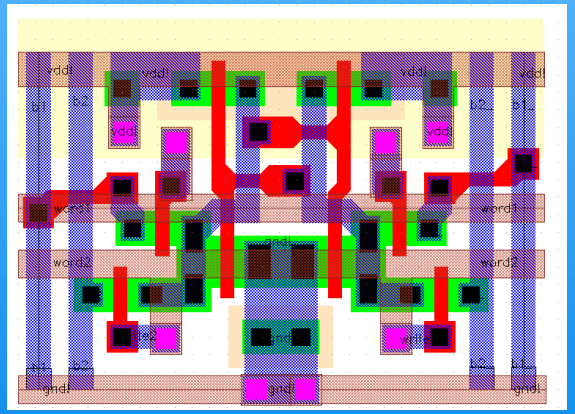
Process node		250 nm	180 nm	130 nm	90 nm	65 nm	45 nm
6T-SRAM	bit cell	7.56	4.65	2.43	1.36	0.71	0.34
	with overhead	11.28	7.18	3.73	2.09	1.09	0.52
1T-SRAM	bit cell	3.51	1.97	1.10	0.61	0.32	0.15
	with overhead	7.0	3.6	1.9	1.1	0.57	0.28
1T-SRAM-Q	bit cell			0.50	0.28	0.15	0.07
	with overhead			1.05	0.55	0.29	0.14

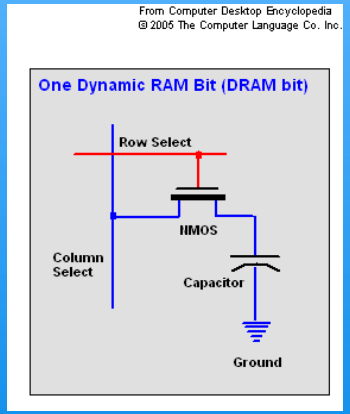
SRAM is faster and more expensive than DRAM; it is typically used for CPU cache while DRAM is used for a computer's main memory.

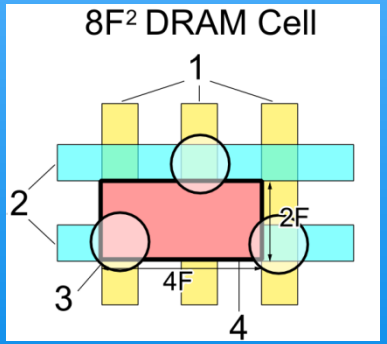
Layouts of SRAM and DRAM

- SRAM – 6T
- DRAM – 1T1C









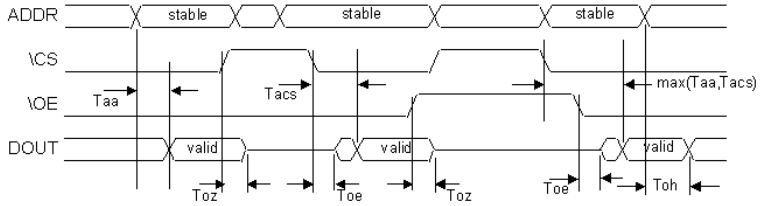
Structure and Timing of SRAM

- SRAM (Static Random-Access Memory) Structure
  - Inputs/Outputs
  - SRAM Structure (RAM Cell):
    - add (decoder), bit-line, word-line
- **SRAM Timing**
  - For read: t_AA, t_ACS, t_OE, t_OZ, t_OH
    - (Access from add/chip_sel/out_en/out_dis/out_hold)
  - For write: t_AS, t_AH, t_CSW, t_WP, t_DS, t_DH
    - (Access_setup_before/acc_hold_after/chip_sel_before/write_pulse/ data_setup_before/data_hold_after)
- Types of SRAM
  - Sync SRAM, Pipeline Burst SRAM

SRAM Timing

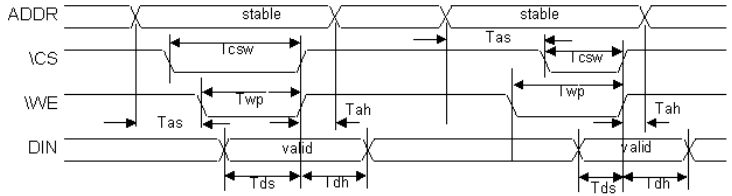
- SRAM *read* timing
- SRAM write timing

SRAM read timing



Taa: access time from address
Tacs: access time from chip select
Toz: output disable time
Toe: output enable time
Toh: output hold time

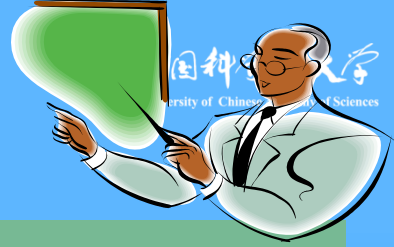
SRAM write timing



Tas: address setup time before write
Tah: address hold time after write
Tcsw: chip select time before write
Twp: write pulse minimum width
Tds: data setup time
Tdh: data hold time

Address must be stable before WE', else invalid data may glitch other cells

Memory IP Designs & Applications



- SRAM and Its Application
- DRAM and Its Application
- Flash Memory and Its Application
- New Memory and HBM
- Discussion

DRAM, SDRAM (synchronous) & ADRAM

Difference by the read style
- Sync. DRAM (SDRAM), Async. DRAM

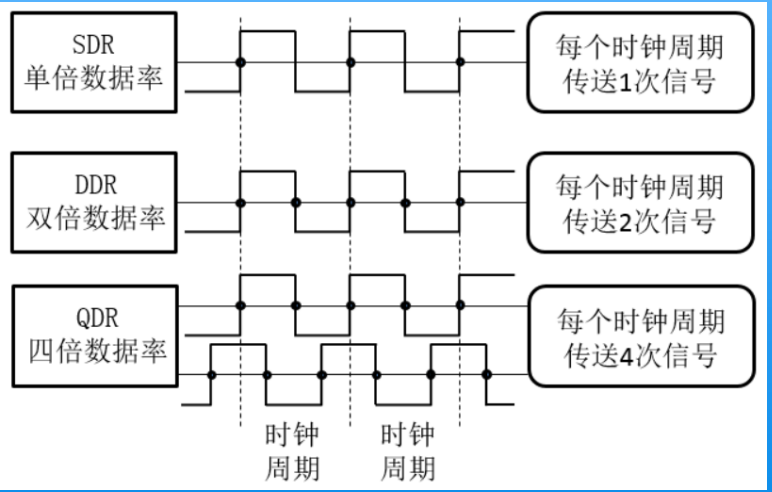
Difference by the datarate, SDRAM has
- Single Data Rate (SDR)
- Double Data Rate (DDR)
- Quad Data Rate (QDR)

From applications Graphics (GDDR)
From power consumption,
- DDR, LPDDR (mDDR)

When DRAM is integrated with other modules (ex. CPU), it is called
- embedded DRAM (eDRAM)

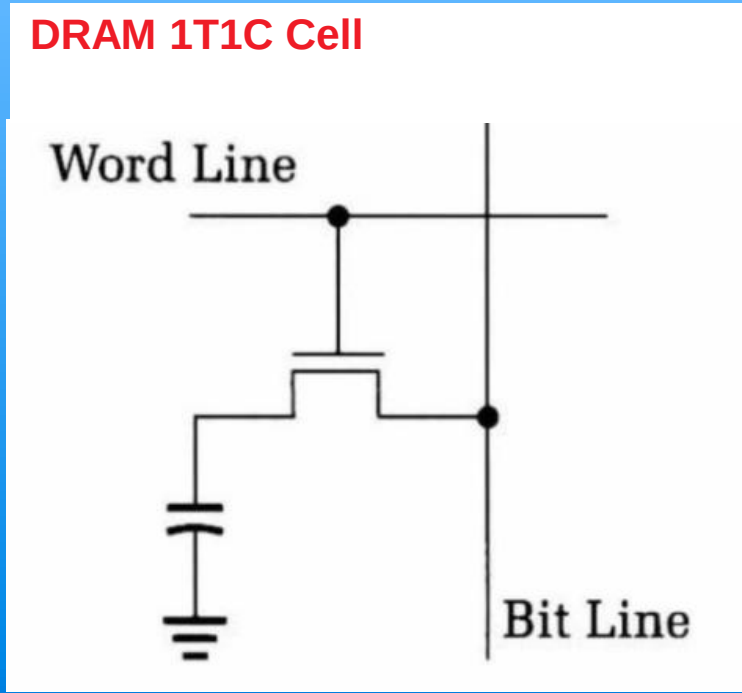
DDRx

- 1998 Gen1 SDR (single datarate SDRAM)
- Gen2 DDR (SDRAM DDR)
- 2004 Gen3 DDR2
- 2007 Gen4 DDR3
- 2014 Gen5 DDR4 3.2Gb/s
- 2020 DDR5 6.4Gb/s

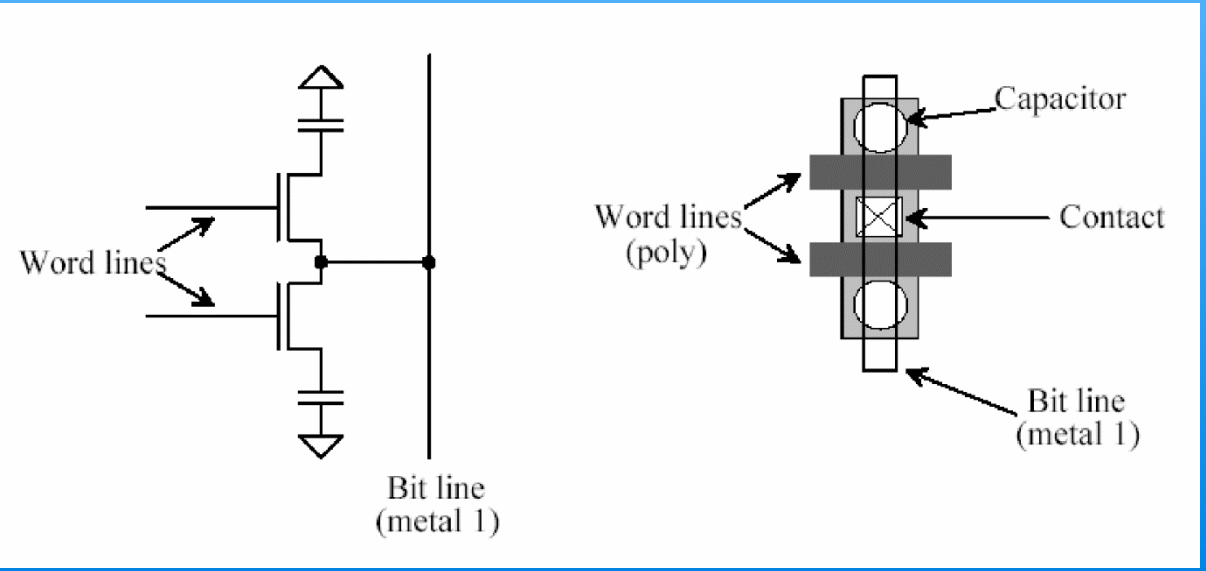


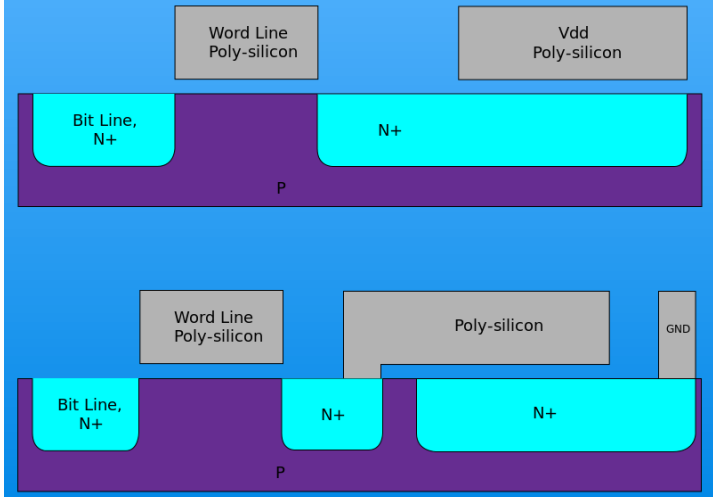
DRAM 1T1C Cell

- Word Line, WL
  - Address, Control, Select
- Bit Line, BL
  - Data, Information, SA
- Sense Amplifier, SA



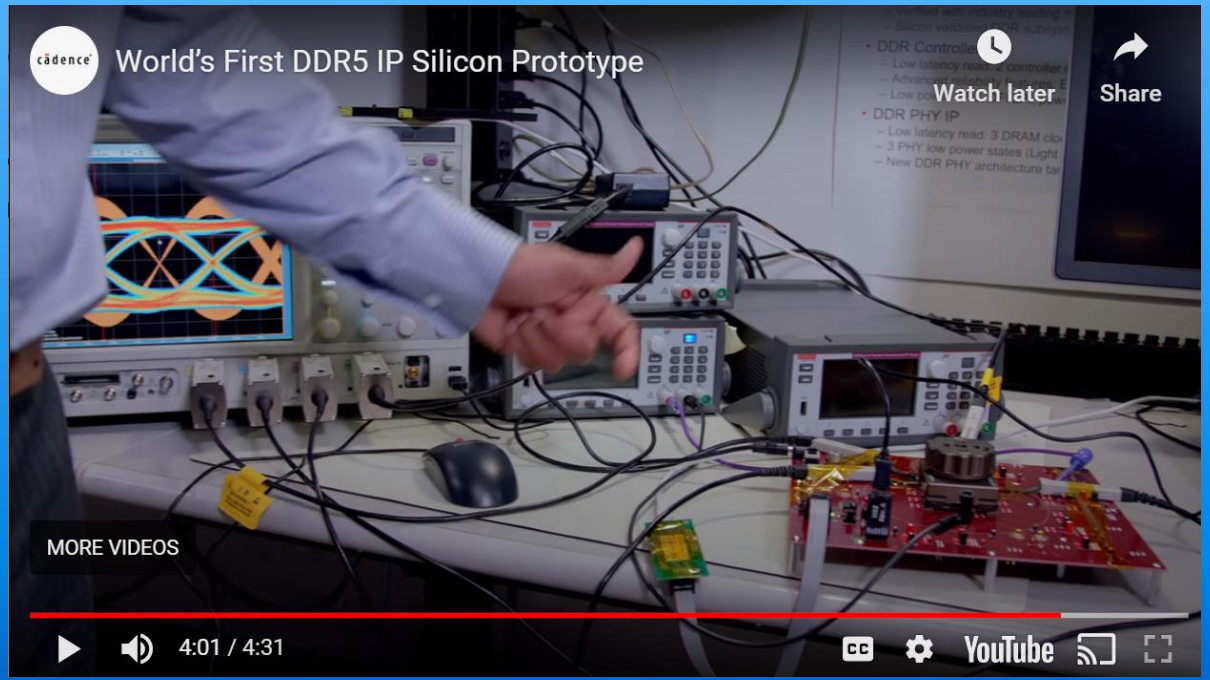
DRAM Layout


Word lines
Bit line
(metal 1)
Capacitor
Word lines
(poly)
Contact
Bit line
(metal 1)


Word Line
Poly-silicon
Vdd
Poly-silicon
Bit Line,
N+
N+
P
Word Line
Poly-silicon
Poly-silicon
GND
Bit Line,
N+
N+
N+
P

World's First DDR5

- DDR IP



Design IP
- Analog IP
- Interface IP
- Denali Memory Interface IP
  - DDR IP
    - DDR/LPDDR PHY IP
    - DDR/LPDDR Controller IP
    - GDDR PHY and Controller IP
    - LPDDR5 PHY IP Solution
    - HBM PHY and Controller IP
  - NAND Flash IP
  - SD/SDIO/eMMC IP
  - Octal and Quad SPI IP
- Systems/Peripherals IP

Tensilica IP

Verification IP

mDDR (LPDDR)

- Low Power DDR， LPDDR
  - Synchronous DRAM, LPDDR SDRAM, used in mobile, mDDR or Mobile DDR
- LPDDR/2/3/4/5

	LPDDR2	LPDDR3	LPDDR4
内部核心时钟 (MHz)	200	200	200
总线时钟 (MHz)	400	800	1600
数据速率 (Mbps)	800	1600	3200
带宽 (GB/s)	6.4	12.8	25.6
预取(prefetch)	4n	8n	16n
电压 VDD2/VDDQ/VDD1	1.2V/1.2V/1.8V	1.2V/1.2V/1.8V	1.1V/1.1V/1.8V
指令/地址总线	10 bits, DDR	10 bits, DDR	6 bits, SDR
Bank 数	4/8	8	8/ch (16)
容量	64Mb~8Gb	4Gb~32Gb	8Gb~32Gb
接口	HSUL_12	HSUL_12(可选ODT)	LVSTL
I/O 类型	x16/x32	x16/x32	2ch x16
封装	MCP/PoP	MCP/PoP	MCP/PoP

注：预取(prefetch)代表缓存大小(buffer size)。例如上表中 LPDDR3，其 prefetch 为 8n，代表每个预取的“数据字数(datawords)”为 8。

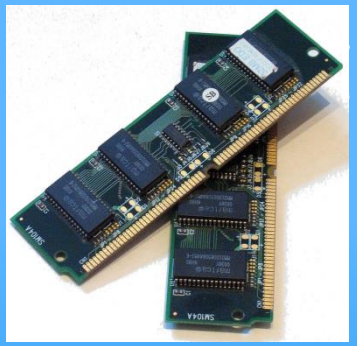
GDDR

- Graphics Double Data Rate, GDDR

性能	GDDR3	GDDR5
电压 VDD, VDDQ	1.5V, 1.35V	1.5V, 1.35V
时钟(MHz)	800/900	1750
数据倍率	2	4
数据传输率(Gbps)	1.6/1.8	6, 7, 8
I/O 宽度	(4), 8, 16	32/16
存储块(Bank)数	8	16
预取	8n	8n
突发长度	4(Burst Chop), 8	8
典型颗粒容量(Gb)	1~2	2~8
循环冗余检验(Cyclic Redundancy Check, CRC)	~~N/A~~ 不适用	~~Yes~~ 适用
封装	BGA-78/96	BGA170

- Types of DRAM
  - EDO DRAM, ext'd data out
  - BEDO DRAM, burst EDO
  - SDRAM, synchronous
  - RDRAM, Rambus
  - DDR SDRAM, double data rate

Prices of six generations of DRAMs

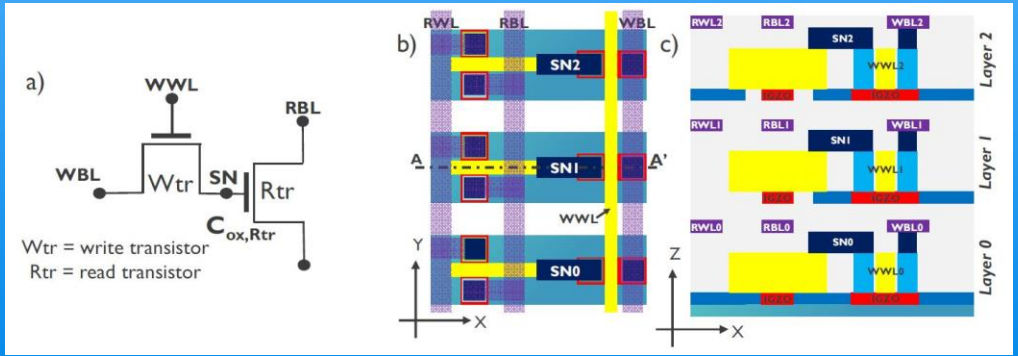


- DRAM 16kbit → 64Mbit (1978-2001)
  - $1/1977 ~ $2.95/2001
- DRAM/ DDR
  - DRAM 1Mbit: >$5000/1977 → $0.35/2000 → $0.08/2001 (≈1977)
  - DDR4 8Gb: $4-$4.5 (May 2019)
- DRAM usage
  - PC DRAM, Server DRAM, Mobile DRAM, Consumer DRAM, DRAM ASP (Ref: DRAMeXchange)

Amazingly

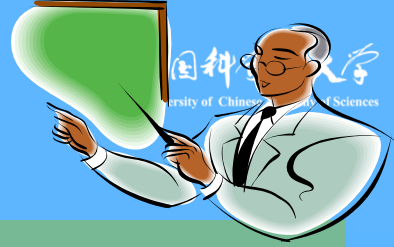
DRAM replacement?

- DFM (Dynamic Flash Memory; SGT)
  - 2T0C (vs 1T1C), Fujio Masuoka, Unisantis Electronics
- IMEC IGZO-TFTs



- VLT
- Z-RAM

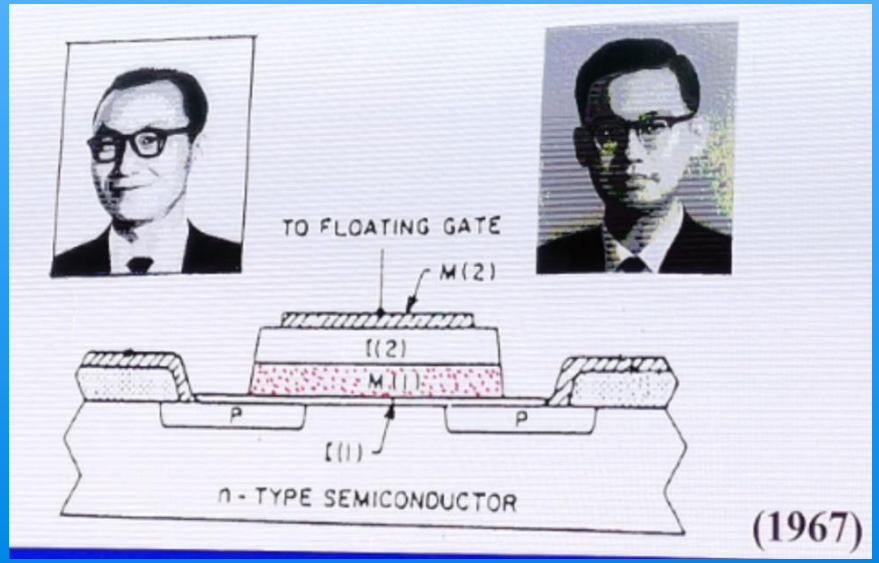
Memory IP Designs & Applications



- SRAM and Its Application
- DRAM and Its Application
- **Flash Memory and Its Application**
- New Memory and HBM
- Discussion

Discovery of The Floating-Gate Memory Effect

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



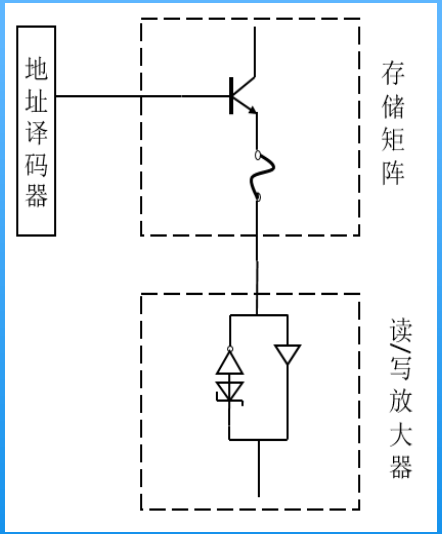
YEAR	DEVICE	DISCOVERER (S)/INVENTOR(S)	ORGANIZATION
1967	Floating-Gate Concept	Kahng and Sze	Bell Labs
1971	EPROM-FAMOS	Frohman-Bentchkowsky	Intel
1976	EEPROM-SAMOS	Iizuka et al.	Toshiba
1984	Flash Memory	Masuoka et al.	Toshiba
1987	NAND Flash	Masuoka et al.	Toshiba
1995	Multilevel Cell	Bauer et al.	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

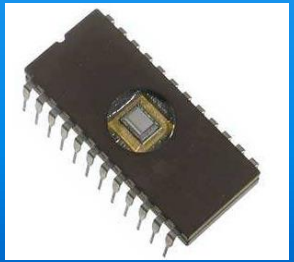
Classification of FGM

- 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)

OTP - PROM

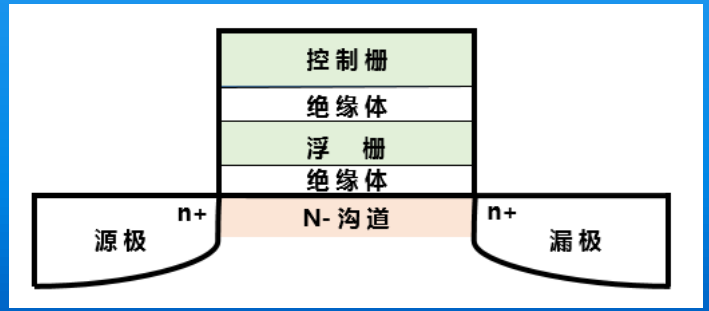
- **One-Time Programmable (OTP)**
- Typical OTP: PROM
  - Wen Tsing Chow in 1956
- New OTP
  - Resistive RAM (RRAM | ReRAM)





MTP – EPROM & EEPROM

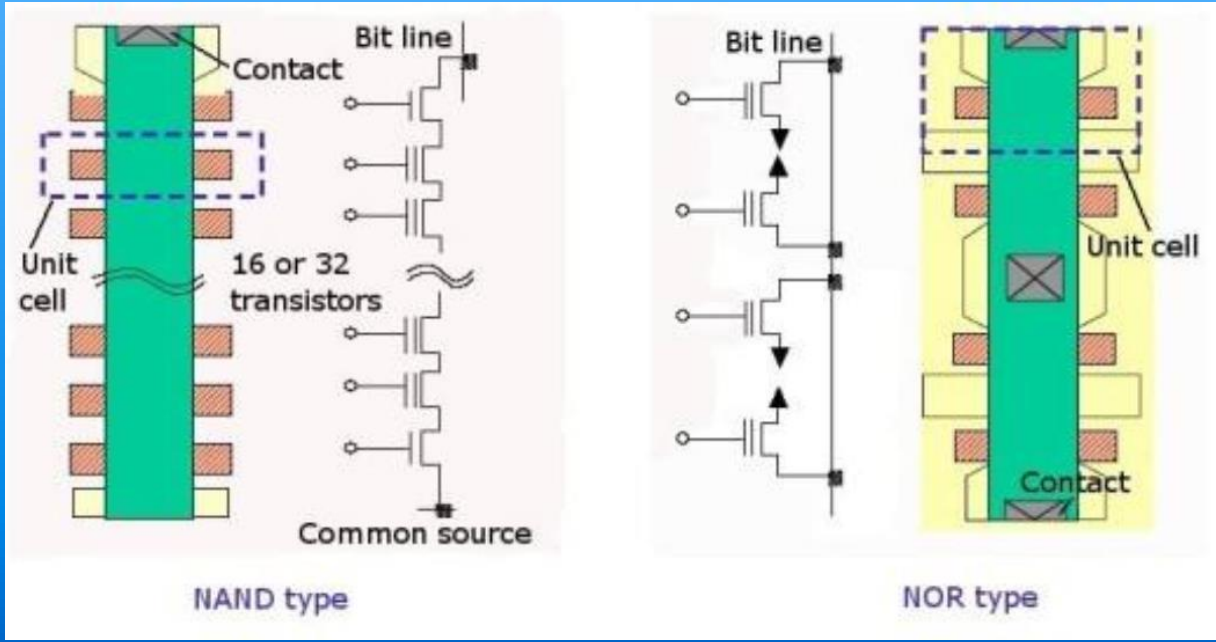
- **Multi-Time Programmable (MTP) Memory**
- Erasable Programmable Read-Only Memory，EPROM
- Electrically Erasable Programmable Read-only memory，EEPROM
  - EEPROM or E^2PROM, are Floating-Gate Transistors
  - Floating-Gate tech by Simon M. Sze & Dawon Khang, 1967
  - FRAM (or FeRAM) , MRAM are new



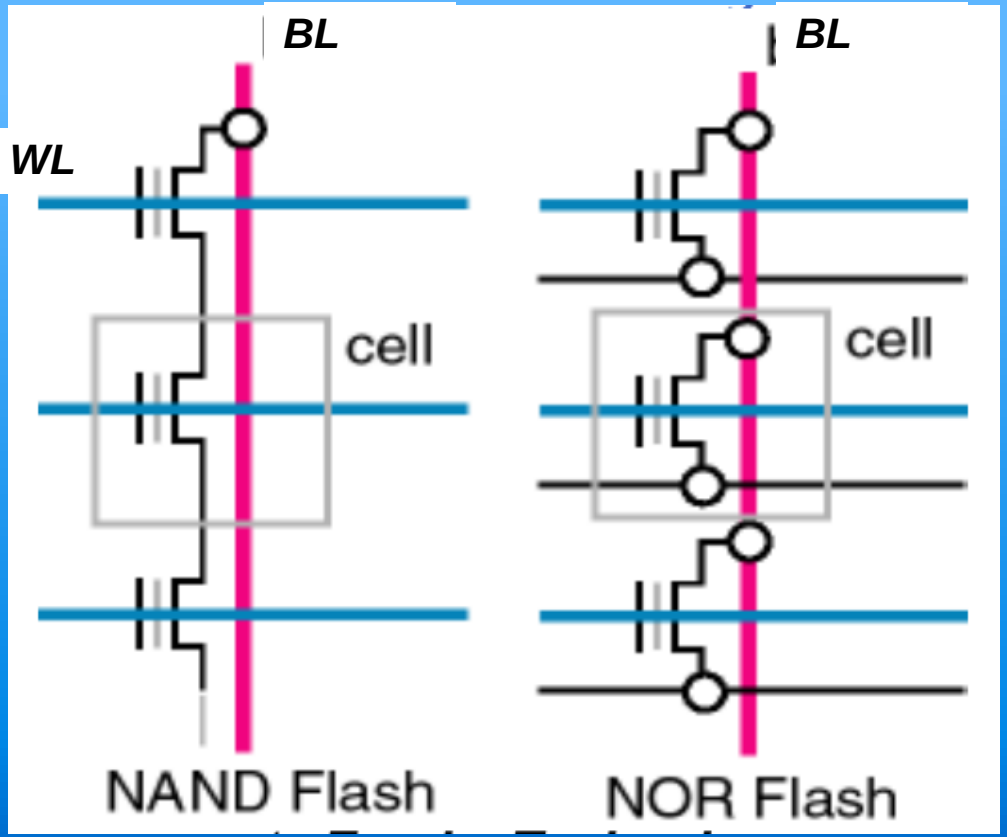
Invention of NOR Flash and NAND Flash

NOR and NAND by & 1987

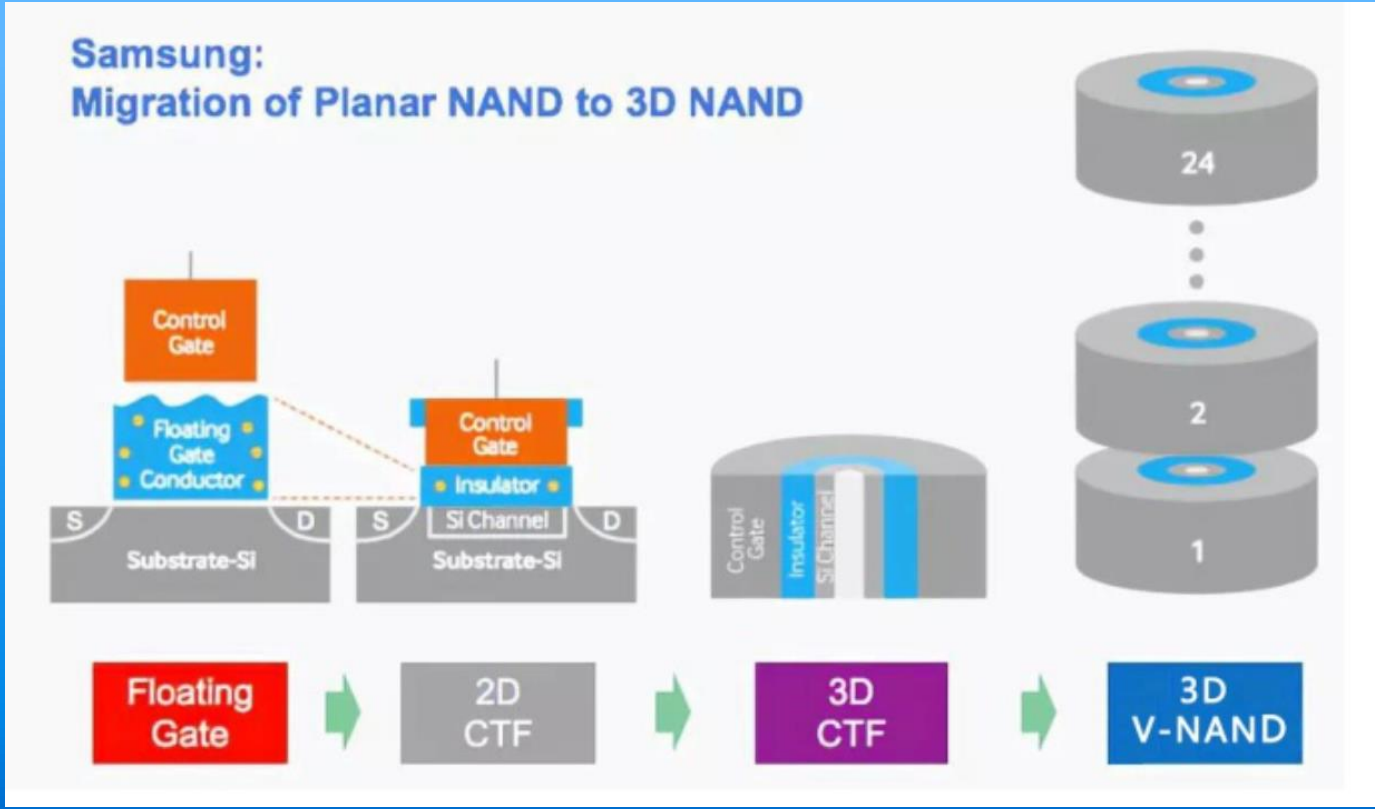
- **NAND** Flash
  - Toshiba, 1989; ONFI
- **NOR Flash**
  - Intel, 1988; XIP Tech



Schematic of Flash Memory



2D NAND and 3D NAND



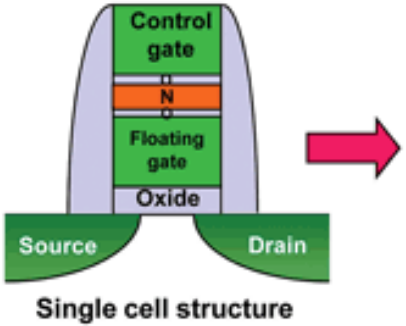
NAND Flash

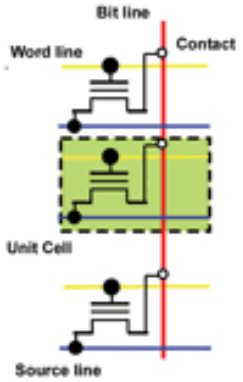
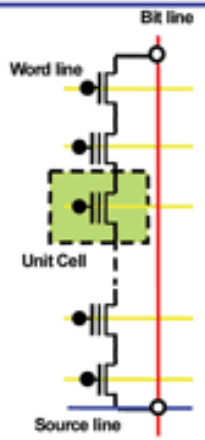
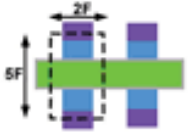
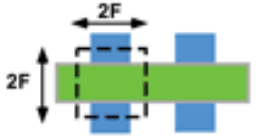
- SRAM
- DRAM
- Flash Memory
  - NOR
  - NAND, by Dr. Fujio Masuoka
    - App: SSD, consumer, medical, automotive

NOR Flash

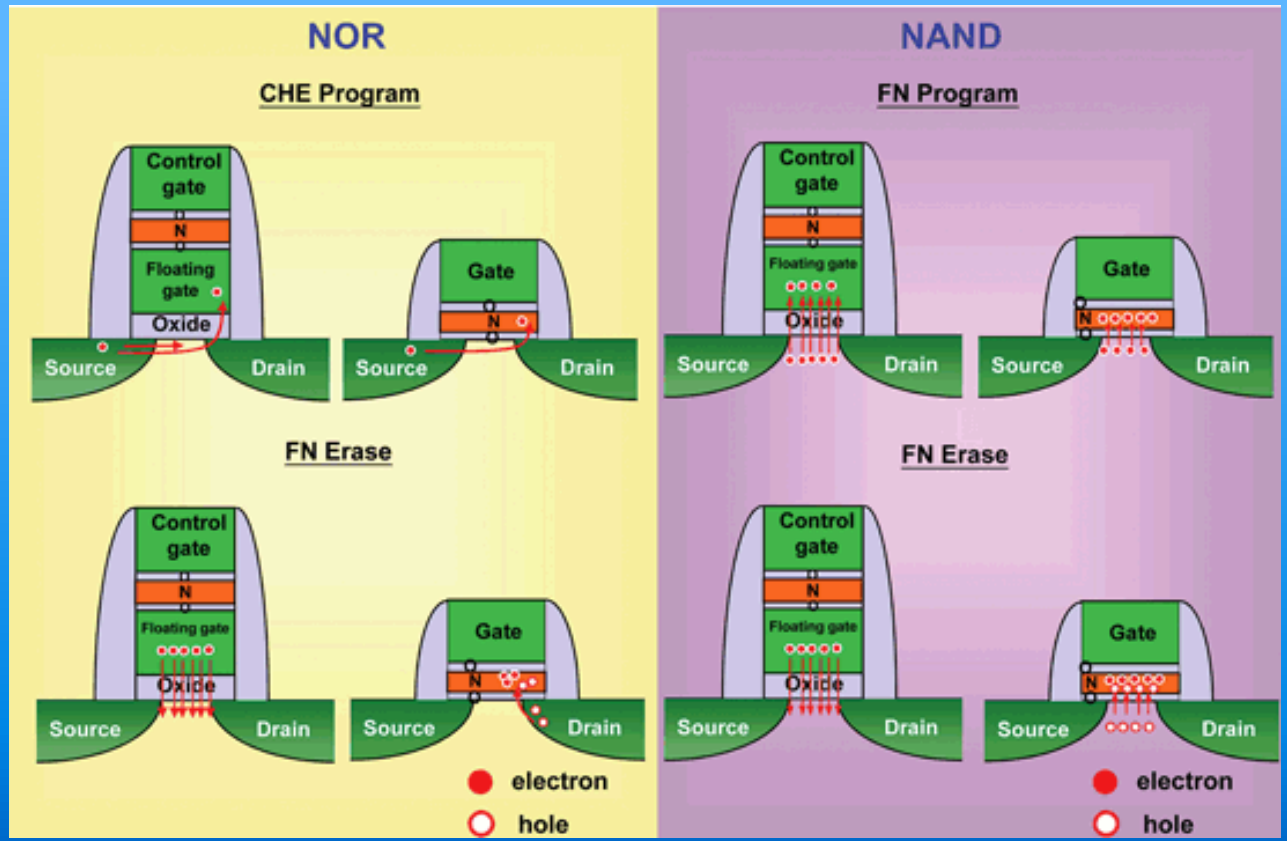
- SRAM
- DRAM
- Flash Memory
  - NOR, by Dr. Fujio Masuoka at Toshiba 1985 (US 4,531,203)
    - App: AMOLED, TWS BT earpods, TDDI , Automotive, 5G
  - NAND

Cross Section of Flash Memory

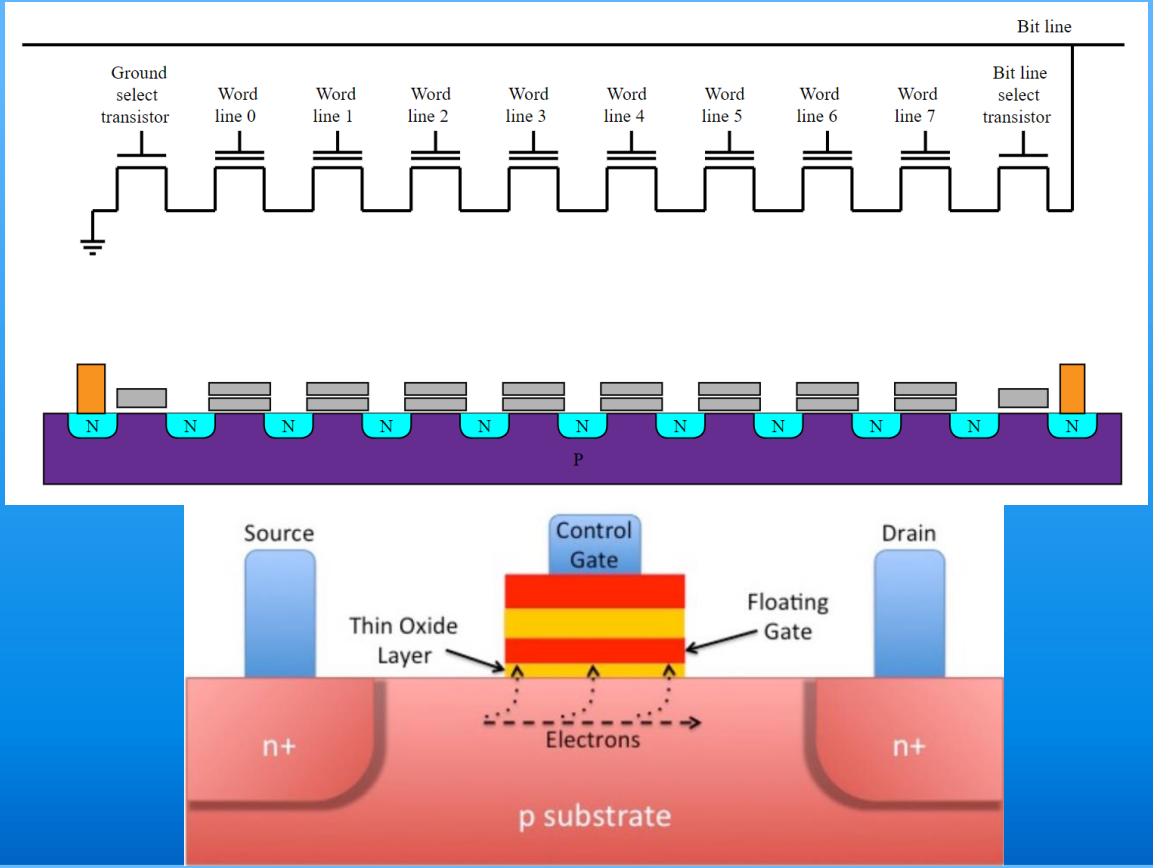


	NOR	NAND
Cell Array		
Layout		
Cross section	 $10F^2$	 $4F^2$

Comparison of NOR and NAND

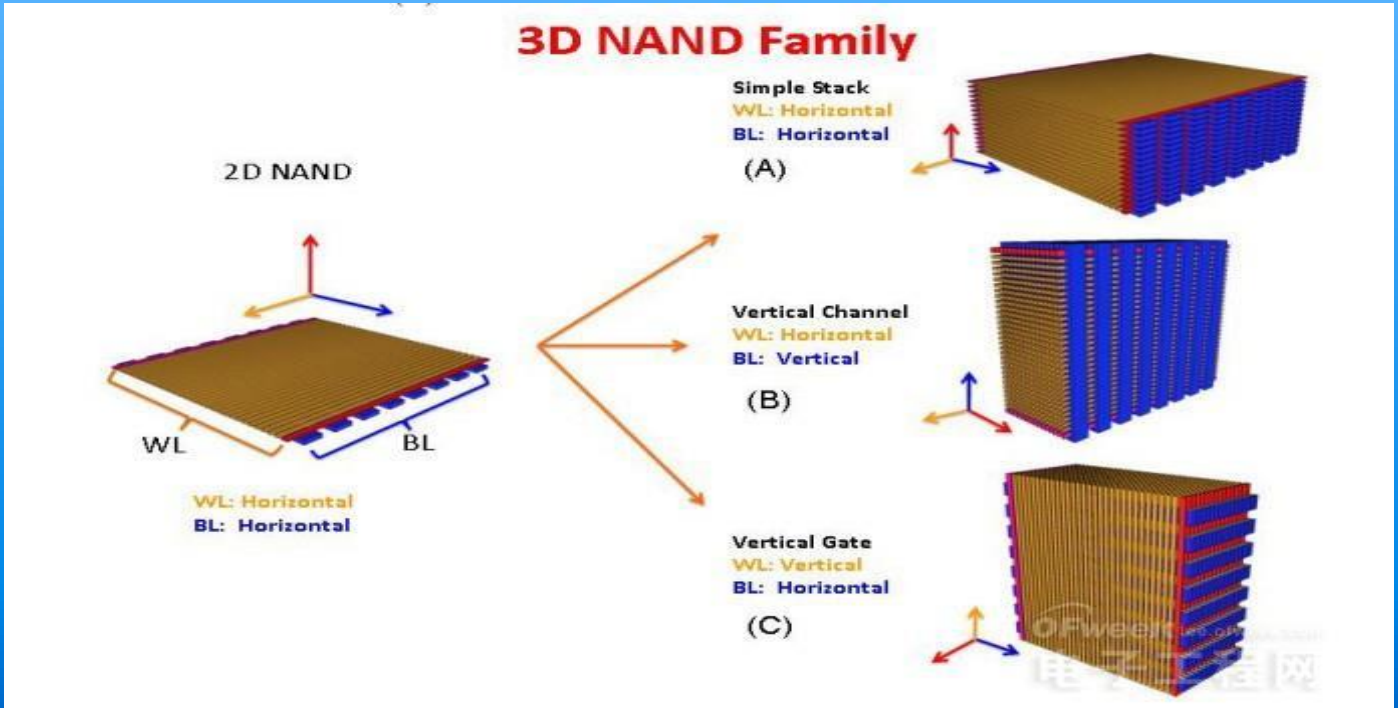


NAND Flash Architecture

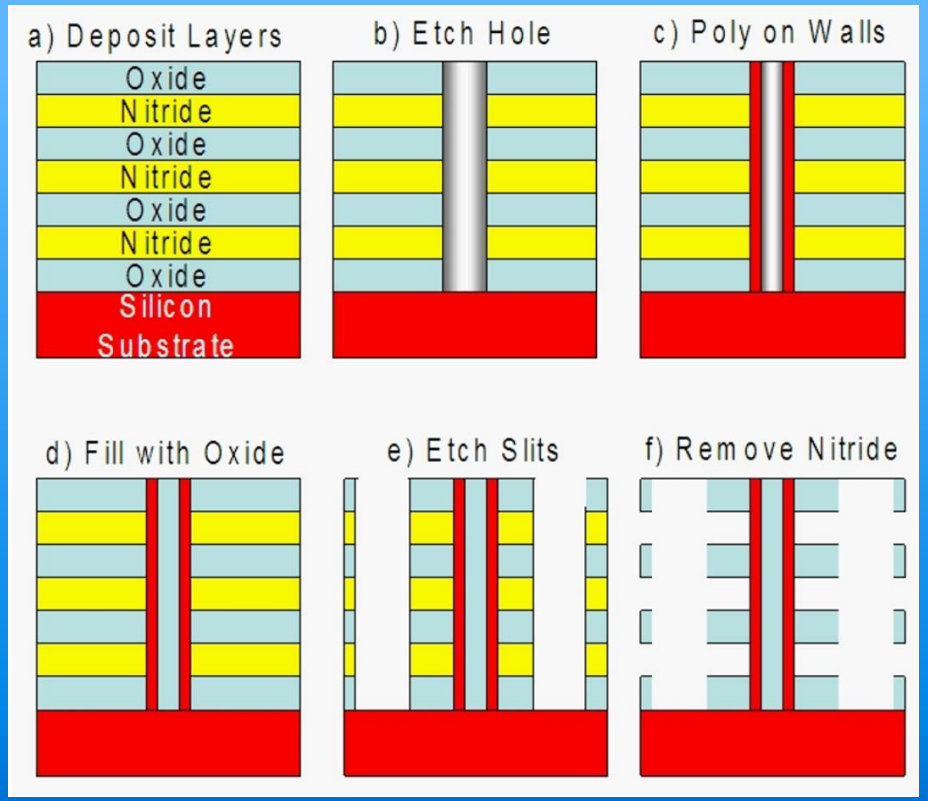


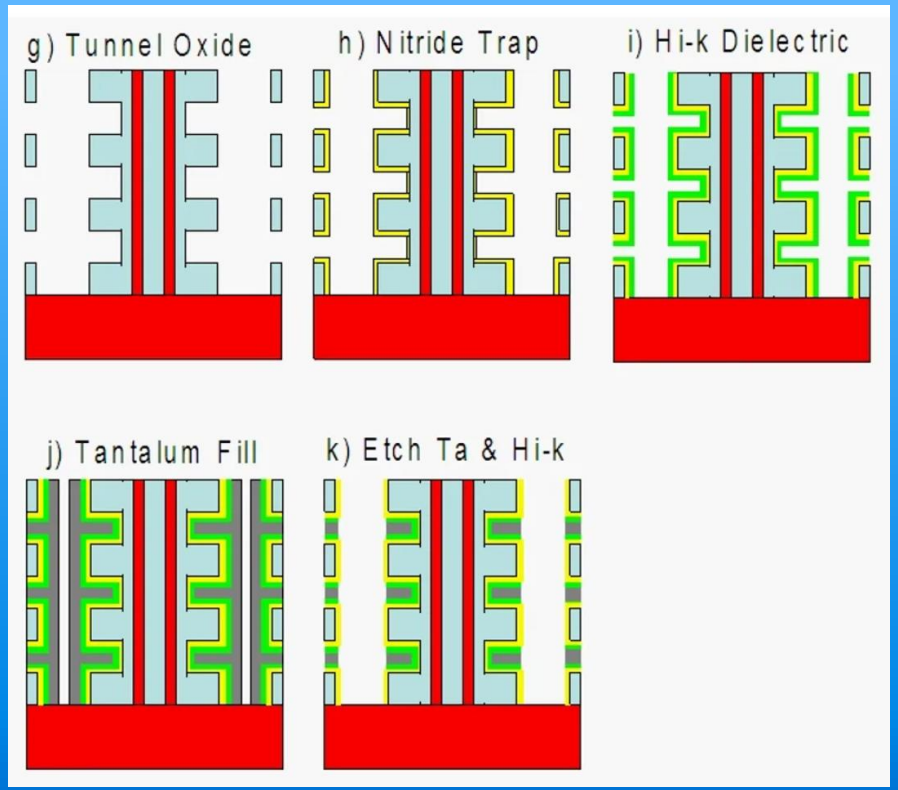
3D NAND

- 3D NAND Family



3D NAND Process is Complex





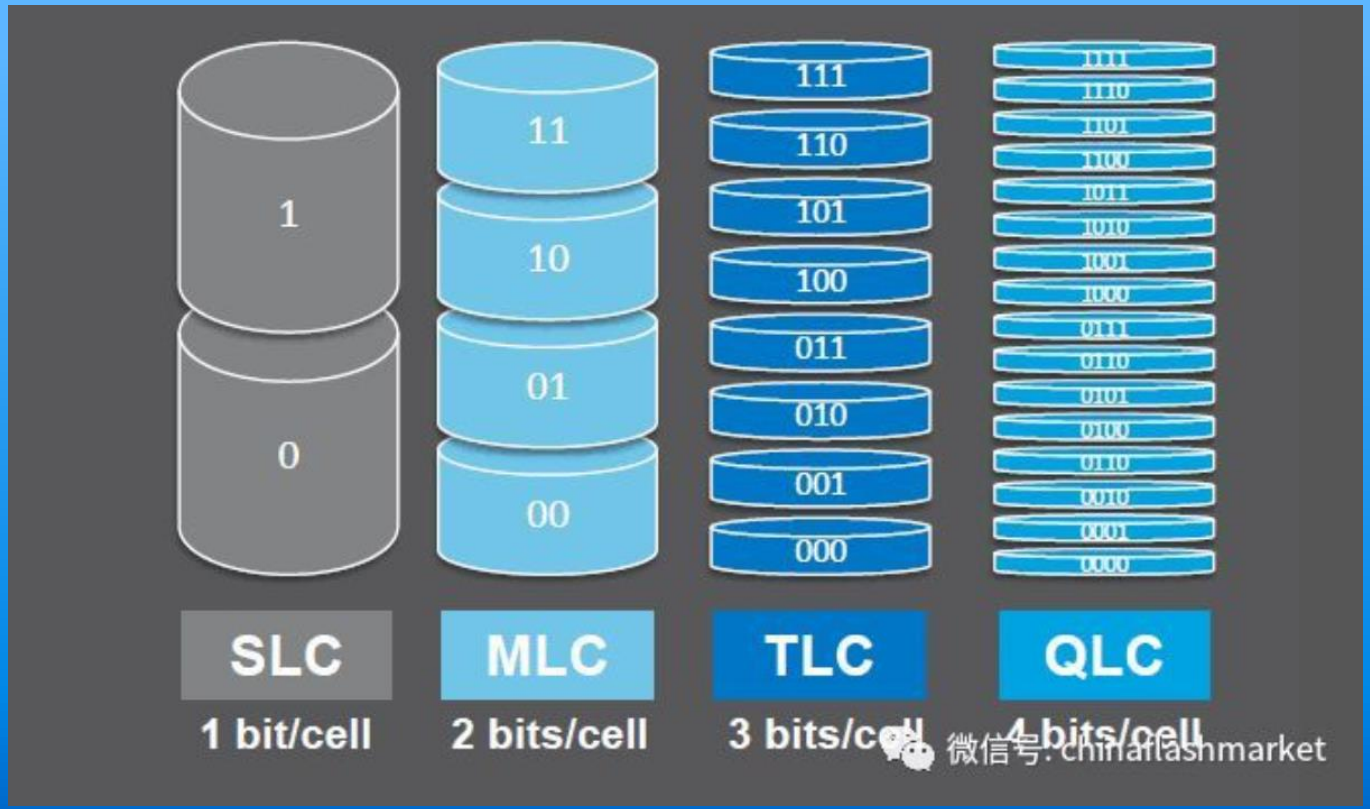
Multi-Level Cell (NAND Flash)

- SLC, Single-level cell, 1 bit/cell, 50k-100k times
- MLC, Multi-level cell, ECC needed
- TLC, Triple-level cell, used in SSD
- QLC!
- Comparison

The **floating-gate MOSFET** (**FGMOS**)

3D NAND Cell Structure

- QLC



3D NAND Controller IC

	型号	制程	DRAM	支持 Flash	Flash 通道	支持接口	接口协议	ECC
Marvell	88NV1160	28nm	N	2D/3D TLC	4	PCIe3.0 x2	AHCI & NVMe1.3	LDPC
	88NV1120	28nm	N	2D/3D TLC	2	SATA 3.0	AHCI	LDPC
	88SS1084	28nm	Y	3D TLC/QLC	4	PCIe3.0 x4	NVMe1.3	LDPC
	88SS1100	28nm	Y	3D TLC/QLC	8	PCIe3.0 x4	NVMe1.3	LDPC
慧荣	SM2263XT	28nm	N	3D TLC/QLC	4	PCIe3.0 x4	NVMe1.3	LDPC
	SM2258XT	40nm	N	3D TLC	4	SATA 3.0	AHCI	LDPC
	SM2262EN	28nm	Y	3D TLC/QLC	8	PCIe3.0 x4	NVMe1.3	LDPC
群联	PS5008-E8T	40nm	N	2D/3D TLC	4	PCIe3.0 x2	NVMe1.2	SmartECC
	PS3111-S11	40nm	N	2D/3D TLC	2	SATA 3.0	AHCI	LDPC
	PS5012-E12	28nm	-	3D TLC/QLC	-	PCIe3.0 x4	NVMe1.3	LDPC

来源：中国闪存市场网 www. chinaflashmarket.com

Flash Memory Top Players

- **3D NAND Flash 2019 Top Players**
  - Samsung Elec., Toshiba/SanDisk, SK Hynix Semi., Micron Tech., Intel Corp
- **NOR Flash Market 2018-2023**
  - Cypress, Samsung, Winbond, Micron, Macronix, ISSI, Eon, Microchip, GigaDevice
- *Storage Patents (25%) owned by*
  - Micron, Samsung Elec., Toshiba, IBM
  - and Intel

NAND

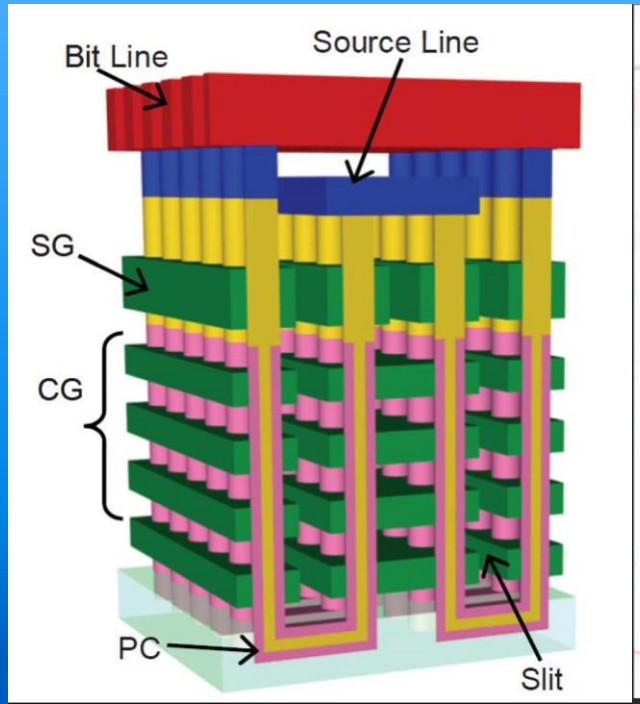
x	y	F
0	0	1
0	1	1
1	0	1
1	1	0

NOR

x	y	F
0	0	1
0	1	0
1	0	0
1	1	0

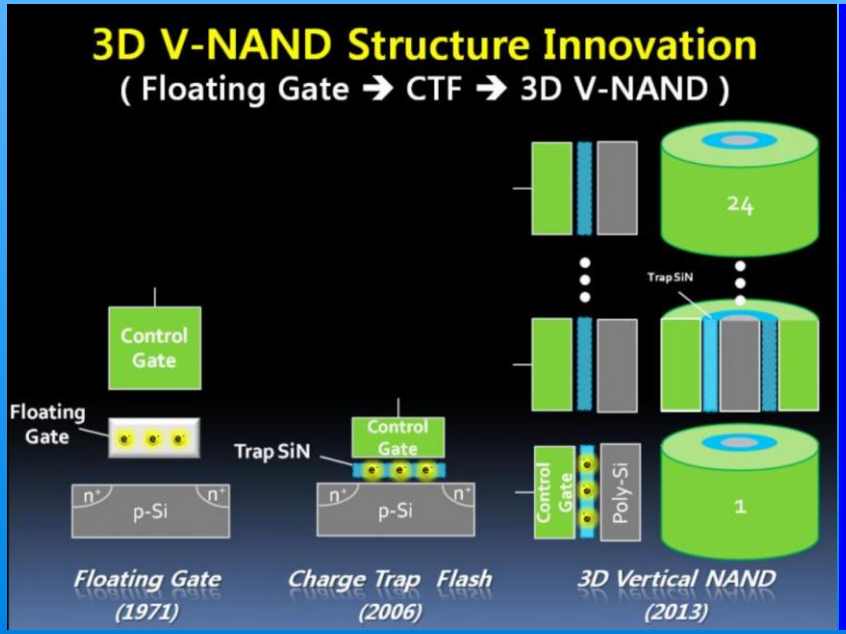
Toshiba/Kioxia and WD

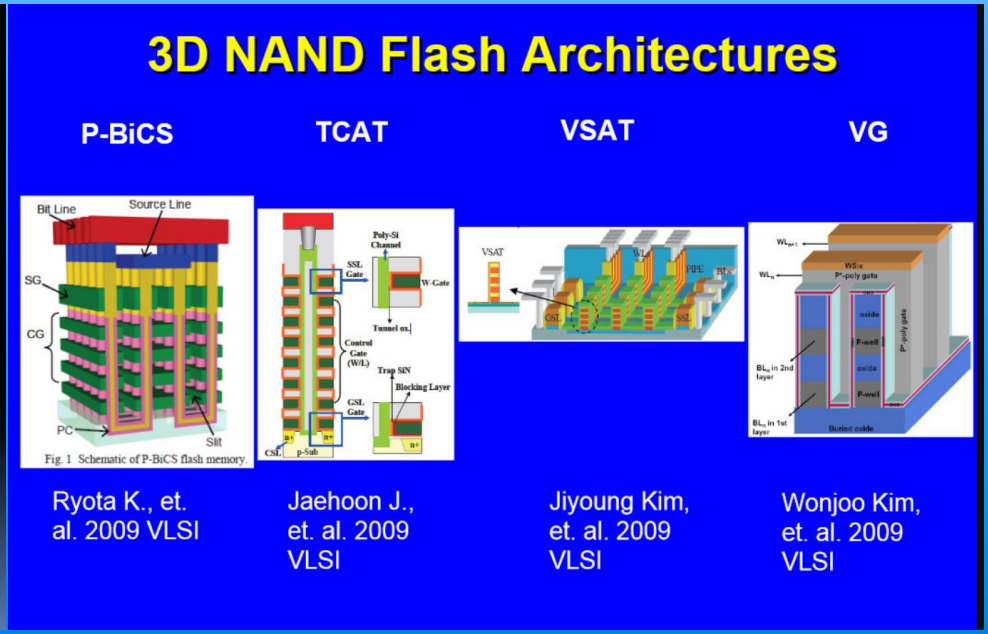
- BiCS, 2007, *"bit cost scalable“*
- BiCS, 2017, *"bit column stacked“ for 3D NAND*



Samsung (HK Hynix 3D FG)

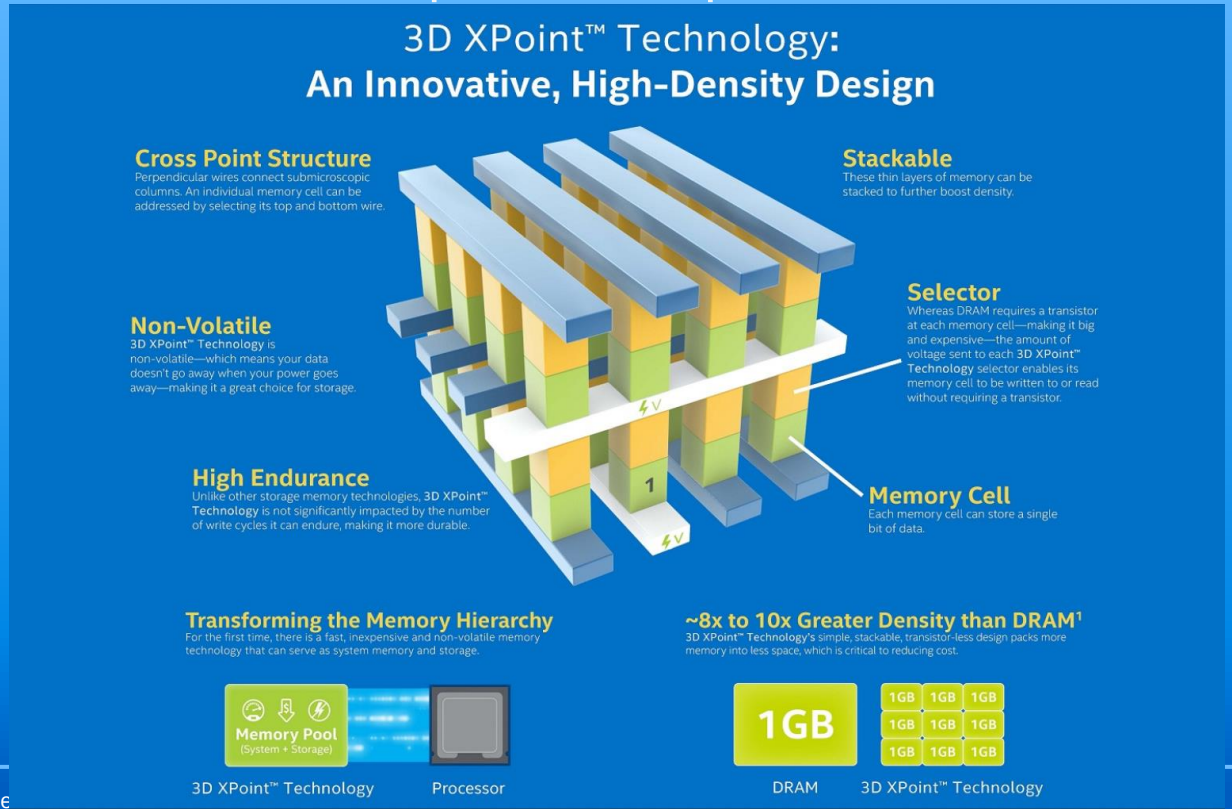
- TCAT, V-NAND





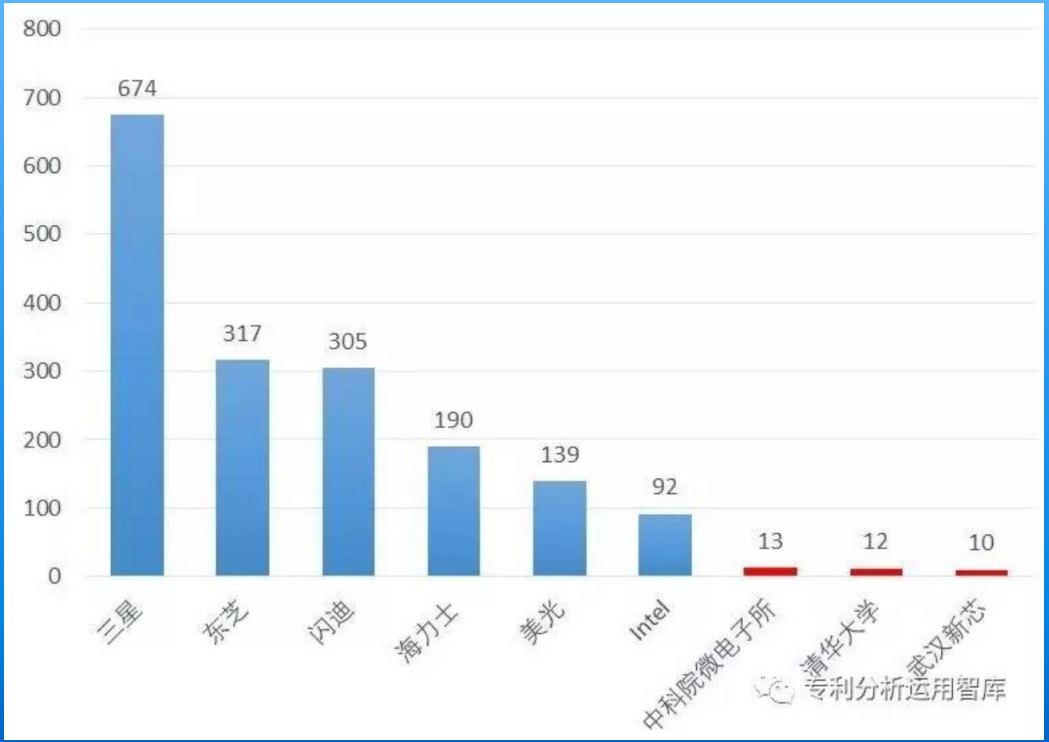
Intel/Micron XPoint

- July 2015, announced; April 2017, Optane

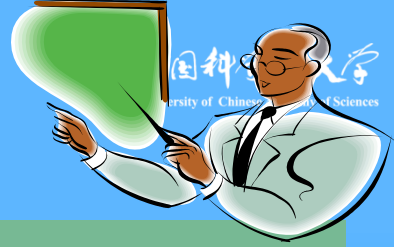


3D NAND Vendors

- Global Number of Leading Inventors on 3D NAND Flash



Memory IP Designs & Applications



- SRAM and Its Application
- DRAM and Its Application
- Flash Memory and Its Application
- New Memory and HBM
- Discussion

New Memory: NVRAM

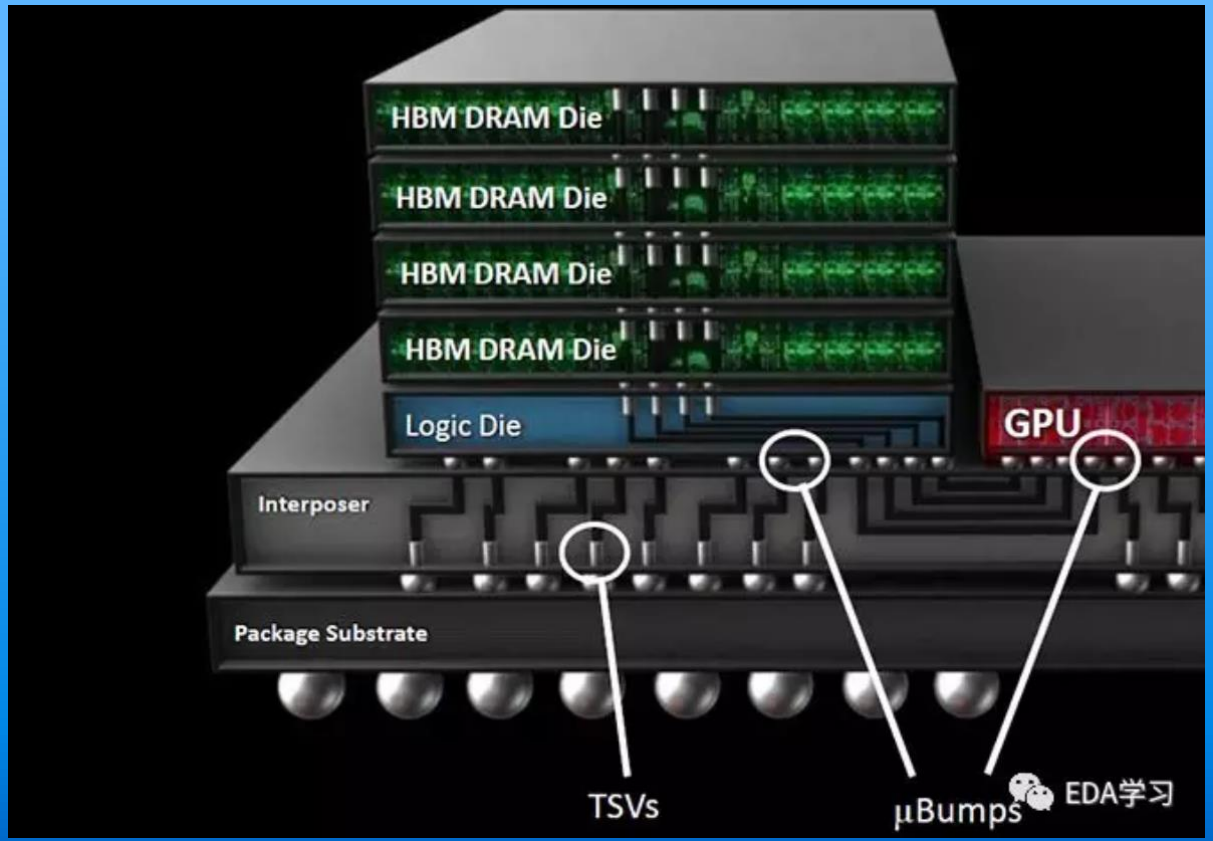
- Ferroelectric RAM (FRAM or FeRAM) → E2PROM
  - popular high-κ gate dielectric HfO_2 (Sharp, Panasonic)
    - CBRAM (conductive-bridging RAM)
    - PCM (phase-change memory)
- Magnetic RAM (MRAM) → E2PROM
  - In production by Everspin Technologies, GF, Samsung
- PCM (PCRAM), an emerging technology

HBM via 3D IC and 2.5D IC

- Existing 4 Packaging Standards
  - Wide-IO (Samsung); HBM (AMD, Nvidia, Hynix);
  - HMC (Micron); **CoWoS & InFO** (TSMC)
- HBM via 3D IC and HBM via 2.5D IC
  - 3D IC (TSV + µBumps); 2.5D IC (Interposer, w/ KGD)
  - Components: DRAMs (DDR, SDRAM), GPU, …
  - FC and BGA
- HBM via FPGA (ex. Xilinx)

HBM via 3D IC and 2.5D IC

Cross Sectional View

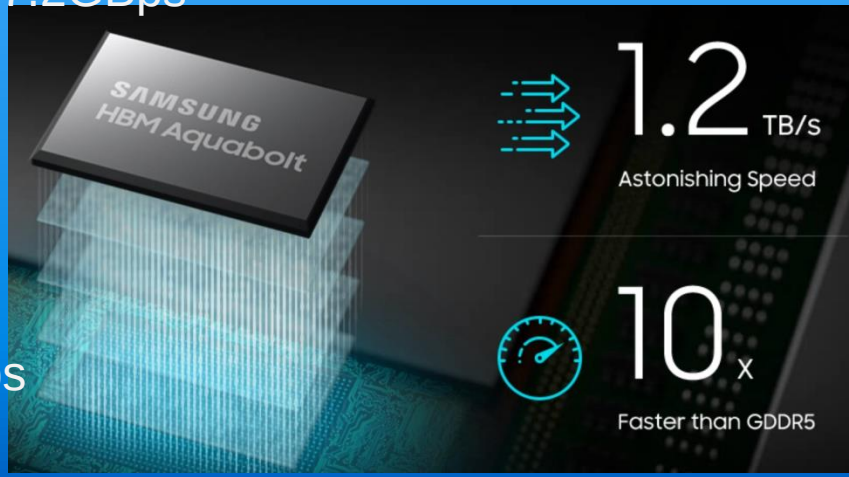


HBM for AI (RAM+GPU/CPU) Available

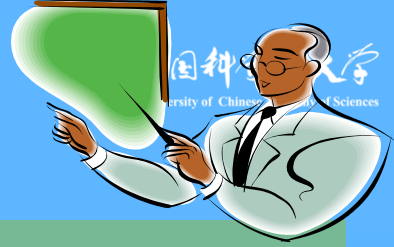
- HBM for CPU/GPU
- Gen1
- Gen2
  - 1024-bit word? 256 Gbps band
- Gen3? X4
  - 4096-bit word? 1024 Gbps band (>1 Tbps)

HBM at Samsung

- Samsung HBM2 Flarebolt's performance: 2.0Gbps at 1.35V
- HBM2 - GDDR5 data transmission speed calculation
  - 8GB HBM2 package's data bandwidth:
    - 2.4Gbps per pin x 1024bit bus = 307.2GBps
- 4 HBM2 package in a system:
  - 307.2GBps x 4 = appr. 1.2TBps
  - 8Gb GDDR5 die's data bandwidth:
    - 8Gbps per pin x 32bit bus = 32GBps



Memory IP Designs & Applications



- SRAM and Its Application
- DRAM and Its Application
- Flash Memory and Its Application
- New Memory and HBM
- Discussion

Emerging Technologies – Persistent Memory

Technology	FRAM	MRAM	ReRAM	PCM	3D XPoint	NRAM	NVDIMM-N
Density	4K-4Mb	256Mb	TBD	128Mb	128Gb	16Gb	32GB
Endurance	10^{12}	10^{12}	10^{5}	10^{8}	10^{12}	10^{11}	∞
Writes	Byte	Byte	Byte	Byte	Byte	Byte	Byte
Read Latency	70-100ns	70ns	25ns	20ns	100ns	100ns	40-180ns
Write Latency	70-100ns	70ns	12µs	65ns	500ns	100ns	40-180ns
Power	Low	Med-Low	Low	Med	Med	Low	High
Interface	DRAM	DDR3 DDR4	Flash-Like	Unique	Unique	DDR4	DDR3 DDR4
Availability	Limited	Prod'n	Alpha	Limited	Samp.	Samp.	Volume

Applications of Memory Cells

- DRAM Market
  - Old: PC 10% decrease
  - New: Mobile, Server, Auto
  - Future: AI Edge Computing (Video Survelliance)
    - MEC (Mobile Edge Computing)
    - Related: Grid Computing
- NAND Market
  - SSD
  - Mobile

SSD

- Application Areas
  - Notebooks, Desktops, Consumer Electronic
  - Embedded market (vehicle, industry control, commercial)
  - Servers, Storage in D. Ctr, Enterprise
  - Mil, Aero, Med
- Interfaces, indcluding
  - Fibre Channel (128 Gbit/s, in Servers)
  - **PCI Express** (PCIe Gen3 x 4, 31.5 Gbit/s)
  - SAS (Serial Attached SCSI, 12.0 Gbit/s, Servers)
  - **USB** (10 Gbit/s)、**Serial ATA** (SATA3.0, 6.0 Gbit/s) etc.

Summary