Introduction to System-on-Chip and its Applications

Introduction to Memory

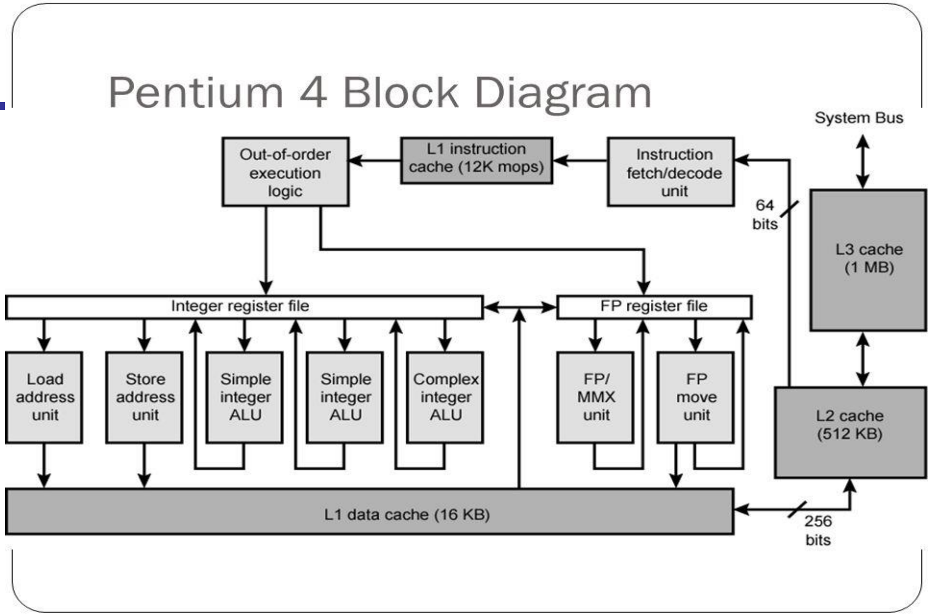
Instructor:Ching-Te Chiu

Outline

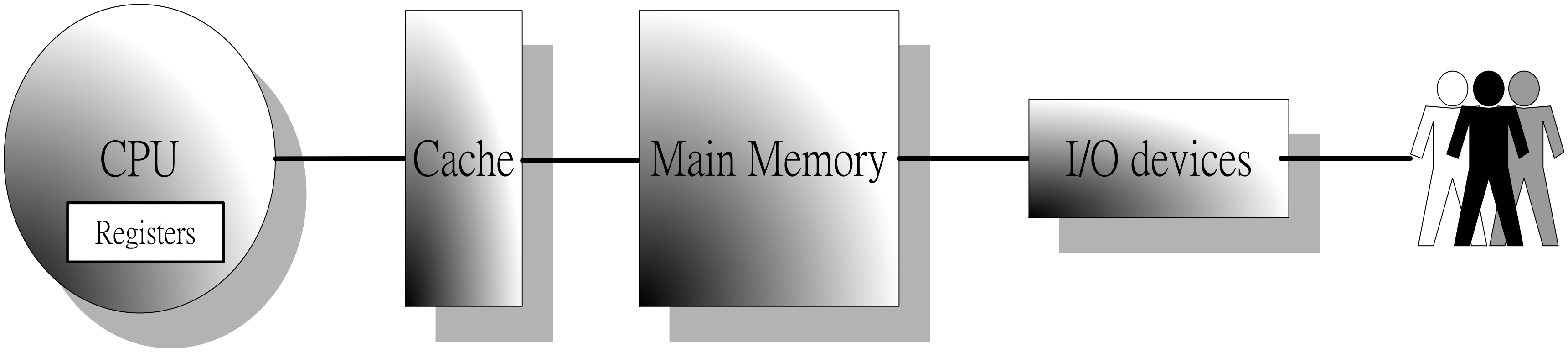
- SRAM (Static Random Access Memory)
- DRAM (Dynamic Random Access Memory)
- Flash memory
- Hard Disk Drive (HDD)
- Solid State Drive (SSD)
- Emerging non-volatile RAM

Memory architecture in desktop computer

- Register (SRAM)
- Cache (SRAM)
- Main Memory (DRAM)
- External Memory (HDD/SSD)



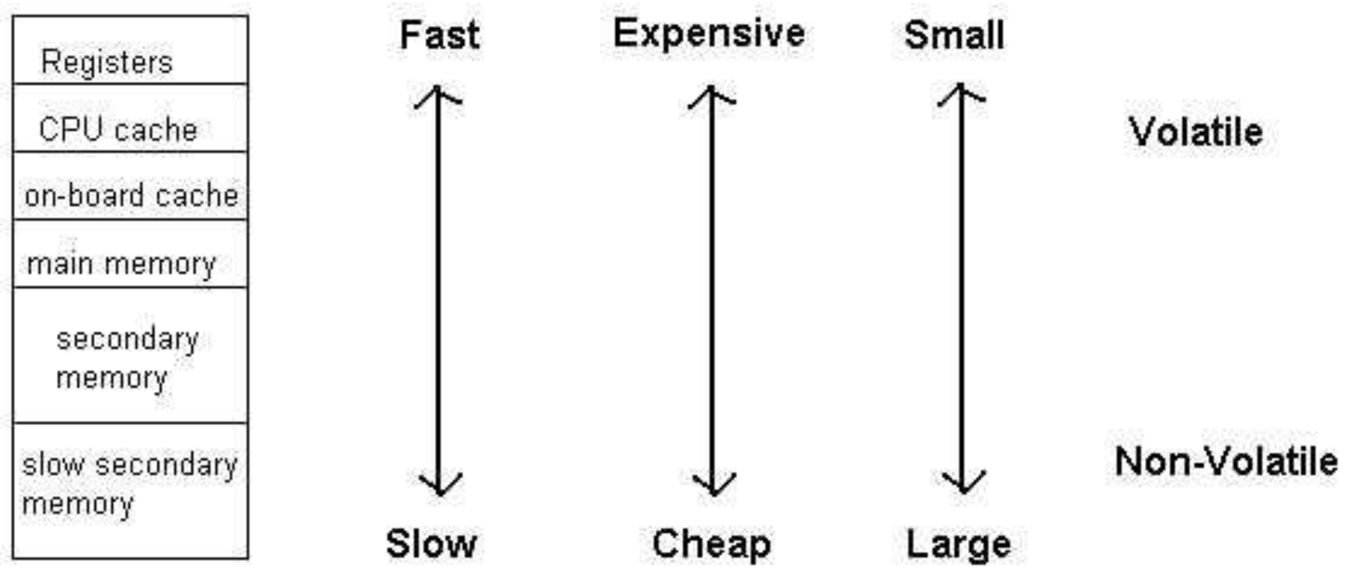
Pentium 4 Block Diagram



Memory hierarchy

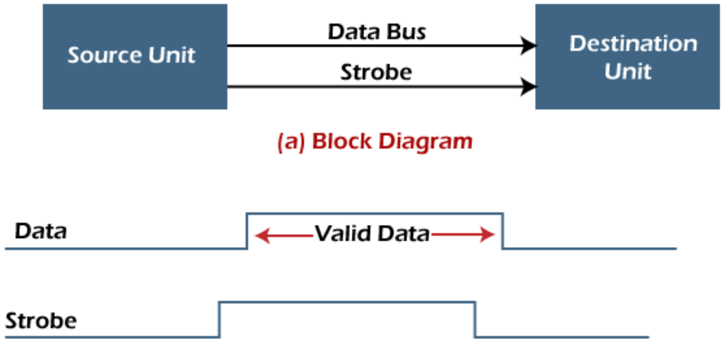
Memory access speed

- **The access speed of memory is generally:**
- CPU register > cache memory> DRAM> hard disk> optical disk> floppy disk



Memory Types

Definitions



(a) Block Diagram

(b) Timing Diagram

Asynchronous

Memory Interfaces for Acessing Data

- Asynchronous (unclocked):

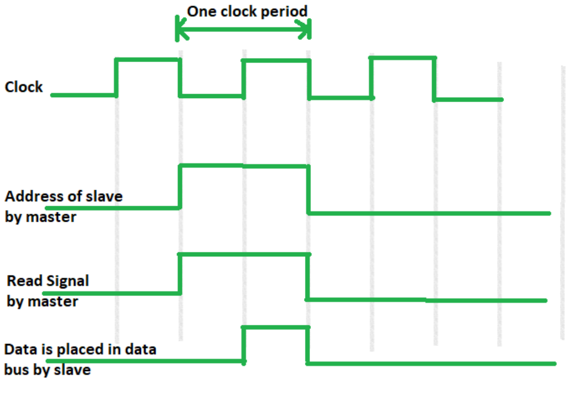
  A change in the address results in data appearing

- Synchronous (clocked):

  A change in address, followed by an edge on CLK results in data appearing or write operation occuring.

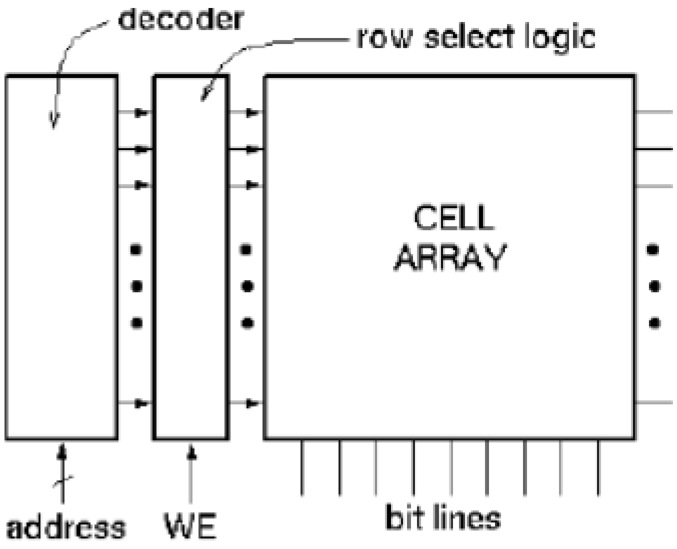
- **Volatile:**

  Looses its state when the power goes off.


One clock period
Clock
Address of slave by master
Read Signal by master
Data is placed in data bus by slave

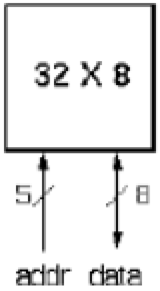
Timing diagram for Synchronous Read Operation

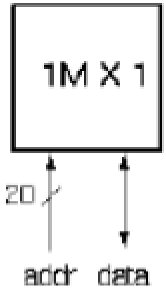
Standard Internal Memory Organization



- Special circuit tricks are used for the cell array to improve storage density.

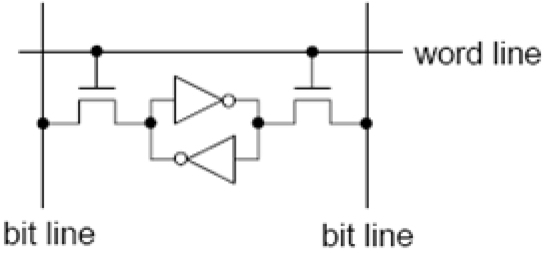
- RAM/ROM naming convention:
  - examples: 32 X 8, "32 by 8" => 32 8-bit words
  - 1M X 1, "1 meg by 1" => 1M 1-bit words

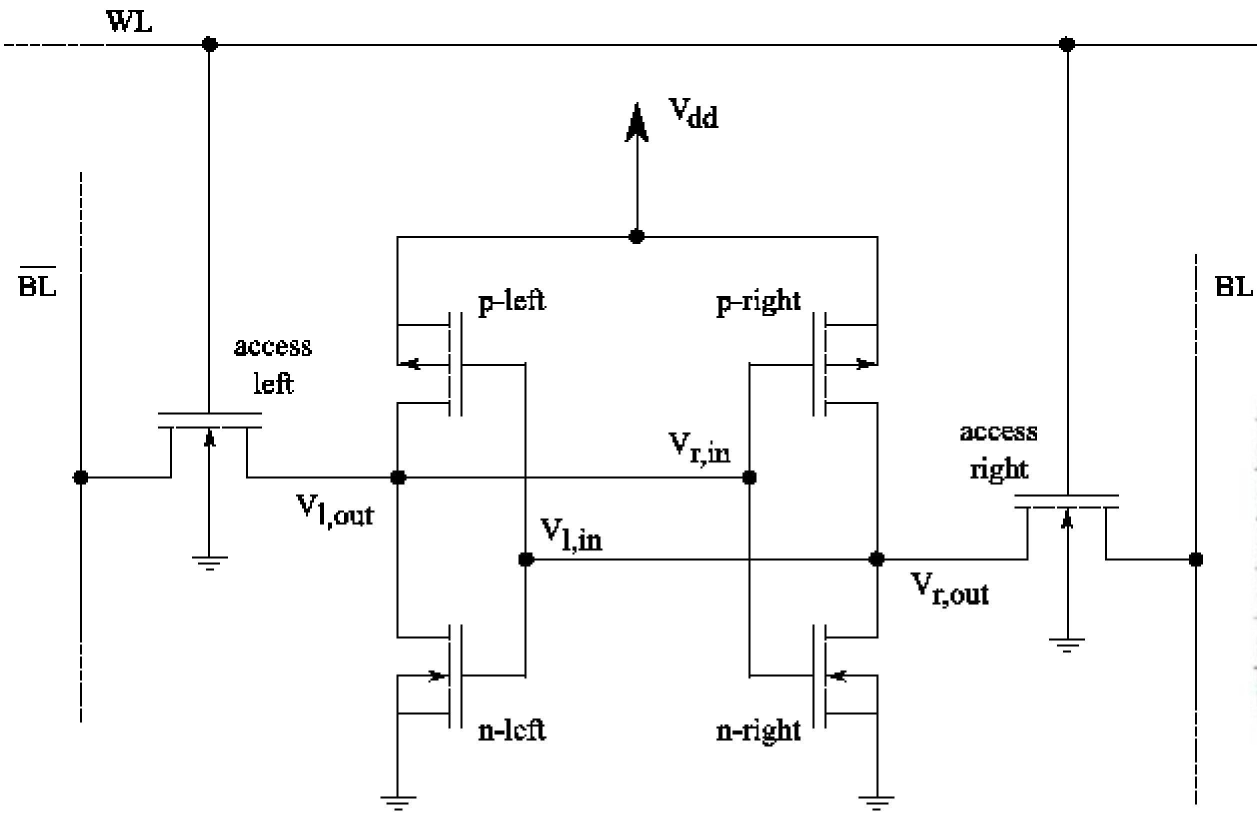


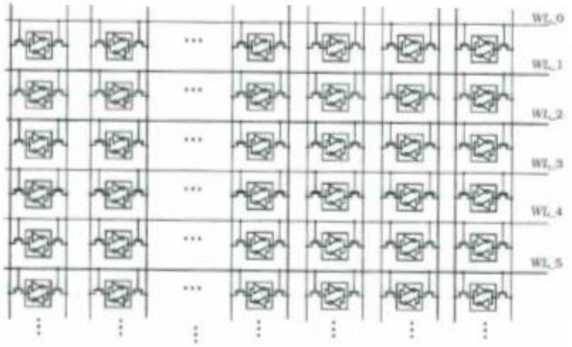


SRAM Cell Circuit Module

- **SRAM (Static Random Access)(6T)**
  - ➔ Six transistors to store one bit
  - ➔ Registers/ Cache





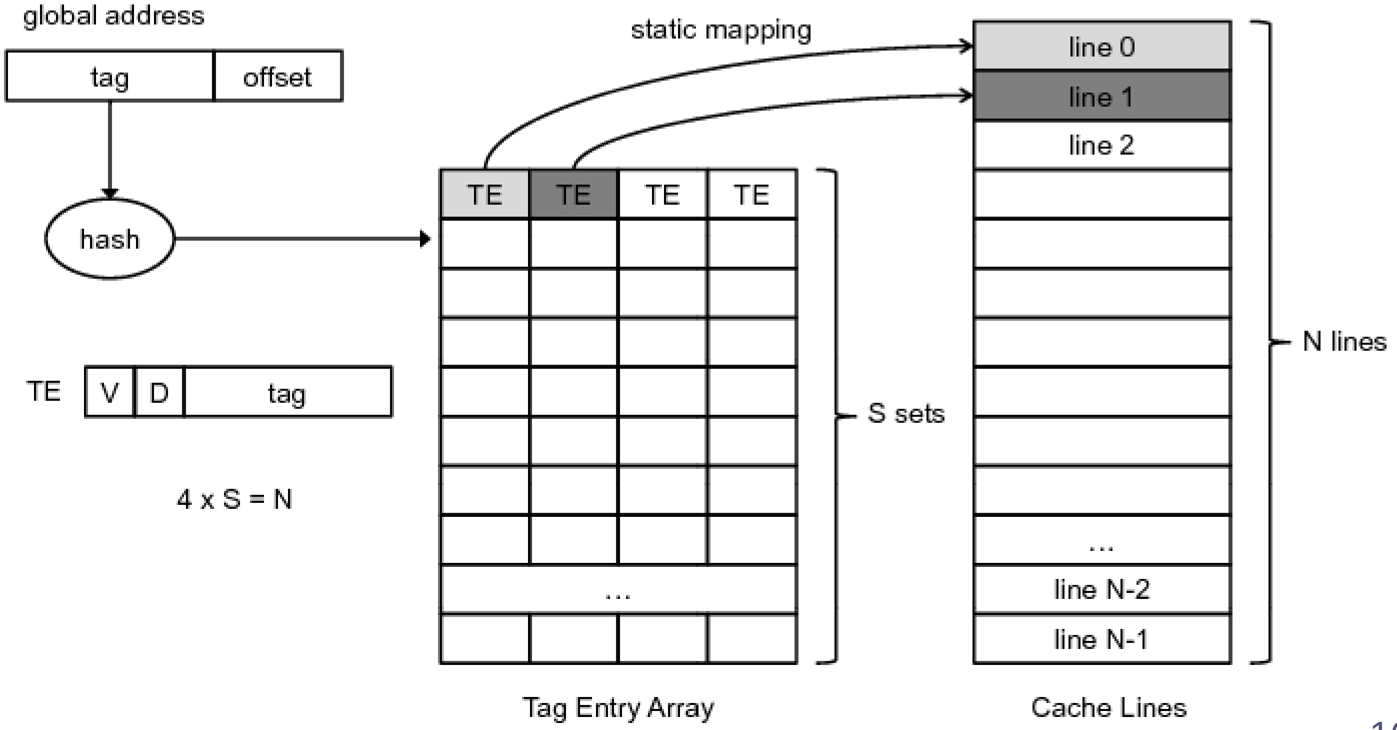


Pentium 4 Cache (Static RAM)

- **80386**
  - no on chip cache
- **80486**
  - 8k using 16 byte lines and four way set associative organization
- **Pentium (all versions)**
  - two on chip L1 caches
    - Data & instructions
- **Pentium 4**
  - L1 caches
    - 8k bytes
    - 64 byte lines
    - four way set associative
  - L2 cache
    - Feeding both L1 caches
    - 256k
    - 128 byte lines
    - 8 way set associative

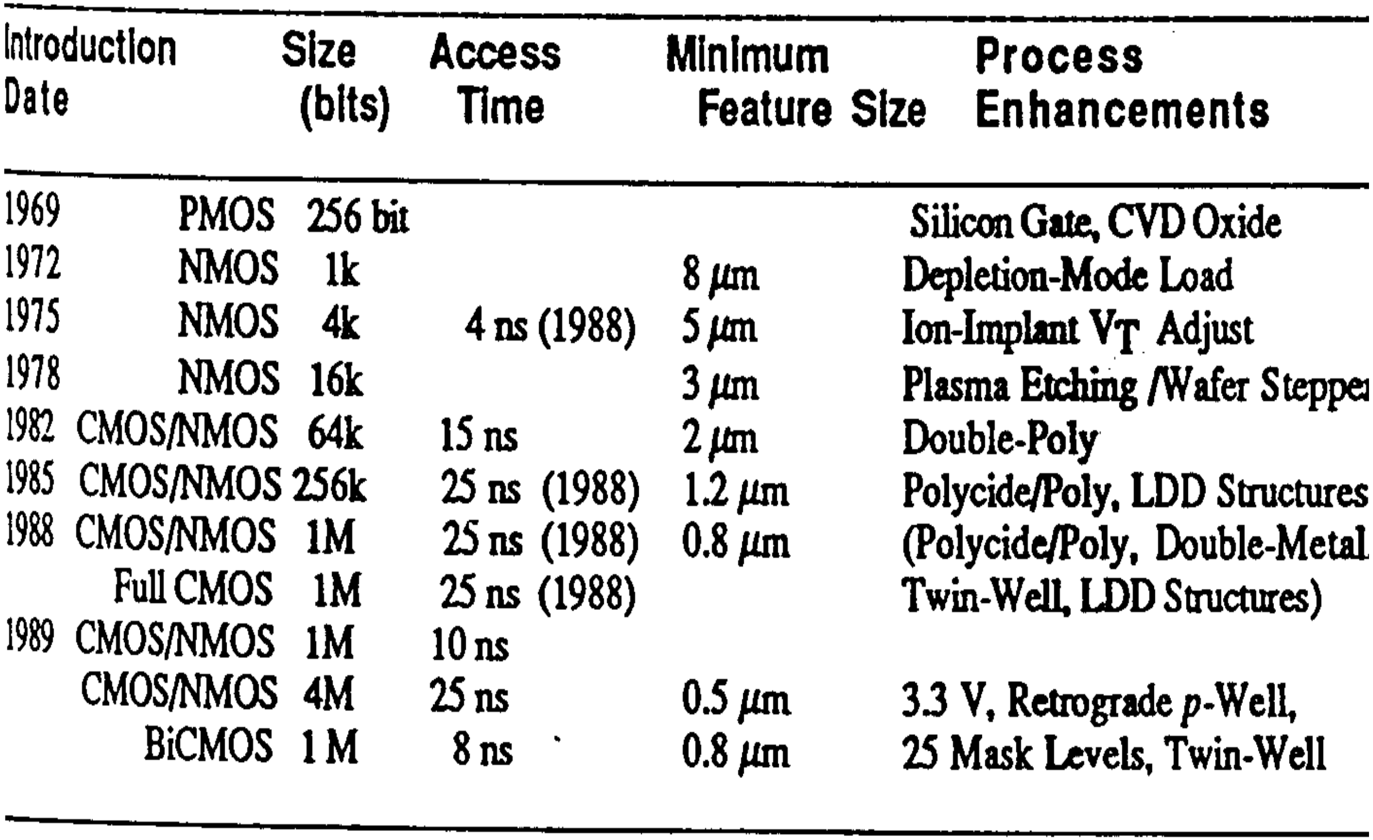
Set Associate Memory

- Mapping in a cache system,
  - ➔ **direct mapping** maps each block of main memory into only one possible cache line.
  - ➔ **set-associative mapping**, the cache is divided into a number of **sets** of cache lines; each main memory block can be **mapped** into any line in a particular **set**
- **Avoid memory collision or looping problem**



SRAM Technology Evolution

- 2022 AMD Ryzen 9 TSMC N5 Cache L1/L2/L3 (1MB/16MB/64MB)



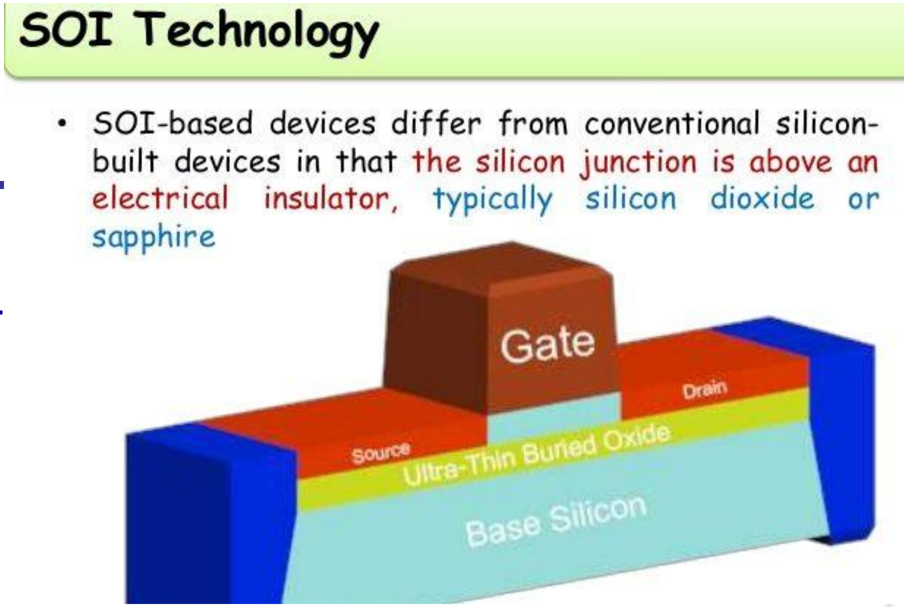
Introduction Date		Size (bits)	Access Time	Minimum Feature Size	Process Enhancements
1969	PMOS	256 bit			Silicon Gate, CVD Oxide
1972	NMOS	1k		8 μm	Depletion-Mode Load
1975	NMOS	4k	4 ns (1988)	5 μm	Ion-Implant V_T Adjust
1978	NMOS	16k		3 μm	Plasma Etching /Wafer Steppe
1982	CMOS/NMOS	64k	15 ns	2 μm	Double-Poly
1985	CMOS/NMOS	256k	25 ns (1988)	1.2 μm	Polycide/Poly, LDD Structures
1988	CMOS/NMOS	1M	25 ns (1988)	0.8 μm	(Polycide/Poly, Double-Metal
	Full CMOS	1M	25 ns (1988)		Twin-Well, LDD Structures)
1989	CMOS/NMOS	1M	10 ns		
	CMOS/NMOS	4M	25 ns	0.5 μm	3.3 V, Retrograde *p*-Well,
	BiCMOS	1 M	8 ns	0.8 μm	25 Mask Levels, Twin-Well

SRAM

- High-performance multimedia processors to drive the embedded SRAM in a single die.
- low-power consumption in SRAM

  -- in mobile and hand-held devices

  --environmental and biomedical sensors

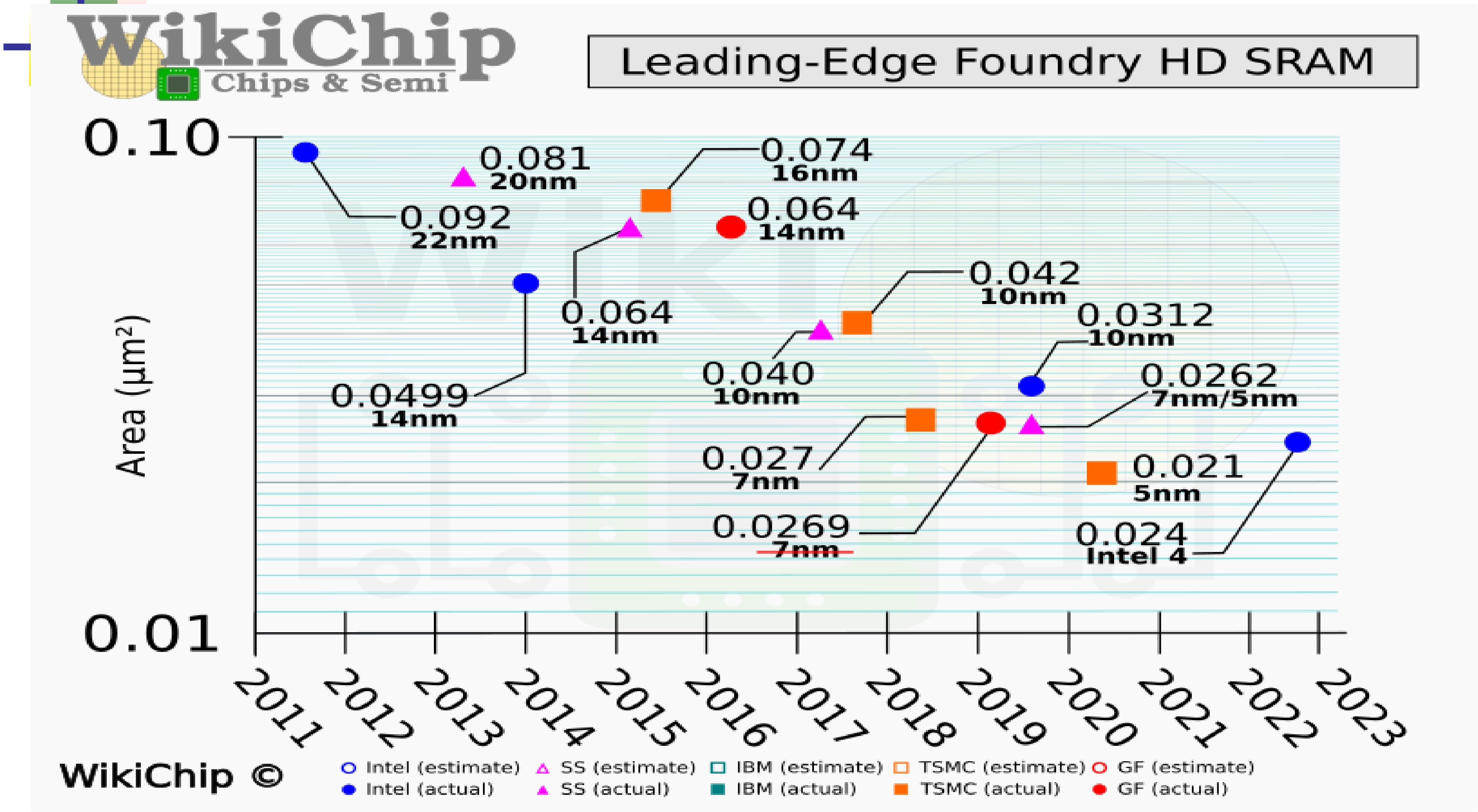


SOI Technology

- SOI-based devices differ from conventional silicon-built devices in that the silicon junction is above an electrical insulator, typically silicon dioxide or sapphire

- **Current status**
- in a 65nm silicon-on-insulator (SOI) technology with a large-signal ripple-domino sensing scheme to achieve up to 6GHz operation at 1.3V
- SOI advantages-**Reduced Source and Drain to Substrate Capacitance**
- At read, the sense amplifier at the end of the two complimentary bit-lines amplify the small voltages to a normal logic level.

High Density (HD) SRAM (Leading-Edge Foundry)

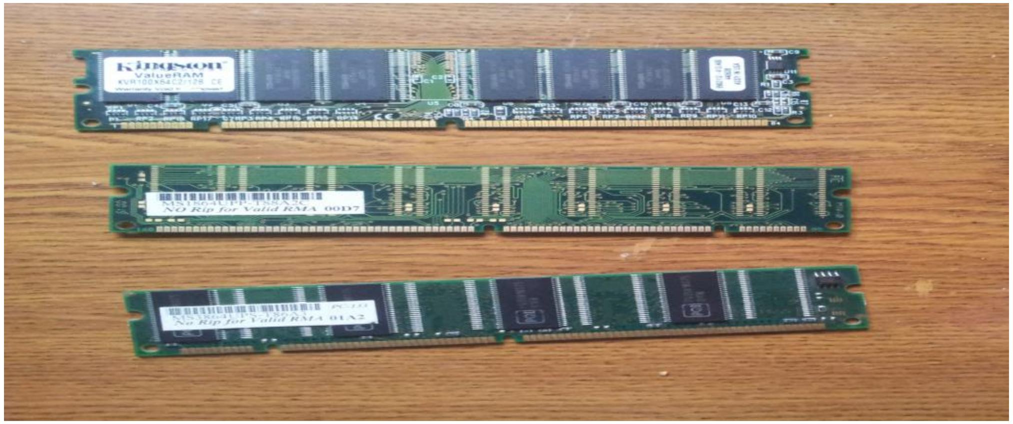


Intel 4 HDC produces a memory density of around 27.8 Mib/mm². Compared to TSMC N5 SRAM which boasts a density of 31.8 Mib/mm², Intel is roughly 14.5% less dense.

https://fuse.wikichip.org/news/6720/a-look-at-intel-4-process-technology/4/

Main Memory Types-DRAM

- **Main memory**
  - ➔ Plugged into the motherboard
  - ➔ The more popular ones currently on the market that plug into the motherboard have the following specifications:
    - ➲ SDRAM (Synchronous Dynamic RAM)
    - ➲ DDRAM (Double Data Rate RAM)
    - ➲ RamBUS ram
    - ➲ SO-DIMM (notebook only)(small outline dual in-line memory module)(pin number is half of the DIMM)



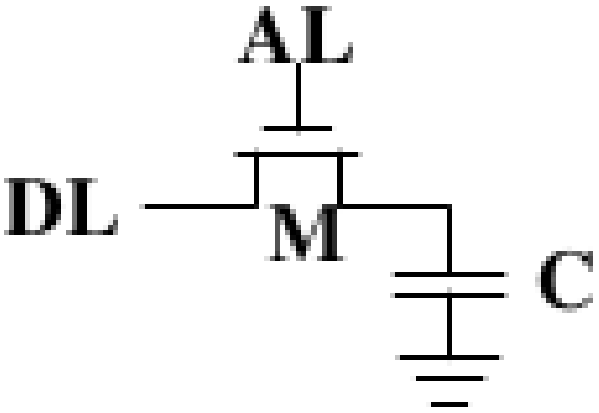
DRAM (Dynamic Random Access Memory)

- **Two-way memory that can be read and stored。**
- **Temporary data or program。**
- **After shutting down, the data disappeared。**
- **Generally speaking, computer capacity refers to RAM。**
- **RAM capacity has 512MB、1GB, 4GB, 16GB (2^n)**
  - ➔ Kingston (John Tu and David Sun)
  - ➔ Micron
  - ➔ Samsung
- 2022 AMD Ryzen 9 uses DDR5-5200 dual memory 32GB

DRAM Cell (Dynamic Random Access Memory)

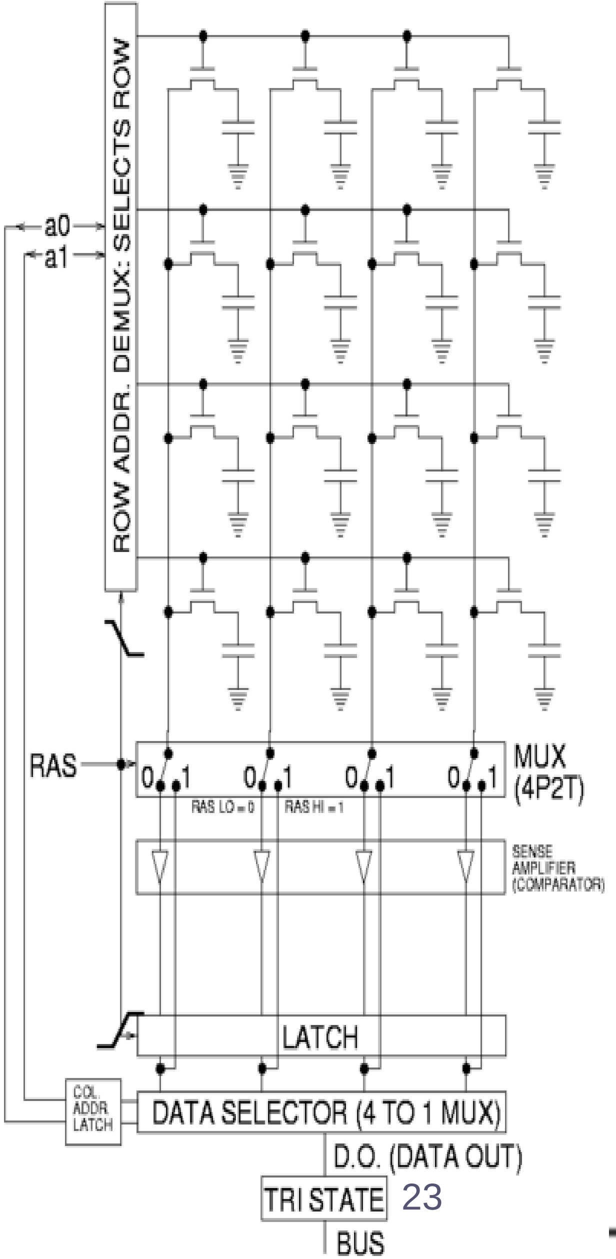
- **A Dynamic RAM Cell (1T1C)**
  - ➔ capacitor C
  - ➔ transistor M
  - ➔ access line AL
  - ➔ data line DL
  - ➔ Data values by

    electrical charges stored at capacitor

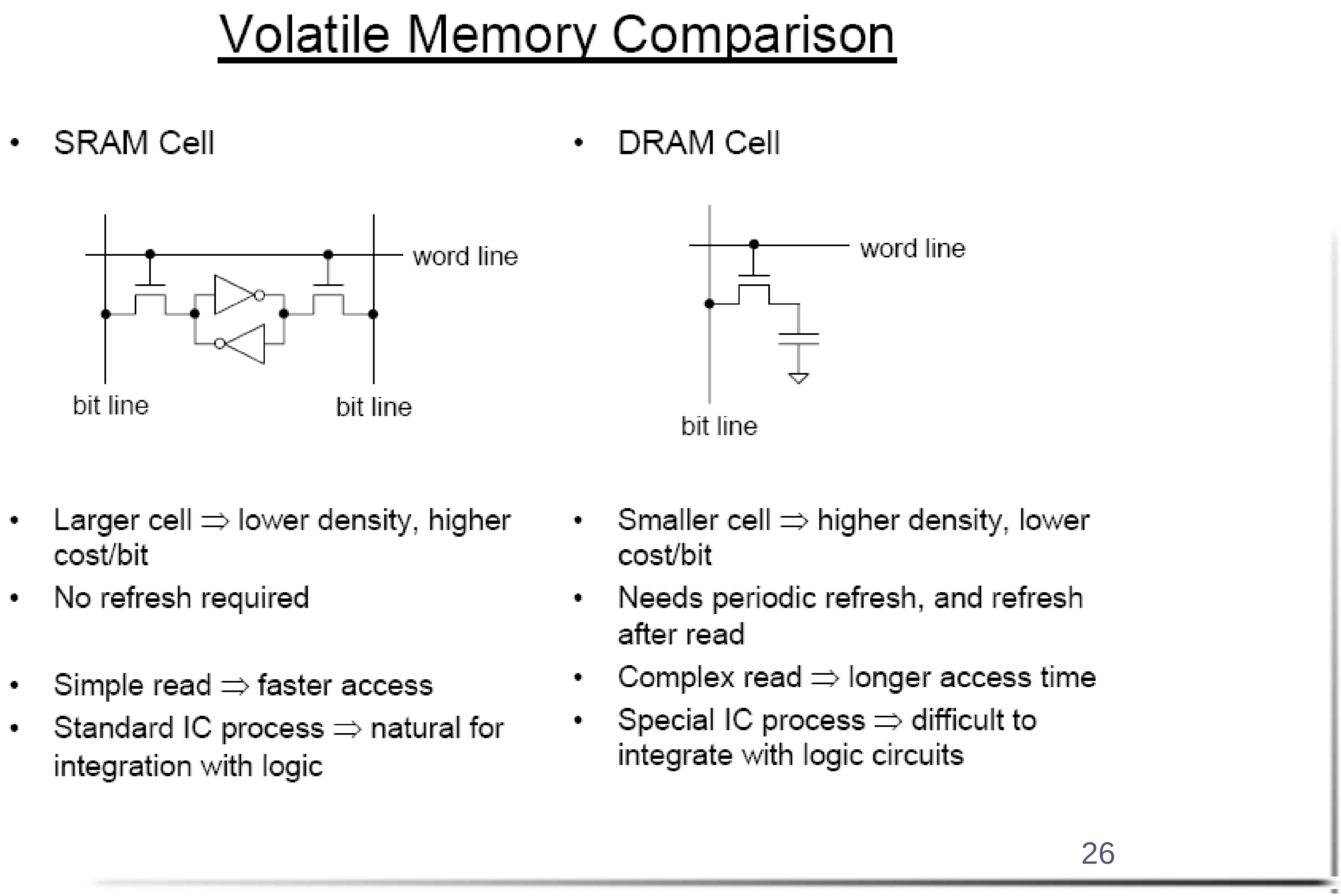


Main Memory DRAM

- Main Memory
- stores each bits of data in a separate capacitor
- Why it’s called dynamic?
- Real capacitors leak charge， the capacitor charge is refreshed periodically



SRAM and DRAM Comparison



Volatile Memory Comparison

- SRAM Cell

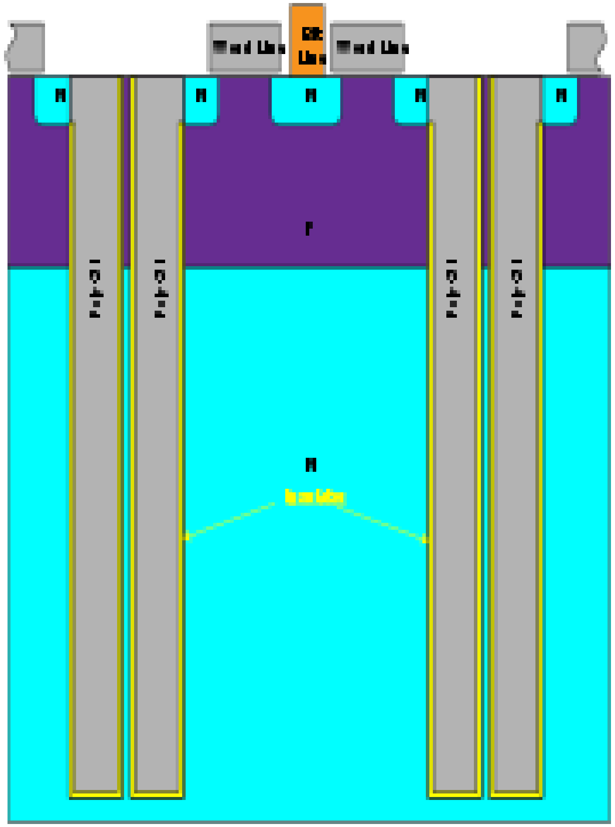
- DRAM Cell

- Larger cell ⇒ lower density, higher cost/bit
- No refresh required
- Simple read ⇒ faster access
- Standard IC process ⇒ natural for integration with logic

- Smaller cell ⇒ higher density, lower cost/bit
- Needs periodic refresh, and refresh after read
- Complex read ⇒ longer access time
- Special IC process ⇒ difficult to integrate with logic circuits

DRAM Process Types -I

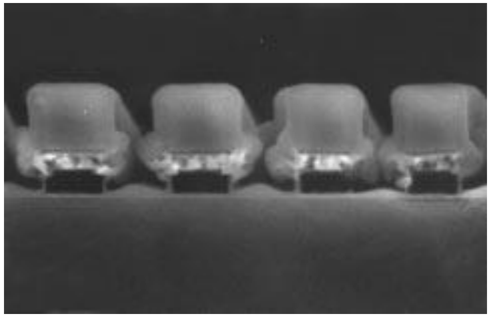
- **Trench**
  - ➔ Digging **trenches** on the surface of the wafer to expand the surface area and **expand the capacitance**.
  - ➔ Advantage
    - ➲ the number of die per wafer is about 10% more than that of the stacked type,
  - ➔ Disadvantage
    - ➲ Physical properties involved in the high-end process are high,
    - ➲ which may affect the yield,
    - ➲ the process below 50nm will be affected.

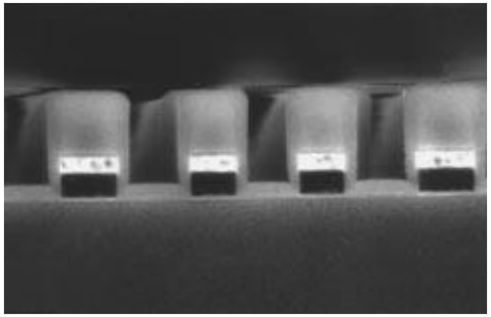


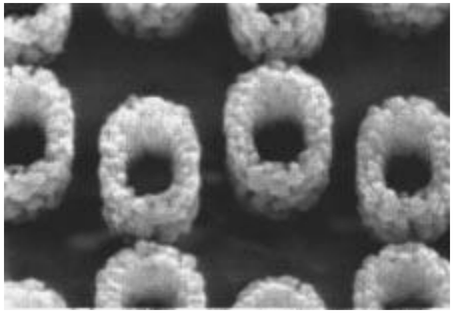
DRAM Process Types - II

- **Stack**
  - ➔ The stacked type uses the upper stacking method to increase the surface area and increase the capacitance.
  - ➔ Advantage
    - ➲ capacitance is good, and the physical limitations of the high-end manufacturing process are easy to overcome
  - ➔ Disadvantage
    - ➲ it is not possible to the development of system single chips.
  - ➔ ref:”process simplicifcation in DRAM manufacturing”,

IEEE trans. on electron devices vol. 45, No 3, March 1998.

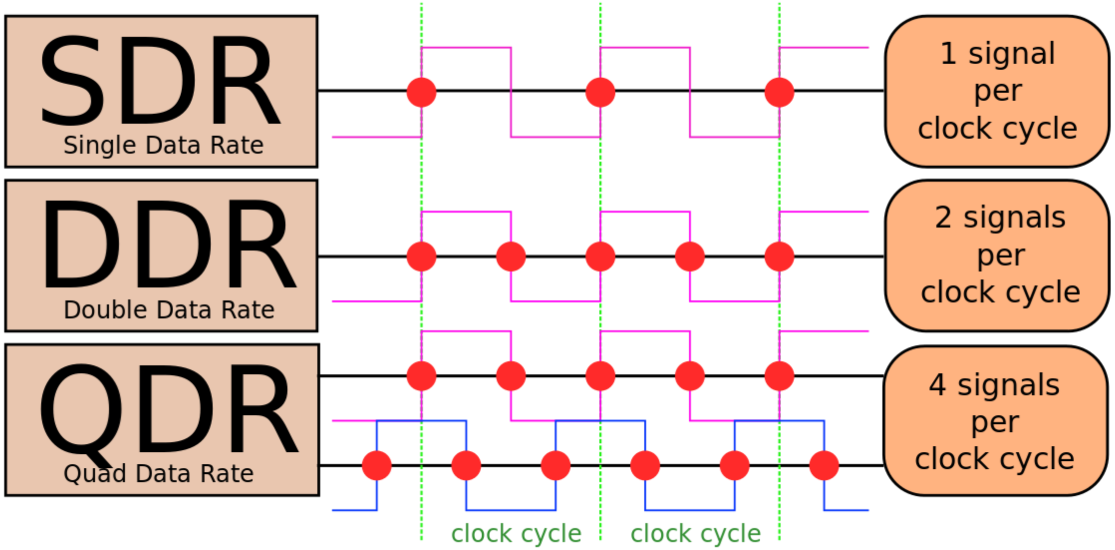




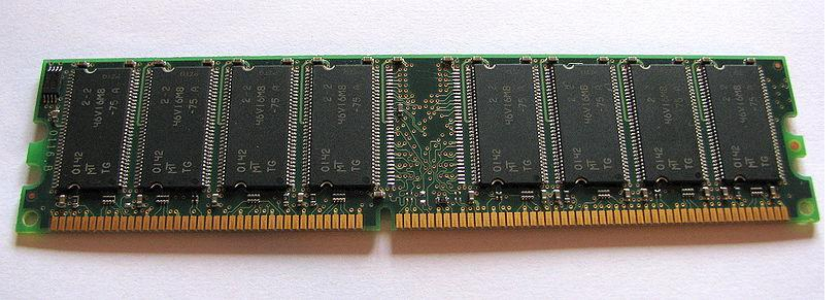


DRAM types

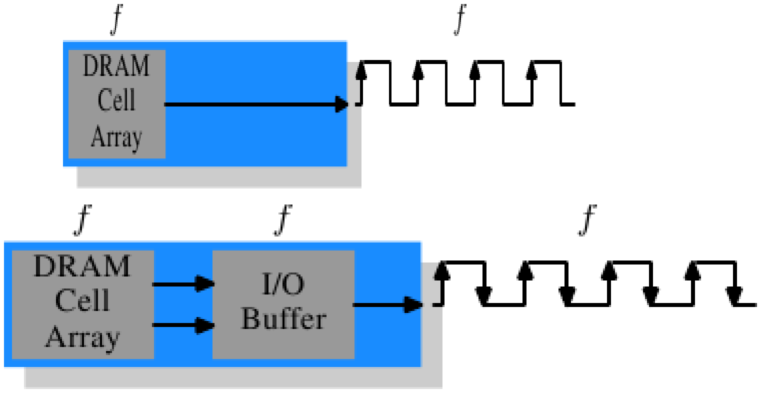
- **SDRAM**
  - ➔ **synchronous dynamic random access memory (SRAM)**
  - ➔ synchronized with the computer's system bus
- **Data Per Cycle**

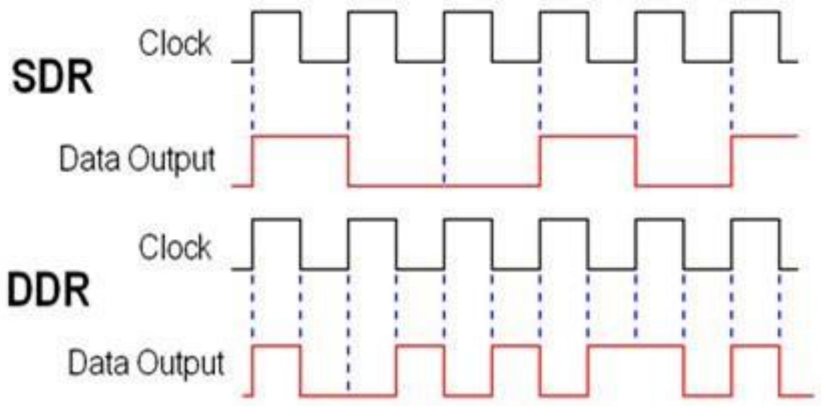


DRAM types



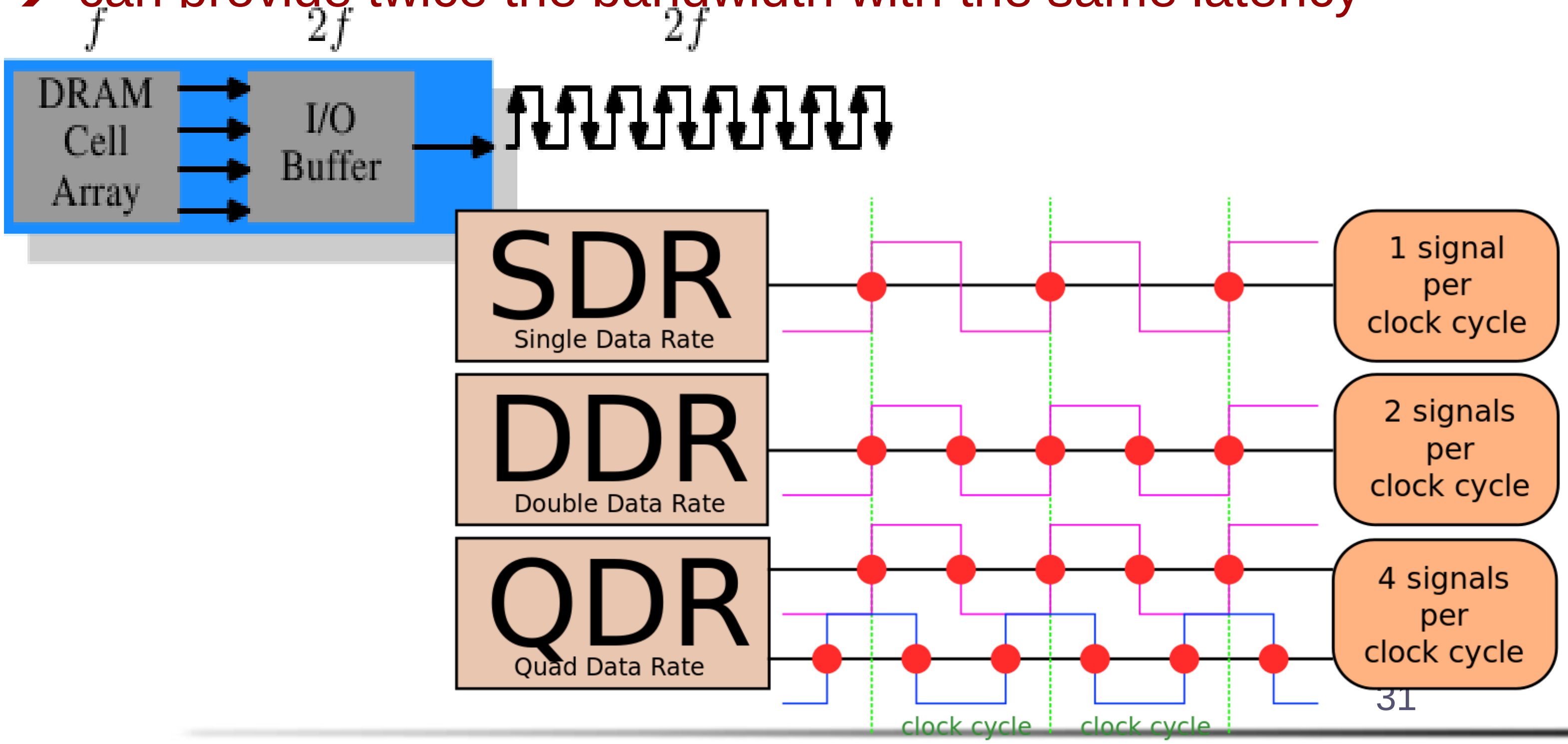
- **Single Data Rate**
  - ➔ can accept one command and transfer one word of data per clock cycle
- **Double Data Rate (DDR) SDRAM**
  - ➔ **double-data-rate synchronous dynamic random access memory**
  - ➔ double pumping (transferring data on the rising and falling edges of the clock signal) without increasing the clock frequency



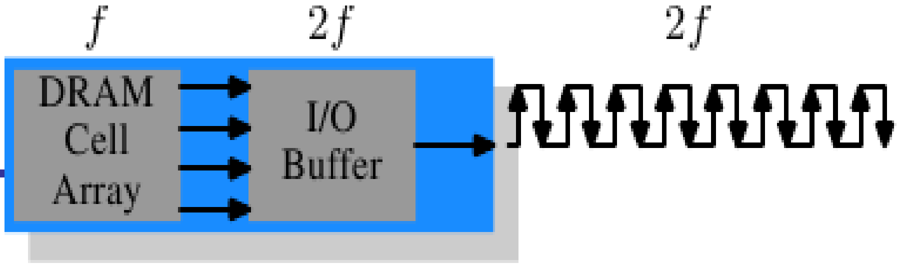


DDR2 DRAM

- **DDR2 DRAM**
  - ➔ the bus is clocked at twice the rate of the memory cells, so four bits of data can be transferred per memory cell cycle
  - ➔ at half the clock rate (one quarter of the data transfer rate)
  - ➔ can provide twice the bandwidth with the same latency

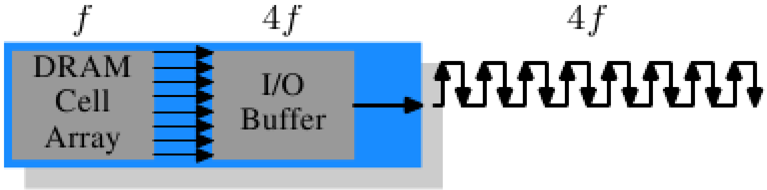


DDR3 DRAM



DDR2

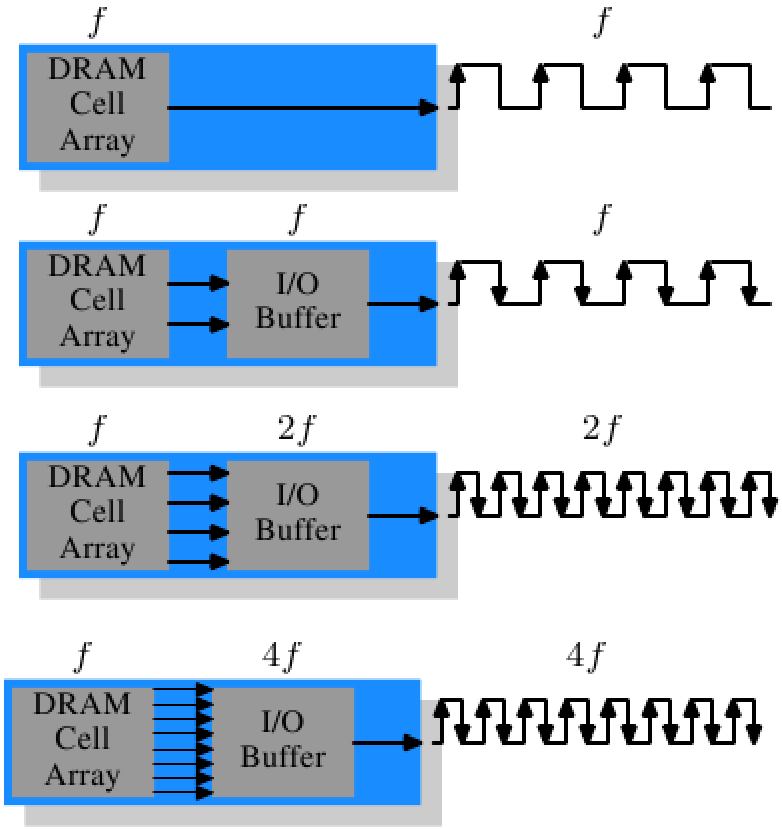
- **DDR3 DRAM**
  - ➔ transfer twice the data rate of DDR2 (I/O at 8× the data rate of the memory cells it contains)
  - ➔ a reduction in power consumption of 30% compared to DDR2 modules due to DDR3's 1.5 V supply voltage, compared to DDR2's 1.8 V or DDR's 2.5 V



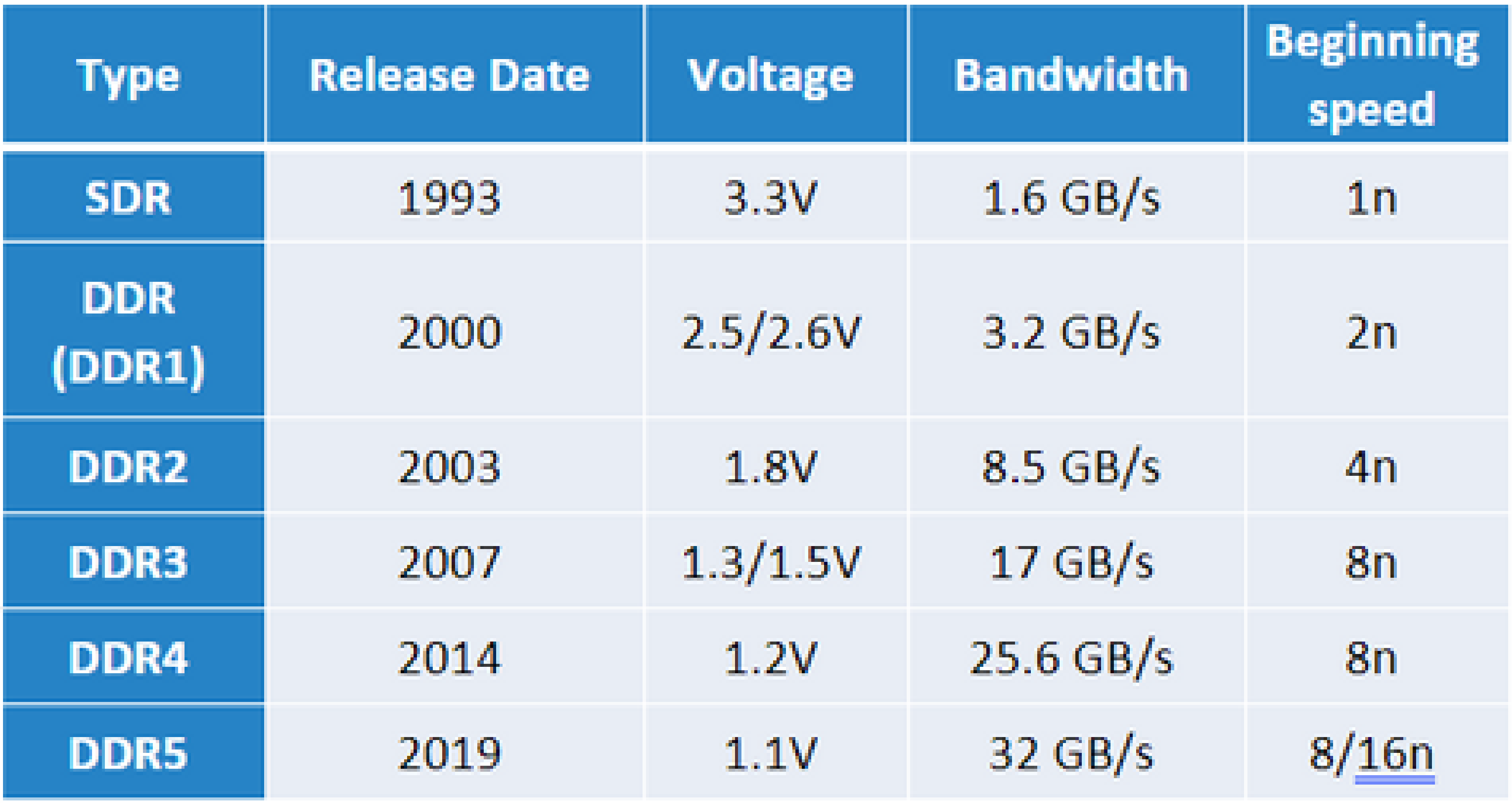
DDR3

DDR Families

- SDR (Single Data Rate) SDRAMs
- DDR
- DDR 2
- DDR 3



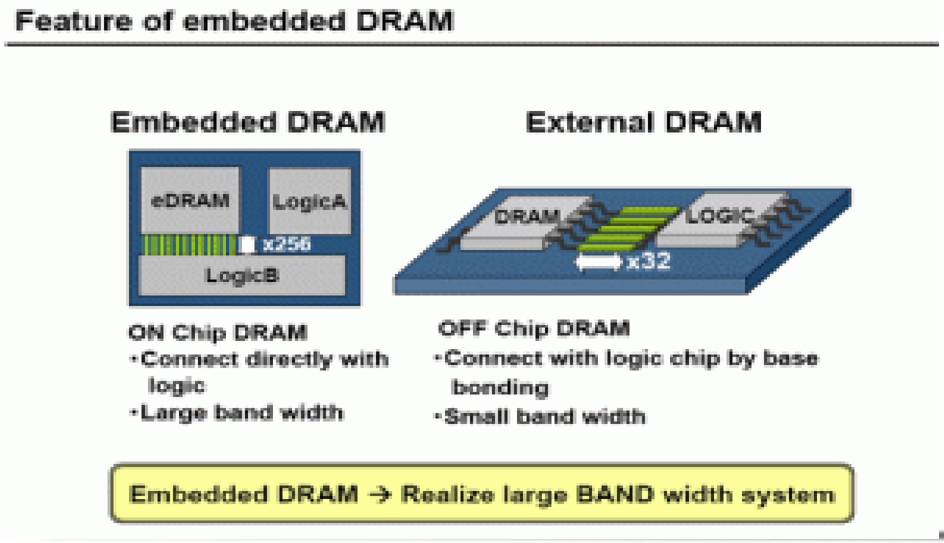
DDR Comparison



Type	Release Date	Voltage	Bandwidth	Beginning speed
SDR	1993	3.3V	1.6 GB/s	1n
DDR (DDR1)	2000	2.5/2.6V	3.2 GB/s	2n
DDR2	2003	1.8V	8.5 GB/s	4n
DDR3	2007	1.3/1.5V	17 GB/s	8n
DDR4	2014	1.2V	25.6 GB/s	8n
DDR5	2019	1.1V	32 GB/s	8/16n

DRAM and eDRAM

- New most popular applications are the next-generation, high-resolution game-consoles.
- High-speed embedded-DRAM (eDRAM) has evolved as a serious contender to embedded-SRAM (eSRAM)
- Good test and repair solutions reduce cost and facilitate higher levels of integration.
- known-good-die (KGD) is a pre-tested die that guarantees the full-spec operation, which is indispensable for cost-effective multi-chip package (MCP) applications.
- GDDR4 standard (graphic double data rate) 4Gb/s/pin

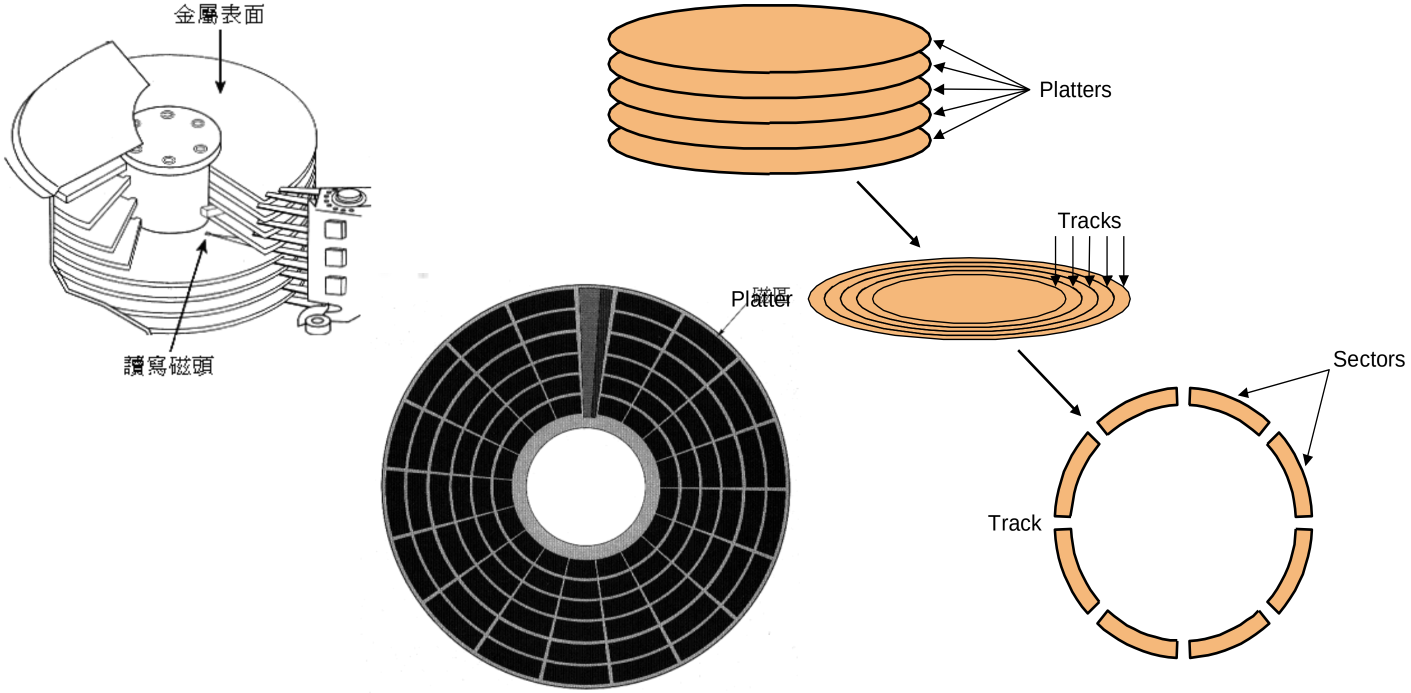

Feature of embedded DRAM

ON Chip DRAM
- Connect directly with logic
- Large band width

OFF Chip DRAM
- Connect with logic chip by base bonding
- Small band width

Embedded DRAM → Realize large BAND width system

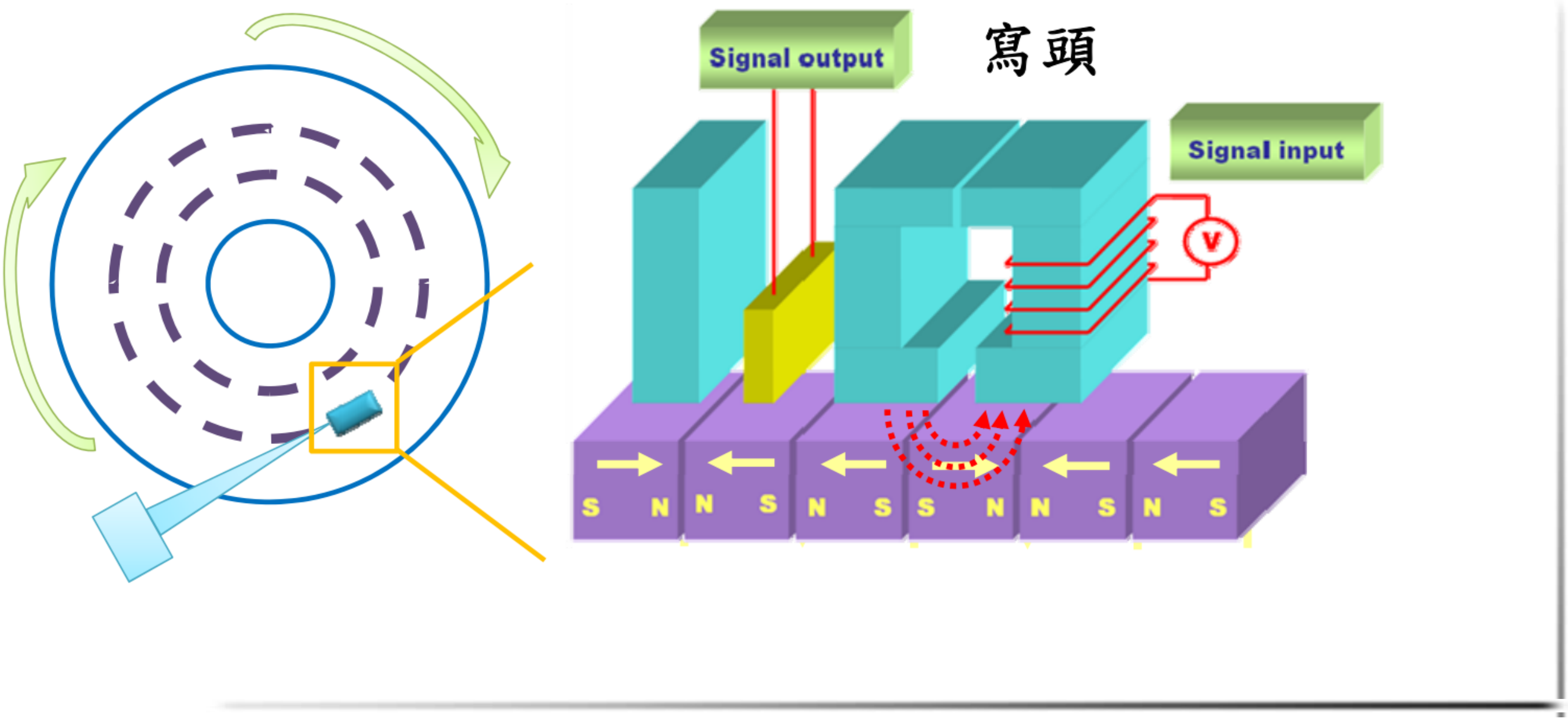
Hard Disk Drive



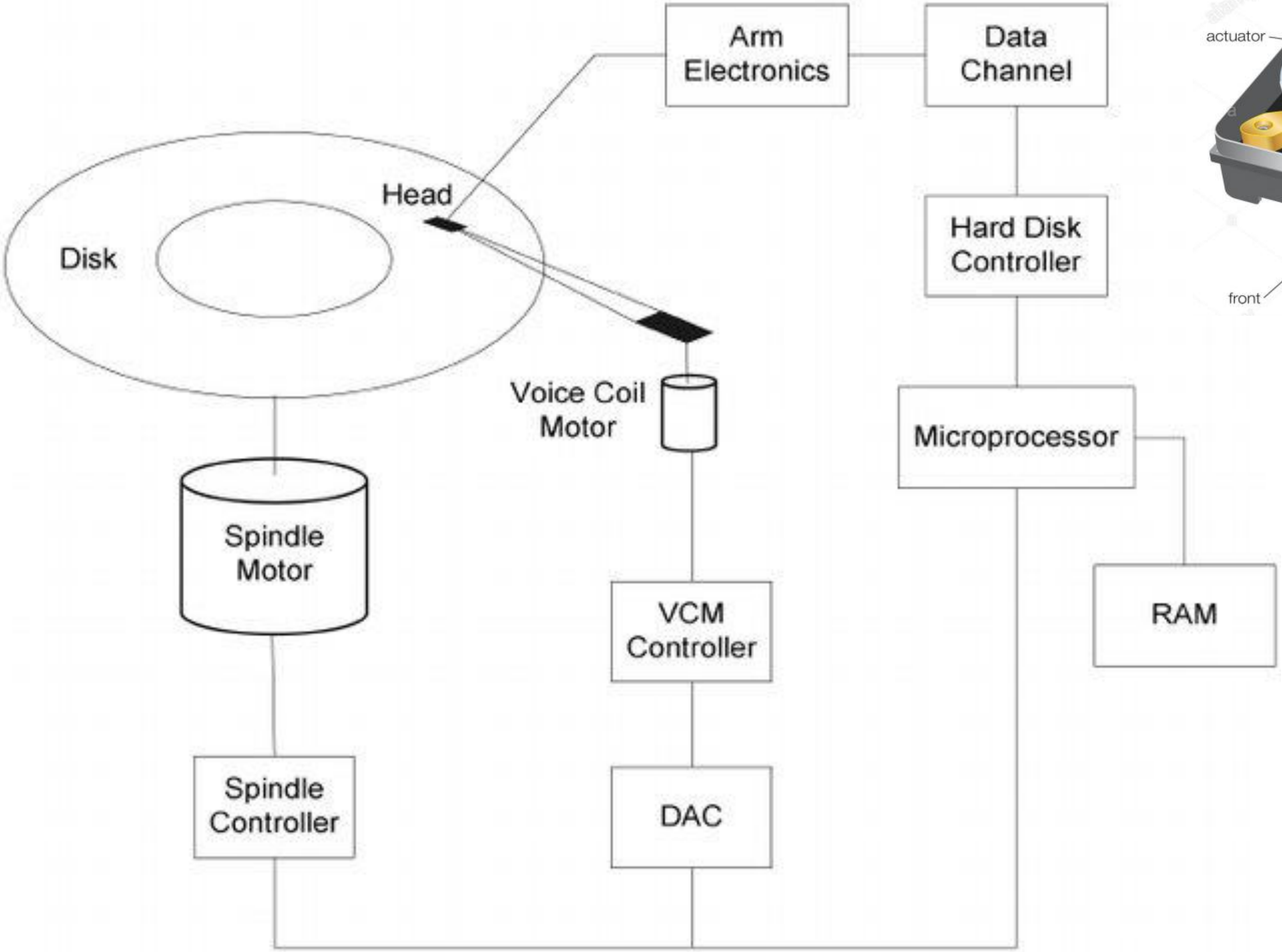
As of July 2010, the highest capacity consumer HDDs are 3 TB. **"Desktop HDDs"** typically store between 120 GB and 2 TB and rotate at 5,400 to 10,000 rpm, and have a media transfer rate of 0.5 Gbit/s or higher. (1 GB = 10^9 bytes; 1 Gbit/s = 10^9 bit/s) Interface -SATA

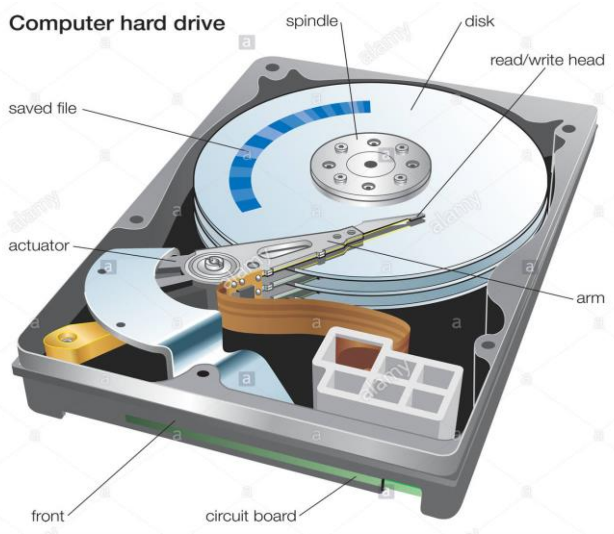
HDD Principle

- A modern HDD records data by magnetizing a thin film of ferromagnetic material on both sides of a disk.
- Sequential changes in the direction of magnetization represent binary data bits

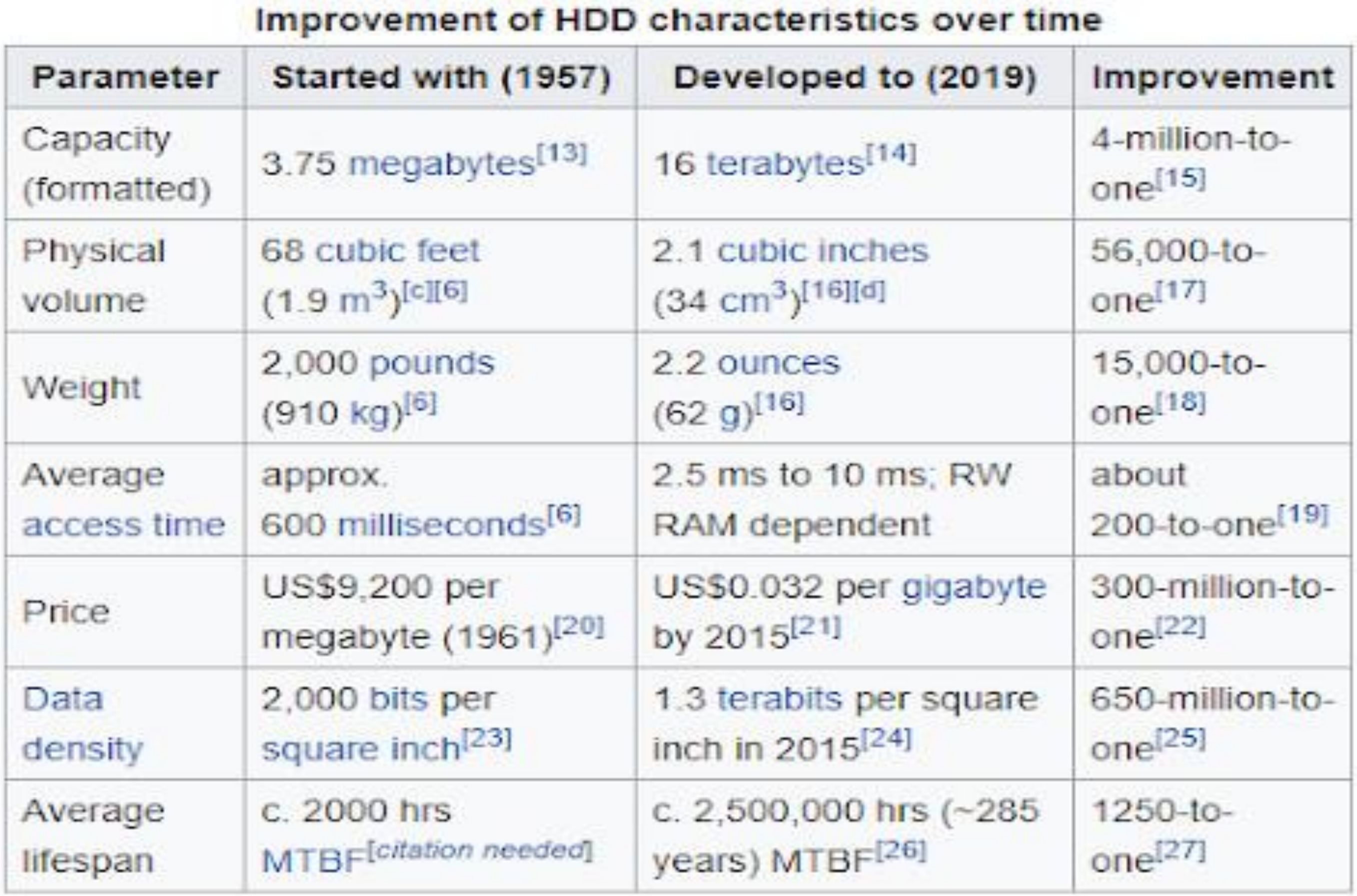


HDD System Diagram





HDD Improvement



Improvement of HDD characteristics over time

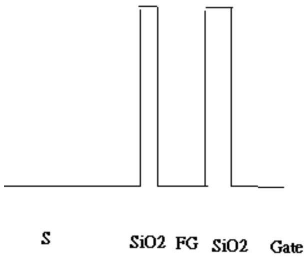
Parameter	Started with (1957)	Developed to (2019)	Improvement
Capacity (formatted)	3.75 megabytes[13]	16 terabytes[14]	4-million-to-one[15]
Physical volume	68 cubic feet (1.9 m^3)[c][6]	2.1 cubic inches (34 cm^3)[16][d]	56,000-to-one[17]
Weight	2,000 pounds (910 kg)[6]	2.2 ounces (62 g)[16]	15,000-to-one[18]
Average access time	approx. 600 milliseconds[6]	2.5 ms to 10 ms; RW RAM dependent	about 200-to-one[19]
Price	US$9,200 per megabyte (1961)[20]	US$0.032 per gigabyte by 2015[21]	300-million-to-one[22]
Data density	2,000 bits per square inch[23]	1.3 terabits per square inch in 2015[24]	650-million-to-one[25]
Average lifespan	c. 2000 hrs MTBF[citation needed]	c. 2,500,000 hrs (~285 years) MTBF[26]	1250-to-one[27]

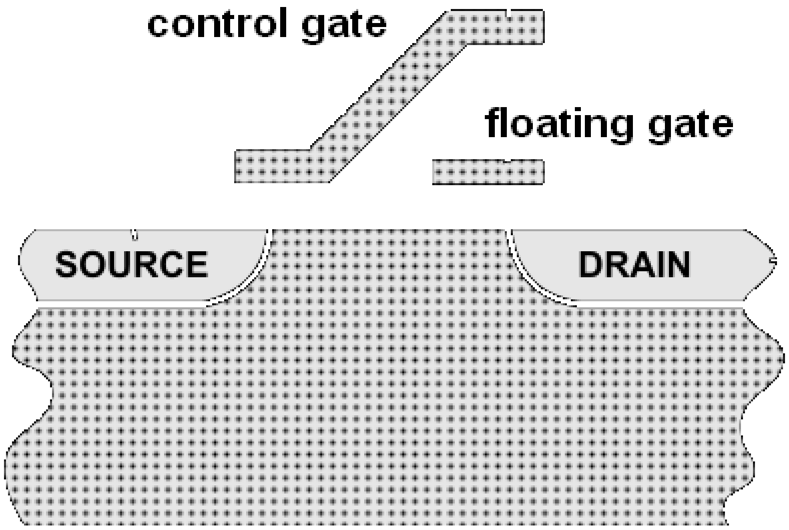
Peripheral storage device I/O

- **Flash**
  - It has the characteristics of small size, high capacity and easy portability.
  - There are the following types:
    - CompactFlash Type I & II（CF）
    - Microdrive（MD）
    - SmartMedia（SM）
    - Memory Stick（MS）
    - MagicGate（MG）
    - MultiMedia Card（MMC）
    - Secure Digital（SD）
    - PC card hard disk
    - ATA flash memory card
    - xD Picture Card（xD）

Operation principle of Flash Memory

- The storage unit is similar to a standard MOSFET
- The difference is that there is another floating gate (FG) covered with a layer of silicon oxide insulator under the control gate (CG).





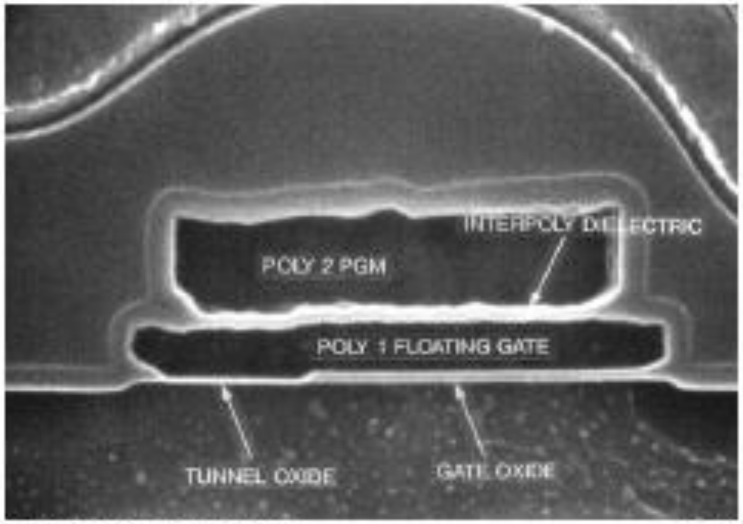


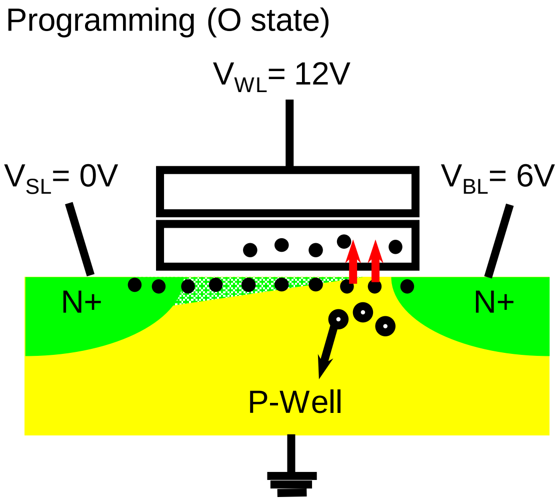
Photo by ICE, "Memory 1997" 22481

Operation principle of Flash Memory

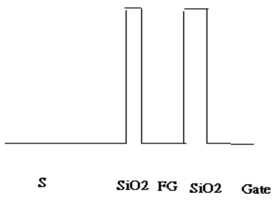
- **There are two ways to let negative electrons in and out of the floating gate**
  - ➔ Channel Hot Electron, CHE
  - ➔ Fowler-Nordheim tunneling, FN

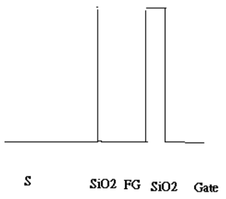
Channel hot electronic programming

- **Channel Hot Electron，CHE**
  - ➔ This method applies a **high voltage to the control gate,** so that the conduction **electrons** break through the barrier of the insulator and **enter the floating gate** under the action of the electric field, and vice versa, to complete the writing or erasing action.

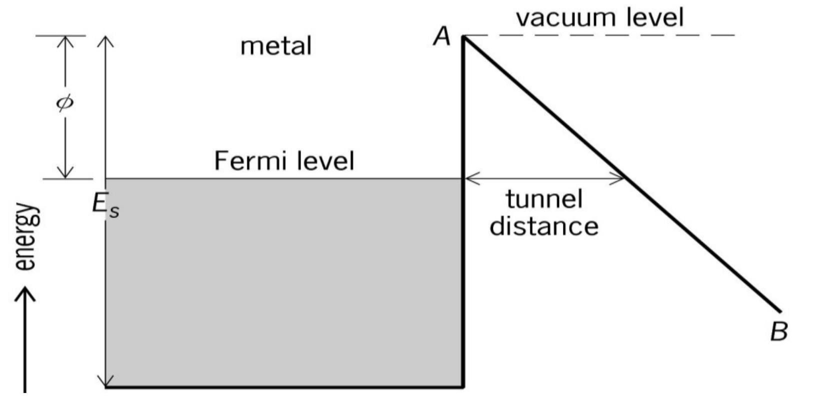


Fowler-Nordheim





- **Fowler-Nordheim(FN)**
  - ➔ It directly applies **high voltage on both sides of the insulating layer** to form a **high-strength electric field** to help electrons enter and exit the **floating gate through the oxide layer** channel.



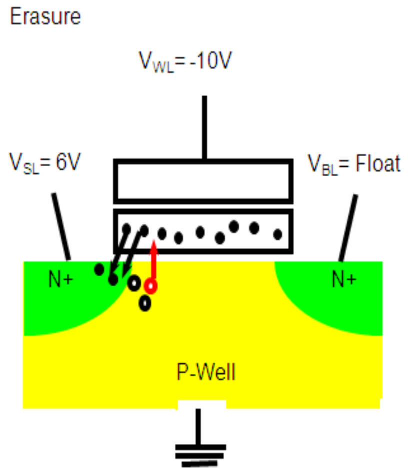


Diagram of the energy-level scheme for field emission from a metal at absolute zero temperature

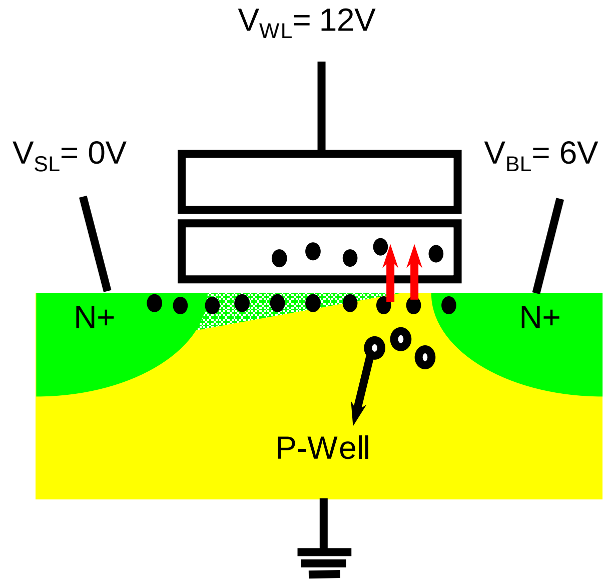
Write and Erase

Program(1 to 0) with CHE (channel hot electron) injection
or Flowler Nordheim (FN) electron tunneling

Erase(0 to 1): with FN (Fowler-Nordheim) tunneling

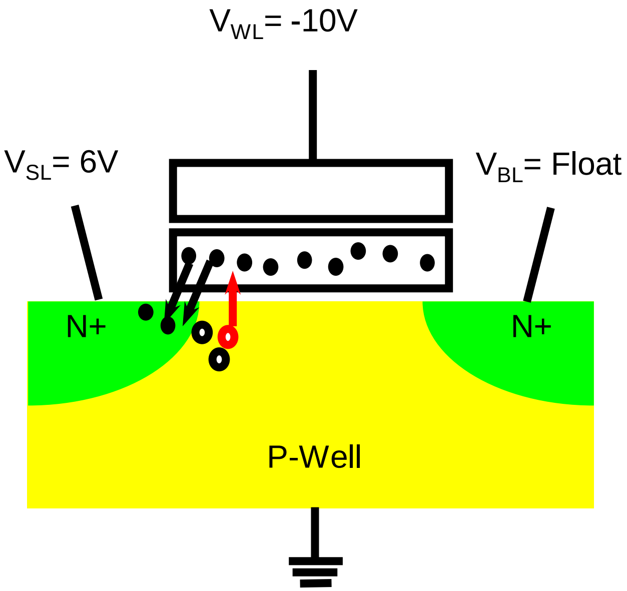
Programming: CHE injection

(0 state) Write 0

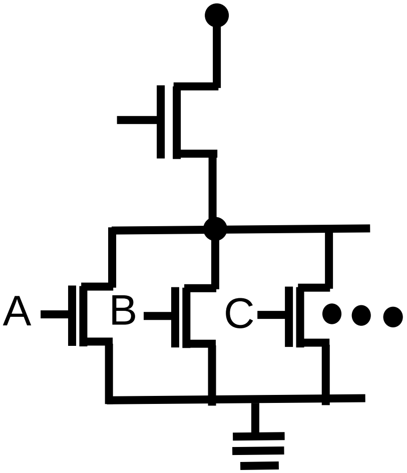


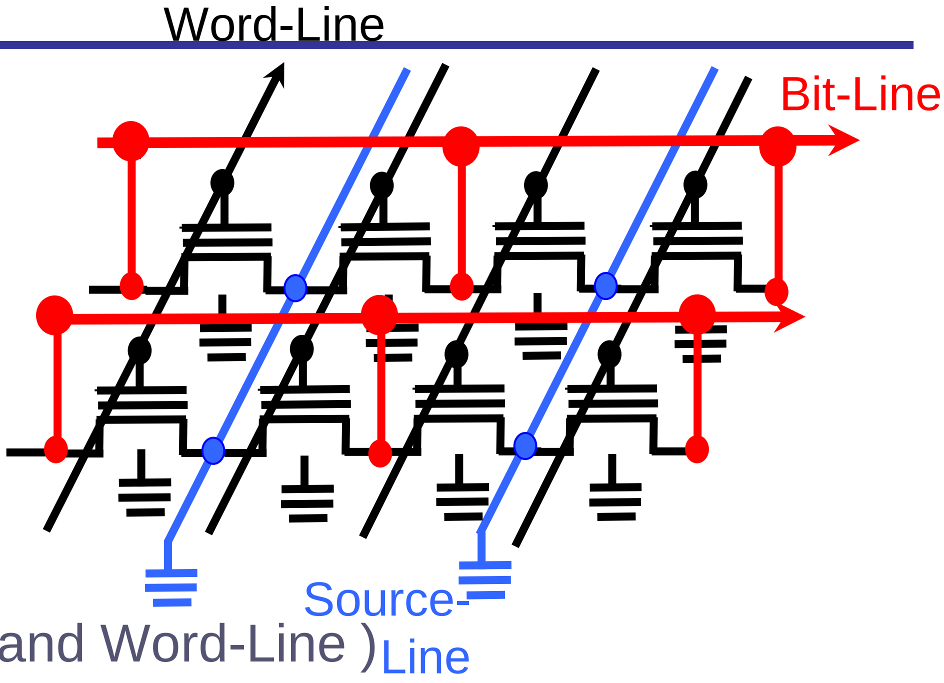
Erasure: FN tunneling

(1 state) Write 1



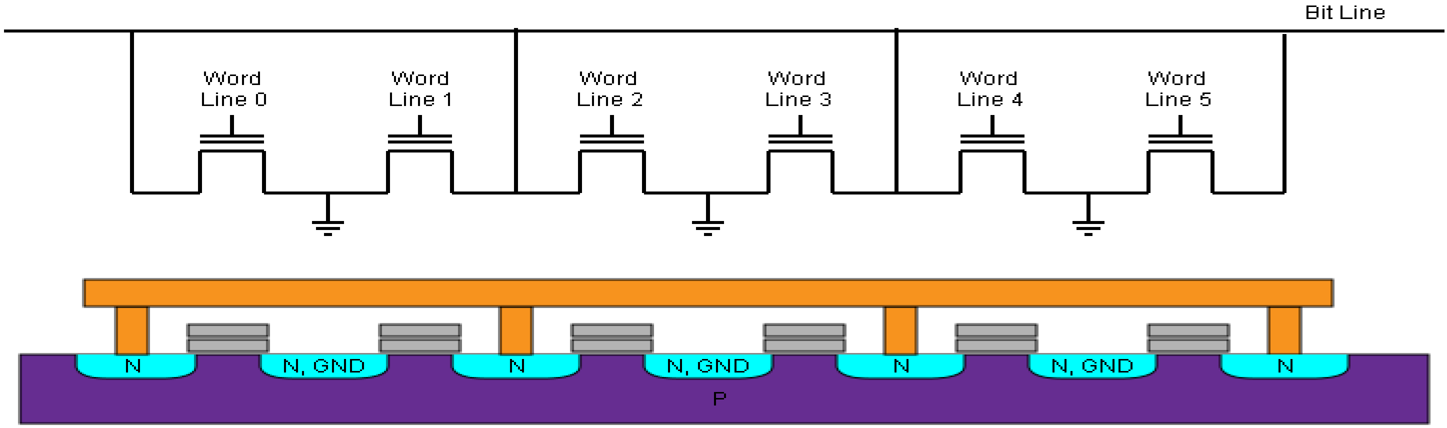
NOR Flash



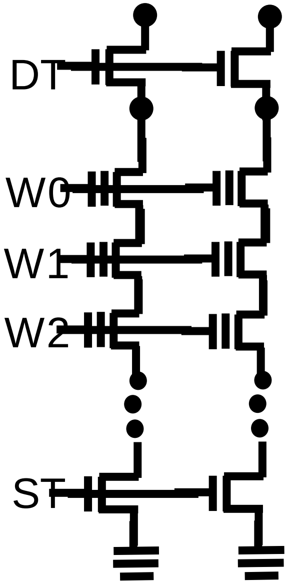


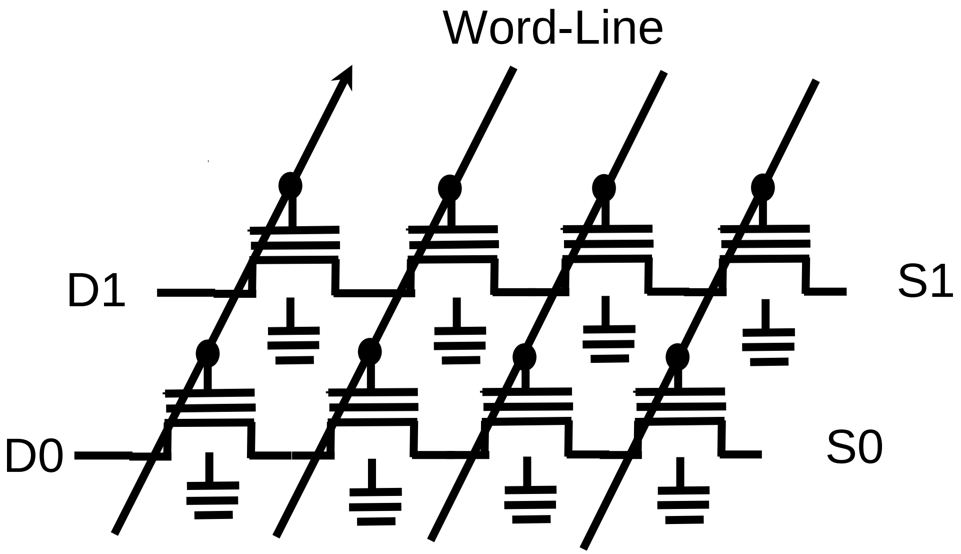
(Bit-Line) ⊥ (Source-Line and Word-Line)

Parallel circuit, read random, program random



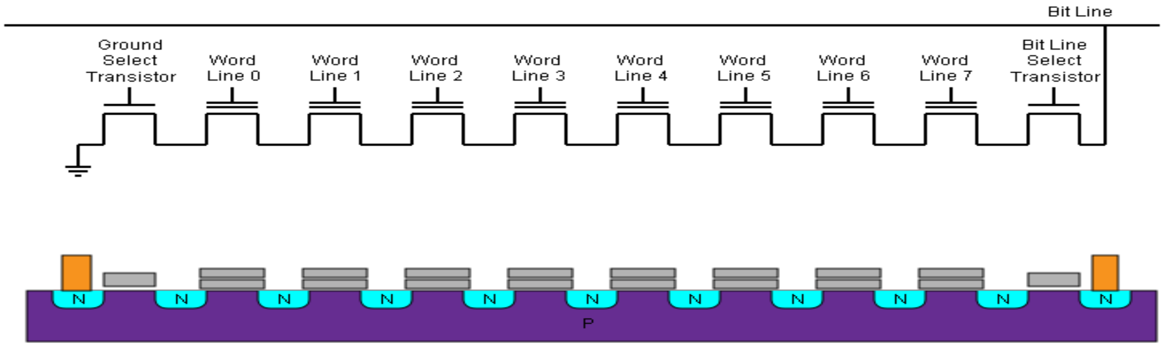
NAND Flash



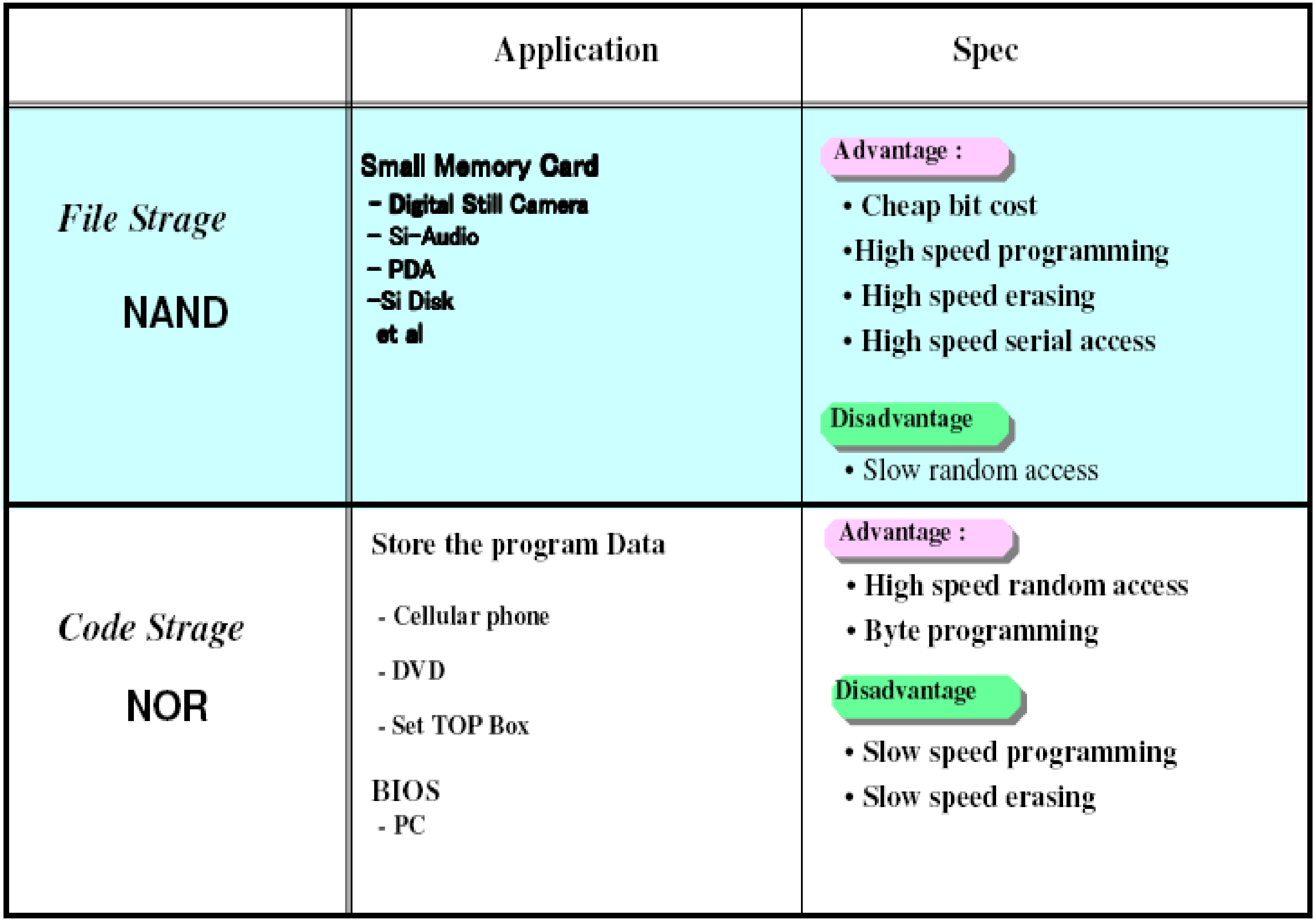


(Word-Line) ⊥ (Source-Line and Drain-Line)

Cell in series and share Drain & Source to each other

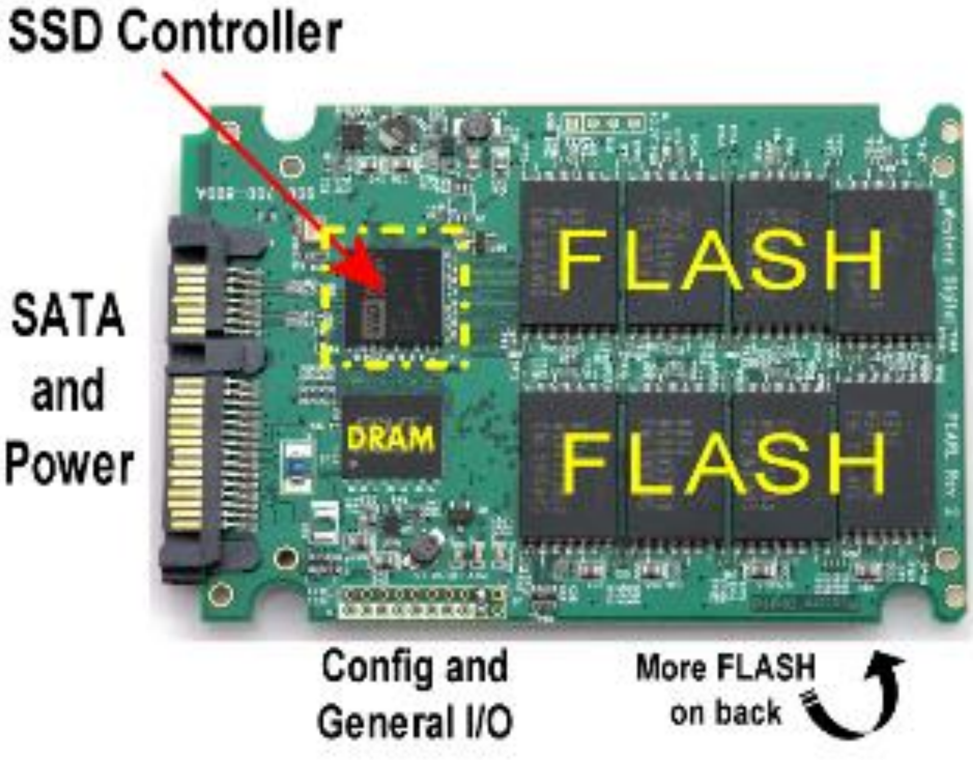


Flash Memory Performance



	Application	Spec
File Strage NAND	**Small Memory Card** **– Digital Still Camera** **– Si-Audio** **– PDA** **–Si Disk** **et al**	Advantage : • Cheap bit cost •High speed programming • High speed erasing • High speed serial access Disadvantage • Slow random access
Code Strage NOR	Store the program Data - Cellular phone - DVD - Set TOP Box BIOS - PC	Advantage : • High speed random access • Byte programming Disadvantage • Slow speed programming • Slow speed erasing

SSD System

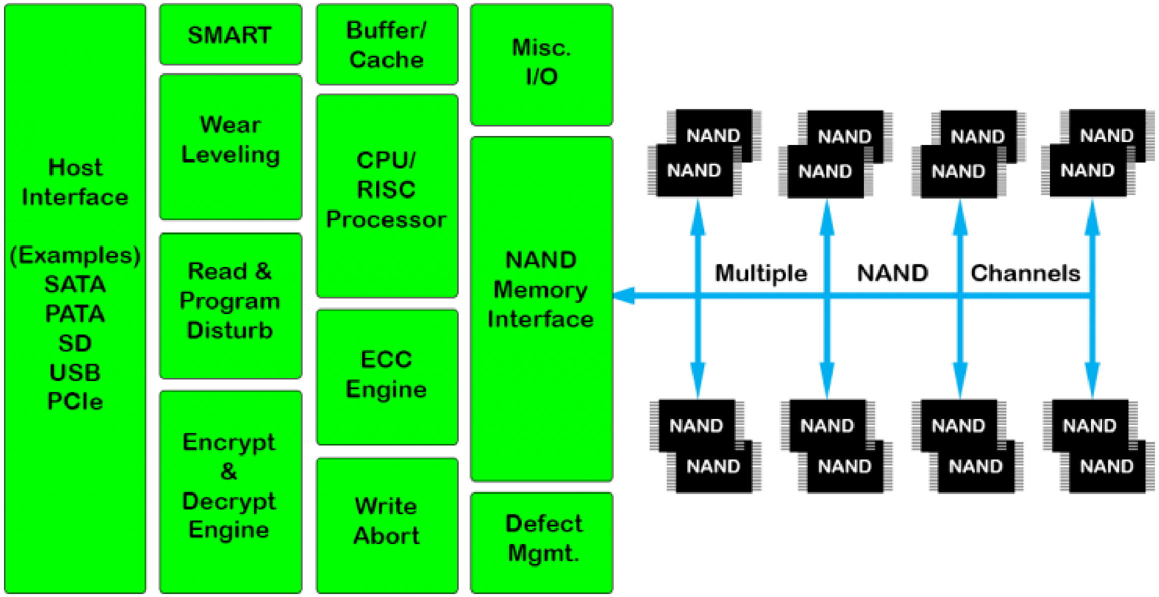


SATA interface flash SSD

Controller Function Diagram

- Important function
  - SMART (Self-Monitoring, Analysis and Reporting Technology):
  - ECC (error correction code) Engine
  - Write Abort
  - Read disturb
  - Wear-leveling

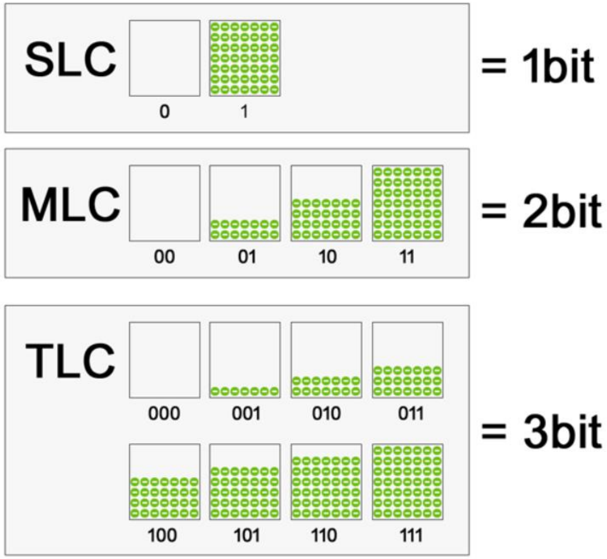
Typical Controller Functions and Blocks

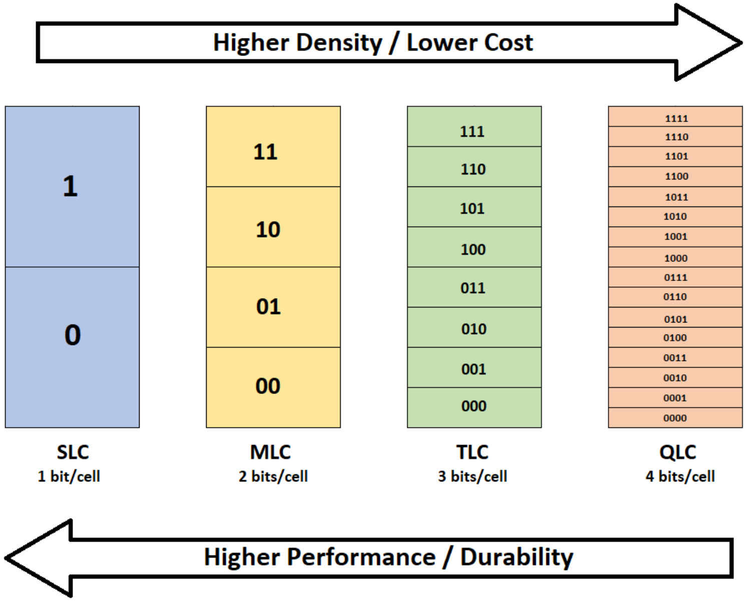


Generic Solid State Drive (SSD) Controller Architecture

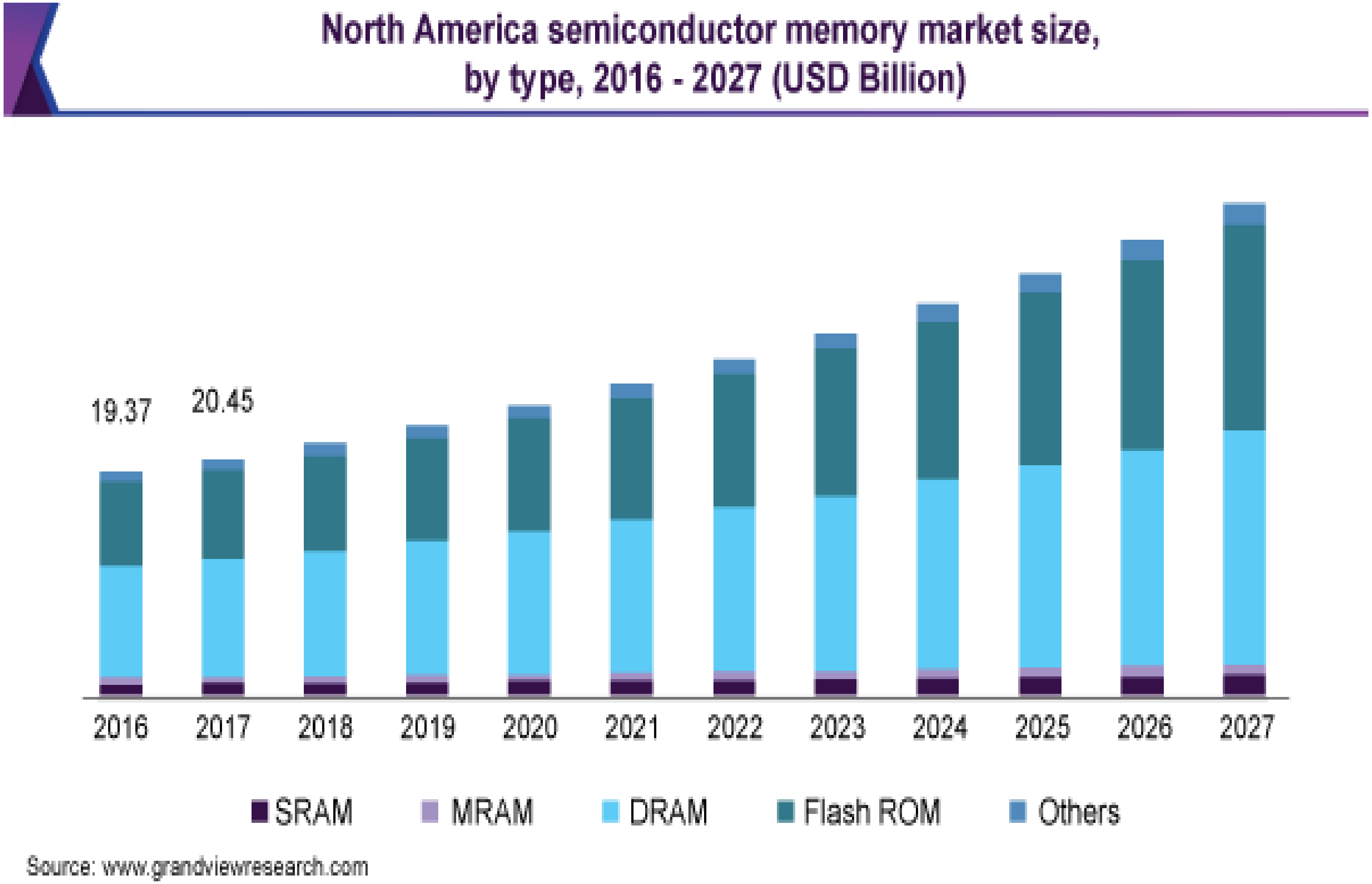
Multi-level cell

- In electronics, a **multi-level cell** (**MLC**) is a memory cell/element capable of storing more than a single bit of information
- The primary benefit of MLC flash memory is its lower cost per unit of storage due to the higher data density, and memory-reading software can compensate for a larger bit error rate The higher error rate necessitates an error correcting code(ECC)



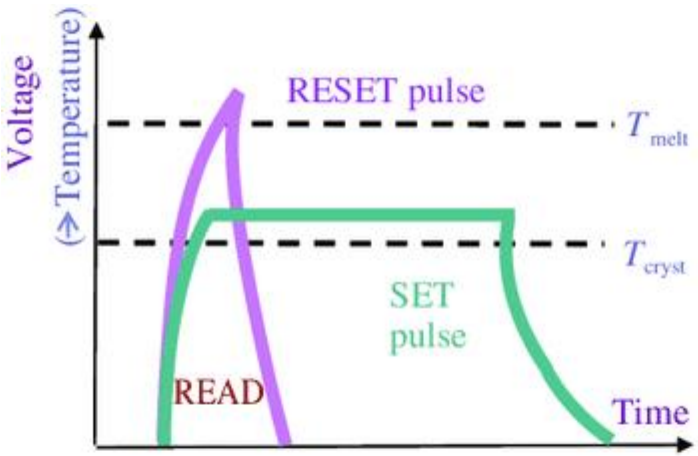


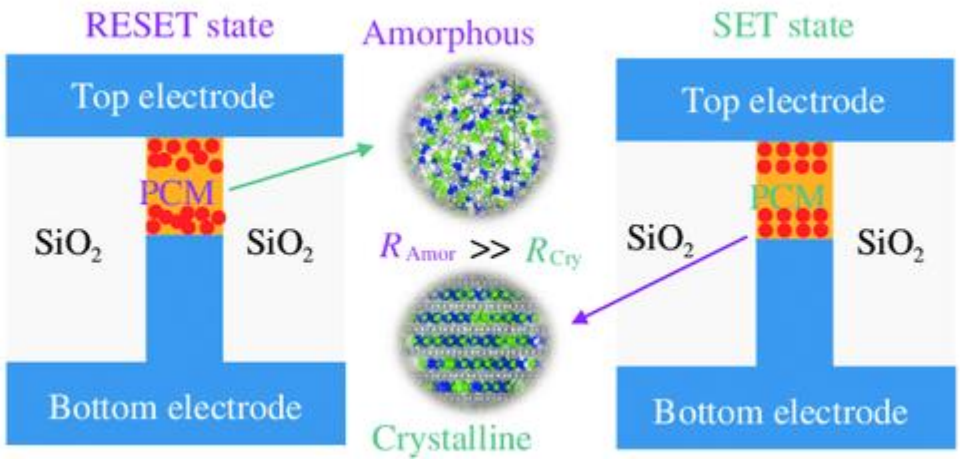
Semiconductor Memory Market Size



Phase Change Memory(PCM)

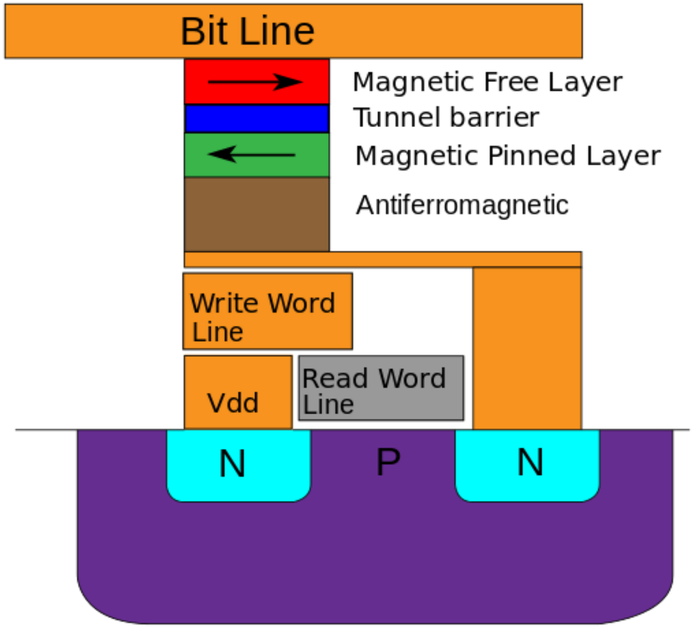
- Phase change material: $Ge_2Sb_2Te_5$(GST)
- Using temperature or laser to switch between amorphous and crystalline state.
  - ➔ Amorphous state (RESET) → high resistance → “0”
  - ➔ Crystal state (SET) → low resistance → “1”

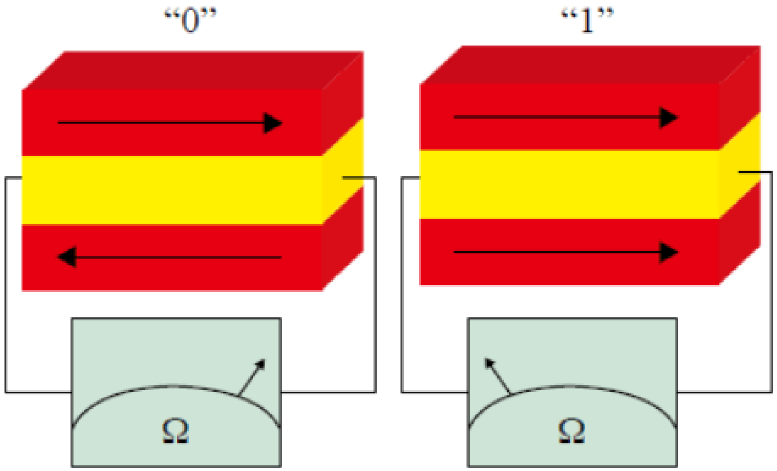




Magnetic RAM(MRAM)

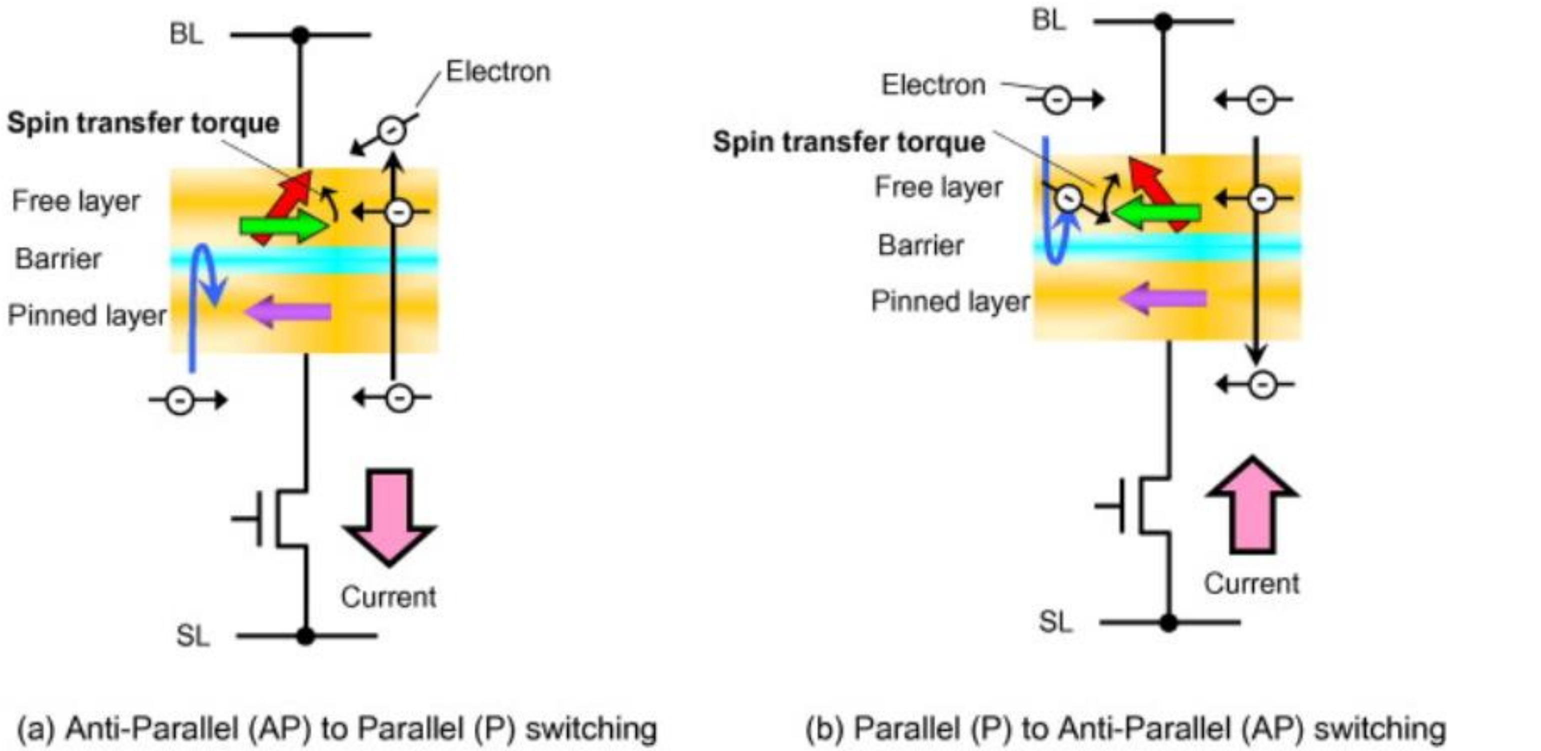
- **Using magnetoresistance changes caused by the different magnetization directions to store data.**
  - ➔ Parallel→ low resistance → “1”
  - ➔ Antiparallel→ high resistance → “0”





Spin-transfer torque MRAM(STT-MRAM)

- **Use the spin polarized current to change the magnetic orientation of the information storage layer in the MTJ element.**



(a) Anti-Parallel (AP) to Parallel (P) switching

(b) Parallel (P) to Anti-Parallel (AP) switching

Memory Comparison

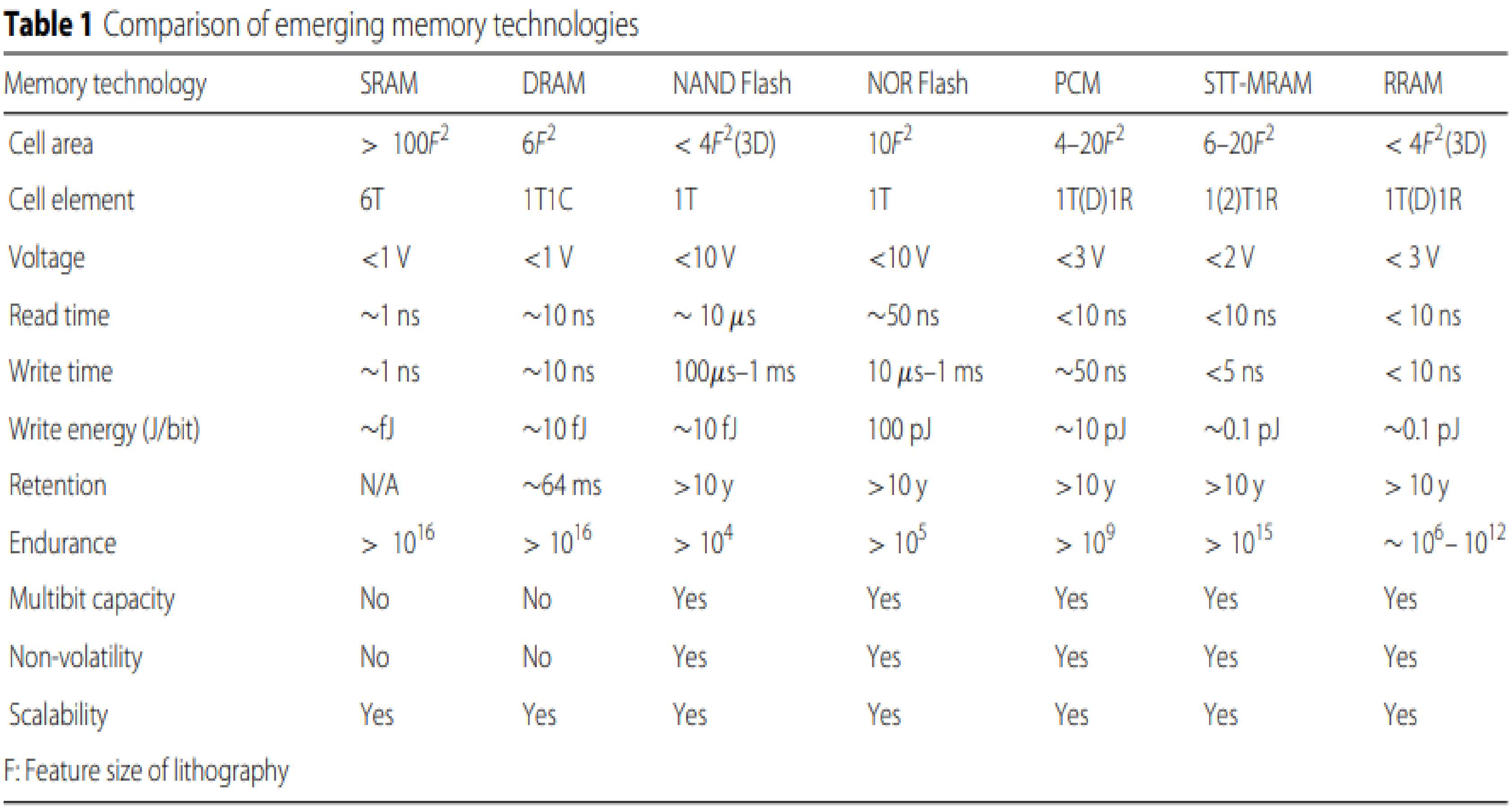


Table 1 Comparison of emerging memory technologies

Memory technology	SRAM	DRAM	NAND Flash	NOR Flash	PCM	STT-MRAM	RRAM
Cell area	$> 100F^2$	$6F^2$	$< 4F^2$(3D)	$10F^2$	$4–20F^2$	$6–20F^2$	$< 4F^2$(3D)
Cell element	6T	1T1C	1T	1T	1T(D)1R	1(2)T1R	1T(D)1R
Voltage	<1 V	<1 V	<10 V	<10 V	<3 V	<2 V	< 3 V
Read time	~1 ns	~10 ns	~ 10 μs	~50 ns	<10 ns	<10 ns	< 10 ns
Write time	~1 ns	~10 ns	100μs–1 ms	10 μs–1 ms	~50 ns	<5 ns	< 10 ns
Write energy (J/bit)	~fJ	~10 fJ	~10 fJ	100 pJ	~10 pJ	~0.1 pJ	~0.1 pJ
Retention	N/A	~64 ms	>10 y	>10 y	>10 y	>10 y	> 10 y
Endurance	$> 10^{16}$	$> 10^{16}$	$> 10^4$	$> 10^5$	$> 10^9$	$> 10^{15}$	$\sim 10^6 – 10^{12}$
Multibit capacity	No	No	Yes	Yes	Yes	Yes	Yes
Non-volatility	No	No	Yes	Yes	Yes	Yes	Yes
Scalability	Yes	Yes	Yes	Yes	Yes	Yes	Yes

F: Feature size of lithography

References

- “Semiconductor Memory” class lecture, by Ya-King Chin
- “Flash Memory” class lecture, by Chron-Jung Lin
- “Advanced Nonvolatile Memory” class lecture, by Riichiron Shirota
- http://stock.yam.com/rsh/article.php/326937
- http://www.toshiba.co.jp/index.htm
- SoC設計探索善用儲存架構特性 提昇快閃記憶體系統效能by郭大維/吳晉賢 http://203.66.123.22/ne/magazine/magazine_article.asp?Id=634
- http://esslab.tw/wiki/index.php/NFTL