Cx1106
Computer Organization and Architecture

Computer Memory Introduction

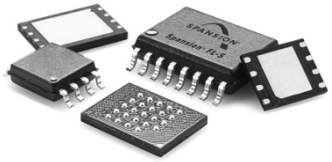
Oh Hong Lye
Lecturer
School of Computer Science and Engineering, Nanyang Technological University.
Email: hloh@ntu.edu.sg

Computer Memory

Semiconductor-based

Solid-State Drive

SODIMM

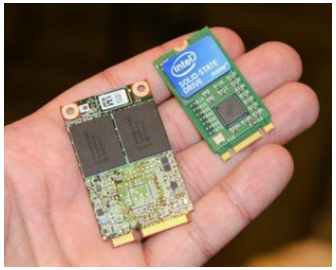


Memory IC Chips

CF, SD, miniSD, microSD, MMC

Solid-State Drive

Solid-State Drive

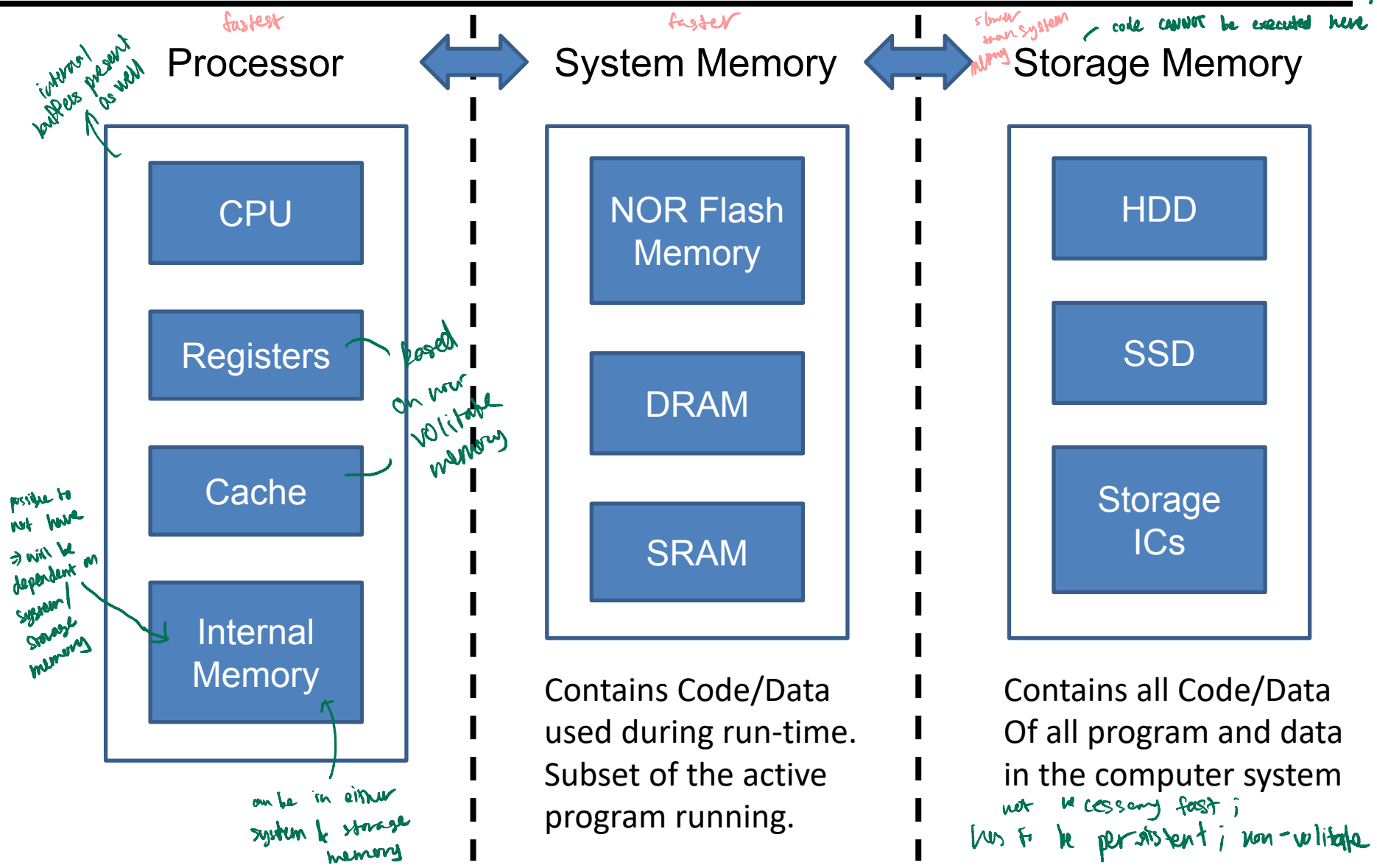


Thumb Drive

Magnetic-based

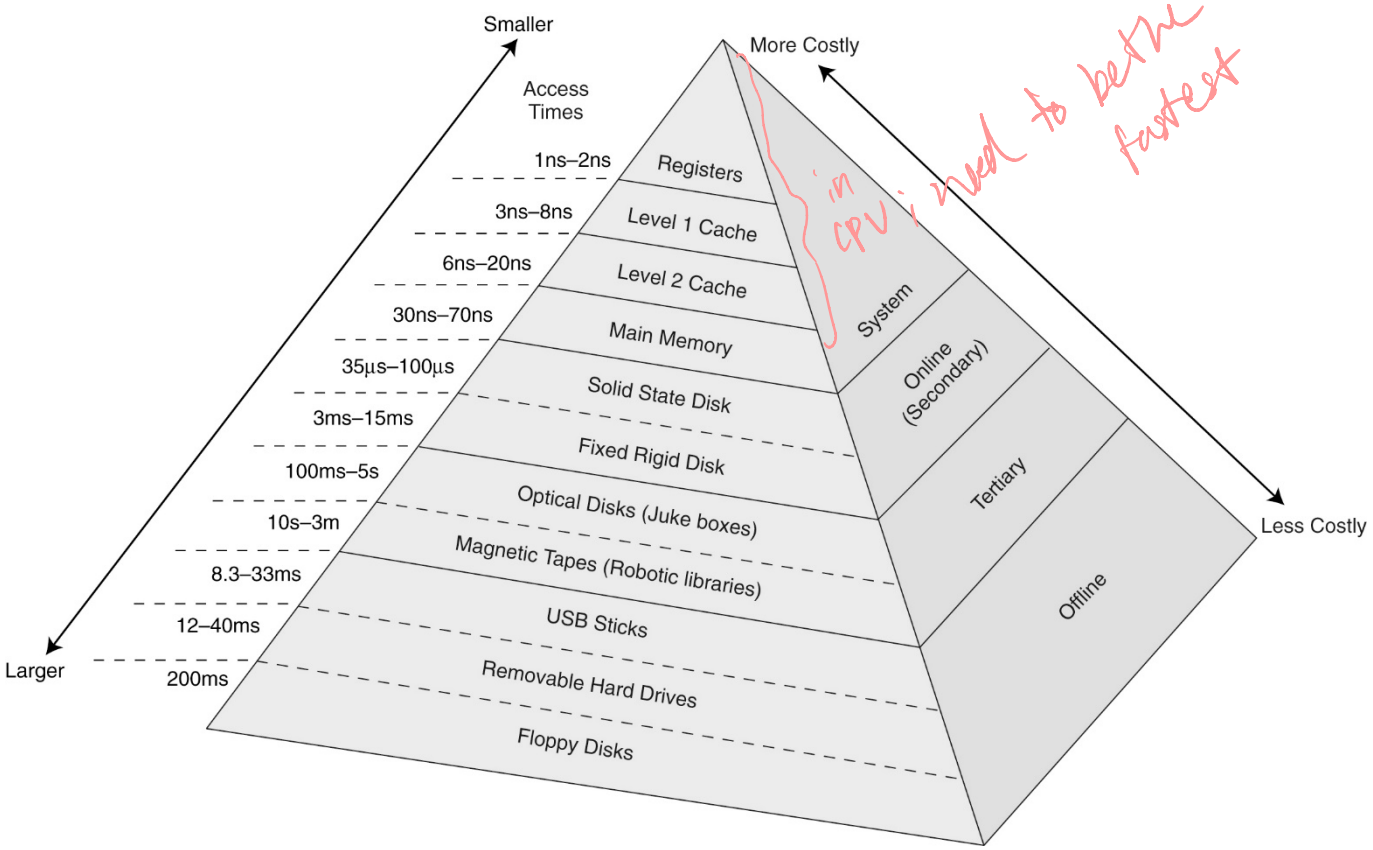
Magnetic HD

Computer Memory (Functional View)



Memory – Cost vs Function Trade Off

- The memory and storage devices may be organized like a pyramid.
- The pinnacle has the fastest access time, but is also more costly.
- Memory Access Time below are only for illustration. These changes with improvement in technology.



Volatile and Non-Volatile Memory

- Volatile
  - Data is lost when electric power is removed.
  - Temporary storage.
  - Typically used as system memory.
  - We will look at Random Access Memories such as Static-RAM (SRAM) and Dynamic-RAM (DRAM).

- Non-volatile
  - Data is retained even if electric power is removed.
  - Permanent storage.
  - Typically used as main storage.
  - We will look at FLASH, magnetic hard-disk specifically.

CE1006/CZ1006
Computer Organization and Architecture

Semiconductor Memories

Oh Hong Lye
Lecturer
School of Computer Science and Engineering, Nanyang Technological University.
Email: hloh@ntu.edu.sg

Semiconductor Memories

- Memories based on semiconductor integrated circuits (IC).
- Used as processor internal memories and system memory in a computer system
- Processor internal memories
  - Registers
  - Buffers
  - Cache
  - Internal System and Storage Memory (SRAM, DRAM, Flash based)
- Types of memory
  - SRAM
  - DRAM
  - Flash Memory
  - Solid State Drives (based on Flash Memories)

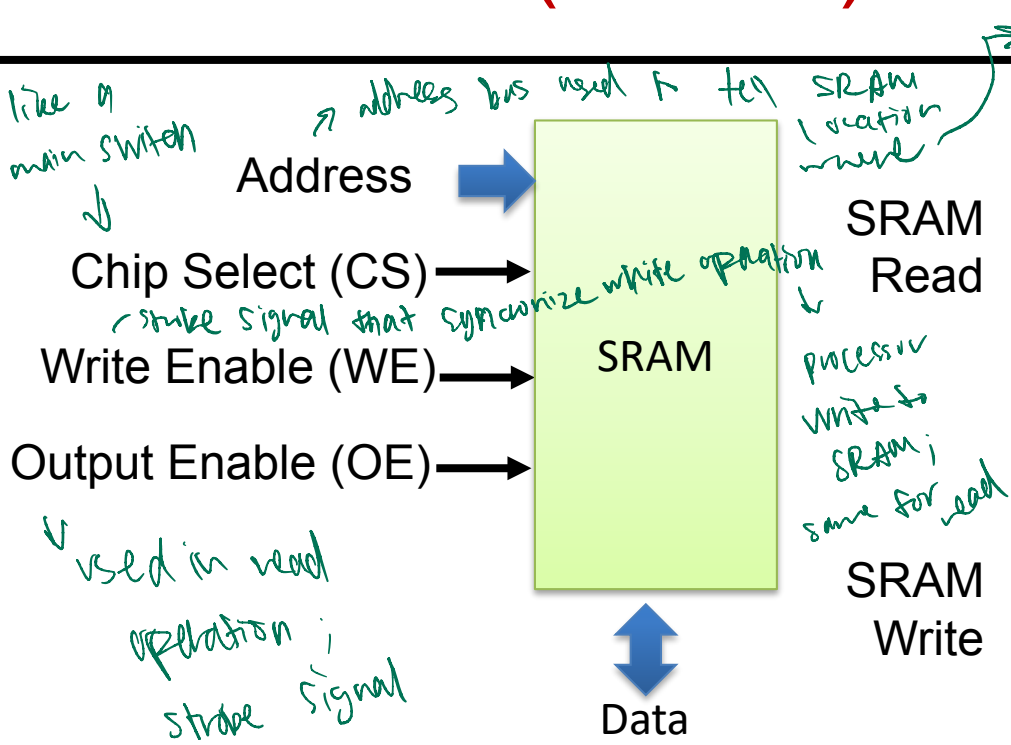
VOLATILE MEMORY

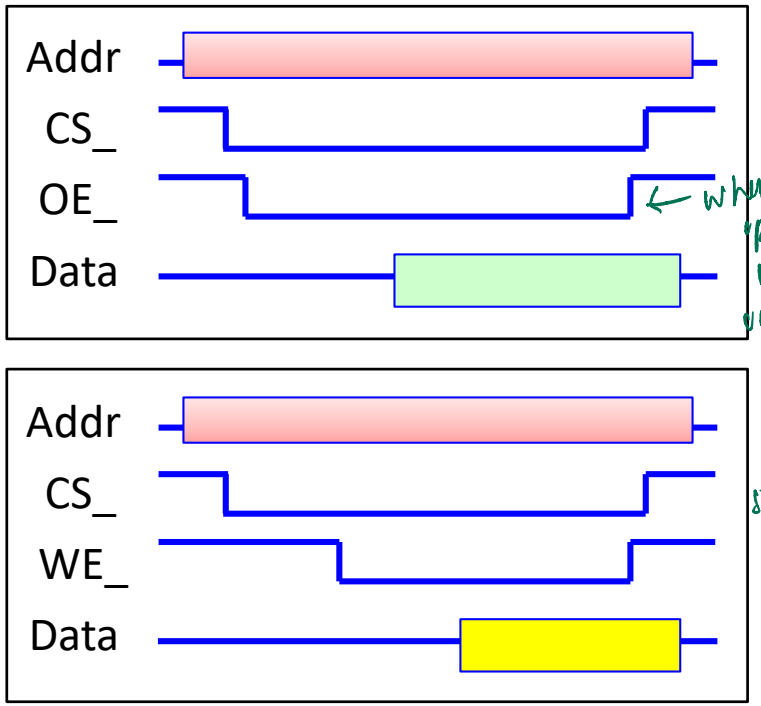
Random Access Memory (RAM)

- Static RAM (SRAM)
  - Static Random Access Memory
  - Data stored as long as supply is applied.
  - Large (4 to 6 transistors per cell).
  - Fast.
  - Low power consumption (active and standby)

- Dynamic Random Access Memory (DRAM)
  - Periodic refresh required.
  - Small (1 to 3 transistors per cell).
  - Slower.
  - Higher power consumption due to need for periodic refresh operation to maintain data integrity of memory cells.

Static RAM (SRAM) Access





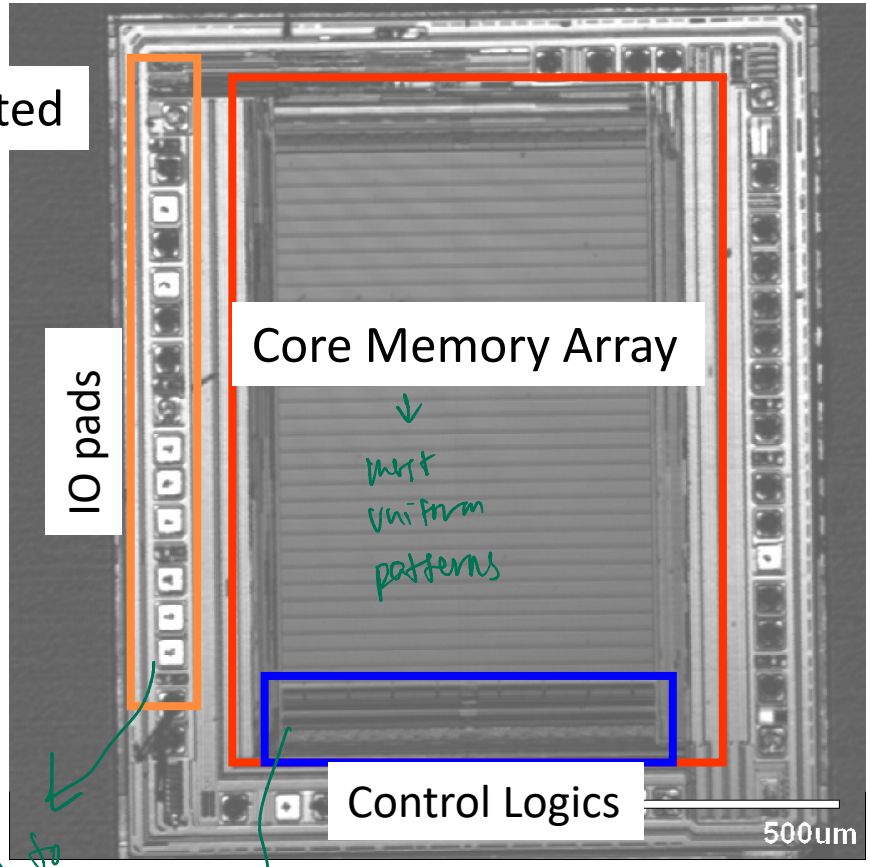
Addr	Specifies Address of memory location.	
Data	Data to be read/written.	Typically 8,16 or 32 bits.
CS	Control signals	Chip Select (Enables or disables the chip).
WE		Write Enable (Allows data to to be written)
OE		Output Enable (Allows SRAM to output data)

SRAM Chip DeCap

Commercial SRAM Chip
Cypress CY62128, 1 Mbit

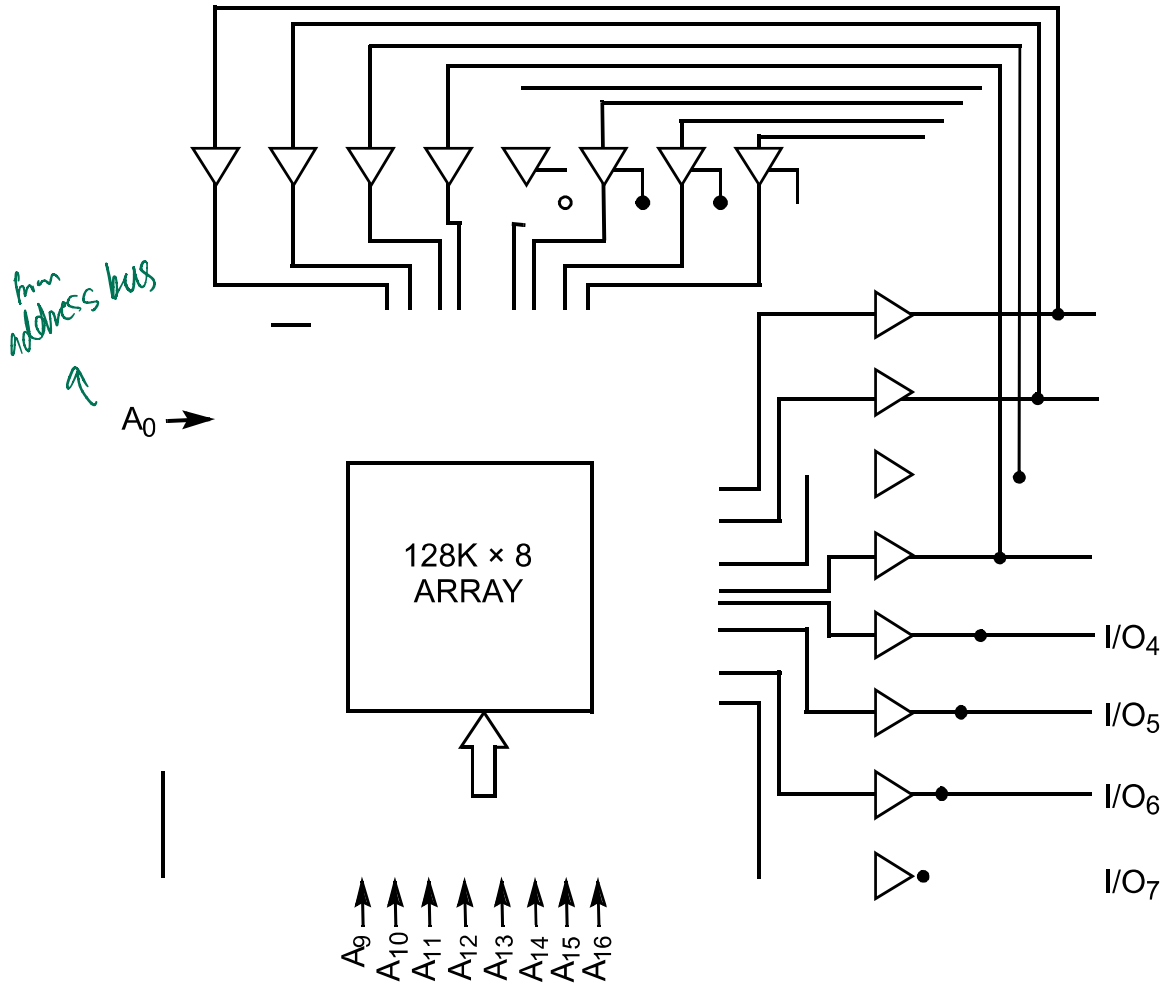
Decapsulated

Decapsulated silicon die
Laser scanned image, 5x magnification



- **Memory** array takes up **most** of the **real estate**.
- Leading Edge Semiconductor Process is usually first tested with Memory Design.

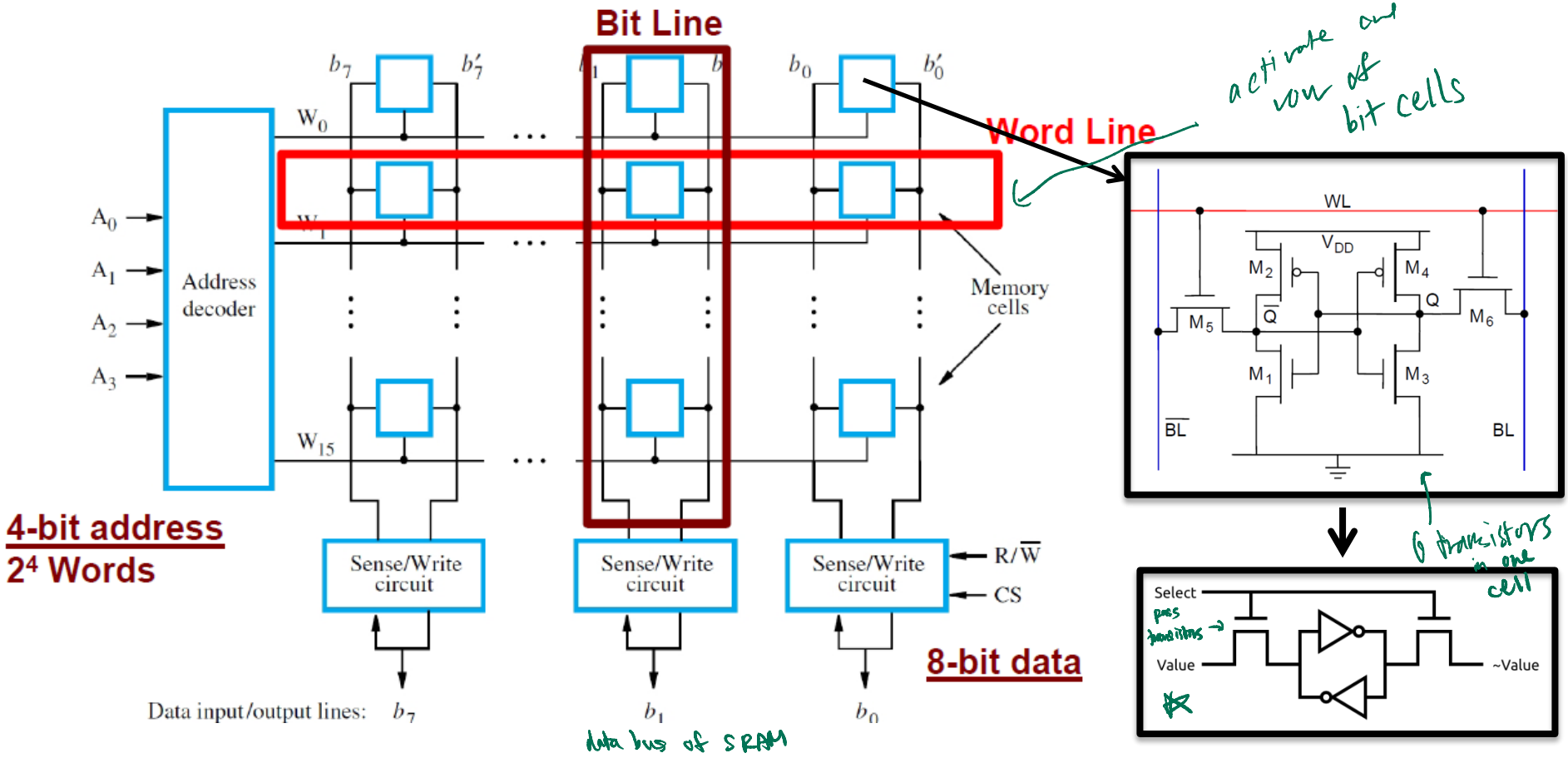
SRAM Internal Circuitry



- Memory array
- 1M SRAM cells for the chip shown.
- Control circuitry
- Decoders
- Sense amplifiers
- Input/Output multiplexers

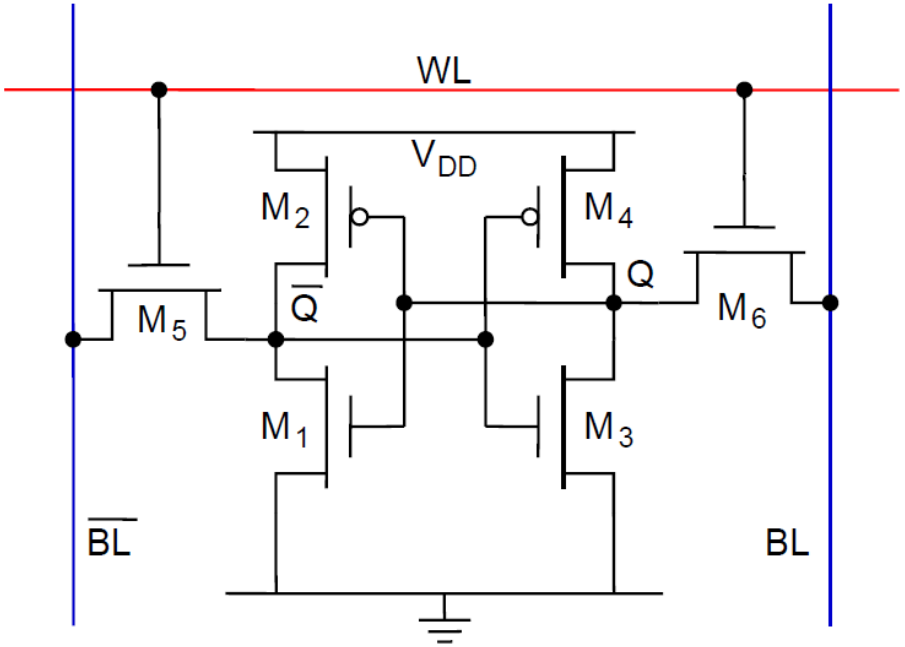
SRAM Cell

- Memory cells are organized in arrays (rows), and are accessed via the word lines and bit lines.
- Each individual SRAM Bit consist of 6 Transistors (typical design).



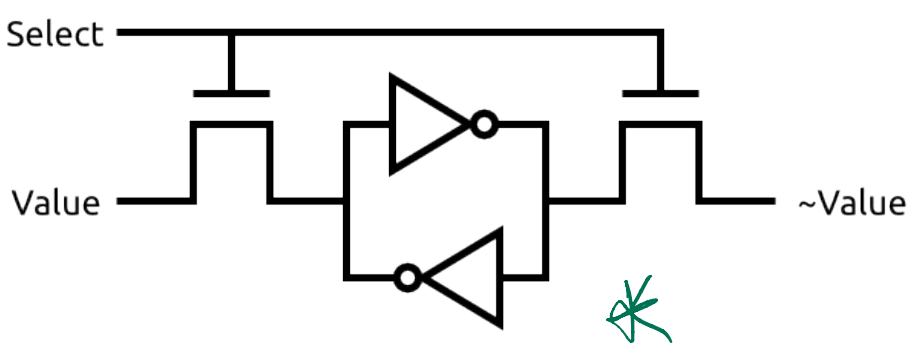
SRAM Cell

M1 - M4 : Storage purposes

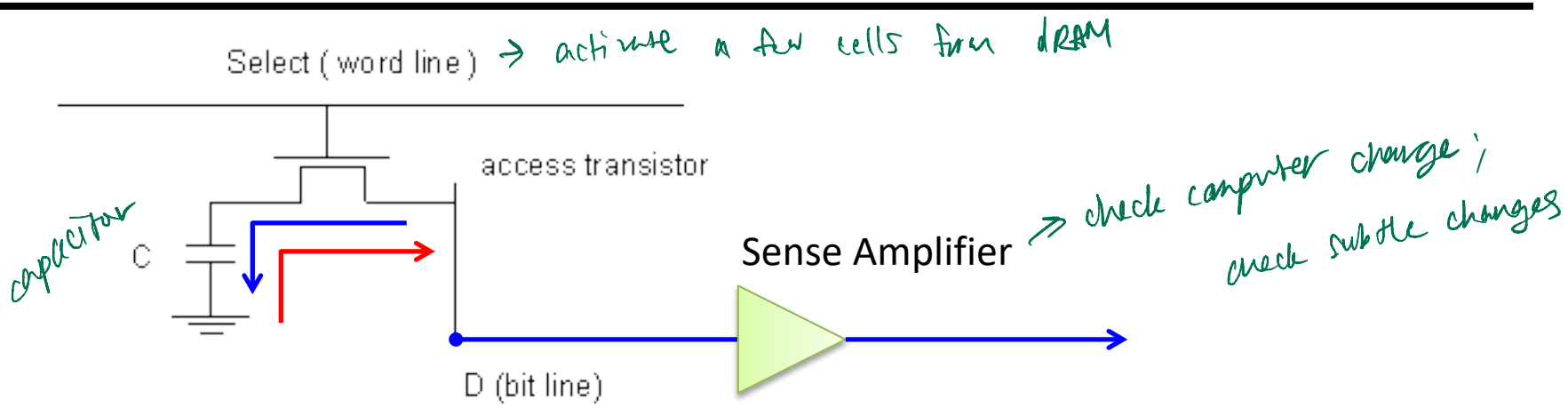


- There are six transistors
- M1, M2, M3, M4, M5 and M6
- Word line (WL) is derived from Address Decoder Output. It controls the Read/Write process
- The actual data are placed on the differential bit lines (BL and BLB). This is connected to the Data Bus.

- The data bit is stored in M1, M2, M3 and M4 (which is equivalent to two invertors connected as shown on the right).
- M5 and M6 are pass transistors.



Dynamic RAM (DRAM)



- Single Transistor Design
- DRAM uses a capacitor as its storage element.
- The Transistor is used to control charges flowing in and out of the capacitor during the Read and Write process.
- Write Process
  - To store a Logic '1': Enable the Transistor, transfer charge into capacitor.
  - To store a Logic '0': Enable the Transistor, discharge the capacitor.
- Read Process
  - Enable the Transistor. Measure the capacitor charge using a sense amplifier.

DRAM – Maintaining Data Integrity

DRAM is the defacto memory choice where system memory is concerned, e.g. laptop; offers very good trade off between cost & performance

- DRAM Read Process destroys information stored on capacitor
- The process of measuring charges on a capacitor also effectively discharged it i.e. data is destroyed.
- Hence, the original data has to be re-written back after every read.
- Periodic refresh is needed as the stored charge “leaks” with time.
- The basic DRAM is more or less obsolete in the market today. It is replaced by its synchronous version called Synchronous DRAM (SDRAM).
- Difference between SDRAM and DRAM is that the former make use of a clock signal from the host to synchronize data transfer, enabling faster transfer rate. SDRAM also has a pipeline architecture that allow faster, overlapping operations.
- Other enhancements of SDRAM includes its double date rate versions DDR, DDR2, DDR3 SDRAM, which could reach transfer speed of more than 2G transfers per second.

Focus:

flash memory

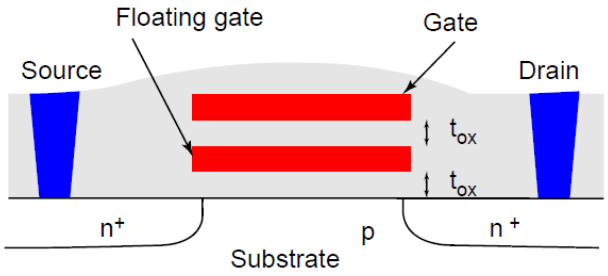
solid state drive – based on flash memory

NON-VOLATILE MEMORY

EPROM and EEPROM

- EPROM (Erasable Programmable ROM)
  - Earliest floating gate transistors are implemented as Erasable Programmable ROM (EPROM) devices.
  - Need to put device under ultra-violet (UV) light to erase the stored program.

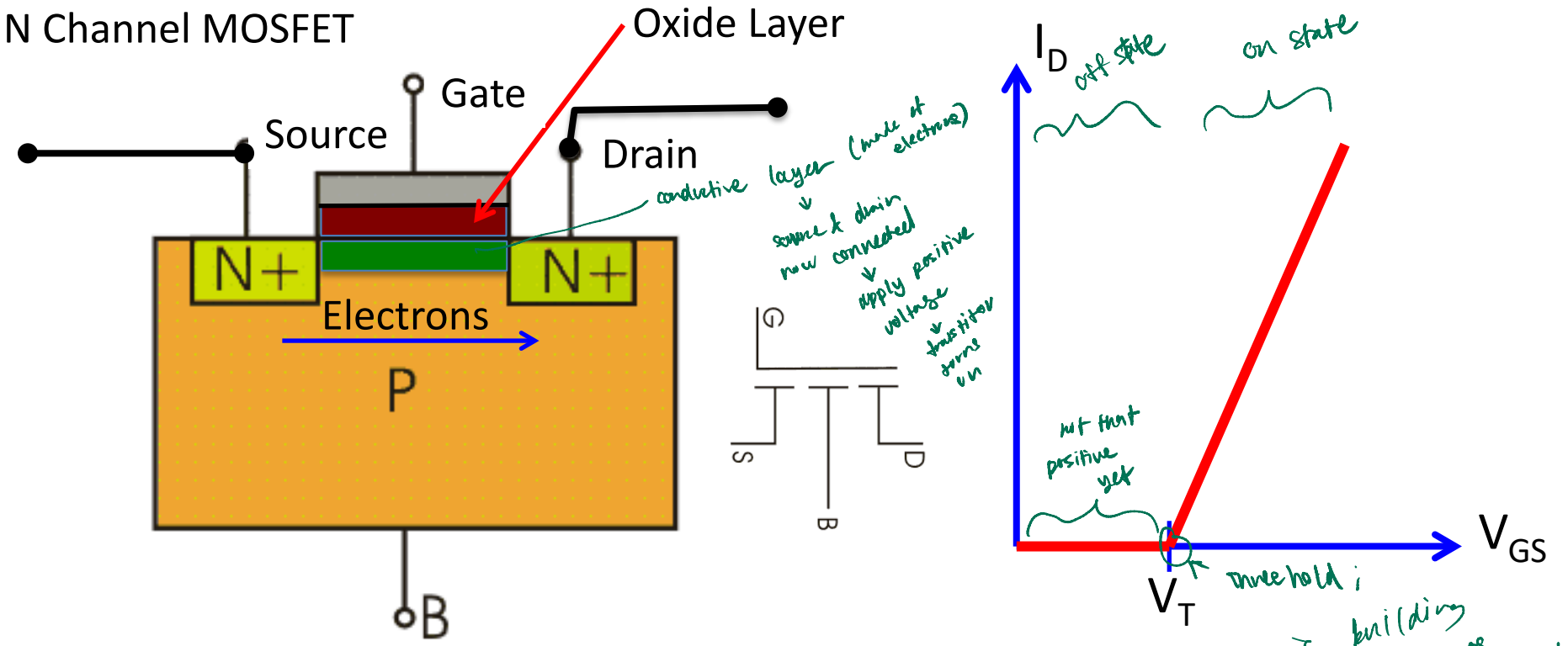
[Source] www.old-computers.com



Device cross-section

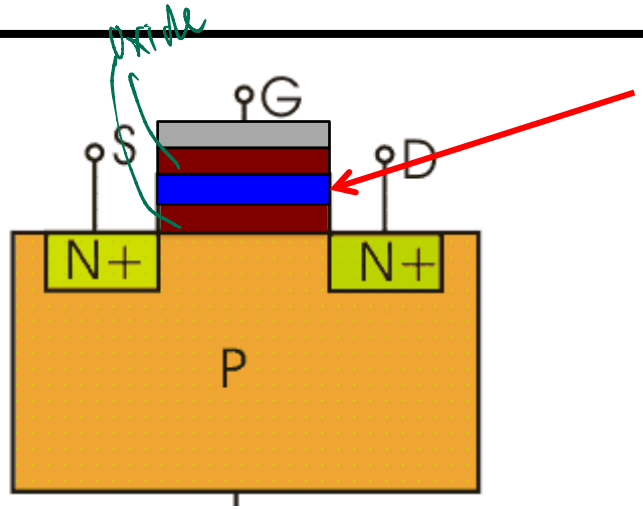
- EEPROM (Electrically Erasable PROM)
  - Advancement in process technologies make it possible to reduce the oxide thickness (t_{ox}).
  - Can electrically program or erase device.
  - Hence, named Electrically Erasable Programmable ROM (E^2PROM).

MOSFET

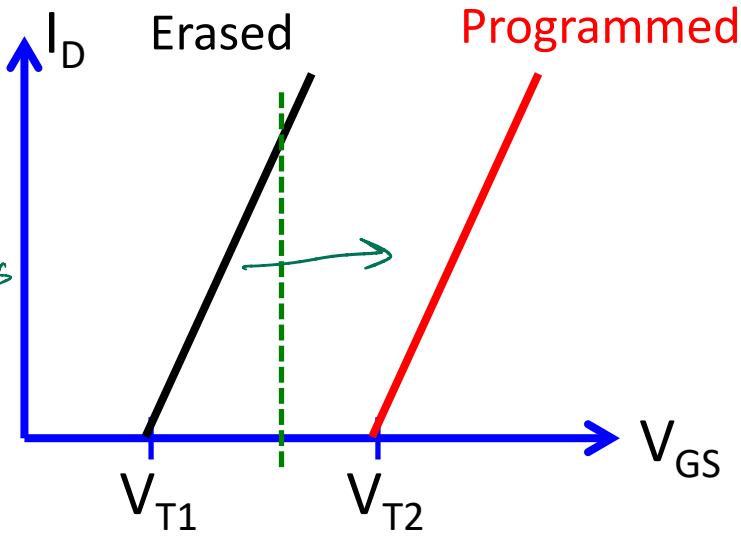


- MOSFET (Metal-Oxide-Semiconductor Field Effect Transistor)
- For N-Channel MOSFET, if Gate-Source Voltage (Vgs) > Threshold Voltage (Vt), a conductive channel of electrons (inversion layer) will be formed and current will flow if a positive voltage is applied across Drain and Source.

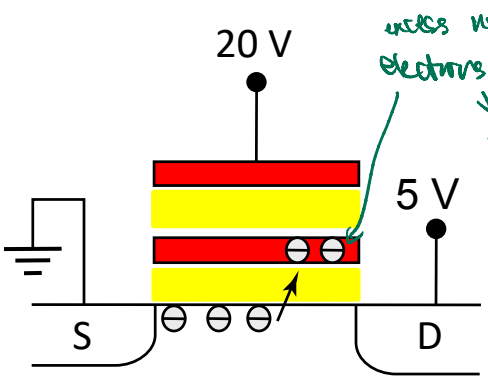
Floating Gate Transistor (FGT)



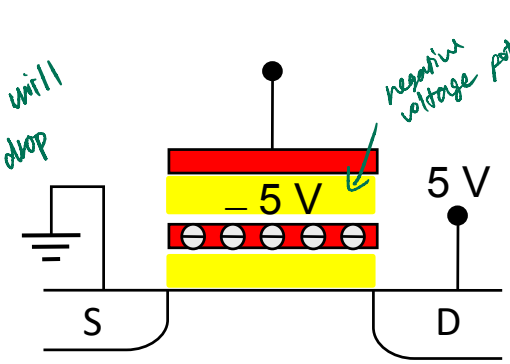
Floating Gate



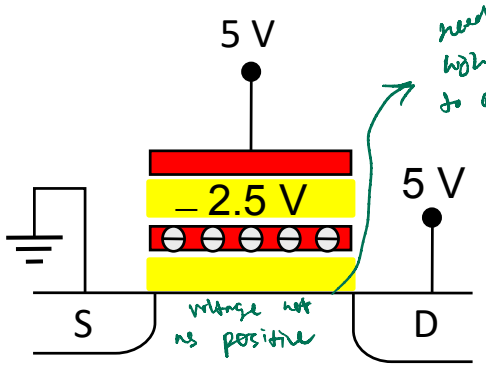
"Programming" results in altered threshold voltage of Floating Gate Transistor



Avalanche injection (Programming)

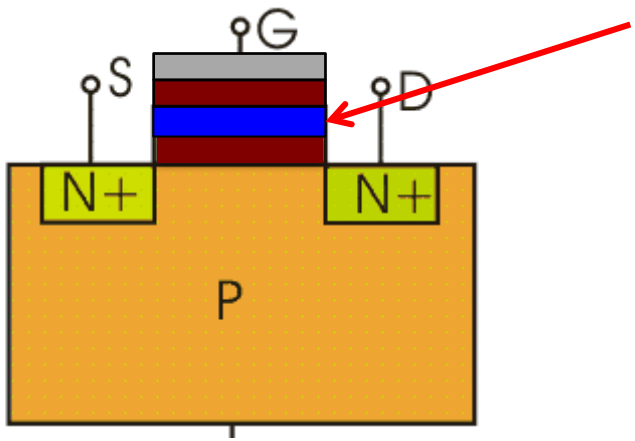


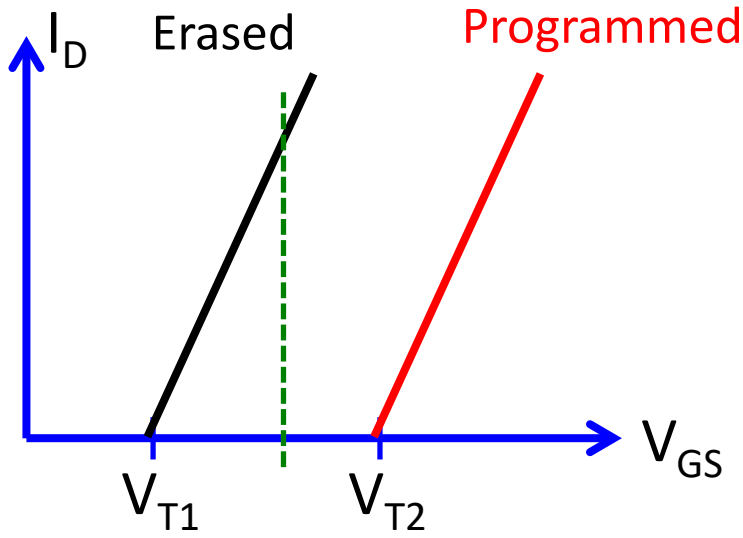
Removing programming voltage leaves charge trapped



Programming results in higher V_T.

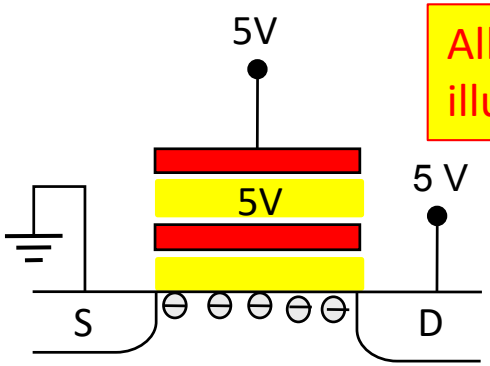
Floating Gate Transistor (FGT)



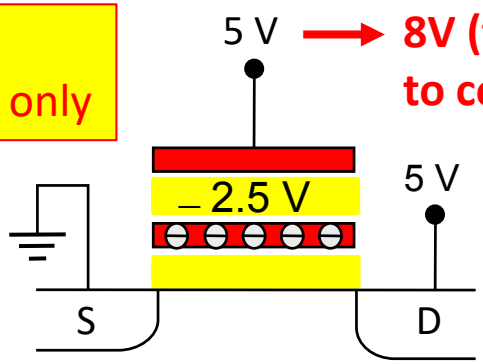


"Programming" results in altered threshold voltage of Floating Gate Transistor

All values are for illustration purpose only



Erased State
No excess electrons in Floating Gate

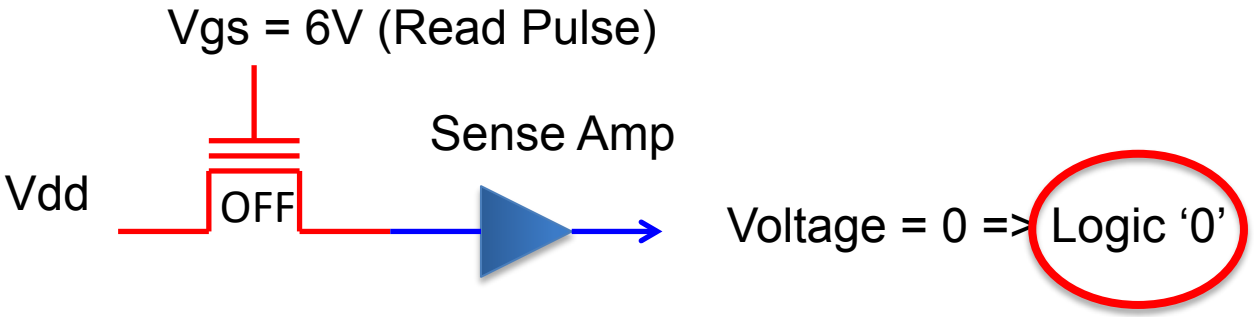


Programmed State

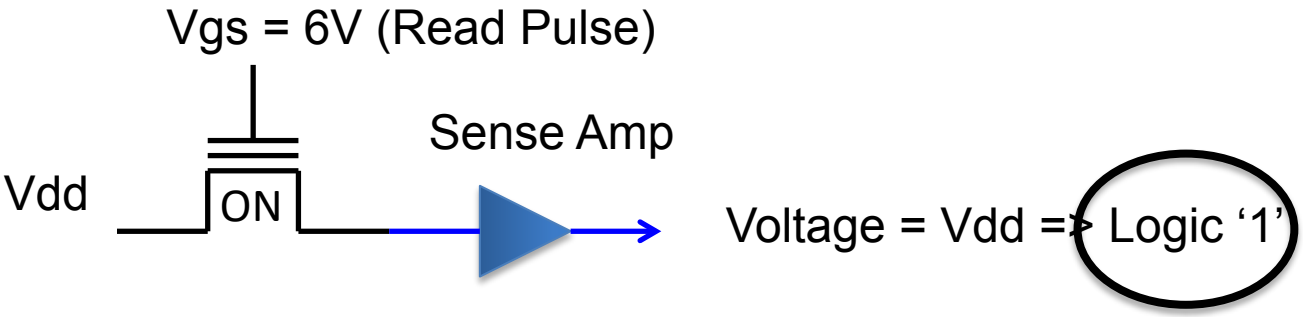
Reading of Stored Data in Floating Gate

- With a positive Gate Voltage (Vgs) = 6V
  - Transistor OFF if floating gate is programmed (contains charges).
  - Transistor ON if floating gate is erased (no charges).
- 6V value below is for illustration only. Actual Vt value in real world may varies depending on the doping of the transistors.

Programmed State
Transistor OFF
(Vt > 6V)

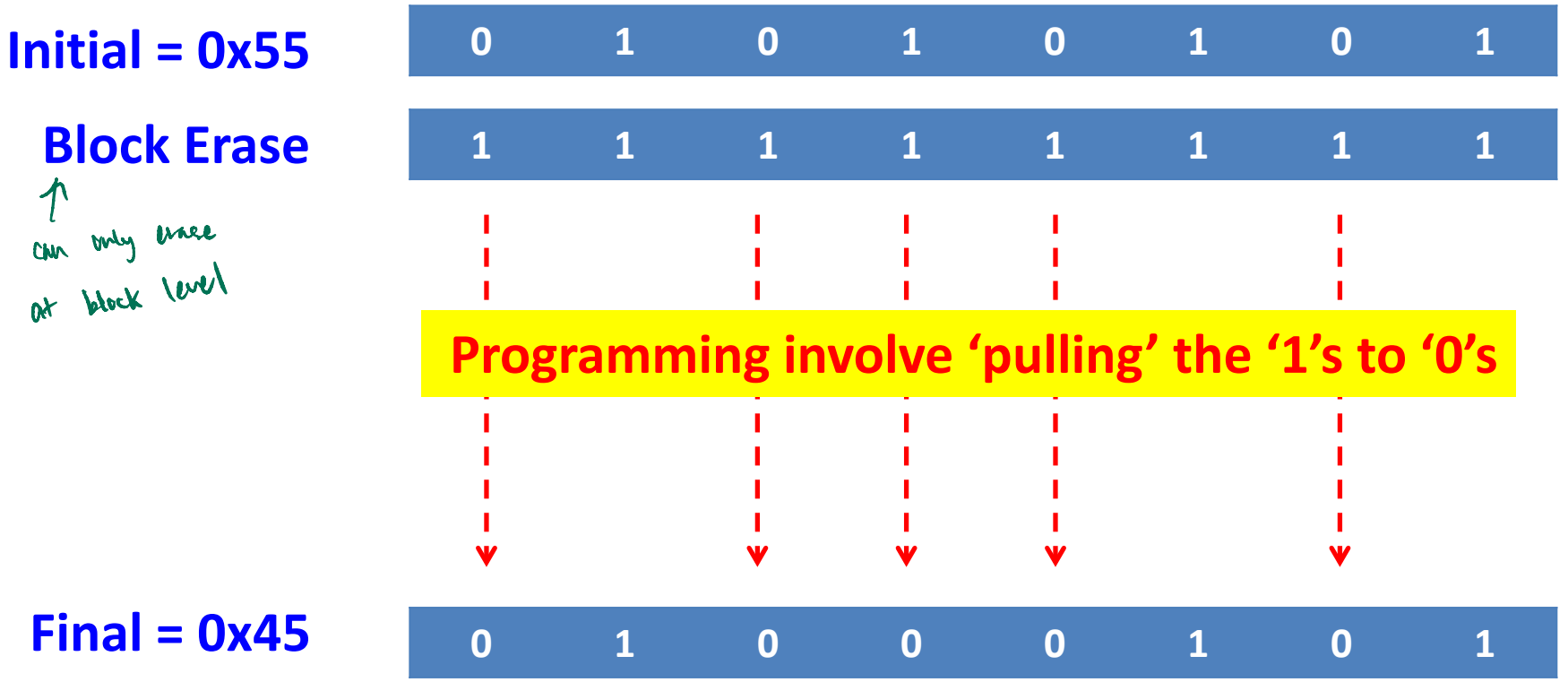


Erased State
Transistor ON
(Vt < 6V)



Example of Programming a FGT based memory

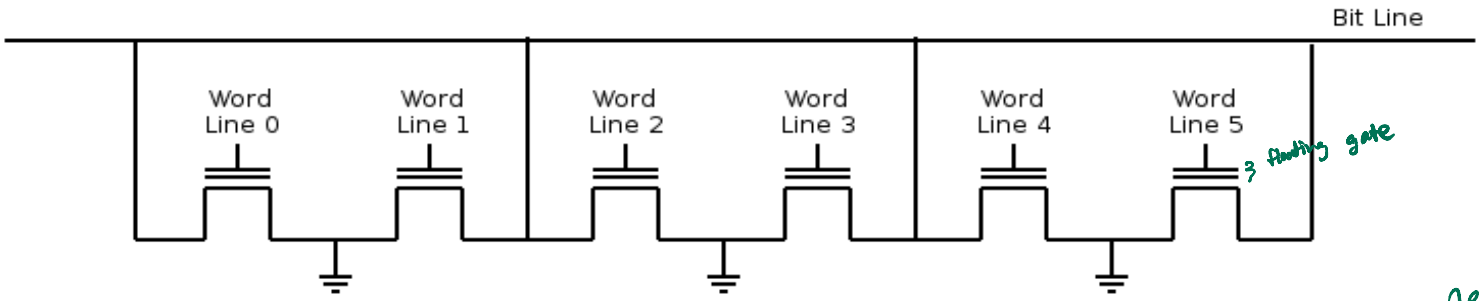
- Flash 'programming' can only modify the cell content from '1' to '0'
- To modify the cell content from '0' to '1', an erase operation has to be done at block level → multiple bits have to be erased at the same time
- Example: To Program a value of '0x45' when initial value in flash memory is '0x55'



Flash

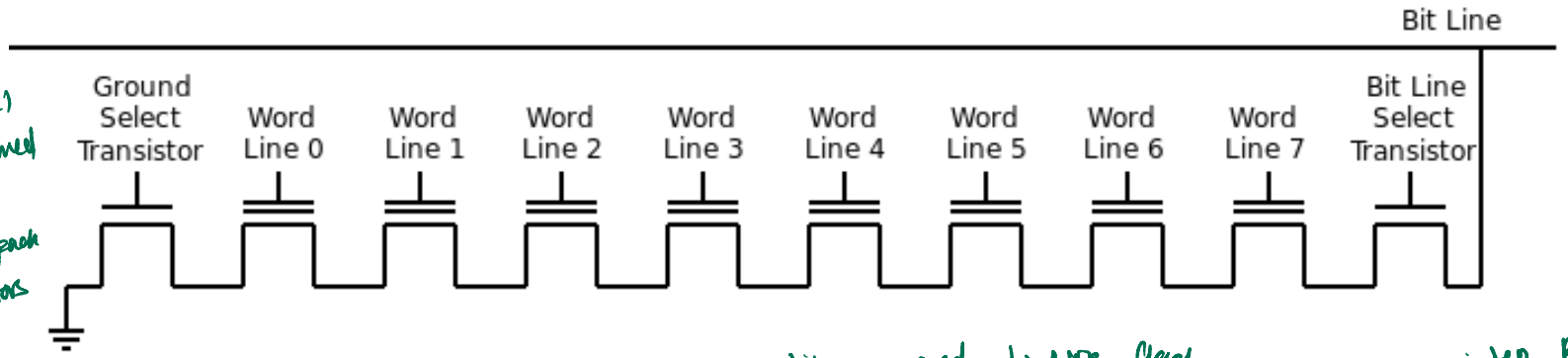
- Based on similar floating gate technology as EEPROM.
- Can be erased in larger blocks size compared to EEPROM (which typically has smaller page/block size).
- Since Erase cycle is comparatively slower than other operations (Read/Write), Flash memory has a faster speed than EEPROM when performing write operations for large block of data.
- Flash also cost less than EEPROM. → due to overhead logics like clearing & erasing of memory cells
- Suitable for system requiring large amount of non-volatile memory. ↓ storage memory
- Two main types of Flash in the market
  - NAND Flash
  - NOR Flash

NOR Flash



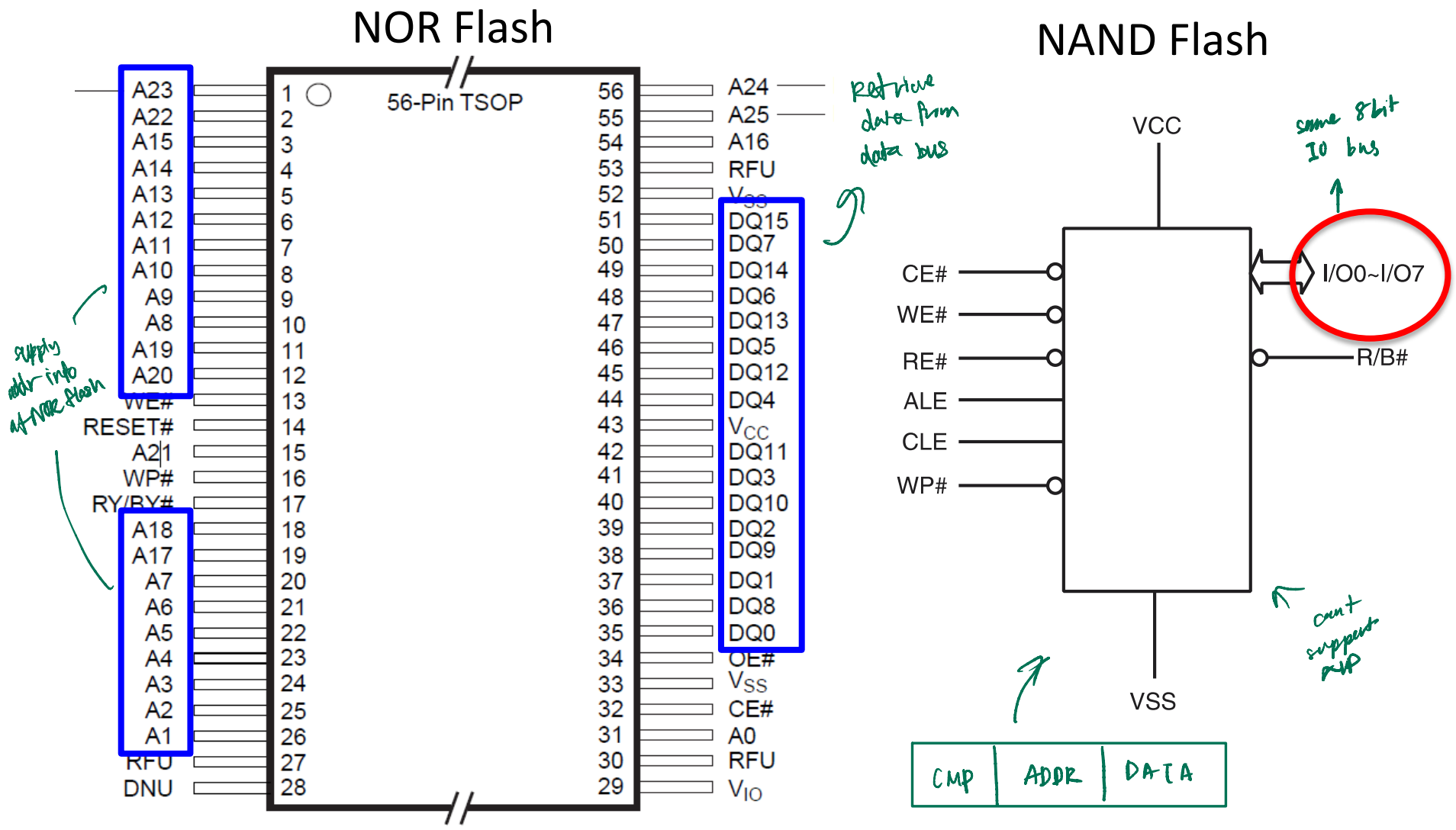
- Cell behaves like a NOR gate. When any one of the Word Line > Vt(Prog), Bit Line output = 0.
- Supports Execute in Place, i.e. Program code stored in Parallel NOR Flash can be executed directly without the need to transfer to internal RAM first.
  - Allows Random Reading of memory data using only Address information (no additional Commands needed).
- Need Special Commands (issue in Write mode) in order to perform operations other than Data Read. E.g. Program, Erase etc.
- Allows random word/byte programming. But erasure has to be done at block level. Typical Block size: 64KByte, 128KByte, 256KByte
- Use as system memory to store program code or general storage memory

NAND Flash



- Cell behaves like NAND gate.
  - Bit line output = '0' only when ALL Word Line > Vt(Prog).
- Does not support execute in place operation.
  - Data has to be accessed one page at a time.
  - Command issued to open a particular page, followed by which byte(s) is/are needed in the page.
  - NAND chips uses a single bus to carry Address and Data.
- Lower Cost per Bit than NOR. Used mainly as Main Storage Memory.
  - On Board Main Storage, USB Flash Drive, SSD etc

NOR and NAND Flash chip pin-out



System vs Storage Memory

- System Memory
  - Used to store runtime program and code is executed directly from the system memory
  - This typically refers to
    - Internal SRAM/DRAM or NOR Flash
    - External memory that supports XIP (NOR Flash, SRAM, DRAM)
  - DRAM technically speaking does not support XIP by itself but processors that has a DRAM interface will have a DRAM controller that handle the necessary translation to allow execution of code directly from the DRAM.
- Storage memory
  - Used to store all program and data in a computer system
  - Cannot run code directly from storage memory, needs to be transferred to system memory before code execution
  - All memory types can be used as storage memory

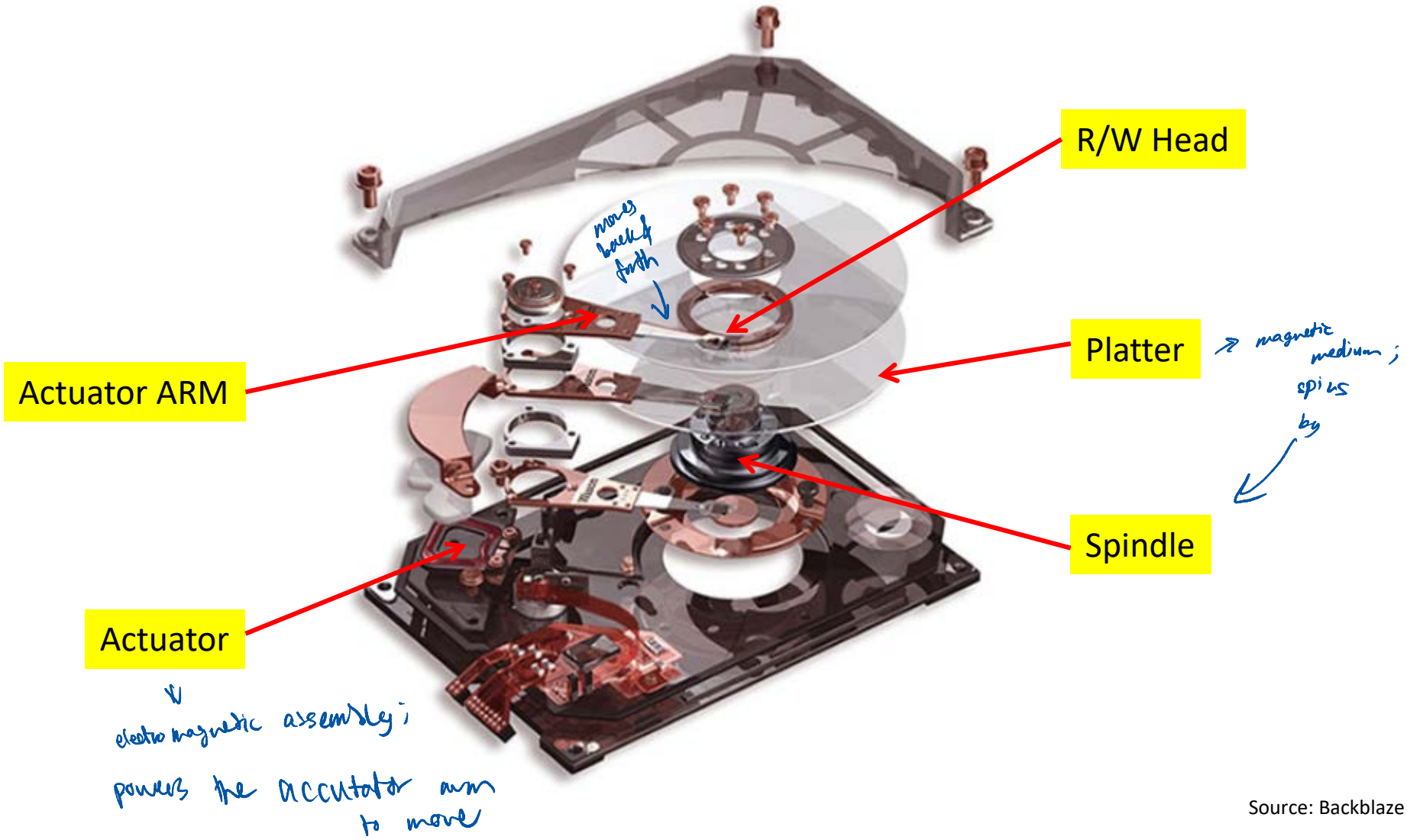
Cx1106
Computer Organization and Architecture

HDD and SSD

Oh Hong Lye
Lecturer
School of Computer Science and Engineering, Nanyang Technological University.
Email: hloh@ntu.edu.sg

Magnetic Hard Disk

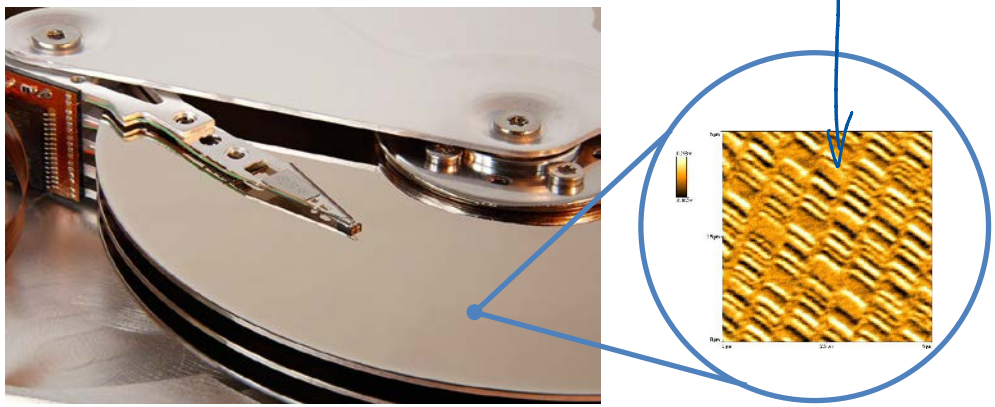
- mainly on desktop & data centres



Source: Backblaze

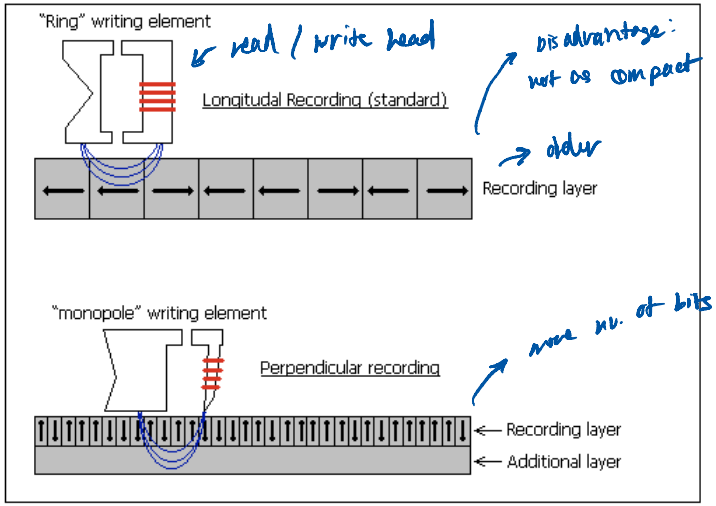
Magnetic Hard Disk

magnetic crystals

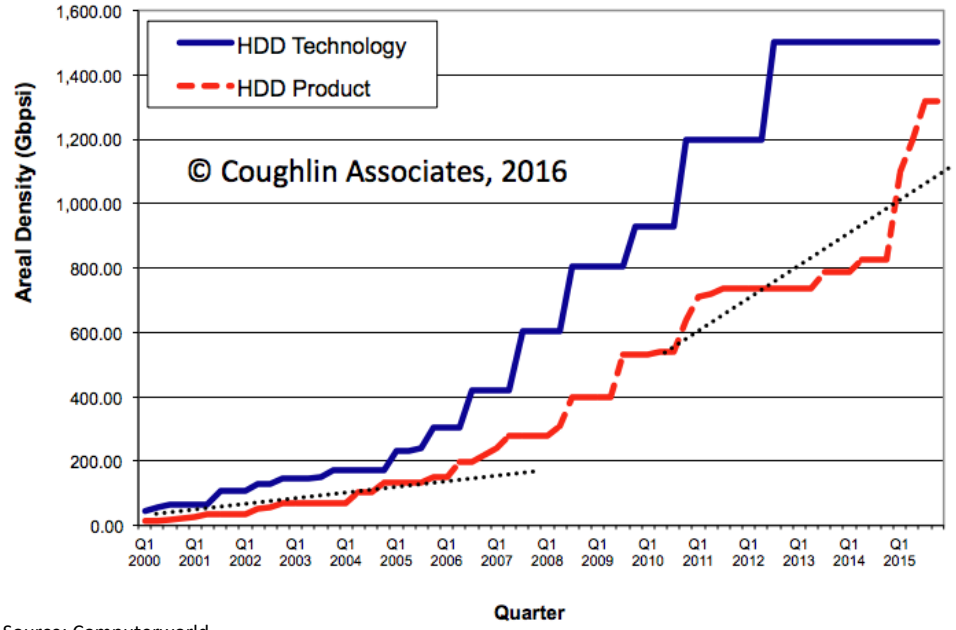


[Source] http://en.wikipedia.org/wiki/Disk_read-and-write_head

Longitudinal vs Perpendicular Recording



[Source] Robert Fontana et al, IBM Areal Density Comparison Paper, 2010



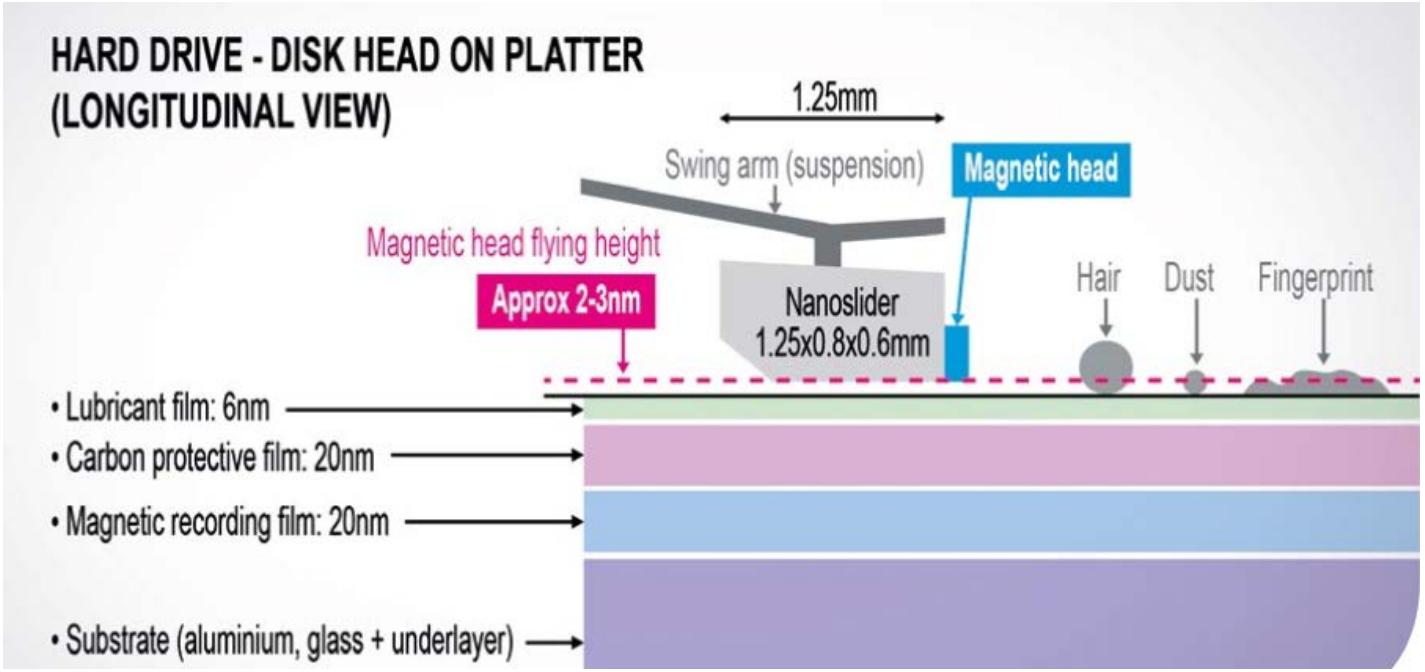
Source: Computerworld

- Stores data by magnetizing a thin film of ferromagnetic media on the circular disk known as platter.
- Video on Introduction to Hard Disk. https://www.youtube.com/watch?v=kdmLvl1n82U

HDD Technology

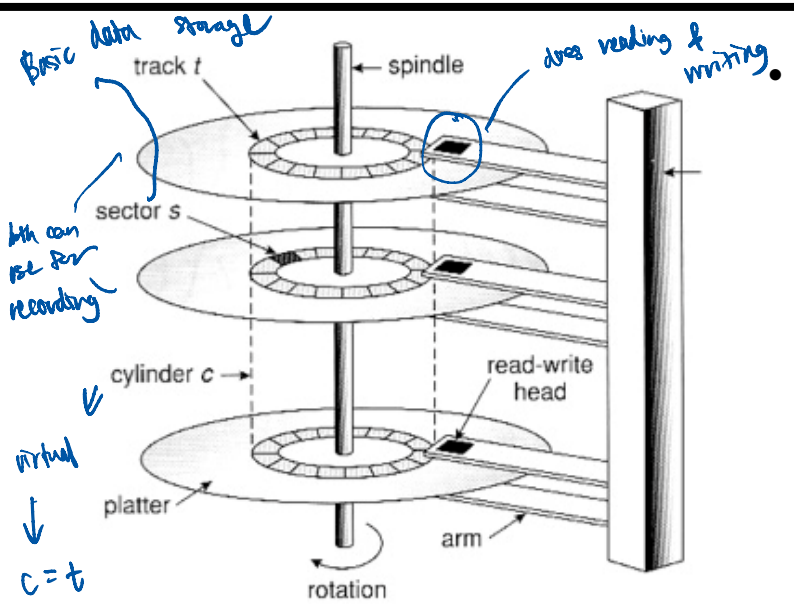
- The magnetic head uses the wind pressure generated by the high speed rotation of the disk to fly above the disk surface, similar to the principles employed by the aircraft.
- Fly height is lower than a finger print mark.

Load/unload sensitive to movement



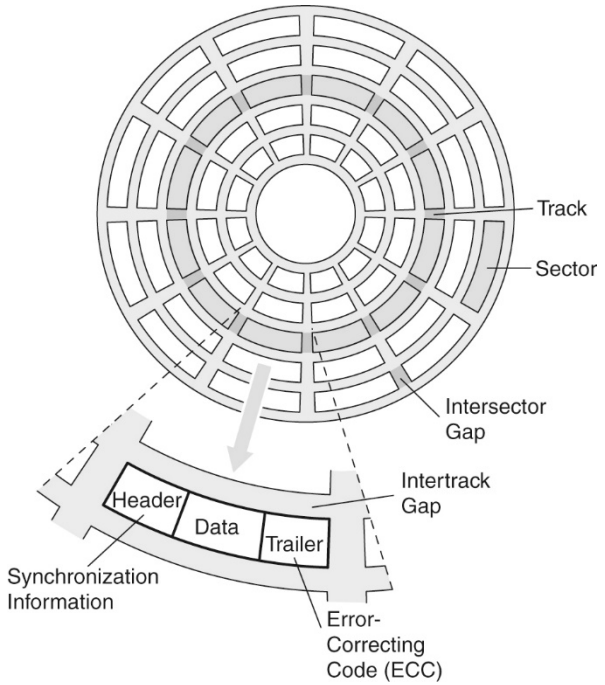
Source: datarecoverydublin

Magnetic HDD Data Organization



- Computers often use magnetic hard disks for large secondary storage devices.
  - One or more platters on a common spindle.
  - Platters are covered with thin magnetic film.
  - Platters rotate on spindle at constant rate.

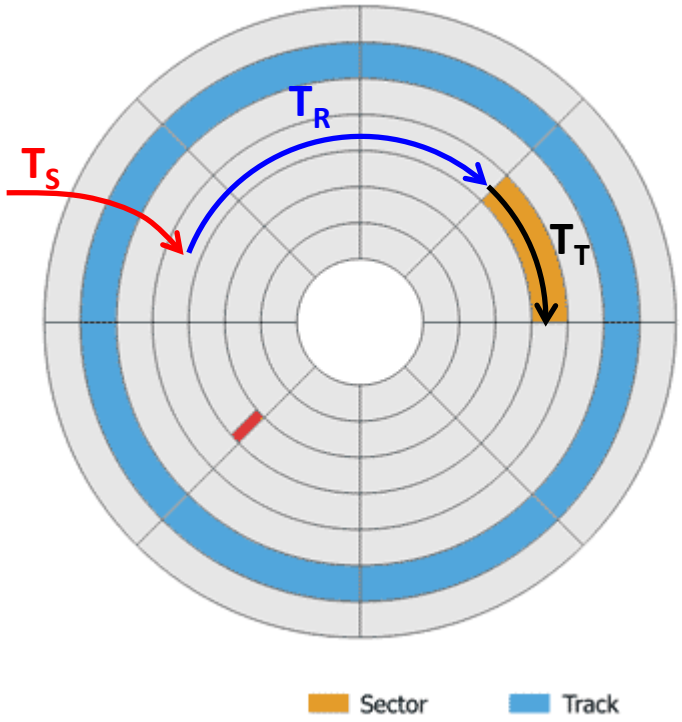
- Data is stored on the surface of the platter in concentric rings called tracks.
  - Gaps between tracks to minimize interferences from adjacent tracks.
- Tracks are divided into sectors.
  - Minimum data block size is a sector.
- Stored data consists header, data and trailer.



HDD Transfer Rate

- Seek time (T_S)
  - Time taken for the head to move to the correct track.
- Rotational Delay (T_R)
  - Time taken for the disk to rotate until the read/write head reaches the starting position of the target sector.
- Access Time (T_A)
  - Time from request to the time the head is in position ($T_S + T_R$)
- Transfer Time (T_T)
  - Time required to transfer the required data after the head is positioned.

y can have multiple sectors



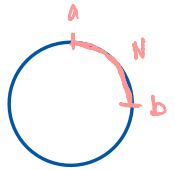
HDD Transfer Rate

- T_R is dependent on the rotational speed, Revolutions Per Minute (RPM), of the disk. For calculations, RPM is usually converted to Revolutions Per Second (RPS). i.e. RPS = RPM/60
- For a random section, average rotational delay $T_{R,AV}$ may be calculated as

$$T_{R,AV} = \frac{0.5}{RPS} \quad \text{seconds}$$

- T_T is dependent on the rotational speed of the disk, the Track Density D_T (number of sectors per track), Sector Density D_S (number of bytes per sector) and the number of bytes N for the transfer.

$$T_T = \frac{N}{RPS * D_T * D_S}$$



entire circumference

HDD Transfer Rate Example

- A magnetic hard disk rotates at 15000 RPM, with the following properties:
  - Average Seek Time, T_S = 4ms
  - Track density, D_T = 500 sectors per track
  - Sector density, D_S = 512 bytes per sector
- Calculate the total time T_{TOTAL} it takes to read a 3 KB file stored in consecutive sectors on the same track?

If data stored back to back on the track

[Solution] RPM to RPS

RPS = 15000/60 = 250 per second

Access time $T_A = T_S + T_R$ = 4 ms + (0.5/250) = 6 ms

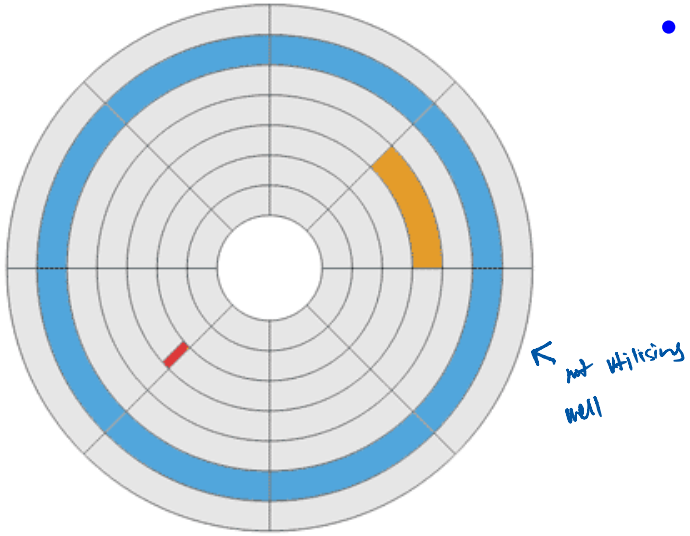
Transfer time T_T = 3072/(250*500*512) = 48 µs

Total time, $T_{TOTAL} = T_A + T_T$ = 6.048 ms

File storage is fragmented

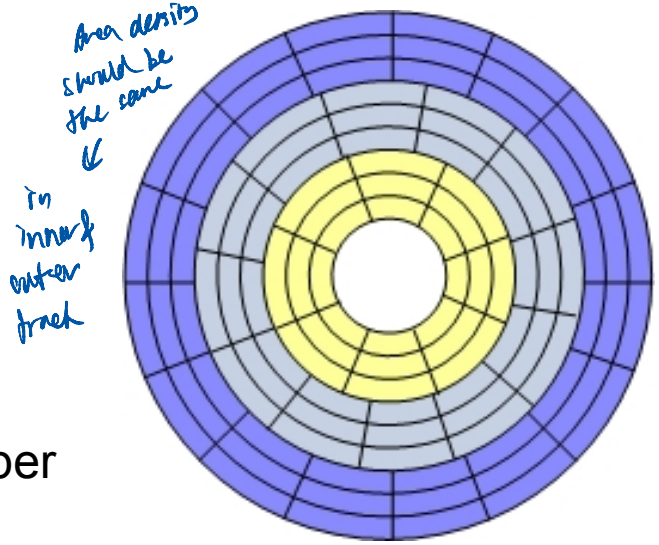
Notice $T_A >> T_T$. This example shows that accessing file whose data is distributed across sectors on different tracks on the magnetic hard disk would potentially incur more time.

Physical Layout



- Early HDD
  - Physical Layout refers to the actual layout design on the HDD.
  - Equal number of sectors per track and each sector has the same data size.
  - Same-numbered Tracks from different surfaces formed a cylinder.
  - The early hard disks were implemented using this topology to simplify the controller design.

- Modern HDD
  - Having equal number of sectors per track means that sectors at the outer tracks are wider.
  - Waste physical space as bit density of those sectors are not optimal.
  - Zone bit recording technique addresses space wastage.
  - Tracks are divided into zones, with differing number of sectors per track for different zones.



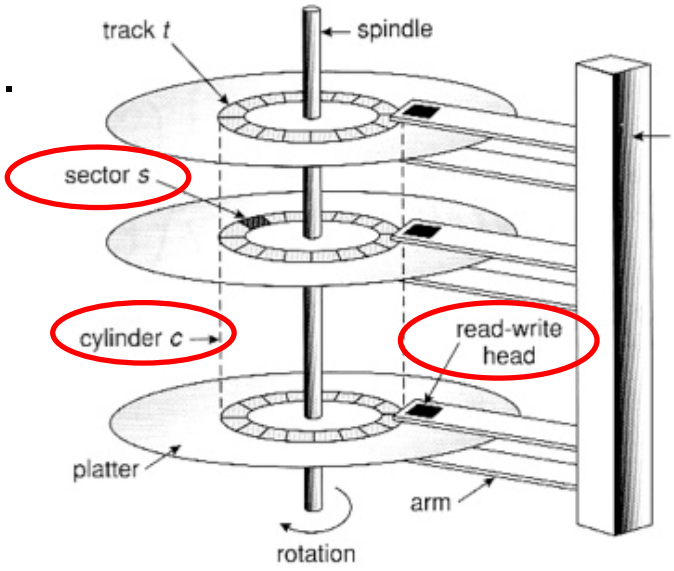
Logical Layout

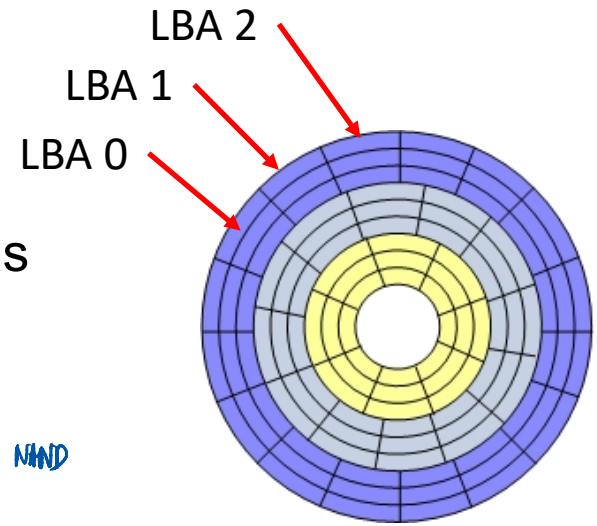
need to conform to logical layout standard;
only need to write software once;
and able to use for many different models of laptop, etc.

- How the software see and address the HDD data.
- Address translations are needed to map the physical to logical locations. Two addressing scheme are CHS and LBA.
- Cylinder-Head-Sector (CHS)
  - Legacy scheme using the old HDD physical structure.
  - Made Obsolete in recent standards due to addressing limitation and more complex formatting.
- Logical Block Addressing (LBA)
  - Simple linear addressing starting from LBA=0 as first block, LBA=1 as second block and so on.
  - 48-bit LBA standard allows addressing up to 128PByte. 1PByte=2^50 Bytes.

taken over by

compatible with NAND flash





Modern Day HDD

Hard drive → mini computer system

e.g. no. of heads, surfaces, tracks

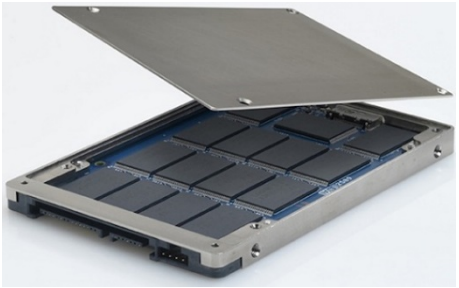
- Detail physical layout not given due to complex zone bit recroding.
- Only the average transfer rate is given these days.
- Use of Cache and Buffers to speed up data transfer.
- However, concept and limitation of the magnetic HDD's physical design still valid
  - Actuator ARM is fixed and access has to be sequential, once data is missed, it has to wait for a disc revolution.
  - Seeking from track to track takes time as the R/W head needs to align to the starting sector.
  - More efficient to read in blocks rather than random access
  - Vulnerable to motion

key disadvantage

due to the dominant time being T_A

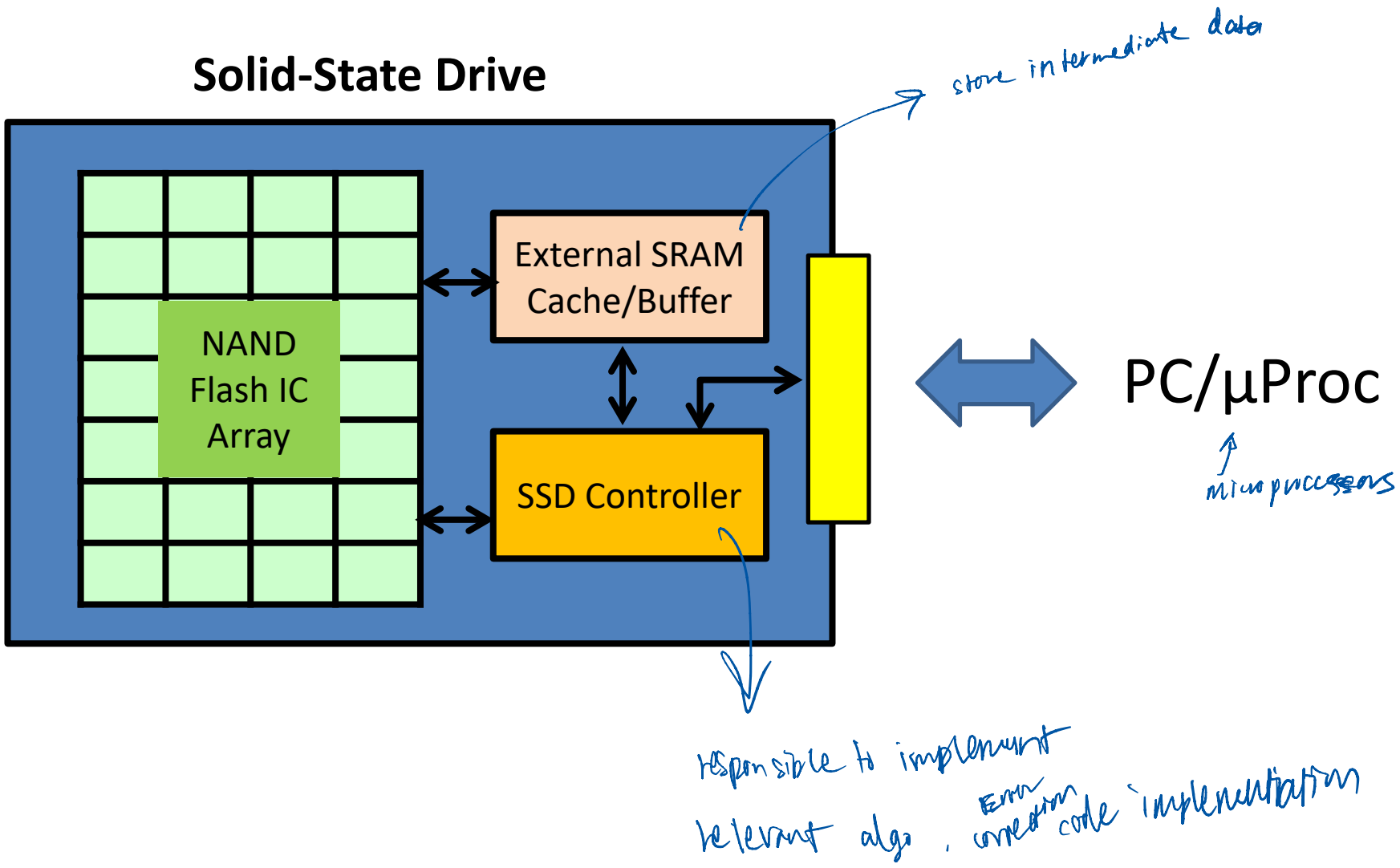
$T_s + T_r$

Solid State Drive (SSD)

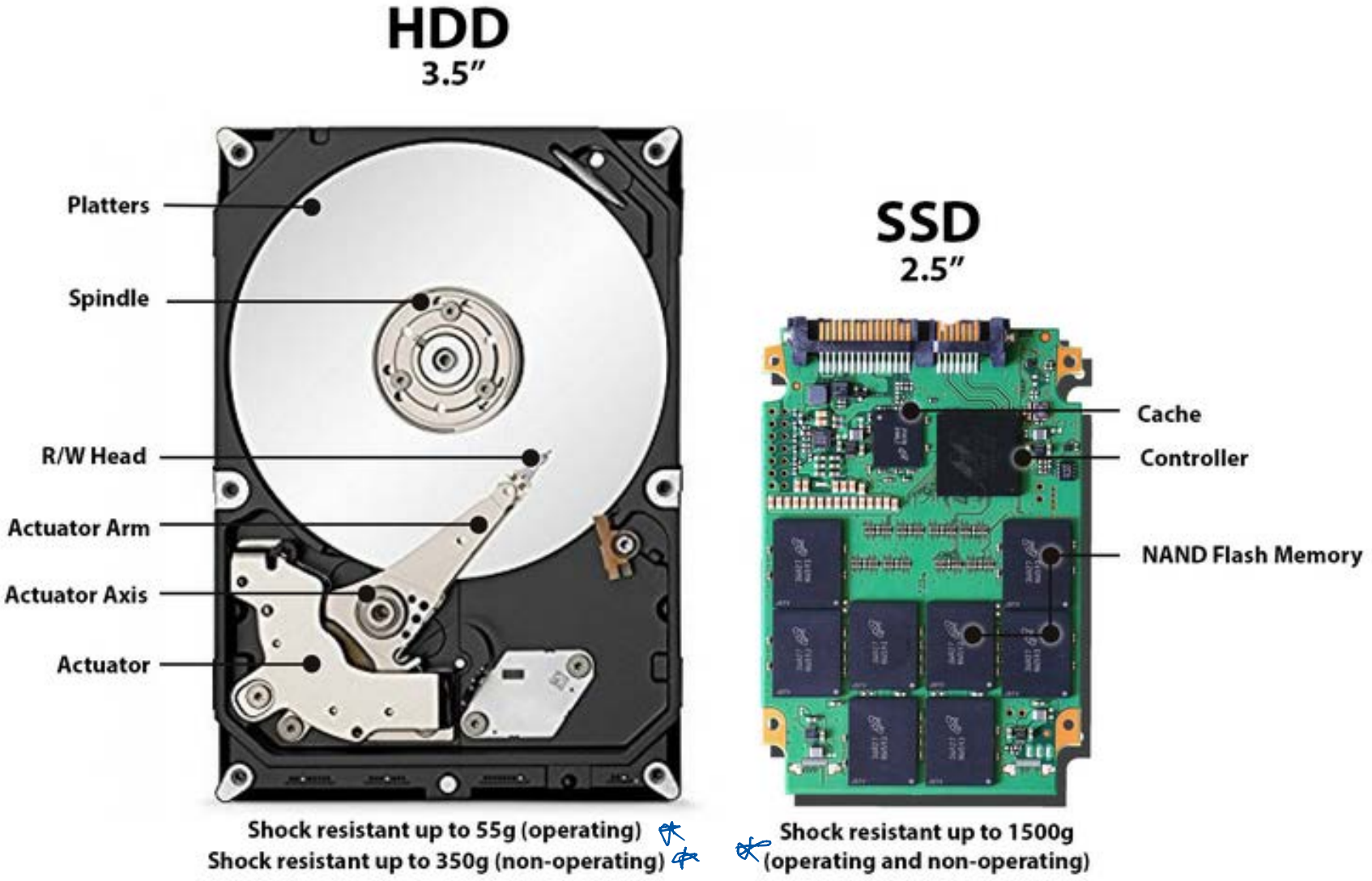


- Solid-state drives (SSD) are becoming popular
  - Memory array based on NAND-FLASH or NOR-FLASH.
  - Still more expensive than magnetic hard-disk.
- Flash memory has limited program-erase cycles (3,000 to 1,000,000).
- Various techniques used to extend the life of SSD
  - Wear levelling is a technique used to extend the life of the SSD disk by distributing data erase/write operations evenly over the entire disk.
  - Use External RAM as buffer to minimize the number of writes to Flash in SSD.
  - Error Correction Code. Ability to recover from one or more bits of error in media.
- MTBF (Mean Time To Failure) of recent SSD is comparable to that of HDD.

SSD System Block Diagram



HDD vs SSD



Source: Backblaze

HDD vs SSD

- HDD
  - Pros
    - Lower Cost Per Bit,
    - Almost infinite Erasure cycles
  - Cons
    - Consist of Moving Mechanical Parts so more prone to crashing if HDD is dropped or shaken.
    - Heavier and larger physical profile.
    - Slower transfer rate
- SSD
  - Pros
    - No moving parts. More robust to movement.
    - Lighter and occupy less space
    - Higher Transfer rate
  - Cons
    - More Costly compared to HDD
    - Finite number of Erasure cycles

→ why data centers use HDD over SSD

↓ due to mechanical limitations, e.g. spinning of platters

DATA CENTER STORAGE

Criteria for storage element

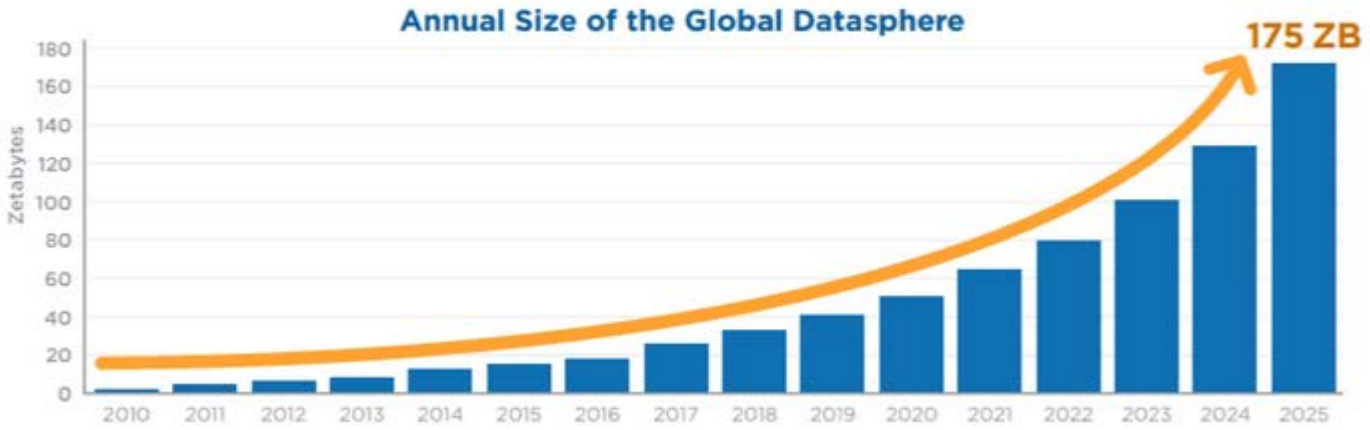
- Power consumption
  - Data centers consumed a lot of power, which not only translate to direct cost in utility and cooling measures, but also limit its choice of location.
  - Centers are commonly found near natural cooling elements such as large natural water bodies which offer a low cost and reliable source of cooling.
- Speed
  - Beneficial for caching databases and other data affecting overall application or system performance.
- Robustness
  - Tolerance to various form of mechanical movement/interference increases reliability and reduces need and cost of maintenance.
  - Drive housing structure shock absorption requirement is also reduced.
- Heat production
  - The less heat generated the less cooling and power required in the data center.
- Size
  - Data centers will be able to store more data in less space, which increases efficiency in all areas (power, cooling, etc.)

Criteria

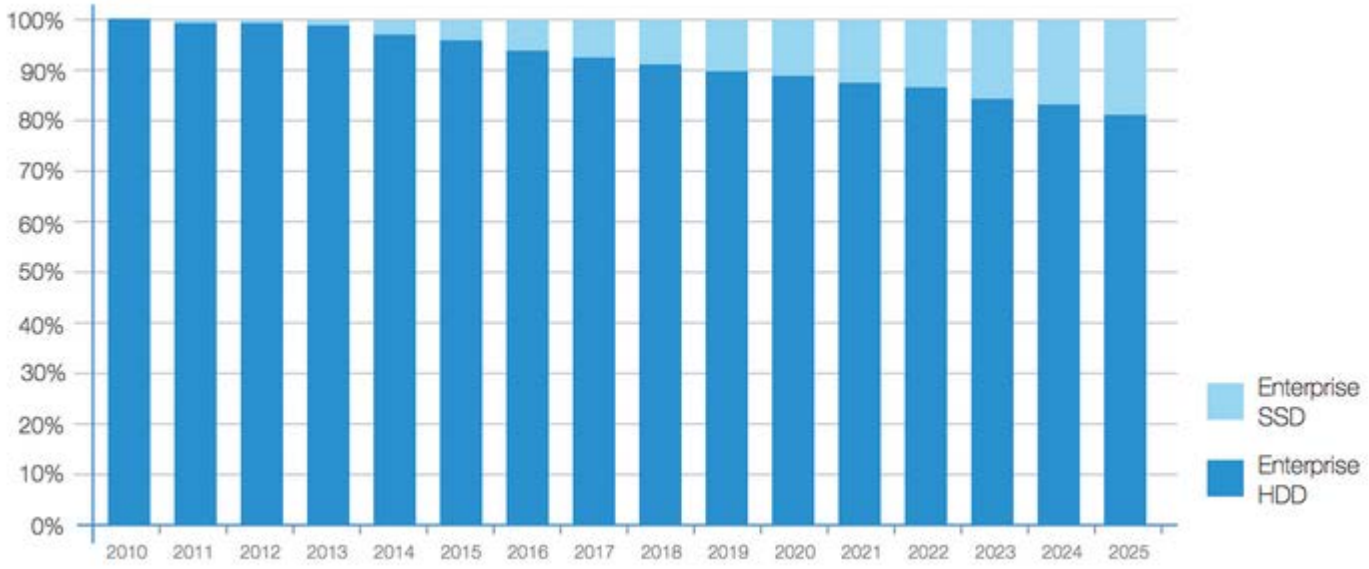
- Based on the criteria in the previous slide, SSD wins HDD hands down.
- So why is HDD still the dominant Storage in Data Centers?

COST

Global Storage Consumption and Shipment Projection

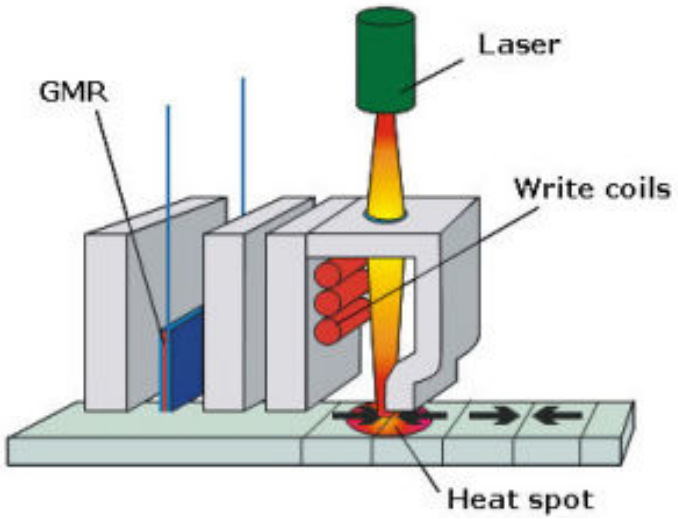


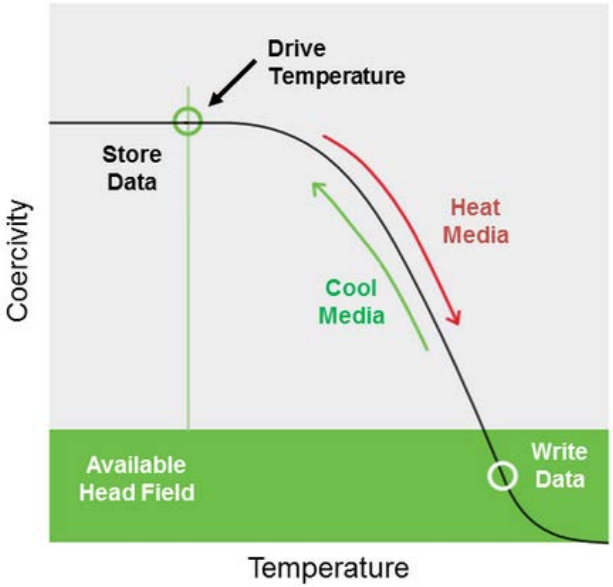
1 ZB = 10^{21} bytes



Source: IDC, Seagate.

HMAR (Heat Assisted Magnetic Recording)





- The areal density of the HDD was stagnant for a while and manufactures had been shipping drives with 2TByte/Platter.
- But with the HMAR, the areal density is expected to increase again.
- A small laser is attached to a recording head, designed to heat a tiny spot on the disk where the data will be written. This allows a smaller bit cell to be written as either a 0 or a 1.
- Current projections are that HAMR can achieve 5 Tbpsi, enabling hard drives with capacities higher than 100 TB.

HMAR

- The major problem with packing bits so closely together on conventional magnetic media is that the data bits become unstable and may flip (0 → 1 or 1 → 0).
- To make the media maintain their stability to store bits over a long period of time, the recording media needs to have a higher coercivity.
- Higher coercivity implies the media is magnetically more stable during storage, but it would also be more difficult to change the magnetic characteristics of the media when writing.
- For that, a laser is employed to heat a tiny region of several magnetic grains for a very short time (~1 ns) to a temperature high enough to lower the media's coercive field to below that of the write head's magnetic field. This is the write process.
- Immediately after the heat pulse, the region quickly cools down and the bit's magnetic orientation is frozen in place. The data is stored in the media and would be stable due to the high coercivity of the recording media.
- See the graph in the previous slide for the visual illustration.

16/3 Lecture

multiple * means multiple answers

**What is/are the effects of DRAM using single transistor cum capacitor design for storage?

- ✓ A. Slower as memory needs to be refreshed
- B. Simpler interface design as less transistors are used
- ✓ C. Smaller layout compared to SRAM as fewer number of transistors are used
- D. Larger layout compared to SRAM as capacitor occupies a larger area compared to transistors

→ SRAM: at least 6 transistors
DRAM: 1 transistor

**What is/are the difference between SRAM and DRAM if both use same process technology?

- ✓ A. SRAM is faster than DRAM — cause of overhead due to refresh
- B. DRAM is faster than SRAM
- ✓ C. SRAM has a larger layout than DRAM
- D. SRAM uses capacitor as a storage element
- ✓ E. DRAM requires periodic refresh to keep its data integrity

Why does DRAM requires constant refresh operation to keep its data integrity?

A. DRAM is a type of volatile memory

✓B. DRAM uses capacitor as its storage element

C. DRAM uses less transistors compared to SRAM so less function build in

D. DRAM is slower than SRAM so needs to periodically refresh its content

The charged capacitor in DRAM design has a natural leakage path to ground, so charges is lost over time and data integrity is compromised.

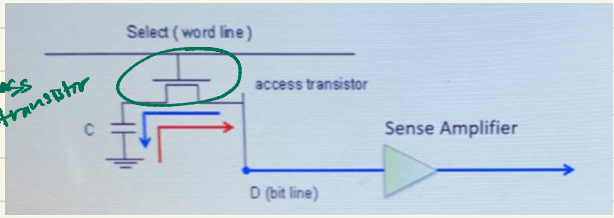
What is the function of the pass transistor in the DRAM Design?

A. Store data

✓B. Control flow of charges to and from the capacitor

C. Control read/write operation mode

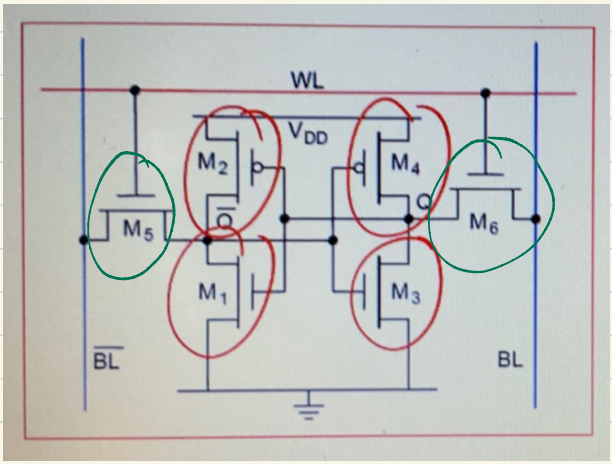
D. Logics to charge and discharge the capacitor



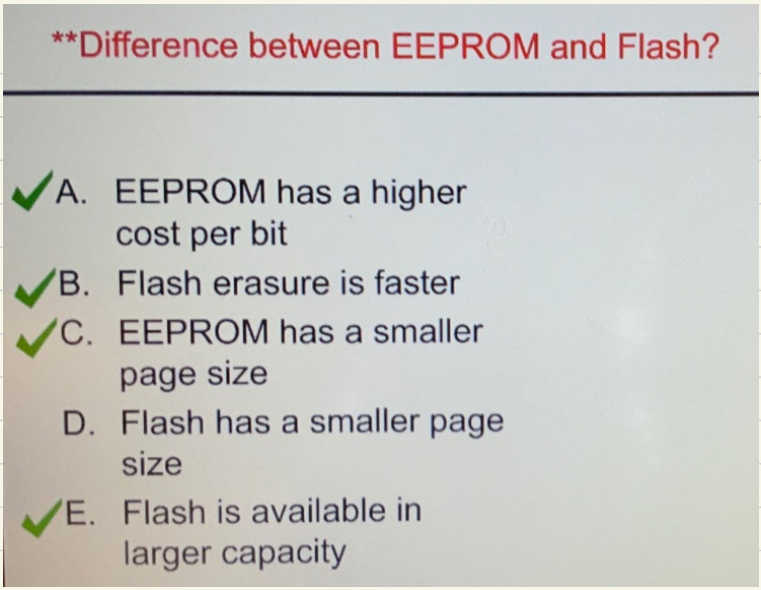
What is the function of transistors in SRAM design?

A. Store Data
B. Control Read/Write mode
C. Store Data and Control logic for signal transfer between SRAM cells and internal data lines
D. Control logic for signal transfer between SRAM cells and internal data lines

No need to know how draw, but need to know functions & logic equivalent



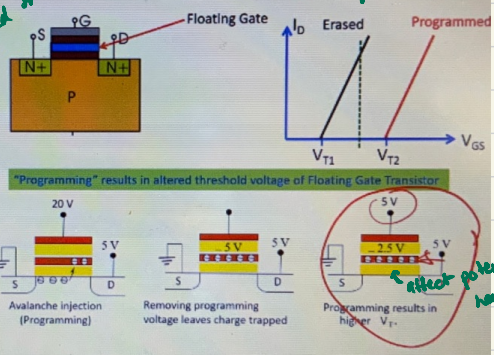
* Pass transistors
* store data & control.



bigger block size → erase faster

Which state of the floating gate transistor will yield a higher threshold voltage (Vt)?

A. No Electrons in the floating gate
B. Presence of Electrons in the floating gate
C. Electrons injection into the floating gate, reducing the gate's voltage potential.
D. Electrons removal from the floating gate, increasing the gate's voltage potential



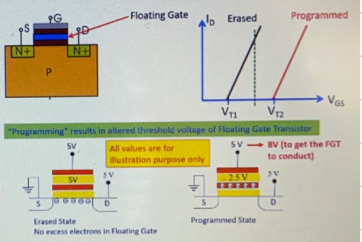
Why is Flash memory able to retain its data after its power is removed?

A. Flash memory is made up of transistors which is used to store data.

B. SRAM and Flash has different design and SRAM is non-volatile.

✓C. The electrons stuck in the floating gate remain there as there is no electrical connection between the gate and to allow electron discharge.

D. Threshold voltage of Flash memory at programmed state is higher than that in erased state.



Flash memory is non-volitile memory

Correction:

Why is Flash memory able to retain its data after its power is removed?

A. Flash memory is made up of transistors which is used to store data.

B. SRAM and Flash has different design and SRAM is ~~non-~~ volatile.

✓C. The electrons stuck in the floating gate remain there as there is no electrical connection between the gate and to allow electron discharge.

D. Threshold voltage of Flash memory at programmed state is higher than that in erased state.

[A], [B] and [D] are true statement by itself but doesn't explain why Flash memory is non-volatile.

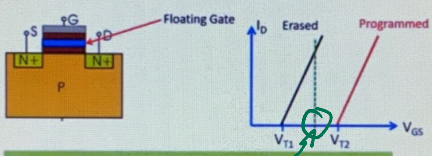
The two oxide layers that sandwich the floating gate are bad conductors of electricity so electrons will not get leaked out of the floating gate. As the electrons are retained, the attributes they induced (increase in Threshold Voltage) is also retained, meaning the data 'stored' is retained when power is off.

What is the logic state of the flash content if the flash is 'Erased'?

A. Logic '0'

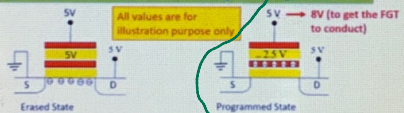
✓B. Logic '1'

can be user defined

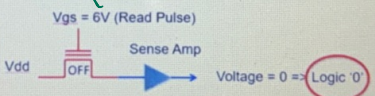


"Programming" results in altered threshold voltage of Floating Gate Transistor

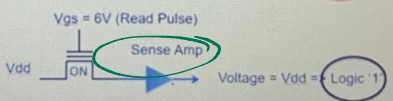
All values are for illustration purpose only

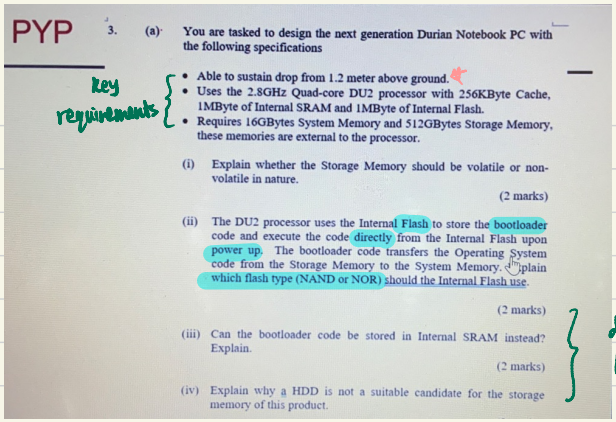


Programmed State
Transistor OFF
(Vt > 6V)



Erased State
Transistor ON
(Vt < 6V)





PYP

3. (a) You are tasked to design the next generation Durian Notebook PC with the following specifications

key requirements

- Able to sustain drop from 1.2 meter above ground.
- Uses the 2.8GHz Quad-core DU2 processor with 256KByte Cache, 1MByte of Internal SRAM and 1MByte of Internal Flash.
- Requires 16GBytes System Memory and 512GBytes Storage Memory, these memories are external to the processor.

(i) Explain whether the Storage Memory should be volatile or non-volatile in nature.

(2 marks)

(ii) The DU2 processor uses the Internal Flash to store the bootloader code and execute the code directly from the Internal Flash upon power up. The bootloader code transfers the Operating System code from the Storage Memory to the System Memory. Explain which flash type (NAND or NOR) should the Internal Flash use.

(2 marks)

(iii) Can the bootloader code be stored in Internal SRAM instead? Explain.

(2 marks)

(iv) Explain why a HDD is not a suitable candidate for the storage memory of this product.

deals with XIP & volatility

(a)(i) non-volatile : able to retain the memory even while system is powered down.

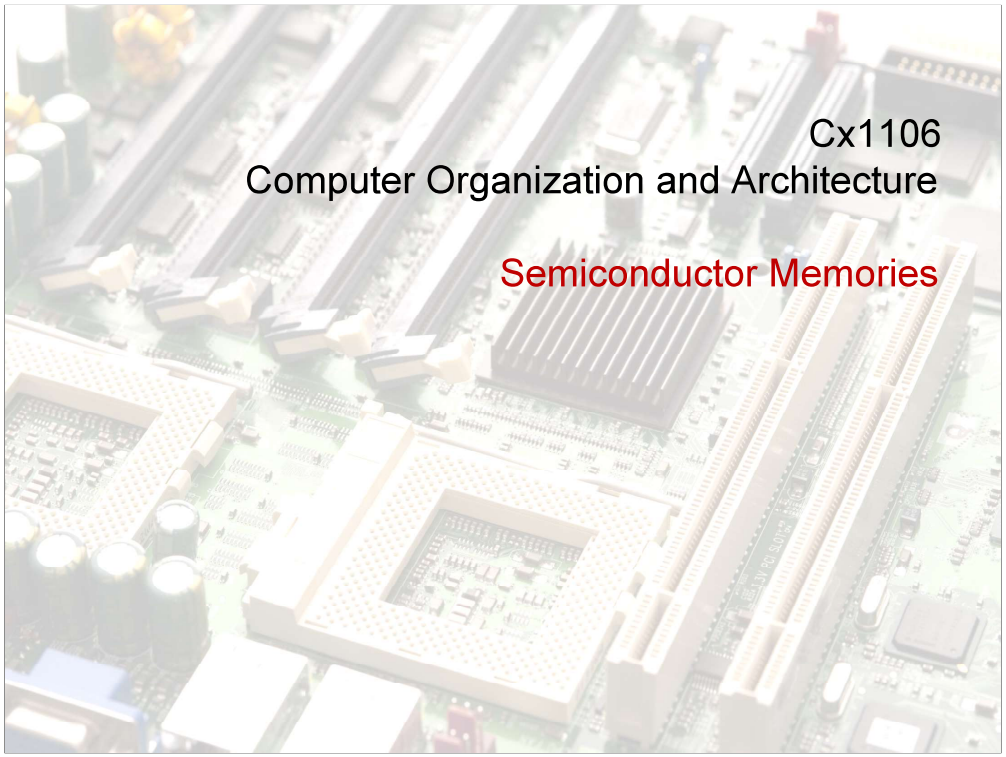
(ii)

(iii) [recall characteristics of SRAM]

(iv) hard disk prone to crashing due to a lot of movement.

Need to understand particular properties & attributes of memory and which is best used in what case.

e.g. volatile / non-volitile, support XIP or not, fast, lower cost compared to other memory type.



Cx1106
Computer Organization and Architecture

Semiconductor Memories

**Which of the following memory are volatile in nature?

A. Magnetic HDD
✓B. SRAM
✓C. DRAM
D. EEPROM
E. NOR Flash

**Which of the following memory devices are non-volatile?

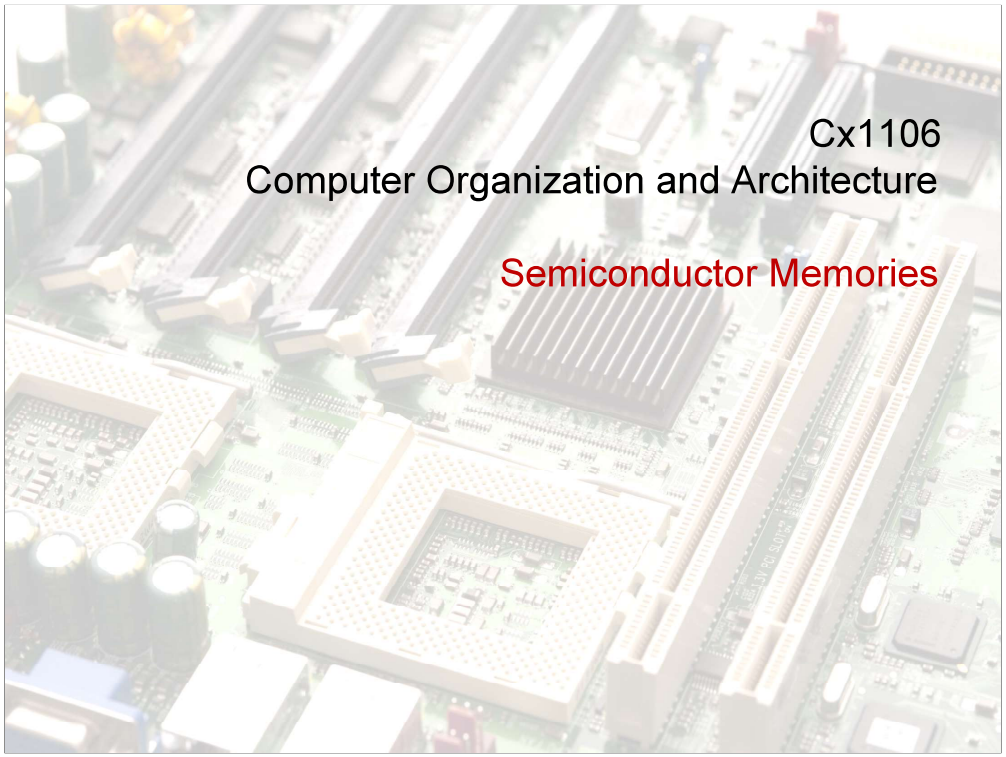
✓A. Magnetic HDD
B. Dynamic RAM
C. Static RAM
✓D. EEPROM
✓E. NAND Flash

**What is/are the effects of DRAM using single transistor cum capacitor design for storage?

✓A. Slower as memory needs to be refreshed
B. Simpler interface design as less transistors are used
✓C. Smaller layout compared to SRAM as less number of transistors are used
D. Larger layout compared to SRAM as capacitor occupies a larger area compared to transistors

Less transistors doesn't imply simpler interface design. DRAM requires more periodic refresh to maintain its data integrity, hence the interface controller has to include such refresh operation on top of the usual data transfer operation.

The larger number of transistors required to build one SRAM memory bit cell result in larger layout (area) needed compared to the DRAM, which only need one transistor + capacitor.



Cx1106
Computer Organization and Architecture

Semiconductor Memories

**Which of the following memory are volatile in nature?

A. Magnetic HDD
✓B. SRAM
✓C. DRAM
D. EEPROM
E. NOR Flash

**Which of the following memory devices are non-volatile?

✓A. Magnetic HDD
B. Dynamic RAM
C. Static RAM
✓D. EEPROM
✓E. NAND Flash

**What is/are the effects of DRAM using single transistor cum capacitor design for storage?

✓A. Slower as memory needs to be refreshed
B. Simpler interface design as less transistors are used
✓C. Smaller layout compared to SRAM as less number of transistors are used
D. Larger layout compared to SRAM as capacitor occupies a larger area compared to transistors

SRAM : at least 6 transistors
DRAM: 1 transistor

Less transistors doesn't imply simpler interface design. DRAM requires more periodic refresh to maintain its data integrity, hence the interface controller has to include such refresh operation on top of the usual data transfer operation.

The larger number of transistors required to build one SRAM memory bit cell result in larger layout (area) needed compared to the DRAM, which only need one transistor + capacitor.

**What is/are the difference between SRAM and DRAM if both use same process technology?

↗ cause of overhead due to refresh

- ✓ A. SRAM is faster than DRAM
- B. DRAM is faster than SRAM
- ✓ C. SRAM has a larger layout than DRAM
- D. SRAM uses capacitor as a storage element
- ✓ E. DRAM requires periodic refresh to keep its data integrity

Using same process technology means the transistor will be of similar size and speed.
[A] and [C] is a direct result of the SRAM/DRAM design.
[E] is an operation required to maintain data integrity of DRAM. SRAM design doesn't use capacitor so doesn't suffer from natural leakage of electrical charges, hence no need for any periodic refresh overhead operation.

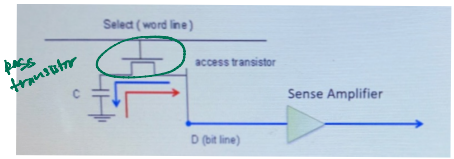
Why does DRAM requires constant refresh operation to keep its data integrity?

- A. DRAM is a type of volatile memory
- ✓ B. DRAM uses capacitor as its storage element
- C. DRAM uses less transistors compared to SRAM so less function build in
- D. DRAM is slower than SRAM so needs to periodically refresh its content

The charged capacitor in DRAM design has a natural leakage path to ground, so charges is lost over time and data integrity is compromised.

What is the function of the pass transistor in the DRAM Design?

- A. Store data
- ✓ B. Control flow of charges to and from the capacitor
- C. Control read/write operation mode
- D. Logics to charge and discharge the capacitor



DRAM design consist of one transistor and a capacitor. The capacitor stores a Logic '1' or '0' based of whether they are charged or discharged. To test the state of the capacitor, it needs to be connected to the bit line, the path to the bit line is gated by a transistor know as the "Pass Transistor". Check the lecture notes for graphical illustration.

What is the function of transistors in SRAM design?

- A. Store Data
- B. Control Read/Write mode
- ✓ C. Store Data and Control logic for signal transfer between SRAM cells and internal data lines
- D. Control logic for signal transfer between SRAM cells and internal data lines

SRAM design consist of six transistors. Four transistors implemented two NOT gates interlocked with each other, two transistors control to flow of charges in and out of the memory cell (the two NOT gates).

**Difference between EEPROM and Flash?

Flash cost less than EEPROM
⇒ due to overhead logic

- ✓A. EEPROM has a higher cost per bit
- ✓B. Flash erasure is faster
- ✓C. EEPROM has a smaller page size
- D. Flash has a smaller page size
- ✓E. Flash is available in larger capacity

Check the lecture notes for details on property of a EEPROM.

EEPROM is also based on Floating Gate Transistor but the design of the memory cells and the connection between the memory cells are different from Flash memory.

bigger block size → erase faster

Why is Flash memory able to retain its data after its power is removed?

↳ non-volatile

A. Flash memory is made up of transistors which is used to store data.
B. SRAM and Flash has different design and SRAM is ~~non~~-volatile.
✓C. The electrons stuck in the floating gate remain there as there is no electrical connection between the gate and to allow electron discharge.
D. Threshold voltage of Flash memory at programmed state is higher than that in erased state.

[A], [B] and [D] are true statement by itself but doesn't explain why Flash memory is non-volatile.

The two oxide layers that sandwich the floating gate are bad conductors of electricity so electrons will not get leaked out of the floating gate. As the electrons are retained, the attributes they induced (increase in Threshold Voltage) is also retained, meaning the data 'stored' is retained when power is off.

Which state of the floating gate transistor will yield a higher threshold voltage (Vt)?

- * A. No Electrons in the floating gate
- * B. Presence of Electrons in the floating gate
- ✓C. Electrons injection into the floating gate, reducing the gate's voltage potential.
- D. Electrons removal from the floating gate, increasing the gate's voltage potential

[A], [B]. Electrons are always present in all materials, electrical neutral material, the number of electrons is equal to the number of protons.
[D] Its correct that Electrons removal from FGT will increase the gate potential. But that will lead to the decrease in threshold voltage instead.
[C] Excess electrons in FGT reduces the gate potential, that means a larger positive voltage has to be applied at the gate in order to attract sufficient electrons to form the conductive layer that connects the drain and source terminal (turning ON the transistor).

What is the logic state of the flash content if the flash is 'Erased'?

can be user defined

A. Logic '0'
✓B. Logic '1'

It's a default definition adopted industry wide.

* A: all material have electrons → neutral cases have same number of electrons & protons → therefore, charges cancel out

* B: doesn't mean anything i need to be access electrons

Between NAND and NOR flash, which one has a higher cost per bit? Why?

A. NAND Flash. NAND gate is more complex than NOR gate.
B. NOR Flash. NOR gate is more complex than NAND gate.
C. NAND Flash. NAND layout has more wires, so more difficult to compact the layout.
✓D. NOR Flash. NOR layout has more wires, so more difficult to compact the layout.

Which Flash memory would you choose to build a 256GByte Solid-State Drive?

A. NOR Flash
✓B. NAND Flash
C. EEPROM
D. EPROM

NAND flash design result in lowest cost per bit compared to the other memory types listed here.

Which Non Volatile memory would you choose for the system memory of a processor system?

✓A. NOR Flash
B. NAND Flash
C. EEPROM

System memory is the memory where processor execute the program code directly during run time.
In order to support direct code execution, the memory needs to support Execute in Place (XIP).
Of the three memory types listed, only NOR flash supports XIP.

Which memory would you choose for the system memory of a smart phone ?

A. NOR Flash
B. NAND Flash
✓C. DRAM
D. SRAM

System memory of a smart phone runs a complex OS with many user applications. Depending on what applications is required and the use case, these applications has to be loaded and removed from the system memory. A lot of data has to be read, modified and re-written back to the system memory.
The abundance of read/write operations means Flash memory is not suitable due to finite erase cycle limit.

Between DRAM and SRAM, DRAM is chosen as the amount of system memory required by smartphone OS (iOS or Android) requires Gbytes of memory. SRAM is too costly with such memory capacity.

Which memory would you choose for the cache memory of a micro-processor?

A. NOR Flash
B. NAND Flash
C. DRAM
✓D. SRAM

Cache memory is a SMALL (Kbytes, Mbytes) and FAST memory used to boost the overall performance of a computer system.
A key requirement is FAST so SRAM is an ideal choice.
Since size is small so absolute cost increase is not as significant even if SRAM is used.
Non volatile memory is not suitable since there are a lot of read/write operations involved.

Which memory would you choose to store user configuration of a smart phone?

✓A. NAND Flash
B. DRAM
C. SRAM

User configuration needs to be retained when power is OFF.
NAND flash is the only non-volatile memory in the list.

What do we mean when we say a memory supports Execute-in-Place (XIP)? How does NOR flash support XIP?

A. XIP means the code could be execute directly from the memory itself without having to transfer into the internal RAM. NOR flash support XIP because the cells behaves like a NOR gate.

✓B. XIP means the code could be execute directly from the memory itself without having to transfer into the internal RAM. NOR flash supports XIP because it allow random reading of its content with only address information supplied.

C. XIP means data could be executed directly from the memory itself without having to transfer into the internal RAM. NOR flash supports XIP because it allow random reading of its content with only address information supplied.

D. XIP means data could be executed directly from the memory itself without having to transfer into the internal RAM. NOR flash support XIP because the cells behaves like a NOR gate.

The memory cells behaving like a NOR gate has nothing to do with NOR flash being XIP capable.

'Data' cannot be executed. Only 'Code' can be executed.