Memory Technologies

Paolo PRINETTO
Paolo.Prinetto@polito.it

Elena Ioana VATAJELU
Elena.Vatajelu@imag.fr

Rel. 26.03.2019

License Information

This presentation is licensed under the Creative Commons BY-NC License

To view a copy of the license, visit:
http://creativecommons.org/licenses/by-nc/3.0/legalcode

Disclaimer

- We disclaim any warranties or representations as to the accuracy or completeness of this material.
- Materials are provided “as is” without warranty of any kind, either express or implied, including without limitation, warranties of merchantability, fitness for a particular purpose, and non-infringement.
- Under no circumstances shall we be liable for any loss, damage, liability or expense incurred or suffered which is claimed to have resulted from use of this material.

Outline

- Memory device *yesterday*
  - ROM Memories
- Memory device *today*
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- Memory device *tomorrow*
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

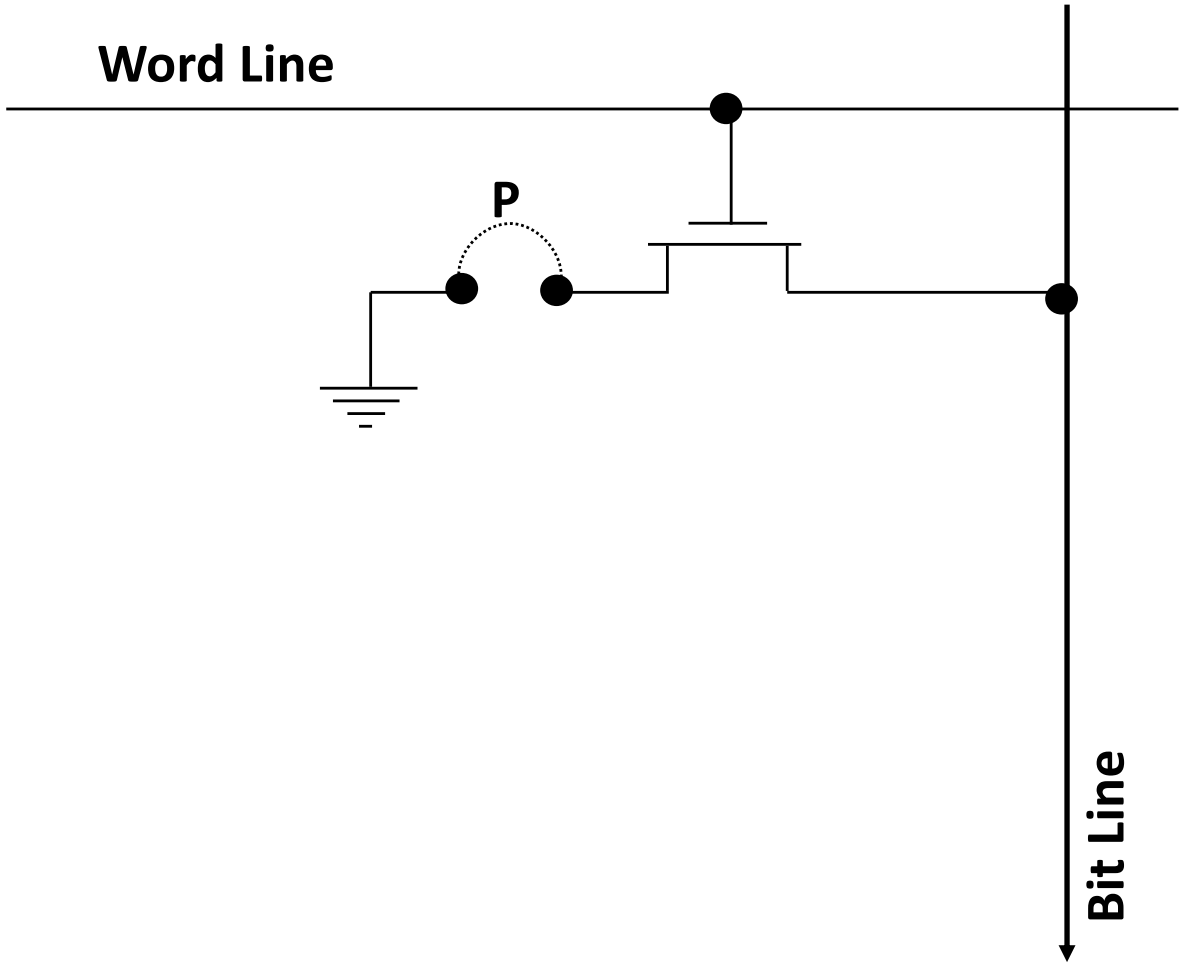
Outline

- Memory device *yesterday*
  - ROM Memories
- Memory device *today*
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- Memory device *tomorrow*
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

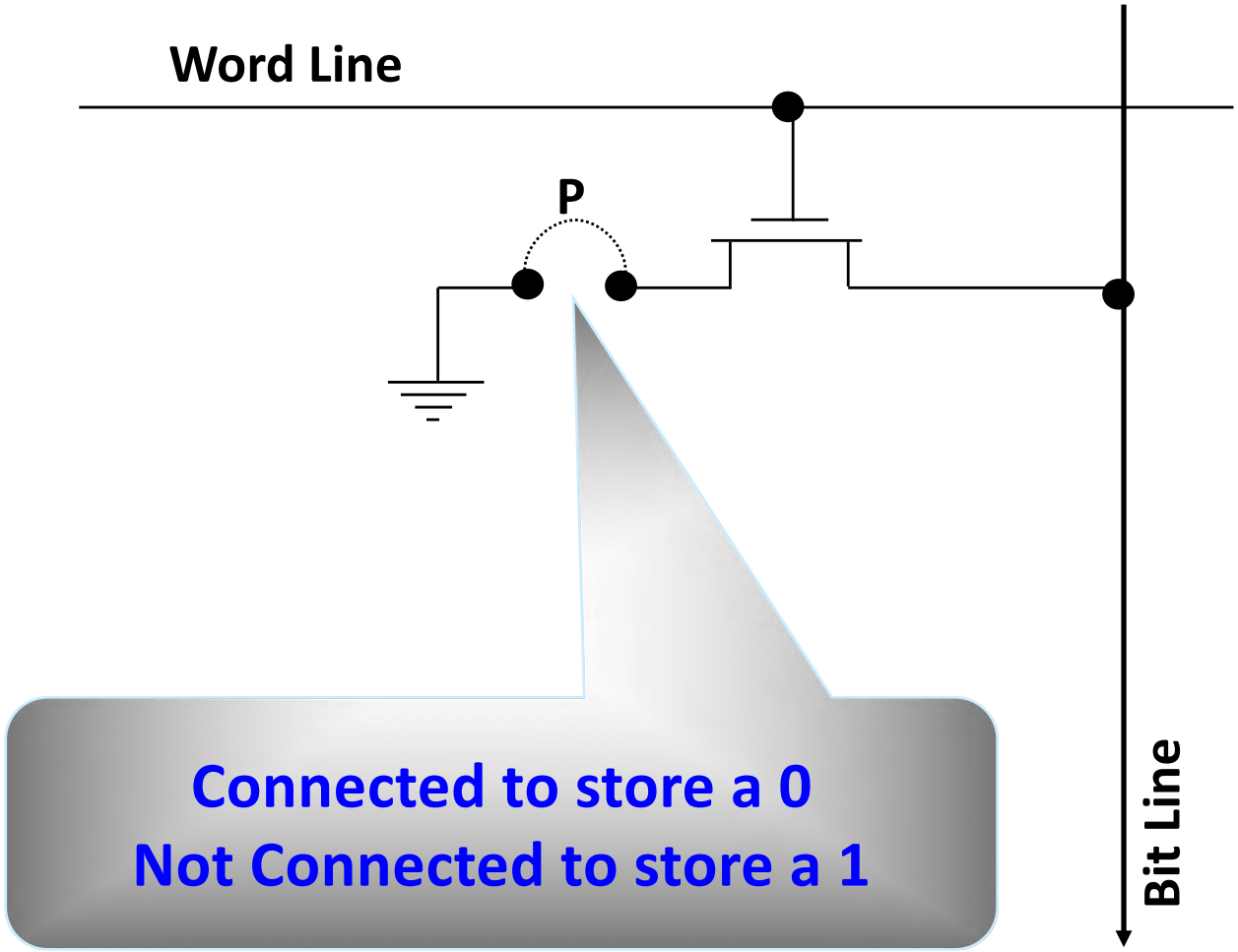
ROM memories

- *Read-Only Memory (ROM)*
  - Information stored as the presence or absence of transistor at manufacturing time
  - Information remains even when power is shut off
- *Erasable, Programmable Read-Only Memory (EPROM)*
  - Can be programmed in the field
  - Can be fully erased by applying ultraviolet light (~20 min)
- *Electrically Erasable, Programmable Read-Only Memory (EEPROM)*
  - Words can be selectively erased by electrical means

ROM Cell



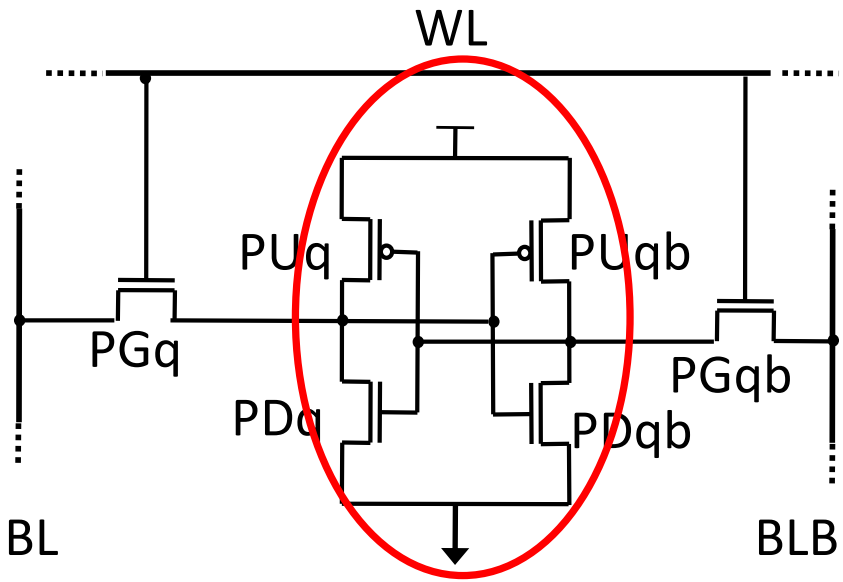
ROM Cell



Outline

- Memory device *yesterday*
  - ROM Memories
- **Memory device *today***
  - **SRAM – Static Random Access Memories**
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- Memory device *tomorrow*
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

SRAM



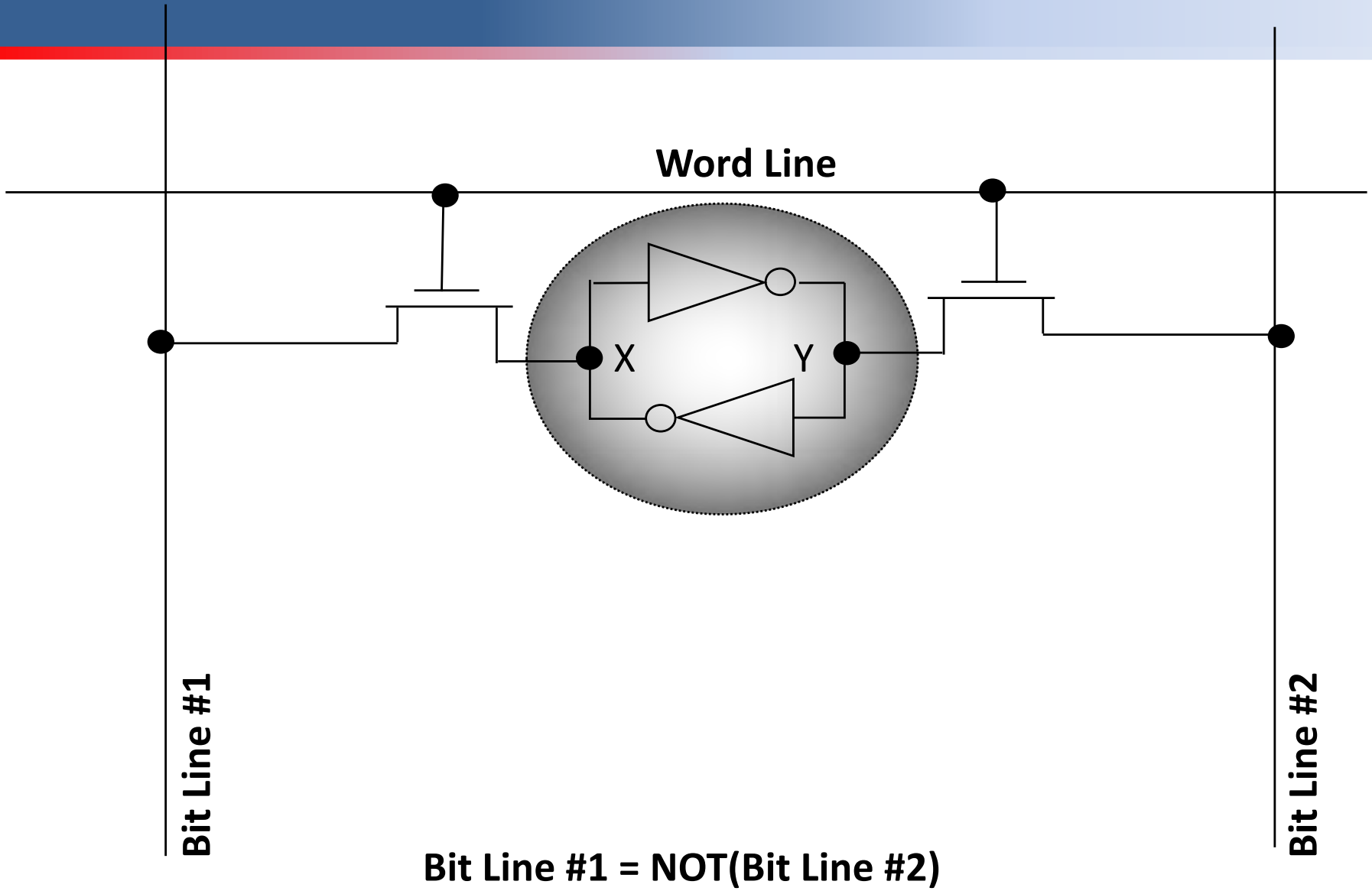
PG – Pass-gate transistor
PU – Pull-up transistor
PD – Pull-down transistor
WL – Word line
BL – Bit line

- ☹ Cell size $120F^2$
- ☹ Scalability
- 😀 Access Time <1ns
- 😐 Power Consumption
- 😀 Endurance $>10^{15}$
- ☹ Data Reliability
- Yes Volatility
- ☹ Fabrication Cost

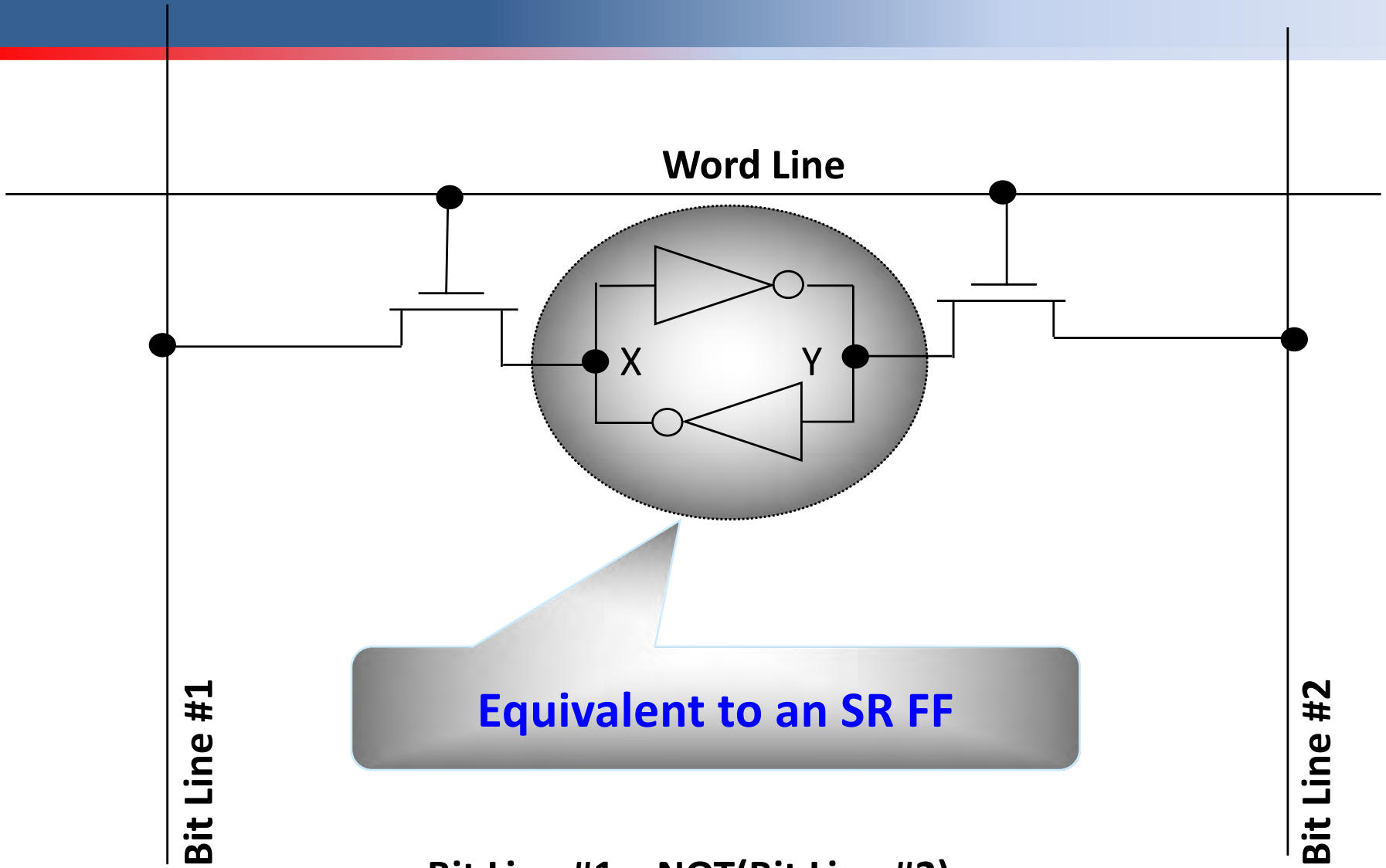
SRAM Memories

- A SRAM cell is composed of a flip-flop (formed by 2 cross-coupled inverters), connected to the Bit Lines via two pass transistors
- Once the bit-cell stores a bit, it keeps that value until the opposite value is written in it.

SRAM cell

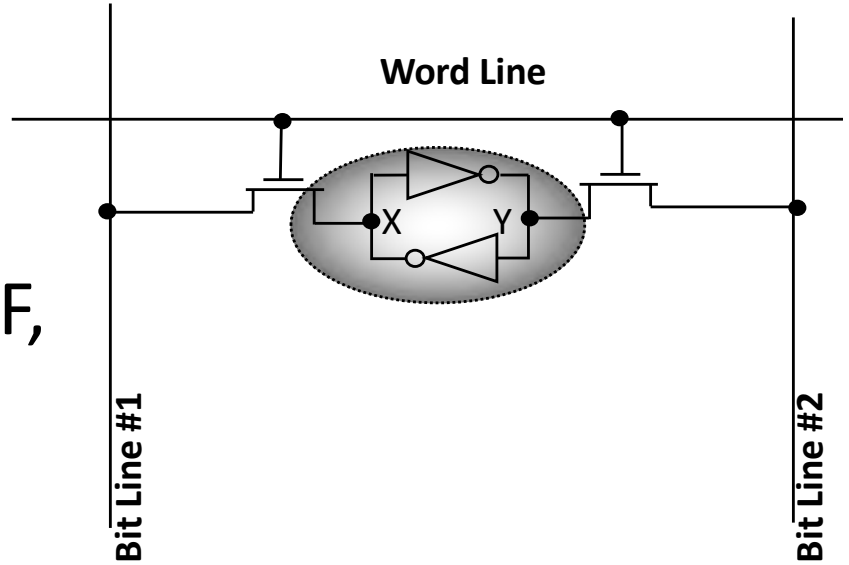


SRAM cell



Basic Principles

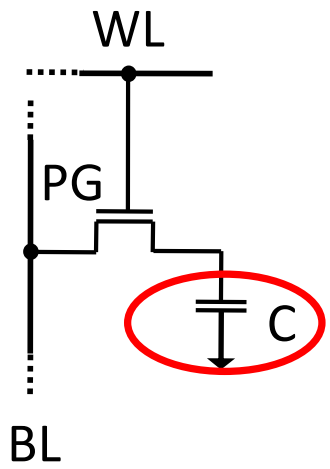
- When the corresponding Word Line is not asserted the FF (i.e., the cell) holds its value
- To access a cell, the corresponding Word Line must be asserted. Then one can:
  - *Read* the (opposite) valued forced by the FF on the two Data Lines (Read operation)
  - *Write* a new value into the FF, by forcing opposite values on the two Data Lines.



Outline

- Memory device *yesterday*
  - ROM Memories
- **Memory device *today***
  - SRAM – Static Random Access Memories
  - **DRAM – Dynamic Random Access Memories**
  - Flash Memories
  - SSD
  - Comparisons
- Memory device *tomorrow*
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

DRAM



PG – Pass-gate transistor
C – Data storage capacitor
WL – Word line
BL – Bit line

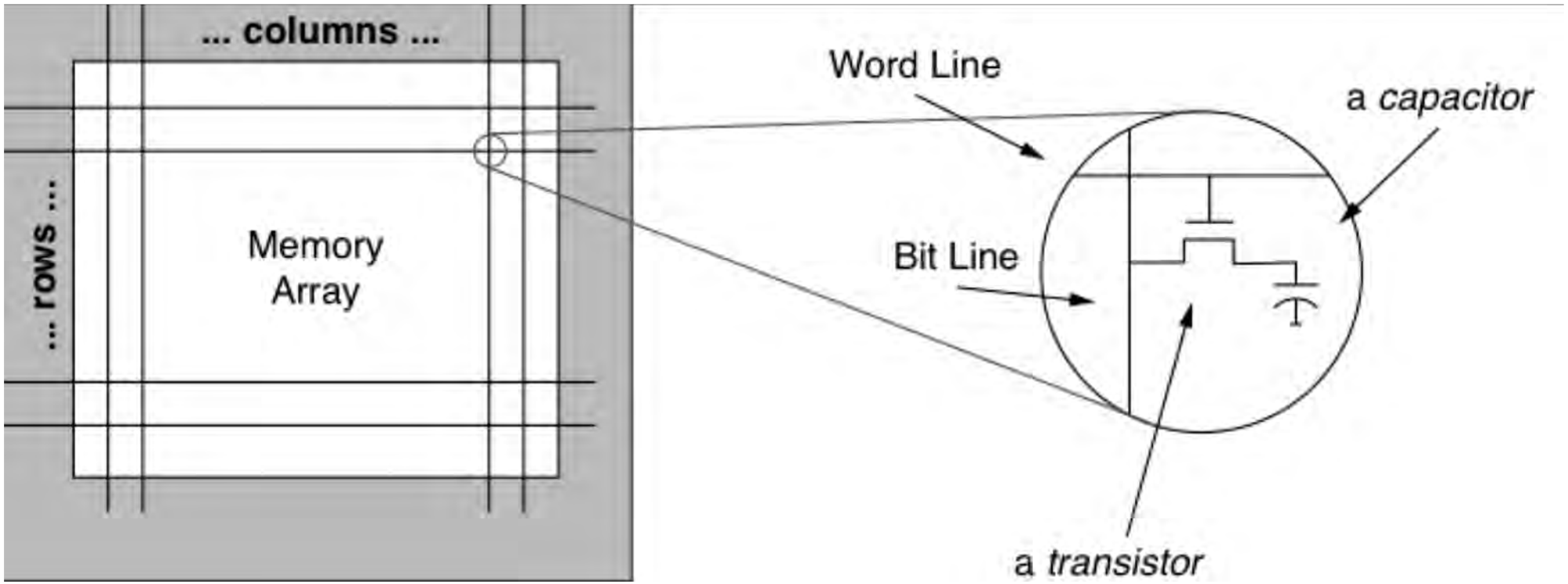
- Cell size — $4 – 6F^2$
- Scalability
- Access Time — *20ns*
- Power Consumption
- Endurance — $>10^{15}$
- Data Reliability
- Yes — Volatility
- Fabrication Cost

Rel. 26.03.2019

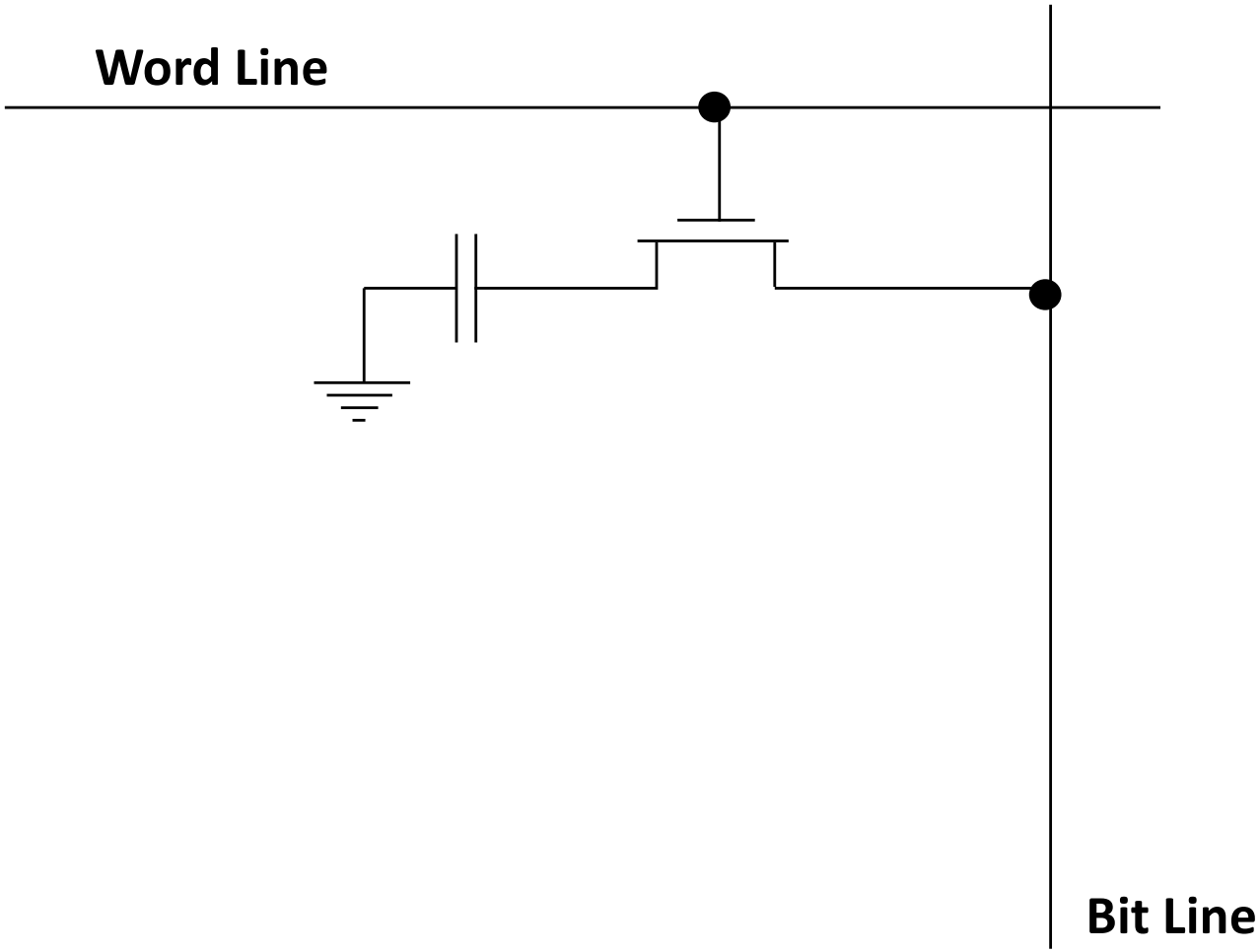
DRAM

- Dynamic RAM (DRAM) stores each bit in an electrical capacitor (implemented in CMOS as a gate capacity), using a transistor as a switch to charge or discharge the capacitor.
- Information (data) is stored as a charge (i.e., voltage at capacitor's pin).

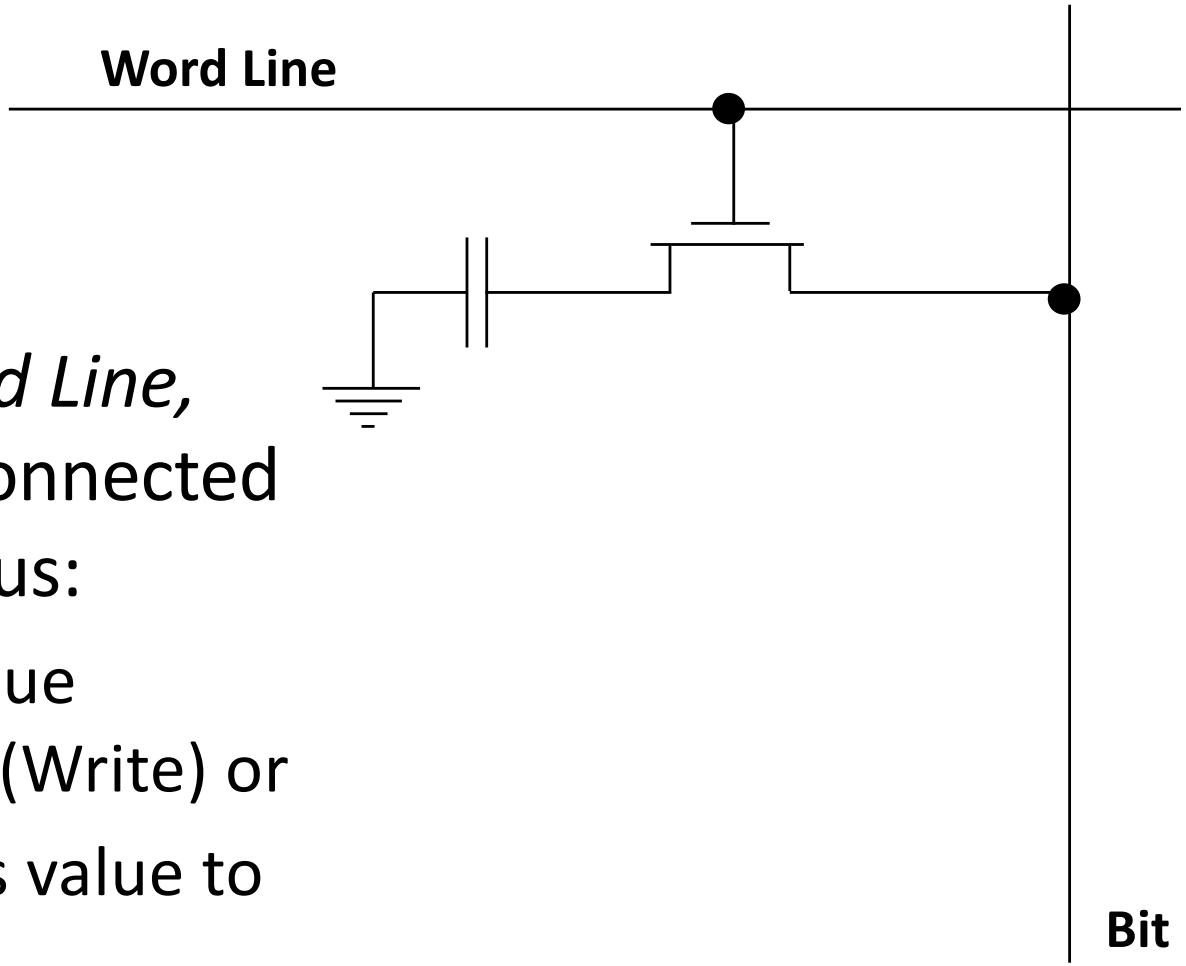
DRAM Cell



DRAM Cell



DRAM Cell



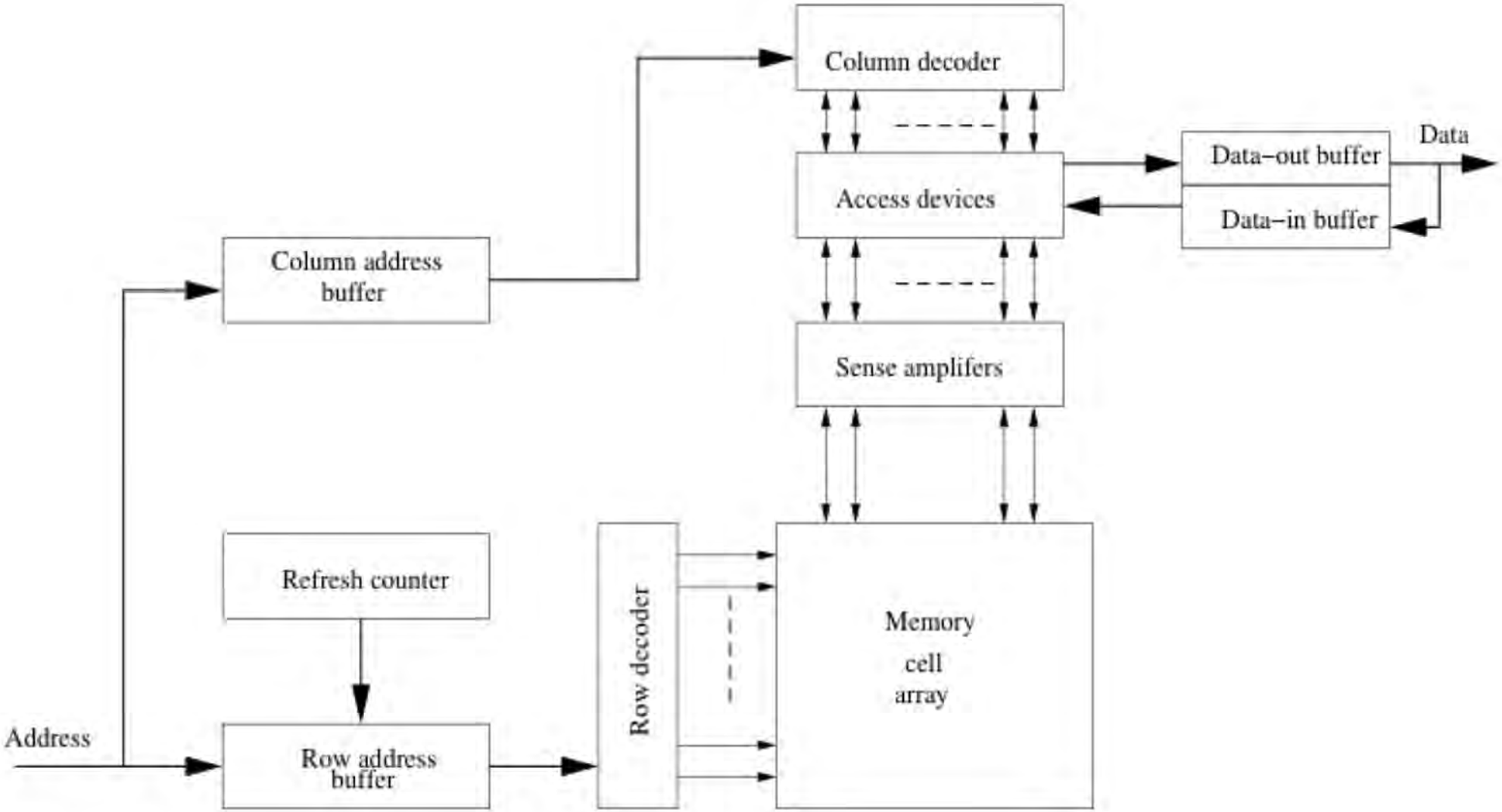
Enabling the *Word Line,* the capacitor is connected to the *Bit Line,* thus:

- It gets the value present on it (Write) or
- It transfers its value to it (Read)

Refresh

- Capacitor discharges even when powered-up, thus loosing the stored data
- To preserve stored data, a “refresh” mechanism is needed
- It consist in reading and writing back the stored value before it disappears
- Implemented by ad-hoc hardware, transparent to the final user
- Refresh operations take just few % of memory timing and have highest priority

DRAM Architecture

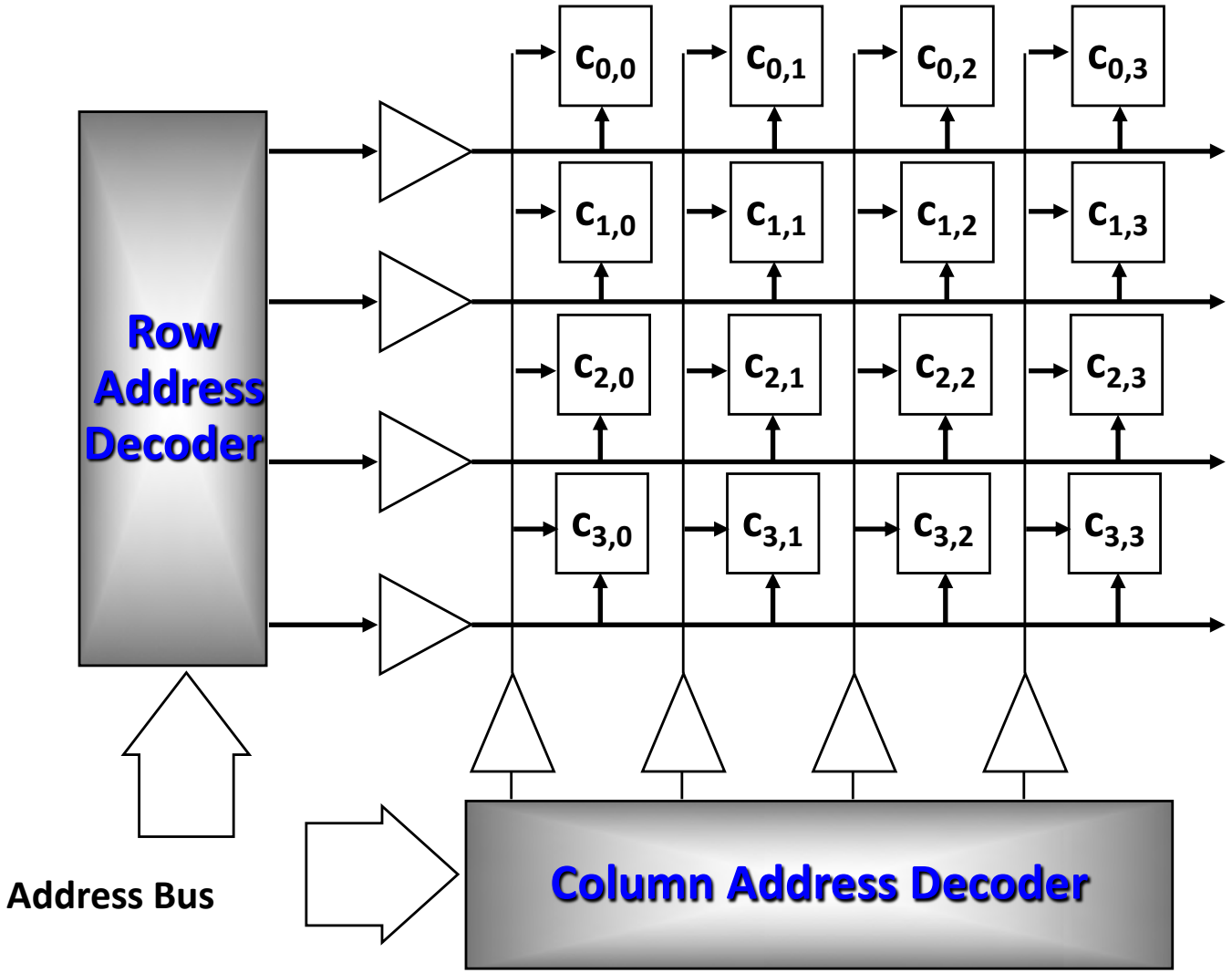


Matrix Organization

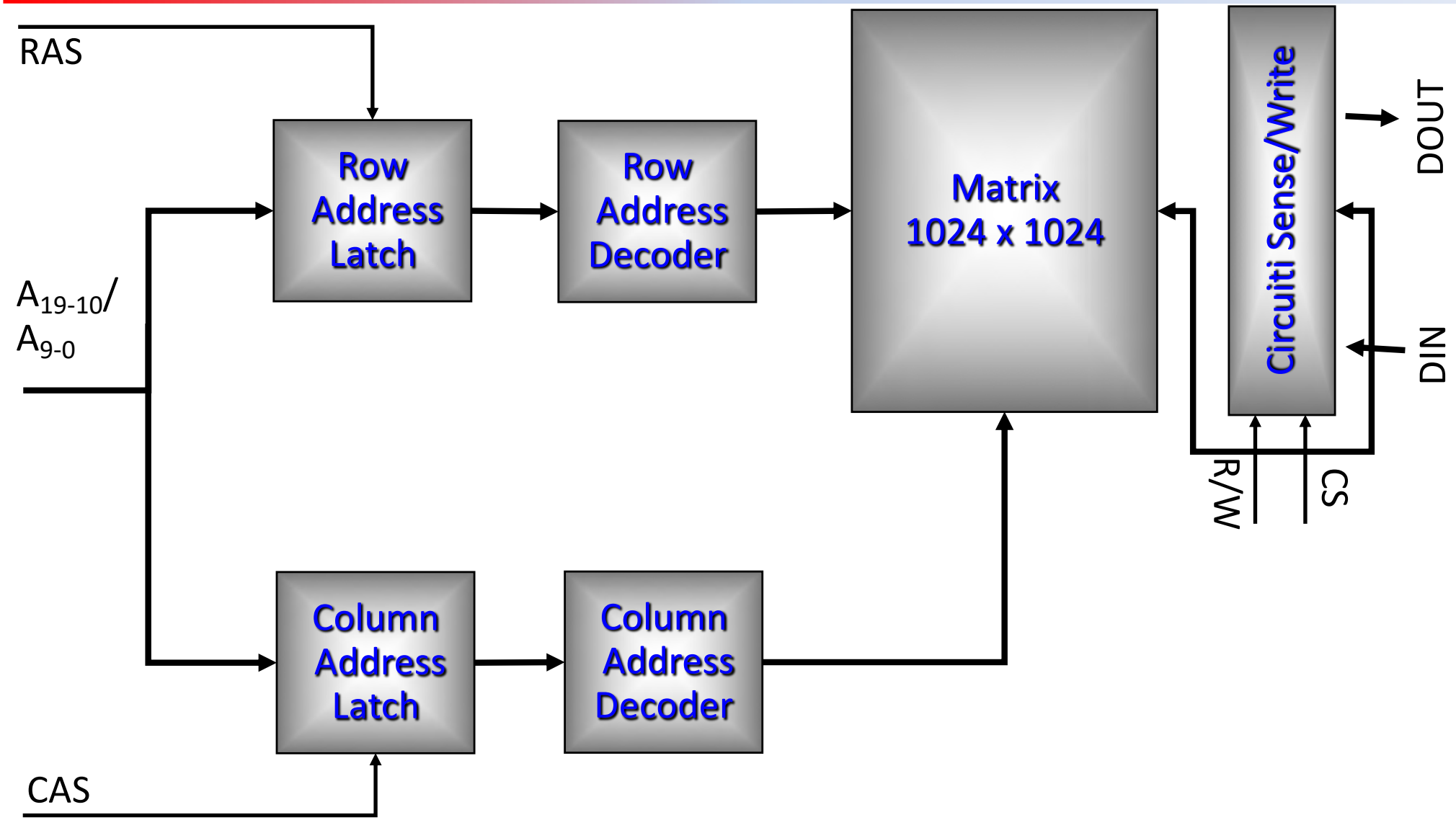
- Address is split in 2 parts:
  - Row Address
  - Column Address

  to select, respectively, the row and the column the selected cell belongs to
- In some cases, to minimize address bus width, the 2 parts are send sequentially:
  - Row Address first, controlled by the *RAS (Row Address Strobe)* signal
  - Column Address second, controlled by the *CAS (Column Address Strobe)* signal

Matrix Organization

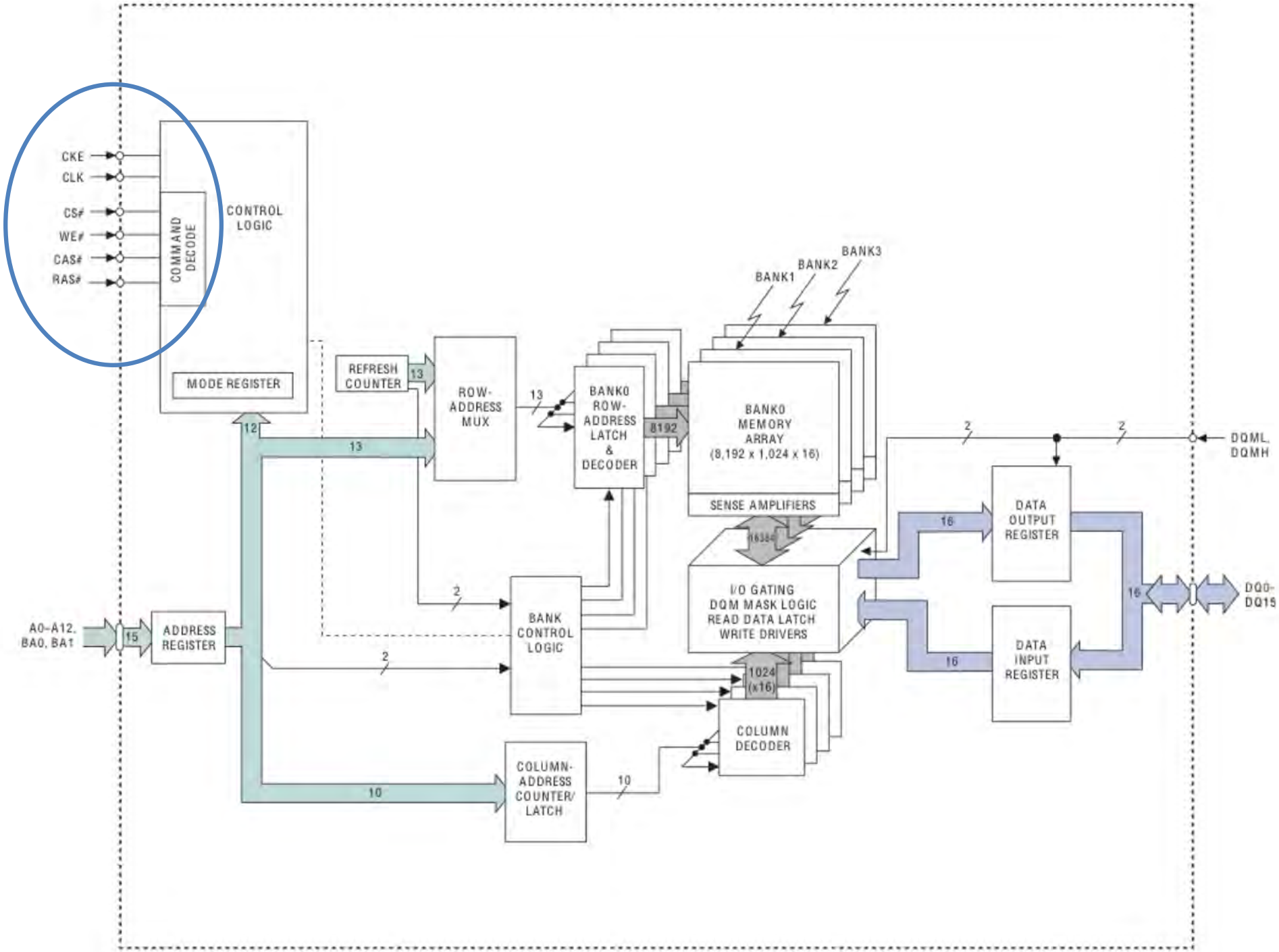


Organization (example)

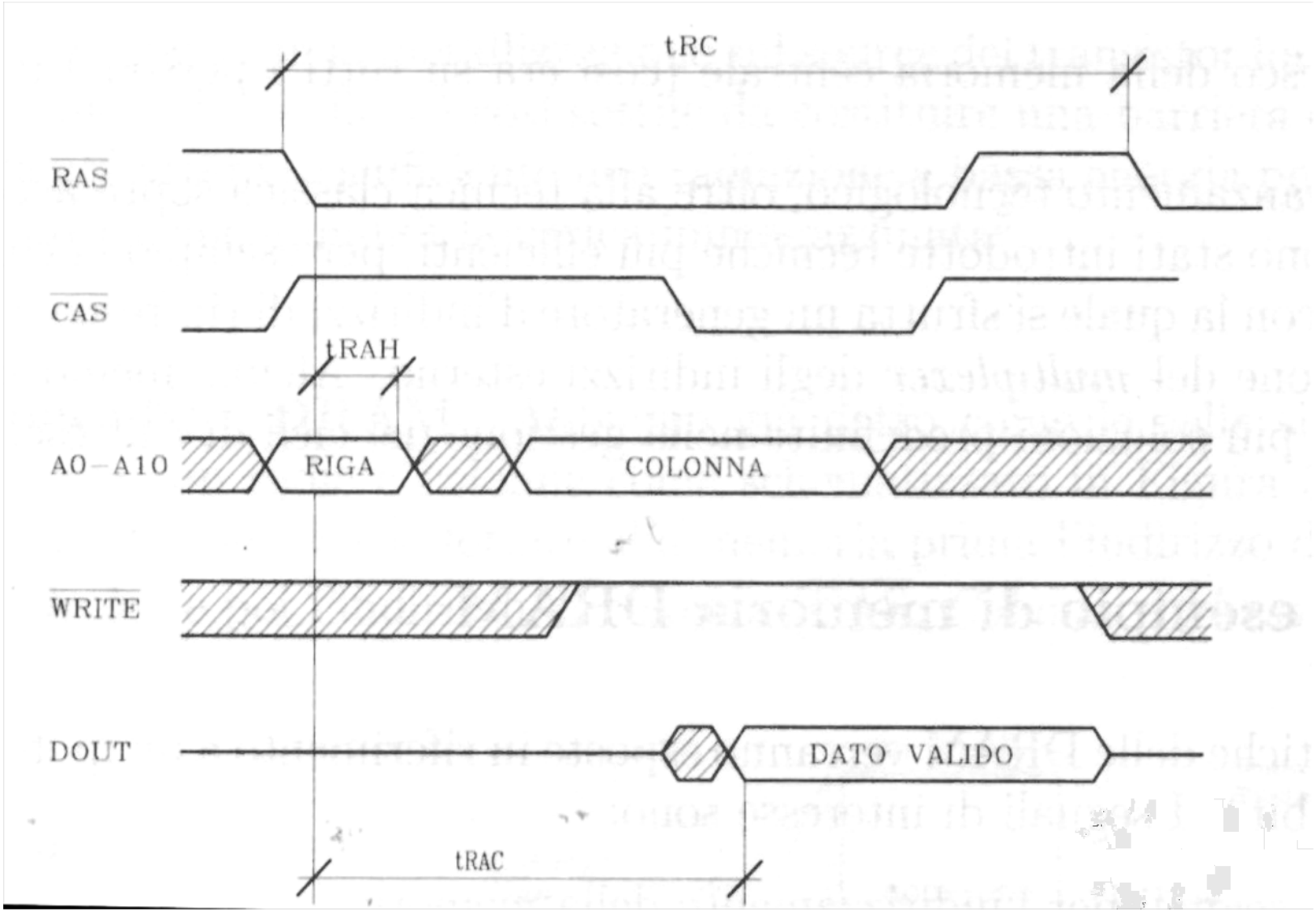


Example: DRAM AS4C1M16E5 (Alliance Semiconductor)

- 1 M × 16 bits
- Organized as a square matrix
- Pins:
  - A0 – A9: Address
  - RAS, CAS (active low)
  - DQ1 – DQ16: data
  - WE: Write Enable (active low)
  - OE: Output Enable (active low)



Timing



Read operation

- Put row address on A0 – A9
- Assert RAS
- Wait for an interval time = t_{RAH}
- Put column address on A0 – A9
- Assert CAS
- After a t_{RAC} interval, data are available on DQ

Note

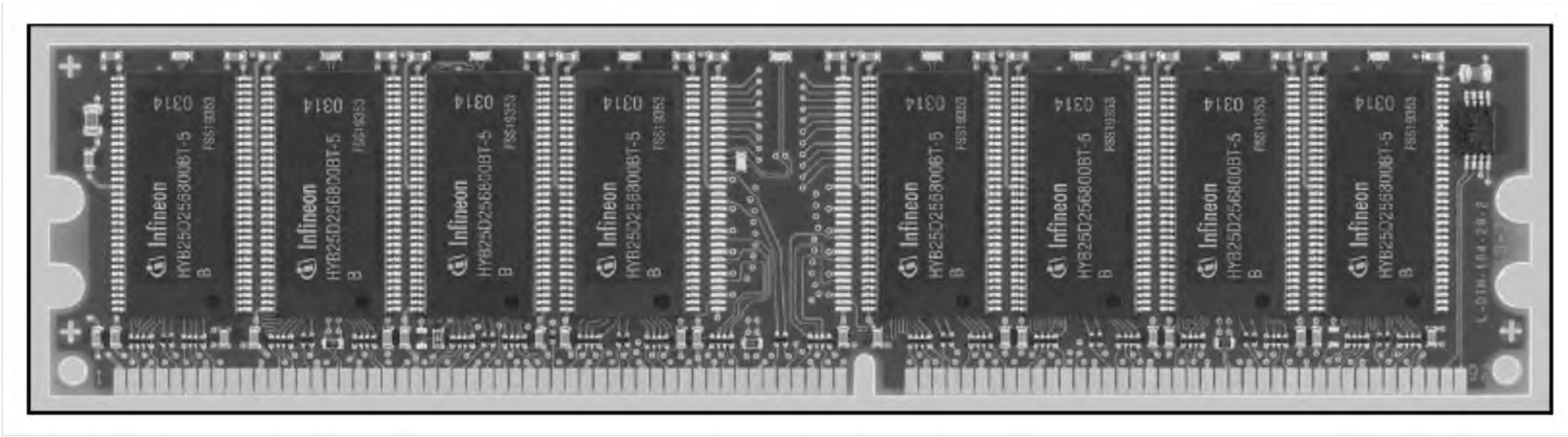
- To start a new Read or Write cycle one has to wait for an interval = t_{RC}
- Every 16 ms a complete refresh must be done: it means executing 1,024 operations with increasing row address

DRAM Modules

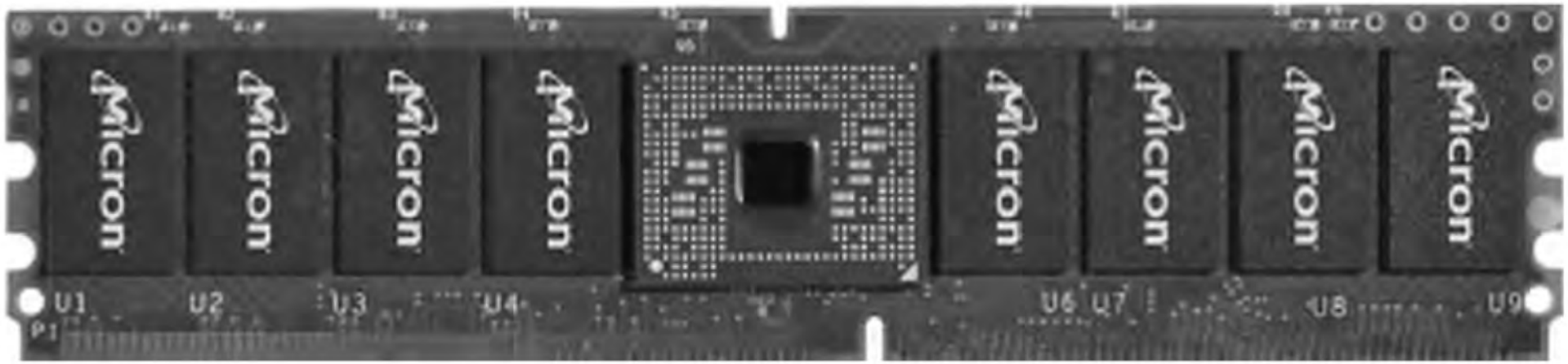
- PC users usually deal with *DRAM Modules* (often referred to as *DRAM card* or *DRAM stick*, as well), i.e., with small boards to be connected to the buses

SIMM module

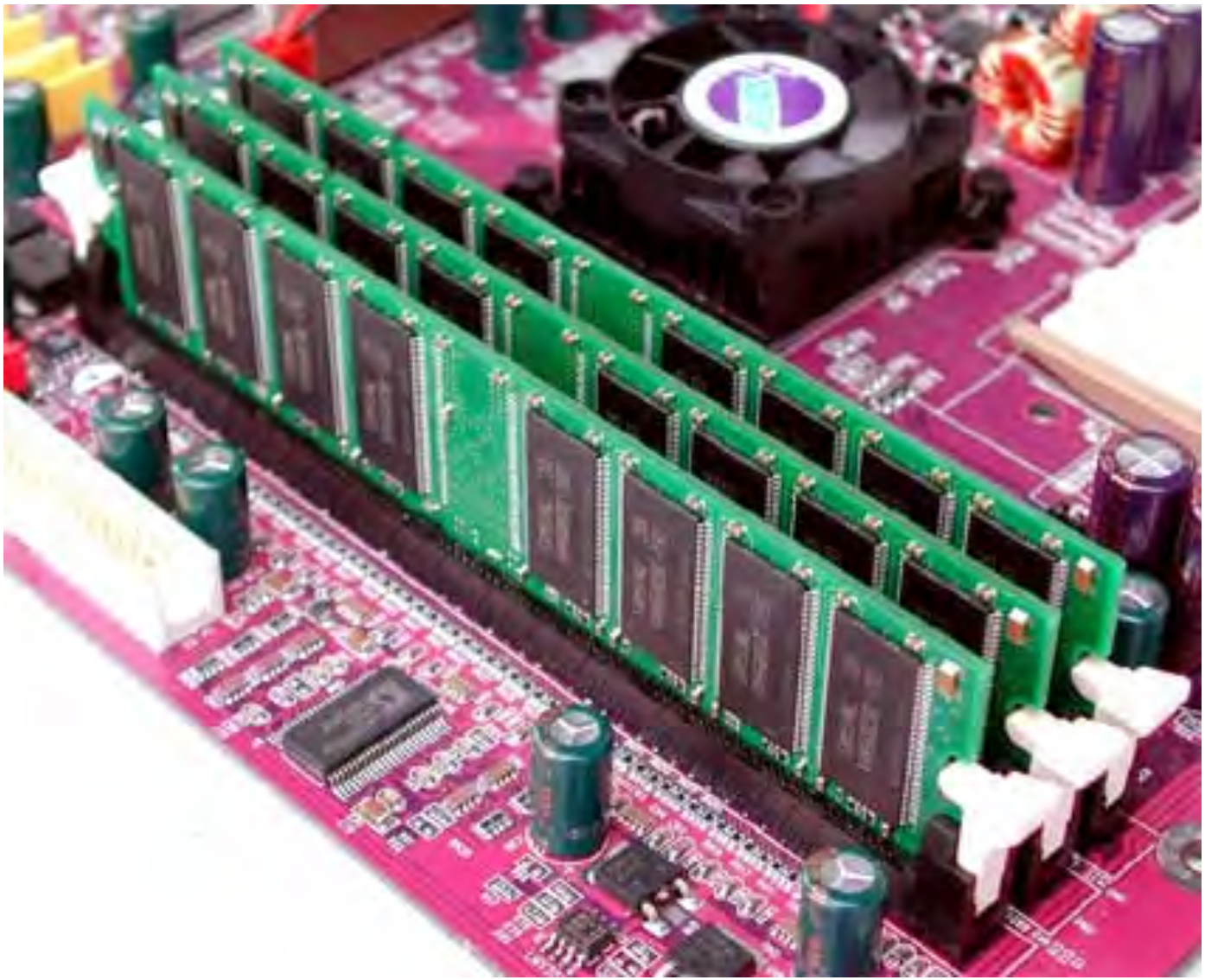
DIMM module of DRAM TSOP memories (Infineon)



DIMM module of DRAM TBGA memories (Micron)



Example of DIMMs on a PC board



DRAM extensions

- *FPM DRAM (Fast Page Mode DRAM)*
  While standard DRAM requires that a row and column be sent for each access, FPM works by sending the row address just once for many accesses to memory in locations near each other, improving access time.

DRAM extensions

- *EDO DRAM (Extended Data-Out DRAM)*
  A type of DRAM that is faster than conventional DRAM.

  Unlike conventional DRAM which can only access one block of data at a time, EDO RAM can start fetching the next block of memory at the same time that it sends the previous block to the CPU.

DRAM extensions

- *SDRAM (Synchronous DRAM)*

  DRAM synchronized with the system bus.

  Classic DRAM has an asynchronous interface, which means that it responds as quickly as possible to changes in control inputs.

  SDRAM has a synchronous interface, meaning that it waits for a clock signal before responding to control inputs and is therefore synchronized with the computer's system bus.

DRAM extensions

The clock is used to drive an internal finite state machine that pipelines incoming commands.

The data storage area is divided into several banks, allowing the chip to work on several memory access commands at a time, interleaved among the separate banks.

This allows higher data access rates than an asynchronous DRAM.

DRAM extensions

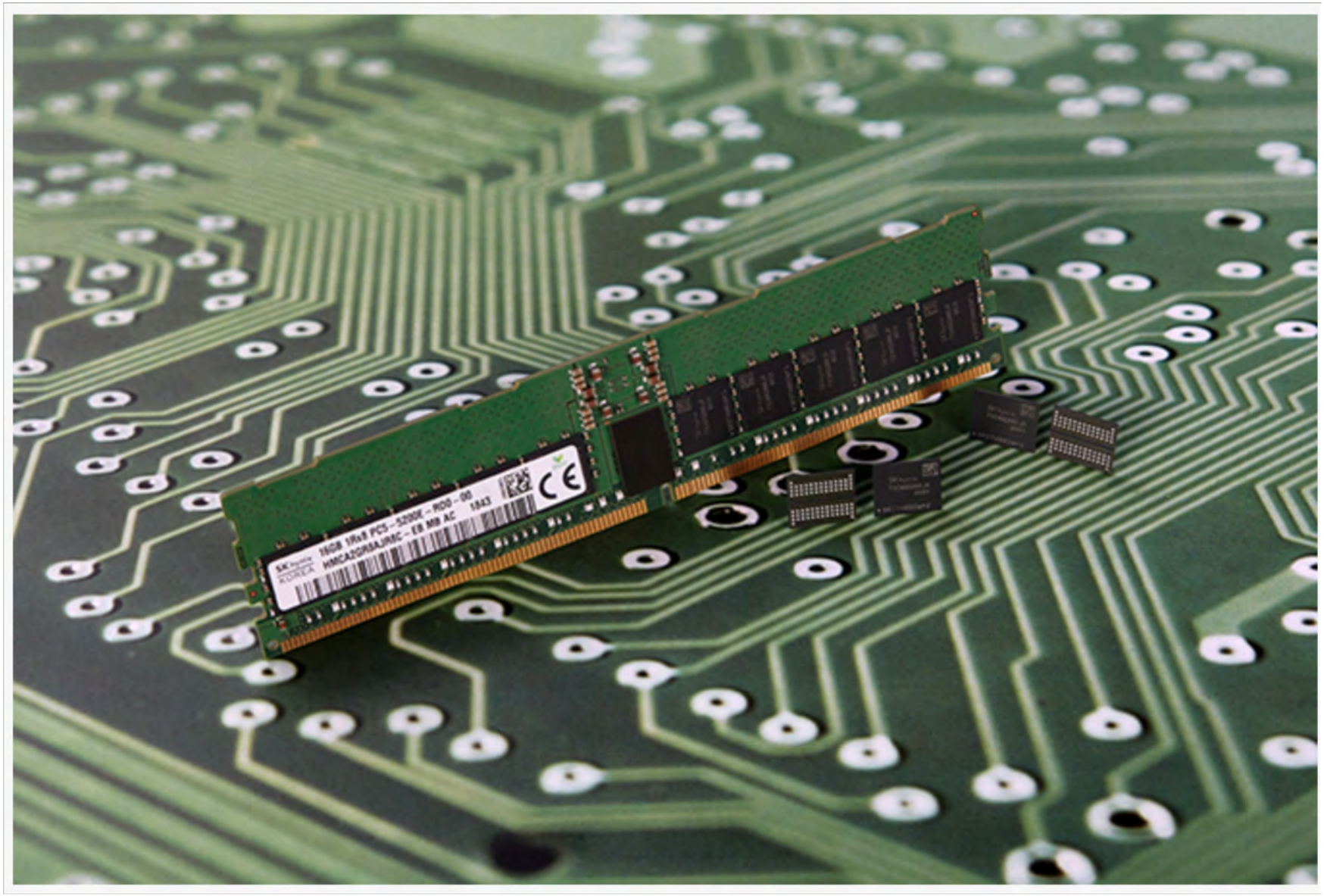
- *DDR SDRAM (Double Data Rate Synchronous DRAM)*

  They hit the mainstream computer market around 2002.

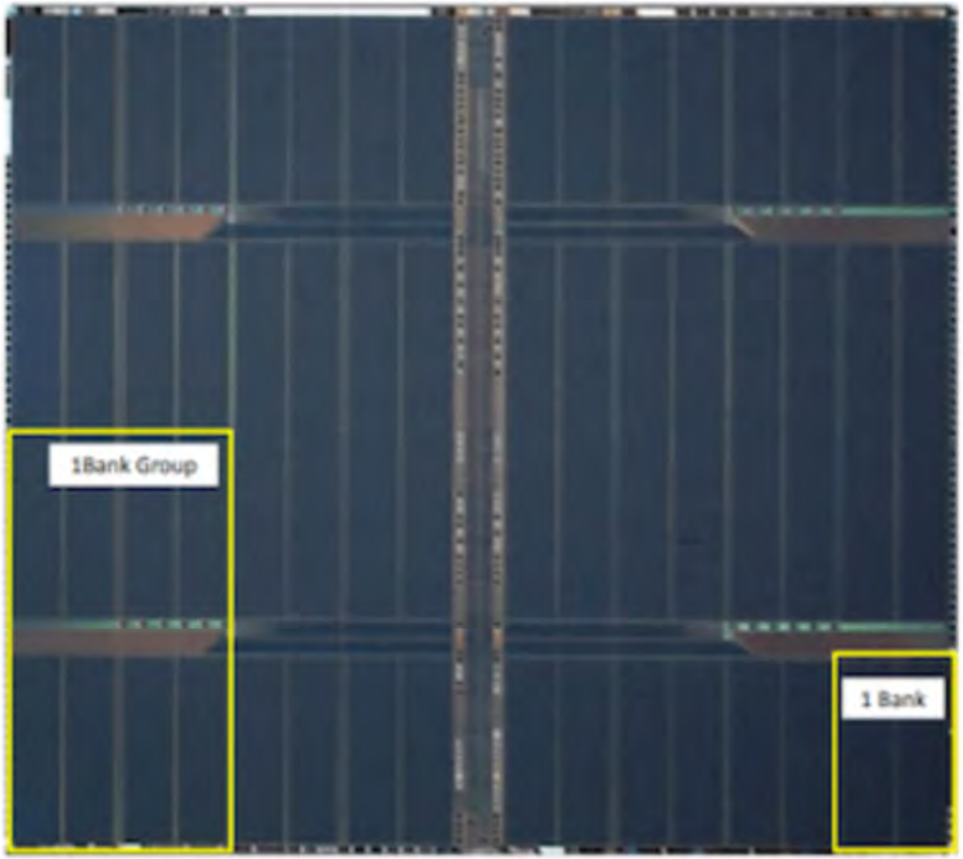
  They reads data on both the rising and falling edges of the clock signal, enabling a DDR memory module to transfer data twice as fast as an SDR memory module

- *DDR5 today available*

SK Hynix DDR5-6400

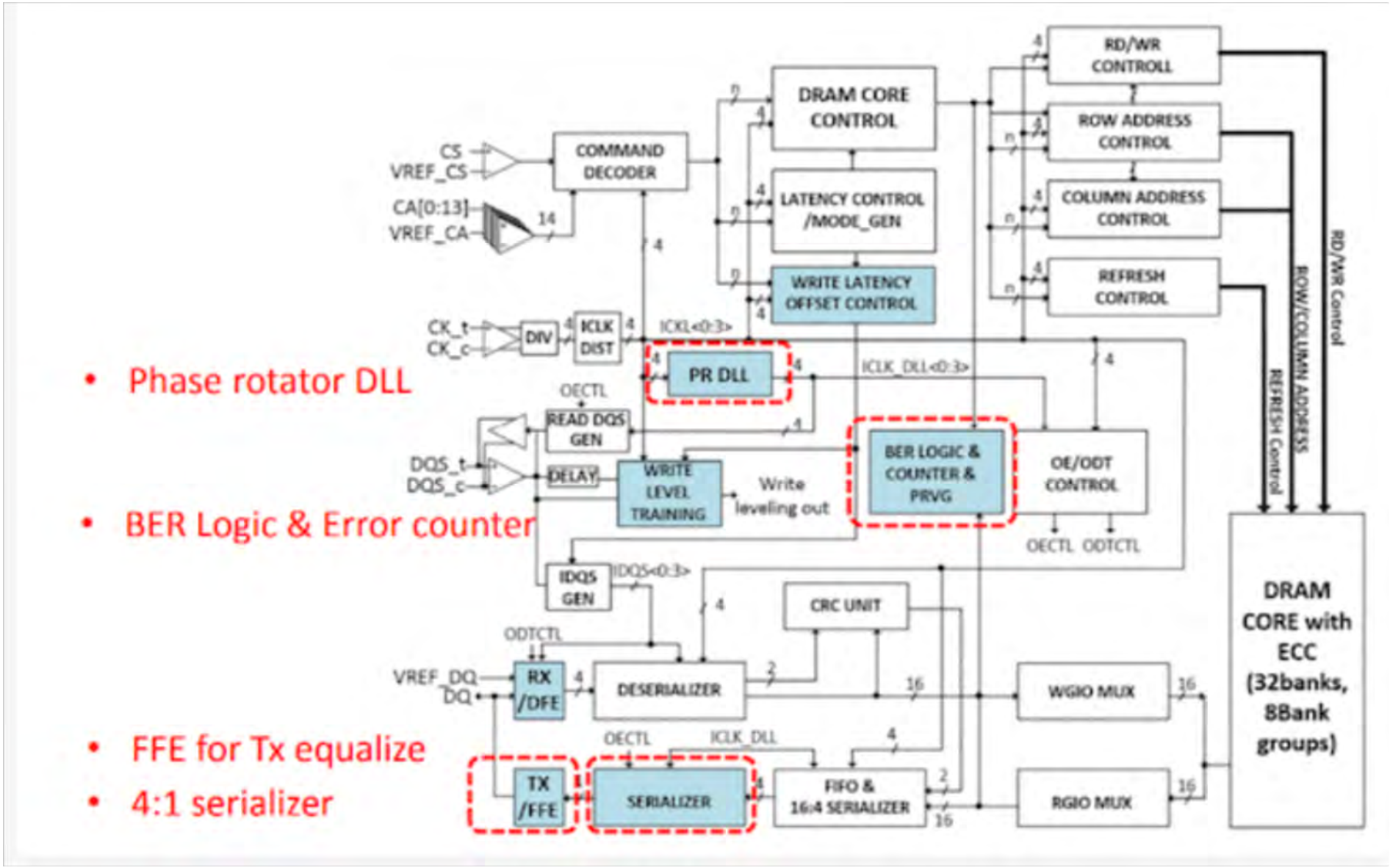


SK Hynix DDR5-6400

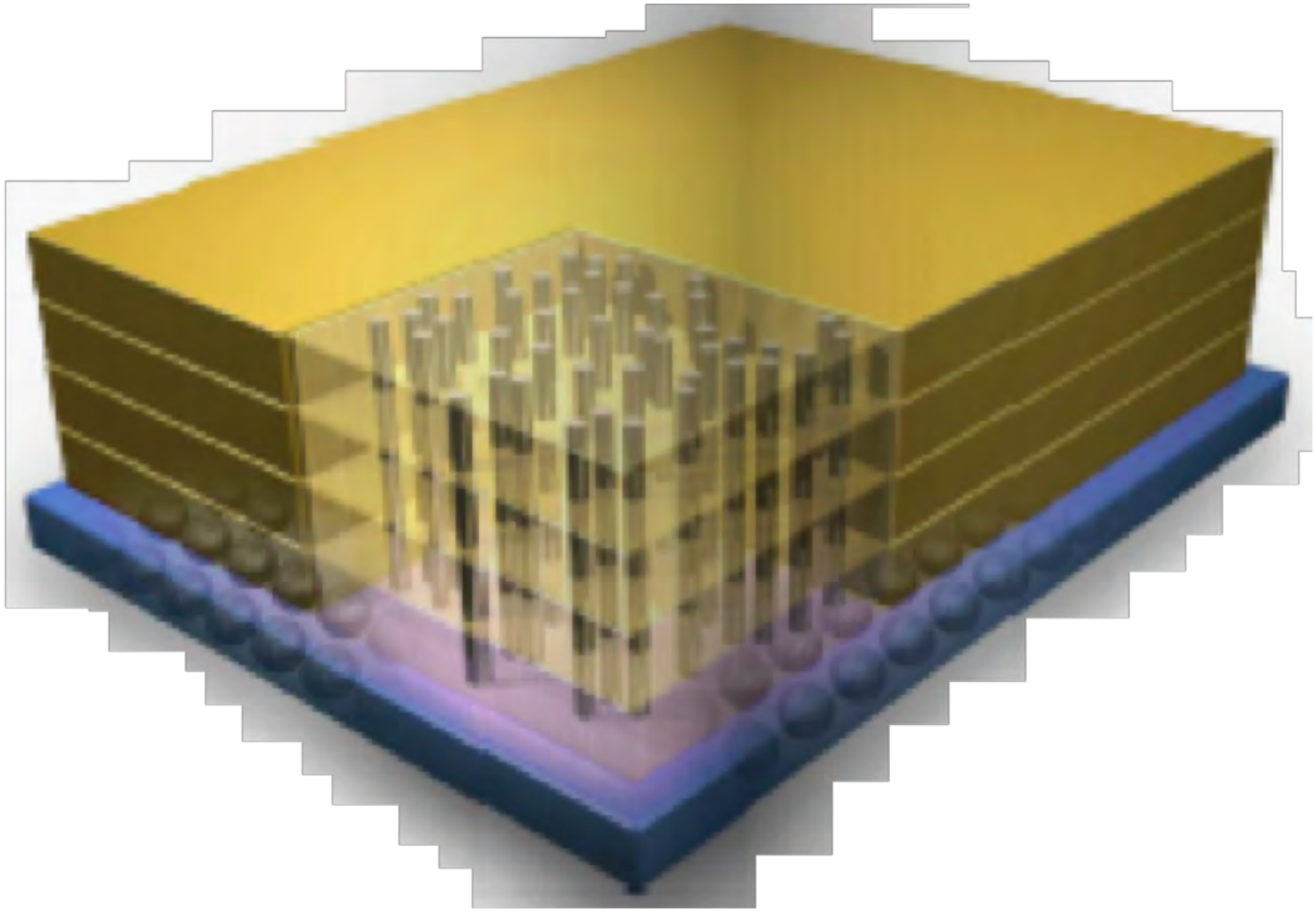


Chip summary	
Technology	1y-nm 4-metal DRAM process
Data Rate	6.4 Gbps/pin
IO configuration	X4/x8/x16
Burst length	BC8, BL16 on-the-fly
Chip area	76.22mm²
Supply voltage	VDD/VDDQ 1.1V, VPP 1.8V
RAS feature	In DRAM ECC
Equalize scheme	DFE/FFE

SK Hynix DDR5-6400



Hybrid Memory Cube

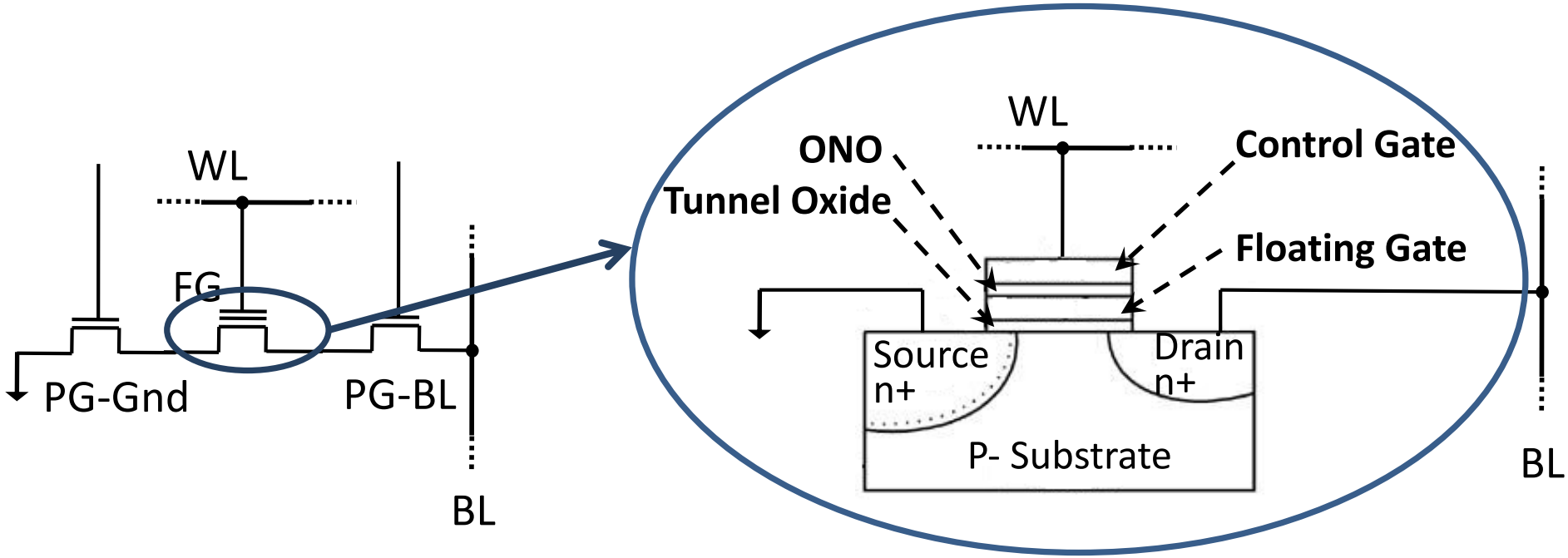


http://hybridmemorycube.org/technology.html

Outline

- Memory device *yesterday*
  - ROM Memories
- **Memory device *today***
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - **Flash Memories**
  - SSD
  - Comparisons
- Memory device *tomorrow*
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

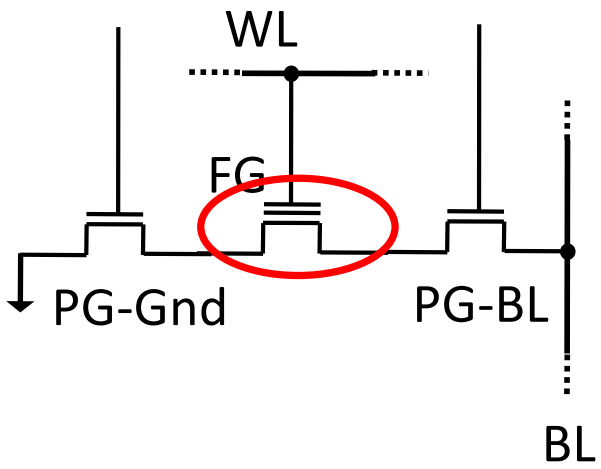
Flash



PG-Gnd – Ground select transistor
PG-BL – BL select transistor
FG – Floating gate transistor
BL – Bit line
WL – Word line

ONO – Oxide-Nitride-Oxide
SLC – single-level cell – stores one bit per cell
MLC – multi-level cell - stores more than one bit per cell

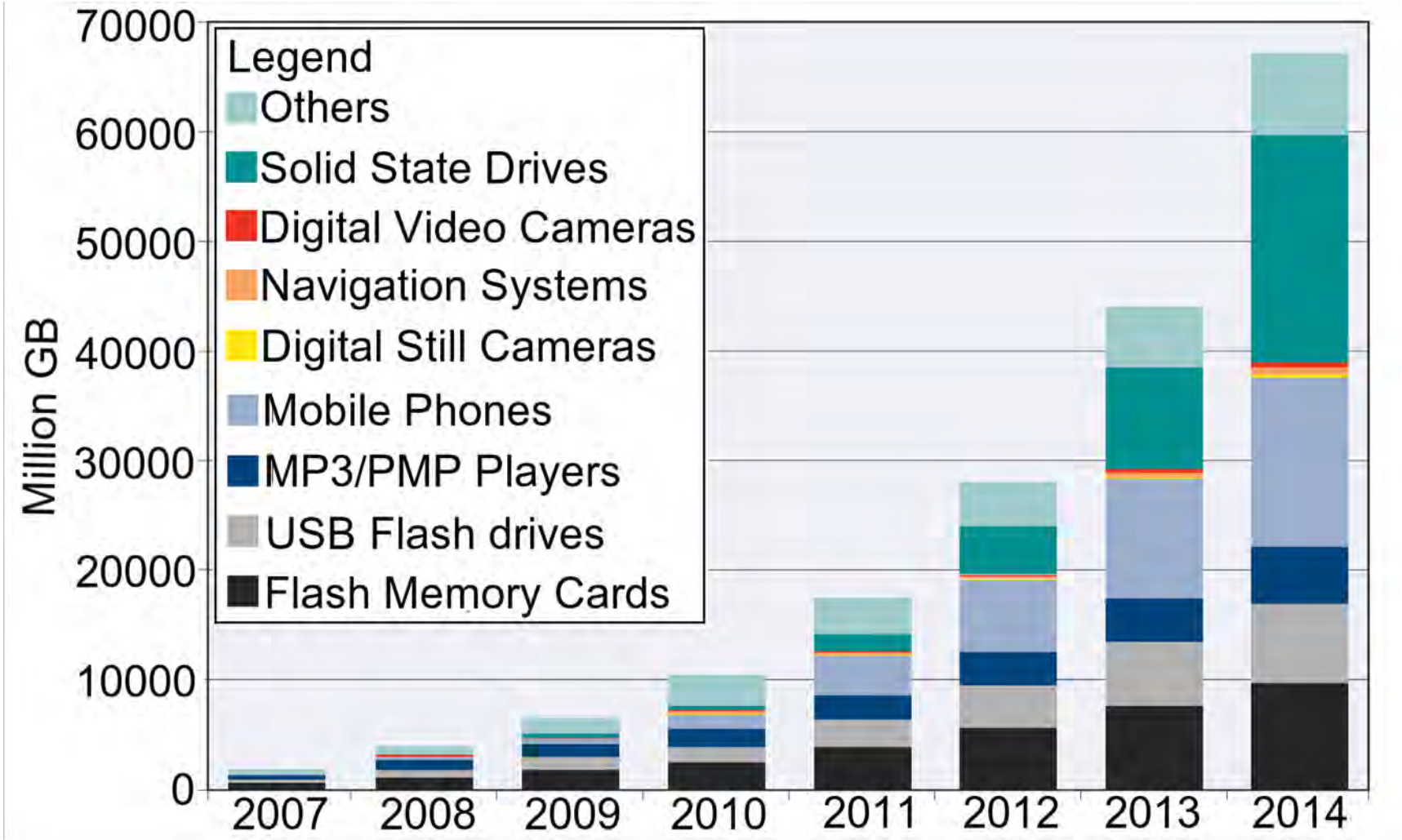
Flash



PG-Gnd – Ground select transistor
PG-BL – BL select transistor
FG – Floating gate transistor
BL – Bit line
WL – Word line

- Cell size — *4 – 5F²*
- Scalability
- Access Time — *Rd: 25µs, Wr: 200µs*
- Power Consumption
- Endurance — *10⁴*
- Data Reliability
- **No** Volatility
- Fabrication Cost

Flash memories



FLASH memories

- Both NOR and NAND, Invented in 1984 by F. Masuoka (Toshiba)
- The great potential of the NOR type flash memory was exploited by Intel, who introduced the first commercial chip in 1988

FLASH memories

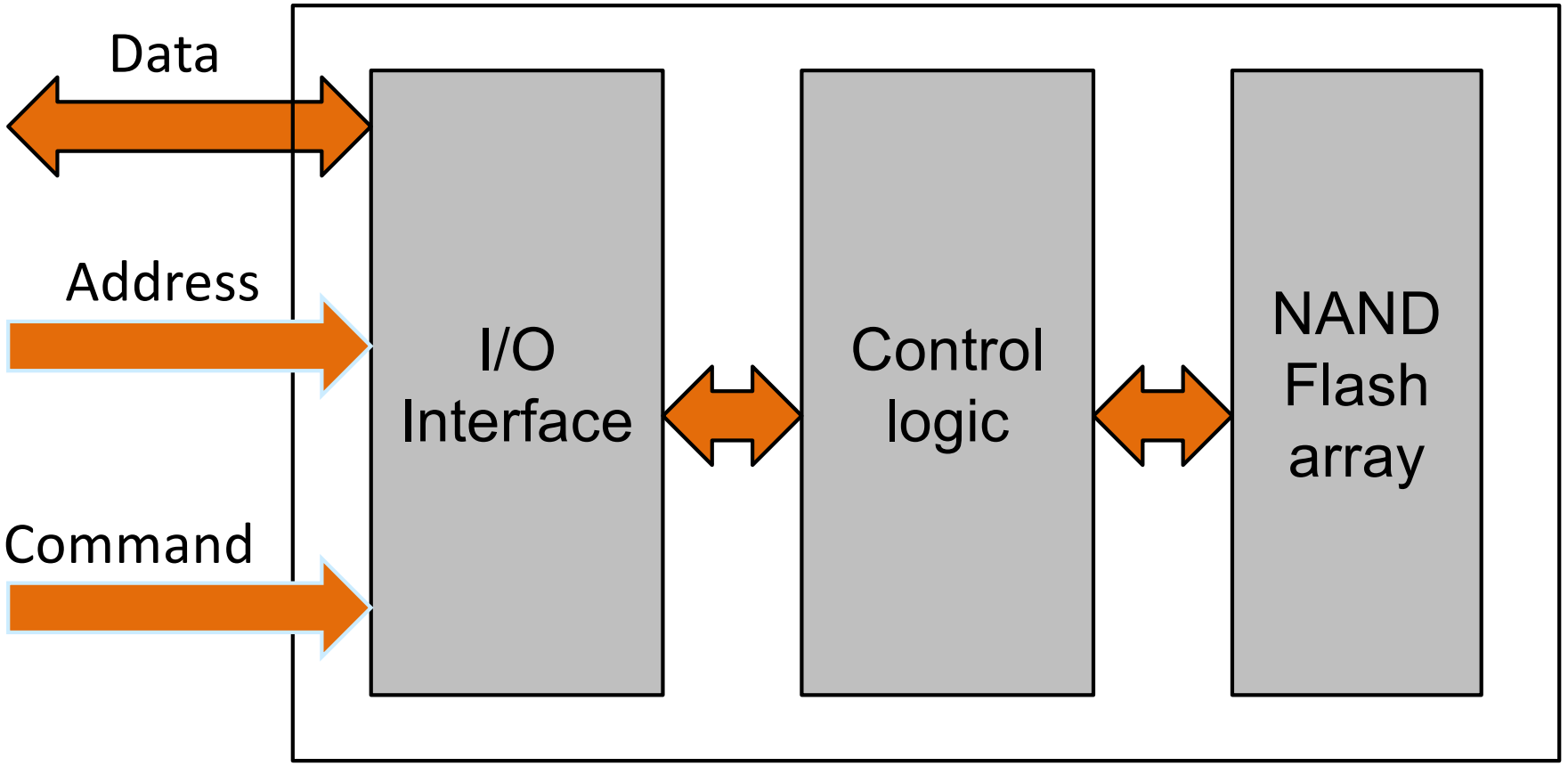
- W.r.t. the other memories, Flashes have peculiar characteristics:
  - Read & Write operations are not symmetric: writing requires a preliminary erase operation
  - They wear out after a given number of erase operations

Flash memories' features

- Main features:
  - low power consumption
  - light weight
  - robustness
  - high cost/information capacity ratio (w.r.t. other devices)
  - access limitations (erasure required before programming)
  - limited usage cycle (reliability issue).

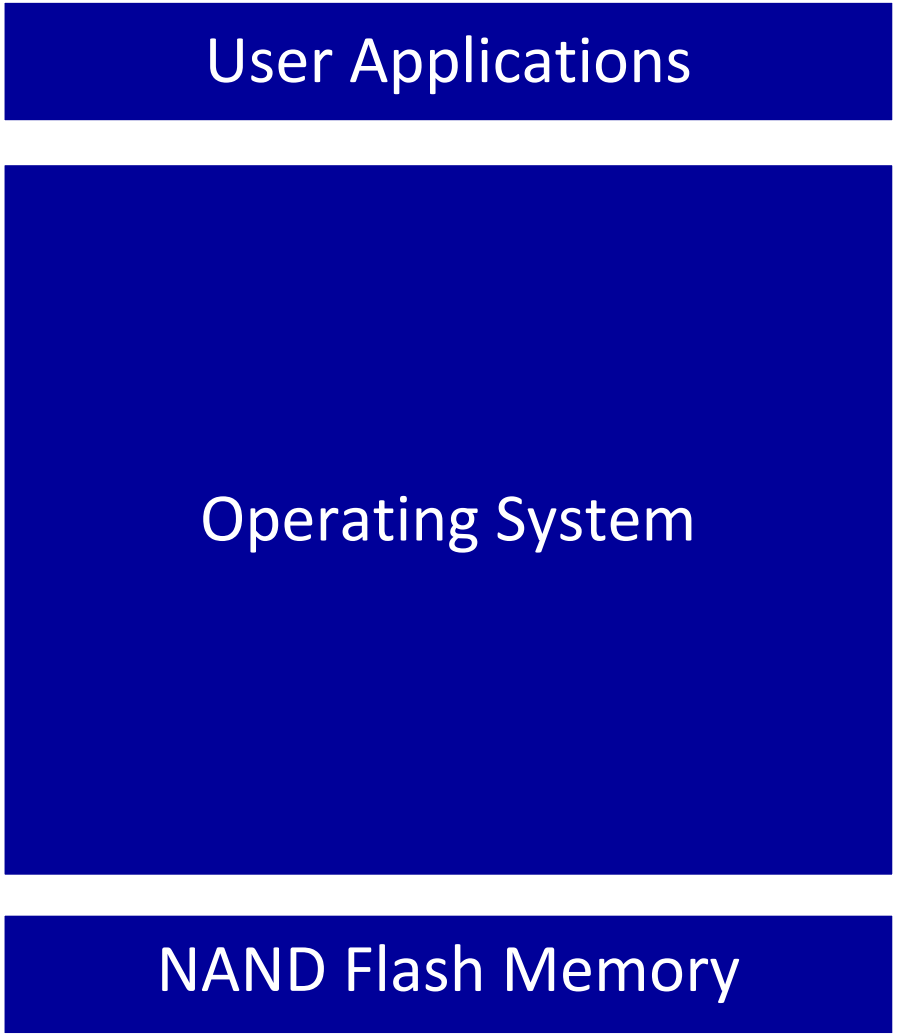
Flash memories overview

- The architecture of a Flash memory is similar to the classic RAM memory architecture:



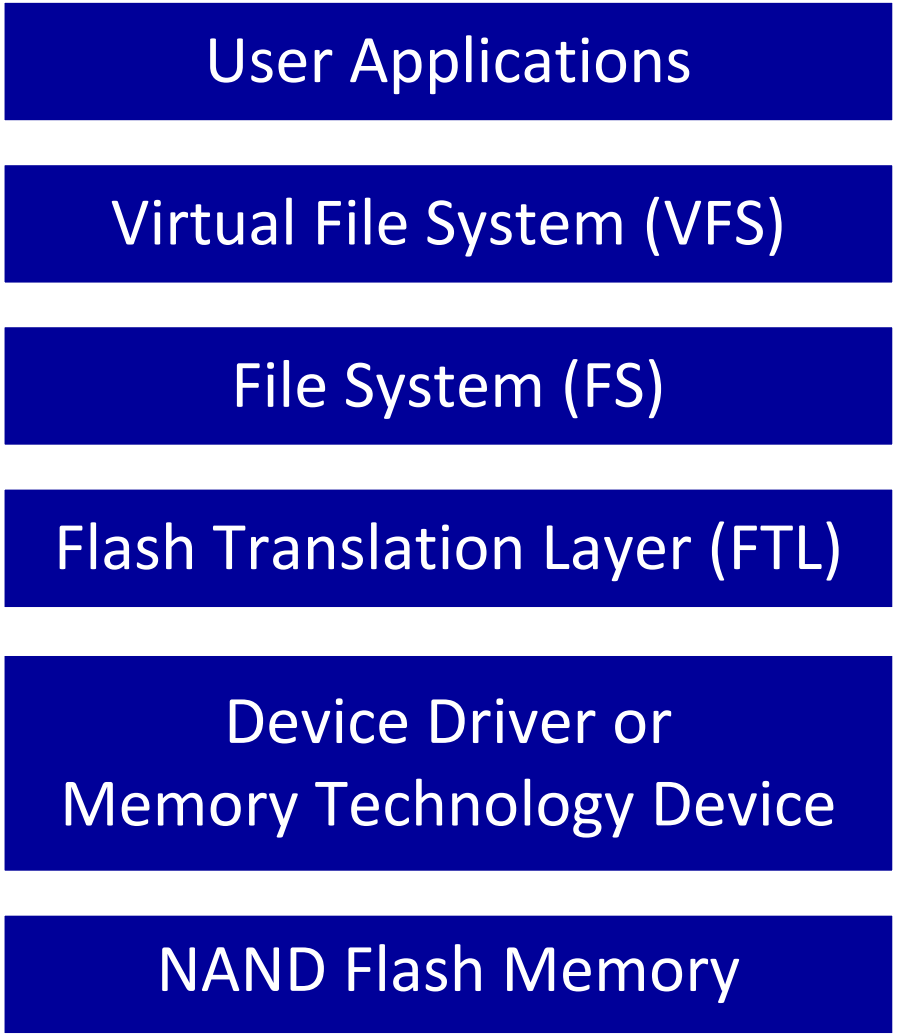
SW Stack

- To let the User Applications use a Flash Memory as a storage device some intermediate levels are needed.



SW Stack

- In particular the Flash Translation Layer is needed to make the Flash Memory be seen, by the upper layers, as a simple block device and solve reliability issues



Flash Memory technologies

- From a technological point of view, Flash memories can be classified as:
  - NOR
  - NAND

NOR Flash

- Allow a random access to each cell
- Read is very fast
- Both Erase and Write are very slow
- Mostly used to store executable code in embedded systems (mainly smartphones, tablets, …)

NAND Flash

- Both Read and Write are very fast
- High-density
- Low-cost data storage
- Low-power consumption
- Mostly used to store huge amounts of data
- Basic components of SSDs (Solid State Drives)

NAND Flash today

- Samsung Electronics, one of the leading developers of advanced memory technology, announced in August 2018 its fifth-generation V-NAND memory chips.
- The new V-NAND chips claim the fastest data transfers and the industry's first use of the Toggle DDR4.0 interface. With data speeds as high as 1.4 gigabits per second (Gbits/s), the 256-Gbit V-NAND chips are 40% faster than the previous 64-layer product.

Outline

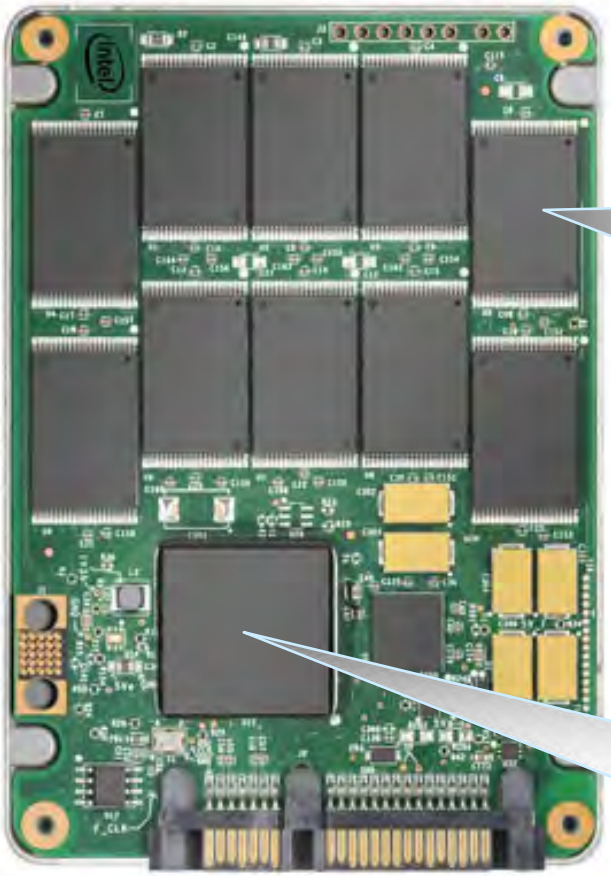
- Memory device *yesterday*
  - ROM Memories
- **Memory device *today***
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - **SSD**
  - Comparisons
- Memory device *tomorrow*
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

Solid State Drive (SSDs)

Yesterday

Solid State Drive (SSDs)

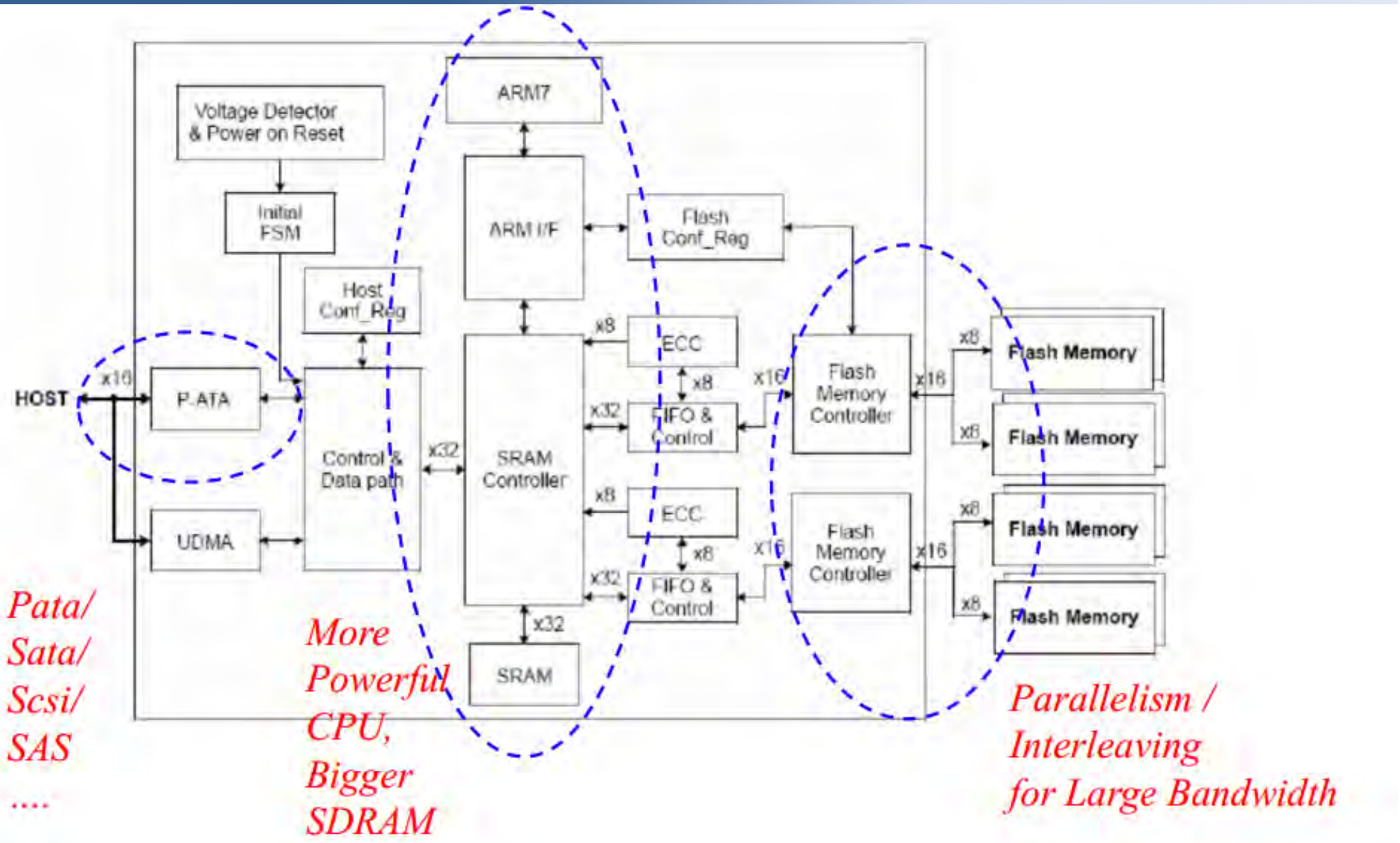
Yesterday **Today**

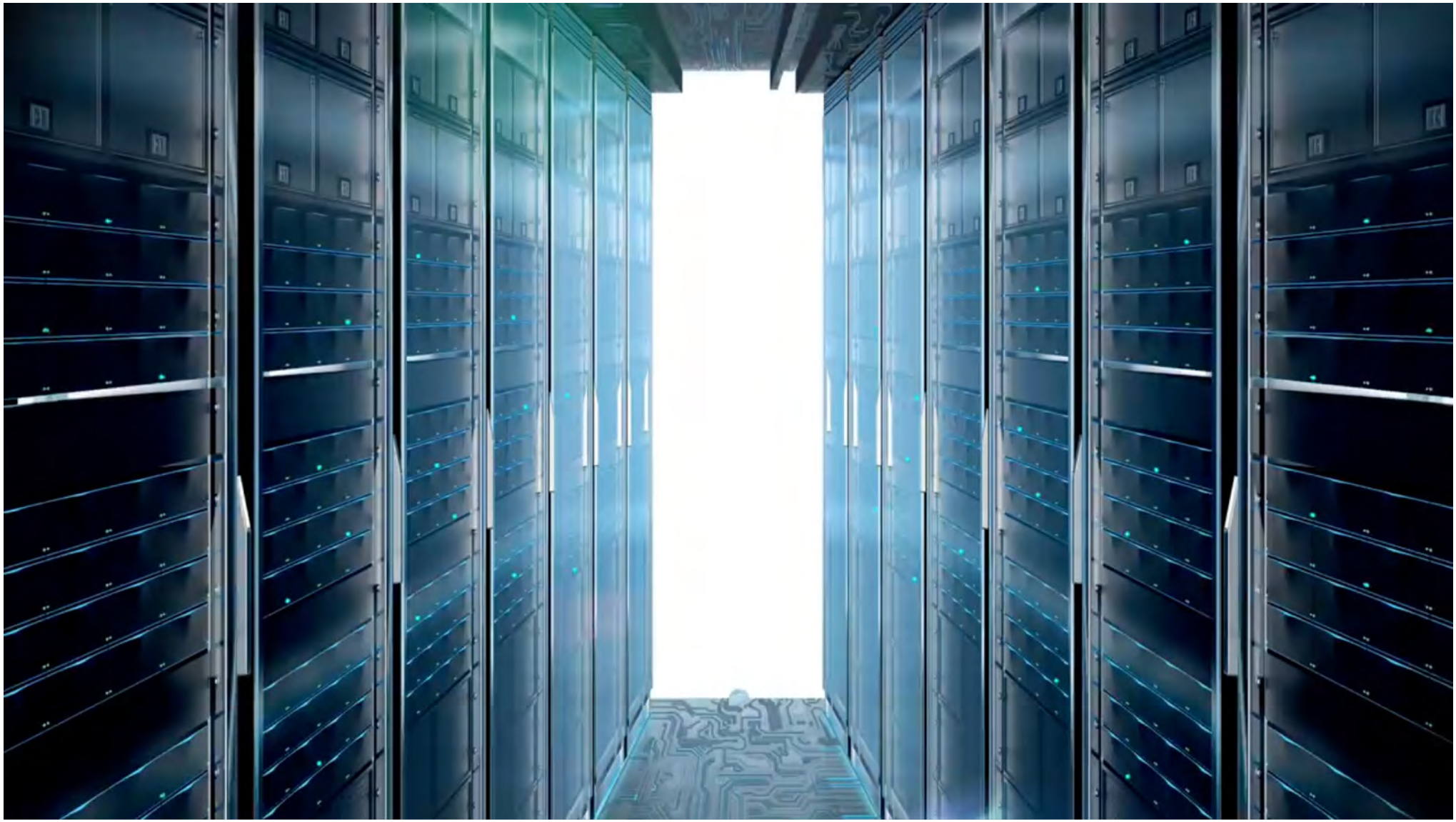


NAND Flash Memories

Memory Controller (ARM-based)

SSD = computing platform



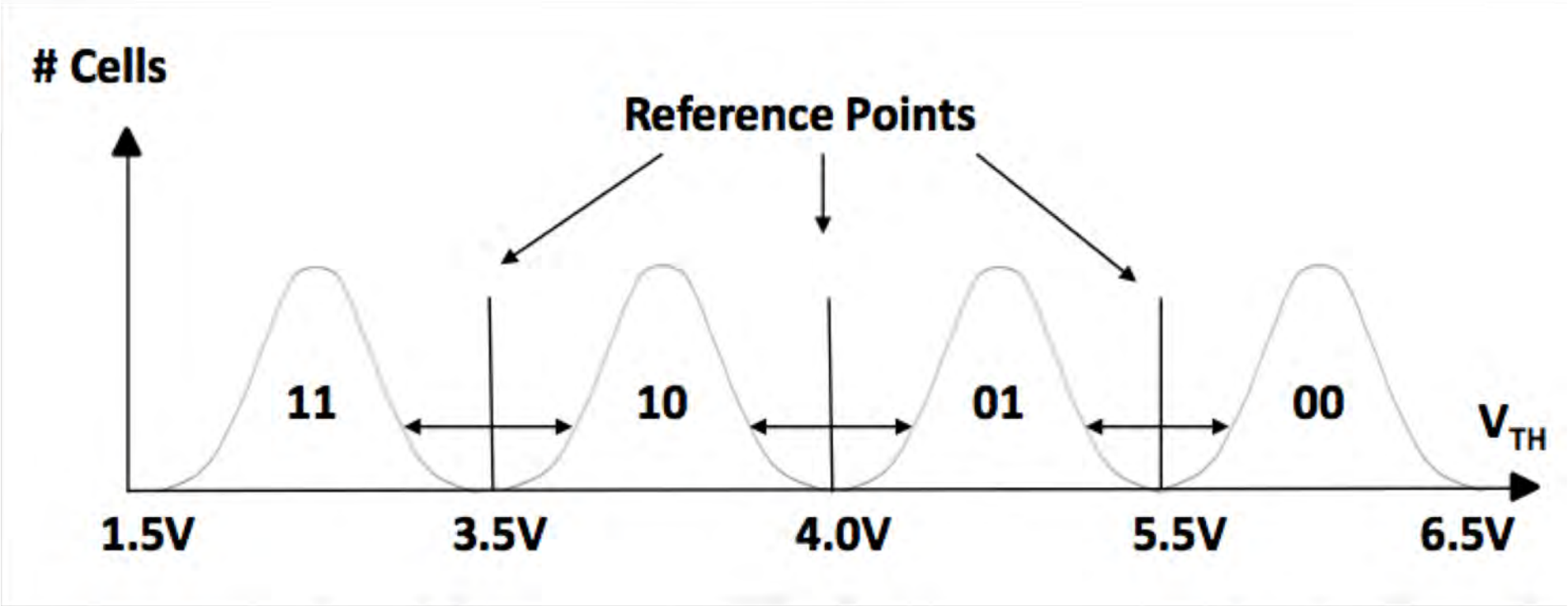


SLC vs MLC

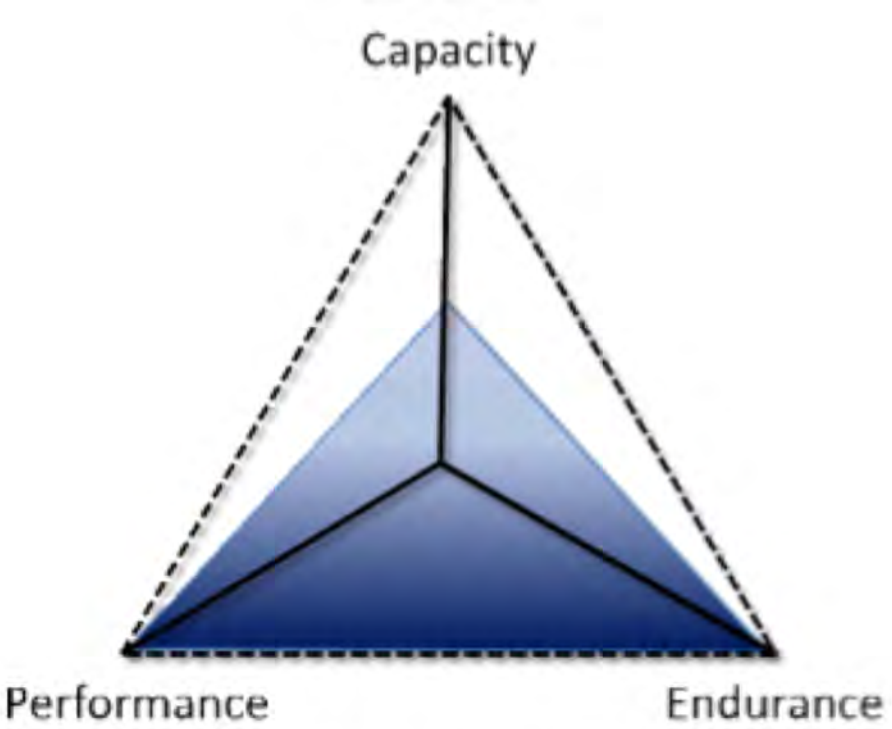
- NAND Flashes clustered as:
  - Single Level Cell (SLC) : store 1 bit / cell
  - Multiple Level Cell (MLC) : store more bits / cell

SLC vs MLC

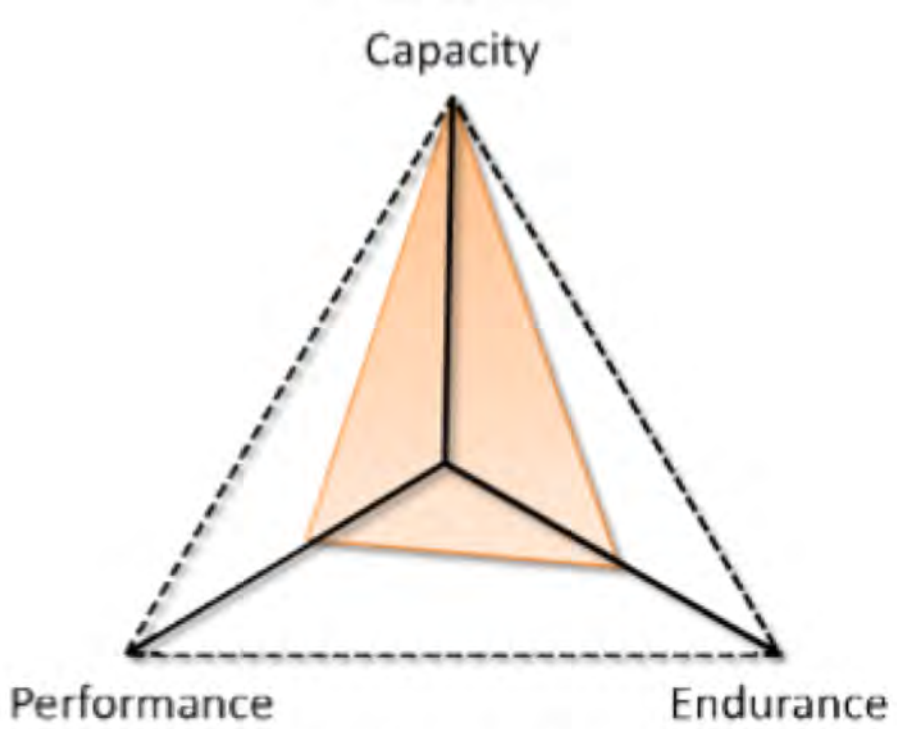
- NAND Flashes clustered as:
  - Single Level Cell (SLC) : store 1 bit / cell
  - Multiple Level Cell (MLC) : store more bits / cell



SLC vs MLC



(a) SLC NAND Flash



(b) MLC NAND Flash

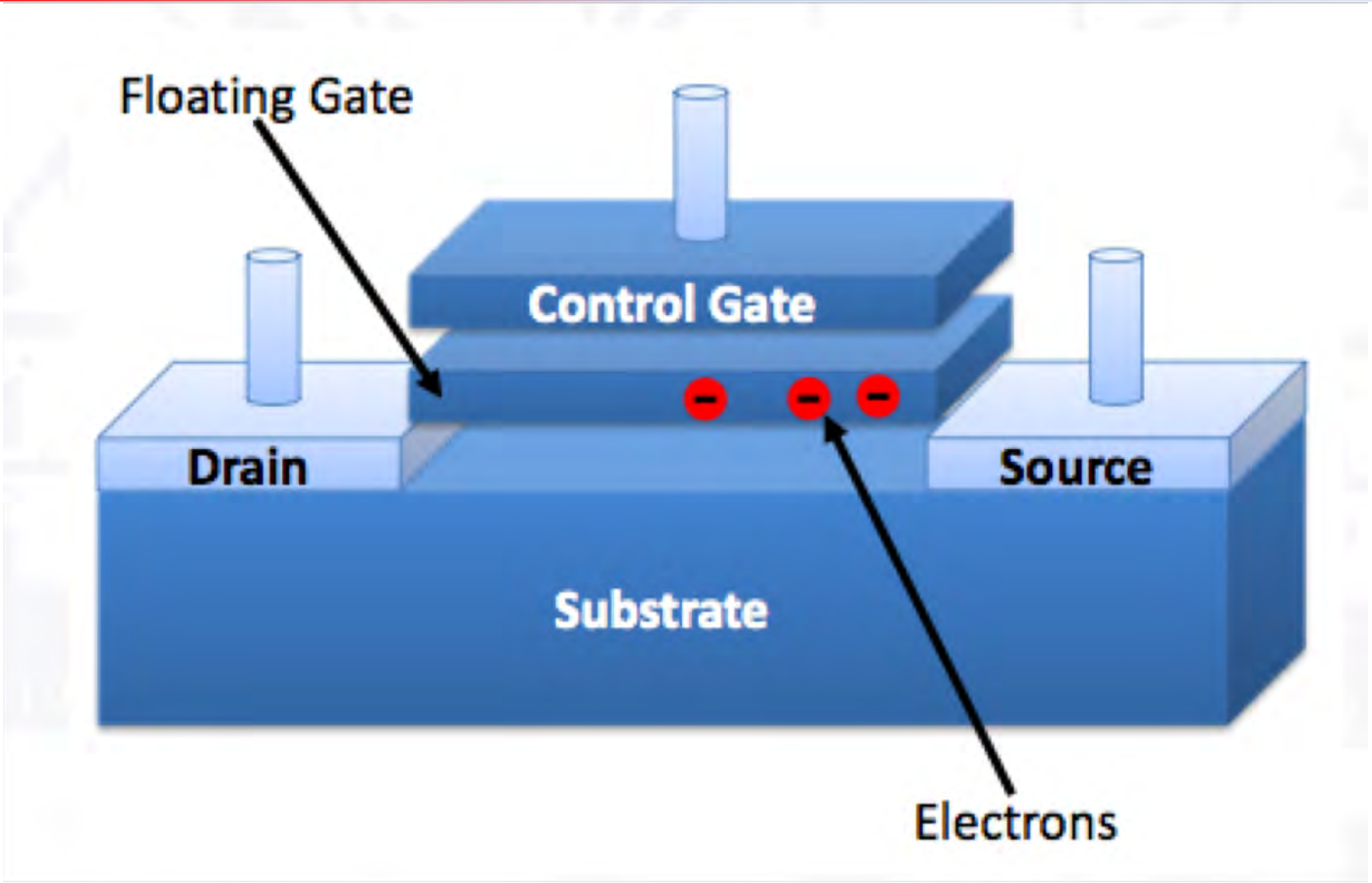
SLC vs MLC – Some data

	Features			*Architecture*			*Reliability*		*Array Operations*			
	Bits /cell	**Volt-age**	**Bus width**	**Pla-nes**	**Page size**	**Pages /block**	NOP	ECC-**512B**	**Endu-rance**	t_{READ} **(max)**	t_{PROG} **(avg)**	t_{ERASE} **(avg)**
SLC	1	3.3V, 1.8V	x8, x16	1 or 2	2,112B	64+	1	1	$<10^5$	25us	200 - 300us	1.5 – 2ms
MLC	2+	3.3V	x8	2+	4,314B+	128+	4+	4+	$<10^4$	50us	600 - 900us	2ms

Next generations (< 40 nm)

- At such fine geometries old fashioned measurements of transistor gate width, line width and the half-pitch of repeated structures become hard to measure.
- The term 3Xnm node is the nomenclature being adopted by the memory industry to denote a manufacturing process with minimum geometries of somewhere between 30nm and 39nm
- 2Xnm node = > between 20nm and 29nm
- 1Xnm node = > between 10nm and 19nm

Floating Gate MOS



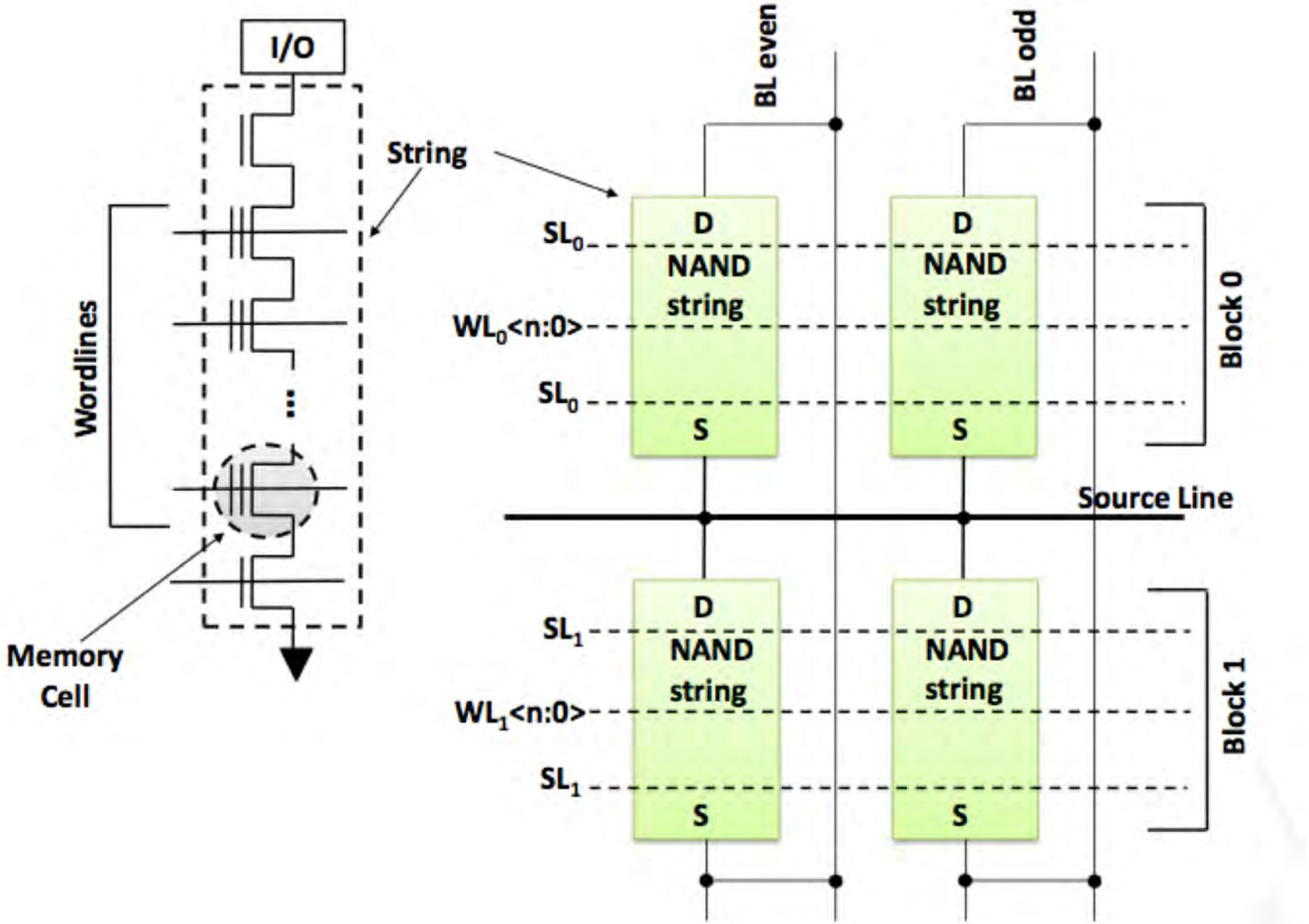
Flash Cell

- A MOS transistor is built with two overlapping gates rather than a single one: the first one is completely surrounded by oxide, while the second one is contacted to form the gate terminal. The isolated gate constitutes an excellent trap for electrons, which guarantees charge retention for years.

Program & Erase

- The operations performed to inject and remove electrons from the isolated gate are called *program* and *erase*, respectively.
- These operations modify the threshold voltage VTH of the memory cell.
- Applying a fixed voltage to cell's terminals, it is then possible to discriminate two storage levels (e.g., whenever gate voltage is higher than the cell's VTH , the cell is on ("1"), otherwise it is off ("0")).

NAND string (left), NAND array (right)



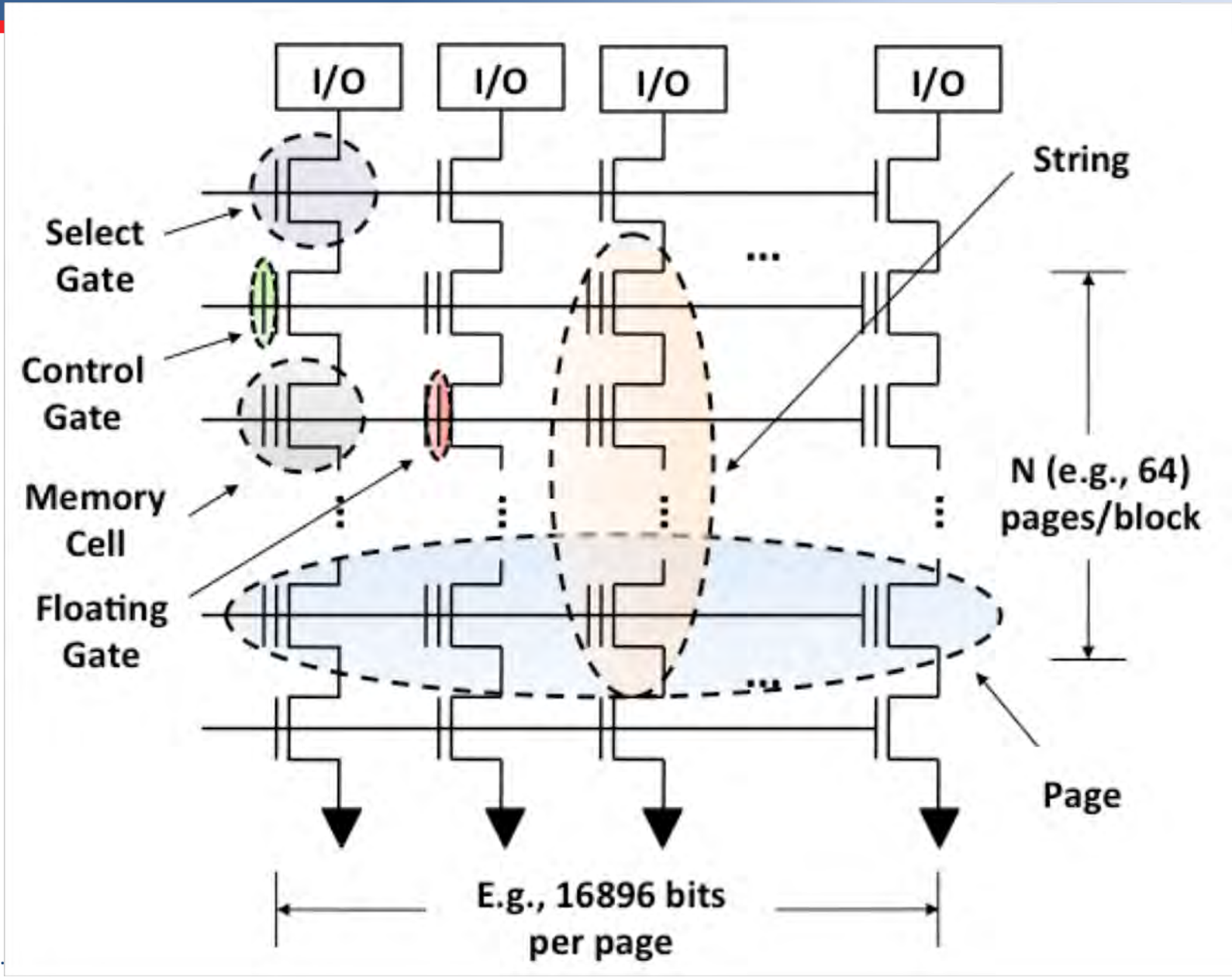
Flash Page

- Logical pages are composed of cells belonging to a same wordline
- The page is the smallest storage unit when performing read and programming operations
- Pages that already contain data must be erased before writing new data
- Typical page sizes:
  - 2 kB (data) + 64 B (spare) for SLC
  - 4 kB (data) + 128 B (spare) for MLC

Flash Block

- A block is a set of Pages
- It's the minimum portion of the memory that can be erased
- To erase a page you have to erase the whole Block it belongs to
- Typical Block size = 64 or 128 pages

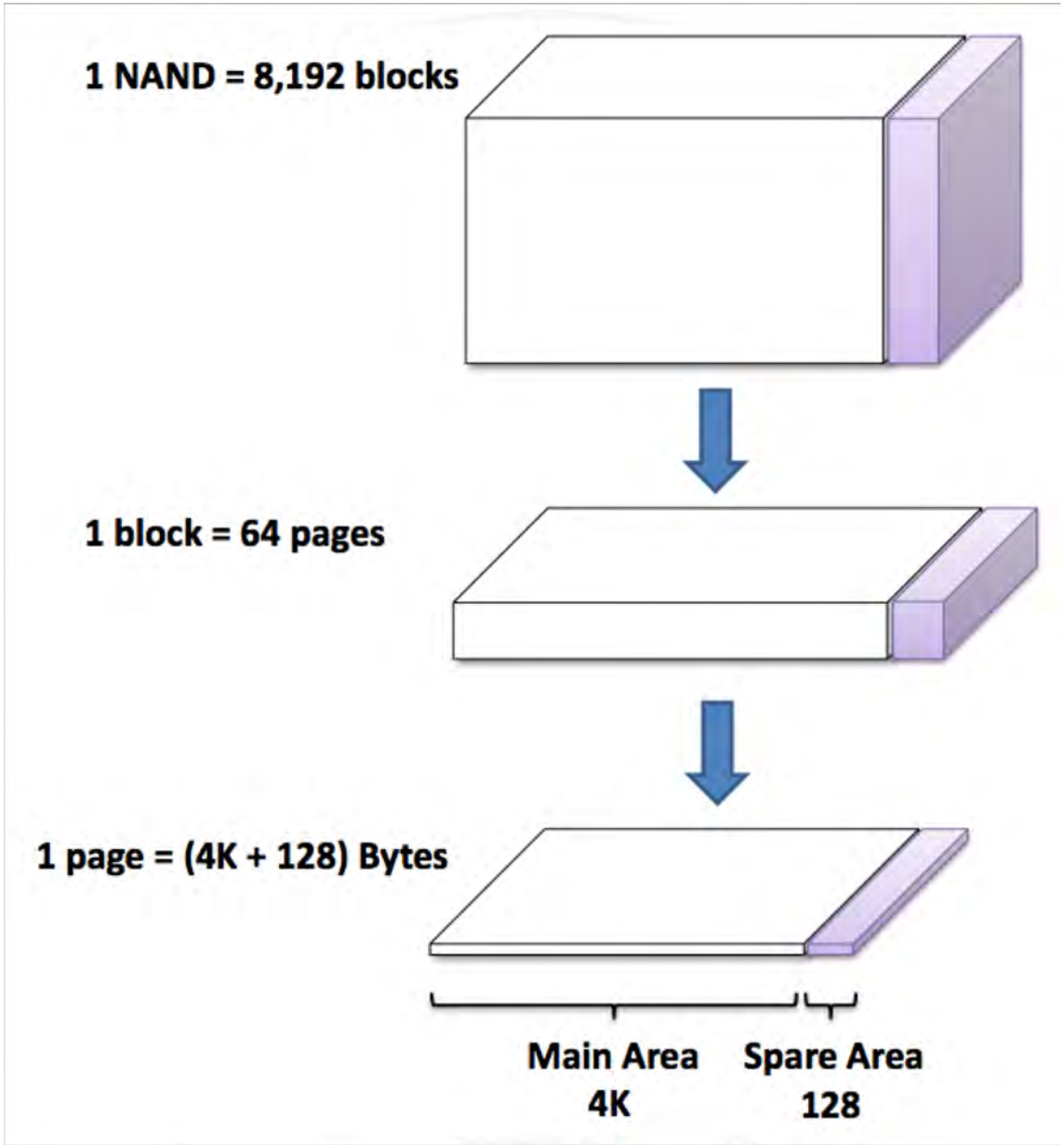
Block structure in a SLCNAND



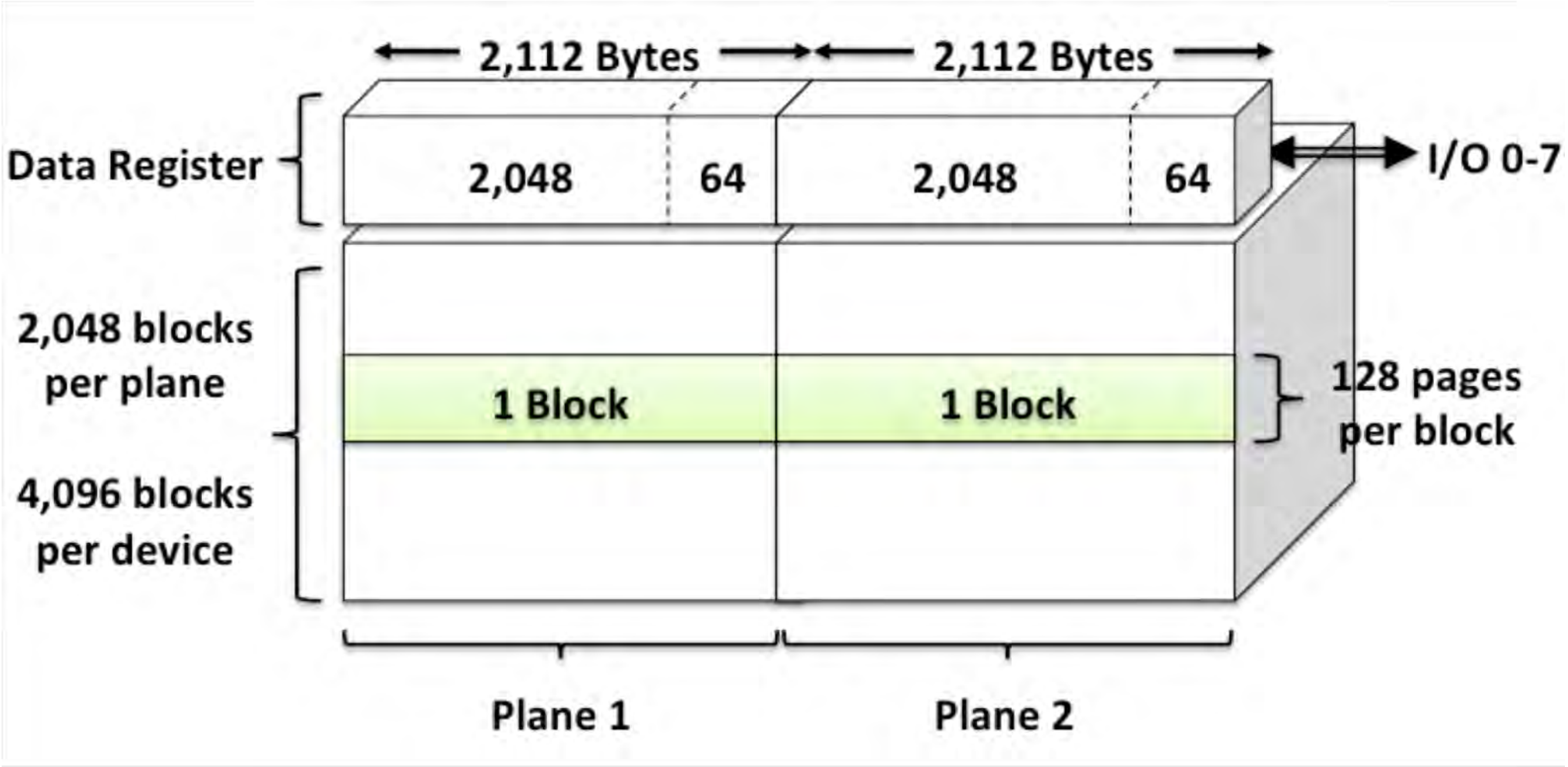
Flash Plane

– Blocks are clustered in Planes : a memory with N planes can concurrently read (or write, or erase) N pages/block, one in each plane

Logical Organization of a NAND



Organization of 2 GB NAND



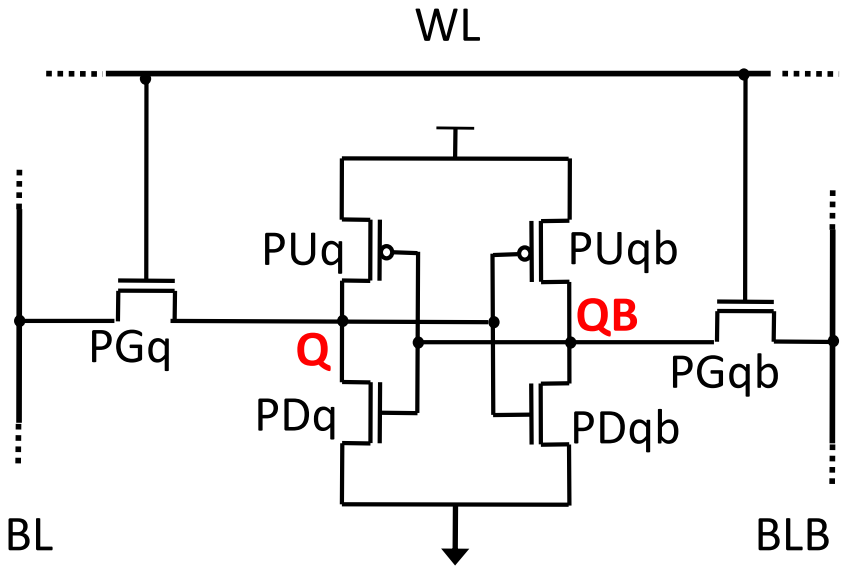
Outline

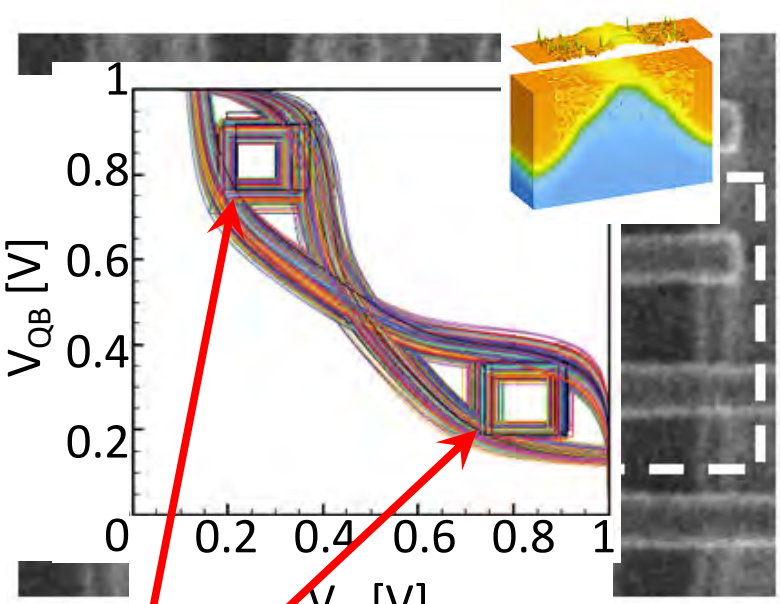
- Memory device *yesterday*
  - ROM Memories
- **Memory device *today***
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - **Comparisons**
- Memory device *tomorrow*
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

Memories Today

	SRAM	DRAM	NOR-Flash	NAND-Flash
Cell Size	$120F^2$	$4\text{-}6F^2$	$10\ F^2$	$4\text{-}5\ F^2$
Read Latency	<1ns	20ns	5,000ns	25,000ns
Write Latency	<1ns	20ns	1,000,000ns	200,000ns
Static power	YES	YES	NO	NO
Endurance	$>10^{15}$	$>10^{15}$	10^4	10^4
Non-volatility	NO	NO	YES	YES

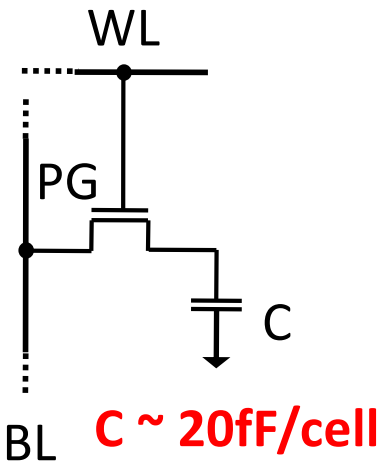
Memory Limitations
SRAM cell





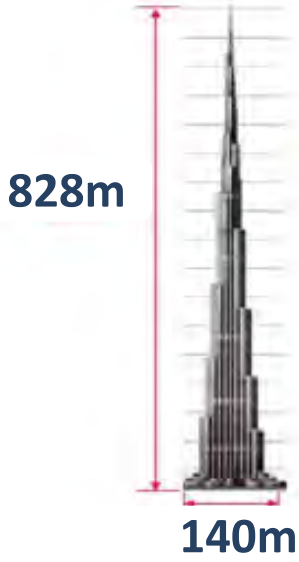
SNM – Static Noise Margin

Memory Limitations
DRAM cell



C ~ 20fF/cell

Aspect ratio: 25

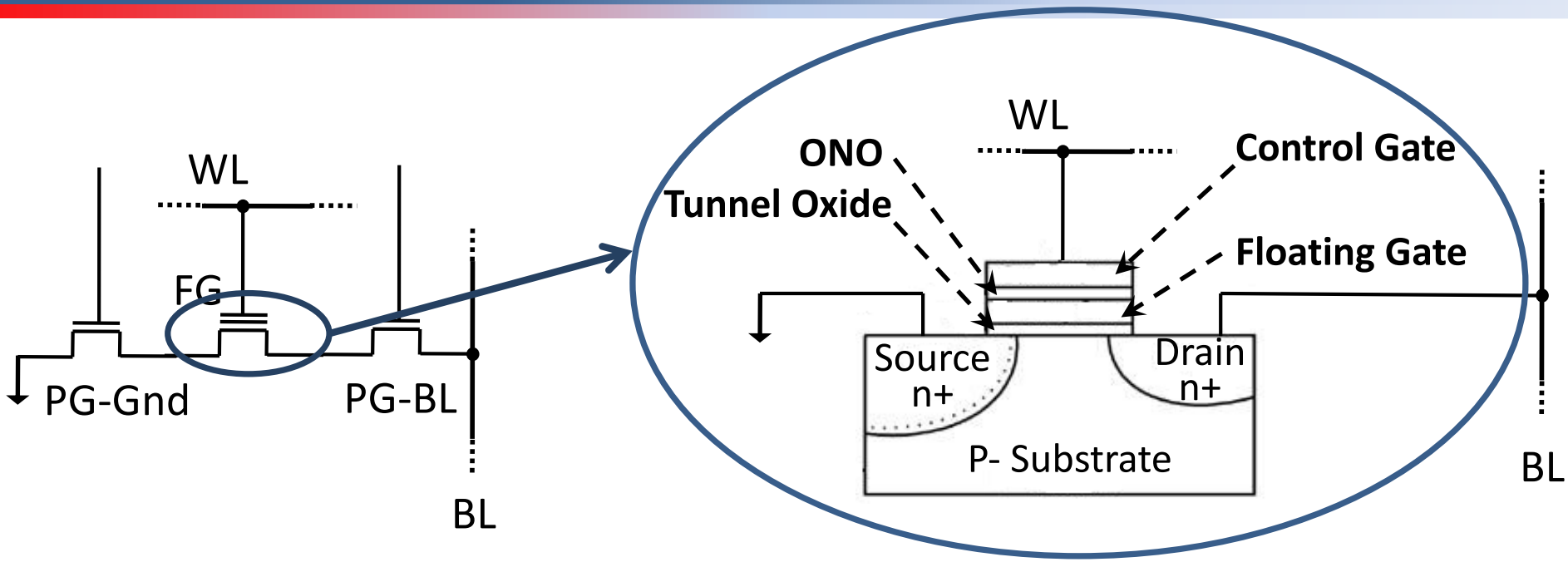


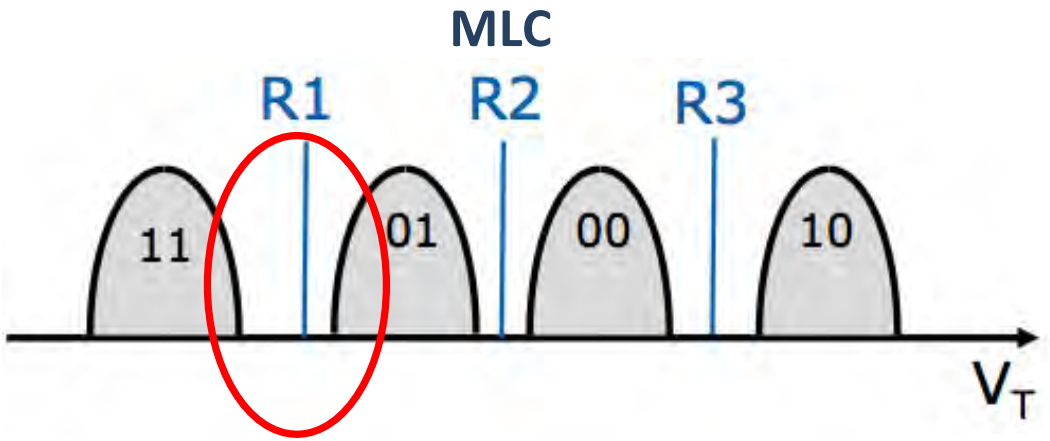
Aspect ratio: 6

- Very high-k materials (k = the dielectric constant)
- Very tall structures

Memory Limitations

Flash cell





Outline

- Memory device *yesterday*
  - ROM Memories
- Memory device *today*
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- **Memory device *tomorrow***
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

Emerging Memory Technologies

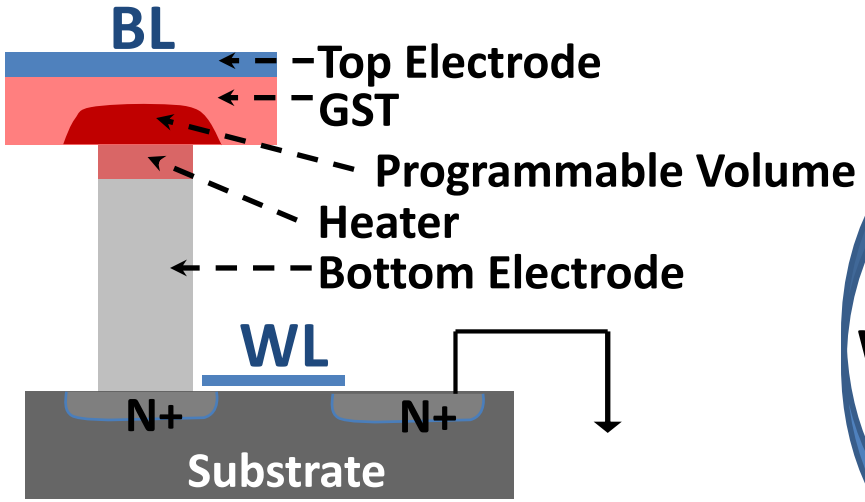
"...the ERC/ERM working groups identified the Spin Transfer Torque MRAM and Redox RRAM as emerging memory technologies recommended for accelerated research and development leading to scaling and commercialization of non-volatile RAM to and beyond the 16nm generation".

[ITRS Roadmap 2010]

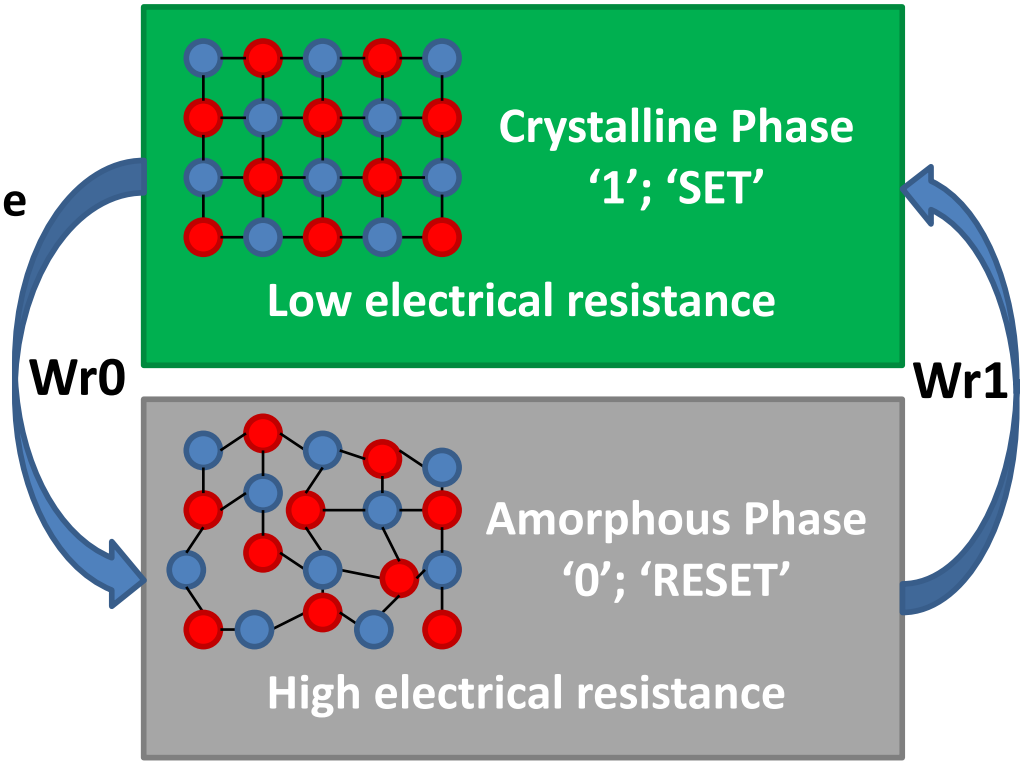
Outline

- Memory device *yesterday*
  - ROM Memories
- Memory device *today*
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- **Memory device *tomorrow***
  - **Phase Change RAM**
  - Resistive RAM
  - Magnetic RAM
  - Comparisons

Phase Change Memory



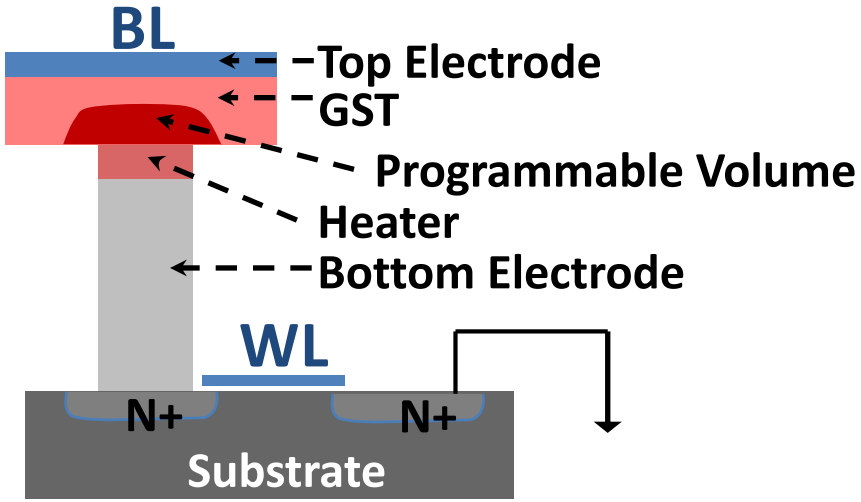
GST - Germanium-Antimony-Tellurium
(chalcogenide glass)

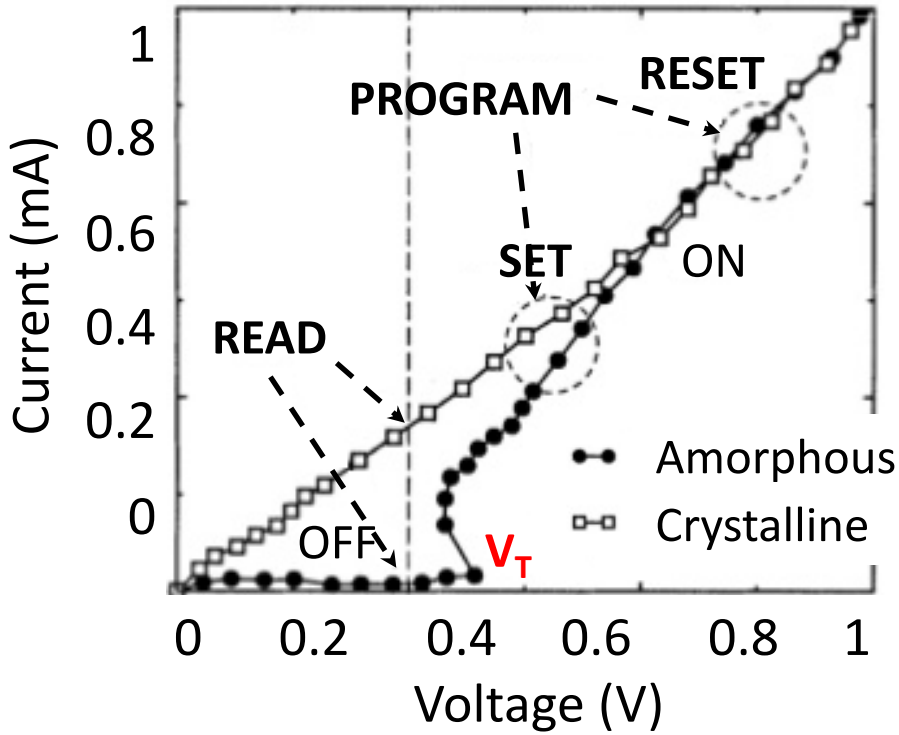


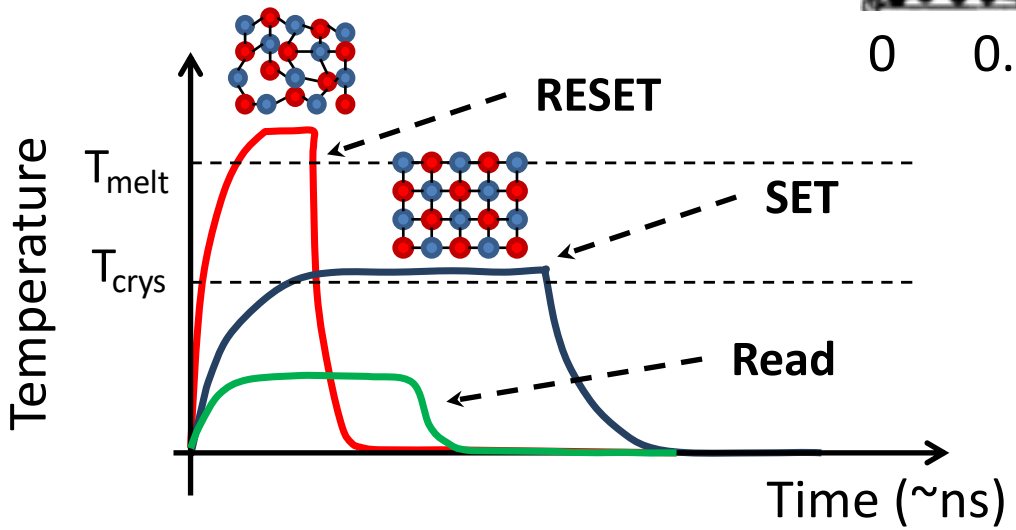
The phase of the GST glass:

- **determines the *resistance* of the material**
- **can be changed by injecting *current*.**

Phase Change Memory

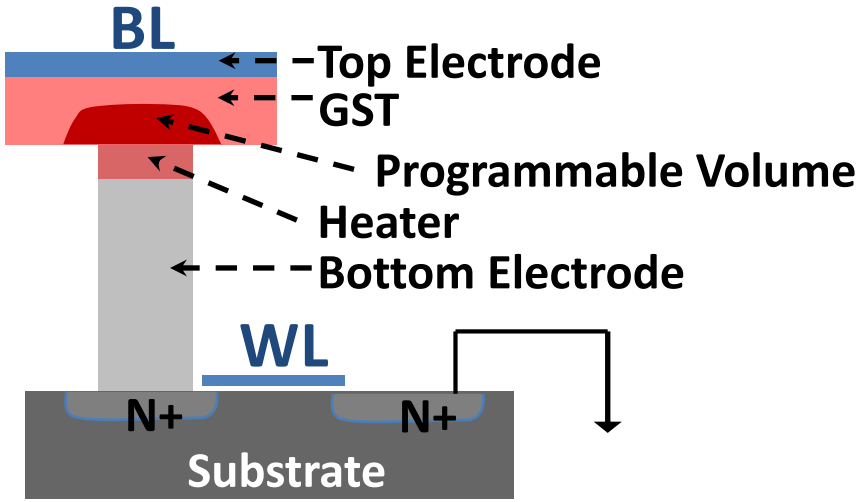


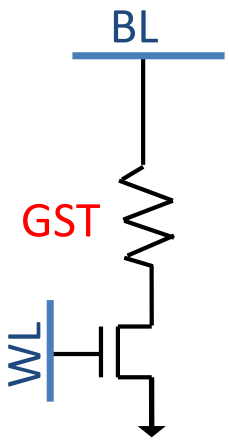




[M. Rudan, VTS, 2012]

Phase Change Memory



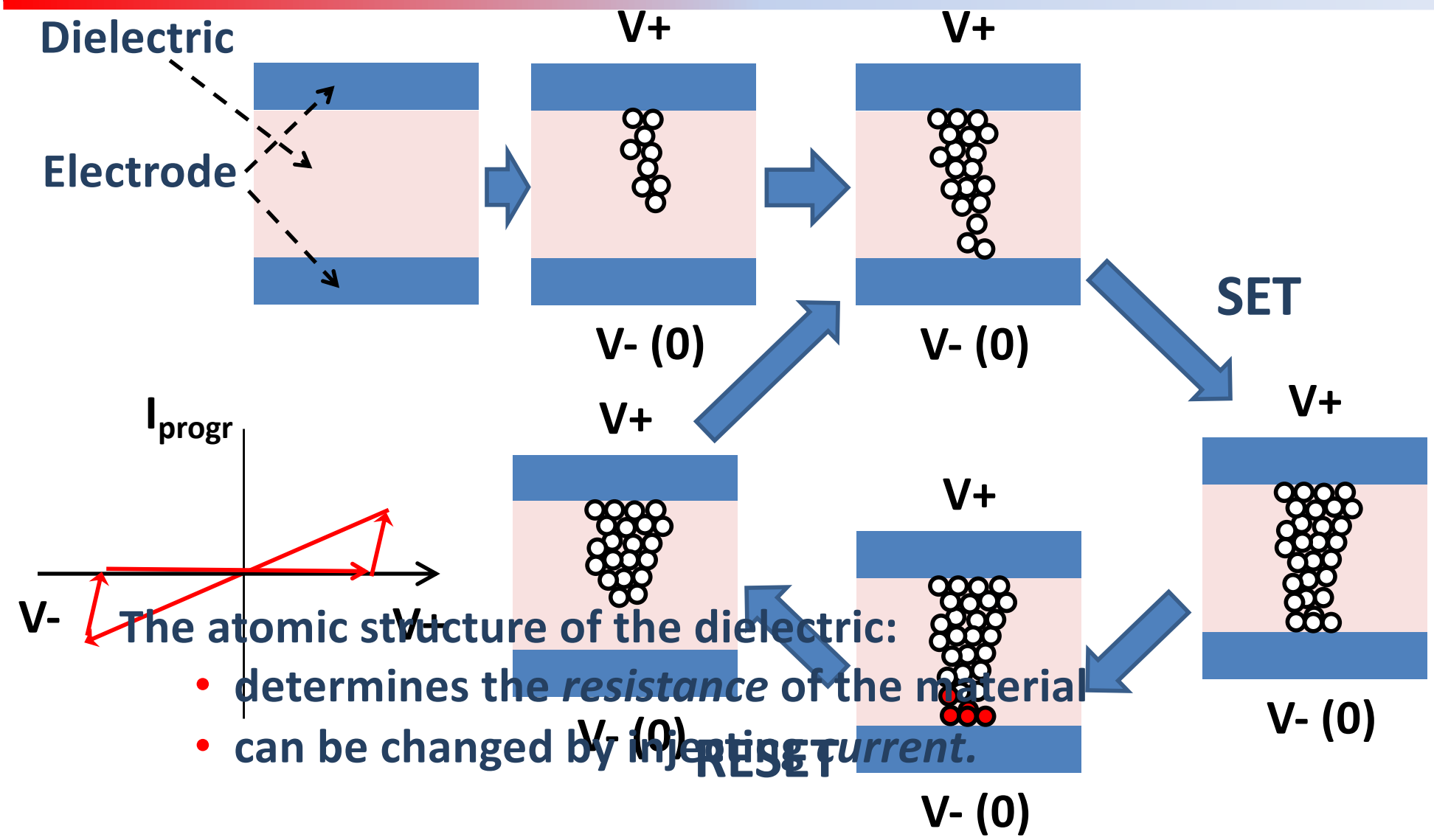


- Cell size — *9 – 12F² SLC*
- Scalability
- Access Time — *Rd: 50ns, Wr1: 150ns, Wr0: 80ns*
- Power Consumption
- Endurance — *10⁸*
- Data Reliability
- No Volatility
- Fabrication Cost

Outline

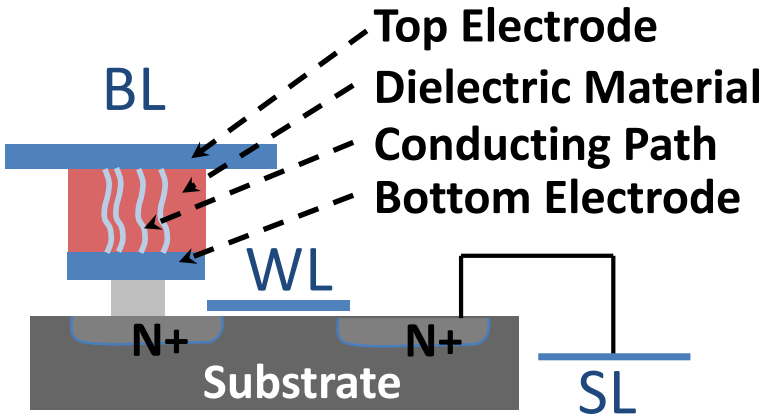
- Memory device *yesterday*
  - ROM Memories
- Memory device *today*
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- **Memory device *tomorrow***
  - Phase Change RAM
  - **Resistive RAM**
  - Magnetic RAM
  - Comparisons

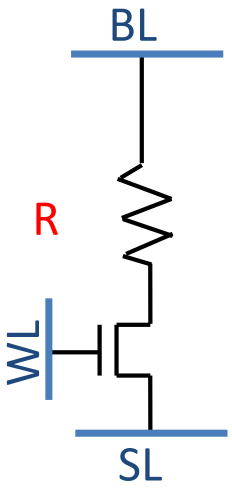
Resistive RAM



- The atomic structure of the dielectric:
  - determines the *resistance* of the material
  - can be changed by injecting *current*.

Resistive RAM



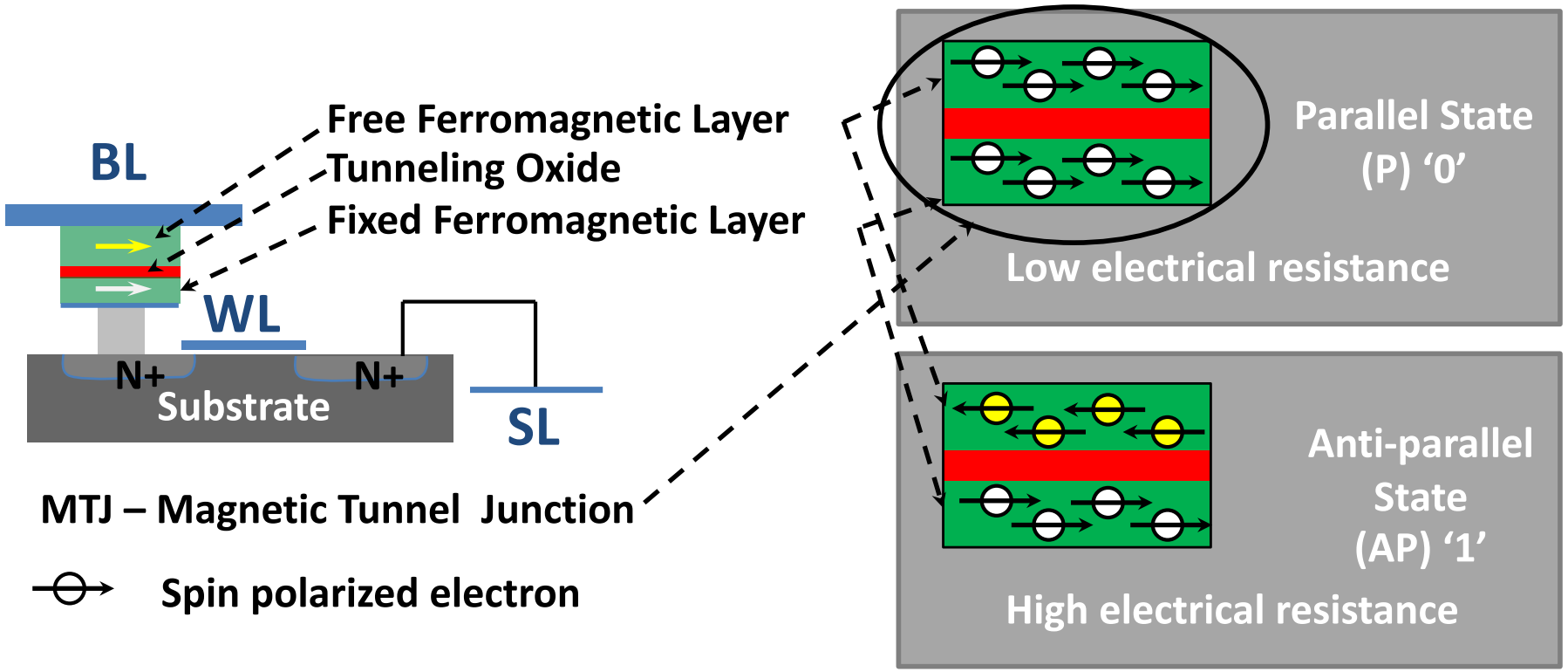


- Cell size — *4 – 6F²*
- Scalability
- Access Time — *Rd: 10ns, Wr: 40ns*
- Power Consumption
- Endurance — *>10¹²*
- Data Reliability
- No Volatility
- Fabrication Cost

Outline

- Memory device *yesterday*
  - ROM Memories
- Memory device *today*
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- **Memory device *tomorrow***
  - Phase Change RAM
  - Resistive RAM
  - **Magnetic RAM**
  - Comparisons

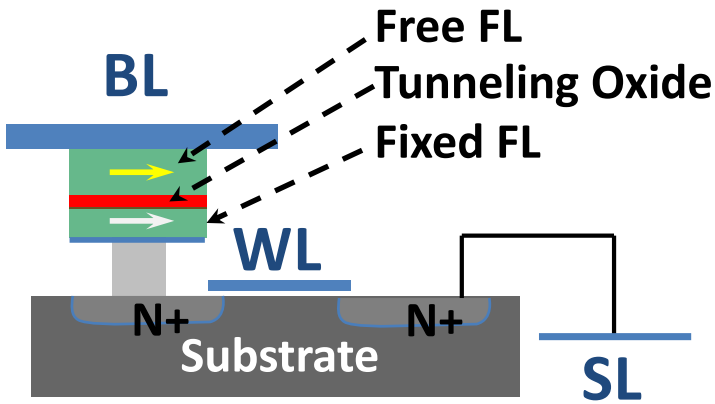
Magnetic RAM



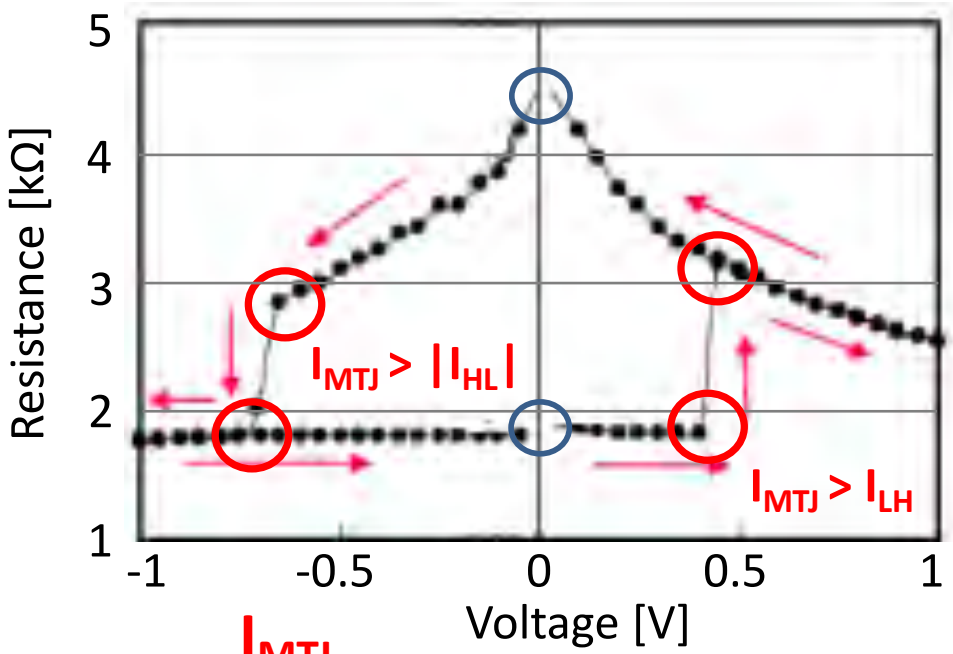
The orientation of the free layer:

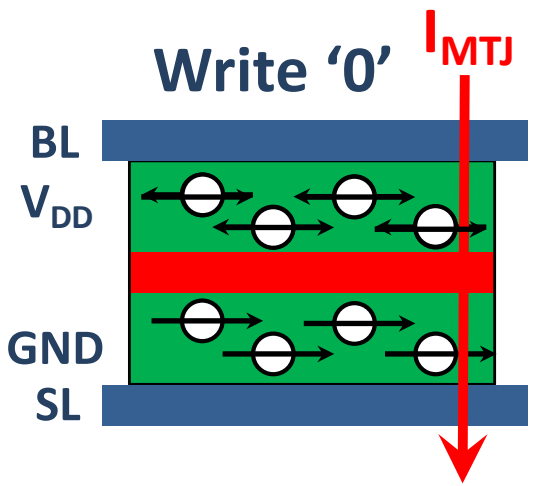
- **determines the *resistance* of the material**
- **can be changed by injecting *current*.**

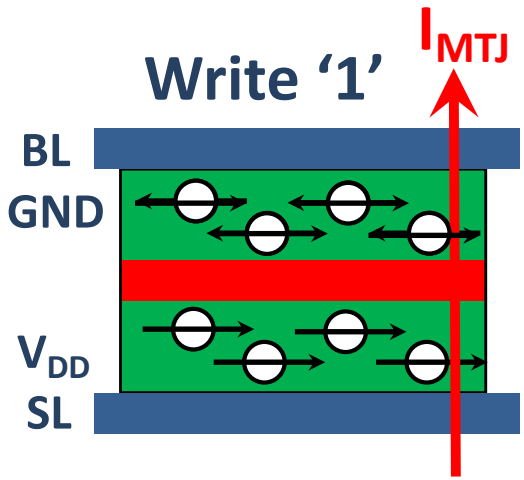
Magnetic RAM
Spin-Transfer-Torque (STT)



FL – Ferromagnetic Layer





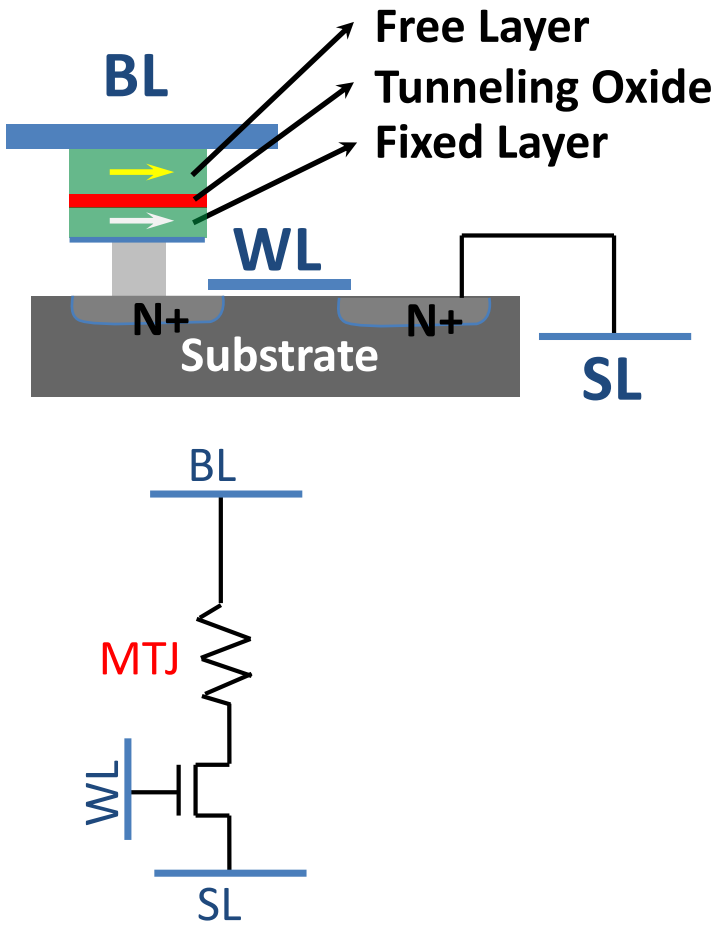


I_{HL} –high to low transition

I_{LH} –low to high transition

[M. Hosomi, IDEM, 2005]

Magnetic RAM
Spin-Transfer-Torque (STT)



- Size — $4 – 6F^2$
- Scalability
- Access Time — Rd: 10ns, Wr1: 40ns, Wr0: 10ns
- Power Consumption
- Endurance — $>10^{15}$
- Data Reliability
- No Volatility
- Fabrication Cost

Outline

- Memory device *yesterday*
  - ROM Memories
- Memory device *today*
  - SRAM – Static Random Access Memories
  - DRAM – Dynamic Random Access Memories
  - Flash Memories
  - SSD
  - Comparisons
- **Memory device *tomorrow***
  - Phase Change RAM
  - Resistive RAM
  - Magnetic RAM
  - **Comparisons**

Emerging RAM compared

	PCRAM	STT-MRAM	ReRAM
Cell Size	9-12F^2	4-6F^2	4-6F^2
Read Access Time	50ns	10ns	10ns
Write1 Access Time	150ns	20ns	40ns
Write0 Access Time	80ns	40ns	40ns
Endurance	10^8	>10^{15}	10^{12}
Non-volatility	YES	YES	YES

What we have learned

	SRAM	DRAM	Flash	PCRAM	STT	ReRAM
Cell Size	$120F^2$	$4\text{-}6F^2$	$4\text{-}5\ F^2$	$9\text{-}12F^2$	$4\text{-}6F^2$	$4\text{-}6F^2$
Read Access Time	<1ns	20ns	25,000ns	50ns	10ns	10ns
Write1 Access Time	<1ns	20ns	200,000ns	150ns	20ns	40ns
Write0 Access Time	<1ns	20ns	200,000ns	80ns	40ns	40ns
Endurance	$>10^{15}$	$>10^{15}$	10^4	10^8	$>10^{15}$	10^{12}
Non-volatility	NO	NO	YES	YES	YES	YES


Малые Автюхи, Калинковичский район, Республики Беларусь