micron Educator Hub

Introduction to Memory

Reviewed: 2025

micron

Copyright guidelines

By using any content provided by the Micron Educator Hub, you acknowledge that Micron Technology, Inc. (“Micron”) is the sole owner of the content and agree that any use of the content provided by the Micron Educator Hub must comply with applicable laws and require strict compliance with these Guidelines:

1. Credit shall be expressly stated by you to Micron for use of the content, including any portion thereof, as follows:
    a. “*© 2020-2025 Micron Technology, Inc. All Rights Reserved. Used with permission.*”
2. You may not use the content in any way or manner other than for educational purposes.
3. You may not modify the content without approval by Micron.
4. You may not use the content in a manner which disparages or is critical of Micron, its employees, or Micron’s products/services.
5. Permission to use the content may be canceled/terminated by Micron at any time upon written notice from Micron to You if You fail to comply with the terms herein.
6. You acknowledge and agree that the content is provided by Micron to You on an “as is” basis without any representations or warranties whatsoever, and that Micron shall have no liability whatsoever arising from Your use of the content. Micron shall ensure that the content does not violate any statutory provisions and that no rights of third parties are infringed by the content or its publication. Otherwise, liability of the parties shall be limited to intent and gross negligence.
7. You acknowledge and agree that the content is the copyrighted material of Micron and that the granting of permission by Micron to You as provided for herein constitutes the granting by Micron to You of a non-exclusive license to use the content strictly as provided for herein and shall in no way restrict or affect Micron’s rights in and/or to the content, including without limitation any publication or use of the content by Micron or others authorized by Micron.
8. Except for the above permission, Micron reserves all rights not expressly granted, including without limitation any and all patent and trade secret rights. Except as expressly provided herein, nothing herein will be deemed to grant, by implication, estoppel, or otherwise, a license under any of Micron’s other existing or future intellectual property rights.

How to cite sources from the Micron Educator Hub

- Micron is committed to collaborate with Educators to make semiconductor memory education resources available through the Micron Educator Hub.
- The content in the Micron Educator Hub has been identified by Micron as current and relevant to our company.
- Please refer to the table on the right for proper citation.

Use case	How to cite sources
a) Whole slide deck or whole document Description: User uses the whole slide deck or whole document AS IS, without any modification	No additional citation required.
b) Full slide or full page Description: User incorporates a full slide or a full page into their own slide deck or document.	"*© 2020-2025 Micron Technology, Inc. All Rights Reserved. Used with permission.*"
c) Portion of a slide or portion of a page Description: User copies a portion of a slide or a portion of a page into a new slide or page	This is not allowed.

Table of Contents

Introduction to Memory - Goal and Objectives

Participants will be able to understand basic concepts of semiconductor memory technologies.

Objectives:

1. Explain what is Semiconductor Memory
2. Describe some of the fundamental semiconductor devices used in memory (resistor, diode, capacitor, transistor)
3. Describe the basic structure and operation of a MOSFET transistor
4. Describe the basic structure and operation of DRAM memory
5. Describe the basic structure and operation of Flash memories

Introduction to Memory - Target Audience

- This Introduction to Memory module covers the basic structure and operation of DRAM and Flash memory
- Interns, NCGs (New College Grads), and new employees in technical roles need to understand these concepts
- Examples of critical target audience roles at Micron that utilize these concepts:
  - Process Technicians
  - Design Engineer
  - Product Engineer
  - Verification Engineer
  - Process Engineer
  - Process Integration Engineer
  - Test Engineer
  - Probe Engineer
  - Characterization Engineer
  - Reliability Engineer
  - Signal Integrity Engineer
  - Quality Engineer

Pro Tip: everyone interviewing at Micron can use this presentation to prepare for the interview by learning foundational information about memory. Check out the candidate guides for Engineering, Technician and Business roles:

- Micron engineering candidate guide
- Micron technician candidate guide
- Micron business candidate guide

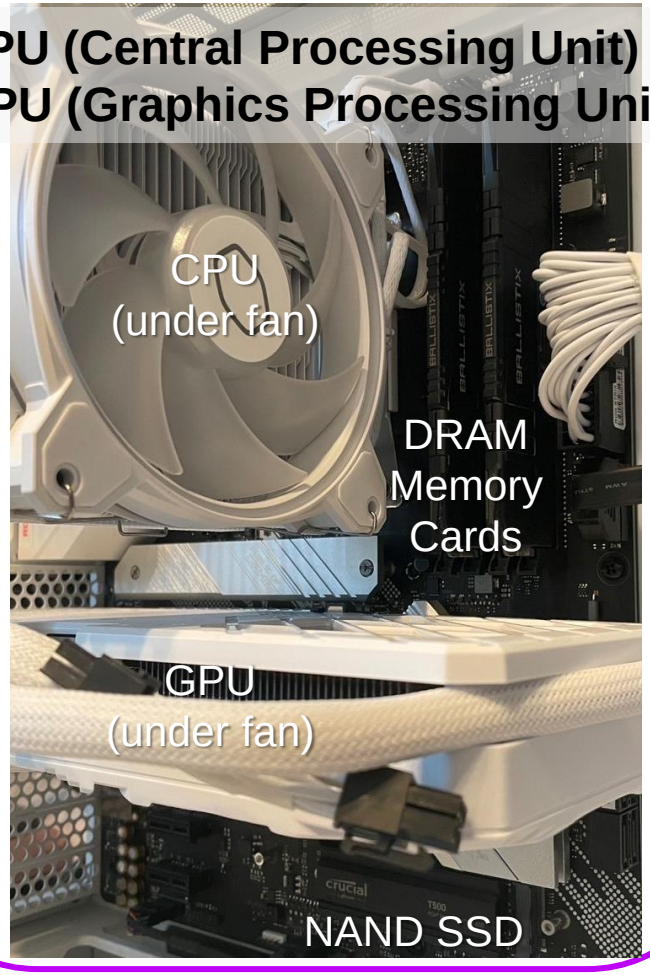
2. Introduction to Semiconductor Memory

micron

Types of Semiconductor Devices

LOGIC DEVICES

CPU (Central Processing Unit)
GPU (Graphics Processing Unit)



MEMORY DEVICES

DRAM

Flash

NAND

NOR

MICRON Products

SPECIALTY DEVICES

CMOS Image Sensors

LEDs

CPUs can have GPU cores integrated on same chip

Logic vs Memory

LOGIC

- A logic chip performs a function on given data
  - A processor or controller chip
  - Logic implies "processing"
    - a fixed operation is performed
    - a set of instructions is executed that may vary

MEMORY

- A memory chip allows you to store & retrieve data
  - A chip that holds programs and data (temporarily or permanently)
    - DRAMs are for temporary workspace
    - Flash memory is used like a disk drive (permanent until erased)

Basic Memory Operations

WRITE OR PROGRAM

WRITE or PROGRAM or store information (1s and 0s) in the memory

Example: Snap a picture with a phone. The picture is “written/programmed” in memory.

READ

READ information from the memory

Example: You look at a picture you took yesterday. You are “reading” it from the memory.

ERASE

ERASE or delete information from the memory

Example: You don’t like the picture, so you “erase” it from the memory!

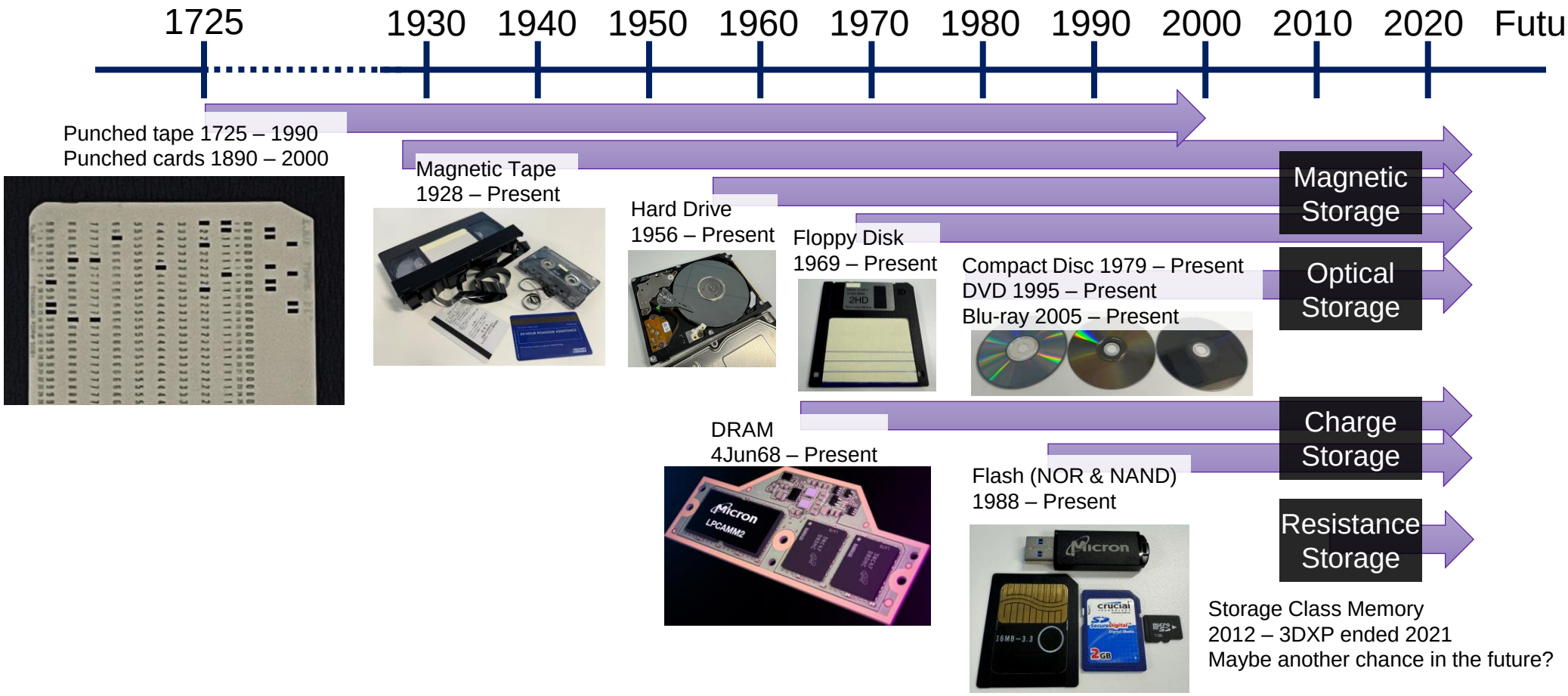
How Memory Works – Binary language

- Electronic systems store information using digital technology
- Digital technology uses BITs and BYTEs to encode information (1 BYTE = 8 BITs)
  - Each BIT is a single piece of information that can have one of two values: 0 or 1
  - Groups of BITs are used to represent whole pieces of information, such as a letter, a number, or a color
  - 1 BYTE of information is required to represent a letter
  - The name Micron is represented by six BYTEs as follows

M	i	c	r	o	n
01001101	01101001	01100011	01110010	01101111	01101110

Memory Storage

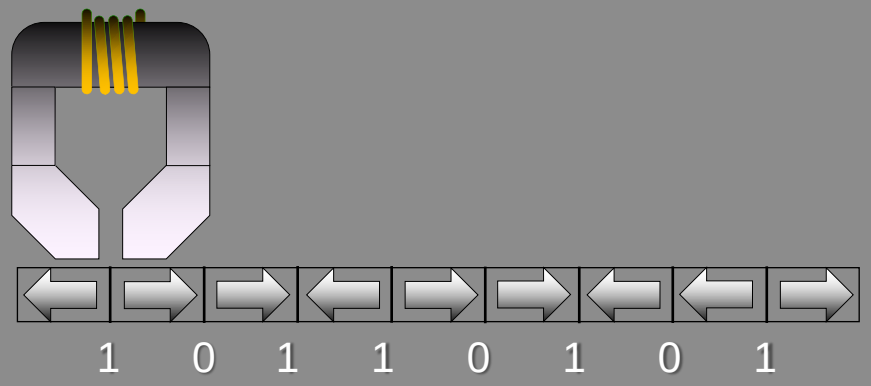
- Memory (data) storage is not a modern invention.
- Over the years many different approaches have been used to store information…



Memory Storage Methods

Magnetic | Optical | Charge | Resistance

Hard Disk Drive



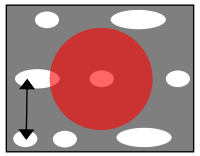
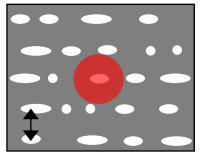
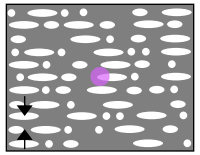
Read head senses changes in direction of domains within the thin magnetic film: change = 1, no change = 0

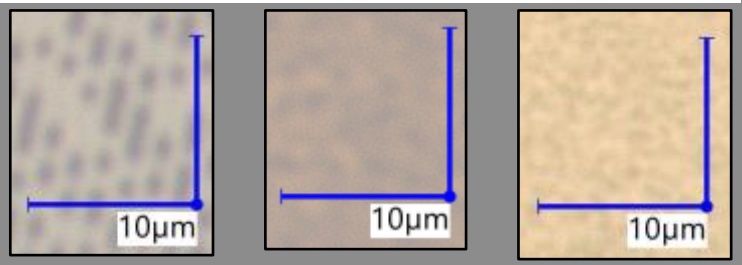
(binary code for “μ”)

Memory Storage Methods

Magnetic | Optical | Charge | Resistance

CD ROM, DVD, Blu-Ray

Compact Disc 780nm Red Laser	**Digital Video Disc** 650nm Red Laser	**Blu-Ray Disc** 405nm Blue Laser
		
Track pitch 1.6µm	Track pitch 0.74µm	Track pitch 0.30µm



Optical microscope photos

Storing Binary 1's and 0's with indentations in a metallic film

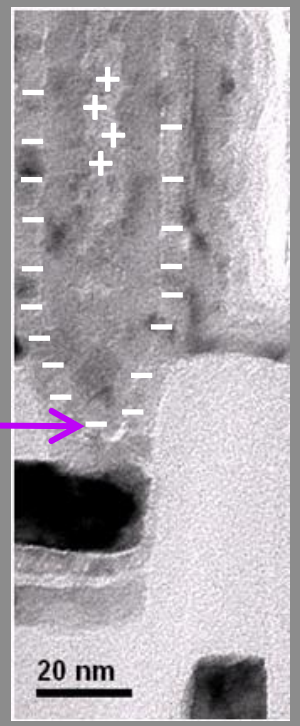
Storage Capacity:
750MB x 6 = 4.5GB x 5.6 = 25GB

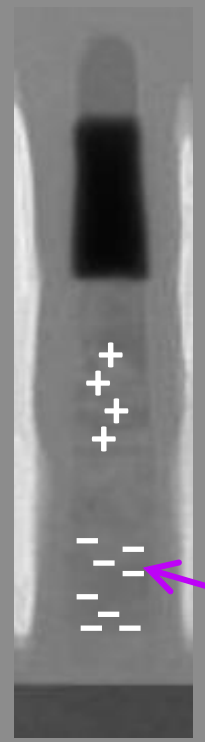
Memory Storage Methods

Magnetic | Optical | Charge | Resistance

DRAM | NAND Flash

Storing Binary 1's and 0's based on how much charge is kept in the storage node

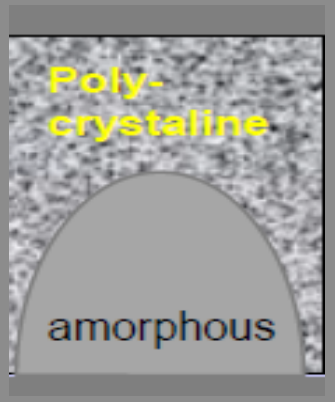




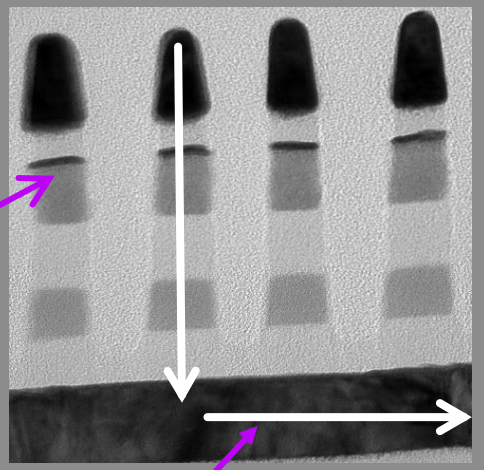
Memory Storage Methods

Magnetic | Optical | Charge | Resistance

Cross-Point



Storage Node



Electrical Current

Storing Binary 1's and 0's based on how much resistance there is to current flowing through the storage node

Major Types of Memory

- **Memory can be classified as Volatile or Non-Volatile**
- **Volatile** (example: DRAM)
  - Memory that will lose stored information when power is removed from the part
- **Non-Volatile** (example: NOR & NAND Flash)
  - Memory that will retain stored information even when power is removed from the part

Quiz: What type of memory (volatile or non-volatile) do you want for the memory that stores the pictures you take in a smart phone?

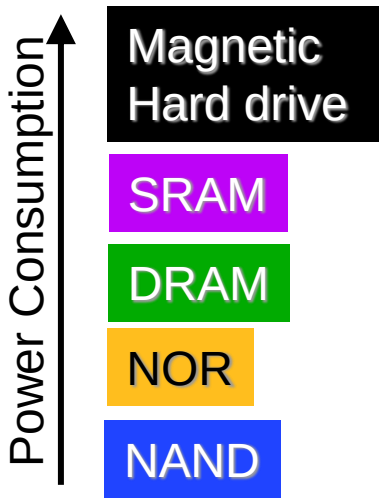
Summary of Silicon-Based Memory Technologies

Short Name	Full Name	Description	Type	Speed	Density	Primary Applications
SRAM	Static Random Access Memory	Data is stored as the state of a digital flip-flop. Does not require refresh.	Volatile	Very Fast Read and Write	Low	Cache between CPU and other memory types.
DRAM	Dynamic Random Access Memory	Data is stored as charge on a capacitor. Charge is quickly lost and must be constantly “refreshed”.	Volatile	Fast Read and Write	High	“Scratch Pad” for a CPU. Computers, servers, laptops, tablets, PDA’s, cell phones, AI applications, etc.
NAND	NAND Flash Memory	Data is stored by “trapping” charge in a film. Data can be stored for up to 10 years without need for refresh.	Non-Volatile	Slow Read, Very Slow Write	Very High	Long term storage: digital cameras, cell phones, MP3s, memory sticks, SSDs, etc.
NOR	NOR Flash Memory	Similar to NAND Flash, but the array is configured for faster read/write.	Non-Volatile	Faster access time than NAND	Medium	Applications that require fast access time like code execution. Example: OS (Operative System)

NAND/NOR are not acronyms:
NAND refers to the series arrangement of memory cells similar to NAND (Not AND) digital logic
NOR memory cells are arranged in parallel like NOR (Not OR) digital logic

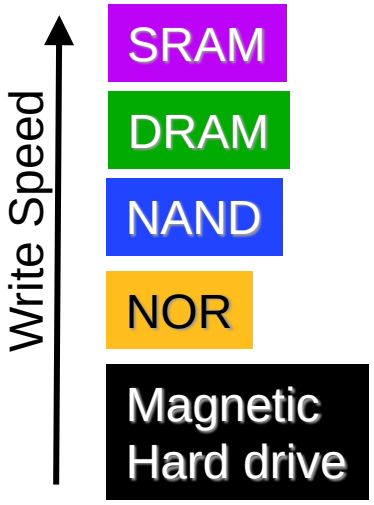
Comparison of Memory Types

Power Consumption



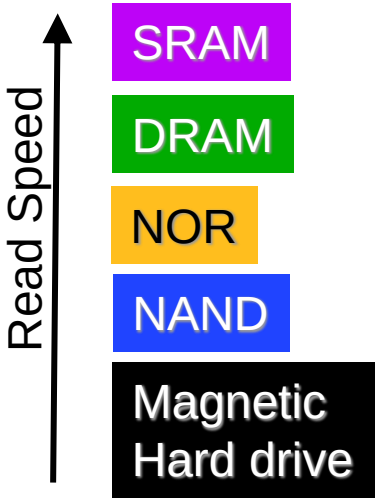
Lower is better
to extend battery charge
in mobile devices and reduce
operating costs in data centers

Write Speed



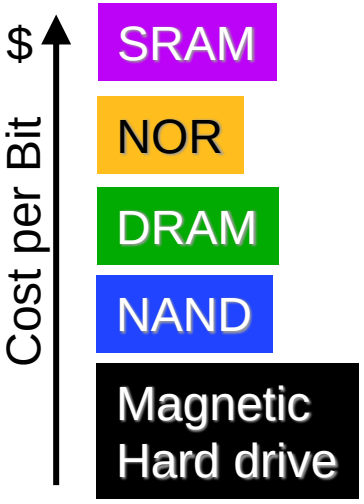
Higher is better
for faster operation

Read Speed



Higher is better
for faster operation

Cost per Bit



Lower is better
for higher density
at similar price point

DRAM = Dynamic Random Access Memory
DDR = Double Data Rate
LPDDR = Low Power Double Data Rate
GDDR = Graphics Double Data Rate
HBM = High Bandwidth Memory

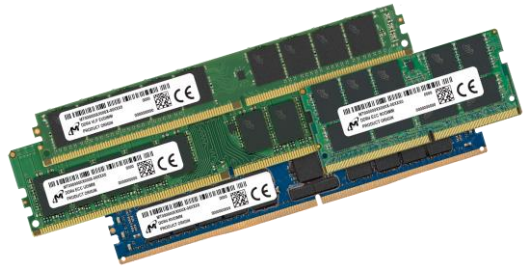
DRAM (DDR, LPDDR, GDDR & HBM)

Smartphone | Automotive ADAS | Data Center | Personal Computing | PC Gaming | Game Console

Content Creation

Smart factory & Robotics

- High-performance volatile memory
- Modules & components
- Low power consumption
- High-bandwidth
- System level solutions



AR/VR Applications

Work from Anywhere

Industrial IoT Applications | Telco & Edge | Aerospace and Defense | Database/ Hypervisor | Generative AI | Medical Equipment

Flash Memory (NAND and NOR)

ADAS = Advanced Driver Assistance Systems
AI = Artificial Intelligence
AR = Augmented Reality
IoT = Internet of Things
VR = Virtual Reality

Smartphone

Automotive ADAS

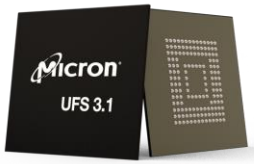
Data Center

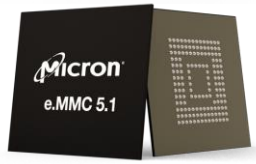
Personal Computing

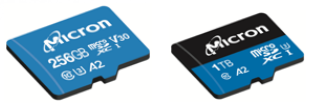
Portable Storage

Game Console

Video Surveillance







- Long-term memory storage
- Security Protection
- Low Power Consumption
- Reprogrammable
- Booting Software

AR/VR Applications

Smart factory & Robotics

Point of Sales / Wearable

Industrial IoT Applications

Networking

Aerospace and Defense

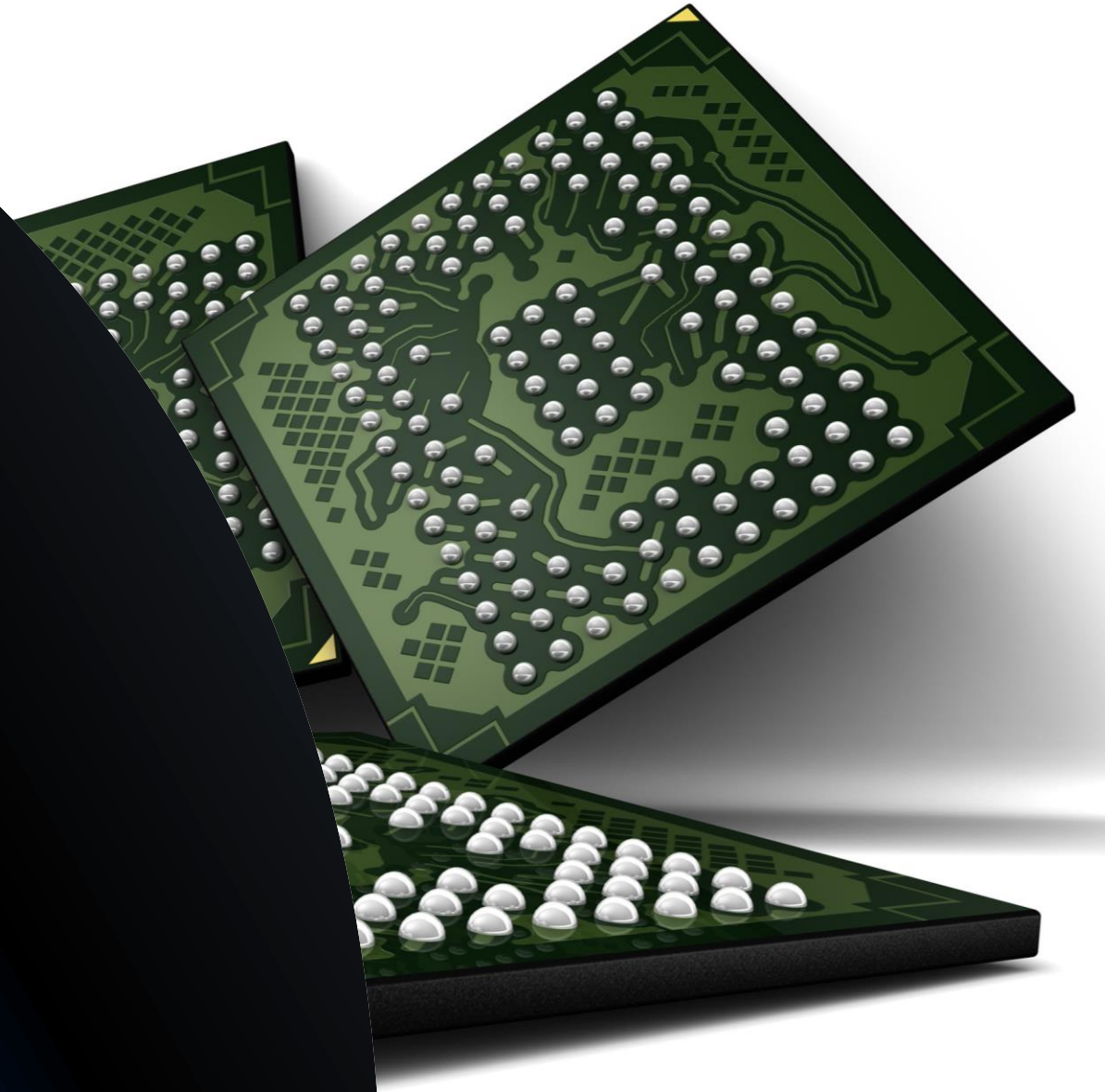
Energy

Drones and Transport

Medical Equipment

3. Semiconductor Devices Overview

Fundamentals | Capacitors | Transistors



micron

Conductors, Insulators & Semiconductor Materials

Conductors

- Allow electrical current to flow easily
- Conductors used in integrated circuits are mostly metals: Aluminum, Tungsten and Copper

Insulators

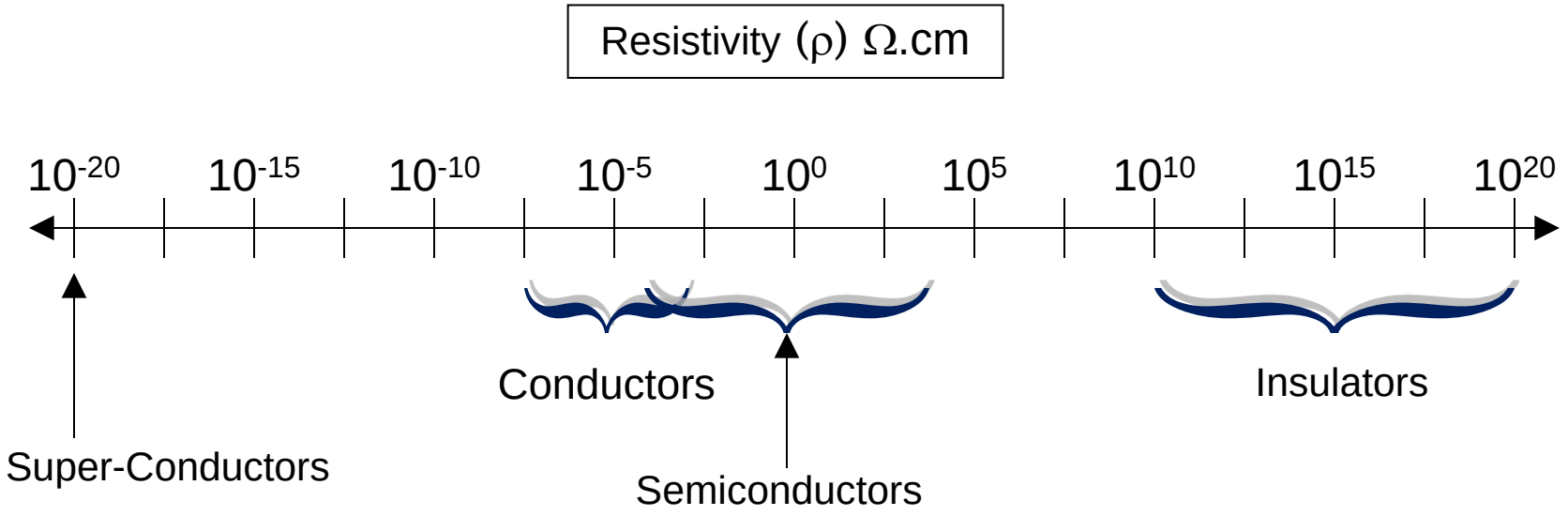
- Very resistive to electrical current flow
- Used to electrically isolate one part of the circuit from another
- Some insulators used in integrated circuits: Silicon dioxide and Silicon nitride

Semiconductors

- Between conductors and insulators on the resistivity scale
- The semiconductor used in Micron is Silicon
- Silicon resistivity can be changed via “doping” (see upcoming slide)

Resistivity Scale

- Electrical resistivity is a measure of how strongly a material opposes the flow of electric current.
- Low resistivity indicates a material that readily allows the movement of electrical charge.



Semiconductor Doping

Intrinsic (pure) Silicon

- Very resistive to electrical current flow
- Intrinsic silicon has very few applications in memory chips
- Silicon has 4 valence electrons (in outer shell)

Doped Silicon

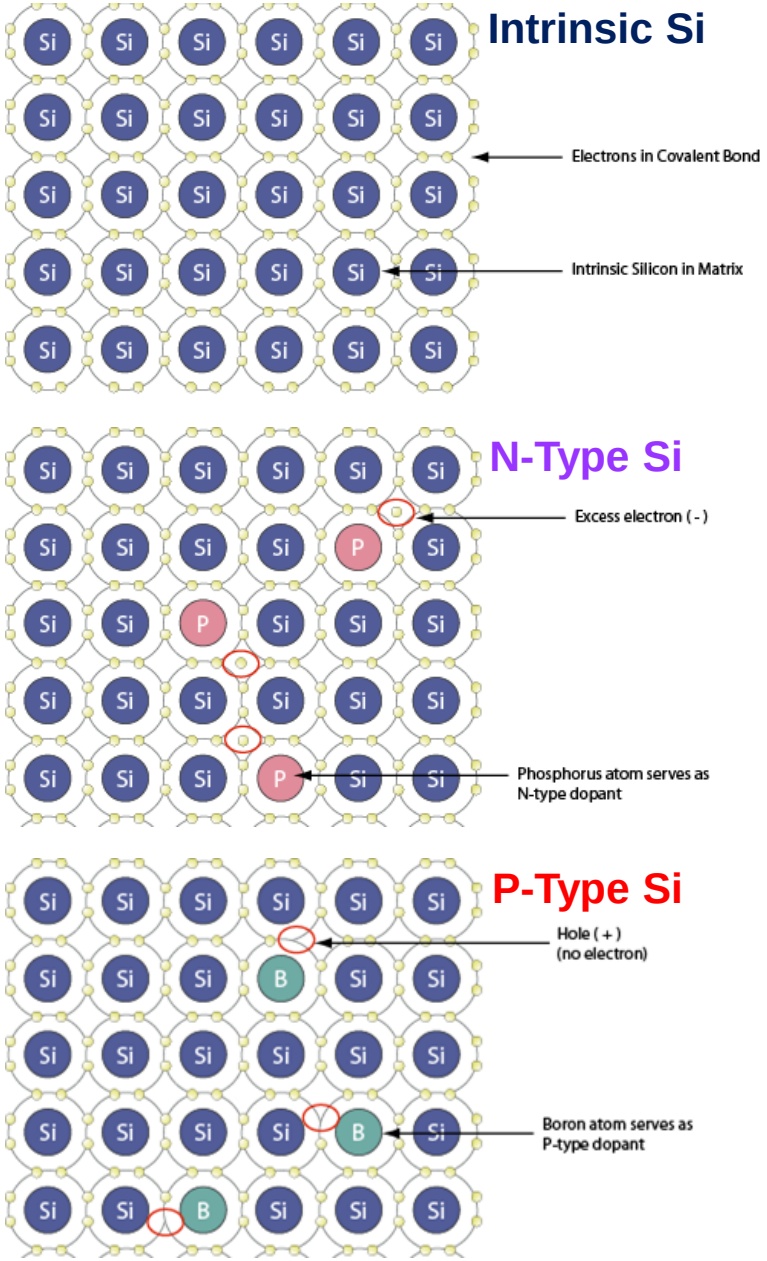
- A small percentage of impurities (“dopants”) are added to the silicon
- Dopants reduce the resistance of the silicon
- Dopants can be “N-type” or P-type”

N-Type Silicon

- Most common N-Type Dopants are Arsenic (As) or Phosphorous (P)
- Current is carried by electrons that have a negative charge (N-Type)
- Phosphorous and Arsenic have 5 valence electrons (in outer shell)

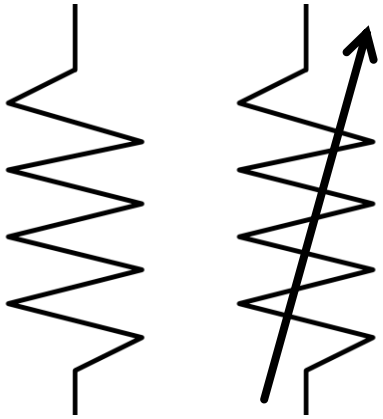
P-Type Silicon

- Most common P-Type Dopants is Boron (B)
- Current is carried by “holes” that have a positive charge (P-Type)
- Boron has 3 valence electrons (in outer shell)



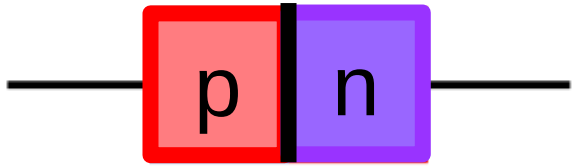
Important Semiconductor Components

Resistors



- Reduces or limits current flow

Diodes

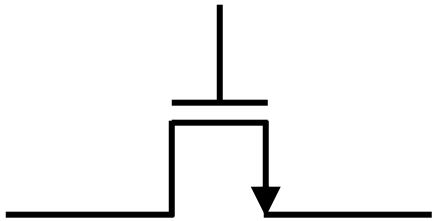


- Allows current to flow in one direction
- Can be used to electrically isolate adjacent devices from each other

Capacitors

- Stores an electrical charge
- Has two conductive plates separated by an insulator

Transistors



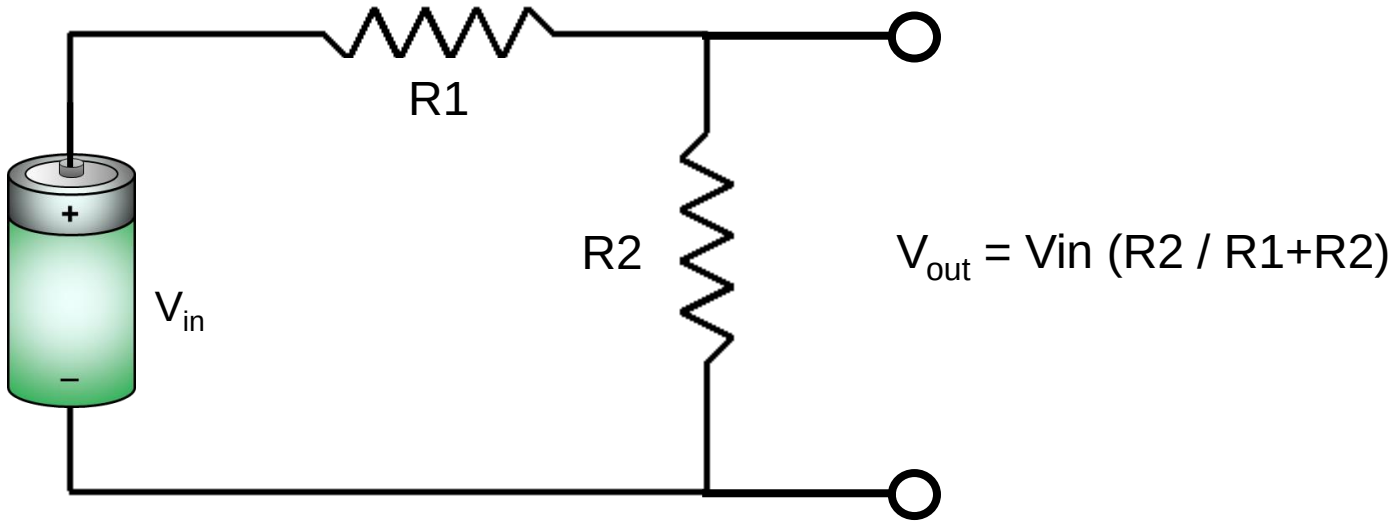
- Controls current flow
- Can function like a switch or valve

Resistors

Definition: A device that reduces or limits current flow

- A resistor is created by using a material less conductive (more resistive) anywhere along the path of current flow
- Resistors can have a fixed  or variable  resistance value

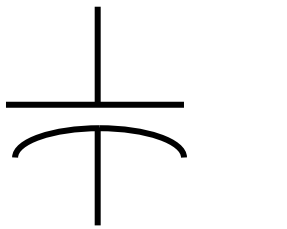
Voltage Divider Circuit



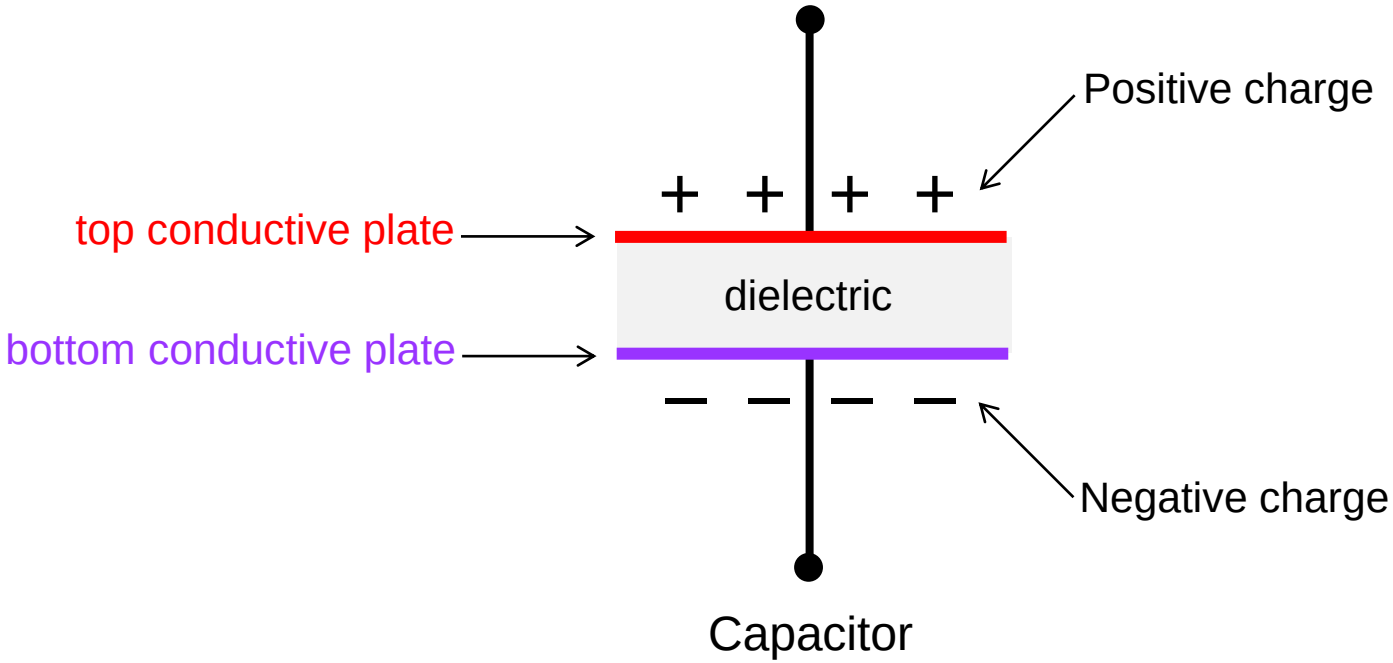
Capacitors

Definition: An electronic device used to store electric charge

- A capacitor consists of:
  - Two conductive “plates” – we arbitrarily call them “top” and “bottom” plates
  - An insulating layer known as a “dielectric” that electrically separates the two plates
- Opposite charges on the plates are held in place by their mutual attraction, but they cannot cross the dielectric layer

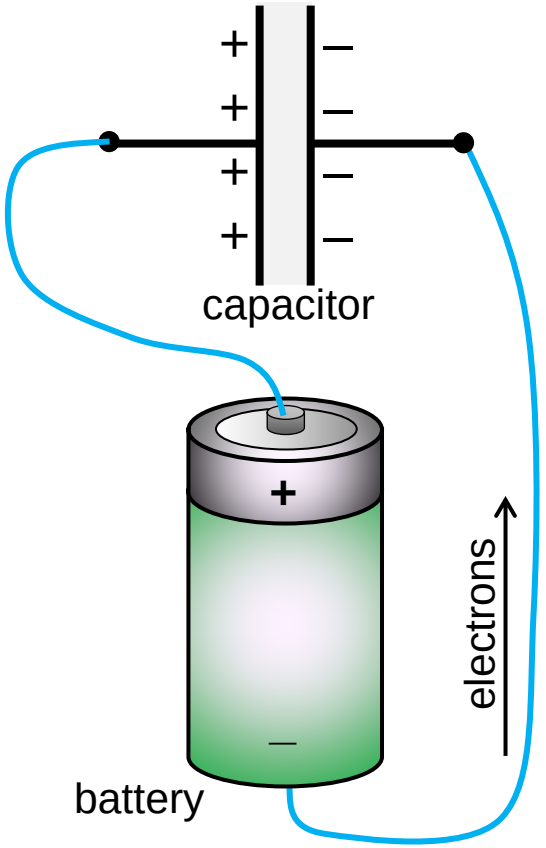


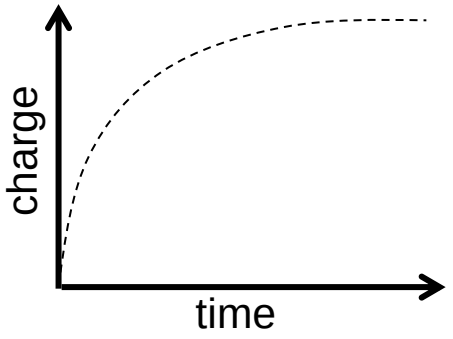
Circuit symbol for a capacitor



Capacitors

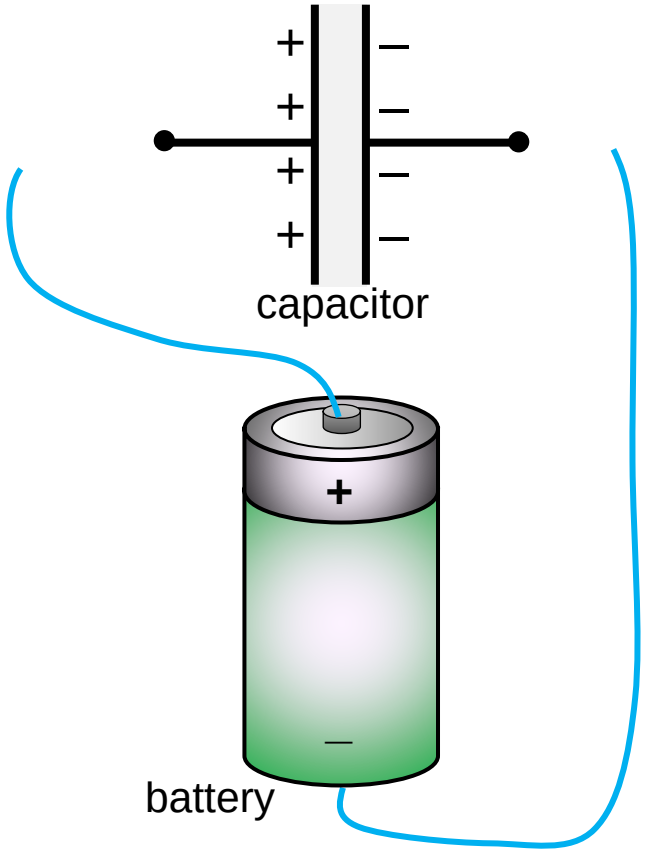
- If a capacitor is hooked up to a power source, such as a battery, the capacitor will “charge up” until it can hold no more charge
- Analogy: This is like filling a bucket with water until it can hold no more water

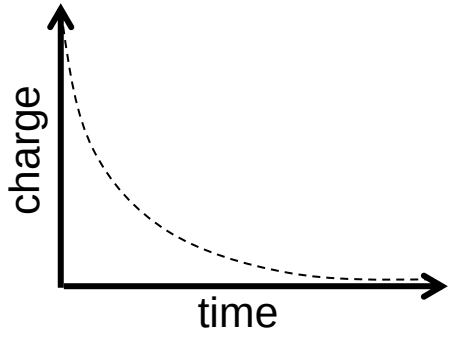


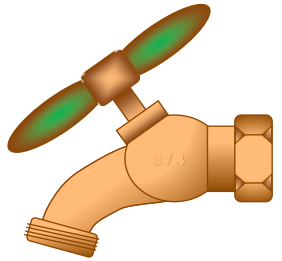


Capacitors

- When the battery is disconnected, the capacitor will gradually begin to lose the charge. This is known as “discharging”.
- Analogy: This is like water slowly leaking out of small holes in a bucket





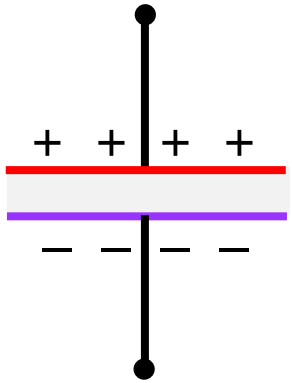


Capacitors

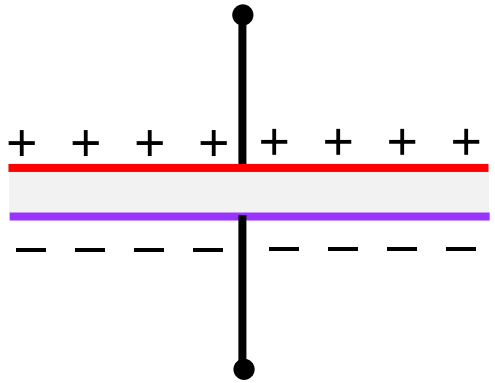
How much charge can a capacitor hold?

C ≈ k x plate area / spacing

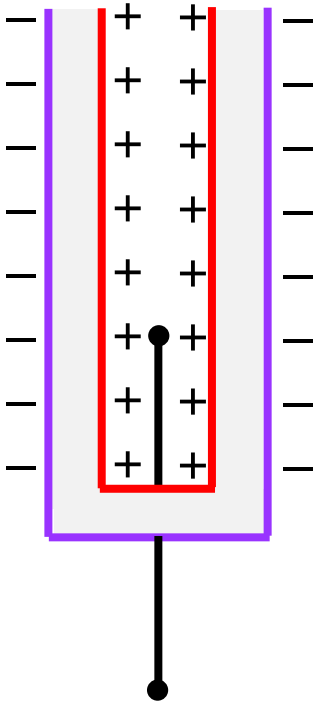
- As the surface area of the conductive plates is increased, the capacitor can hold more charge
- Note: The plates do not have to be flat!
- The amount of charge or energy storage in a capacitor is called the “capacitance”



Small flat plates → less charge can be stored

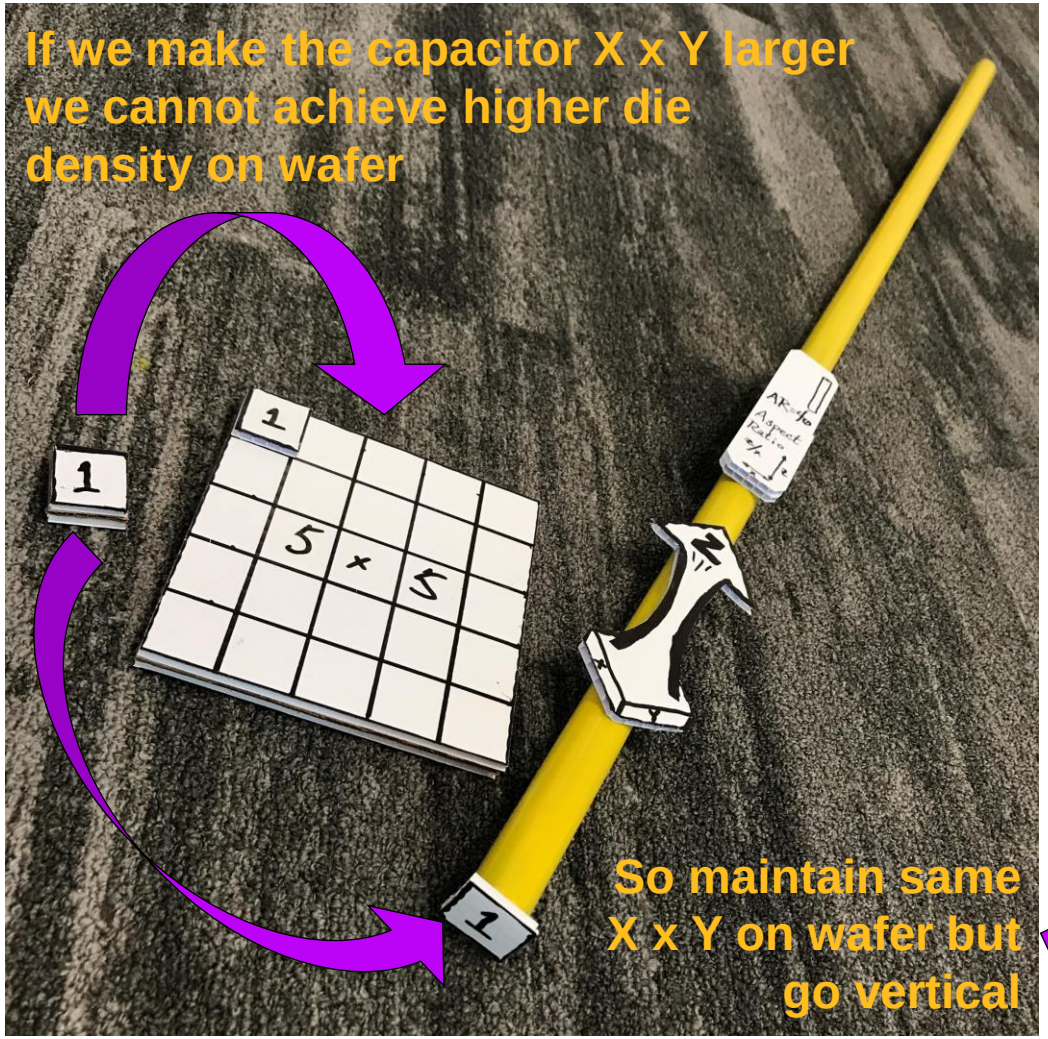


Large flat plates → more charge can be stored but require *more* X x Y area on wafer surface



Large cylindrical plates → even more charge can be stored with *less* area on the wafer!

Into the Z Dimension!

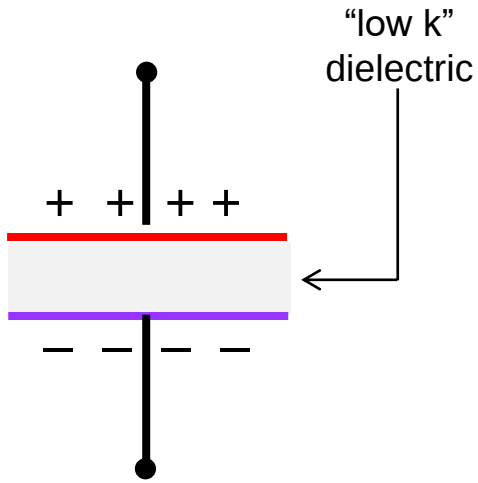


Capacitors

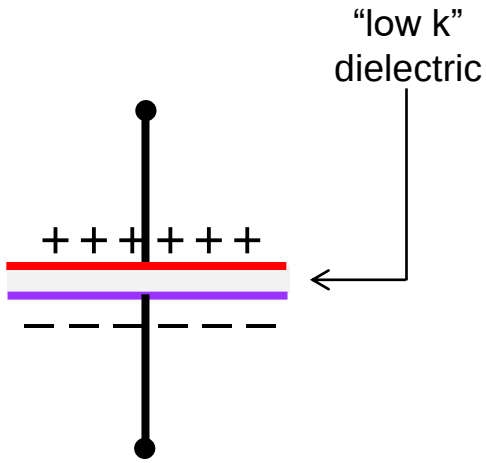
How much charge can a capacitor hold?

$C \approx k \times \text{plate area} / \text{spacing}$

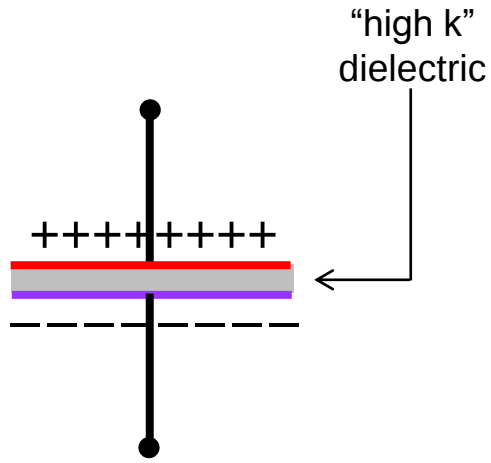
- As the spacing between the plates decreases, the capacitor can hold more charge. This spacing can also be thought as the dielectric thickness.
- The choice of dielectric material can also increase the amount of charge that can be stored.
  - "High k" dielectric materials can store more charge.



Wide space between plates → less charge can be stored



Narrow space between plates → more charge can be stored



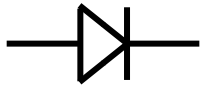
Narrow space, with special dielectric → even more charge can be stored

Dielectric Constant k	
SiO_2	3.9
Si_3N_4	7.5
Ta_2O_5	22
HfO_2	25
ZrO_2	25
TiO_2	80

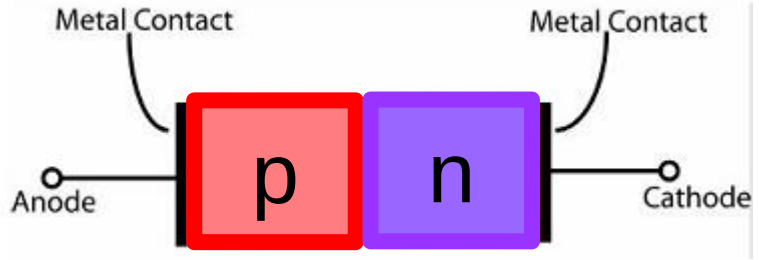
Diodes

Definition: An electronic device that allows current to flow in one direction only

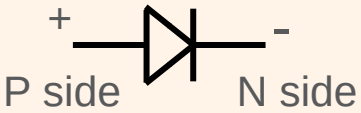
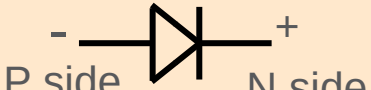
- Formed by connecting a P-type material with an N-type material



Schematic symbol for a diode

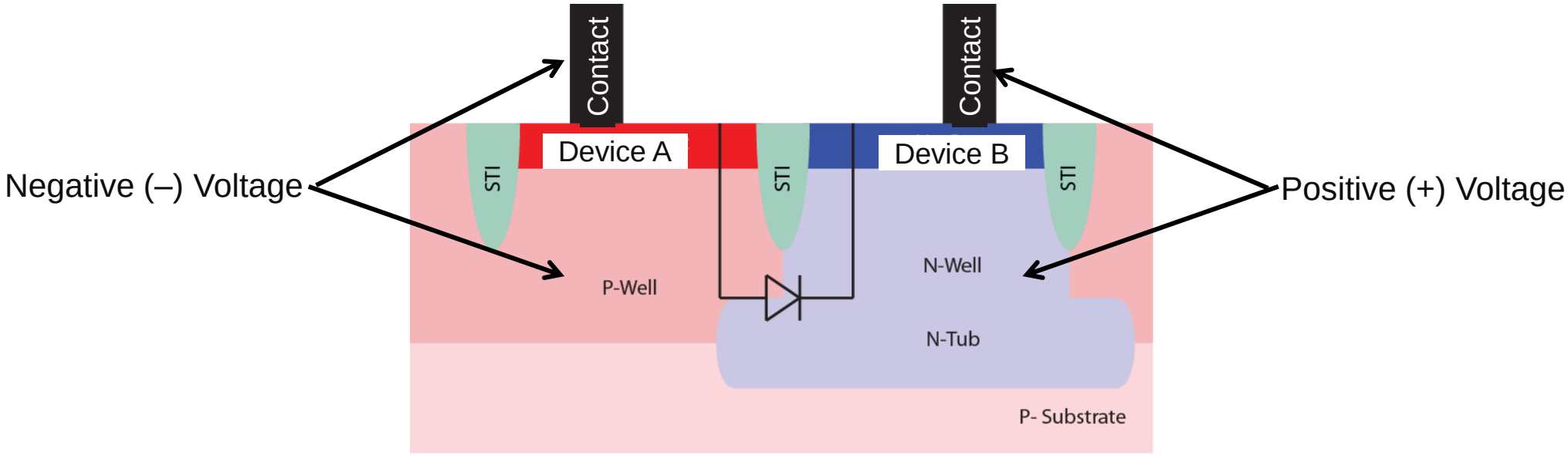


Simplified physical structure of a diode

Diodes	Forward biased:		Current flows
	Reverse biased:		Negligible current flows

Diodes

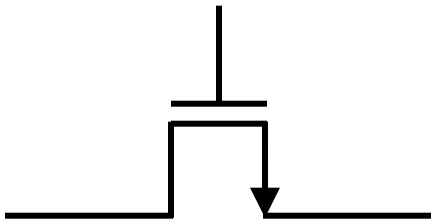
- Diodes can be used to electrically isolate adjacent devices from each other
  - In the example below, the P-Well and N-Well form a diode where they meet
  - By applying a negative voltage to the P-Well and a positive voltage to the N-Well, the diode is reverse-biased, and current cannot flow between devices A and B



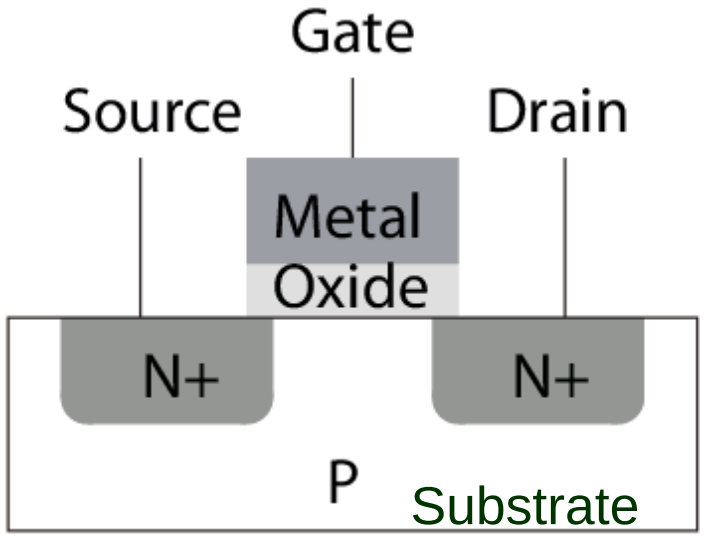
Transistors

Definition: An electronic device that can control the current running though it – like an On/Off switch

- The most common type is the *Field Effect Transistor* (FET)
  - Used in digital applications
  - Low-power device
  - Still being scaled below 10 nm



Circuit symbol of a
MOSFET Transistor



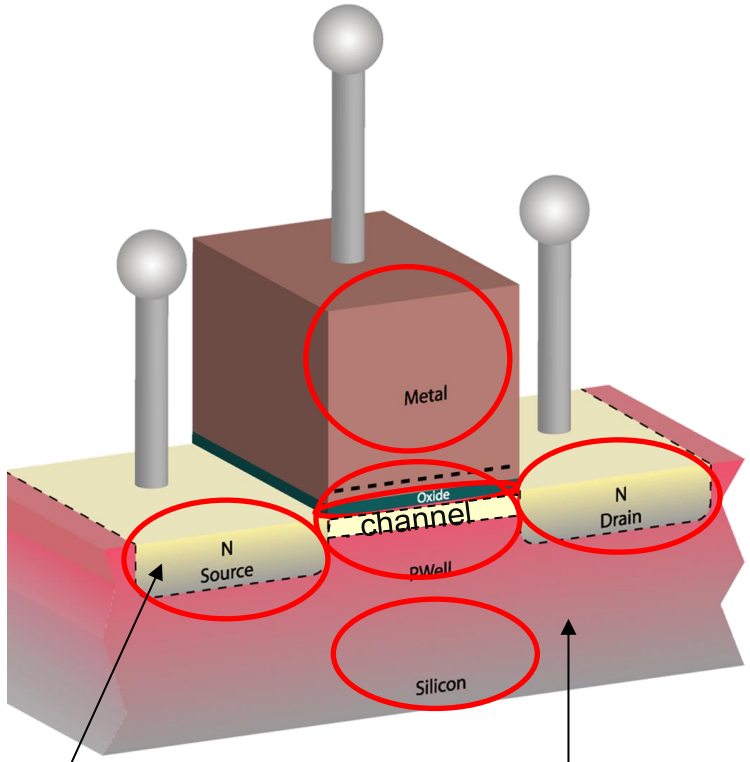
Structure of a MOSFET Transistor

MOSFET: Metal Oxide Semiconductor Field Effect Transistor

MOSFET Overview

Structure of a MOSFET

- Doped semiconductor (silicon) substrate
- Doped source and drain regions (oppositely doped from substrate)
- Insulating gate oxide
- Metal control gate on top of oxide is used to apply voltages that turn the transistor ON or OFF
- The “channel” is the region beneath the gate where electrons will flow from source to drain under the right conditions



Doped N-type in this example

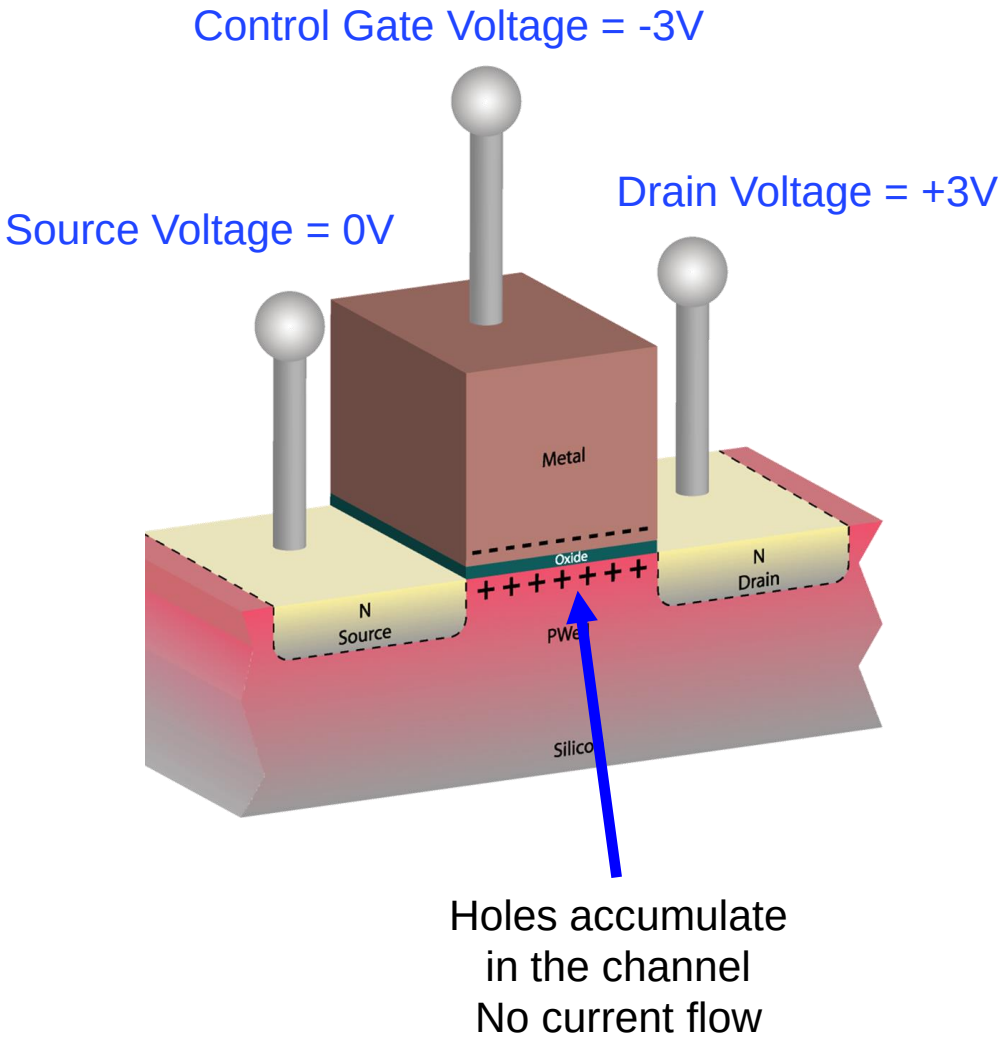
Doped P-type in this example (also called the Pwell)

MOSFET: Metal Oxide Semiconductor Field Effect Transistor

MOSFET Overview

OFF STATE

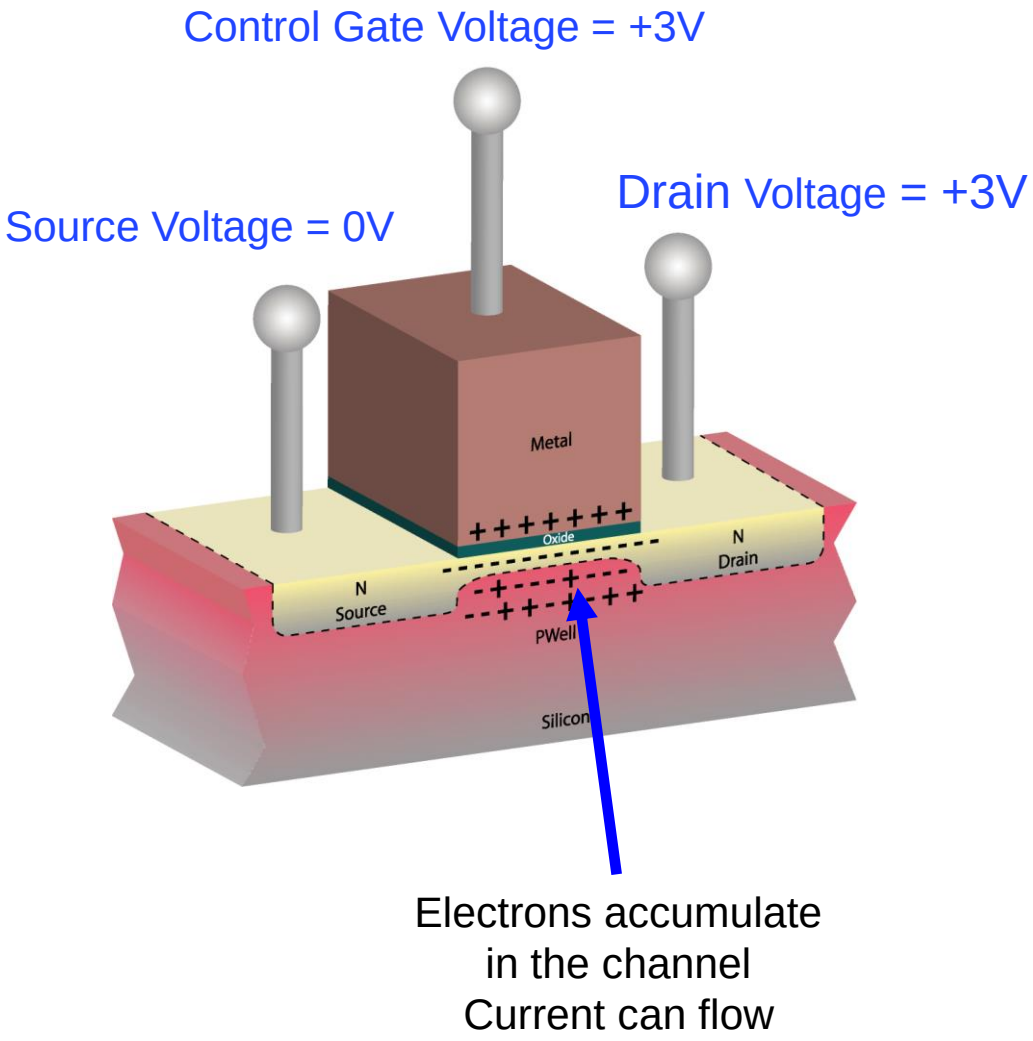
- Drain voltage is higher than source voltage, creating a voltage drop. Current could flow through the device under these conditions if the gate allows.
- Negative voltage applied to the control gate attracts positive charges (“holes”) into the channel region
- The holes in the channel act as a barrier so current cannot flow between the Source and Drain
- The transistor is OFF



MOSFET Overview

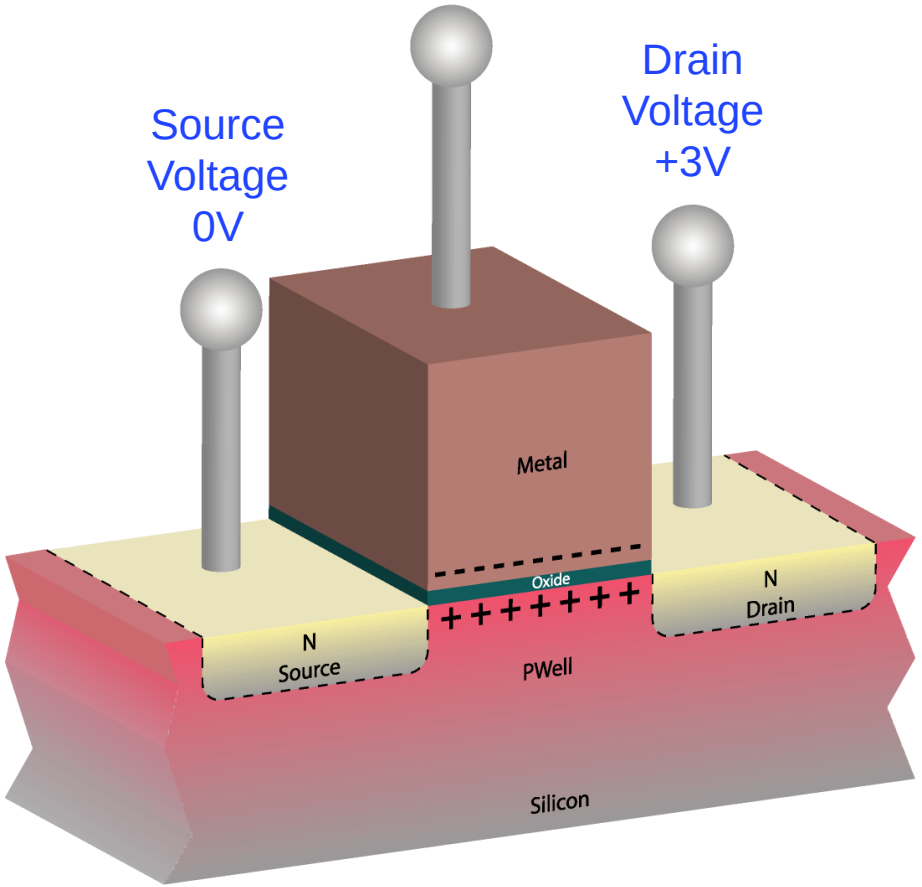
ON STATE

- Drain voltage is higher than source voltage, creating a voltage drop. Current could flow through the device under these conditions if the gate allows.
- Positive voltage applied to the control gate attracts negative charges (electrons) into the channel region
- Electrons in the channel create an electrical connection between the Source and Drain, allowing current to flow
- The transistor is ON



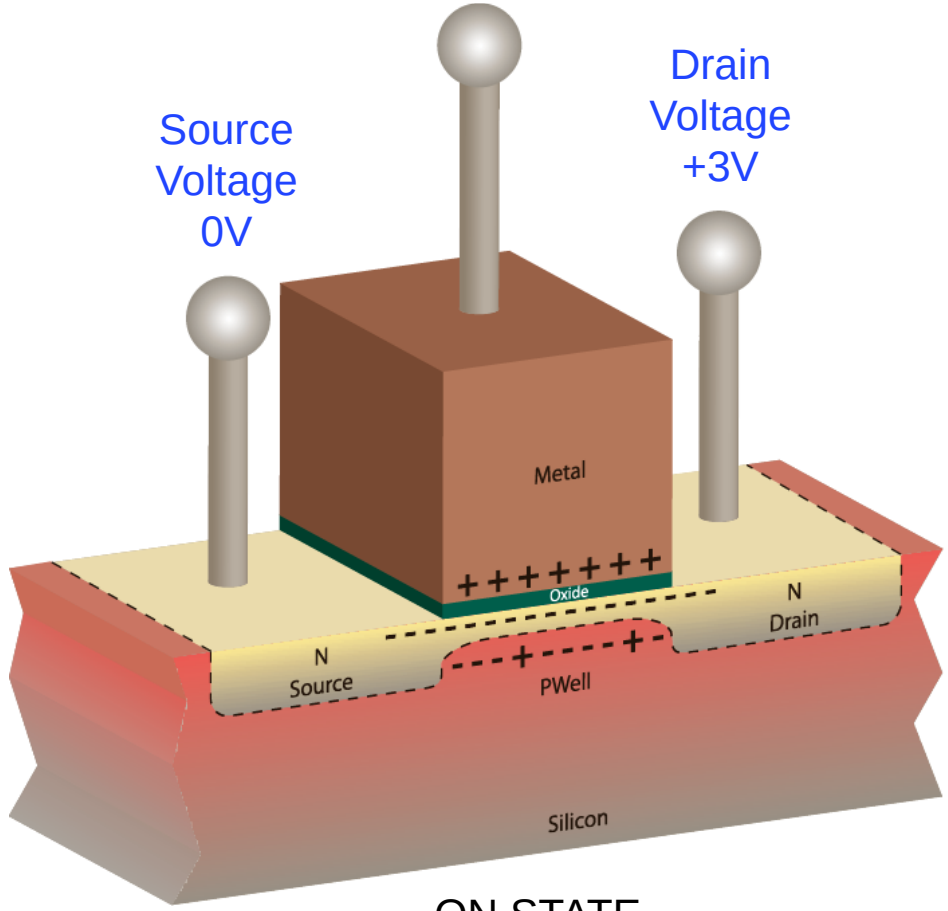
MOSFET States: OFF and ON Summary

Control Gate Voltage = -3V



OFF STATE

Control Gate Voltage = +3V

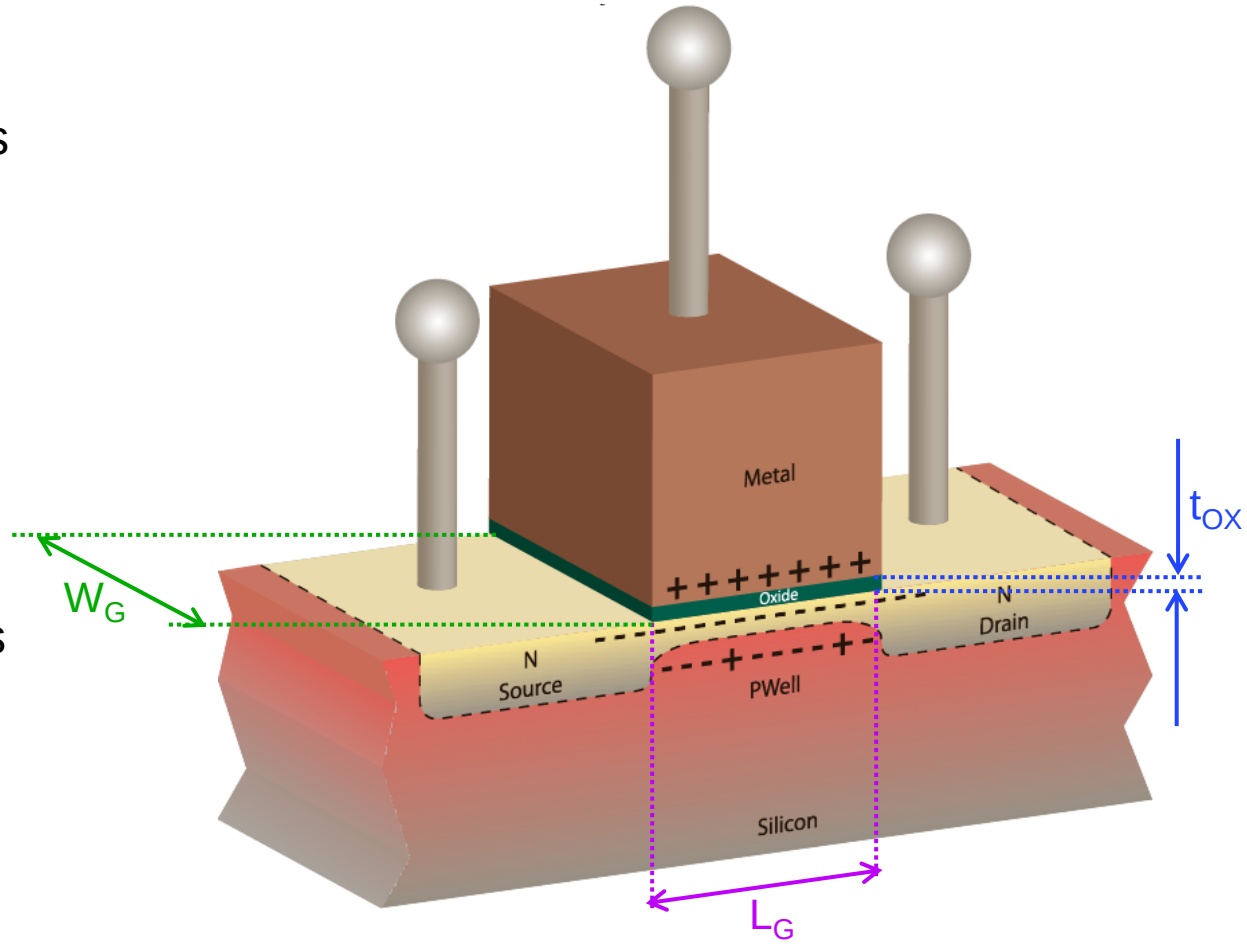


ON STATE

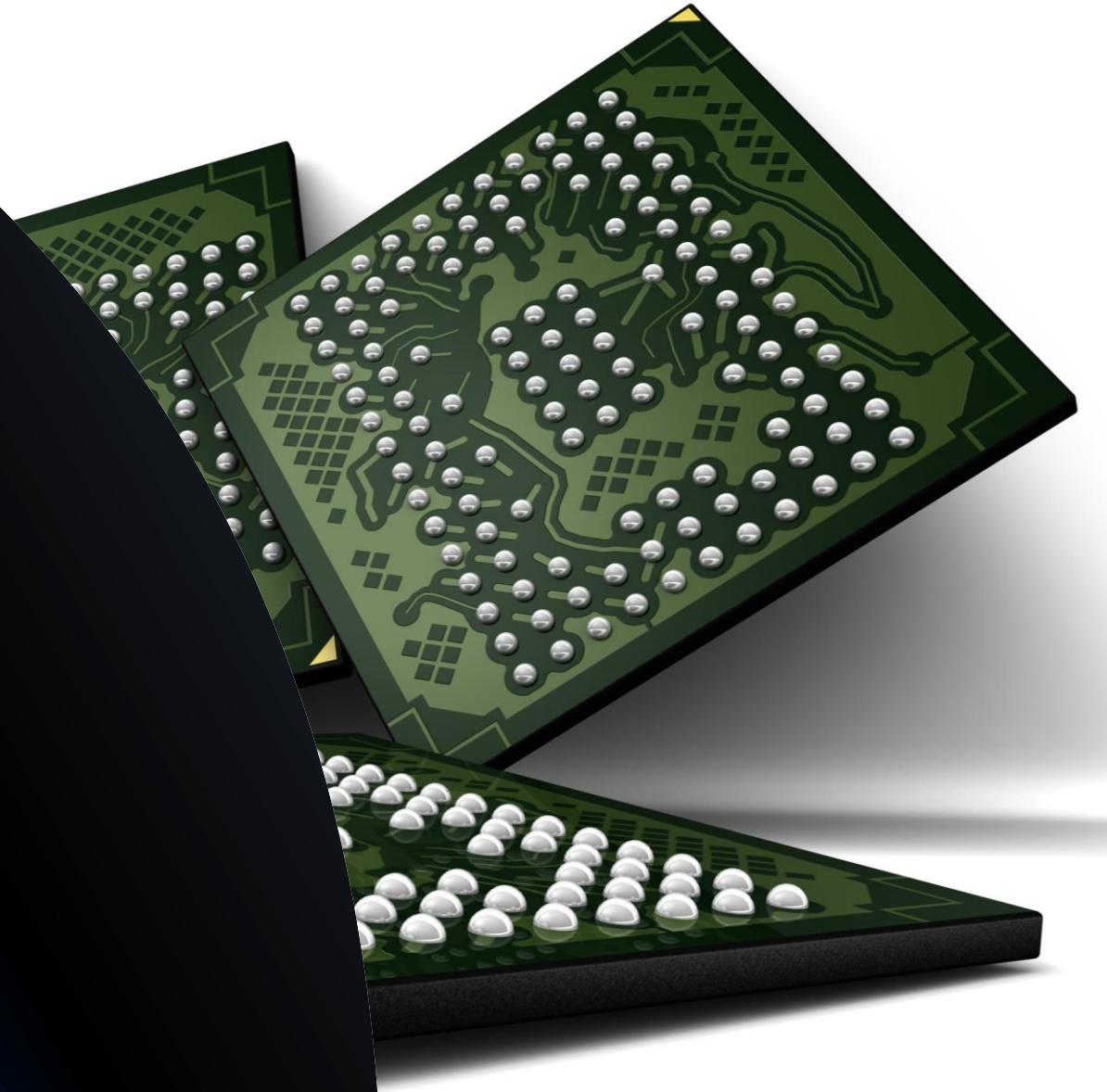
MOSFET Transistor Dimensions

Three Important Dimensions (but many others!):

- Gate Length (L_G): As the gate length (channel) gets shorter, the device can switch faster. The tradeoff is an increase of leakage current in the off-state.
- Gate Width (W_G): As the gate gets wider, more current can flow from drain to source in the on-state which improves circuit speed. The tradeoff is that more silicon area is needed.
- Gate Oxide Thickness (t_{OX}): As the gate oxide gets thinner, the device can switch faster. The tradeoff is an increase of leakage current through the gate.



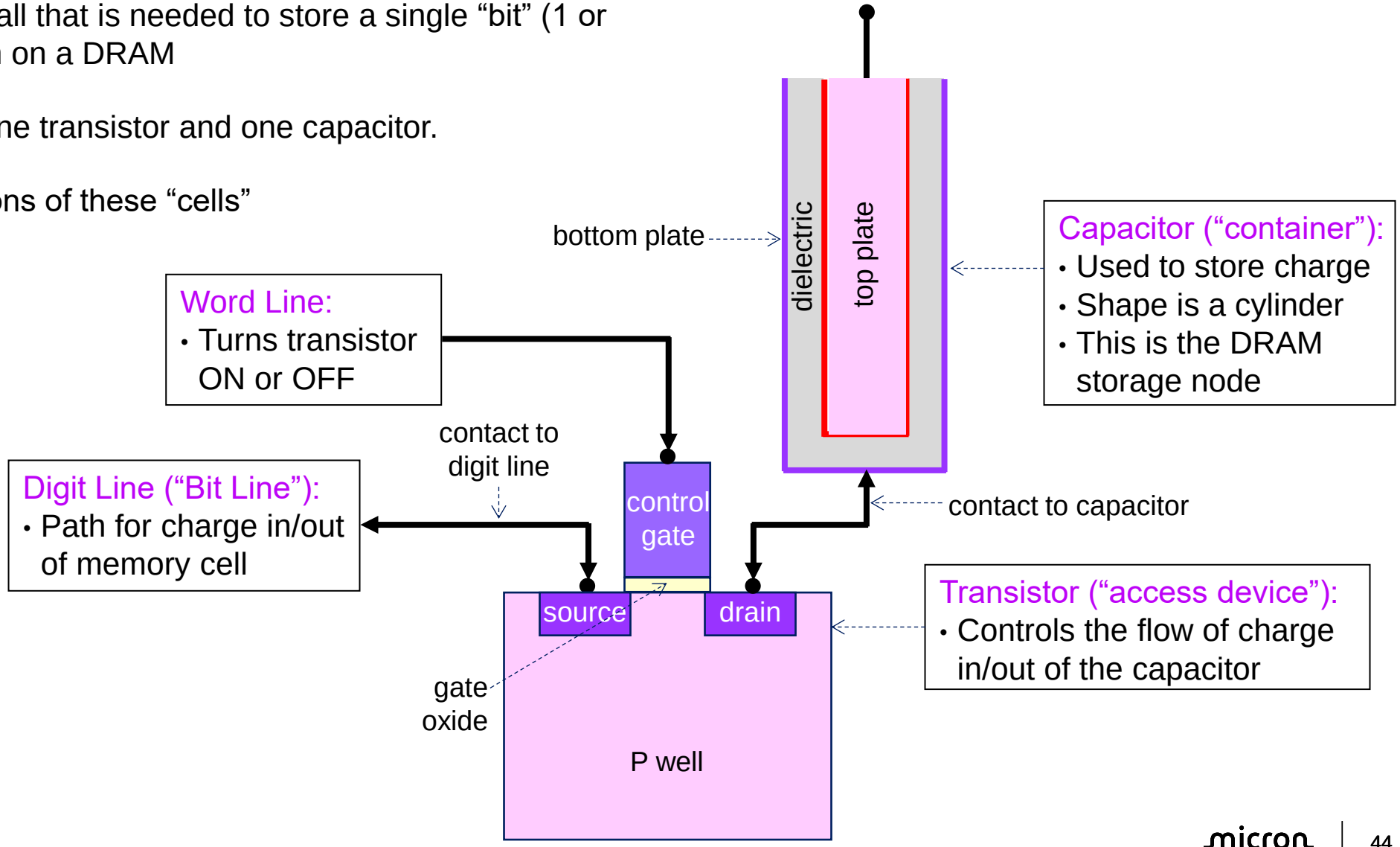
4. Introduction to DRAM



micron

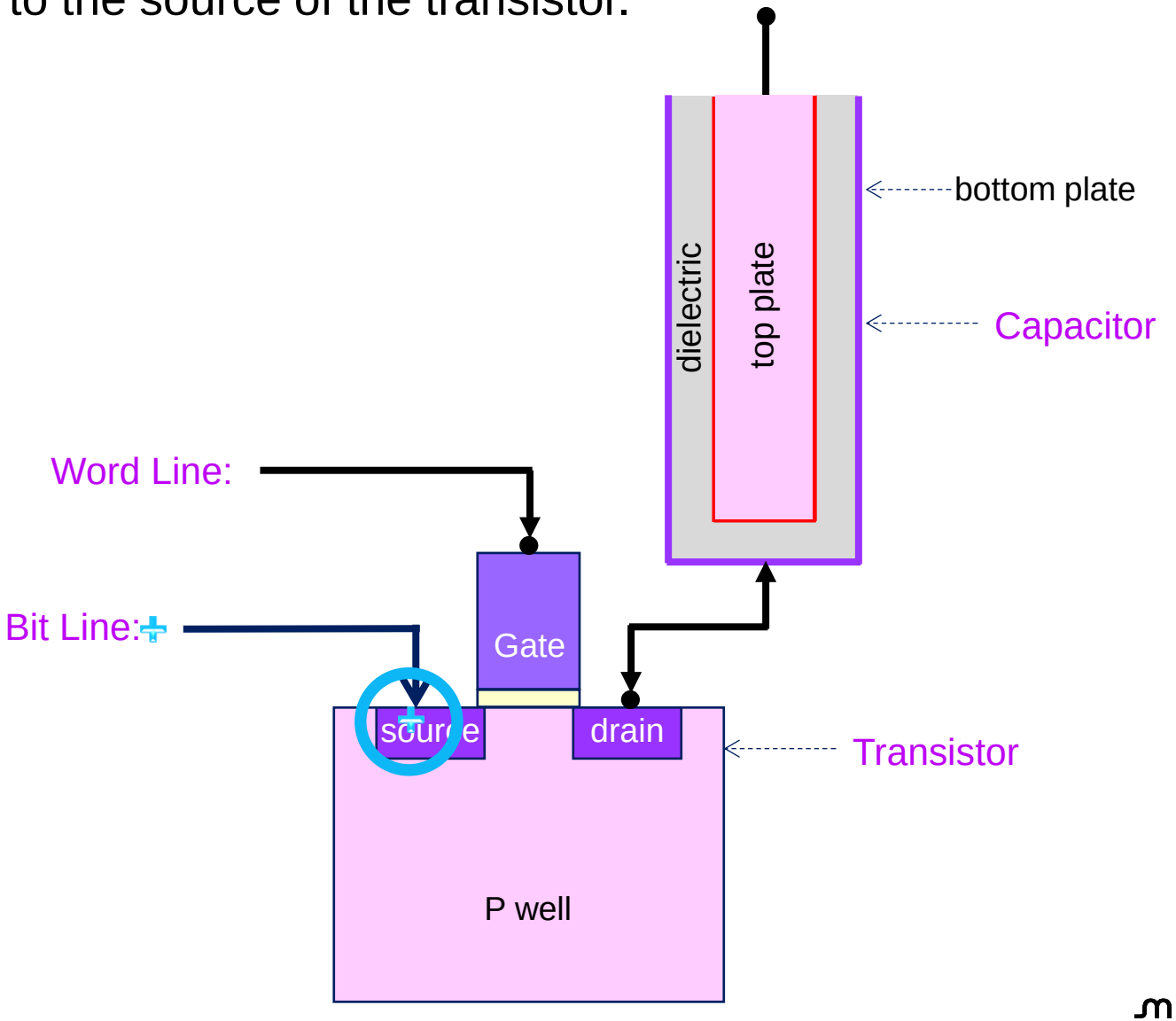
Basic DRAM Memory Cell

- Shown on this slide is all that is needed to store a single “bit” (1 or 0) of digital information on a DRAM
- Each cell consists of one transistor and one capacitor.
- A DRAM chip has billions of these “cells”



DRAM Operation – Write

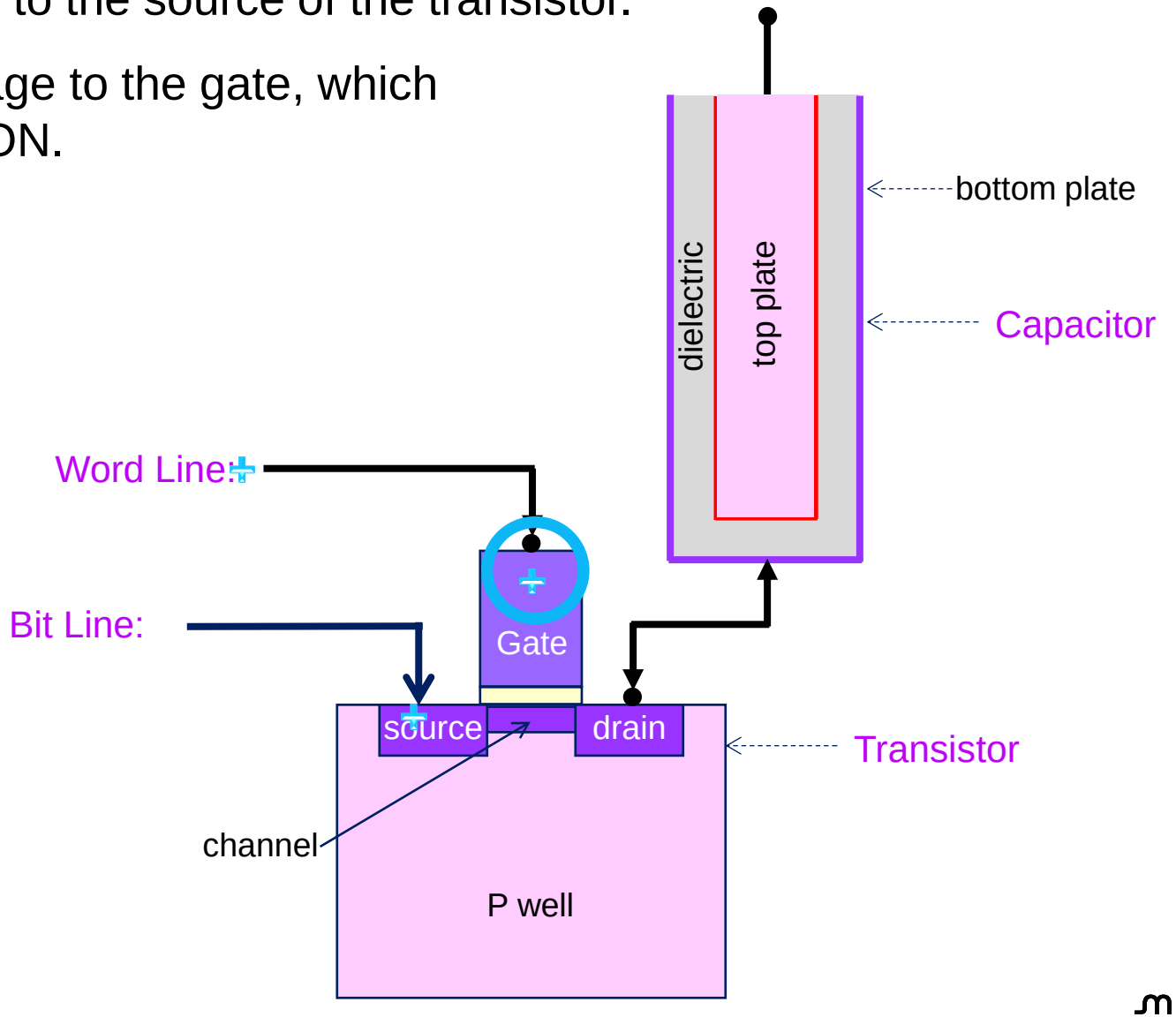
1. Bit Line delivers positive voltage to the source of the transistor.



DRAM Operation – Write

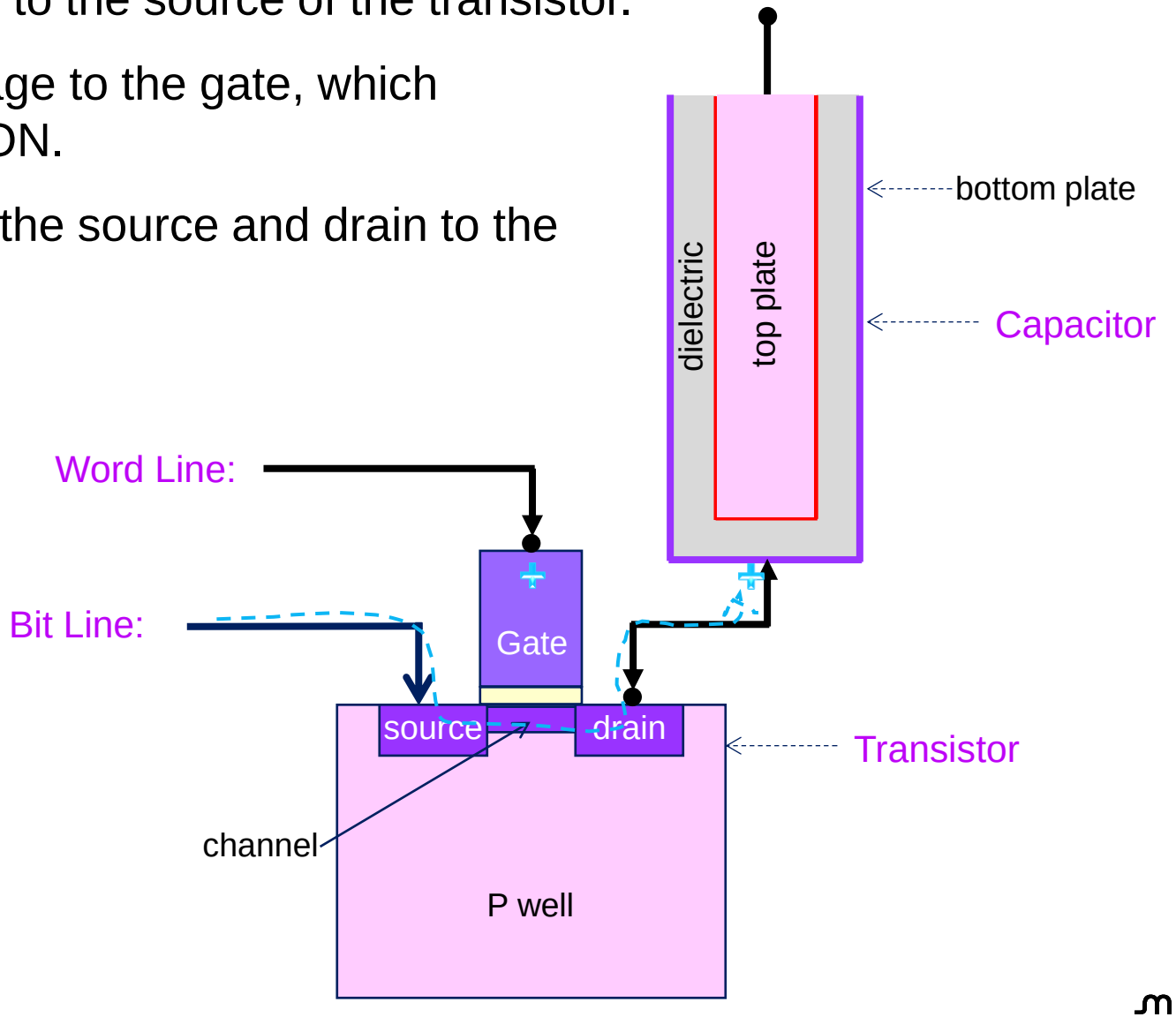
1. Bit Line delivers positive voltage to the source of the transistor.

2. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.



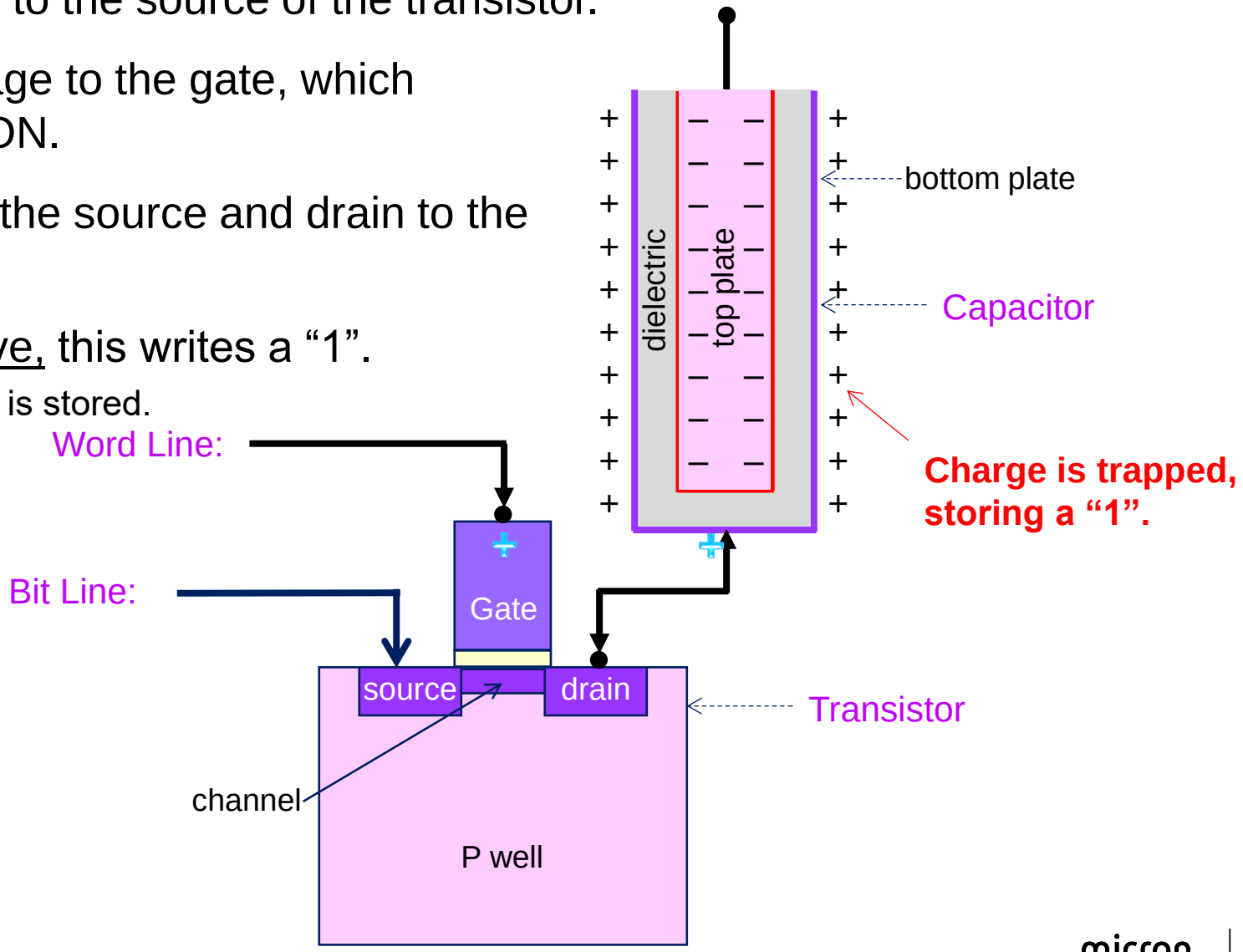
DRAM Operation – Write

1. Bit Line delivers positive voltage to the source of the transistor.
2. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.
3. Bit Line voltage passes through the source and drain to the capacitor.



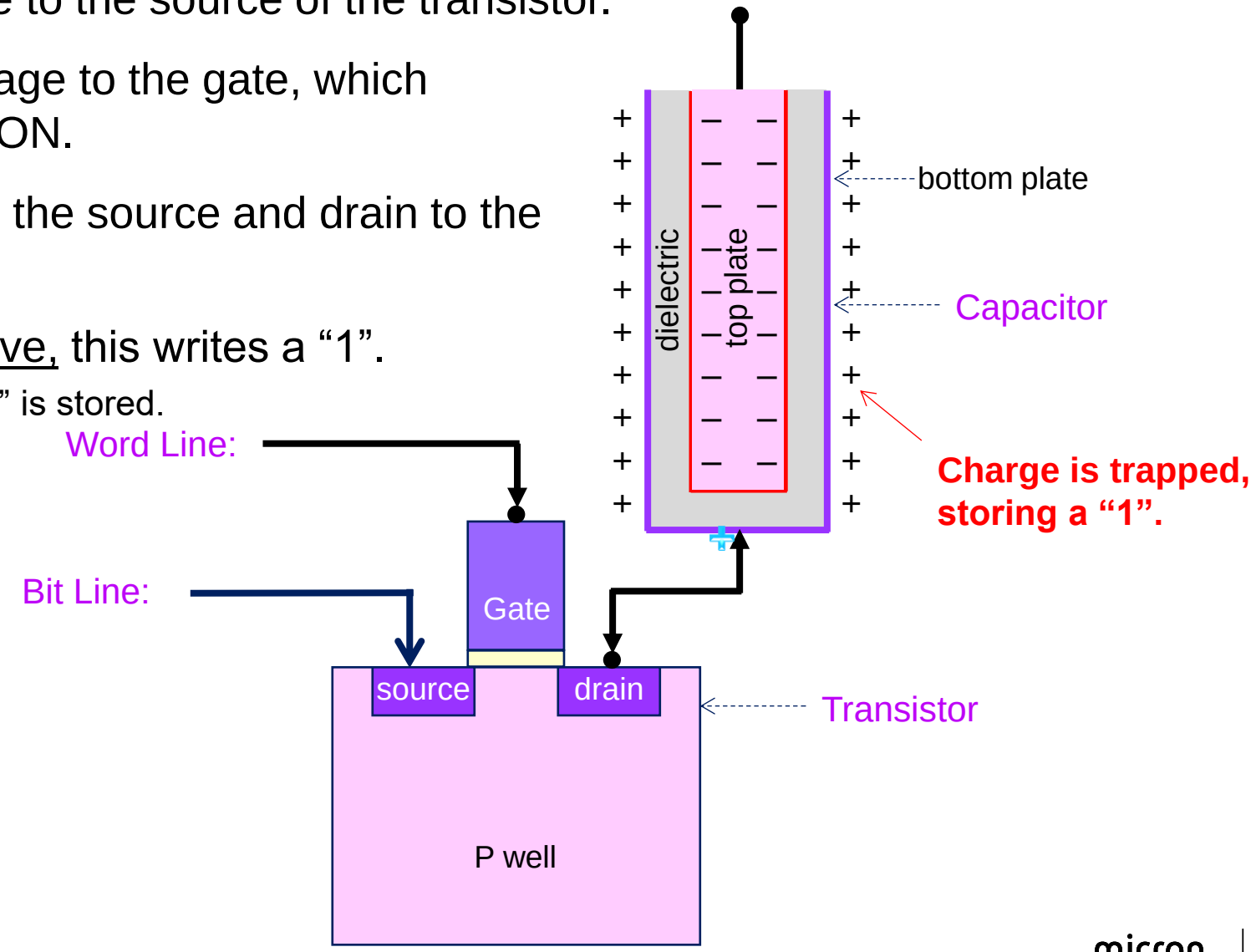
DRAM Operation – Write

1. Bit Line delivers positive voltage to the source of the transistor.
2. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.
3. Bit Line voltage passes through the source and drain to the capacitor.
4. The bottom plate charges positive, this writes a “1”.
   - If bottom plate charges negative, a “0” is stored.



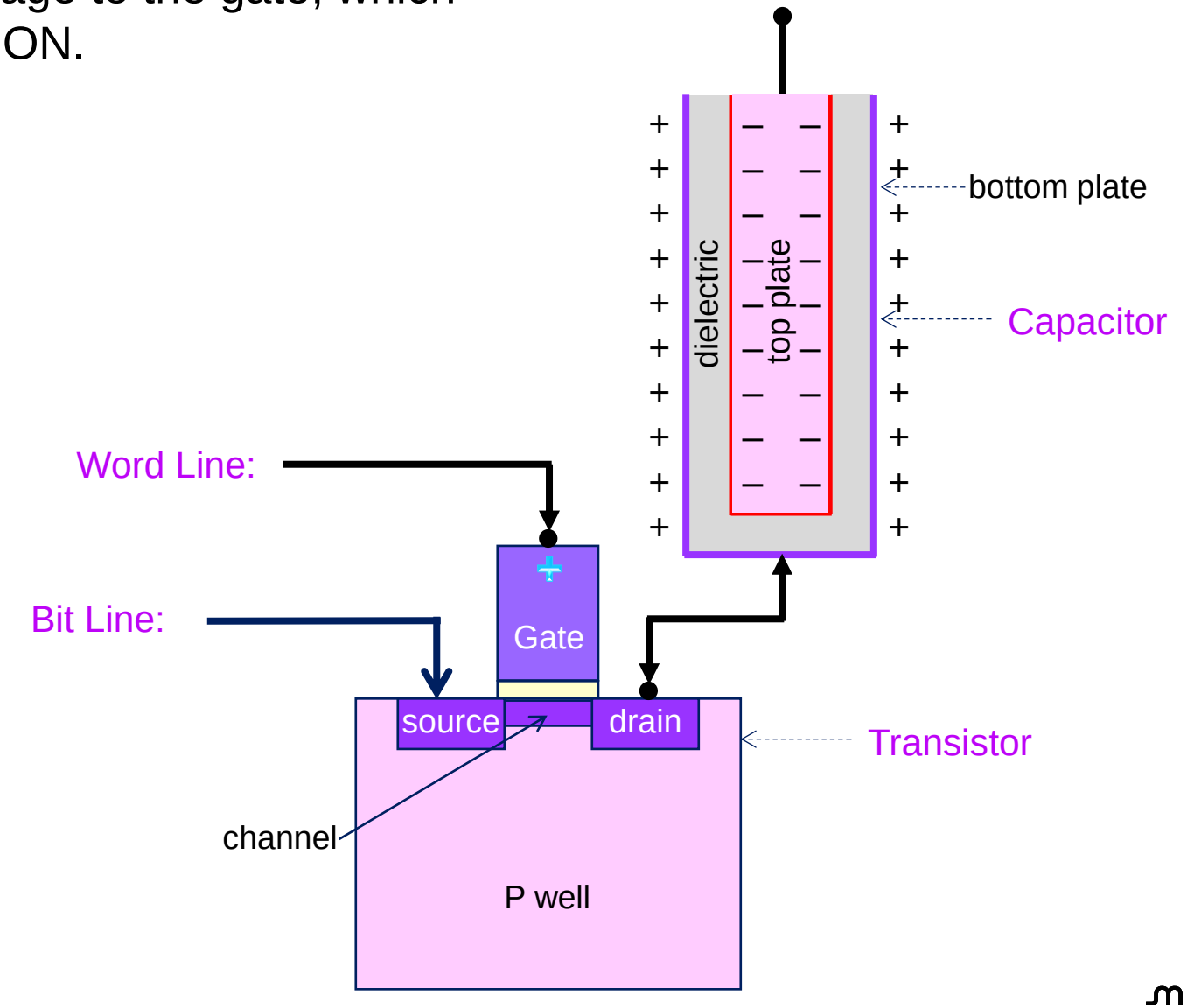
DRAM Operation – Write

1. Bit Line delivers positive voltage to the source of the transistor.
2. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.
3. Bit Line voltage passes through the source and drain to the capacitor.
4. The bottom plate charges positive, this writes a “1”.
   - If bottom plate charges negative, a “0” is stored.
5. Word Line turns off transistor, “trapping” the charge on the capacitor.



DRAM Operation – Read

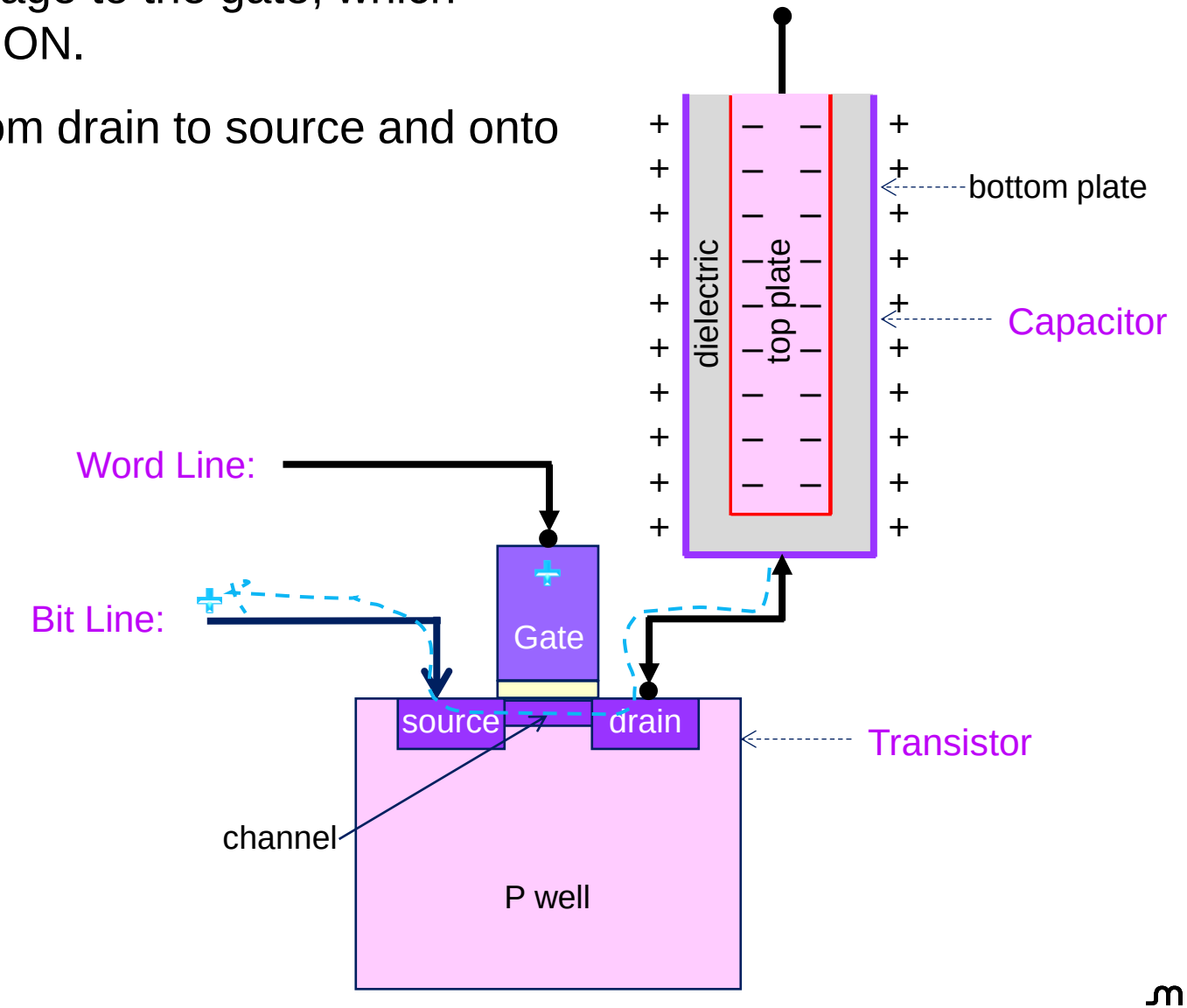
1. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.



DRAM Operation – Read

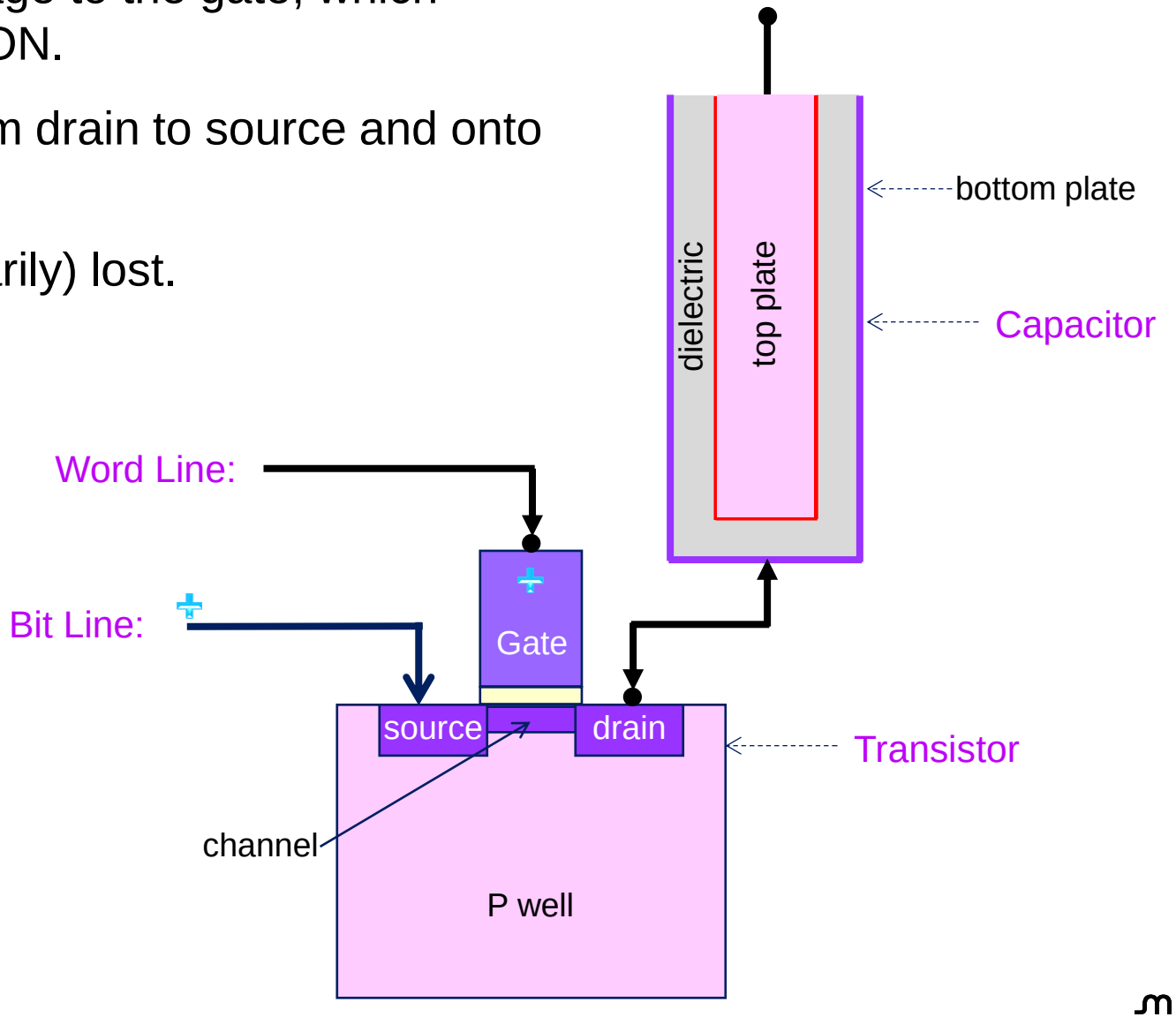
1. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.

2. Charge on capacitor passes from drain to source and onto the Bit Line.



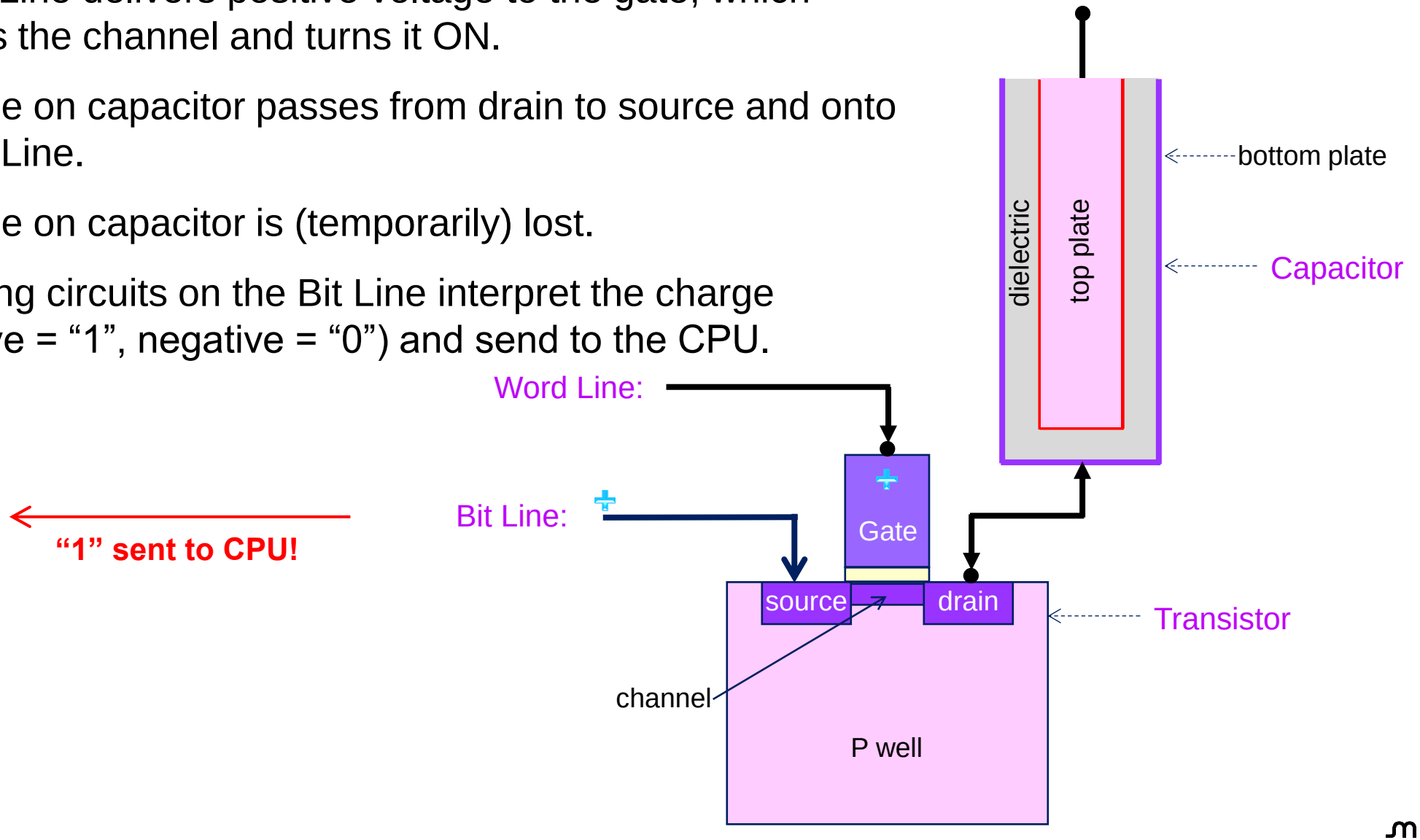
DRAM Operation – Read

1. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.
2. Charge on capacitor passes from drain to source and onto the Bit Line.
3. Charge on capacitor is (temporarily) lost.



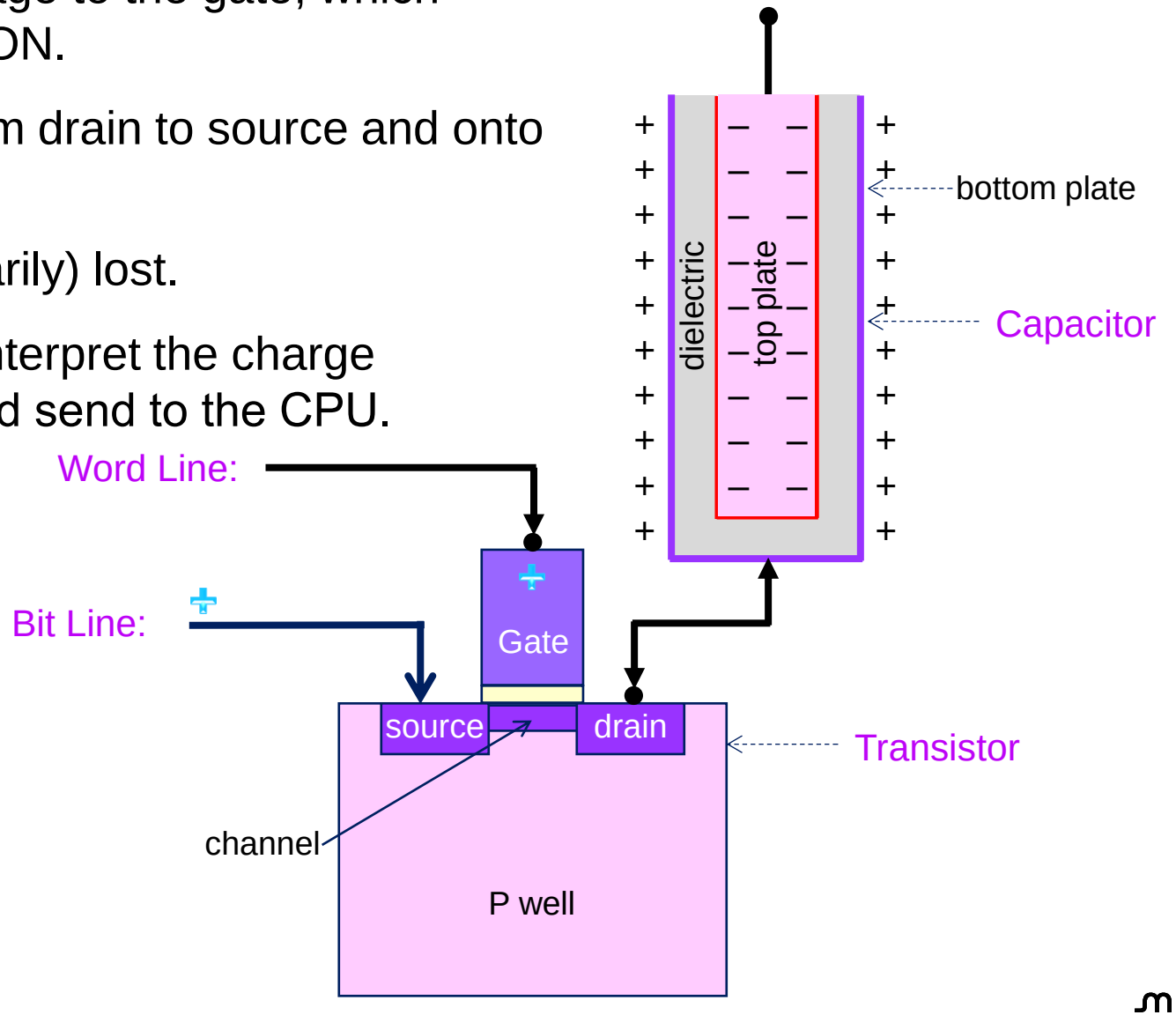
DRAM Operation – Read

1. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.
2. Charge on capacitor passes from drain to source and onto the Bit Line.
3. Charge on capacitor is (temporarily) lost.
4. Sensing circuits on the Bit Line interpret the charge (positive = “1”, negative = “0”) and send to the CPU.

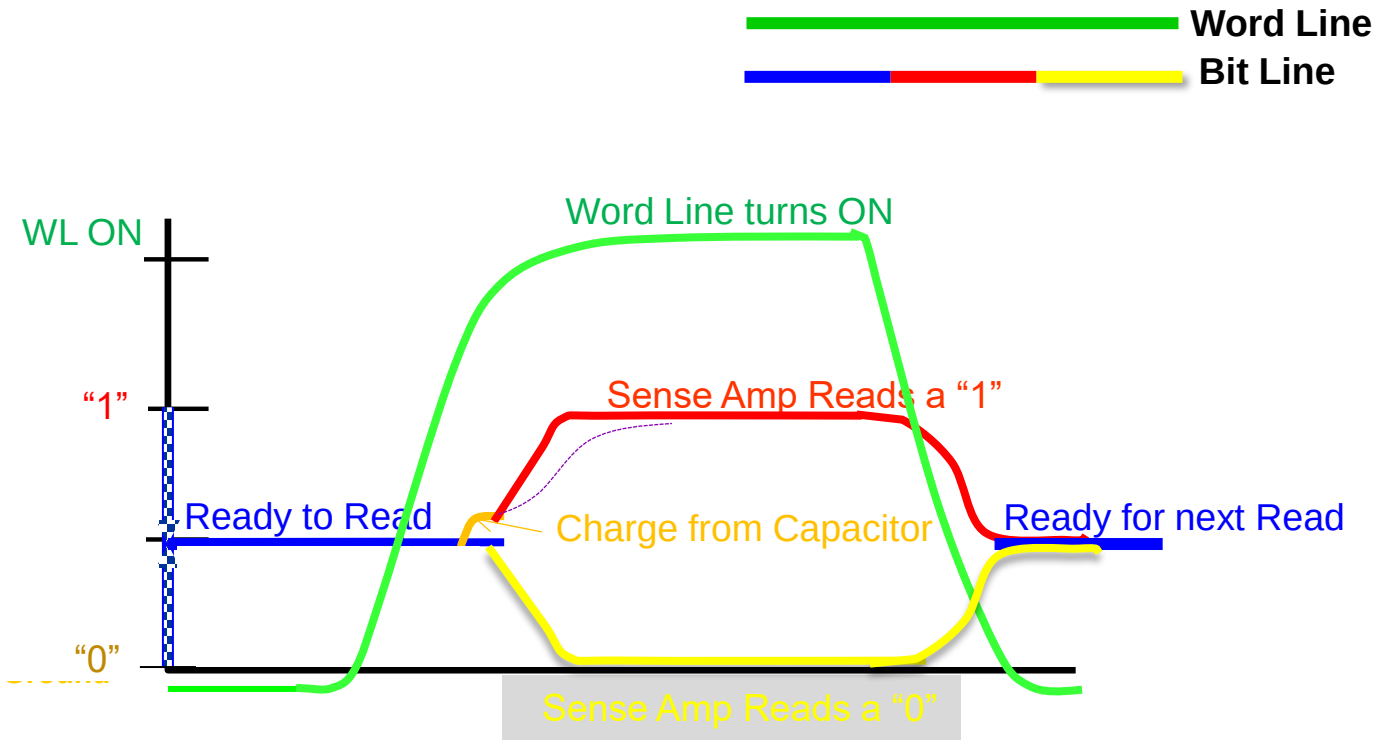


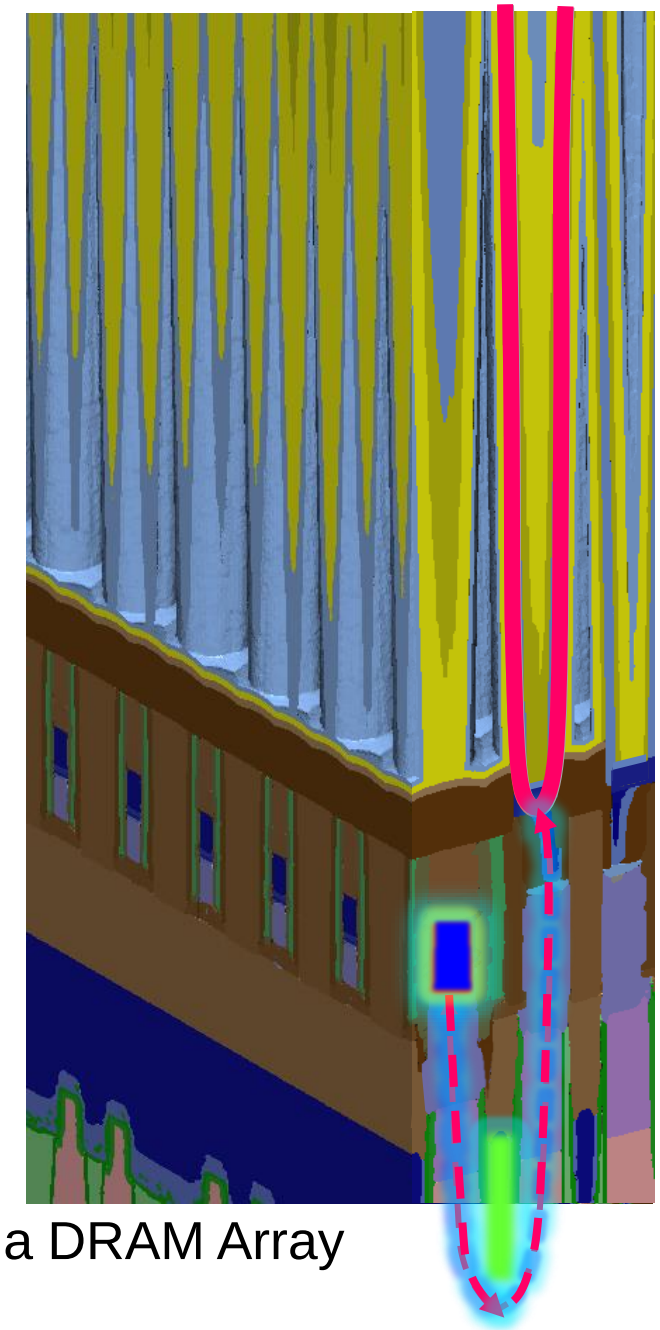
DRAM Operation – Read

1. Word Line delivers positive voltage to the gate, which creates the channel and turns it ON.
2. Charge on capacitor passes from drain to source and onto the Bit Line.
3. Charge on capacitor is (temporarily) lost.
4. Sensing circuits on the bit line interpret the charge (positive = “1”, negative = “0”) and send to the CPU.
5. Lost charge on the capacitor must be “refreshed” (same as write operation).

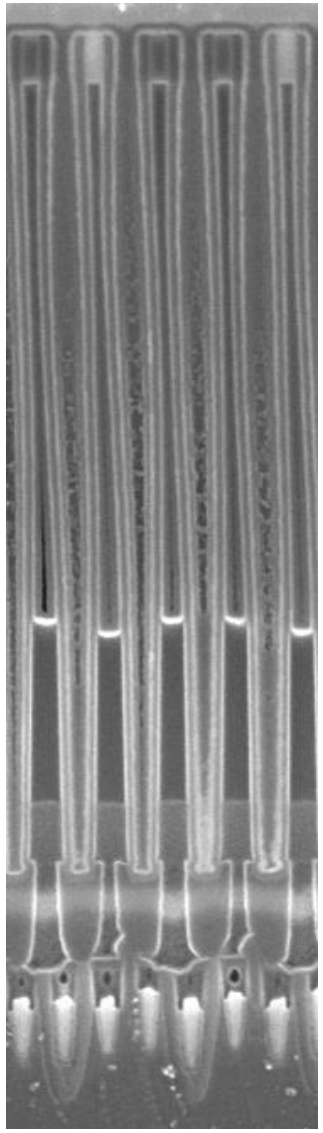


Read (& Refresh)



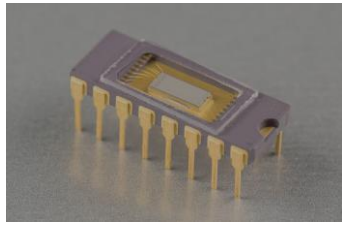


3D Model of a DRAM Array

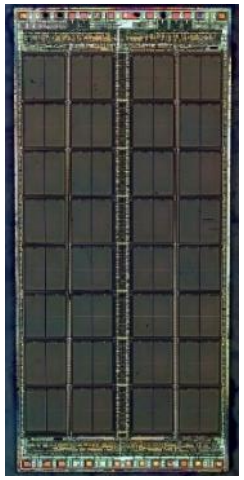


Container depth ~1um
human hair ~70um

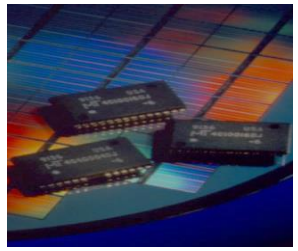
Evolution of DRAM



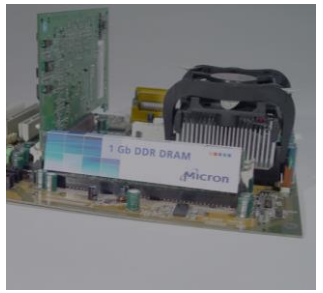
1981
64K DRAM



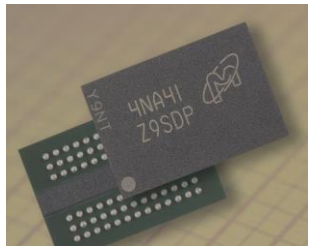
1987
1Mb DRAM



1992
16Mb DRAM



2002
1Gb DRAM



2014
8Gb DRAM

2020
16Gb DRAM

2022
Up to 128Gb

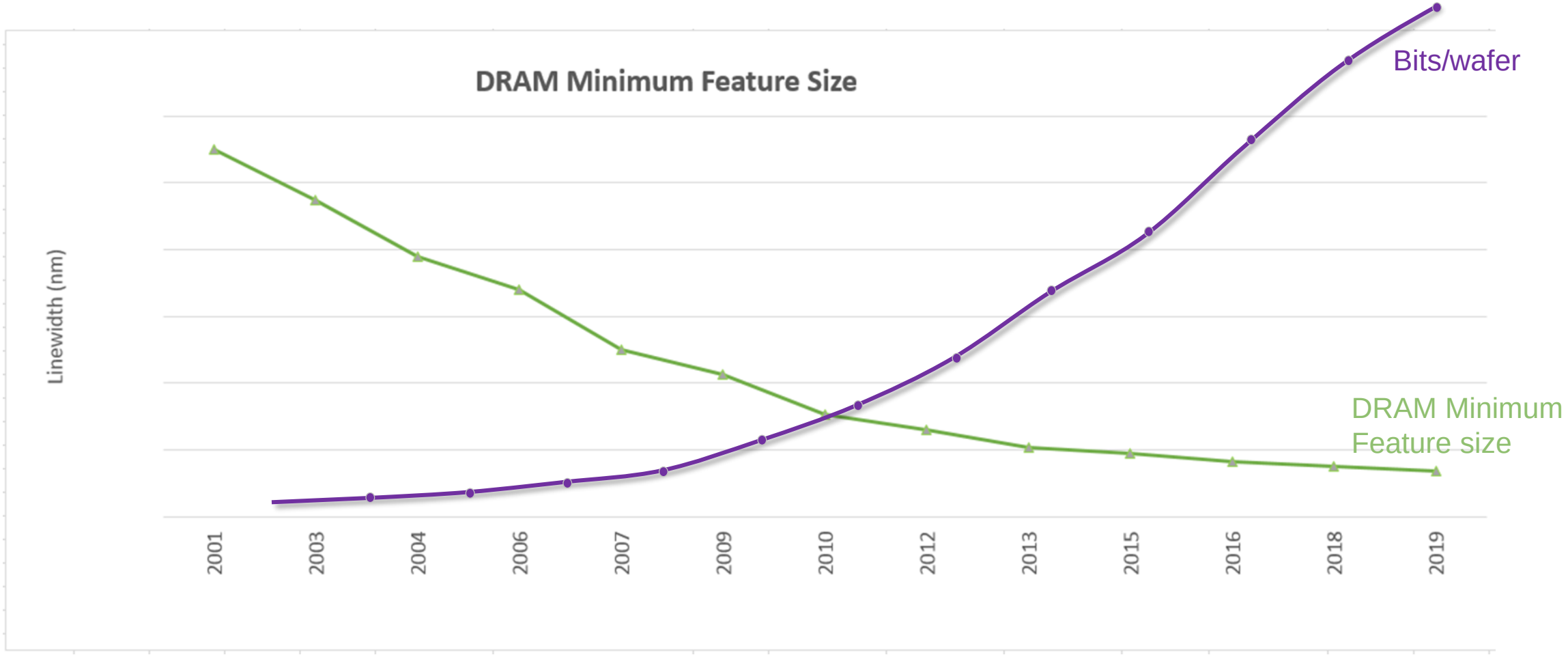
41 years
2,000,000x density increase

1α
Micron's World's Most Advanced DRAM Process Technology (in 2020)

1β
Micron's Next World's Most Advanced DRAM Process Technology (in 2022)

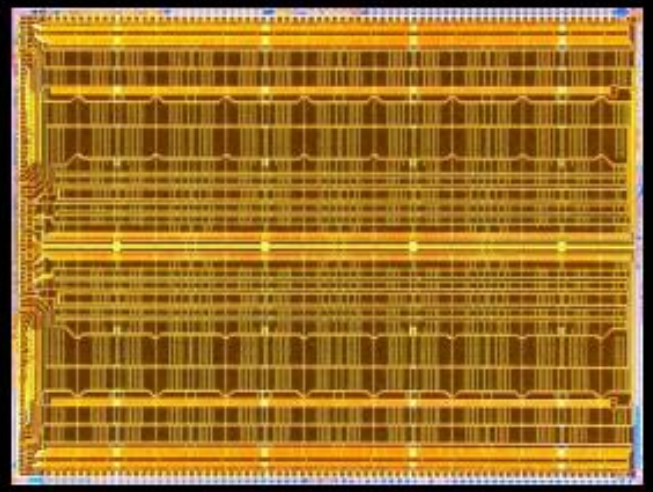
DRAM Scaling Trend

With each generation, by reducing the feature sizes of our memory cells on each die we can increase density, achieve higher bits per wafer, and lower cost per bit.



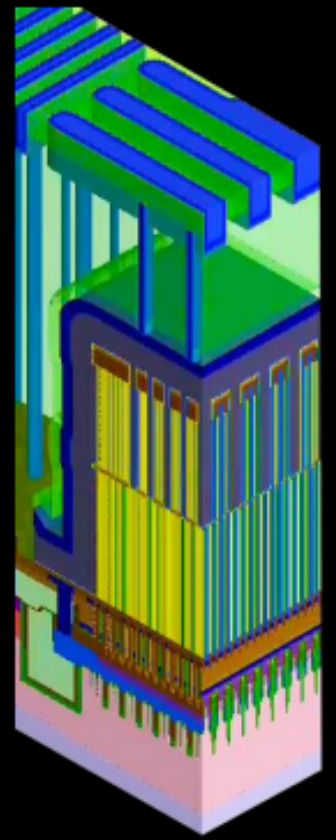
1β: The world's most advanced DRAM

Industry leadership sustained – starting ramp in late CY22



16Gb LPDDR5

- Advanced pattern multiplication lithography
- Extend DDR5 and LPDDR5 leadership
- Delivers graphics, HBM3 and automotive excellence
- Leadership in power, performance and cost



1β DRAM

1β = name of a Micron DRAM technology
CY = Calendar Year
DRAM = Dynamic Random Access Memory
DDR = Double Data Rate
Gb = Gigabit
HBM = High Bandwidth Memory
LPDDR = Low Power Double Data Rate

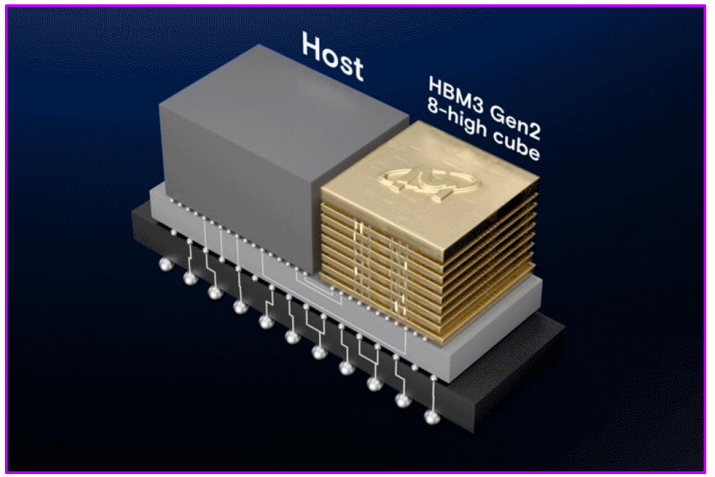
July 26, 2023 at 9:02 AM EDT

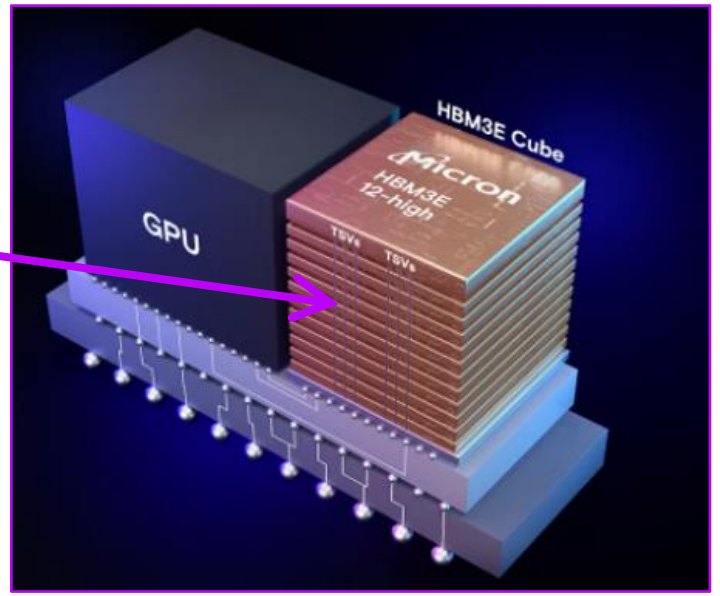
Micron Delivers Industry's Fastest, Highest-Capacity HBM to Advance Generative AI Innovation

First in industry to launch 8-high 24GB HBM3 Gen2 with bandwidth over 1.2TB/s and superior power efficiency enabled by advanced 1β process node

BOISE, Idaho, July 26, 2023 (GLOBE NEWSWIRE) -- Micron Technology, Inc. (Nasdaq: MU) today announced it has begun sampling the industry's first 8-high 24GB HBM3 Gen2 memory with bandwidth greater than 1.2TB/s and pin speed over 9.2Gb/s, which is up to a 50% improvement over currently shipping HBM3 solutions. With a 2.5 times performance per watt improvement over previous generations, Micron's HBM3 Gen2 offering sets new records for the critical artificial intelligence (AI) data center metrics of performance, capacity and power efficiency. These Micron improvements reduce training times of large language models like GPT-4 and beyond, deliver efficient infrastructure use for AI inference and provide superior total cost of ownership (TCO).

The foundation of Micron's high-bandwidth memory (HBM) solution is Micron's industry-leading 1β (1-beta) DRAM process node, which allows a 24Gb DRAM die to be assembled into an 8-high cube within an industry-standard package dimension. Moreover, Micron's 12-high stack with 36GB capacity will begin sampling in the first quarter of calendar 2024. Micron provides 50% more capacity for a given stack height compared to existing competitive solutions. Micron's HBM3 Gen2 performance-to-power ratio and pin speed improvements are critical for managing the extreme power demands of today's AI data centers. The improved power efficiency is possible because of Micron advancements such as doubling of the through-silicon vias (TSVs) over competitive HBM3 offerings, thermal impedance reduction through a five-time increase in metal density, and an energy-efficient data path design.





This 24GB HBM3 Gen2 contains 8 of the leading-edge 1Beta 24Gbit DRAM die! Those HBM3-specific die are interconnected vertically by Thru-Silicon Vias (TSVs) that are formed in the wafer during in-fab processing

From Micron News Release 26Jul23 LINK

5. Introduction to Flash

micron

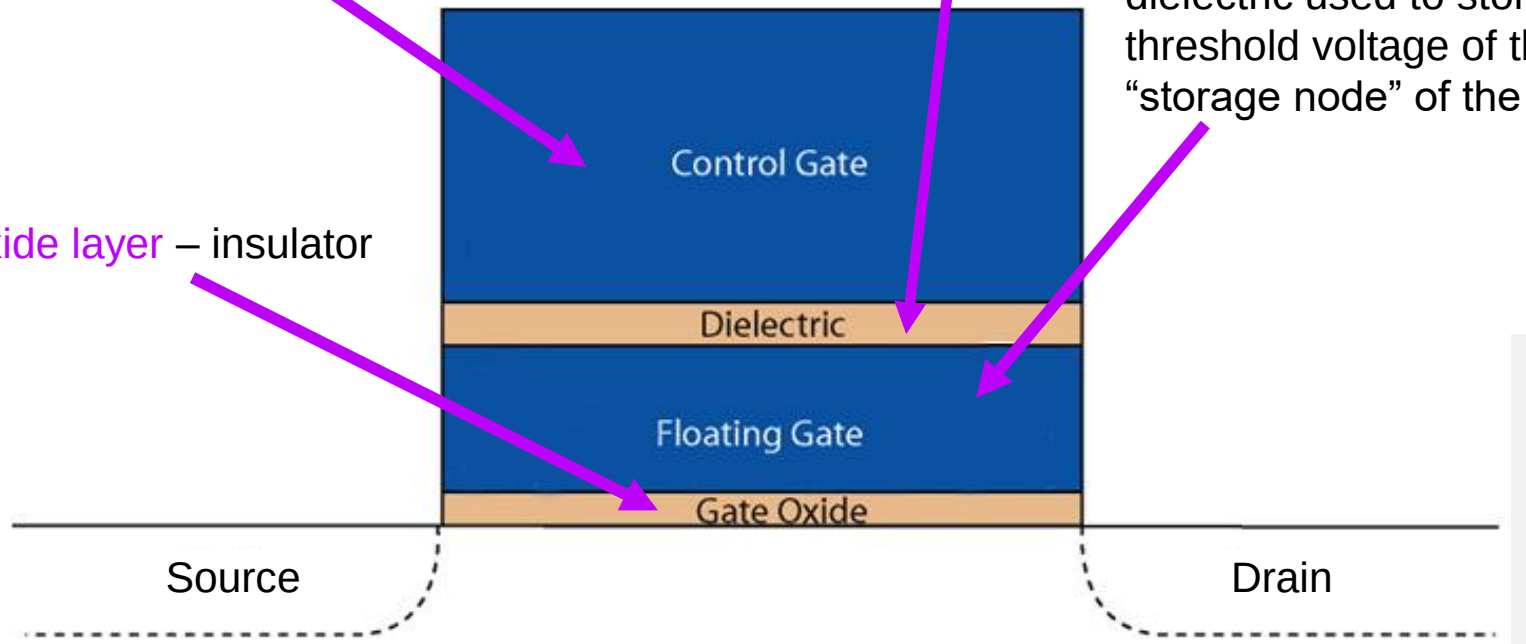
Structure of a Basic Flash Cell

Dielectric – insulator

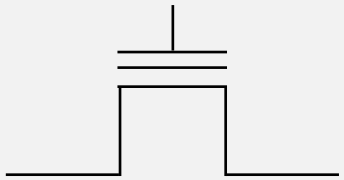
Control Gate – where a controlling voltage is applied

Floating Gate – conducting or semi-conducting layer that is surrounded by a dielectric used to store charge and alter the threshold voltage of the device. This is the "storage node" of the NAND cell.

Gate (Tunnel) oxide layer – insulator

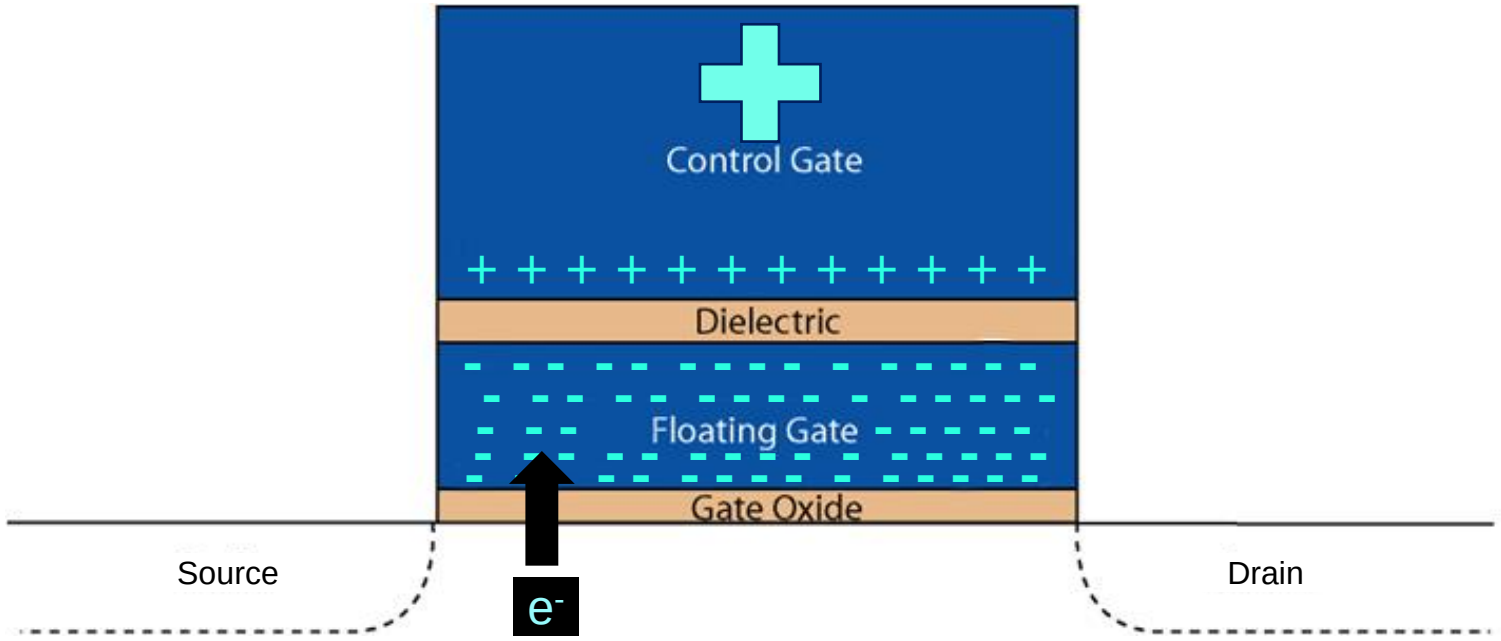


Symbol of the dual gate transistor



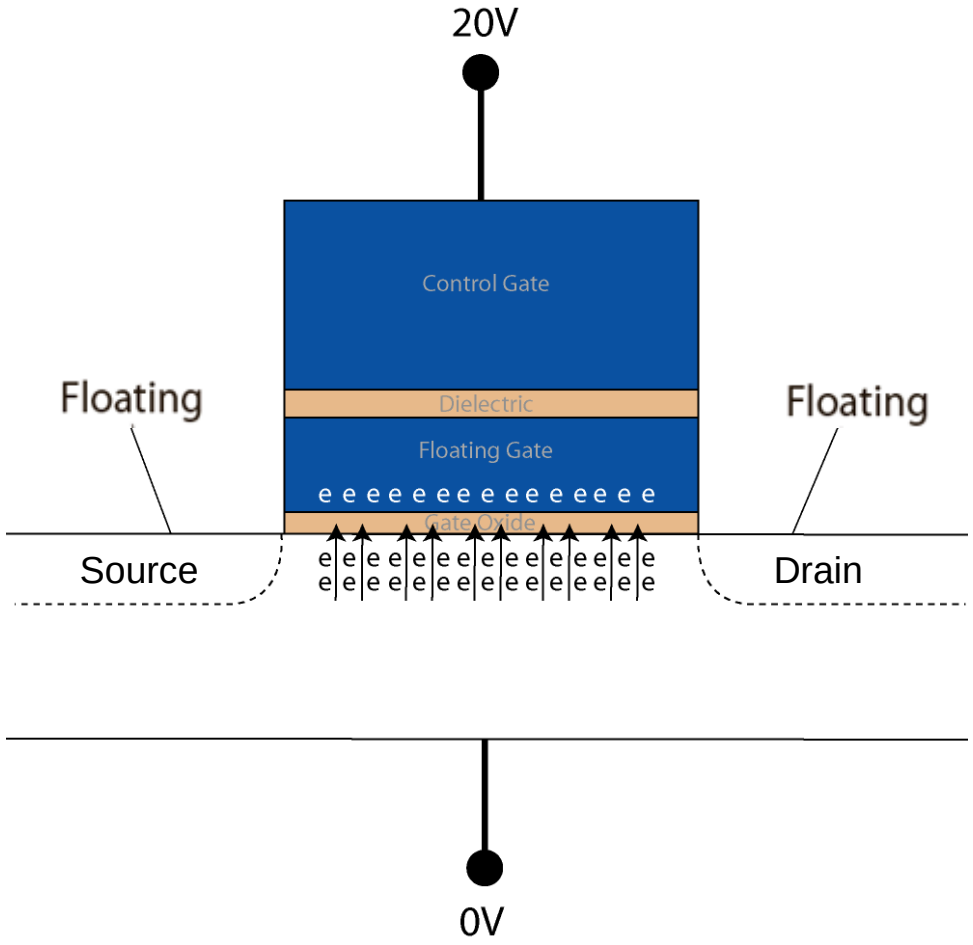
NAND Function

- High voltage on the Control Gate causes electrons to tunnel through the gate oxide and get trapped within the Floating Gate



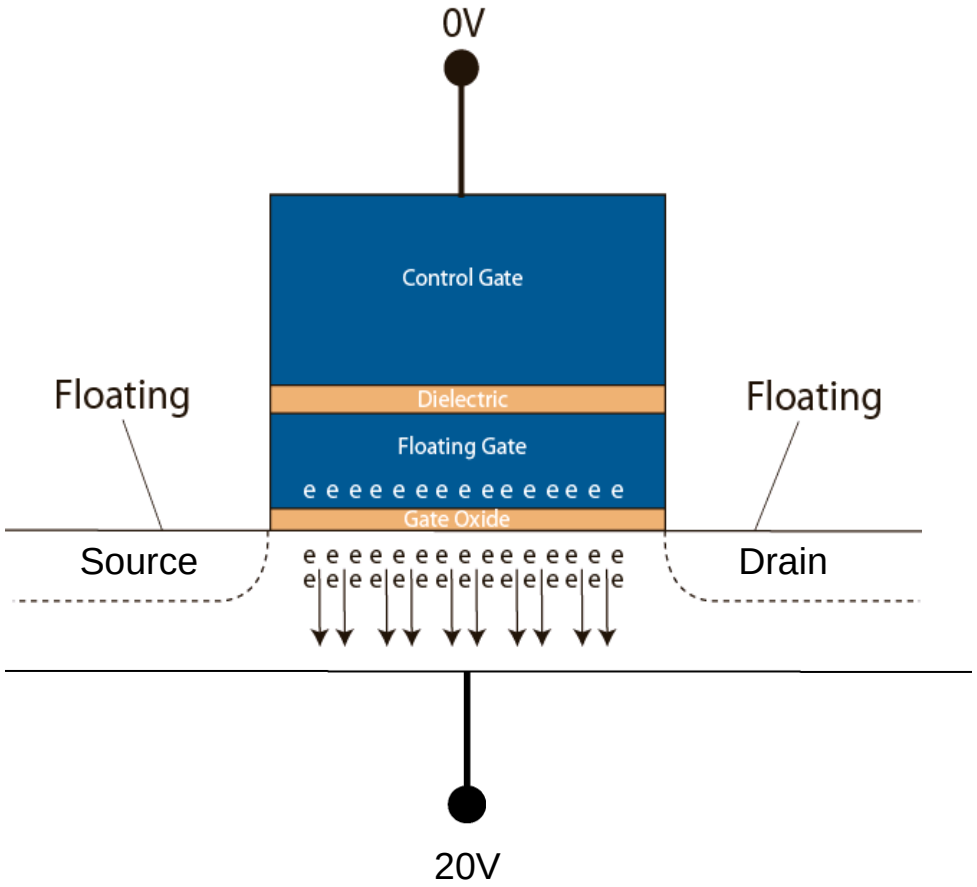
Programming / Writing a NAND Cell

- Ground the substrate
- Apply ~+20V to the control gate
- Electrons in the substrate tunnel through the gate oxide and relocate in the floating gate
- When voltage is released, electrons are trapped in the floating gate making NAND Flash a non-volatile memory
- For some of our parts, data is guaranteed for 10 years



Erasing a NAND Cell

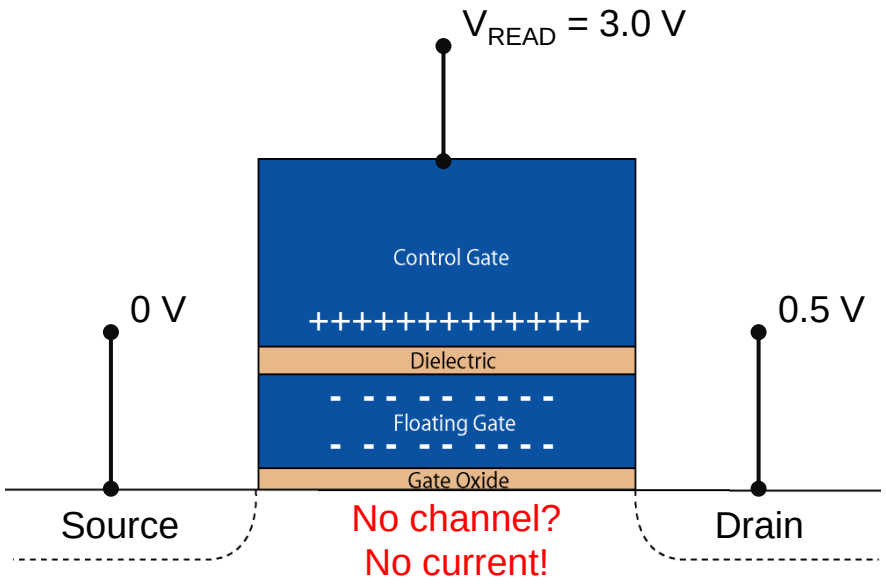
- Ground the control gate
- Apply ~+20V to the substrate
- Electrons trapped within the floating gate tunnel back through the gate oxide into the substrate



Reading a NAND Cell

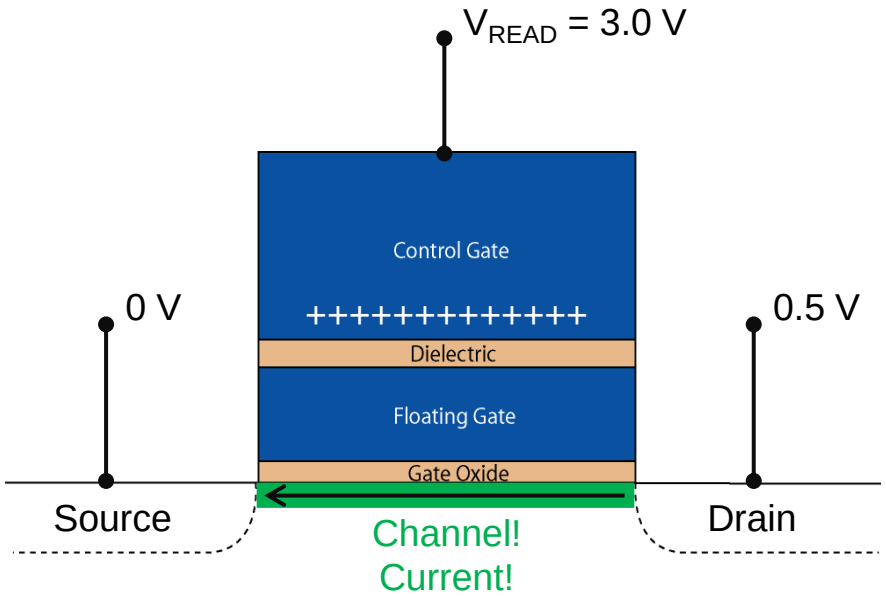
Reading a Programmed Cell:

- 0.5V on drain, 0V on source create a voltage drop
- V_{READ} = 3.0V on the control gate attempts to create a channel to turn transistor ON
- Electrons trapped in the floating gate prevent the channel from forming so transistor remains OFF (V_{READ} not sufficient to form channel)
- No current flows between source and drain so a logic “0” is read



Reading an Erased Cell:

- 0.5V on drain, 0V on source create a voltage drop
- V_{READ} = 3.0V on the gate attempts to create a channel to turn transistor ON
- With no electrons trapped in the floating gate, the channel is formed, and transistor turns ON
- Current flows between source and drain so a logic “1” is read



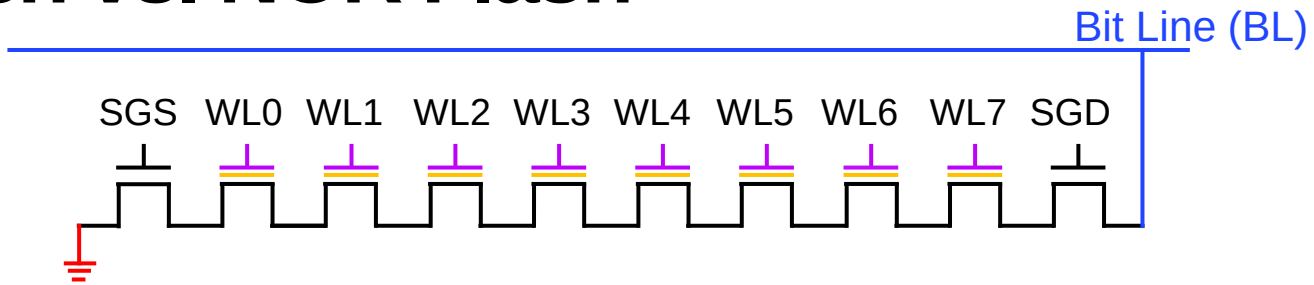
Flash Function: Logic States

# of e^- Trapped in Floating Gate	Cell Threshold Voltage	State of the Cell	Current in Channel	Logic State
Low	Low	Not Programmed	Yes	1
High	High	Programmed	No	0

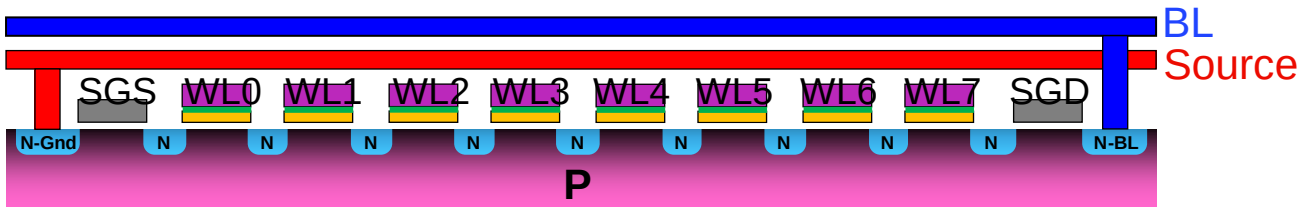
NAND Flash vs. NOR Flash

NAND

- Higher Density but Slower

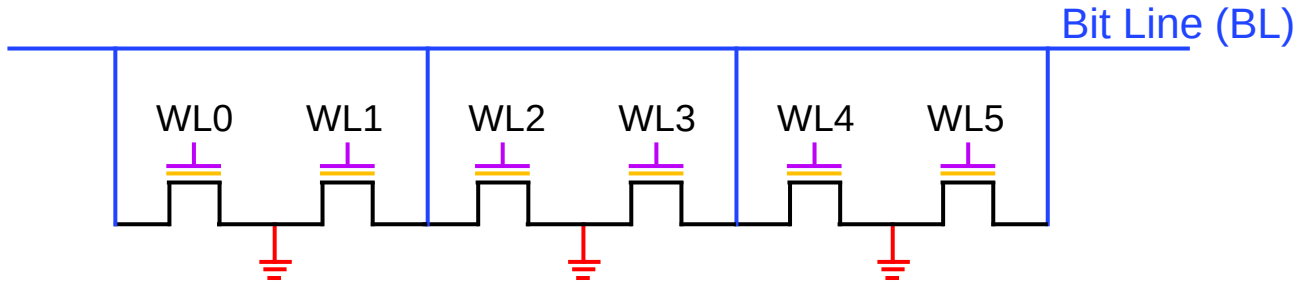


Shared Select Gate Source (SGS) and Bit Line connections for a string of NAND Flash Memory Cells

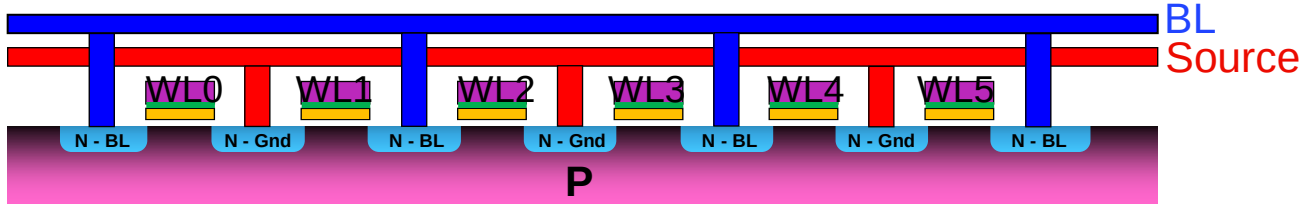


NOR

- Lower Density but Faster



Individual Bit Line connections for each NOR Flash Memory Cell



Key
BL: Bit Line
SGS: Select Gate Source Transistor
SGD: Select Gate Drain Transistor
WL: Word Line

Multi Level Cell (MLC) Technology

- So far we have seen a NAND cell that can store 1 bit per cell where the cell is in one of two possible states:
  - Programmed (Logic 0) or cup is full
  - Erased (Logic 1) or cup is empty
- Continuous innovations in NAND technology now allow us to store more than 1 bit per cell.
- Micron has NAND parts that can store 2, 3, or 4 bits in each cell.
  - SLC: 1 bit per cell (Single level Cell)

Erased
Logic 1

Programmed
Logic 0

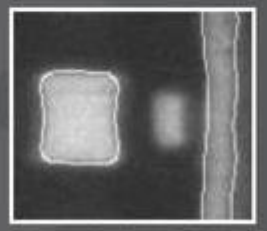
Multi Level Cell (MLC) Technology

- So far we have seen a NAND cell that can store 1 bit per cell where the cell is in one of two possible states:
  - Programmed (Logic 0) or cup is full
  - Erased (Logic 1) or cup is empty
- Continuous innovations in NAND technology now allow us to store more than 1 bit per cell.
- Micron has NAND parts that can store 2, 3, or 4 bits in each cell.
  - SLC: 1 bit per cell (Single level Cell)
  - MLC: 2 bits per cell (Multi Level Cell)
  - TLC: 3 bits per cell (Triple Level Cell)
  - QLC: 4 bits per cell (Quad Level Cell)
- These new cells are allowed to have "partially full" states
- Consider this MLC example ➔

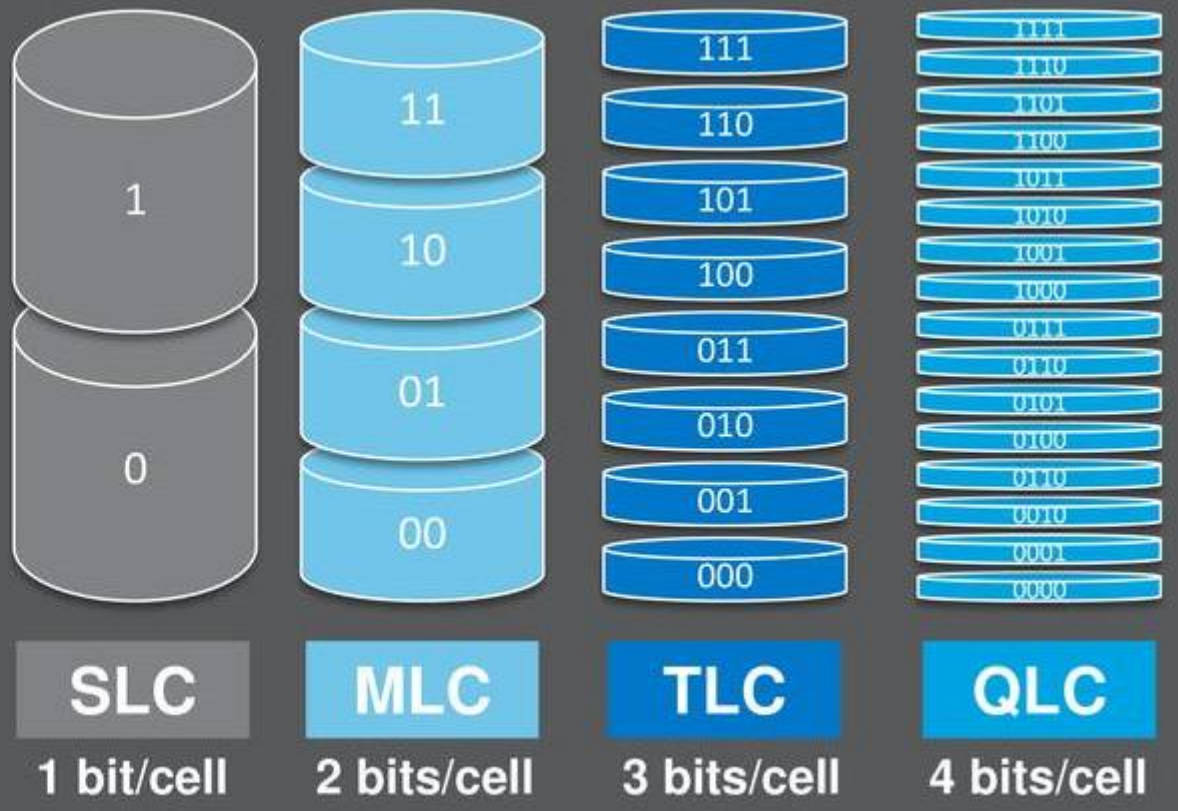
Erased	Programmed	Programmed	Programmed
Level 0	Level 1	Level 2	Level 3

NAND: Leading the Industry in 4 Bits per Cell (QLC)

- First 64 Layer QLC (4 bits/cell) SSD solution
- Industry's first 1Tb die
- Achieves 33% density increase over TLC



Single NAND Cell



MLC = Multi Level Cell (2 bits per cell)
QLC = Quad Level Cell (4 bits per cell)
SLC = Single Level Cell (1 bit per cell)
SSD = Solid State Drive
Tb = Terabit (one trillion bits or 1,000 gigabits)
TLC = Triple Level Cell (3 bits per cell)

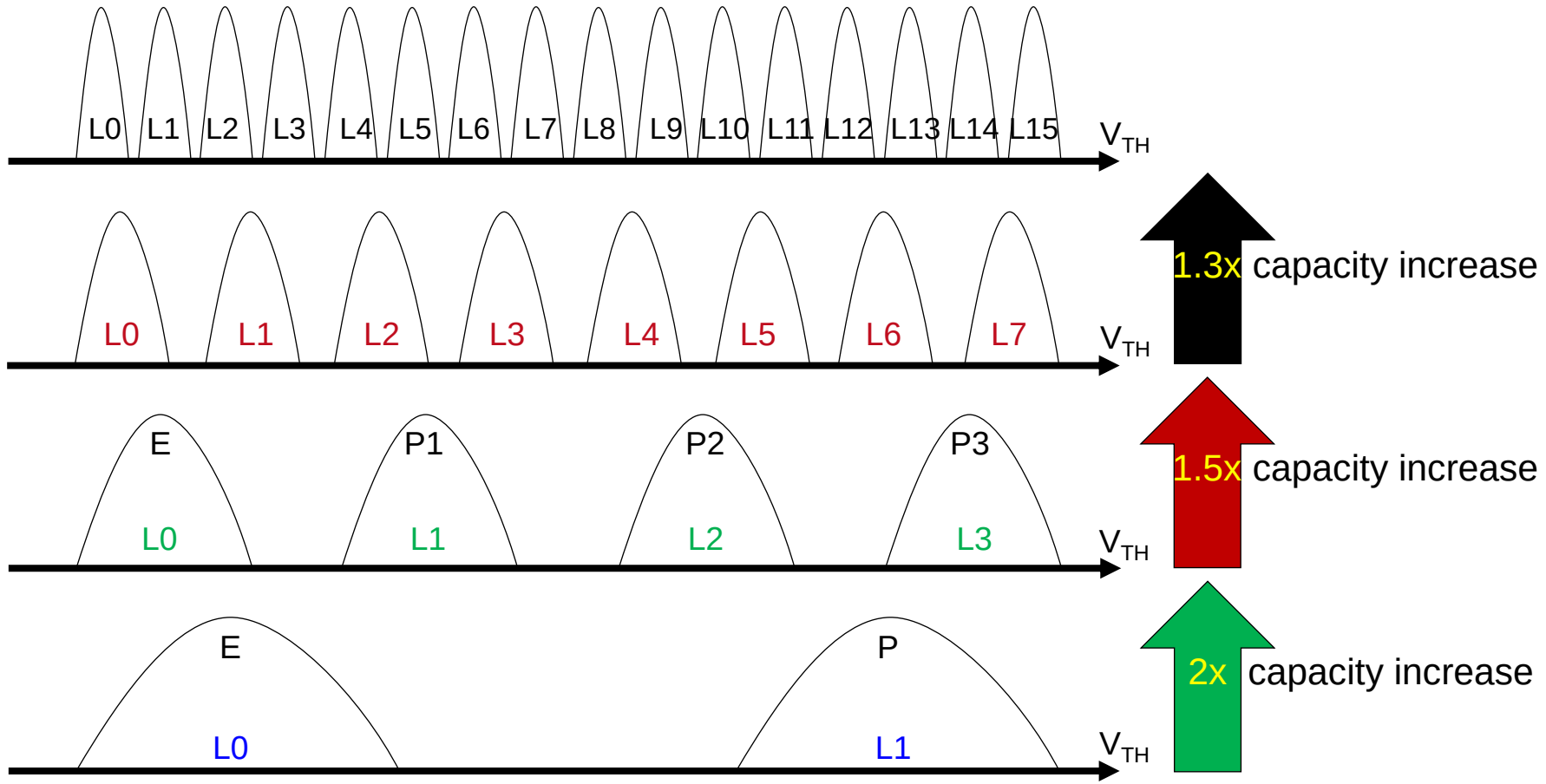
Charge States in NAND SLC/MLC/TLC/QLC Memory

QLC - Quad Level Cell
16 Charge States:
1 Erased + 15 Programmed
4 Bits per cell

TLC - Triple Level Cell
8 Charge States:
1 Erased + 7 Programmed
3 Bits per cell

MLC - Multi-Level Cell
4 Charge States:
1 Erased + 3 Programmed
2 Bits per cell

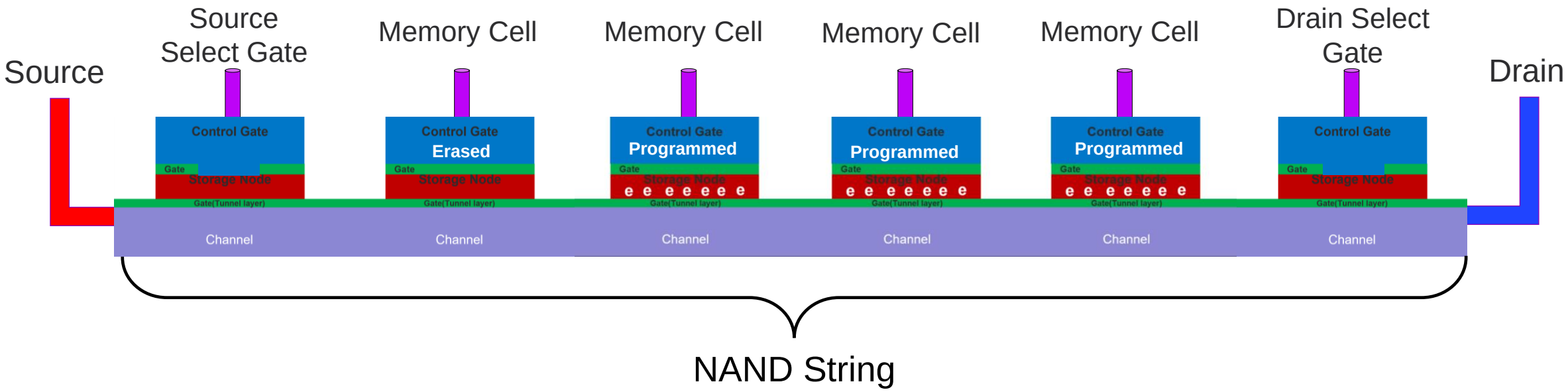
SLC - Single Level Cell
2 Charge States:
1 Erased + 1 Programmed
1 Bit per cell



Graphic is in arbitrary x/y scale

2D NAND of Years Past

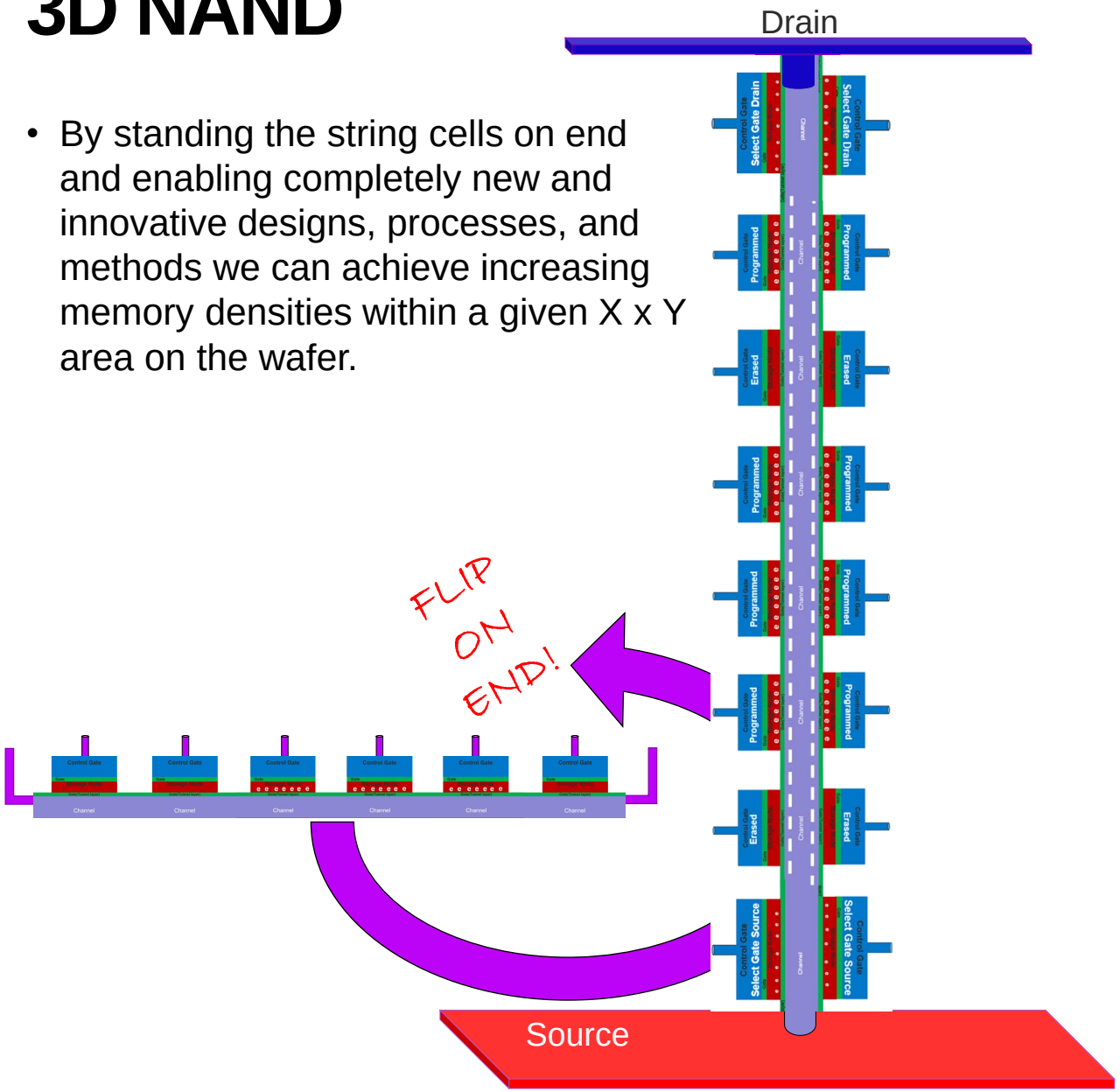
- Early NAND generations arranged the cells horizontally

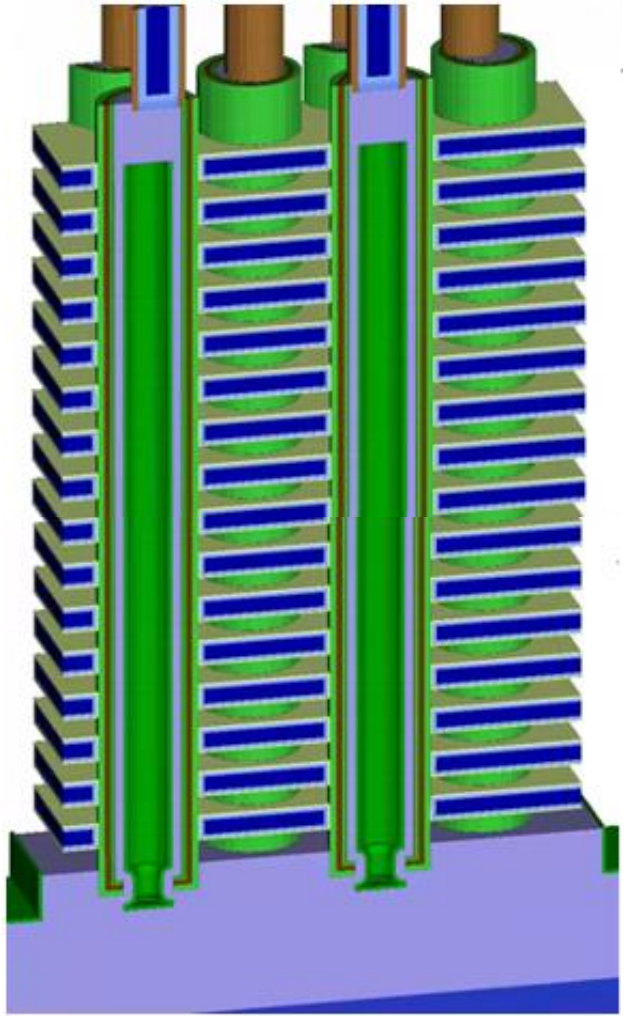


Into the Z Dimension!

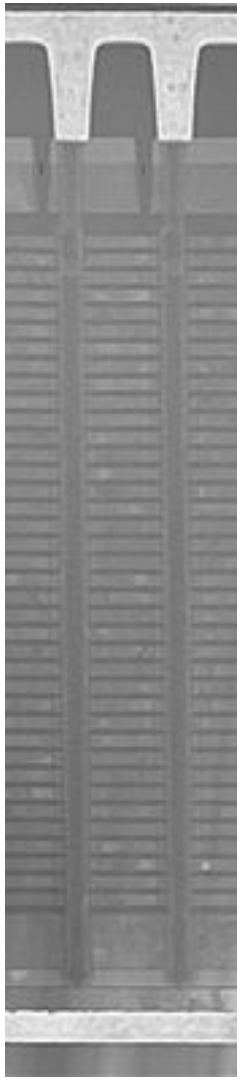
3D NAND

- By standing the string cells on end and enabling completely new and innovative designs, processes, and methods we can achieve increasing memory densities within a given X x Y area on the wafer.





3D Model of a 3D NAND Array

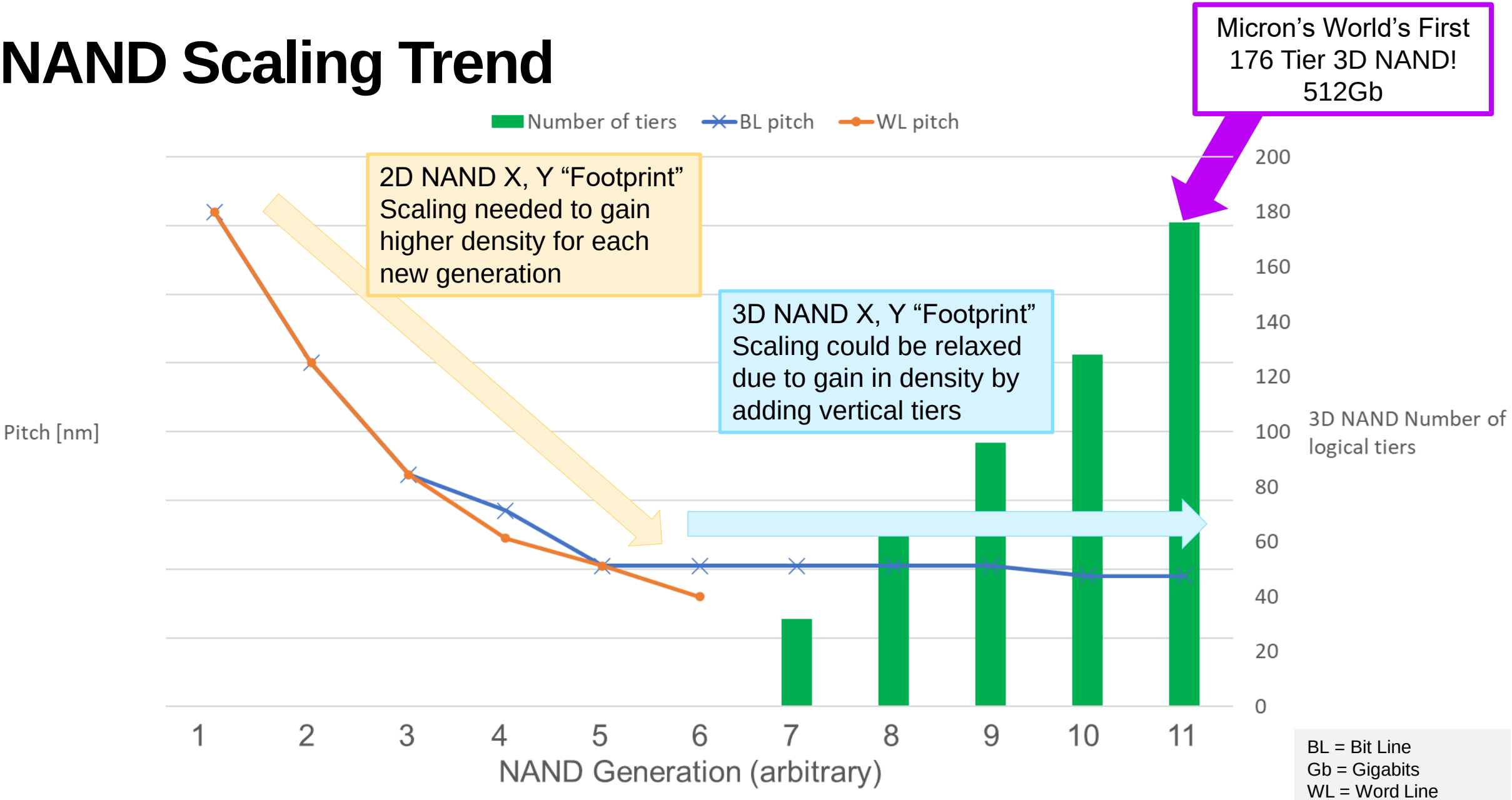


Only ~38 tiers shown in this SEM

Stacked tiers ~10um
human hair ~70um

SEM = Scanning Electro Microscope

NAND Scaling Trend

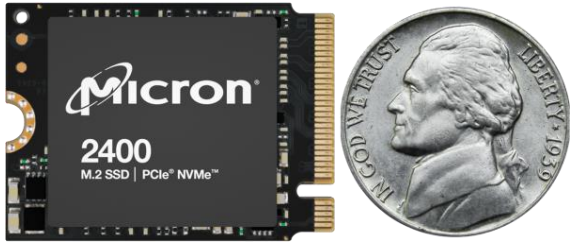


BL = Bit Line
Gb = Gigabits
WL = Word Line

January 11, 2022

Micron Ships the Industry's First 176-Layer QLC NAND in Volume and Unveils the 2400 PCIe Gen4 Client SSD

Groundbreaking technology enables world's first 2TB 22x30mm SSD optimized for client applications



BOISE, Idaho, Jan. 11, 2022 (GLOBE NEWSWIRE) -- Micron Technology, Inc. (Nasdaq: MU), today announced it has begun volume shipments of the world's first 176-layer QLC NAND SSD. Built with the most advanced NAND architecture, Micron's 176-layer QLC NAND delivers the industry's leading storage density and optimized performance for a broad range of data-rich applications. Designed for use cases spanning client and data center environments, Micron's transformative new NAND technology is now available with the introduction of the Micron 2400 SSD, the world's first 176-layer PCIe Gen4 QLC SSD for client applications. The new 176-layer QLC NAND will also be incorporated into select Micron Crucial consumer SSDs, and available as a component for system designers.

Micron's groundbreaking 176-layer QLC NAND provides a layer count and density unprecedented in QLC NAND flash and follows Micron's delivery of the industry's first 176-layer TLC NAND. Additionally, Micron's 176-layer QLC NAND enables 33% higher I/O speed[1] and 24% lower read latency[2] than Micron's prior generation solution. Its replacement-gate architecture is the only mass production QLC flash storage that combines charge trap with a CMOS-under-array design. These improvements are driving adoption of QLC SSDs in the client PC market, which is expected to triple QLC adoption by 2023, exceeding 35%, and reaching nearly 80% bit share in 2025.[3]

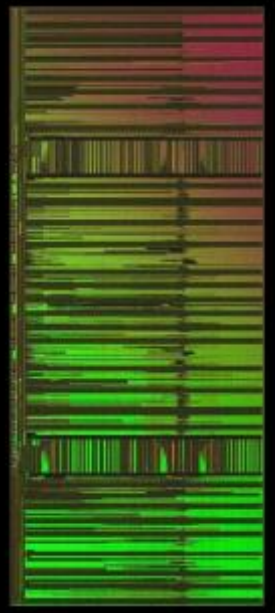
This 2TB SSD contains 16 of the 176-layer 1Tb die!

From Micron News R

I/O speed = Input/Output speed
PCIe = Peripheral Component Interconnect Express
QLC = Quad Level Cell (4 bits per cell)
SSD = Solid State Drive
Tb = Terabit
TB = Terabyte

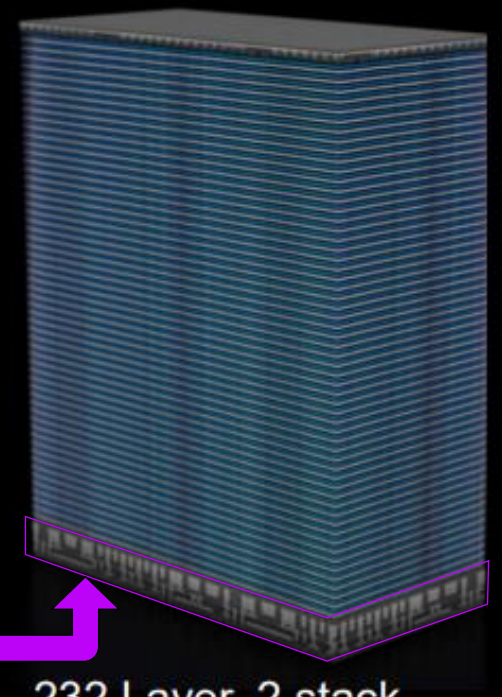
232-layer: The world's most advanced NAND

Industry leadership sustained – starting ramp in late CY22



1Tb TLC NAND

- Extending CuA and 2 array stack process architecture
- Optimized for leadership in managed NAND and SSD
- Combination of external and optimized internal controllers
- Increased density, power and bandwidth node-over-node



232 Layer, 2 stack, CuA* NAND

*CuA – CMOS Under Array - another Micron innovation – by locating all support circuitry underneath the memory cell tiers, the die size can be further minimized, thereby gaining even higher bit density (like all supporting utilities and parking garage installed beneath a skyscraper!)

CuA = CMOS Under Array
CY = Calendar Year
SSD = Solid State Drive
Tb = Terabit
TLC = Triple Level Cell (3 bits per cell)

From 12May22 Micron Investor Day 2022

The world's fastest TLC NAND[1]

Micron is now the first to ship the industry's 9th-generation (G9) 3D NAND in an SSD, and for the third-generation in a row[2]

- Advanced building block for cutting-edge storage
- Design flexibility with ultra-compact, ultra-dense storage
- High performance for devices from your PC to the edge and into the AI-enabled cloud

Peak performance.[1] Density dominance.[1]

3.6 GB/s
performance[1]

Up to **99%** better read[1]

Up to **88%** better write[1]

World's densest NAND is now shipping in the Micron 2650 SSD[1]

Up to **73%** denser NAND[1]

Up to **28%** more space efficient[1]

Ideal for the most demanding high performance and data intensive workloads

- Data center
- Client
- Mobile
- Automotive and embedded

Learn more at micron.com/G9

Sources

[1] Competitors are identified as SK Hynix, Solidigm, Kioxia, WD and Samsung Semiconductor. The comparisons of I/O speed and design are based on the specifications detailed in the datasheets for NAND shipped in an SSD at the time of Micron's G9 NAND product announcement. The evaluations of read/write bandwidth performance and density are derived from tests conducted in Micron's laboratories, utilizing NAND that was commercially available.

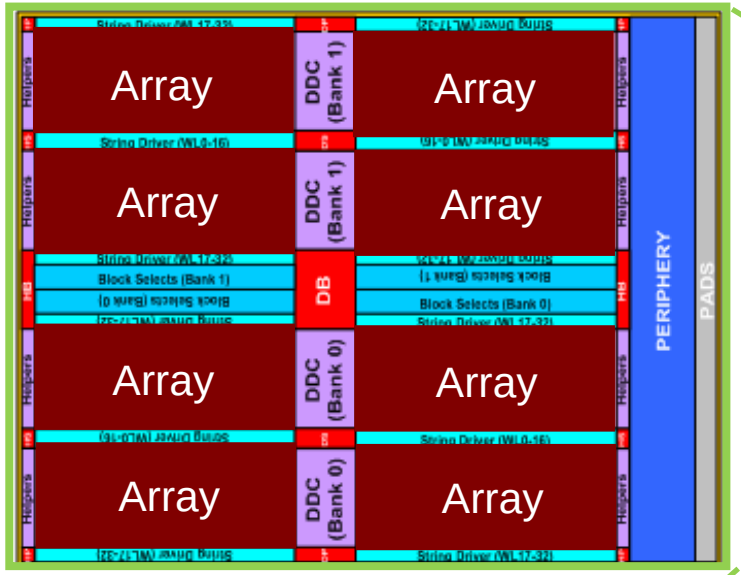
[2] Micron announced shipment of 9th-generation NAND (G9) in the Micron 2650 NVMe, a first for industry 9th-generation NAND. Micron was previously first to announce the industry's 7th- and 8th-generation NAND shipping in an SSD in 2020 and 2022, respectively, and now is first to ship again per footnote 1. See **https://investors.micron.com/news-releases/news-release-details/micron-ships-worlds-first-176-layer-nand-delivering-breakthrough** and **https://investors.micron.com/news-releases/news-release-details/micron-ships-worlds-first-232-layer-nand-extends-technology**.

From 30Jul24 Micron Press Release LINK

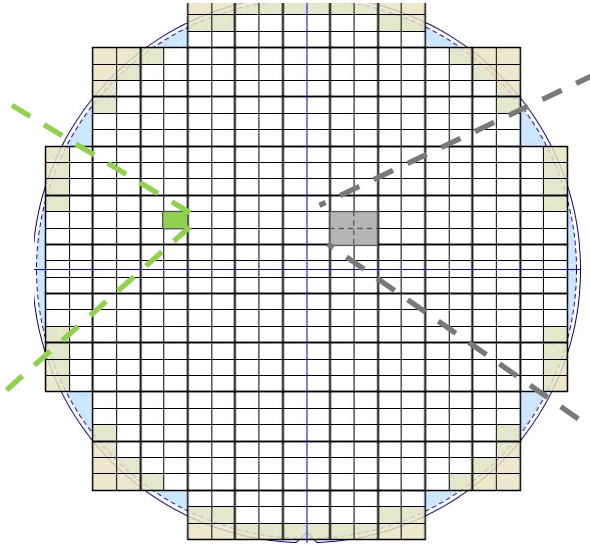
6. Key Terminology/ Glossary

Terminology: wafer, die, array, periphery, scribe

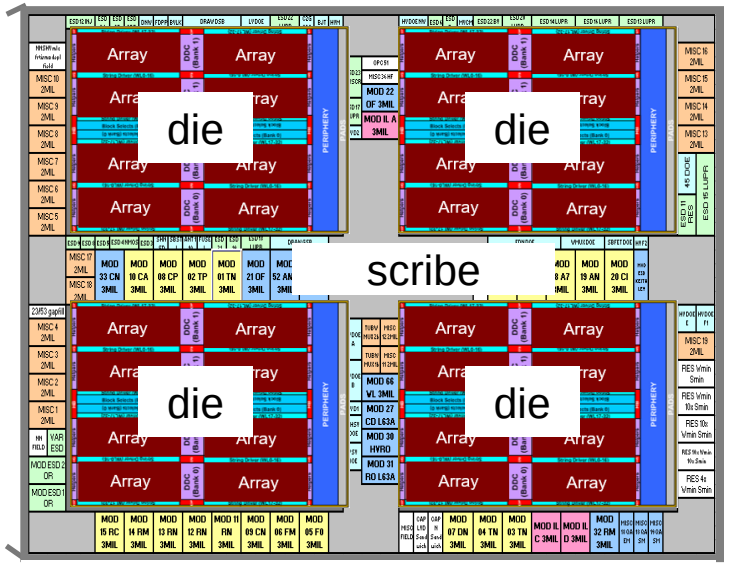
- Memory is fabricated on a 300 mm diameter silicon **wafer**. We try to maximize our die per wafer.
- A **die** is the memory chip which has a memory **array** and a **periphery**
  - The **array** (DRAM cells or NAND cells) stores information
  - The **periphery** has many circuits that operate with the array (pumps, regulators, IO (Input/Output), ESD, etc.)
- The **scribe** or **frame** is the region between die. It contains hundreds of alignment and metrology structures and electrical test circuits to enable in-line measurements and electrical testing. Die gets cut apart through the die singulation process for packaging.



Example of a die floorplan. Array is the dark red. Periphery shown in other colors.



Wafer



Example of scribe circuits around die (not to scale)

Glossary

Term or Acronym	Definition/Description
Array	The die has a memory array and a periphery. The memory array is where the data (1s and 0s) is stored.
Capacitor	An electronic device used to store electric charge. A capacitor consists of two conductive “plates” and an insulating layer known as a “dielectric” that electrically separates the two plates.
Conductor	A material that allows electrical current to flow easily. Examples of conductors used in integrated circuits are metals like aluminum, tungsten and copper.
CuA	CMOS Under Array - a Micron innovation. An architecture that consists on locating all support circuitry (periphery) underneath the memory cell array. CuA allows the die size to be further minimized, thereby gaining higher bit density (like all supporting utilities and parking garage installed beneath a skyscraper!)
Die	A memory die is a semiconductor memory chip in its state before it is packaged. At Micron, memory die are fabricated on a silicon substrate. Hundreds of memory die are built on the wafer, and after fabrication is completed, each memory die gets cut apart through the die singulation process. Individual memory die are very delicate, so they next go through the Packaging process.
Diode	An electronic device that allows current to flow in one direction only. Formed by connecting a P-type material with an N-type material. Diodes can be used to electrically isolate adjacent devices from each other
DRAM	Dynamic Random Access Memory. This is one of the types of semiconductor memory that Micron designs and fabricates. In DRAM memory data is stored as electrical charge on a capacitor. Charge is quickly lost though, so data must be constantly “refreshed”. DRAM is one of the fastest memories and provides both fast read and write operations. DRAM can be found in applications like computers, servers, laptops, tablets, cell phones, AI applications, etc.
Frame	The frame or scribe is the region of the wafer that separates one die from another die. The frame contains hundreds of alignment and metrology structures to enable in-line measurements, and electrical test circuits to enable electrical testing. These structures are sacrificial as they are cut apart through the die singulation process as each die needs to be cut apart before packaging.
Insulator	A material that is very resistive to electrical current flow. Examples of insulators used in integrated circuits are silicon dioxide (SiO_2) and silicon nitride (Si_3N_4).
MLC	Multi Level Cell. it refers to a NAND architecture that allows to program 2 bits per memory cell.

Glossary

Term or Acronym	Definition/Description
MOSFET	Metal Oxide Semiconductor Field Effect Transistor. A low-power electronic transistor device used in digital applications.
NAND	This is one of the types of semiconductor Flash memory that Micron designs and fabricates. In NAND memory data is stored by “trapping” charge in a film. Data can be stored for up to 10 years without need for refresh. While NAND has a slower read and write than other memory types like DRAM, this memory is a good fit for many non-volatile high-density applications that require large amounts of storage like cellphones, SSDs, etc. NAND is not an acronym; NAND refers to the series arrangement of memory cells similar to the NAND (Not AND) digital logic.
Non-volatile	Refers to the type of semiconductor memory that will retain stored information even when power is removed from the part. Examples of non-volatile memory are NAND Flash and NOR Flash.
NOR	This is one of the types of semiconductor Flash memory that Micron designs and fabricates. Similar to NAND, in NOR the memory data is stored by “trapping” charge in a film. But unlike NAND, the NOR architecture is designed for faster read/write operations. While NOR is slower than DRAM memory, this memory is a good fit for many non-volatile applications and it is used for applications that require fast access time like code execution, for example for storing and booting the Operative System (OS). NOR is not an acronym; NOR refers to memory cells arranged in parallel like a NOR (Not OR) digital logic.
Periphery	A memory die is a memory chip in its state before it is packaged. The die has a memory array and a periphery. The periphery is the region of the die that has the many circuits that allow to operate on the memory array. Some of the periphery circuits are pumps, regulators, IO (Input/Output), ESD (electrostatic discharge circuits), etc.
QLC	Quad Level Cell. it refers to a NAND architecture that allows to program 4 bits per memory cell.
Resistivity	Electrical resistivity is a measure of how strongly a material opposes the flow of electrical current. Resistivity is represented by the Greek letter ρ. Low resistivity indicated a material that readily allows movement of electrical charge.
Resistor	A device that reduces or limits current flow. A resistor is created by using a material less conductive (more resistive) anywhere along the path of current flow.
Scribe	See Frame
Semiconductor	A material that is between conductors and insulators on the resistivity scale. The semiconductor material used at Micron is silicon.

Glossary

Term or Acronym	Definition/Description
SLC	Single Level Cell. In NAND Flash, it refers to a NAND architecture that allows to program only 1 bit per memory cell.
Silicon, doped	A small percentage of controlled impurities or dopants are added to the silicon to reduce its resistance.
Silicon, intrinsic	Intrinsic Silicon is pure silicon. This material has 4 valance electrons and is very resistive. Intrinsic silicon has very few applications in memory chip manufacturing.
Silicon, n-type	An n-type dopant is an element that is added to the silicon that has more valence electrons than the silicon. The most common n-type dopants used at Micron are arsenic (As) and phosphorous (P), both with 5 valence electrons. In the regions of the die where the silicon has been doped n-type there is an excess of electrons so current is carried by electrons.
Silicon, p-type	An p-type dopant is an element that is added to the silicon that has less valence electrons than the silicon. The most common p-type dopant used at Micron is boron (B) with 3 valence electrons. In the regions of the die where the silicon has been doped p-type there is an excess of “holes” so current is thought as carried by “holes” that have a positive charge.
TLC	Triple Level Cell. In NAND Flash, it refers to a NAND architecture that allows to program 3 bits per memory cell.
Transistor	An electronic device that can control the current running though it – like an On/Off switch.
Volatile	Refers to the type of semiconductor memory that will lose stored information when power is removed from the part. Examples of volatile memory are DRAM and SRAM.

7. Document Updates

Document Updates

Date	Description
January 2025	• Added goals, objectives and target audience • Added several new images • Added acronym legends • Added glossary section

Educator Hub