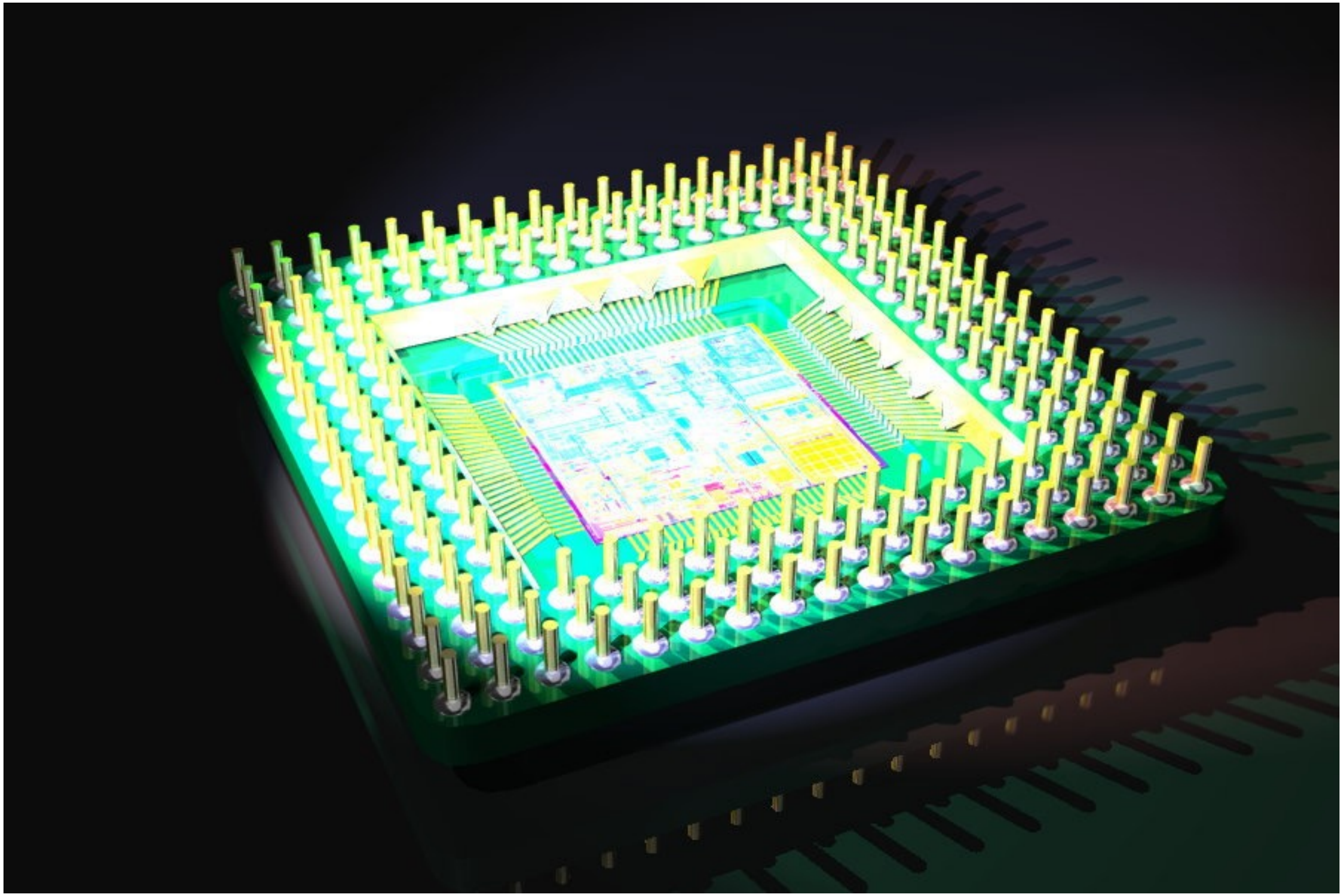
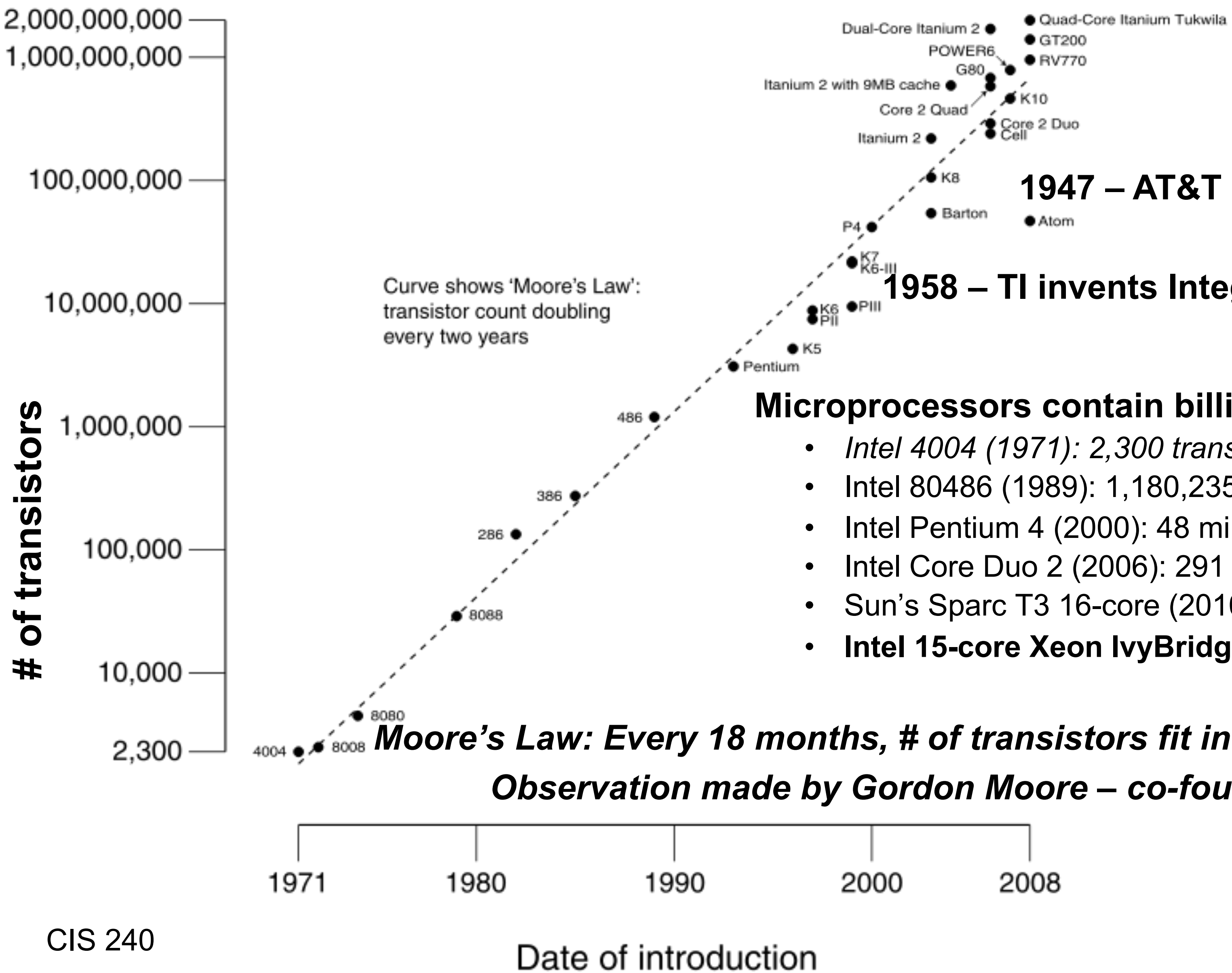
Chapter 3 (Part 0)
Transistors & Gates

Inside a microprocessor



Transistor: Building Block of Computers



1947 – AT&T invents transistor

Shockley

1958 – TI invents Integrated Circuit (IC)

Jack Kilby

Microprocessors contain billions of transistors

- *Intel 4004 (1971): 2,300 transistors*
- Intel 80486 (1989): 1,180,235 transistors
- Intel Pentium 4 (2000): 48 million
- Intel Core Duo 2 (2006): 291 million
- Sun's Sparc T3 16-core (2010): 1 billion
- **Intel 15-core Xeon IvyBridge-EX (2014): 4.3 billion**

Moore's Law: Every 18 months, # of transistors fit into an IC will double

Observation made by Gordon Moore – co-founder of Intel

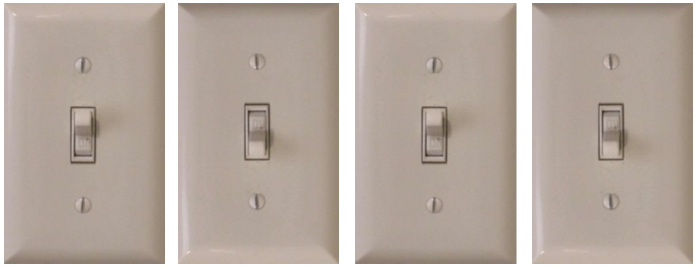
Transistor: Building Block of Computers

So what exactly is a transistor?

- A <u>semiconductor</u> device used to…
  <u>AMPLIFY</u> or <u>SWITCH</u> electronic signals in a circuit
  *CPUs mostly use the "SWITCH" capability

How does a SWITCH form the basis of a computer?

- Previously…we introduced "binary digital system"
  - Two symbols: 0 and 1, represents two states
- Switch also represents two states: OFF and ON



ON	OFF	ON	OFF
1	0	1	0

- Combine them to make logic: AND/OR/NOT
- Even higher: adders/mux/decoders, finally CPUs (LC4)

How Does a Computer Represent Data?

At the lowest level, a computer has electronic "plumbing"

- **Operates by controlling the flow of electrons**

Easy to recognize two conditions

1. **Presence of a voltage – we'll call this state "1"**
2. **Absence of a voltage – we'll call this state "0"**

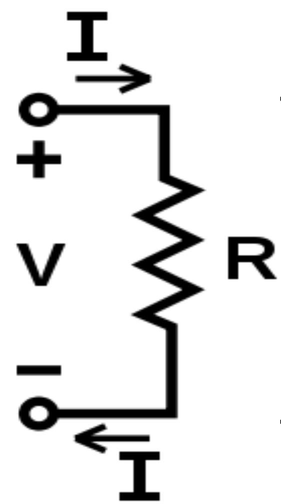
Computer use transistors as switches to manipulate bits

- **Before transistors: tubes, electro-mechanical relays (pre 1950s)**
- **Mechanical adders (punch cards, gears) as far back as mid-1600s**

How do transistors work?
→1st we need a mini-Physics review…

- **Electricity corresponds to the flow of charged particles, typically electrons. (see Ben Franklin)**
  - **Current = flow of charge carried by electrons**
  - **Voltage = *motivation* for current to flow (potential difference)**

Difference In height *motivates* water to flow



Difference in voltage *motivates* current to flow

Ohm's Law: V = IR

Charge

Atoms contain two types of charged particles, protons in the nucleus and electrons that orbit the nucleus. Their charges are equal magnitude and opposite sign

Charge is measured in Coulombs

The charge of an electron is approximatel 1.6e-19 Coulumbs

Like charges repel and opposite charges attract

Current

Current refers to the flow of charged particles

Current is measured in Amperes where a 1 Ampere current refers to a flow of one Coulumb of charge per second

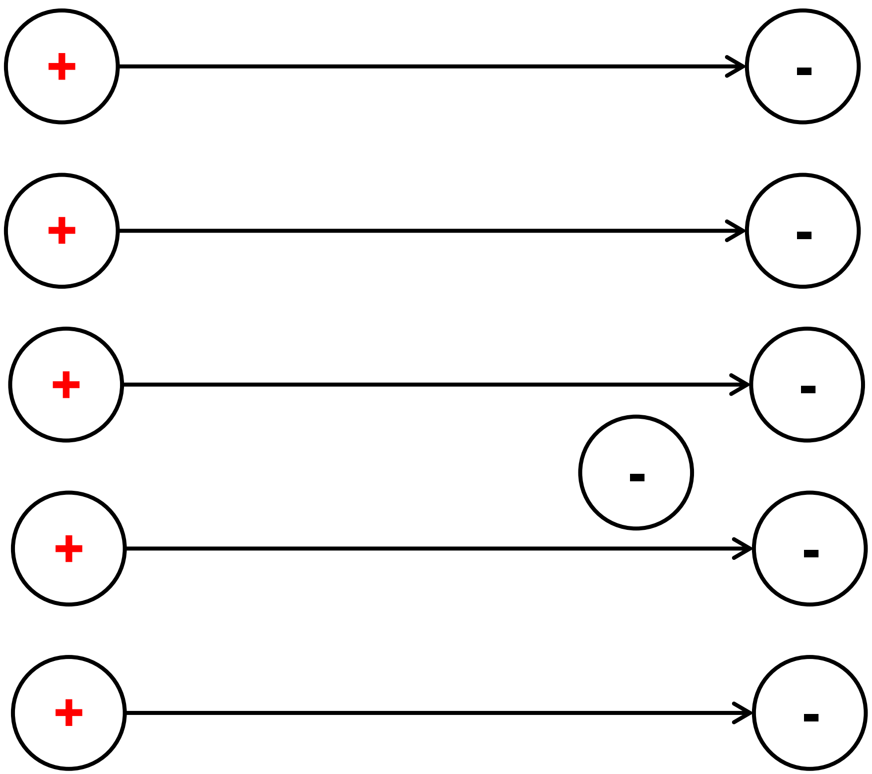
Voltage

The potential difference between two points is refered to as voltage it can be defined as the amount of work required to move one Coulomb of charge between the two locations.

Voltage/Current and Electric Field

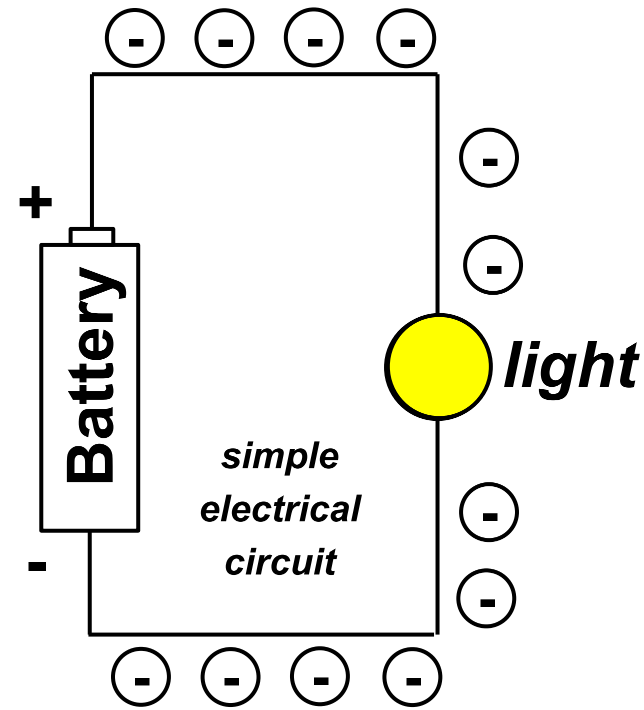
E-field: produces "voltage difference"

Aka: motivation for charge to flow



Battery provides <u>voltage</u>

Aka: voltage difference



Direction of charge carrier (e-)

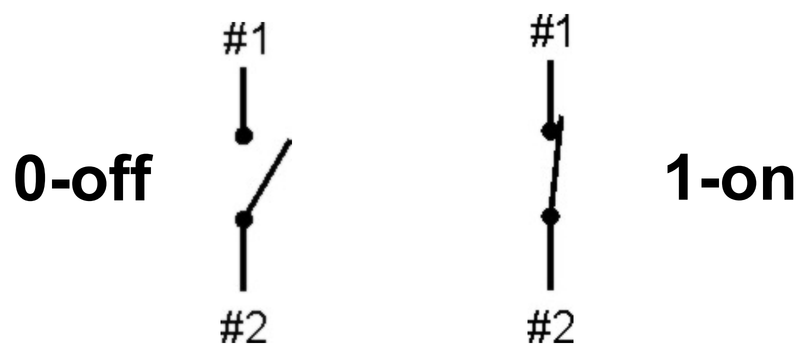
Direction of current

Direction of current

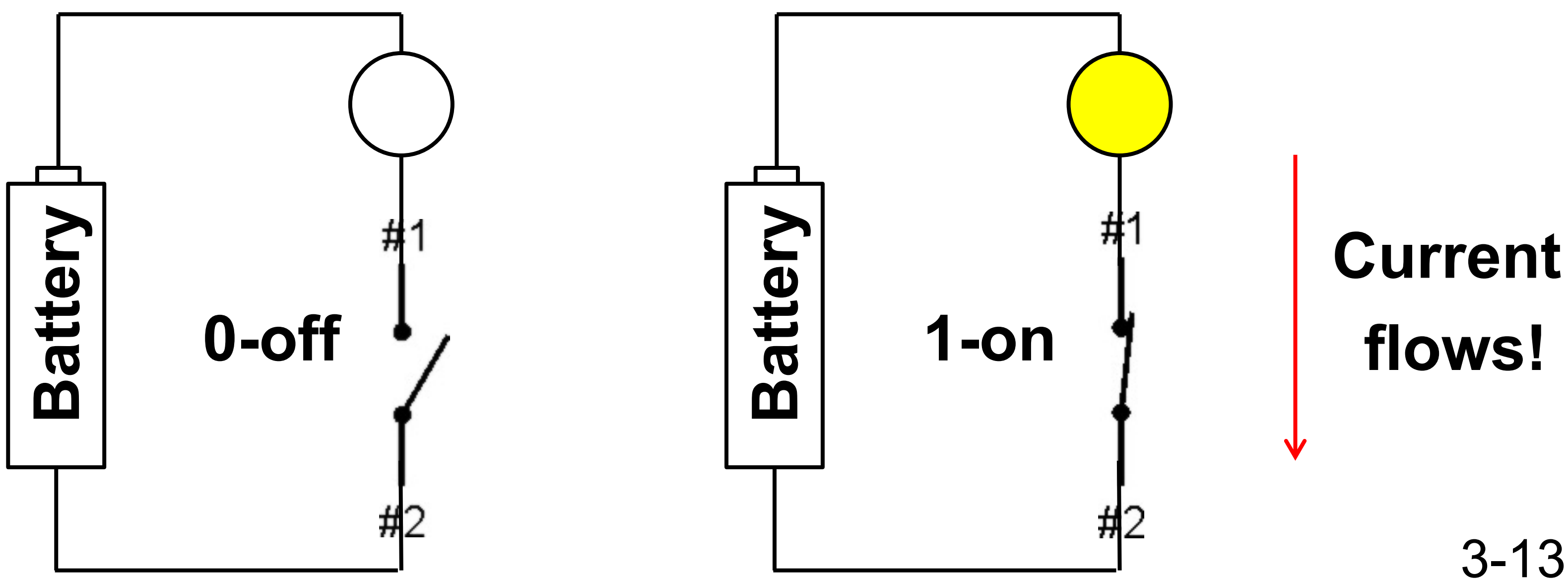
Ohm's Law: V = IR

Switches control flow of current in a circuit

- A switch inherently represents two states, on/off

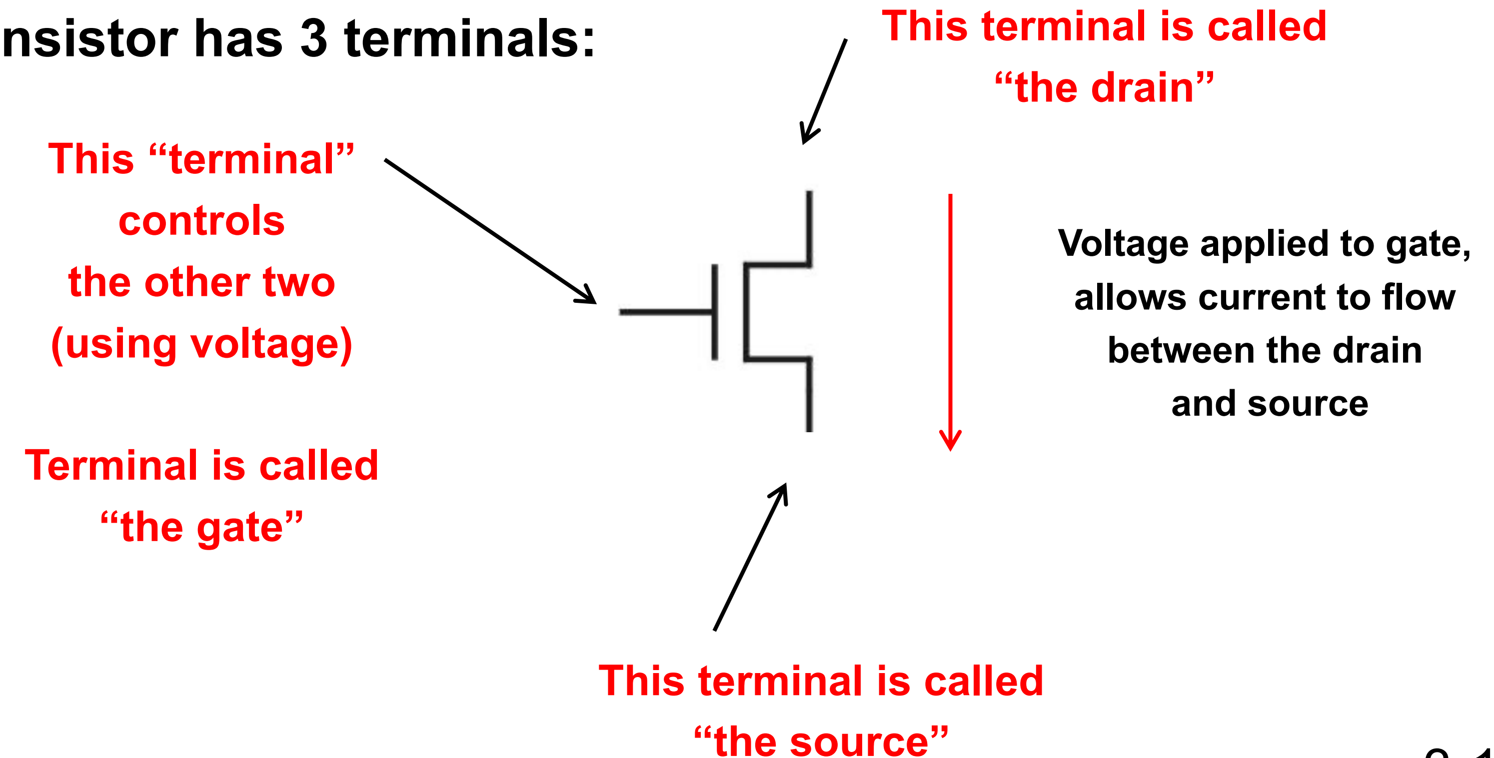


- When put in a circuit, can start/stop current flow



Transistor as electronic switch

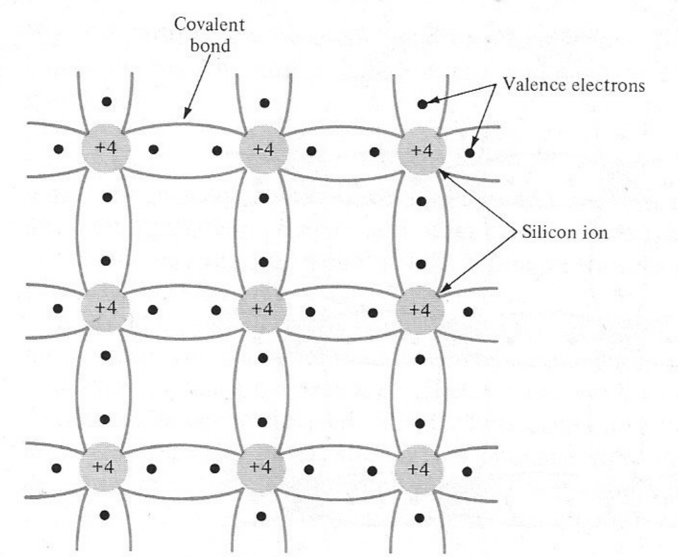
- In the last example, someone must manually "flip" the switches to control the "input" to our circuits
- In a computer we need a way to "remotely" flip the switch
  - ➢ Transistor offers us this capability
  - ➢ We use voltage, to remotely flip the switch
- A transistor has 3 terminals:

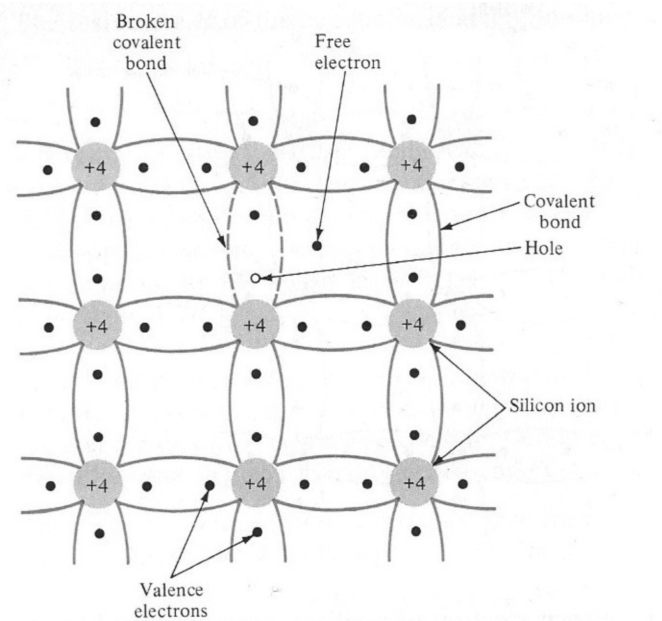


How does a transistor work?

Begin at the beginning (*what is it made of ?*)

- Currently transistors are made of Silicon
  - Atomic symbol: Si – atomic number 14
- In its crystalline state, silicon atoms form covalent bonds with four neighbors using their 4 outer electrons
- At room temperature, Silicon is a semiconductor

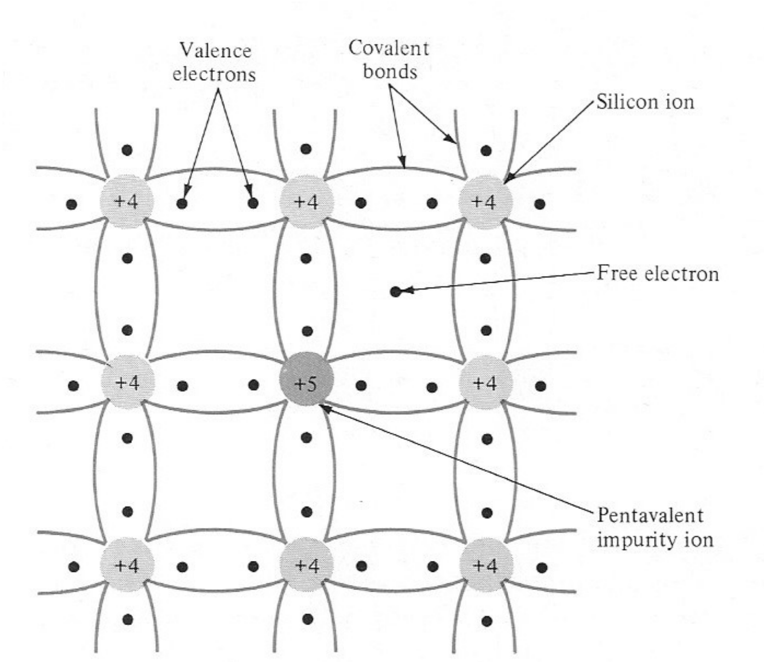


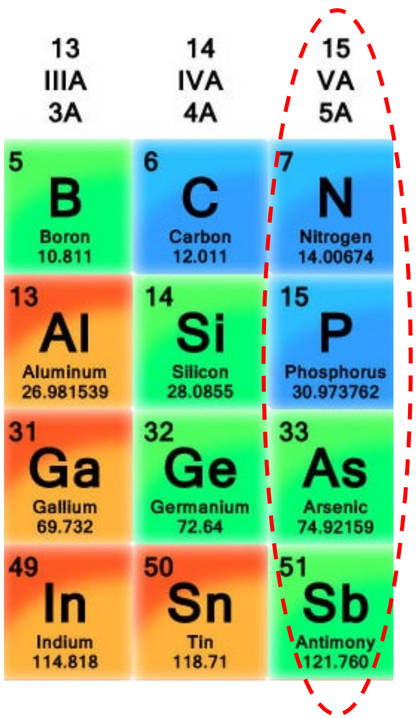


Doping – not what you think

We can improve the conduction of Silicon by *doping* it with other elements.

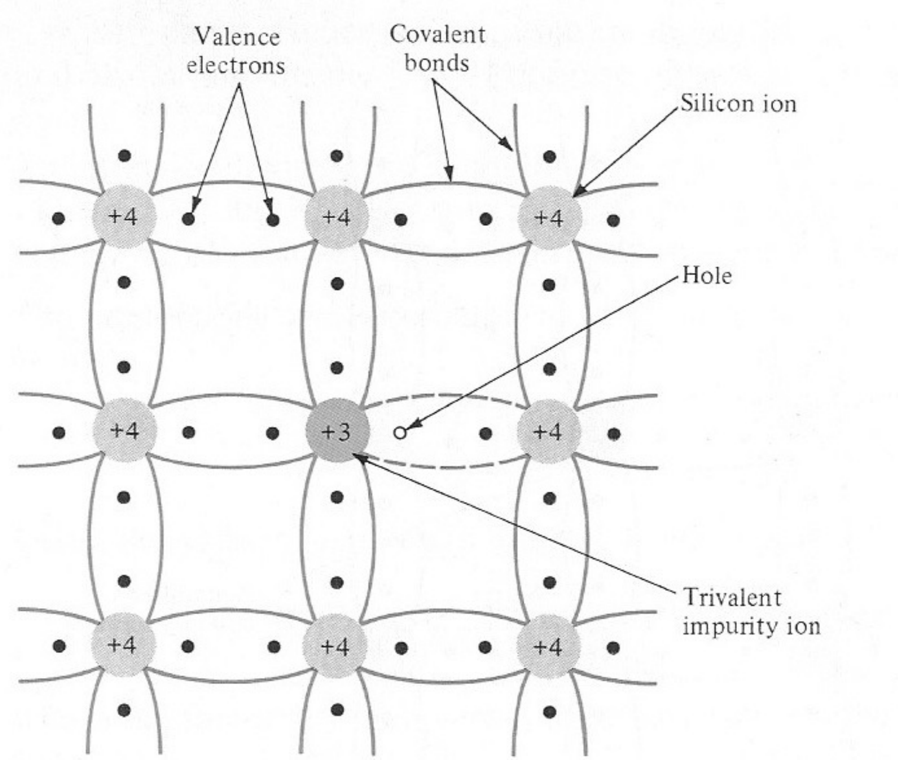
- **N-type regions are formed by adding small amounts of elements that have more than 4 electrons in their outer shell and, these extra electrons can serve as charge carriers.**
- **N-type dopants – antimony (Sb), phosphorus(P), arsenic (As)**

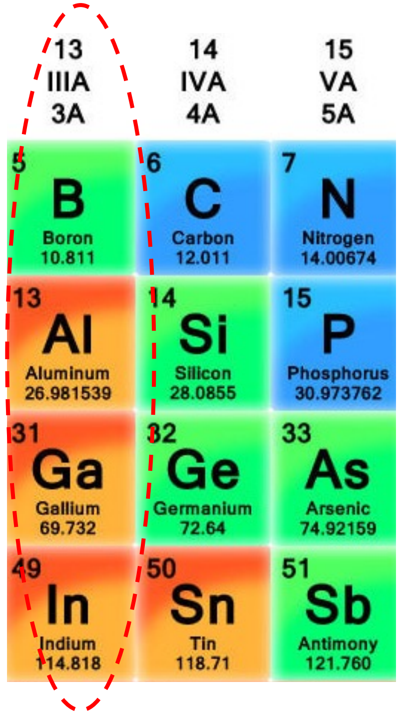




P-type doping

- P-type materials are formed by adding elements that have 3 electrons in their outer valence shell.
- These atoms create spaces in the lattice of covalent bonds into which electrons can flow.
- P-type dopants : boron (B), gallium (Ga), indium (In)





Bottom Line

- **N-type materials are good semiconductors because they have extra electrons which are negatively charged and can be used to carry a current.**
- **P-type materials are good semiconductors because they have extra spaces into which electrons can move. These ‘holes’ can be thought of as positive charge carriers.**

LET'S MAKE A DIODE

Review so far…

Our goal is to make an electronic "switch" (a transistor)

- **A device that will allow us to control the flow of current in a circuit**

Doping:

- **Add an "impurity" to silicon (our semiconductor)…**

  …to make it a better conductor of current

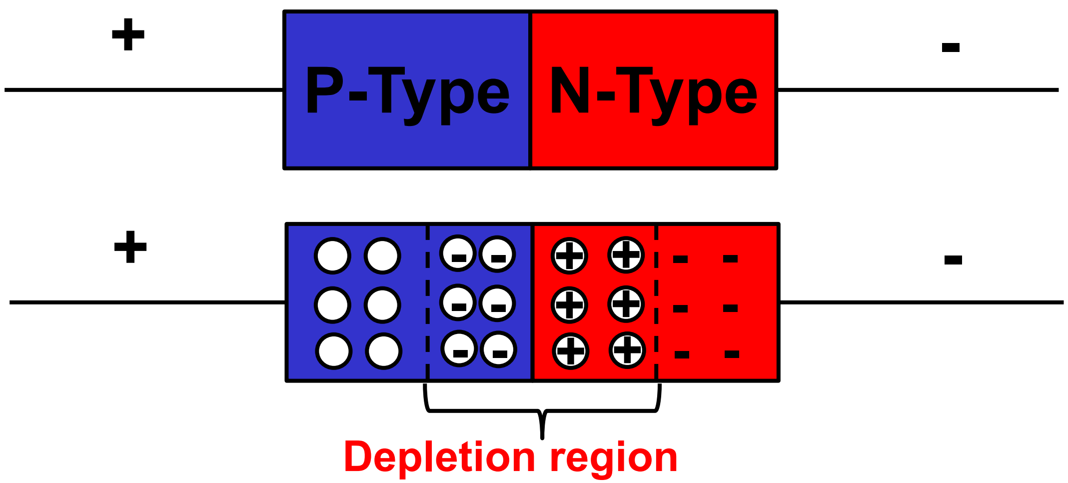
  - **nType – Impurity added, adds extra "electrons" to silicon**
  - **pType – Impurity added, takes away electrons, making "holes"**

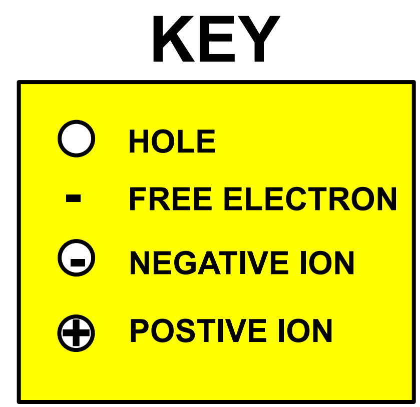
How we'll do this:

- **Use nType and pType Silicon to create a device that controls the direction of current flow in a circuit (like a one-way valve)**
- **Next, we'll configure our valve and add a lever (gate) to it**
  - **So we can turn current flow on/off in a circuit**
  - **Then we'll have a transistor!**

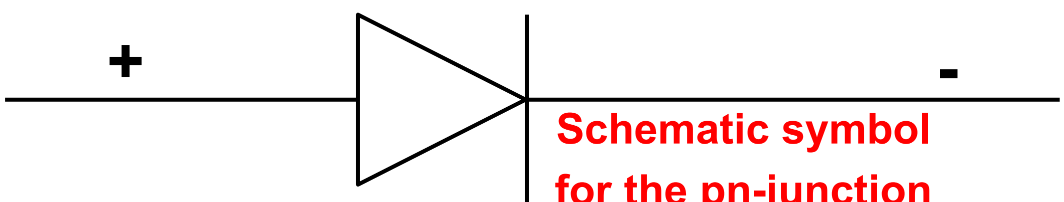
A Diode (a pn-junction)

- A union of P-type and N-type materials
- Functions as a one-way "valve" in an electric circuit
- Only allows current to flow in one direction





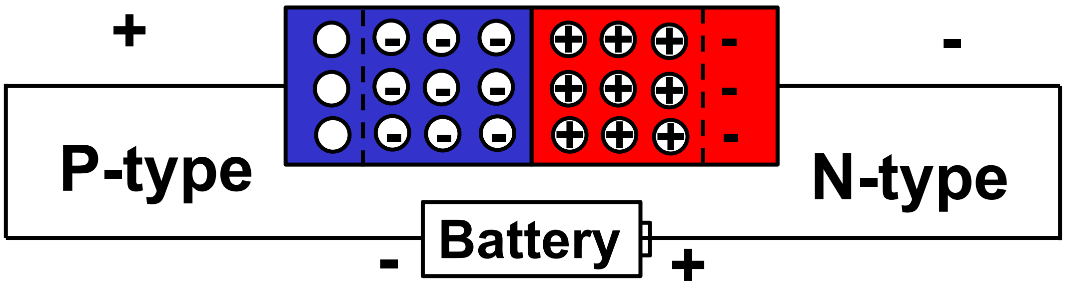
PN Junction reaches "equilibrium" setting up an E-field that stops flow of additional electrons across the junction

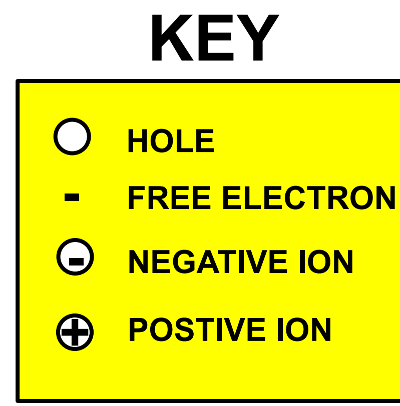


Schematic symbol
for the pn-junction
Call it a DIODE

A Diode (a pn-junction)

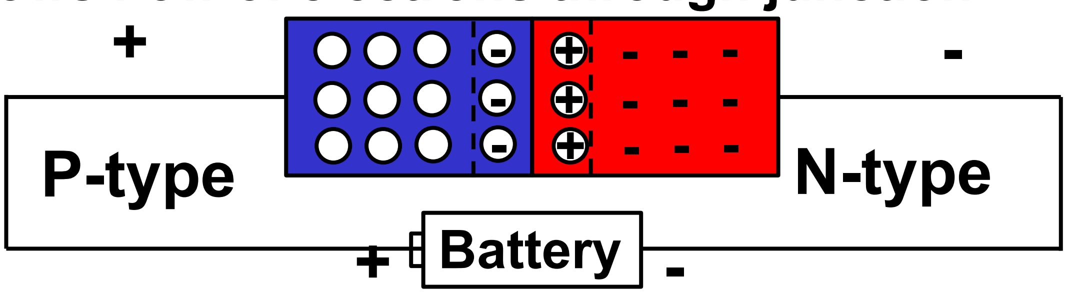
- Reverse bias (battery against the +/- terminal of diode):
  - Depletion region gets bigger (E-field gets stronger)
    - reduces/stops flow of current from + to -
  - Impedes flow of electrons through junction





- Forward bias:
  - Depletion region gets smaller (E-field shrinks)
    - Allows current to flow from + to -
  - Allows flow of electrons through junction

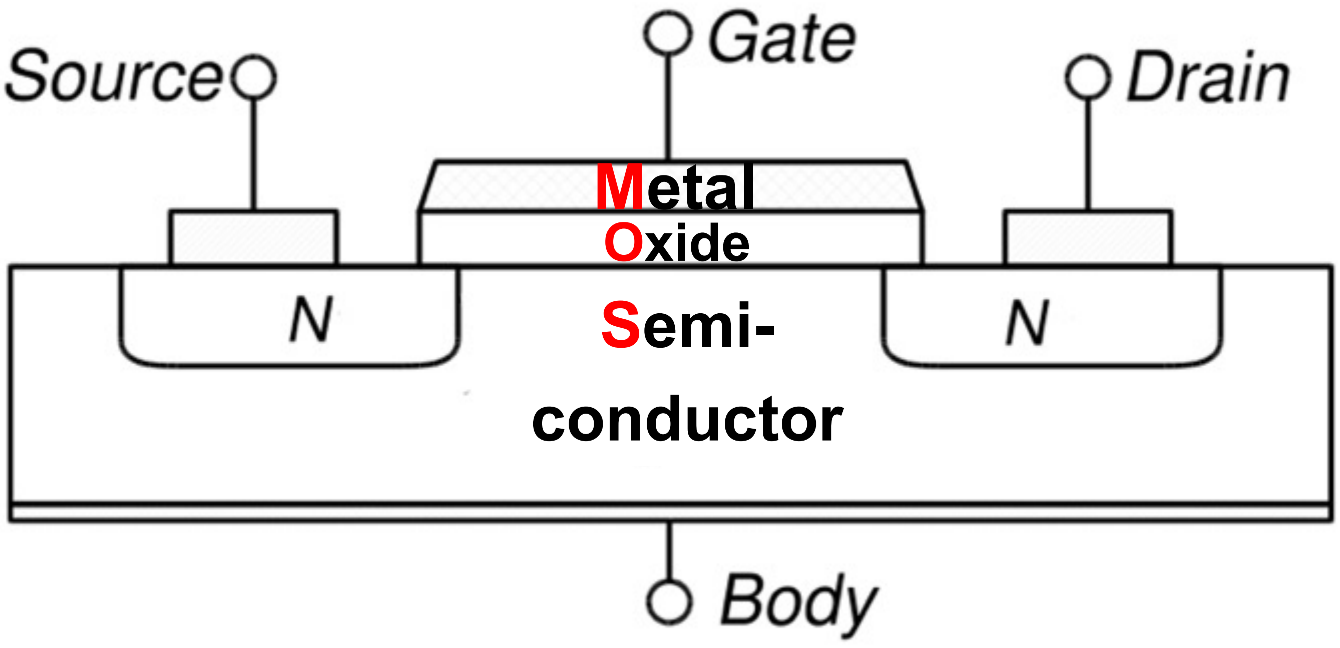
A diode is
Like a 1-way valve
Only lets current
In 1 direction in a
circuit



LET'S MAKE A TRANSISTOR

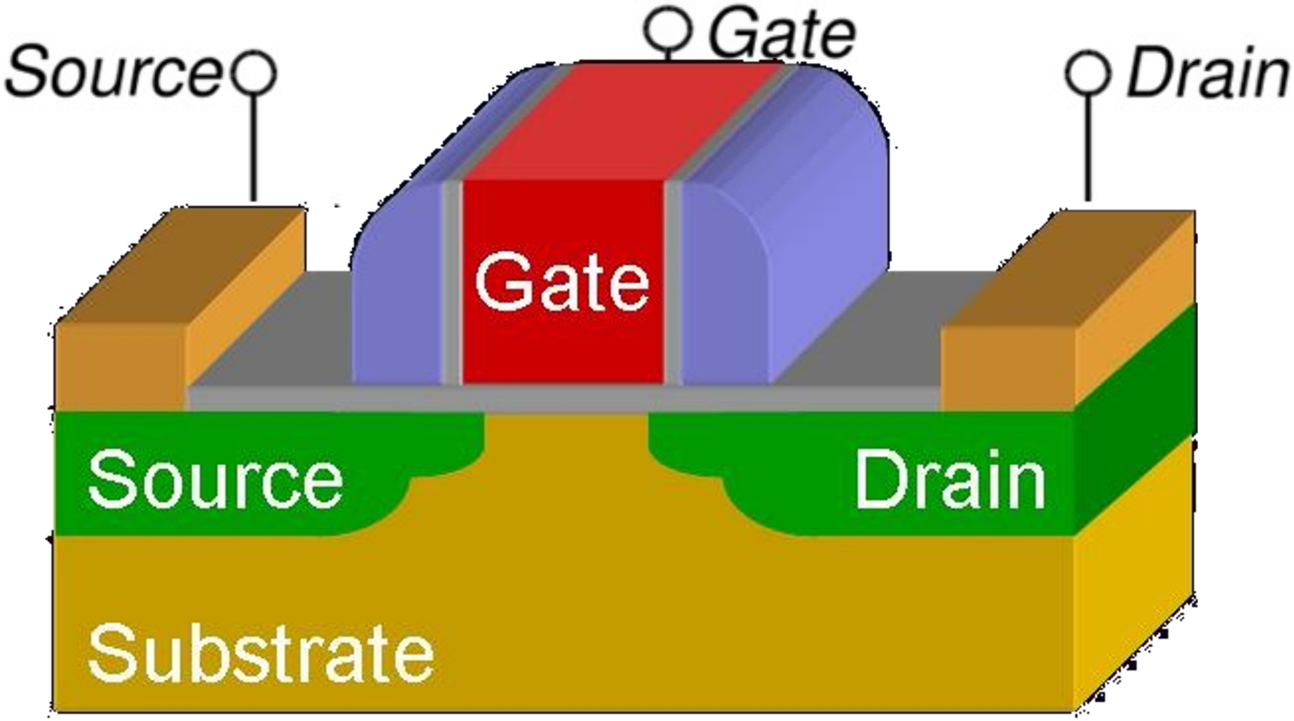
Next up…the MOSFET (your 1st Transistor!)

- **MOSFET : Metal Oxide Semiconductor Field Effect Transistor**
- **Picture shows a cross section of such a device.**
- **Notice it has 4 electrical terminals: Source/Drain/Gate/Body**



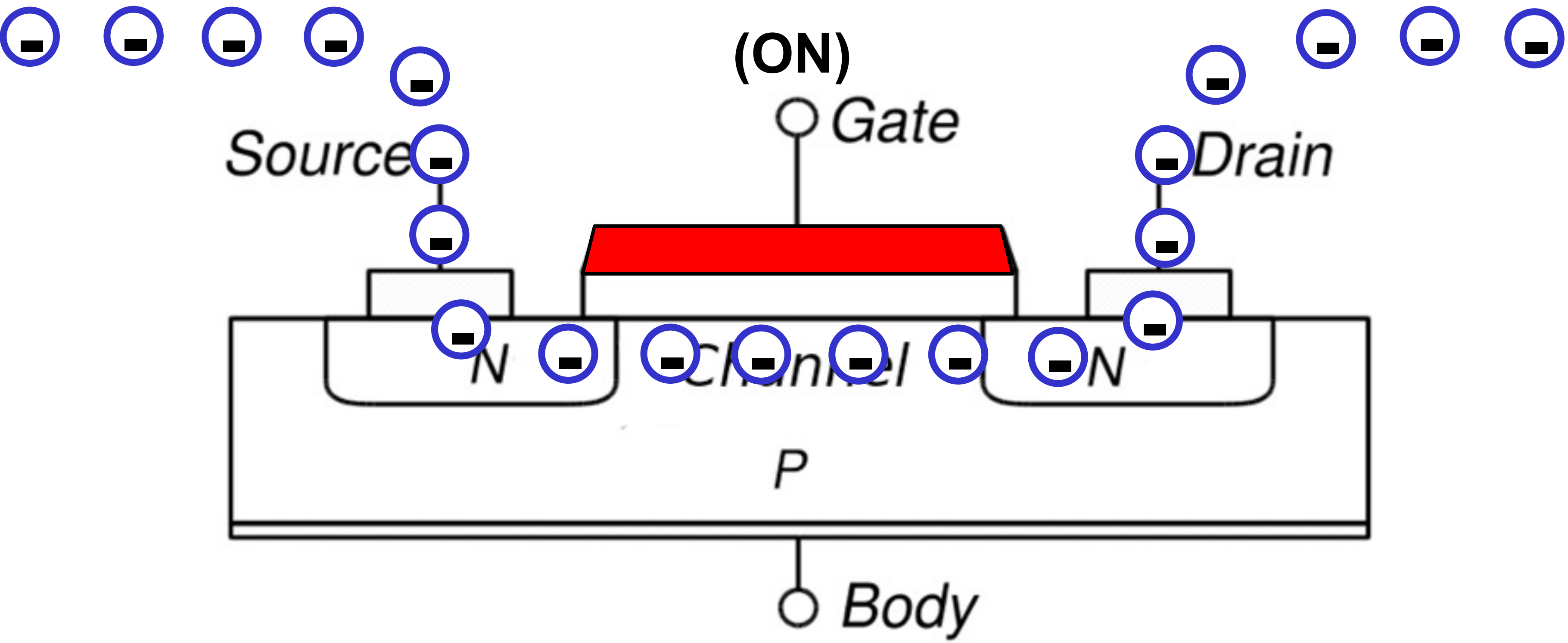
Next up…the MOSFET (your 1st Transistor!)

- **MOSFET : Metal Oxide Semiconductor Field Effect Transistor**
- **Picture shows a cross section of such a device.**
- **Notice it has 4 electrical terminals: Source/Drain/Gate/Body**



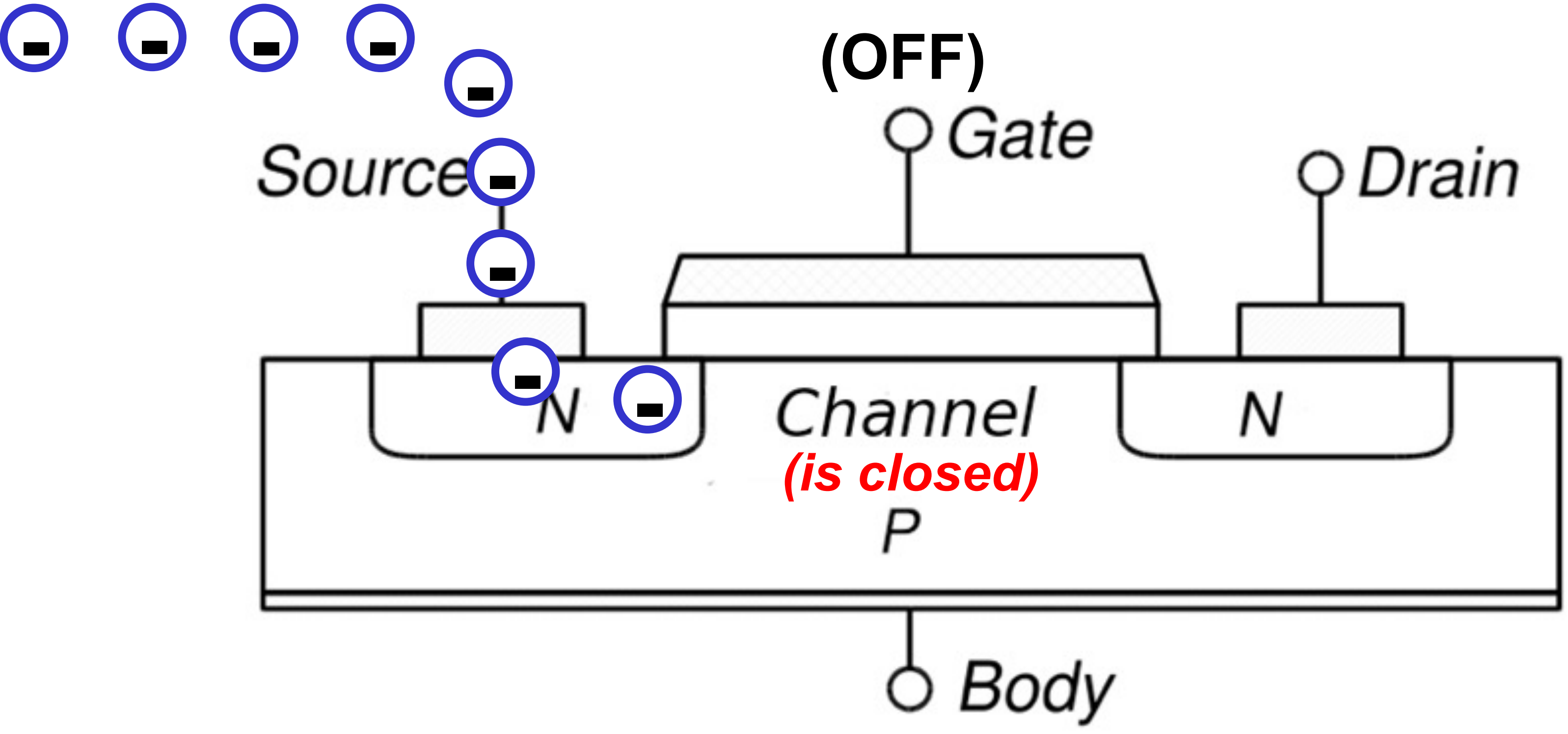
How we want it to work…

- Goal: Pass current through this device (from drain to source)
  - ➢ BUT we want to control that current (using the gate terminal)
    - – If GATE is ON
      - • electrons pass from source to drain through channel



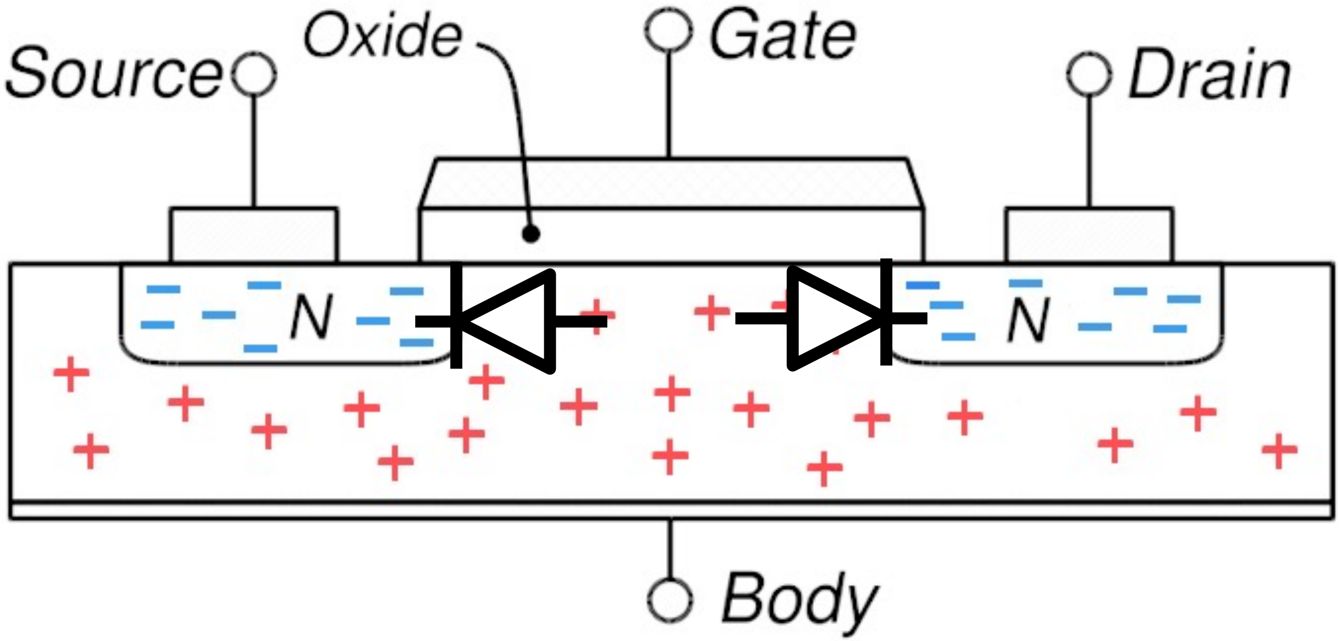
How we want it to work…

- Goal: Pass current through this device (from drain to source)
  - ➢ BUT we want to control that current (using the gate terminal)
    - – If GATE is OFF
      - • electrons cannot pass through channel



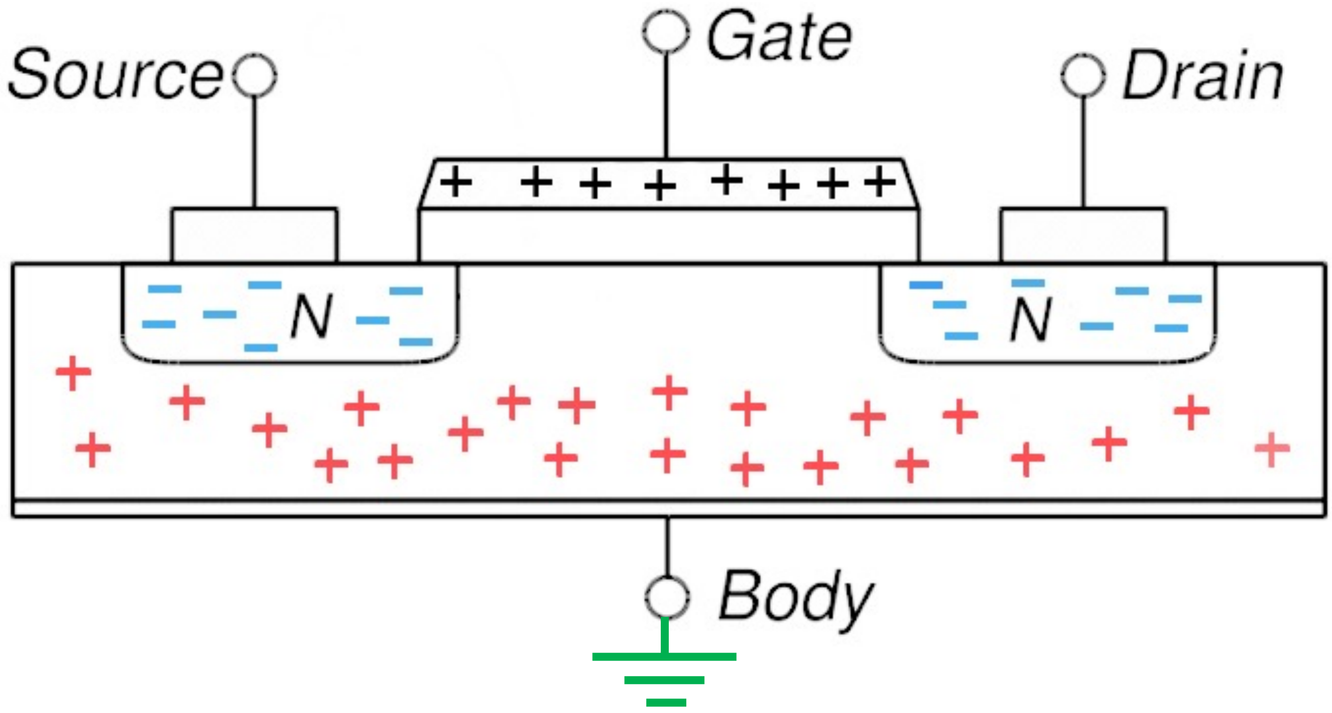
How we achieve this behavior…

- **At "rest" we have (closed state)**
  - **2 n-type spots (source/drain)**
  - **1 p-type spot (channel region)**
  - **2 back-to-back p-n junction diodes!**
    - **Halts flow of electrons through channel (channel doesn't exist!)**



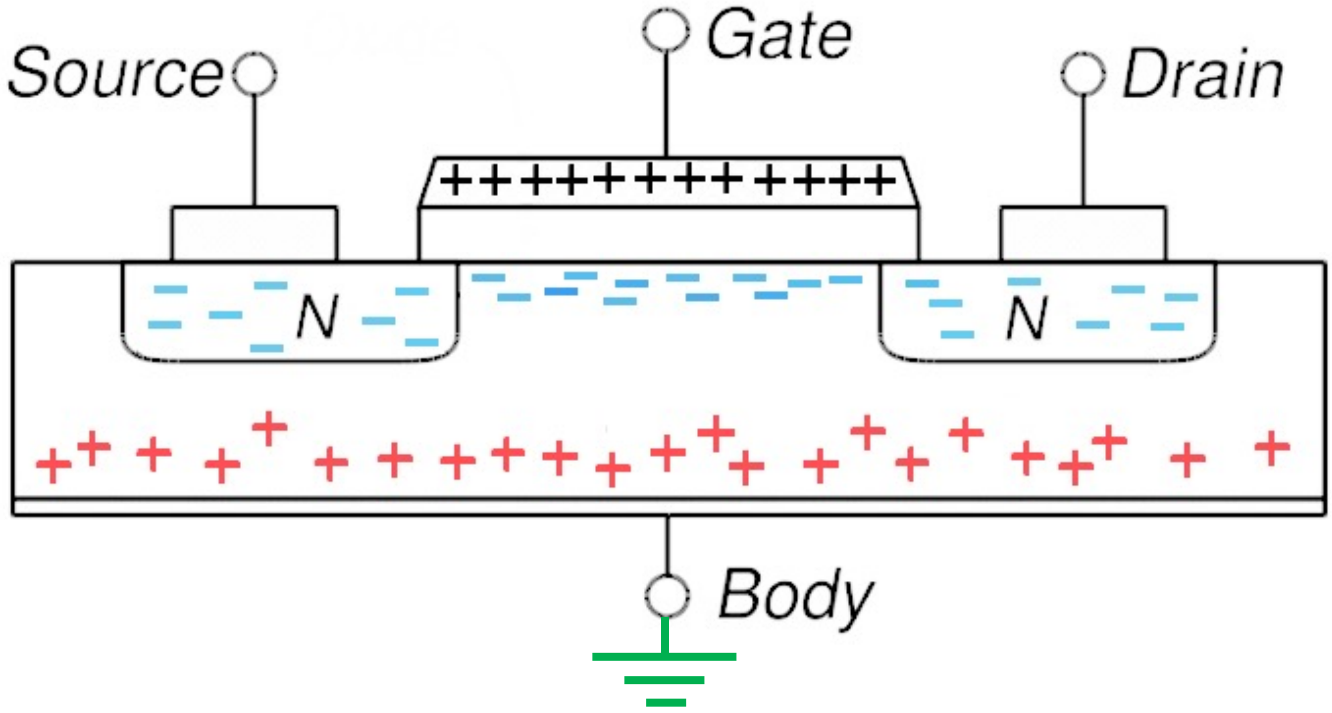
How we achieve this behavior…

- **If we wish to turn device on:**
  - **We apply a "positive" voltage to GATE with respect to BODY**
    - **This positive voltage "repels" holes from under the gate**
    - **"depletes" the future channel region of all its holes**



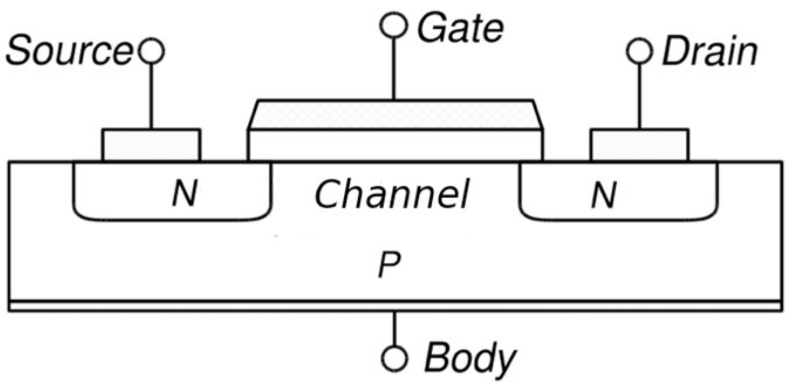
How we achieve this behavior…

- If we go further:
  - ➢ Apply a “very positive” voltage to the gate
    - – Begins to attract electrons (from source & drain)
    - – The channel region has been “inverted”
    - – Connects (electrically) source and drain, so current can flow!

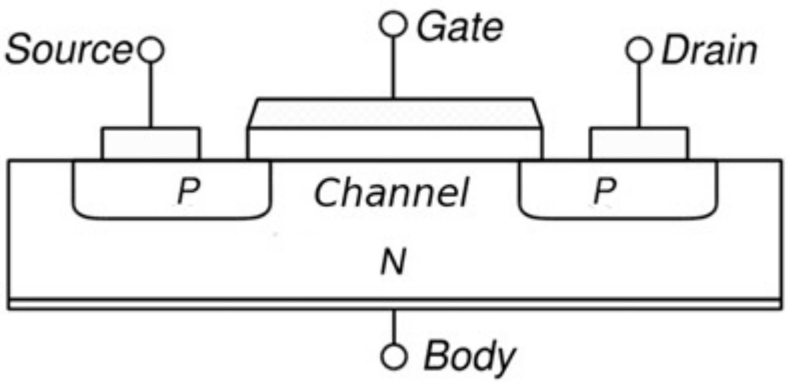


Two types of MOSFETs: nMOSFET and pMOSFET

- **nMOSFET (nMOS): channel carries negative charges (electrons)**
- **pMOSFET (pMOS): channel carries positive charges (holes)**



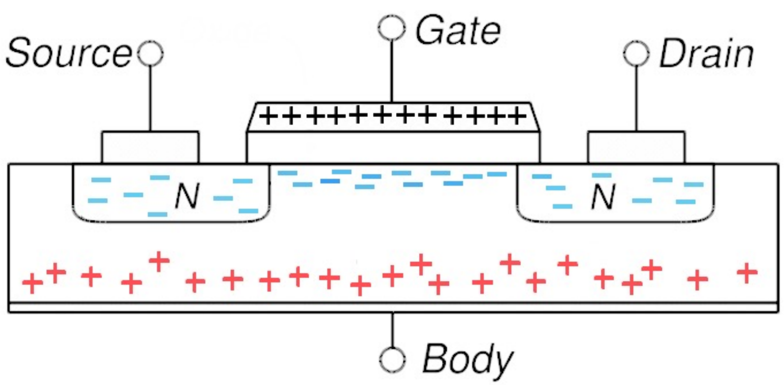
nMOSFET



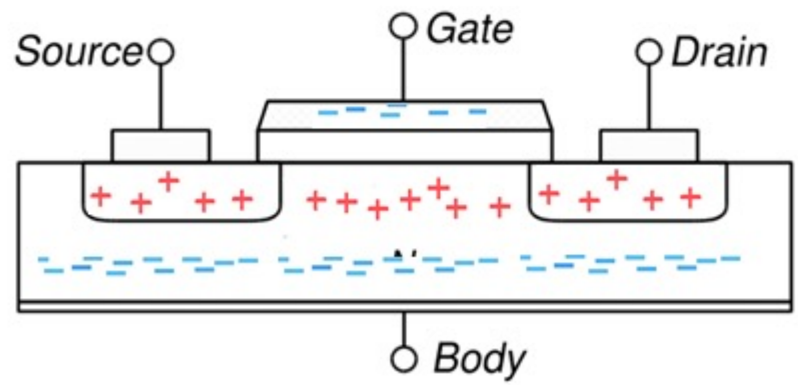
pMOSFET

Two types of MOSFETs: nMOSFET and pMOSFET

- **nMOSFET (nMOS): channel carries negative charges (electrons)**
  - **GATE Voltage MUST BE > Body, Source, Drain to be ON**
- **pMOSFET (pMOS): channel carries positive charges (holes)**
  - **GATE Voltage MUST BE < Body,Source, Drain to be ON**



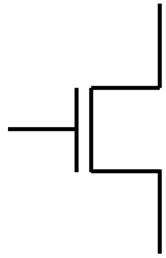
nMOSFET



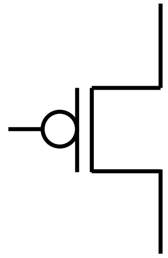
pMOSFET

Transistor Symbols

➢ **The circuit symbols for nmos and pmos transistors are shown below.**



NMOS Transistor Symbol



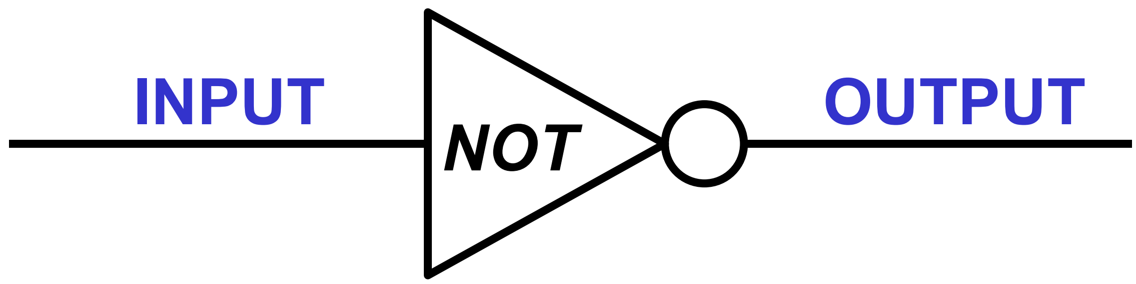
PMOS Transistor Symbol

TRANSISTORS TO LOGIC GATES

Transistor to logic device: GATE

Our first logic device will be an <u>inverter</u>:

2.9 V (HIGH/ON)



0 V (LOW/OFF)

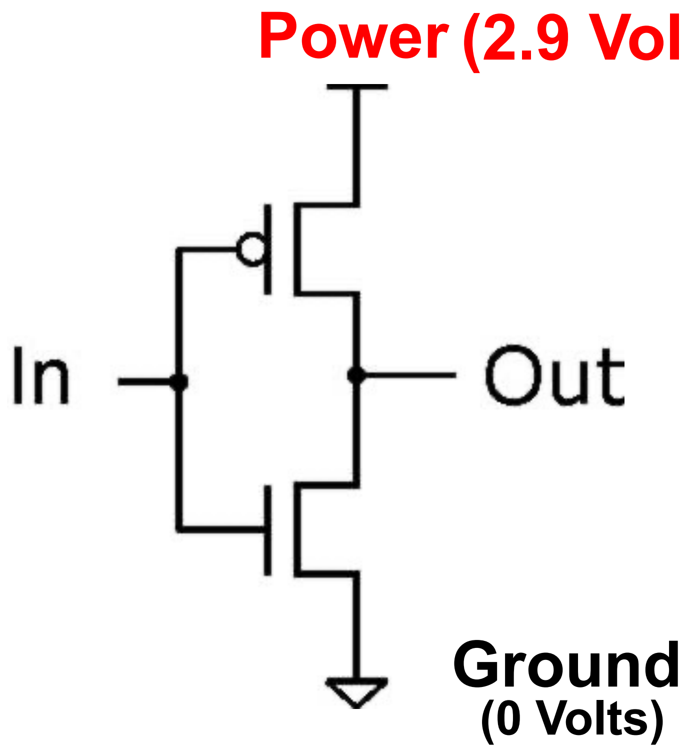
Logical Behavior: "inverts" the incoming signal:

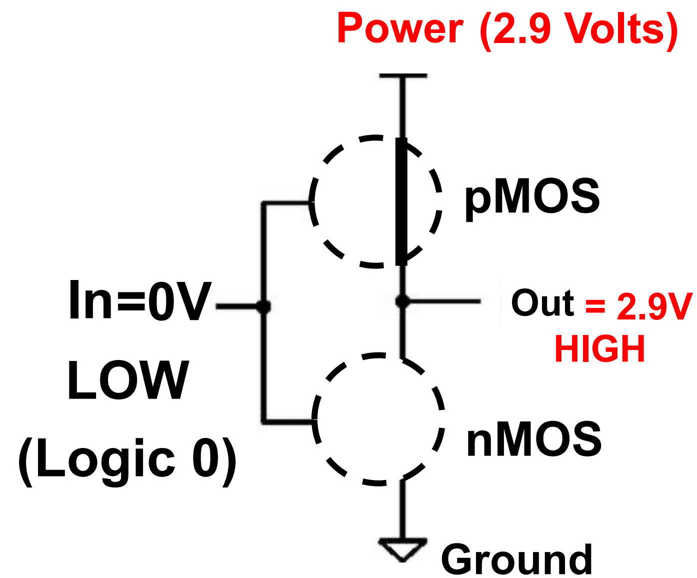
- **Input: LOW-> output: HIGH**
- **Input: HIGH->output: LOW**

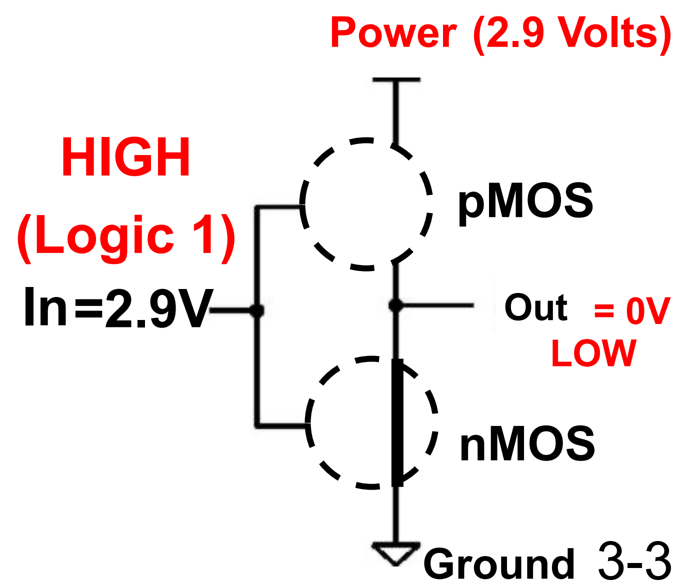
INPUT	OUTPUT
LOW (0)	HIGH (1)
HIGH (1)	LOW (0)

Truth Table All possible Combinations Of inputs

How do we configure transistors to make inverter?



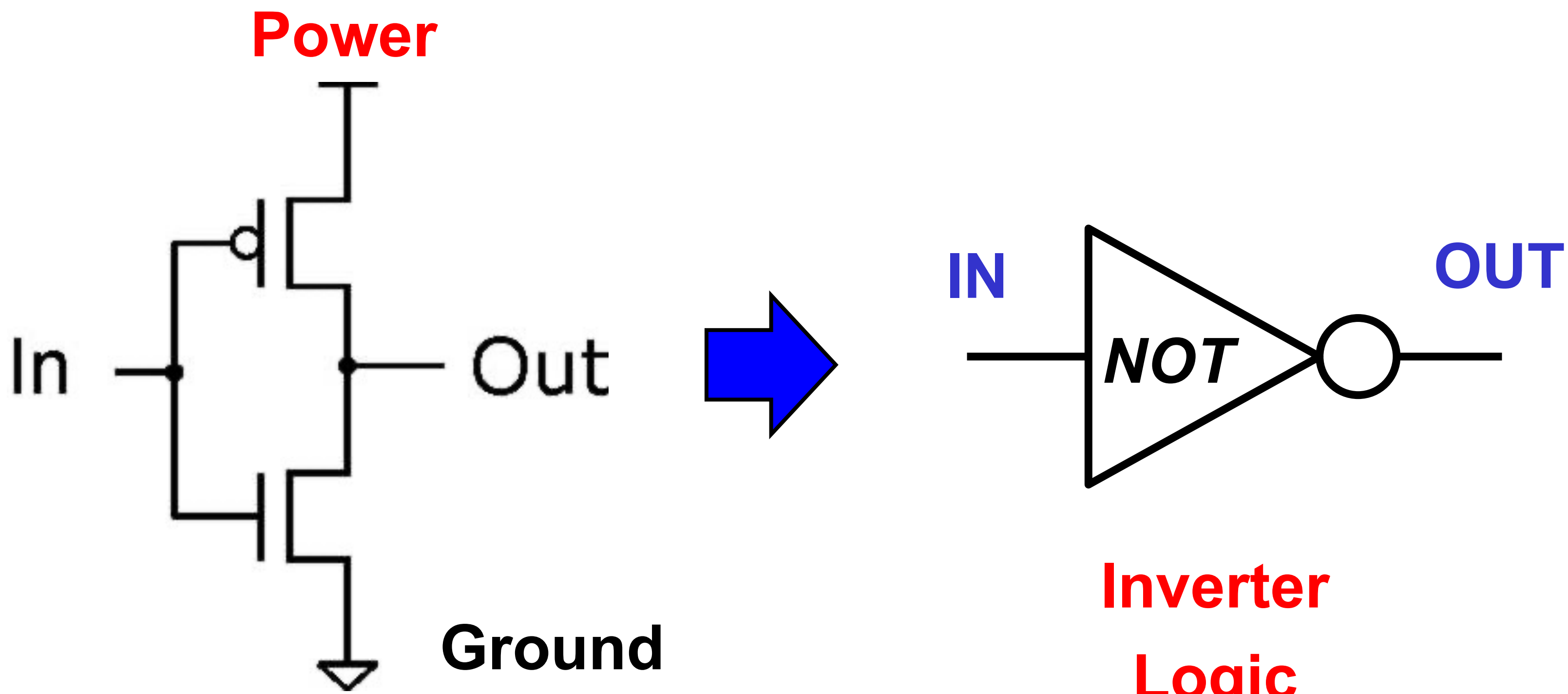




We take advantage of opposing nature!

- **-Setting IN = 0V (LOW=Logic 0)**
  - **nMOS turns OFF**
  - **pMOS turns ON, connects OUT to Power**
  - **OUT=2.9V (HIGH=Logic 1 – inverse of In)**

This configuration is called: CMOS



CMOS = Complimentary MOS Inverter

Inverter Logic Gate

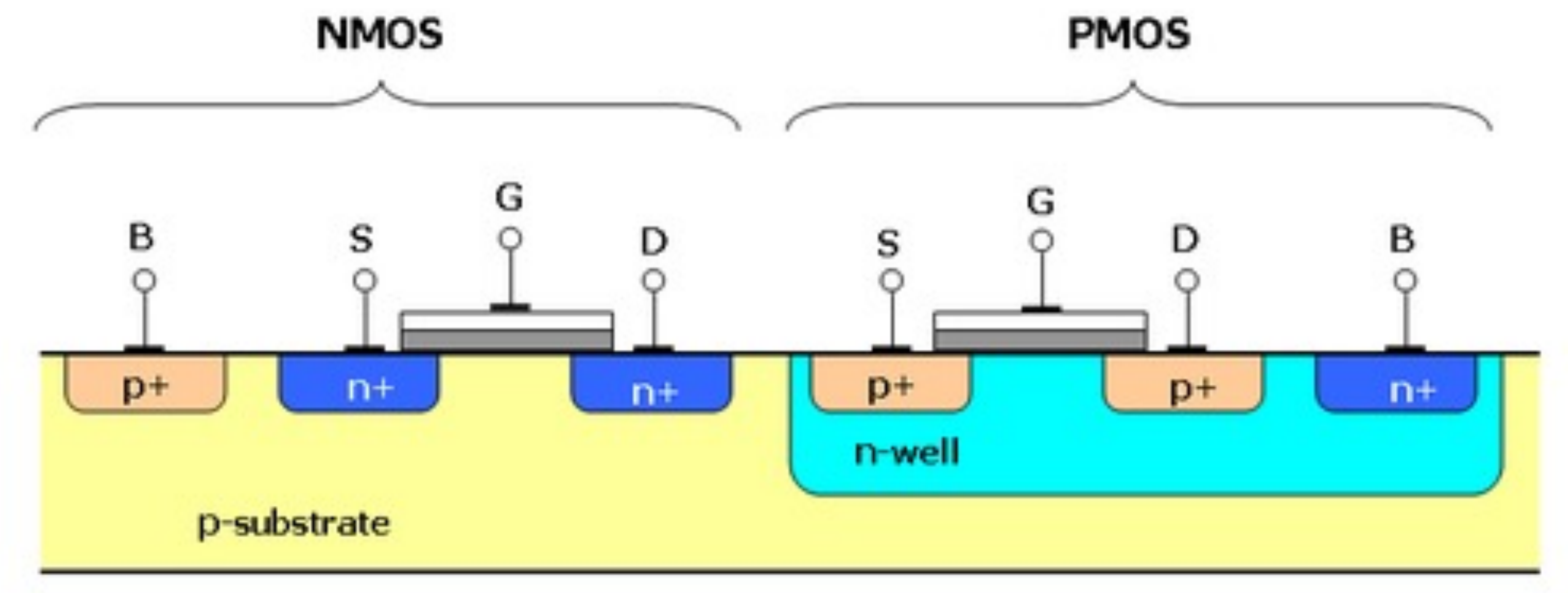
We have "jumped up" 1 level of abstraction

--From transistors to "gate"

--Technology inside the gate (CMOS here) isn't as crucial as its behavior

--could be: transistors, vacuum tubes, biological device, etc.

Side View of Transistors in CMOS process



We can use the same silicon to create both types of transistors!

--The same "p-type" substrate is used to hold both NMOS/PMOS

--NMOS devices can be created right in the p-type substrate

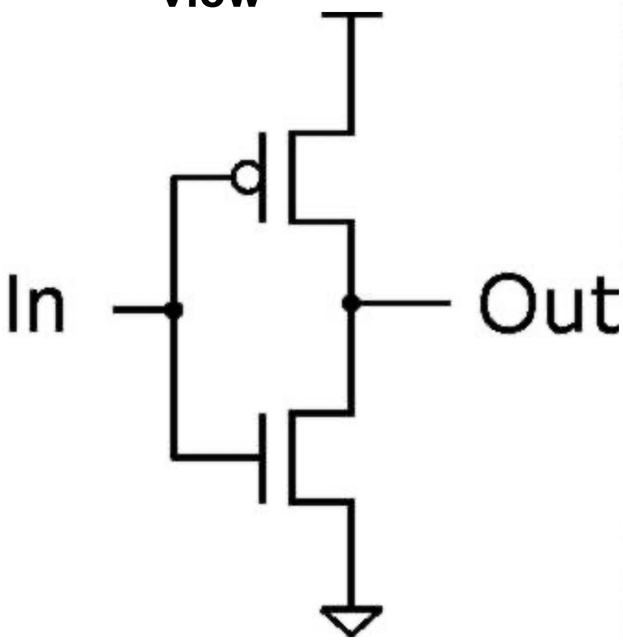
--When a PMOS device is needed, we create an n-type *WELL*

--PMOS devices then have their own n-type "body"

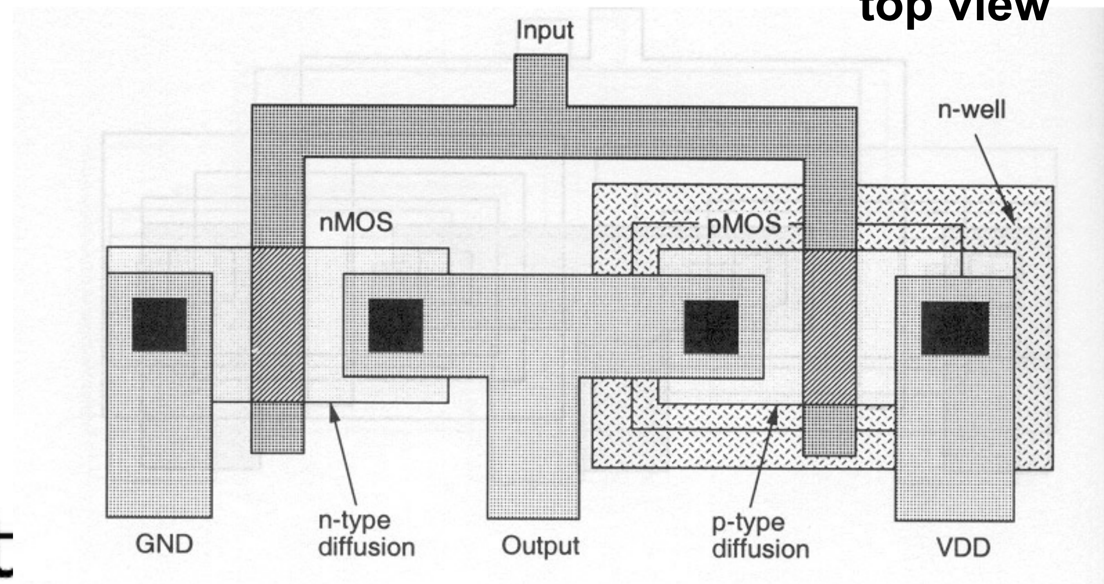
--Notice the body terminal on each transistor

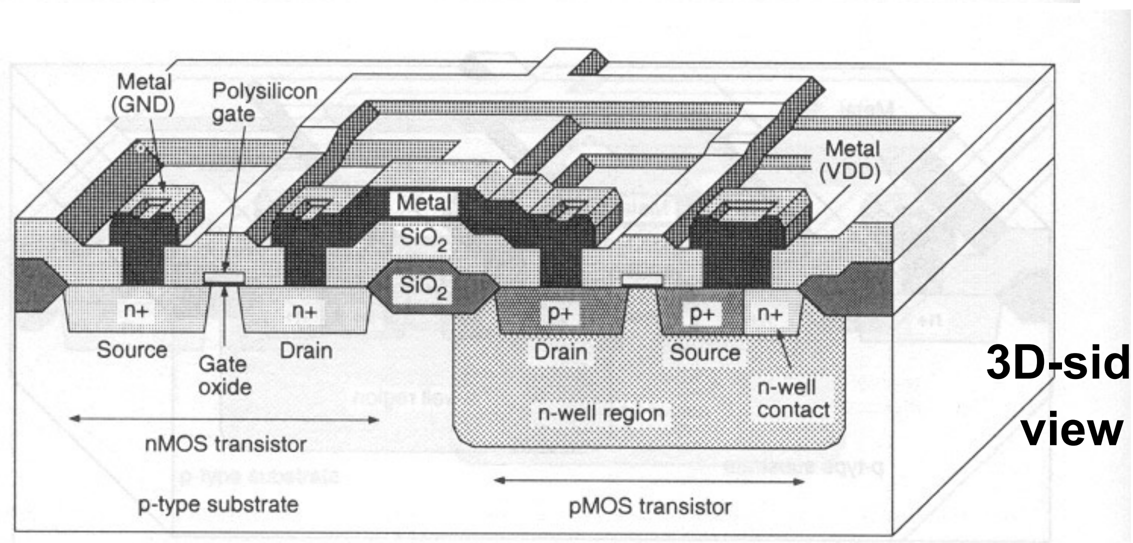
3-D View of CMOS Inverter in Silicon

schematic view



top view

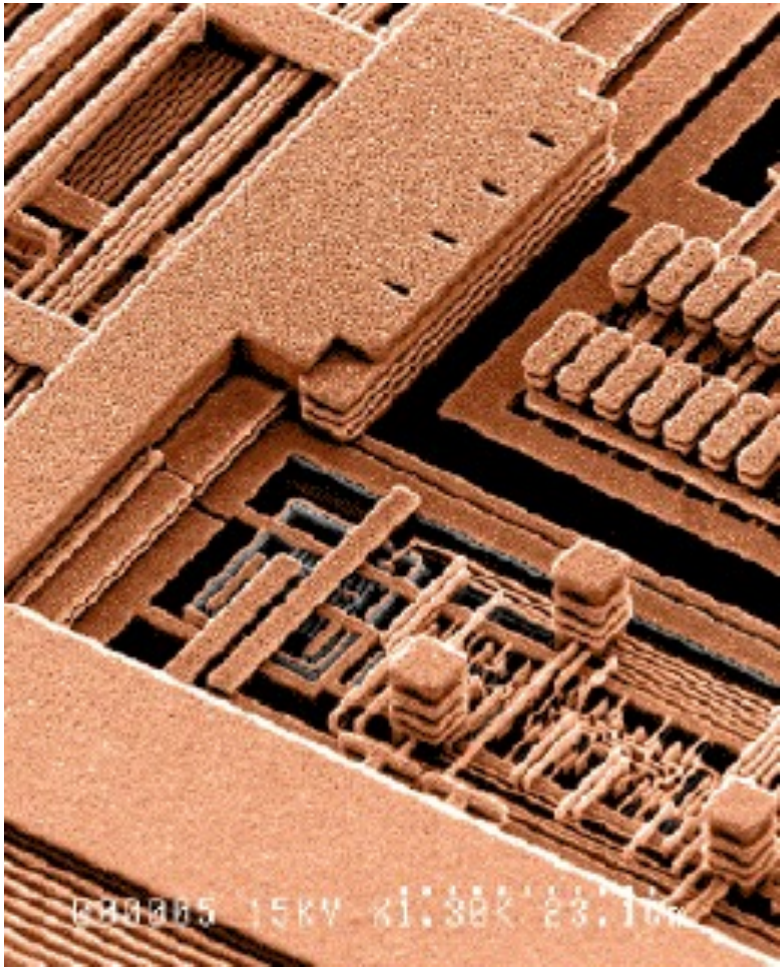




3D-side view

Note: we can make pMOS and nMOS transistor on the same piece of silicon

3-D of larger CMOS circuits



This is an SEM photo

shows all the metal Interconnections On an IC

pMOS/nMOS are at the very bottom

An idea of scale

The Intel Skylake chip uses transistors that are 60 times smaller than the wavelength of light.

Skylake transistor size 14nm. Wavelength of visible light 400-700nm

An idea of scale

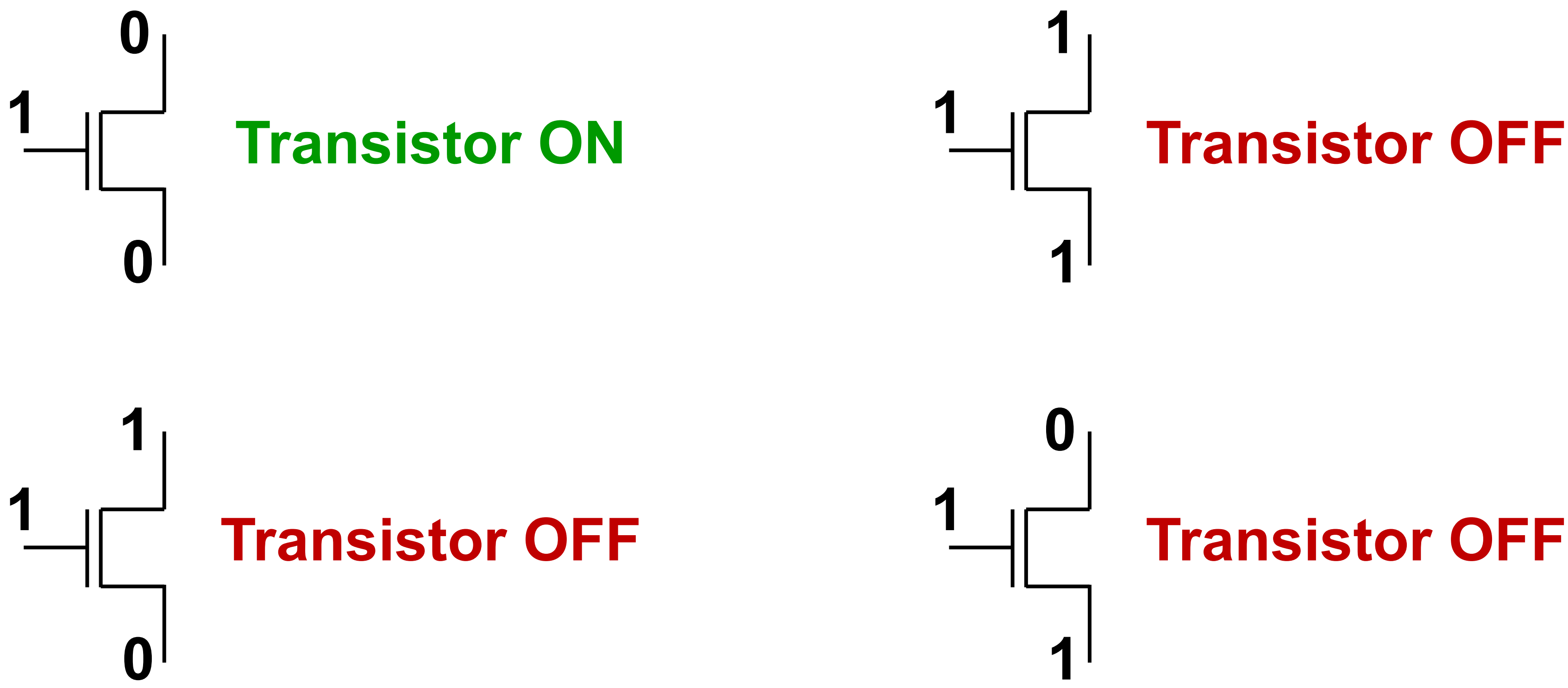
There are now more transistors at work in the world (15 quintillion : 15,000,000,000,000,000,000) than there are leaves on all the trees in the world

(The Perfectionists, Simon Winchester)

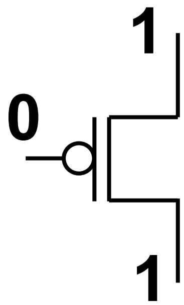
Limitations on CMOS transistors

- In a digital circuit there are only two relevant voltage levels, the low voltage level, GND (0), and the high voltage level, Vdd (1).
- An nmos transistor is only ON when the gate voltage is higher than the source and drain voltages
- A pmos transistor is only ON when the gate voltage is lower than the source and drain voltages
- As a consequence an nmos transistor can only be used to pass a low voltage and a pmos transistor can only be used to pass a high voltage

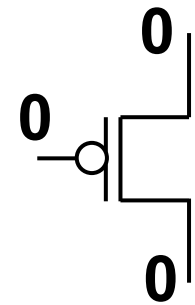
nMOS transistor is only on if the gate voltage is higher than the source and the drain voltages



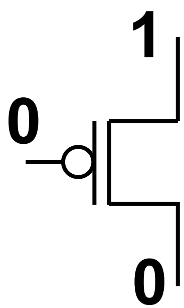
pMOS transistor is only on if the gate voltage is lower than the source and the drain voltages



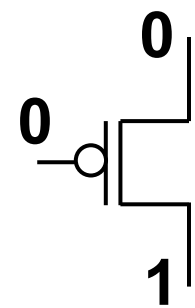
Transistor ON



Transistor OFF



Transistor OFF

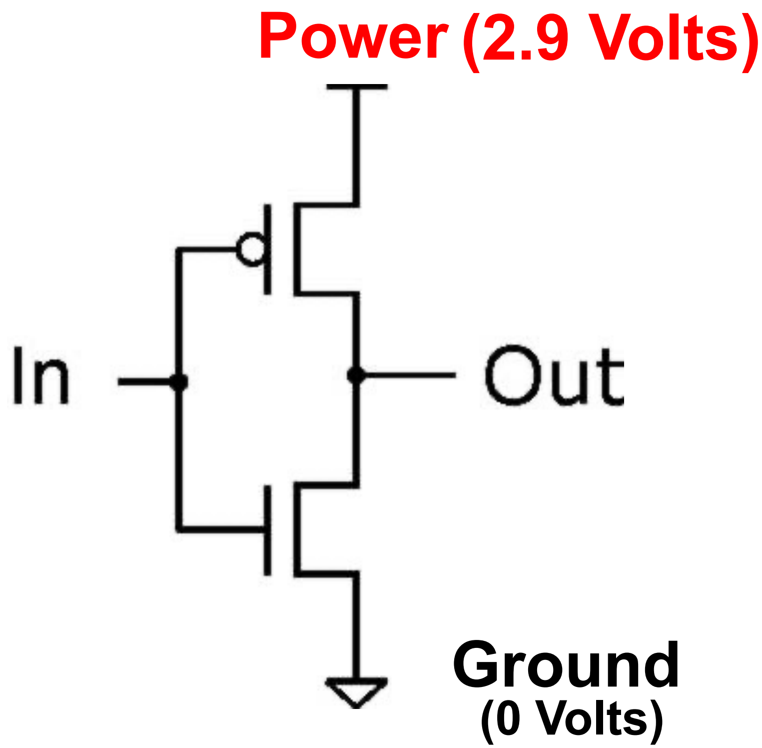


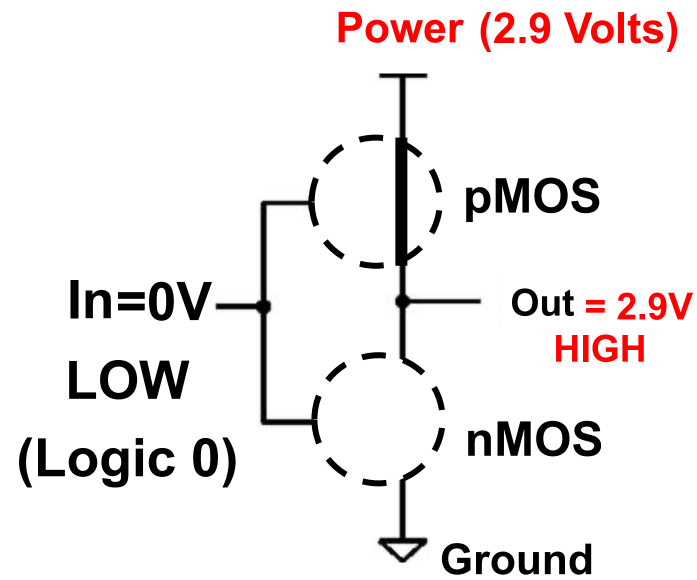
Transistor OFF

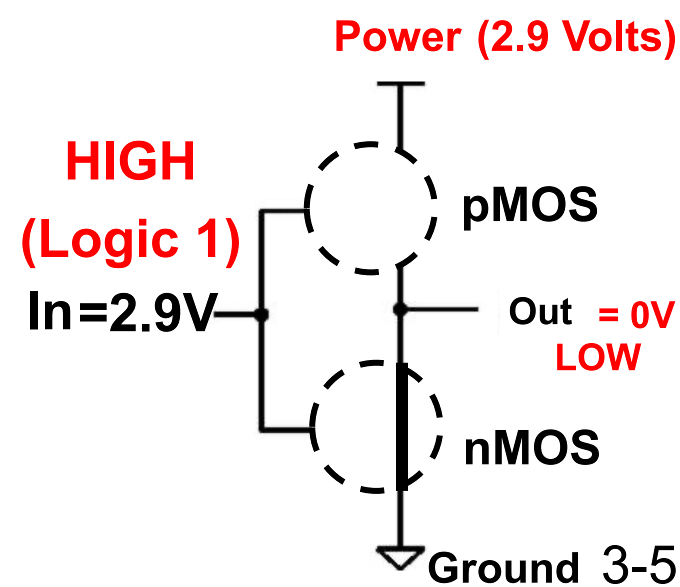
Bottom Line

- **nMos transistors can be used in switching circuits to connect circuit points to the low voltage level – GND (0). These are referred to as <u>Pull Down Networks</u>**
- **pMos transistors can be used in switching circuits to connect circuit points to the high voltage level – VDD (1). These are referred to as Pull Up Networks**

How do we configure transistors to make inverter?







We take advantage of opposing nature!

-Setting IN = 0V (LOW=Logic 0)

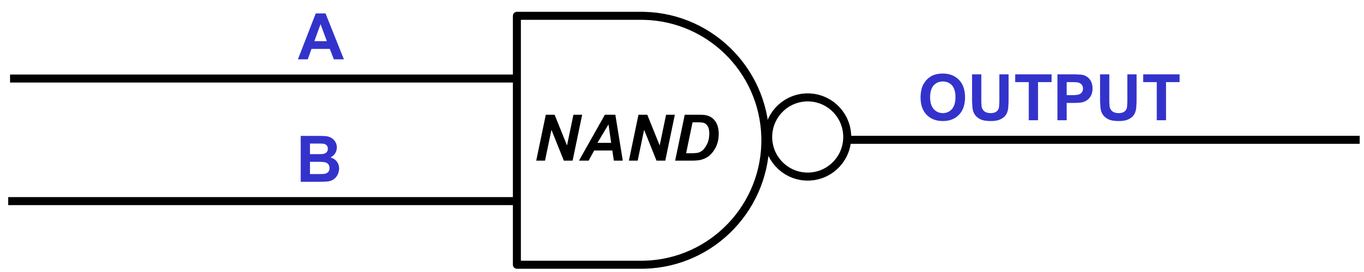
nMOS turns OFF

pMOS turns ON, connects OUT to Power

OUT=2.9V (HIGH=Logic 1 – inverse of In)

Our next "Gate"…the NAND gate

We introduce a 2-input (1-output) gate called: NAND



It is an AND gate with its output INVERTED

A	B	OUT
0 (LOW)	0 (LOW)	HIGH (1)
0 (LOW)	1 (HIGH)	HIGH (1)
1 (HIGH)	0 (LOW)	HIGH (1)
1 (HIGH)	1 (HIGH)	LOW (0)

Truth Table shows all possible combinations
Of inputs and effect on the output

Logic Function:

OUT=NOT (A AND B)

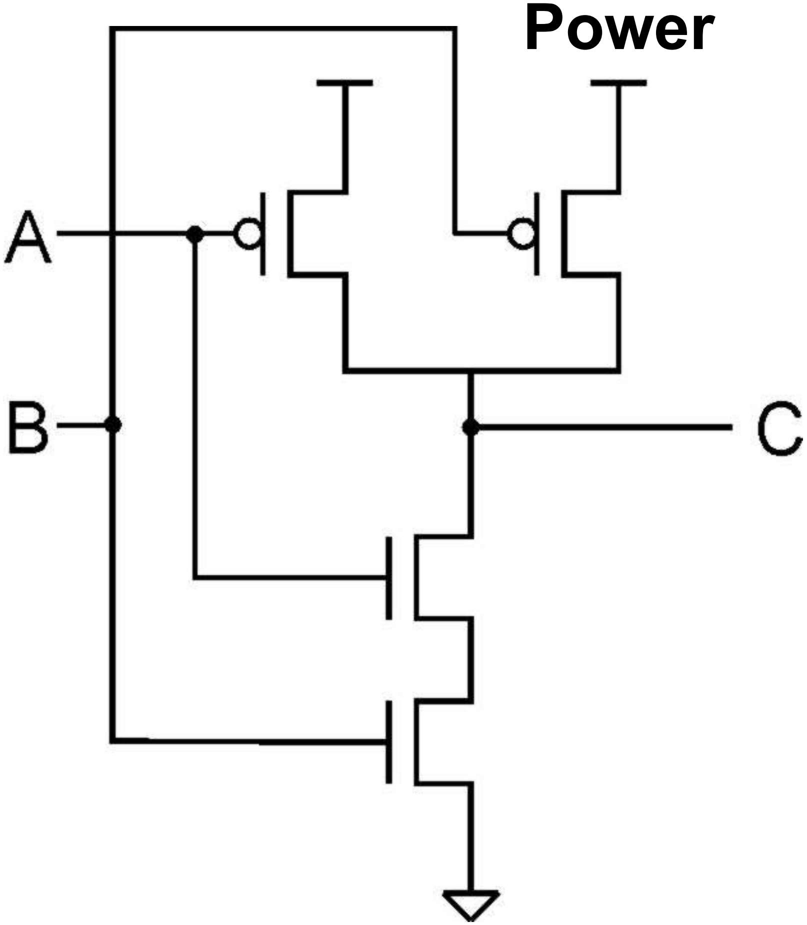
OUT= ¬(A ∧ B)

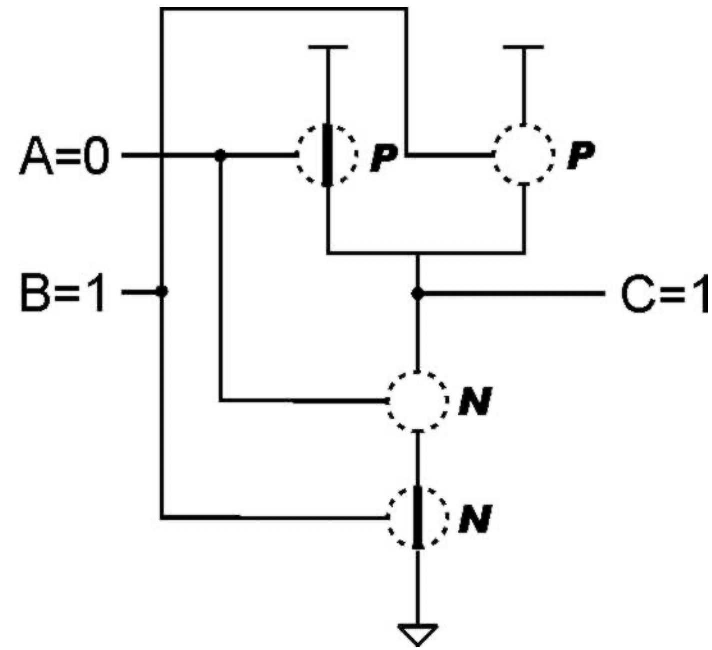
OUT=(AB)'

$OUT=\overline{AB}$

Different notation, but all identical
With function, we can calculate output

NAND Gate (NOT-AND)





A	B	C
0	0	1
0	1	1
1	0	1
1	1	0

Note: parallel structure on top, series on bottom.