COMP 202. Introduction to Electronics

Dr. N. B. Gyan

Central University, Miotso. Ghana

Vacuum Tubes/Thermionic Valves

credit: http://www.nutsvolts.com/magazine/article/vacuum_tube_in_its_100th_year

A Little History

The first vacuum tube / thermionic valve was developed when Ambrose Fleming used a discovery Edison had made that was called the Edison Effect.

Edison had not been able to find any applications for it, but Fleming used this two electrode diode it to rectify radio signals in a new form of radio detector he called his oscillation valve..

Later Lee de Forest added a third electrode to make a triode.

Further developments improved performance and added additional electrodes.

Technology

Vacuum tube or thermionic valve technology is based around the basic concept of thermionic emission.

The concept of thermionic valve or vacuum tubes used the idea that a heated element in a vacuum emitted electrons that would normally remain in the vicinity of this heated element because of the charge attraction.

If a second electrode was placed into the vacuum and a high positive potential placed on it, then the electrons would be attracted away from the heated element towards this element with a high potential. As a result a current would flow in this direction.

As electrons were unable to travel in the reverse direction, this simple valve or vacuum tube acts as a *diode*(a one-way switch for current).

Technology

Diodes are useful if you want to turn alternating (two-way) electric current into direct (one-way) current. Diodes can also be made so they give off light when electricity flows through them. You might have seen these light-emitting diodes (LEDs) on pocket calculators and electronic displays on hi-fi stereo equipment.

It is also possible to place a third element known as a grid into the structure between the structure between the other two electrodes. This electrode is normally formed of a gauze to allow electrons to pass through.

By varying the potential on this electrode, the flow of electrons can be controlled.

A Cathode Ray Tube is a special kind of a Vacuum Tube.

Welcome the Transistor

The transistor was invented at Bell Labs in New Jersey in 1947 by John Bardeen, Walter Brattain and William Shockley.

The transistor is at the heart of almost all electronics and so it is one of the most important inventions of the 20th century.

The second big step, the invention of the integrated circuit, took place simultaneously at Fairchild and Texas Instruments from 1957 to 1959. Jean Hoerni at Fairchild developed the planar transistor then Jack Kilby at Texas Instruments and Robert Noyce at Fairchild developed the integrated circuit.

Welcome the Transistor

This turned out to be the big breakthrough. Until that point transistors were built one at a time and wired together manually. The planar manufacturing process allowed multiple transistors to be created simultaneously and connected together simultaneously. By 1962 Fairchild was producing integrated circuits with about a dozen transistors.

Much has changed in the intervening years but this same basic principle is how we build today's chips with billions of transistors.

How is a Transistor Made?

Transistors are made from silicon, a chemical element found in sand, which does not normally conduct electricity (it doesn't allow electrons to flow through it easily).

Silicon is a *semiconductor*, which means it's neither really a conductor (something like a metal that lets electricity flow) nor an insulator (something like plastic that stops electricity flowing).

If we treat silicon with *impurities* (a process known as doping), we can make it behave in a different way.

How is a Transistor Made?

If we dope silicon with the chemical elements arsenic, phosphorus, or antimony, the silicon gains some extra ”free” electrons—ones that can carry an electric current—so electrons will flow out of it more naturally. Because electrons have a negative charge, silicon treated this way is called n-type (negative type).

We can also dope silicon with other impurities such as boron, gallium, and aluminium. Silicon treated this way has fewer of those ”free” electrons, so the electrons in nearby materials will tend to flow into it. We call this sort of silicon p-type (positive type).

Silicon Sandwiches

If the two different types of silicon are put together in layers, making sandwiches of p-type and n-type material, we can make different kinds of electronic components that work in all kinds of ways.

Suppose we join a piece of n-type silicon to a piece of p-type silicon and put electrical contacts on either side. Exciting and useful things start to happen at the junction between the two materials.

If we turn on the current, we can make electrons flow through the junction from the n-type side to the p-type side and out through the circuit.

Silicon Sandwiches

This happens because the lack of electrons on the p-type side of the junction pulls electrons over from the n-type side and vice-versa. But if we reverse the current, the electrons won't flow at all.

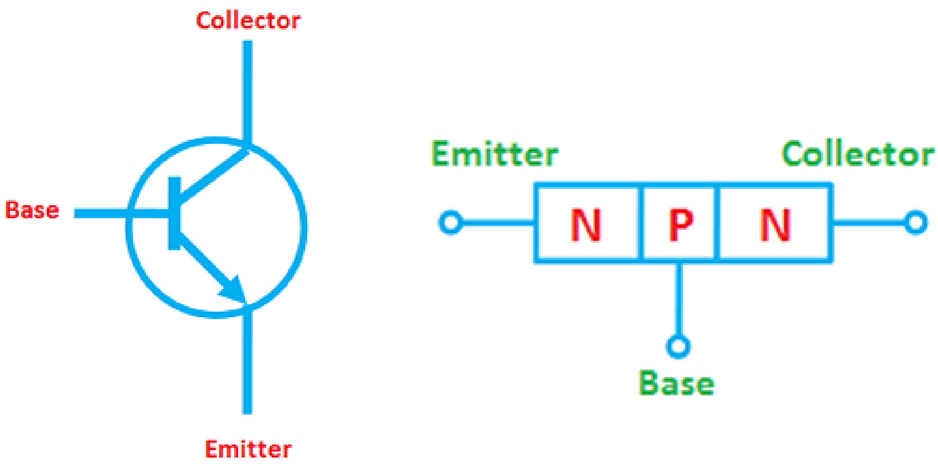
What you get here is called a diode (or rectifier) (remember?).

Junction Transistors

Now suppose we use three layers of silicon in our sandwich instead of two. We can either make a p-n-p sandwich (with a slice of n-type silicon as the filling between two slices of p-type) or an n-p-n sandwich (with the p-type in between the two slabs of n-type).

If we join electrical contacts to all three layers of the sandwich, we can make a component that will either amplify a current or switch it on or off—in other words, a transistor.

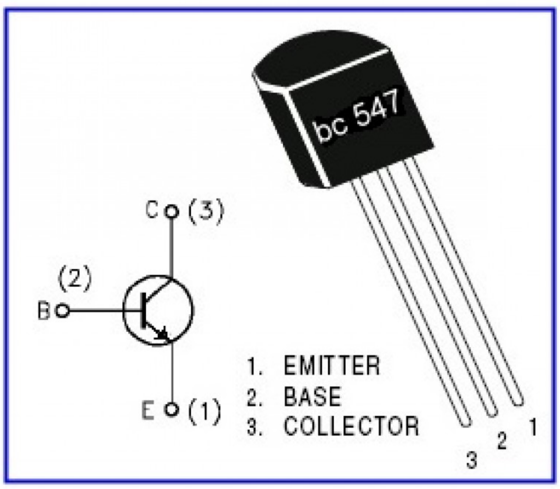
n-p-n Transistor



NPN transistor symbol

From Vacuum Tubes to Semiconductors
CMOS Working Principle and Applications

n-p-n Transistor



n-p-n Transistor

The two contacts joined to the two pieces of n-type silicon are the *emitter* and the *collector*, and the contact joined to the p-type silicon is called the *base*.

When no current is flowing in the transistor, we know the p-type silicon is short of electrons and the two pieces of n-type silicon have extra electrons.

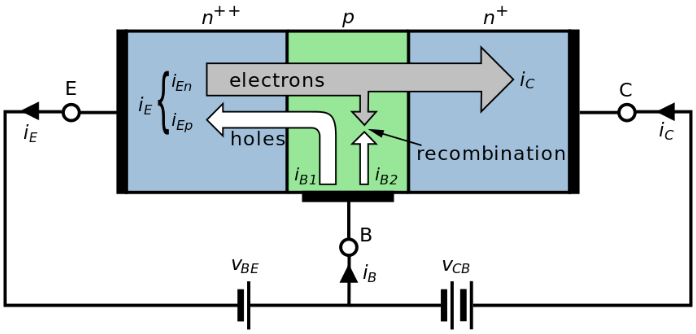
n-p-n Transistor

Another way of looking at this is to say that while the n-type has a surplus of electrons, the p-type has holes where electrons should be. Normally, the holes in the base act like a barrier, preventing any significant current flow from the emitter to the collector while the transistor is in its ”off” state.

A transistor works when the electrons and the holes start moving across the two junctions between the n-type and p-type silicon.

n-p-n Transistor

Suppose we attach a small positive voltage to the base, make the emitter negatively charged, and make the collector positively charged. Electrons are pulled from the emitter into the base—and then from the base into the collector. And the transistor switches to its ”on” state:



From Vacuum Tubes to Semiconductors
CMOS Working Principle and Applications

n-p-n Transistor

The small current that we turn on at the base makes a big current flow between the emitter and the collector. By turning a small input current into a large output current, the transistor acts like an *amplifier*.

But it also acts like a *switch* at the same time. When there is no current to the base, little or no current flows between the collector and the emitter. Turn on the base current and a big current flows.

So the base current switches the whole transistor on and off. Technically, this type of transistor is called *bipolar* because two different kinds (or ”polarities”) of electrical charge (negative electrons and positive holes) are involved in making the current flow.

Forward & Reverse Biases

We can also understand a transistor by thinking of it like a pair of diodes.

With the base positive and the emitter negative, the base-emitter junction is like a *forward-biased* diode, with electrons moving in one direction across the junction (from left to right in the diagram) and holes going the opposite way (from right to left).

The base-collector junction is like a *reverse-biased* diode. The positive voltage of the collector pulls most of the electrons through and into the outside circuit (though some electrons do recombine with holes in the base).

Field Effect Transistors

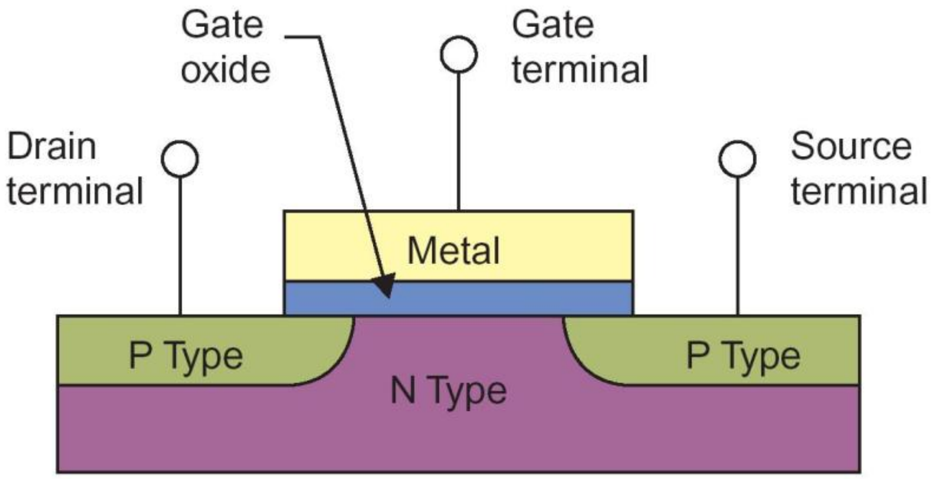
All transistors work by controlling the movement of electrons, but not all of them do it the same way.

Like a junction transistor, a FET (field effect transistor) has three different terminals—but they have the names source (analogous to the emitter), drain (analogous to the collector), and gate (analogous to the base).

In a FET, the layers of n-type and p-type silicon are arranged in a slightly different way and coated with layers of metal and oxide.

That gives us a device called a MOSFET (Metal Oxide Semiconductor Field Effect Transistor).

Field Effect Transistors



Field Effect Transistors

Although there are extra electrons in the n-type source and drain, they cannot flow from one to the other because of the holes in the p-type gate in between them.

However, if a positive voltage is attached to the gate, an electric field is created there that allows electrons to flow in a thin channel from the source to the drain. This“field effect” allows a current to flow and switches the transistor on.

For the sake of completeness, we could note that a MOSFET is a *unipolar* transistor because only one kind (“polarity”) of electric charge is involved in making it work.

CMOS Working Principle and Applications

One of the most popular MOSFET (metal-oxide-semiconductor field-effect transistor) technologies available today is the Complementary MOS or CMOS technology.

This is the dominant semiconductor technology for microprocessors, microcontroller chips, memories like RAM, ROM, EEPROM and application specific integrated circuits (ASICs).

The term CMOS stands for “Complementary Metal Oxide Semiconductor”.

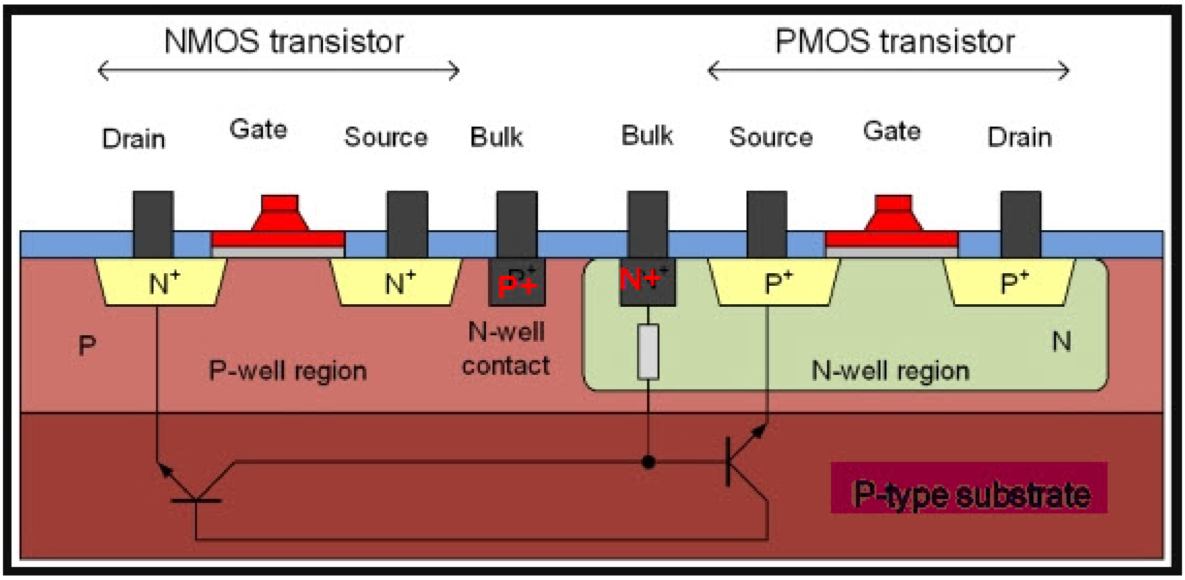
This technology makes use of both P channel and N channel semiconductor devices.

The main advantage of CMOS over other technologies (NMOS and BIPOLAR) is the much smaller power dissipation.

Unlike NMOS or BIPOLAR circuits, a Complementary MOS circuit has almost no static power dissipation. Power is only dissipated in case the circuit actually switches.

This allows integrating more CMOS gates on an IC than in NMOS or bipolar technology, resulting in much better performance. Complementary Metal Oxide Semiconductor transistor consists P-channel MOS (PMOS) and N-channel MOS (NMOS).

CMOS

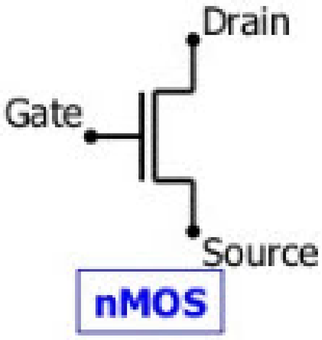


NMOS

NMOS is built on a p-type substrate with n-type source and drain diffused on it.

In NMOS, the majority carriers are electrons. When a high voltage is applied to the gate, the NMOS will conduct. Similarly, when a low voltage is applied to the gate, NMOS will not conduct.

NMOS are considered to be faster than PMOS, since the carriers in NMOS, which are electrons, travel twice as fast as the holes.

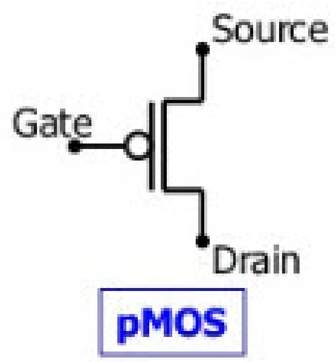


PMOS

P-channel MOSFET consists P-type Source and Drain diffused on an N-type substrate.

Majority carriers are holes. When a high voltage is applied to the gate, the PMOS will not conduct. When a low voltage is applied to the gate, the PMOS will conduct.

The PMOS devices are more immune to noise than NMOS devices.



CMOS Working Principle

In CMOS technology, both N-type and P-type transistors are used to design logic functions.

The same signal which turns ON a transistor of one type is used to turn OFF a transistor of the other type.

This characteristic allows the design of logic devices using only simple switches, without the need for a pull-up resistor.

CMOS Working Principle

In CMOS logic gates a collection of n-type MOSFETs is arranged in a pull-down network between the output and the low voltage power supply rail (Vss or quite often ground).

Instead of the load resistor of NMOS logic gates, CMOS logic gates have a collection of p-type MOSFETs in a pull-up network between the output and the higher-voltage rail (often named Vdd).

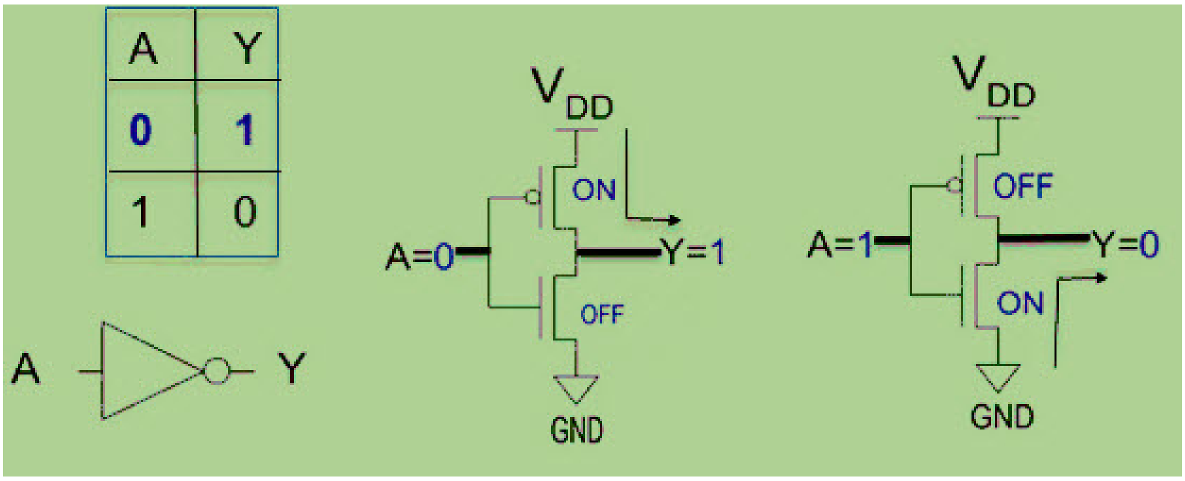
Thus, if both a p-type and n-type transistor have their gates connected to the same input, the p-type MOSFET will be ON when the n-type MOSFET is OFF, and vice-versa.

CMOS Working Principle

The networks are arranged such that one is ON and the other OFF for any input pattern as shown in the figure below.

CMOS offers relatively high speed, low power dissipation, high noise margins in both states, and will operate over a wide range of source and input voltages (provided the source voltage is fixed).

CMOS Inverter



CMOS Inverter

The NMOS transistor has an input from V_{ss} (ground) and PMOS transistor has an input from V_{dd}.

The terminal Y is output.

When a high voltage ($\approx V_{dd}$) is given at input terminal (A) of the inverter, the PMOS becomes open circuit and NMOS switched OFF so that the output will be pulled down to V_{ss}.

Video

Building Logic Gates from MOSFET Transistors