2.0. Introduction to integrated circuit devices

2.1. Introduction

In this chapter the basics of the devices that make up integrated circuits will be presented. The objective of this chapter is to give the reader enough of an understanding of the operation of devices fabricated on integrated circuits (ICs) to be able to understand the process trade-offs and challenges discussed later in the book. For readers who already have a good working familiarity with device physics, this chapter may be skipped.

2.2. Basic electronic concepts

Electronic circuits regulate and control the flow of electric current. Electric current is the flow of electrons, the tiny subatomic particles that surround the nucleus of atoms. Electrons carry a fixed negative electric charge and the movement of electrons carries charge from one location to another - the flow of electrons is referred to as electric current. Electric current is driven by a difference in potential from one location to another location measured in volts. Electric current flows easily through materials that are conductors, and is blocked by materials that are insulators. The amount of resistance that a material presents to the flow of electric current is logically called resistance. Conductors have low resistance to the flow of electric current and insulators have extremely high resistance (essentially infinite until the voltage is so high that the material breaks down). For a given voltage, the higher the resistance the lower the current that will flow and the lower the resistance the higher the current that will flow. Conversely, for a given resistance, the higher the voltage the higher the current that will flow and the lower the voltage the lower the current that will flow.

2.3. Electronic circuit elements

Electronic circuits are made up of a number of elements used to control current flow. There are a wide variety of different circuit elements but for the purpose of this discussion the circuit elements will be restricted to the four most commonly used in ICs; resistors, capacitors, diodes and transistors. Resistors, provide a fixed amount of resistance to current flow. Capacitors, store electronic charge until discharged, somewhat similar to a battery. Diodes, allow current to flow in one direction but not in the opposite direction, a one way valve. Transistors, provides two major modes of action, one, a switch turning current flow on and off, or two, act as an amplifier whereby an input current produces a larger output current.

An IC is nothing more than a number of these components connected together as a circuit all formed on the same substrate.

2.4. Atomic structure and band theory

Atoms are made up of positively charged - protons, neutral - neutrons, and negatively charged electrons. Protons and neutrons are relatively heavy particles with similar masses and electrons are relatively light particles with much lower mass. The nucleus of an atom is made up of protons and neutrons with electrons arrayed around it. The electrons occupy specific allowed energy levels with each level accommodating two electrons. Outer levels may also contain sub levels each capable of holding two electrons.

For silicon, the element of most interest to this publication, the outermost energy level has 3 sub levels with a total capacity of 6 electrons - two per sub

Basic concepts

- Current - the flow of electrons carrying electric charge.
- Voltage - the force driving the flow of current.
- Resistance - a materials resistance to the flow of electric current.
- Conductor - a material that readily supports the flow of electric current.
- Insulator - a material that blocks the flow of electric current.

Circuit elements

- Resistors - resists current flow.
- Capacitors - stores charge.
- Diodes - allows current to flow in only one direction.
- Transistor - switches and or amplifies current.

Sub-atomic particles

- Proton - relatively heavy, positively charge particle present in the nucleus.
- Neutron - relatively heavy, neutral particle present in the nucleus.
- Electron - relatively light, negatively charged particle

level. The actual number of electrons occupying the outer most levels is 2, so there are 4 unfilled levels. In solid silicon, the outer most levels are filled by the four nearest silicon neighbor atoms sharing electrons, see figure 7.

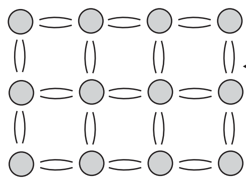

Figure 7. Silicon covalent bonds

When atoms exist in a solid, the proximity of electrons on neighboring atoms leads to a further splitting of allowed energy levels. The outer most energy levels split into two bands of closely spaced energy levels known as the conduction and the valence bands. The two bands are separated from each other by a gap in the allowed energy levels. The width of the energy gap determines the electrical conduction of the material. A large gap of several electron volts (eV) results in an insulating material, a gap of a few tenths of an eV to a few eV is a semiconductor, and a material with overlapping bands is a conductor.

The reason that the energy gap determines the conduction properties of the material is as follows. At zero degrees kelvin (minus 273 centigrade), the valence band is completely filled with electrons and the conduction band is empty. At room temperature the electrons in the material have gained thermal energy. If the energy gap is large, very few if any electrons will have sufficient thermal energy to jump to the conduction band where the electrons are free to move and participate in conduction. If the gap is small or non existent, many electrons will jump to the conduction band and be free to move.

2.5. Intrinsic semiconductors

Materials with energy gaps of a few tenths of an eV to a few eV at room temperature will contain electrons that have jumped from the valence band to the conduction band due to thermal energy. The electrons in the conduction band will be free to move about in the material participating in conduction. In addition, when the electrons jump from the valence band they leave behind a missing electron or hole in the valence band. Electrons from neighboring atoms may move into the hole left behind by the escaping electron. The hole will then effectively have moved to a neighboring atom where it may move once again by another electron from a neighboring atom filling it. Since atoms with an associated hole in the valence band have one more proton than electron, the hole has an effective positive charge equal and opposite to the charge on an electron (atoms without holes have equal numbers of protons and electrons, i.e., they are electrically neutral). Each electron thermally excited into the conduction band results in a mobile negative electron in the conduction band and a mobile positive hole in the valence band. Although holes are less mobile than electrons, they may be thought of as positively charged particles free to move in their own right. Thermally generated holes and electrons are called intrinsic carriers and materials with significant numbers of thermally generated carriers at room temperature are called intrinsic semiconductors.

Energy gap effect on conduction

- Insulators - energy gaps of several eV.
- Semiconductors - energy gaps of a few tenths of an eV to a few eV.
- Conductors - overlapping energy bands with no energy gap.

Intrinsic semiconductors

- Holes - missing electrons in the valence band that “move” by capturing electrons from neighboring atoms thereby switching the hole to the neighboring atom.
- Intrinsic carriers - electron-hole pairs created by thermal energy.
- Intrinsic semiconductor - a material with a significant number of intrinsic carriers at room temperature.

2.6. Extrinsic semiconductors

Free electrons and holes may be created in a semiconductor material by introducing foreign elements known as dopants. Silicon is a group IV element in the periodic table. Group IV elements each have 4 electrons in an outer most energy level and are all semiconductors. Group III elements each have 3 electrons in the outer most energy level and group V elements each have 5 electrons in the outer most energy level. If a group III element is introduced into silicon, the group III element can bond with the four nearest silicon atoms leaving one missing electron or hole. If a group V element is introduced to silicon it can bond with the four nearest silicon atoms introducing an extra electron. Figure 8 illustrates a section of the periodic table containing group III, IV and V elements of interest.

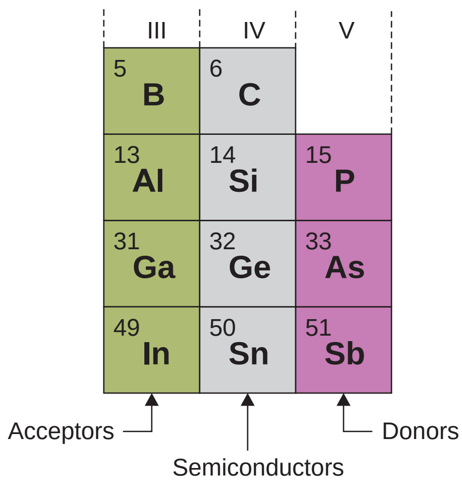

Figure 8. Periodic table section.

Figure 9a illustrates a group III element boron, introduced into silicon. Boron introduces an unfilled energy level into the energy gap near the valence band that can “accept” an electron creating a hole. Figure 9b illustrates a group V element, phosphorus introduced into silicon. Phosphorous introduces an electron into the energy gap near the conduction band. The electron in the energy gap may easily be “donated” to the conduction band creating a free electron.

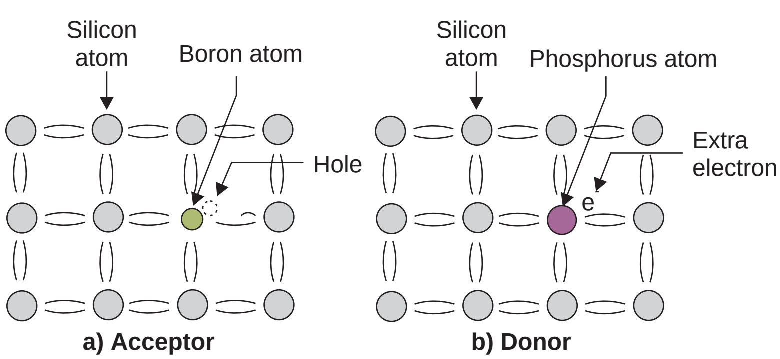

Figure 9. Dopants in silicon.

Group III dopants that accept electrons creating a hole are called p-type dopants for positive dopant. Group V elements that donate an electron are called

Extrinsic semiconductors

- dopant - an impurity that introduces holes or electrons to a semiconductor.
- p-type dopant - a dopant that accepts electrons creating holes, also called an acceptor.
- p-type - silicon doped to have more free holes than electrons.
- n-type dopant - a dopant that donates electrons, also called a donor.
- n-type - silicon doped to have more free electrons than holes.
- Majority carrier - the carrier type in the majority, holes for p-type silicon and electrons for n-type silicon.
- Minority carrier - the carrier type in the minority - electrons for p-type silicon and holes for n-type silicon.

n-type dopants. Throughout this text, p-type dopants will be indicated by green and n-type dopants will be indicated by purple.

If a sufficient number of dopants are introduced into a semiconductor to create more carriers of one type than the intrinsic carrier concentration, the material is said to be an extrinsic semiconductor. Extrinsic semiconductors have more of one type of carrier than the opposite type of carrier. Materials doped with p-type dopants are p-type semiconductors and have more holes than free electrons. Materials doped with n-type dopants are said to be n-type semiconductors and have more free electrons than holes. The carrier type, hole or electron in the majority is called the majority carrier and the type of carrier in the minority is called the minority carrier.

2.7. PN junction

When p-type silicon and n-type silicon are in contact, a PN junction results. Initially electrons flow from the n-type side of the junction to the p-type side of the junction and holes flow from the p-type side of the junction to the n-type side of the junction (opposite charges attract). Eventually the flow of carriers creates a voltage across the junction that prevents further current flow. If a voltage is applied across the pn junction with the positive terminal attached to the p-type side and the negative terminal attached to the n-type side and the voltage is higher than the built in voltage, the junction will turn “on” and electric current will flow. This is referred to as a forward biased junction, see figure 10 top. If a voltage is applied with the positive terminal attached to the n-type side and the negative terminal attached to the p-type side, the negative terminal will attract holes away from the junction and the positive terminal will attract electrons away from the junction resulting in an area at the junction depleted of mobile carriers called the depletion region. As the applied voltage is increased the depletion region will grow and no current will flow until the junction “breaks down” and current begins to flow. This is referred to as a reverse biased junction, see figure 10 bottom.

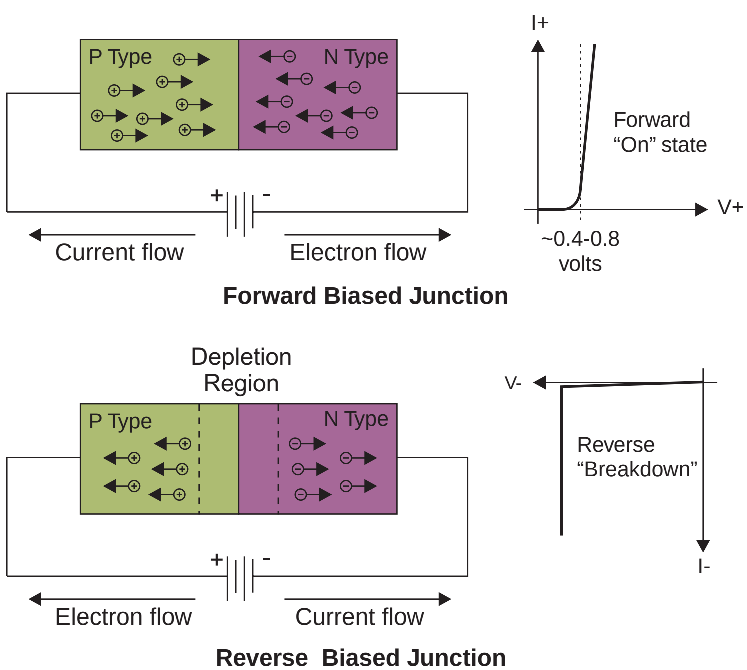

Figure 10. PN junction behavior.

PN junction

- Forward biased junction - a pn junction with a positive voltage applied to the p-type side of the junction and a negative voltage applied to the n-type side of the junction. The junction turns on if the applied voltage is greater than the built in voltage.
- Reverse bias - a pn junction with a negative voltage applied to the p-type side of the junction and a positive voltage applied to the n-type side of the junction.
- Depletion region - a region “depleted” of mobile carriers.
- Breakdown - when a sufficiently high reverse bias is applied to a junction the junction will breakdown and conduct current.
- Impact ionization - carriers with sufficient energy impact atoms in the semiconductor exciting electrons into the conduction band creating electron-hole pairs.
- Avalanche breakdown - a kind of runaway condition where impact ionization creates electron-hole pairs that create new electron-hole pairs by further impact ionization.

The net result of the differing behavior for forward and reverse biased pn junction, is that the junction acts like a one way valve only allowing current flow in the forward direction. Diodes are frequently used in electric circuits to “steer” the current flow.

Although it may not be obvious, pn junction breakdown is a very important phenomena in ICs. The maximum voltage an IC can handle may be determined by junction breakdown so it is important to understand what determines the breakdown voltage of a pn junction.

Breakdown occurs when electrons and or holes in the semiconductor obtain enough energy to create new electron-hole pairs by colliding with atoms in the semiconductor exciting electrons into the conduction band, this is referred to as impact ionization. At a high enough applied voltage the electron-hole pairs created by impact ionization will have enough energy to create new electron -hole pairs creating a cascading runaway of electron-hole creation in the semiconductor. The carriers (electrons and holes) created by impact ionization will flood the depletion region collapsing it and causing a rapidly increasing current to flow, this is referred to as avalanche breakdown.

Most pn junctions have unequal doping on the p-type and n-type side. When a pn junction is reverses biased, a depletion region forms to oppose the applied voltage. The width of the depletion region depends on the applied voltage and the doping concentration in the semiconductor. The following rules determine the width of the depletion region:

1. The higher the applied voltage the thicker the depletion region will be until breakdown is reached.
2. For a given applied voltage the lower the doping concentration the wider the depletion region will be.
3. For a pn junction the region with the lowest doping level will have the widest depletion region, the highest energy carriers and will breakdown first leading to avalanche for the junction.

2.8. Bipolar transistors

A bipolar transistor is a three layer device with two pn junctions and three terminals- see figure 11. In figure 11 the transistor on the top is comprised of an n-type layer sandwiched between two p-type layers - a PNP, and the bottom transistor is comprised of a p-type layer sandwiched between two n-type layers - an NPN.

On the left side of the top of figure 11 is a p-type area referred to as the emitter, in the center is an n-type area referred to as the base and on the right side is a p-type area referred to as the collector. There is an electrical connection made to each of the three areas, emitter, collector and base referred to as terminals. For reasons that will be discussed shortly, the emitter is heavily doped p-type, the base is lightly doped n-type and of relatively narrow width and the collector is relatively lightly doped p-type. This type of bipolar transistor is called a PNP because an n-type layer is sandwiched between to p-type layers. There is also NPN transistors where a p-type layer is sandwiched between two n-type layers - see the bottom of figure 11.

Bipolar Transistor

- PNP - a bipolar transistor where an n-type layer is sandwiched between two p-type layers.
- NPN - a bipolar transistor where a p-type layer is sandwiched between two n-type layers.
- Emitter - a heavily doped layer in a bipolar transistor that “emits” carriers into the base.
- Collector - a layer in a bipolar transistor that “collects” the carriers “emitted” by the emitter.
- Base - a thin - lightly doped layer that modulates the flow of carriers between the emitter and base of a bipolar transistor.

Recombination

- Recombination - when a free electron is captured by a hole annihilating the electron-hole pair.

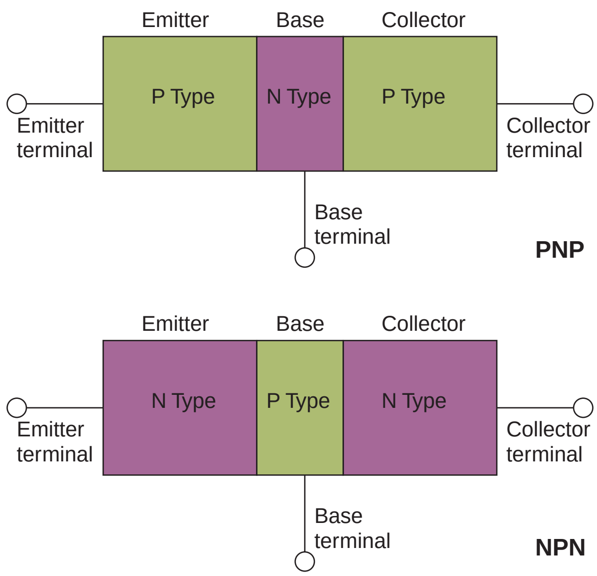

Figure 11. PNP Bipolar transistor.

In the preceding section it was shown that a forward biased pn junction would conduct current.

If a positive voltage is applied to the emitter and a negative voltage is applied to the base of the PNP illustrated in figure 11 - see figure 12, then holes will be injected from the p-type emitter into the n-type base and electrons will be injected from the n-type base into the p-type emitter and current will flow. Under these conditions the emitter-base junction is forward biased. If the emitter is heavily doped relative to the base, under forward bias conditions the emitter will inject many more holes into the base than the base will inject into the emitter.

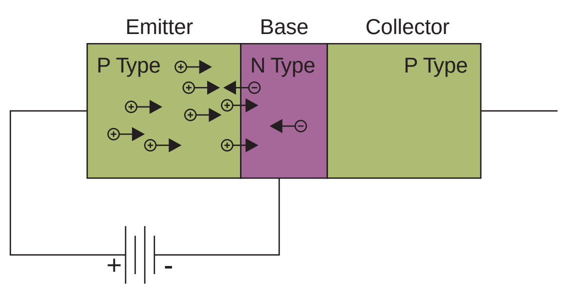

Figure 12. Forward biased emitter-base junction.

If a positive voltage is applied to the base and a negative voltage is applied to the collector, the collector base junction is revere biased and no current will flow - see figure 13.

If the forward biased emitter-base junction exists in the same transistor with the reverse biased collector base junction, then the behavior of the transistor becomes really interesting.

If the base of the bipolar transistor is relatively narrow and lightly doped, then holes injected into the base from the emitter may transit the base without recombing with electrons in the base and reach the collector-base junction. At the collector-base junction the holes are "collected" by the collector and current flows from emitter through the base to the collector. The current flowing from emitter terminal to collector terminal may be controlled by the current at the base terminal and the "gain" of the transistor is given by:

$$gain = \frac{I_C}{I_B} \tag{2}$$

where, I_C is the collector current and I_B is the base current.

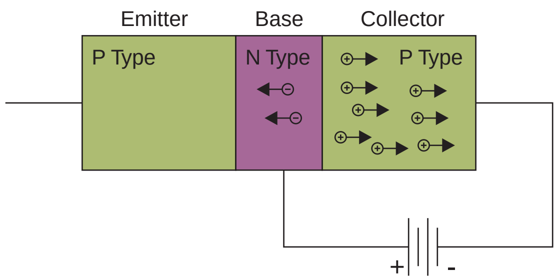

Figure 13. Reverse biased collector-base junction.

The configuration illustrated in figure 14 shows how a small signal connected to a base terminal may be used to modulate the current flowing from emitter to base. If a transistor is properly biased a small signal at the base terminal will be amplified at the collector terminal.

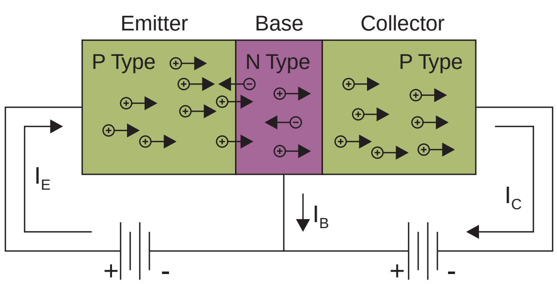

Figure 14. Bipolar transistor conducting.

A bipolar transistor is called bipolar because both carrier types, electrons and holes participate in conduction. In the section 2.10 a unipolar device the MOSFET will be presented. A MOSFET is a device in which only majority carriers participate in conduction.

2.9. MOS Capacitor

The MOS capacitor is the simplest MOS structure. A basic capacitor is formed when two conductive layers are separated by an insulating dielectric layer. If a voltage is applied across a capacitor, opposite charges accumulate on the two conductive layers and are stored there because the insulating layer prevents current flow from conductive layer to conductive layer. Eventually capacitors lose their stored charge by leakage through the dielectric, but for some period of time the capacitor will store charge that can later be released from the capacitor. In this case a capacitor is acting like a battery. The measure of charge a capacitor can store is the capacitance and is given by:

Bipolar Transistor 2

- Gain - the ratio of the collector current to the base current. Gain is often referred to as Beta.
- Bipolar - a device in which both carrier types, holes and electrons take part in conduction.

$$C = \frac{kA}{d} \qquad (3)$$

where, C is the capacitance, k is the dielectric constant, A is the area of the capacitor conductors and d is the thickness of the dielectric layer. The k value of a dielectric is an intrinsic property of the dielectric. From equation 2 it can be seen that capacitance increases with increasing dielectric constant k, increasing area A, or decreasing dielectric thickness d.

For a "standard" capacitor with two conductive plates capacitance is a fixed value versus voltage, however, for an MOS capacitor capacitance varies with voltage. Figure 15 illustrates an MOS capacitor under different voltages and polarities.

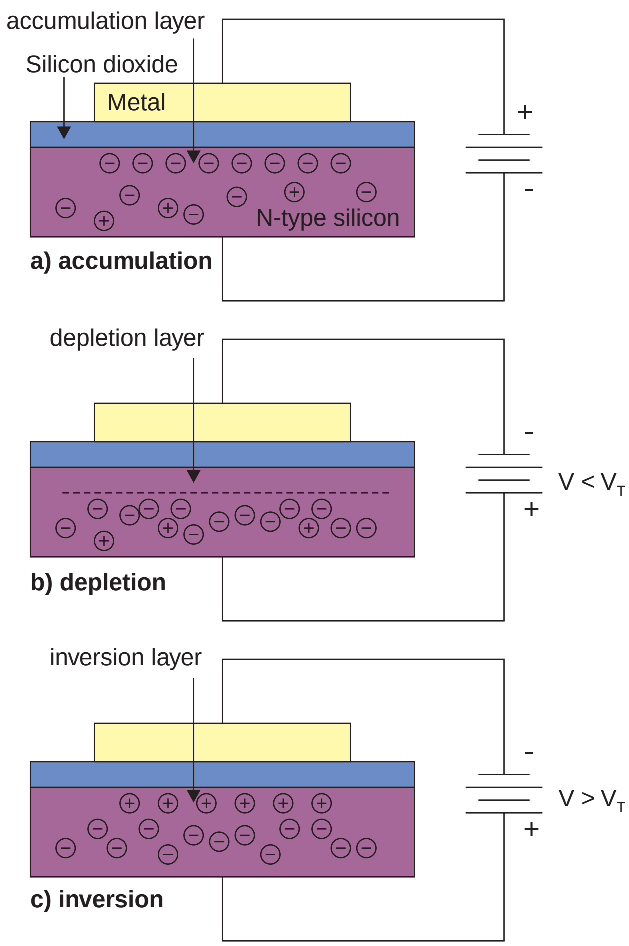

Figure 15. MOS capacitor.

In figure 15, the MOS capacitor consists of a metal top conductor, a silicon dioxide dielectric and an n-type silicon bottom "conductor". When a voltage is applied with the positive terminal attached to the top conductor - figure 15a, the majority n-type carriers are attracted to the silicon-silicon dioxide interface (unlike charges attract). Under these conditions, the positive charges on the metal conductor and the negative charges in the silicon are separated by the minimum distance and from equation 2 this condition leads to the maximum capacitance determined by the silicon dioxide thickness, dielectric constant and the area of the top metal conductor.

MOS Capacitors

- Capacitance - the capacity of a structure to store electric charge.
- Accumulation - an increase in concentration of carriers in a region of a semiconductor brought about by an applied electric field.
- Depletion - a reduction in the concentration of free carriers in a region of a semiconductor brought about by an applied electric field.
- Depletion region or layer - a region of a semiconductor in which no free carriers are present.
- Inversion - when a region of a semiconductor switches type from the majority carrier type to the minority carrier type due to accumulation of minority carriers brought about by an applied electric field.

If a relatively low voltage is applied to the capacitor with the negative terminal attached to the top conductor - figure 15b, the negative carriers in the silicon are repelled by the negative charge on the conductor plate (like charges repel). The majority n-type carriers are repelled from the silicon-silicon dioxide interface leaving behind a layer depleted of charges - a depletion layer forms. The depletion layers acts like a dielectric (insulator) due to the lack of availability of charges to conduct current. The charge on the top metal conductor and the charge in the silicon are now separated by a greater distance than just the thickness of the oxide. Since capacitance goes down as the thickness of the dielectric increases (in this case the effective thickness, not the physical thickness) the capacitance decreases.

If a relatively higher voltage is applied to the capacitor with the negative terminal attached to the upper metal conductor - figure 15c, the few p-type minority carriers that exist in n-type silicon will be attracted to the silicon-silicon dioxide interface while the n-type carriers are repelled. With sufficient voltage and time, the surface of the n-type silicon will “invert” to p-type. Since the carriers are now once again separated only by the thickness of the dielectric layer, the capacitance will again reach a maximum value for the thickness of dielectric, dielectric constant and area of the capacitor. Figure 16 illustrates capacitance versus voltage for the capacitor illustrated in figure 15.

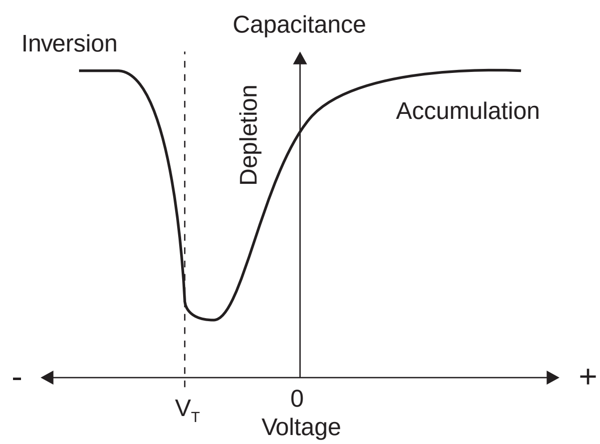

Figure 16. Capacitance voltage plot.

A key concept from figure 16 is the threshold voltage indicated by V_T on the voltage axis. The threshold voltage is the voltage required to invert the surface of the semiconductor. Threshold voltage is effected by the type of metal used for the conductive top plat, the dielectric thickness and dielectric constant and the amount of doping and therefore carrier concentration in the silicon. Foe a common system of aluminum - silicon dioxide - silicon, the thinner the silicon dioxide layer and the lower the silicon doping the lower the threshold voltage. Conversely, thicker silicon dioxide layers with heavier doping levels have higher threshold voltages. Threshold voltage is a critical concept for other MOS devices that will be discussed in the next section.

The entire preceding discussion has centered around n-type silicon. P-type silicon exhibits the same behavior except that p-type silicon accumulates when a negative voltage is applied to the top metal conductor and inverts when a positive voltage is applied to the top conductor, the exact opposite of n-type silicon behavior versus polarity.

Capacitance voltage plot

- Capacitance voltage plot - a plot of capacitance versus voltage for an MOS capacitor. Also called a CV plot.
- Threshold voltage - the voltage at which inversion begins.

2.10. The MOSFET

The Metal Oxide Semiconductor Field Effect Transistor or MOSFET is the basic device building block underlying the vast majority of ICs produced in the world today - see figure 17.

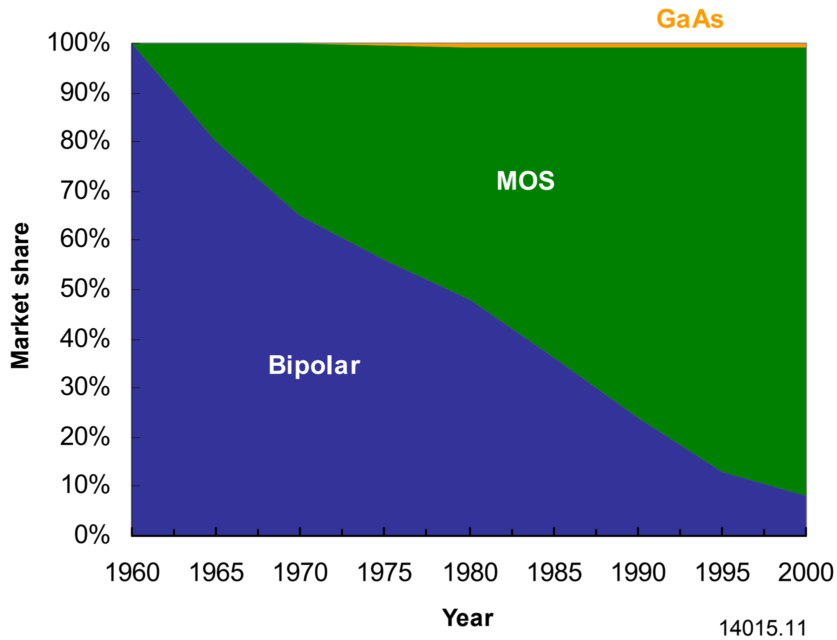

Figure 17. Relative market share of different technologies.

In a MOSFET, an MOS capacitor is used to apply an electric field that controls the transistor switching. The structure of a MOSFET comes in two basic types, NMOS and PMOS - see figure 18.

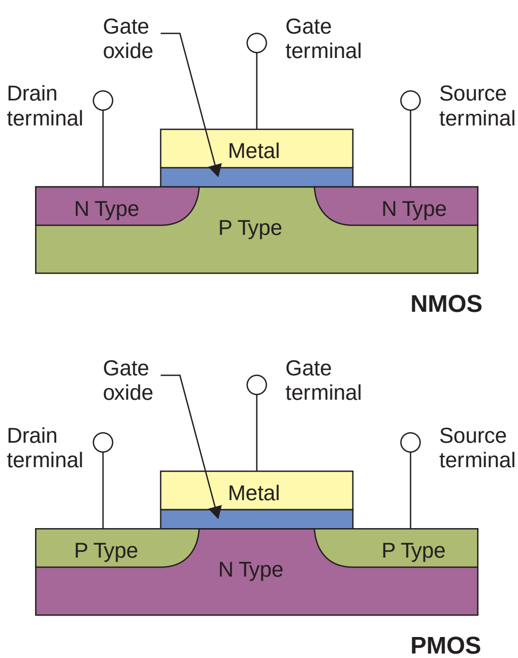

Figure 18. NMOS and PMOS MOSFETs.

The MOSFET

- MOSFET - Metal Oxide Semiconductor Field Effect Transistor
- NMOS - a MOSFET that uses n-type carriers for conduction. NMOS is made up of an n-type source region and an n-type drain region separated by a p-type region.
- PMOS - a MOSFET that uses p-type carriers for conduction. PMOS is made up of a p-type source region and a p-type drain region separated by an

In an NMOS device - figure 18 top, two highly doped n-type areas are separated by a lightly doped p-type with a conductive "gate" over an insulating oxide over the p-type region. In a PMOS device - figure 18 bottom, two highly doped p-type areas are separated by a lightly doped n-type area.

2.10.1. MOSFET off-state

If a positive voltage is applied to the drain of an NMOS device and the gate and source are tied to ground, then the drain to body pn junction is reverse biased and no current flows - see figure 19.

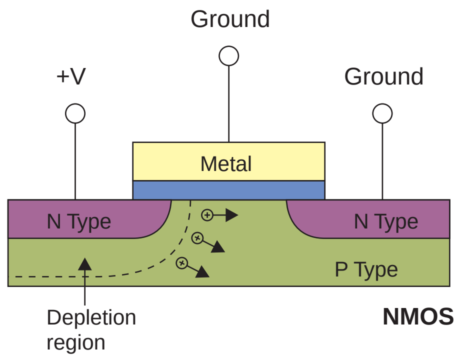

Figure 19. NMOS in the "off-state".

In figure 19, note how the depletion region only spreads into the body of the device and not into the drain, this is because the drain is very highly doped and the body is lightly doped.

The doping of the body of the MOSFET is very important to a number of operating parameters of the device. The doping right under the gate oxide determines the threshold voltage of the device. The doping needs to be light if the device is to have a low threshold voltage - a very desirable quality. On the other hand, low doping in the body of the device means the depletion region spreads further for a given applied voltage. If the depletion region spreads all the way to the source, the device "punches through" and breaks down. For very small MOSFETs, punch-through breakdown limits the operating voltage of the device to a value lower than the expected avalanche breakdown voltage. A great deal of device engineering that will be discussed in the later chapters is dedicated to punch-through control.

2.10.2. MOSFET on-state

If a positive voltage is applied to the drain of an NMOS device, the source is grounded and the gate also has a positive voltage applied greater than the threshold voltage, the surface under the gate will invert connecting the source and drain and allowing current to flow - see figure 20. In figure 20 the n-type source and drain areas become connected by an n-type inversion layer, so only n-type carriers participate in conduction making a MOSFET a unipolar device.

2.11. MOSFET scaling

In figure 17 it was shown that MOSFETs are the dominant type of IC in production today. The reason that MOSFETs have taken over from Bipolar devices is simply put, MOSFETs scale and Bipolar transistors don't.

The MOSFET 2

- Gate - the terminal of a MOSFET used to apply and electric filed to switch the MOSFET on.
- Channel region - the inversion layer that form under the gate oxide of a MOSFET in the on state.
- Source - the region of a MOSFET that injects carriers into the channel region during conduction.
- Drain - the region of a MOSFET that "drains" carriers from the channel region during conduction.

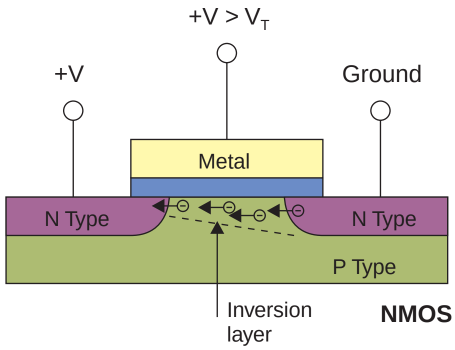

Figure 20. NMOS in the "on-state".

At the 1974 International Electron Devices Meeting Dennard, et.al., presented the classic paper on scaling "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions". Dennard's group had discovered that if a MOSFETs physical dimensions were scaled down while maintaining a constant electric field every other MOSFET performance parameter improved. For example, if a 250nm gate length MOSFET is scaled by 1.4, the new MOSFET has a gate length of 250/1.4 ~180nm. The operating voltage of 1.8 volts is also scaled by 1.4 to achieve a constant electric field and the new MOSFET operates at 1.8/1.4 ~ 1.3 volts. The resulting MOSFET is 1/2 the size, 1.4 times as fast and consumes 1/2 the power.

2.12. Conclusion

In this chapter the four basic device type commonly used in integrated circuits have been presented. In the next chapter we will discuss how NMOS and PMOS devices are used together to make CMOS (Complementary MOS), the structures required for CMOS and the issues that arise as CMOS is scaled to submicron geometries.

Scaling

- Gate length - the distance between on side of the gate electrode and the other side in the direction defined by a line drawn from drain to source of the MOSFET.
- Electric field - voltage per unit length. For example, the voltage applied between drain and source of a MOSFET divided by the distance between the drain and source of the MOSFET.

