Semiconductors and their bandgaps.

The absorption wavelength and emission wavelength depends on their bandgaps.

To grow different single crystal semiconductors on top of one another, it is necessary to also match their lattice constants.

For example, it is easy to grow AlAs on GaAs since they have very similar lattice constants.

Drift and Diffusion:

Drift: electric charges will respond to a field by moving in the direction towards the opposite pole (electrons move to the anode, holes move to the cathode)

Diffusion: Random thermal motion of particles (electrons, holes, dopant atoms, etc.) cause a redistribution of concentration with time to reduce the concentration gradient. The diffusion of particles creates an effective particle current towards the points of lower concentration.

\[
J = -D \frac{dc}{dx} \quad \text{….. Fick’s law}
\]
Diffusion:

Diffusion is a common way of introducing dopants into semiconductors. This processing step is the basis for many p-n junctions.

Atoms will diffuse from regions of high concentration to regions of low concentration.

\[ C = \text{concentration of atoms in units of atoms/cm}^3 \]

\[ C(x) = \text{concentration of atoms as function of distance} \]

\[ F = \text{flux of atoms in the solid} \]

\[ D = \text{diffusion coefficient in atoms/cm}^2\text{s} \]

\[ F = -D \frac{dC}{dx} \]

Another Example

Diffusion causes a rearrangement of the concentrations of these 4 populations with time.

An example would be fleas on 4 dogs running next to each other on a racetrack. The fleas can randomly jump from one dog to another. Say, one dog starts out with 1024 fleas…
Time-dependent one-dimensional diffusion equation (Fick’s second law)

\[ J_1 = -D(n_2-n_1)/dx = -D\frac{dC}{dx} \]
\[ J_2 = -D(n_3-n_2)/dx = -D\frac{dC}{dx} \]

The rate of accumulation in \( n_2 \) is \( J_1-J_2 \). Thus:
\[ \frac{\partial}{\partial x}(D\frac{\partial C}{\partial x})=\frac{\partial C}{\partial t} \]

Diffusion Coefficient

\[ D \frac{\partial^2 C}{\partial x^2} = \frac{\partial C}{\partial t} \]

Second derivative of concentration with respect to distance

Two common solutions to the diffusion equation (Fick’s second law)

For fixed number of particles diffusing from a square concentration profile:

\[ \frac{\partial C(x,t)}{\partial t} = D \frac{\partial^2 C(x,t)}{\partial x^2} \]

\[ C(x,t) = \frac{S}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) \]

For an infinite supply of diffusing particles from a surface:

\[ C(x,t) = C_i \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \]

Note that the bottom (distance) axis is not linear
The \( \text{erf}(x) \) function is very important for furnace diffusion problems where there is a surface source which is replenished continuously.

Here are examples of some \( \text{erf}(x) \) and complementary \( \text{erfc}(x) \) values tabulated for various values of \( x \).

<table>
<thead>
<tr>
<th>( x )</th>
<th>( \text{erf}(x) )</th>
<th>( \text{erfc}(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.01</td>
<td>0.0013</td>
<td>0.9987</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1125</td>
<td>0.8875</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8427</td>
<td>0.1573</td>
</tr>
<tr>
<td>1.5</td>
<td>0.9661</td>
<td>0.0339</td>
</tr>
<tr>
<td>2.0</td>
<td>0.9953</td>
<td>0.0047</td>
</tr>
<tr>
<td>2.25</td>
<td>0.998537</td>
<td>0.001463</td>
</tr>
<tr>
<td>2.50</td>
<td>0.999593</td>
<td>4.670 \times 10^{-4}</td>
</tr>
<tr>
<td>3.00</td>
<td>0.99997791</td>
<td>2.209 \times 10^{-4}</td>
</tr>
<tr>
<td>3.25</td>
<td>0.99999570</td>
<td>4.300 \times 10^{-4}</td>
</tr>
<tr>
<td>3.50</td>
<td>0.999999257</td>
<td>7.430 \times 10^{-4}</td>
</tr>
<tr>
<td>3.75</td>
<td>0.999999884</td>
<td>1.140 \times 10^{-3}</td>
</tr>
</tbody>
</table>

Source: After Grove (1967).

Diffusion profile after diffusing p-impurities into a n-doped sample. Note the position of the p-n junction is where both p and n dopant densities are equal.
\[ \frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \quad \text{Fick's Second Law} \]

1. Set up a grid space.

2. Derive a computational scheme.

\[ C_{i} = C_{i+1} - C_{i} \]

3. Solve for \( C_{i} \) for each step in time.

4. Stabilize the scheme.

5. Use an explicit method.

if \( i = 0 \), \( C_{i,0} \) is imaginary, but the boundary condition can still yield a numerical result.

\[ 1 \text{ step} = \frac{\Delta t}{2 \Delta x^2} \]

we get

\[ \frac{C_{i}^{n+1} - C_{i}^{n}}{\Delta t} = D \left( \frac{C_{i+1}^{n} - 2C_{i}^{n} + C_{i-1}^{n}}{\Delta x^2} \right) \]

Since we know the profile at time \( n \) (i.e., \( C_{\text{old}} \)),

we use \( C_{i}^{n}, C_{i+1}^{n}, \) and \( C_{i-1}^{n} \) to calculate \( C_{i}^{n+1} \):

\[ C_{i}^{n+1} = \frac{\Delta t}{\Delta x^2} \left( C_{i+1}^{n} - 2C_{i}^{n} + C_{i-1}^{n} \right) + C_{i}^{n} \]

This is the basic Forward Finite Difference algorithm.

Remarks:

1. Stability criteria \( \frac{\Delta t}{\Delta x^2} \leq \frac{1}{2} \)

2. New concentration needs to be found \( n+1 \) times the old concentration.
Examples of diffusion profiles after diffusion from (a) a constant source (erfc distribution) and (b) a finite source (Gaussian).

Fig. 4  Impurity diffusion coefficients as a function of temperature for Ge, Si, and GaAs. (After Burger and Donovan, Ref. 19; Kendall and DeVries, Ref. 20.)

Diffusion coefficients of elements in some common semiconductors.
To obtain reasonably fast diffusion processes, Si wafers are typically heated to 1000 degrees C in the presence of a gas source, such as phosphene, arsene, boron, etc...

![Diagram of a typical diffusion furnace](image)

Typical diffusion furnace (a large tube furnace) for doping of semiconductors. The surface concentration is constantly replenished with a gas flow. Contamination with other elements, such as carbon, should be carefully avoided.

What is the surface concentration in a furnace diffusion process?

Depending on the rate limiting step, it often is the solid solubility of the material in the semiconductor (in this case Si). Note that the solubility is strongly dependent on the temperature.
The light source can also be an excimer (fluoride) laser (ArF or KrF).

U-V photolithography relies mainly on 5:1 or 10:1 steppers like this one.

For research applications with odd shaped samples, contact aligners are sometimes used.

This lens may contain many (100) elements.

Throughput = 20 wafers/hour.
Oxidation of Silicon

This process is a tremendously important fabrication step which is used throughout CMOS fabrication for making masks and electrostatic gates. Gate oxide thicknesses have to be controlled often to within less than 1 nm to make small MOSFETs. Much thicker oxide layers are often used to isolate devices or as diffusion masks. These have to be about 1 micron thick.

Ion implantation: A more accurate (and expensive) doping technique

We can measure the ion current and determine the total number of implanted dopants.
Ion implantation profile as a function of depth. The projected range ($R_p$) depends on the energy and mass of the ions and is the midpoint of a Gaussian distribution.

$$N(x) = \frac{\phi}{\sqrt{2\pi}\Delta R_p} \exp \left[ -\frac{1}{2} \left( \frac{x - R_p}{\Delta R_p} \right)^2 \right]$$

Ion dose $\propto$ depth

Number of atoms at any depth $x$

Projected range

Straggle (distribution half width)

Several consecutive implants with different energies can be used to obtain a desired profile shape.

Projected range and straggle for different materials

Projected range

Straggle
After ion implantation, the samples can be annealed at high temperatures to remove some of the ion damage and to smooth out the dopant profile. The concentration profile then changes to:

$$N(x) = \frac{\phi}{\sqrt{2\pi(\Delta R_p^2 + 2Dt)^3}} \exp \left[ -\frac{1}{2} \left( \frac{(x - R_p)^2}{\Delta R_p^2 + 2Dt} \right) \right]$$

Vacuum Technology

- Invention of the Torricelli mercury manometer (1643)
  - Noticed that the length of a column of mercury in an up-ended tube was always the same

- Invention of the Vacuum pump (1647)
  - Otto von Gericke (Magdeburg hemispheres 1654)
Boyle, Charles, Bernoulli, Avogadro, Gay-Lussac

\[ P_1 T_1 = P_2 T_2 \]

Vacuum Technology

- The Ideal Gas law
  - \((PV=NkT)\) is a combination of Avogadro’s law \((V/N=\text{const})\) with Charles’s \((V/T=\text{const})\) and Boyle’s \((PV=\text{const})\) Laws

- Dalton’s Law
  - The total gas pressure of a system is equal to the sum of the partial pressures of the constituent gases
Effusion – Thomas Graham

Thomas Graham (1805-1869)
Graham was professor of chemistry at University College in London and later became Master of the Mint. He is best known for Graham's law (1846) which states that the rate of effusion of a gas is inversely proportional to the square root of its molecular weight. Graham also devised the technique known as dialysis to separate colloids from crystalloids and coined many of the terms used in colloid chemistry.

Speed of gas atoms/molecules

J.C. Maxwell

Maxwell treated gases statistically (1866) and formulated what has become known as the root-mean-square molecular equation \( \langle u \rangle = \sqrt{\frac{3RT}{M}} \). This represents a relationship between molecular mass, average speed, and temperature (\( R \) is the familiar gas constant). Because two gases with two different masses must have the same average kinetic energy at the same temperature, the heavier gas molecules must possess lower average speed.

On another front, Maxwell's mathematical equations expressing the behavior of electric and magnetic fields are considered one of great achievements of the 19th century.
Boltzmann’s distribution

The distribution of velocities among molecules of a gas was first developed by Maxwell (1859) and later generalized by Boltzmann (1871). The Maxwell-Boltzmann theory explained the gas laws in terms of the motion of individual molecules. Previously it was assumed heat flowed from hot to cold. The Maxwell-Boltzmann theory treated molecules at high temperature as having a high probability of moving toward those at low temperature. Consider the distribution of velocities for oxygen at 25°C shown in the figure: Boltzmann worked out a statistical approach to show more molecules moving at 400 m/s than at any other speed. This type of curve is called a Maxwell-Boltzmann distribution. All systems observed to date appear to obey the distribution law.

Vacuum Technology

- **Vulcanization of rubber (1839)**
  - Charles Goodyear started polymer science that developed a large spectrum of inert conformal sealing materials
- **Invention of the Vacuum flask (1892)**
  - Sir James Dewar developed the glass vacuum flask for storing liquid air and its fractions (O, N, Ar)
van der Waals Equation: \((P + \frac{a}{V^2})(V-b) = RT\)

### Vacuum Technology

- **Invention of the Vacuum Tube (1904)**
  - John Fleming (1904) demonstrates vacuum diode, and Lee DeForest (1906) demonstrates the triode

- **Discovery of the plasma discharge (1880)**
  - Sir William Crookes studied the passage of electrical current through “rarified gases”. Ultimately, this led to the discovery of the electron by J.J. Thomson
    - Plasma Processing
    - The term Plasma was first introduced by Langmuir
Vacuum Technology

- Irving Langmuir
  - Joined GE (1909), developed the mercury diffusion pump and started the field of “surface science”
- Invention of the Transistor (1947)
  - Required vacuum processing
  - Developed a good reason for lithographic miniaturization

Vacuum of 1mTorr to 1 Torr (760 Torr = 1 atm)
Could be mechanically pumped or diffusion pumped depending on the vacuum required
Freons are ideal for Si and SiO₂ etching since SiF₄ is a gas
Sometimes the wafer is heated to increase the etch rate
Reactive ion etching or Plasma etching process

Plasma Radicals & Ions
Etch Gas
Vacuum of 1mTorr to 1 Torr (760 Torr = 1 atm)
Freons are ideal for Si and SiO₂ etching since SiF₄ is a gas
Sometimes the wafer is heated to increase the etch rate
Reactive ion etching or Plasma etching process

Pump
Anode (Chamber)
Cathode
RF Power Supply
Wafer with photoresist mask
Sputter deposition is a popular metallization technique in the semiconductor industry.

Argon is ionized in a plasma and accelerated to the cathode, which in this case is a plate of aluminum.

Mechanical pumps: These are displacement pumps in which a piston or vane pushes gas molecules to an exhaust.
Diffusion pump and typical diffusion pumped vacuum system

Water cooling coils
To mechanical pump

The Ion getter pump uses a high field to ionize gases and then accelerates these gases into reactive metals (Ti)

The Turbomolecular pump uses rotating blades to “push” gas from the inlet to a mechanical pump
Scientific applications of vacuum systems

- Mass spectrometers
- Molecular beam apparatus
- Ion sources
- Particle accelerators
- Electron microscopes
- Electron diffraction apparatus
- Vacuum spectrographs
- Low-temperature research
- Production of thin films
- Surface physics
- Plasma research
- Nuclear fusion apparatus
- Space simulation
- Material research
- Preparation for electron microscopy

More general applications of vacuum pumps

- Annealing of metals
- Melting of metals
- Degassing of molten metals
- Steel degassing
- Electron-beam melting
- Electron-beam welding
- Evaporation coating
- Sputtering of metals
- Zone melting and crystal growing in high vacuum
- Molecular distillation
- Degassing of liquids
- Sublimation
- Casting of resins and lacquers
- Drying of plastics
- Drying of textile fibers
- Freeze-drying of foods etc.
- Freeze-drying of pharmaceutical products
- Production of incandescent lamps
- Production of electron tubes
- Production of gas-discharge tubes
Vacuum gauges: We can measure displacement, thermal conductivity or ion currents to determine the pressure.

The best way to measure the pressure is with a mass spectrometer, since it also gives the composition of the residual gas.

Thermocouple (Pirani) gauge
Drift and Diffusion of carriers

In an electric field, the current of charged carriers is controlled by both drift and diffusion. The sum of these two components will determine how much current will flow.

For electrons:
\[ J_n = q\mu_n n \mathcal{E} + qD_n \nabla n \]

For holes:
\[ J_p = q\mu_p p \mathcal{E} - qD_p \nabla p \]

\[ J_{\text{cond}} = J_n + J_p. \]

When we put p- and n-doped semiconductors in contact, the Fermi level of both semiconductors has to become equal.

This results in the bending of the conduction and valence bands (band bending) and the establishment of an internal electrostatic potential \( V_0 \).

When no current is flowing through the junction,
\[ J(\text{drift}) + J(\text{diffusion}) = 0 \]
There have to be an equal number of charges on either side of the junction.

Note that this looks more or less like a capacitor with opposite charges on either side of an insulating region.

Charges must equal on both sides of the junction.

If it's a capacitor, the field depends on the dielectric constant and the number of charges (Poisson's equation).

Simplifying a bit, we get:

\[
qA_{x_p0}N_a = qA_{x_n0}N_d
\]

Next, we integrate to get the field:

\[
\begin{align*}
\varepsilon_0 &= \frac{-q}{\varepsilon} N_d x_{p0} = \frac{-q}{\varepsilon} N_d x_{p0} \\
\end{align*}
\]
\[ \varepsilon_0 = \frac{-q}{\varepsilon} N_a x_{n0} = \frac{-q}{\varepsilon} N_d x_{p0} \]

\[ \varepsilon(x) = \frac{dV(x)}{dx} \quad \text{or} \quad V_0 = \int_{-x_{p0}}^{x_{n0}} \varepsilon(x) dx \]

\[ V_0 = -\frac{1}{2} \varepsilon_0 W = \frac{1}{2} \frac{q}{\varepsilon} N_a x_{n0} W \]

Since \( Ax_{n0}N_a = Ax_{p0}N_d \) and \( W = x_{p0} + x_{n0} \), we can substitute:

\[ V_0 = \frac{1}{2} \frac{q}{\varepsilon} \frac{N_a N_d}{N_a + N_d} W^2 \]

This provides us with an equation which we can use to calculate the depletion width if we know the dopant concentrations of a p-n junction.

**Depletion width equation:**

\[ W = \sqrt{\frac{2\varepsilon V_0}{q} \left( \frac{N_a + N_d}{N_a N_d} \right)} = \sqrt{\frac{2\varepsilon V_0}{q} \left( \frac{1}{N_a} + \frac{1}{N_d} \right)} \]

Depletion width \quad Built-in potential \quad Permittivity of semiconductor

where \( N_a \) and \( N_d \) are the Acceptor and Donor Dopant concentrations

Note: For silicon, \( \varepsilon \), the permittivity, is given as 11.8 \( (8.85 \times 10^{-14} \text{ F/cm}) \)

We can also determine how much of the depletion width is on the p and n side:

\[ x_{p0} = \frac{WN_d}{N_a + N_d} = \frac{W}{1 + N_d/N_a} = \left\{ \frac{2\varepsilon V_0}{q} \left[ \frac{N_d}{N_a(N_a + N_d)} \right] \right\}^{1/2} \]

\[ x_{n0} = \frac{WN_a}{N_a + N_d} = \frac{W}{1 + N_d/N_a} = \left\{ \frac{2\varepsilon V_0}{q} \left[ \frac{N_a}{N_d(N_a + N_d)} \right] \right\}^{1/2} \]
Example: A Si diode is formed from an abrupt junction with \(N_d = 10^{16}\) and \(N_a = 4 \times 10^{18}\). Calculate \(V_0, x_{n0}, x_{p0}, Q_+\), and the field for this junction.

**Built-in potential** \(V_0\)

\[
V_0 = \frac{kT}{q} \ln \frac{N_a N_d}{n_i^2} = 0.0259 \ln \frac{4 \times 10^{34}}{2.25 \times 10^{20}}
- 0.0259 \ln (1.78 \times 10^{14}) = 0.85 \text{ V}
\]

From Eq. (5-23),

\[
x_{n0} = \frac{3.34 \times 10^{-5}}{1 + 0.0025} = 0.33 \text{ \mu m}
\]

\[
x_{p0} = \frac{3.34 \times 10^{-5}}{1 + 400} = 8.3 \times 10^{-7} \text{ cm} = 8.3 \text{ \AA}
\]

Note that \(x_{n0} = W\).

\[
Q_+ = -Q_- = qAx_{n0}N_d = (1.6 \times 10^{-19})(2 \times 10^{-3})(3.33 \times 10^{-5})(10^{16})
= 1.07 \times 10^{-10} \text{ C}
\]

\[
\xi_0 = -\frac{qN_a x_{n0}}{\epsilon} = -\frac{(1.6 \times 10^{-19})(10^{16})(3.3 \times 10^{-5})}{(11.8)(8.85 \times 10^{-1}4)}
= -5.1 \times 10^4 \text{ V/cm}
\]
This relationship is for no external voltage applied to the diode. If we apply a reverse bias voltage $V_a$, we add that voltage to $V_o$ and use $(V_a+V_o)$ instead of $V_o$ here.

\[
W = \frac{2\epsilon V_o}{q} \left( \frac{N_a + N_d}{N_a N_d} \right)^{1/2} = \frac{2\epsilon V_o}{q} \left( \frac{1}{N_a} + \frac{1}{N_d} \right)^{1/2}
\]

Depletion width  
Built-in potential  
Permittivity of semiconductor

where $N_a$ and $N_d$ are the Acceptor and Donor Dopant concentrations.

Note: For silicon, $\epsilon$, the permittivity, is given as $11.8 \times (8.85 \times 10^{-14} \text{ F/cm})$.

Note: As we increase the reverse bias voltage, the depletion width increases.

At positive voltages, the current follows an exponential dependency on voltage:

\[
I = I(\text{gen.}) \left( e^{qV/kT} - 1 \right)
\]

The Boltzmann diode equation

At negative voltages, the current saturates to a saturation current $I_{\text{gen}}$.

At 0 applied Voltage, the current flowing is 0.
We know that the concentrations of holes in the p side and the concentration of holes on the n side of a p-n junction are related by $V_o$:

\[
\frac{p_p}{p_n} = e^{qV_o/kT}
\]

\[
\frac{p(-x_{p0})}{p(x_{n0})} = e^{q(V_o-V)/kT}
\]

Now, we can find the excess minority carrier concentration on both p and n-side of the depletion width by subtracting:

\[
\Delta p_n = p(x_{n0}) - p_n = p_n(e^{qV/kT} - 1)
\]

\[
\Delta n_p = n(-x_{p0}) - n_p = n_p(e^{qV/kT} - 1)
\]
Since we know that the excess carrier concentration drops off to zero as we get further away from the junction, we can define diffusion lengths $L_n$ and $L_p$ for electrons and holes, respectively and assume an exponential decay with distance.

These diffusion lengths depend on the dopant concentration (recombination).

Usually, the n and p regions are long with respect to $L_n$ and $L_p$ and:

$$\delta n(x_p) = \Delta n_p e^{-\frac{x_p}{L_n}} = n_p (e^{qV/kT} - 1)e^{-\frac{x_p}{L_n}}$$

$$\delta p(x_n) = \Delta p_n e^{-\frac{x_n}{L_p}} = p_n (e^{qV/kT} - 1)e^{-\frac{x_n}{L_p}}$$

These are the excess hole and electron concentrations as a function of distance $x_n$ and $x_p$ from the junction

We remember the current density equation, and can simplify it to only contain the diffusion current

$$J_p(x) = q \left[ \mu_p D_p (x) \delta (x) - D_p \frac{dp(x)}{dx} \right] = 0$$

From this, we can define the diffusion current as:

$$I_p(x_n) = -qAD_p \frac{d\delta p(x_n)}{dx_n} = qA \frac{D_p}{L_p} \Delta p_n e^{-\frac{x_n}{L_p}} = qA \frac{D_p}{L_p} \delta p(x_n)$$

The total hole current is:

$$I_p(x_n = 0) = qAD_p \frac{\Delta p_n}{L_p} = qAD_p \frac{D_p}{L_p} p_n (e^{qV/kT} - 1)$$

The total electron current is

$$I_n(x_p = 0) = -qAD_n \frac{\Delta n_p}{L_n} = -qAD_n \frac{D_n}{L_n} n_p (e^{qV/kT} - 1)$$
Finally, the total diode current is the sum of the hole and electron currents across the p-n junction and is given by:

\[
I = I_p(x_n = 0) - I_n(x_p = 0) = \frac{qA D_p}{L_p} \Delta p_n + \frac{qA D_n}{L_n} \Delta n_p
\]

\[
I = qA \left( \frac{D_p}{L_p} p_n + \frac{D_n}{L_n} n_p \right) \left( e^{qV/kT} - 1 \right) = I_0 \left( e^{qV/kT} - 1 \right)
\]

Area of junction    diffusivity    diffusion length    Applied voltage    Temperature (K)

For those interested, here are three ways of deriving the diode equation:

\[
I = -qA \left( \frac{D_p}{L_p} p_n + \frac{D_n}{L_n} n_p \right) = -I_0
\]

\[
I_p(x_n = 0) = qAD_p \frac{\partial n}{\partial x_p} \mid_{x_n = 0}
\]

\[
= -qA \frac{D_p}{L_p} \Delta n_p
\]

\[
I_n(x_p = 0) = -qAD_n \frac{\partial p}{\partial x_n} \mid_{x_p = 0}
\]

\[
= qA \frac{D_n}{L_n} \Delta p_n
\]

\[
I = I_p(x_n = 0) - I_n(x_p = 0) = qA \left( \frac{D_p}{L_p} \Delta p_n + \frac{D_n}{L_n} \Delta n_p \right)
\]

\[
= qA \left( \frac{D_p}{L_p} \Delta p_n + \frac{D_n}{L_n} \Delta n_p \right) \left( e^{qV/kT} - 1 \right)
\]
Diode Breakdown Effects:

When we apply a large reverse bias (negative voltage) onto a diode, it eventually breaks down and conducts again. That voltage is called the reverse breakdown voltage $V_{br}$.

There are two common breakdown mechanisms: Tunneling and Avalanche Multiplication.

Avalanche Multiplication

If a high enough voltage is applied to the depletion region, minority carriers are accelerated through this region and are energetic enough to generate secondary electrons. This results in an “avalanche” effect which amplifies the number of free carriers in the depletion region and results in breakdown.

Note: Avalanche photodiodes are often used as sensitive optical sensors with built-in gain. They are typically biased to 100V.
Higher fields occur at small depletion widths and large reverse biases.

For Si, $3 \times 10^5$ V/cm

$V_{br} = \frac{\varepsilon E_{br}^2}{2qN_d}$

Plot of breakdown characteristics as a function of dopant concentration

Tunneling or Zener breakdown:

At a high enough voltage, if $W$ is narrow enough, it is possible for electrons to tunnel from the valence band of the p-side of the diode directly to the conduction band of the n-side of the diode.

Note: Zener diodes are often used as voltage control devices.

The mechanism for this process is field ionization, which requires very short $W$ and high fields of $>10^6$ V/cm in the junction. There is not enough distance for impact ionization if $N_d$ and $N_a$ are high.

This results in a very sharp increase in current at the tunneling voltage $V_T$. 

$V_T$
Metal semiconductor contacts:
p-semiconductor case:

When a metal is deposited on a semiconductor directly the work function of the metal and the Fermi level of the semiconductor must line up again.

Metal semiconductor contacts:
n-semiconductor case:

In this case, the metal behaves very much like a heavily doped p+ layer, and the semiconductor is depleted of electrons.

The work function of a metal is defined as the energy for its (free) electrons to escape into vacuum.

A depletion layer is formed because of the band-bending.
Again, when we apply a voltage, diffusion current will give rise to an exponential increase of the current.

And in reverse bias, the current flow is limited to $I_{gen}$.

For a Metal-Semiconductor (Schottky) diode, the I-V relationship can be described by:

$$I = ABT \cdot e^{-\frac{q\Phi_B}{kT}} \cdot e^{\frac{qV}{nkt}}$$

- $I$ = Current
- $A$ = Cross-sectional Area
- $B$ = barrier height
- $T$ = temperature
- $n$ = ideality factor
- $N$ = ideality factor which ranges from 1-2
- $\Phi_B$ = Schottky barrier height ~0.85 for a typical Si surface with a Pt contact. This value depends on the surface of Si and the work function of the metal chosen.
- $B$ = constant describing the junction properties

Note that the last part of this equation is very similar to the regular p-n diode current equation.
Surface depletion and Fermi level pinning

Many semiconductors, such as GaAs have dangling bonds on the surface, and this pins the Fermi level to a fixed value of ~0.8eV below the conduction band. This results in band bending and surface depletion even if there is no metal on the surface.

In InAs, the Fermi level is pinned above the conduction band edge. This means that this material is ideal for constructing ohmic contacts.

The Transistor: First proposed by Lilienfeld in 1930 (but he could never really get it to work right because of surface states)

Field-effect transistor:
Using a gate C, Lilienfeld thought that it should be possible to modulate the current from A to B.

This is conceptually very similar to the vacuum triode, which was used as the amplifier at the time.
In 1947, Bardeen and Brattain invented the Ge point contact transistor. They wanted to make the Field-Effect Transistor, but ended up with a Bipolar Transistor, and got the Nobel Prize anyway. Schockley then developed the bipolar junction transistor.

Field effect devices: The electric field of a gate or grid is used to modulate the number of charges (i.e. electron current) moving from the source to the drain.

In a triode, the charges are electrons accelerated through a vacuum.

N-type Si channel
MOS capacitor

\[ \frac{d \varepsilon}{dx} = \rho \frac{\mu}{K_S \varepsilon_0} = -\frac{q N_A}{K_S \varepsilon_0} \quad (0 \leq x \leq W) \]
Charge balance of a MOS capacitor. \( -Q = +Q \)
\[ E_i(\text{surface}) - E_i(\text{bulk}) = 2[E_F - E_i(\text{bulk})] \]

\[ p_n = n_i e^{[E_i(\text{surface}) - E_F]/kT} = n_i e^{[E_F - E_i(\text{bulk})]/kT} = n_{\text{bulk}} = N_D \]
Substitute the definition of $E_i$ and $E_F$

$$P_{\text{bulk}} = n_i e^{(E_i(\text{bulk}) - E_F)/kT} = N_A \quad \ldots \text{if } N_A \gg N_D$$

$$n_{\text{bulk}} = n_i e^{(E_F - E_i(\text{bulk}))/kT} = N_D \quad \ldots \text{if } N_D \gg N_A$$

$$\phi(x) = \frac{1}{q} [E_i(\text{bulk}) - E_i(x)]$$

$$\phi_S = \frac{1}{q} [E_i(\text{bulk}) - E_i(\text{surface})]$$

$$\phi_F = \frac{1}{q} [E_i(\text{bulk}) - E_F]$$

$$\phi_F = \left\{ \begin{array}{ll}
\frac{kT}{q} \ln(N_A/n_i) & \ldots \text{p-type semiconductor} \\
-\frac{kT}{q} \ln(N_D/n_i) & \ldots \text{n-type semiconductor}
\end{array} \right.$$

$\phi_S = 2\phi_F \quad \text{at the depletion–inversion transition point}$
Note that the charge associated with inversion resides in an extremely narrow channel immediately adjacent to the oxide.

As the drain voltage is increased, the inversion voltage is reduced.
The saturation current can be controlled by adjusting the gate voltage.

Threshold voltage

Interfacial and trapped charges

Metal semiconductor workfunction difference

Charges from depleted dopants

\[ V_T = \Phi_{ms} - \frac{Q_i}{C_i} - \frac{Q_d}{C_i} + 2\Phi_F \]
Definition of the threshold voltage

\[ V_T = \Phi_{ms} - \frac{Q_i}{C_i} - \frac{Q_d}{C_i} + 2\Phi_F \]

The threshold voltage is dependent on the dopant concentration too. This is a way the MOSFET \( V_T \) can be tuned.

Crossection throught the gate of a narrow-channel MOSFET
We should also include the Modulation-doped FET or MODFET (also known as HEMT)
Schematic description of the JFET problem

\[ J_N = q\mu_n n \phi + qD_N \nabla n \]

\[ J_N = J_{Ny} = q\mu_n N_D \phi_y = -q\mu_n N_D \frac{dV}{dy} \]
Integrating the drift equation with respect to x and z, we can derive the total current passing through the transistor channel:

\[ I_D = - \int J_{Ny} \, dx \, dz = -Z \int_{W(y)}^{2a-W(y)} J_{Ny} \, dx = 2Z \int_{W(y)}^{a} q \mu_n N_D \, \frac{dv}{dy} \, dx \]

\[ = \frac{q \mu_n N_D a}{L} \frac{dV}{dy} \left( 1 - \frac{W}{a} \right) \]

\[ \int_{0}^{L} I_D \, dy = I_D L = 2q \mu_n N_D a \int_{V(0)=0}^{V(V)=V_D} \left[ 1 - \frac{W(V)}{a} \right] \, dV \]

\[ I_D = \frac{2q \mu_n N_D a}{L} \int_{0}^{V_D} \left[ 1 - \frac{W(V)}{a} \right] \, dV \]

To find the relationship \( W(V)/a \), we need to derive the depletion width as a function of voltage:

\[ W(V) \approx \left( \frac{2K_S \varepsilon_0}{qN_D} (V_{bi} - V_A) \right)^{1/2} = \left[ \frac{2K_S \varepsilon_0}{qN_D} (V_{bi} + V - V_G) \right]^{1/2} \]

\[ a = \left[ \frac{2K_S \varepsilon_0}{qN_D} (V_{bi} - V_P) \right]^{1/2} \]

\[ W(V) = \left( \frac{V_{bi} + V - V_G}{V_{bi} - V_P} \right)^{1/2} \]

Note that \( W \) approaches \( a \) when the gate voltage \( V_G \) approaches the pinchoff voltage \( V_P \).

Next, we substitute into our original equation and obtain:

\[ I_D = \frac{2q \mu_n N_D a}{L} \int_{0}^{V_D} \left[ 1 - \frac{W(V)}{a} \right] \, dV \]

\[ I_D = \frac{2q \mu_n N_D a}{L} \left\{ V_D - \frac{2}{3} (V_{bi} - V_P) \left[ \left( \frac{V_D + V_{bi} - V_G}{V_{bi} - V_P} \right)^{3/2} - \left( \frac{V_{bi} - V_G}{V_{bi} - V_P} \right)^{3/2} \right] \right\} \]

for \( 0 \leq V_D \leq V_{D_{max}} \); \( V_P \leq V_G \leq 0 \)
Note: This relationship only applies below pinchoff

\[ I_D = \frac{2q \mu_n N_D a}{L} \left( V_D - \frac{2}{3} (V_{bi} - V_P) \left[ \left( \frac{V_D + V_{bi} - V_G}{V_{bi} - V_P} \right)^{3/2} - \left( \frac{V_{bi} - V_G}{V_{bi} - V_P} \right)^{3/2} \right] \right) \]

for \( 0 \leq V_D \leq V_{Dat}; \quad V_P \leq V_G \leq 0 \)

To find the relationship above pinchoff, we use:

\[ I_{D\mid V_D>V_{Dat}} = I_{D\mid V_D=V_{Dat}} = I_{Dat} \]

Thus, we derive the saturation current \( I_{Dat} \)

\[ I_{Dat} = \frac{2q \mu_n N_D a}{L} \left( V_{Dat} - \frac{2}{3} (V_{bi} - V_P) \left[ \left( \frac{V_{Dat} + V_{bi} - V_G}{V_{bi} - V_P} \right)^{3/2} - \left( \frac{V_{bi} - V_G}{V_{bi} - V_P} \right)^{3/2} \right] \right) \]

We can simplify this relationship some more by noting that \( W \) approaches a when \( V(L) = V_{Dat} \)

Then,

\[ a = \left[ \frac{2K_S e_0}{q N_D} (V_{bi} + V_{Dat} - V_G) \right]^{1/2} \]

\[ V_{Dat} = V_G - V_F \]

\[ I_{Dat} = \frac{2q \mu_n N_D a}{L} \left( V_G - V_P - \frac{2}{3} (V_{bi} - V_P) \left[ 1 - \left( \frac{V_{bi} - V_G}{V_{bi} - V_P} \right)^{3/2} \right] \right) \]

\[ I_{Dat} = I_D(1 - V_G/V_P)^2 \quad \text{where} \quad I_D = I_{Dat\mid V_G=0} \]
Calculating the pinch-off voltage $V_p$

The total voltage equal to the built-in barrier potential $V_{bi}$ is:

$$d_n = (2\varepsilon V_{bi}/[qN_D])^{1/2}$$

As the reverse voltage increases, a value $V_p$ is reached at which the depletion layer extends all the way across the channel thickness $a$. From this we can obtain:

$$V_p + V_{bi} = (q/2 \varepsilon)N_D a^2$$

Calculating $f_T$

$$f_T = L/v_d = L/\mu E_x$$

Theoretical $I_D$-$V_D$ characteristics assuming $V_{bi}=1V$ and $V_p=-2.5V$ ($I_{D0}=I_{Dsat}$ at $V_G=0$)

Experimental I-V characteristics of a TI 2N3823 n-channel FET
Another equation for the saturation voltage can be given by:

\[ V_{Dsat} = V_P - V_G - V_{th} = \frac{qN_Da^2}{2\varepsilon_s} - V_G - \frac{kT}{q} \ln \left( \frac{N_DN_a}{n_s^2} \right). \]

\[ V_B (\text{breakdown voltage}) = V_D + V_G \]

**Summary for some common characteristics used in JFETs**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Common Factor</th>
<th>Multiplying Factors for Specific Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g_{max} )</td>
<td>( \frac{2Z\mu p a}{L} )</td>
<td>A: All Charge at ( y = 2a )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B: Uniform</td>
</tr>
<tr>
<td>( V_F )</td>
<td>( \frac{4p a^2}{e_s} )</td>
<td>C: All Charge at ( y = 0 )</td>
</tr>
<tr>
<td>( I_P )</td>
<td>( \frac{8Z\mu p a^3}{e_s L} )</td>
<td></td>
</tr>
<tr>
<td>( \frac{g_{max} V_P}{I_P} )</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Circuit symbols for depletion and enhancement JFETs.

\[ I_{D_{\text{sat}}} = I_p \left[ 1 - \left( \frac{V_G + V_{B1}}{V_P} \right)^2 \right]. \]

MESFETs

The channel can either be normally on or normally off, depending on the depletion width at zero bias on the gate:

Normally on

Normally off
Why would we want to choose a material other than Silicon?

\[ v = \frac{\mu E_x}{1 + \mu E_x/v_s} \]

\[ I_D = qN_D \frac{\mu E_x}{1 + \mu E_x/v_s} (a - h)Z. \]

Electric field profiles and current-voltage characteristics of a Si MESFET under various gate and drain biasing conditions.
Channel cross-section, electric field, drift velocity and space-charge distribution of a GaAs MESFET operated in the current-saturation region.

Time-response for a field-effect transistor: We need to use an equivalent circuit of a MESFET:

\[
\tau = \frac{L}{\mu \xi_x} \approx \frac{L^2}{\mu V_D}
\]

\[
f_T = \frac{g_m}{2\pi C_{GS}} \left( = \frac{1}{2\pi \tau} = \frac{v_s}{2\pi L} \right).
\]

\[
f_{\text{max}} \approx \frac{f_T}{2\sqrt{f_1 + f_T f_3}}
\]
Theoretical cutoff frequency $f_T$ as a function of gate length for Si, GaAs and InP.

Current carrying capacity is extremely important for any transistor and will determine the ultimate device geometry.

*Fig. 22* $I-V$ characteristic of a power MESFET; $I_D$ is the drain current under forward gate bias, $V_B$ is the breakdown voltage, and $(I_{DO}, V_{DO})$ is the operating biasing point. (After DiLorenzo and Wiseman, Ref. 48.)
The total current which can be amplified by the transistor depends on the channel doping and the channel depth.

Ultimately, there is a compromise between power output and frequency.

Here we show the state-of-the-art power vs. frequency curve for a GaAs MESFET.
Various gate configurations for GaAs MESFETs

Power MESFET with plated heat sink and interdigital source and drain fingers for high current operation
Various gate configurations to improve the device performance

Crossectional view of a double heterostructure MESFET.

Band Diagram of the device
V-groove FET showing the geometry and the operation principle of this very short-channel FET.
As the drain voltage is increased, the inversion voltage is reduced.

Complementary MOS (or CMOS) provides the basis of modern Silicon electronics.
The saturation current can be controlled by adjusting the gate voltage.

Ideal threshold voltage dependence: The threshold voltage should only depend on the number of dopants and the dielectric constant. However, there are extra charges in the oxide and on the surface, which make our lives somewhat more complicated.

\[
V_T = 2\phi_F + \frac{K_s x_0}{K_O} \sqrt{\frac{4qN_A}{K_s \varepsilon_0}} \phi_F \quad \text{... ideal } n\text{-channel (p-bulk) devices}
\]

\[
V_T = 2\phi_F - \frac{K_s x_0}{K_O} \sqrt{\frac{4qN_D}{K_s \varepsilon_0}} (-\phi_F) \quad \text{... ideal } p\text{-channel (n-bulk) devices}
\]
Impurities in oxide and at interface provide extra charges which need to be considered.

As the dopant concentration is changed, the workfunction difference is also changed.
Threshold voltage for a MOSFET transistor

\[ V_T = \Phi_{ms} - \frac{Q_i}{C_i} - \frac{Q_d}{C_i} + 2\Phi_F \]

Threshold voltage
Interfacial and trapped charges
Metal semiconductor workfunction difference
Charges from depleted dopants

\[ V_T = \Phi_{ms} - \frac{Q_i}{C_i} - \frac{Q_d}{C_i} + 2\Phi_F \]

Definition of the threshold voltage

The threshold voltage is dependent on the dopant concentration too. This is a way the MOSFET \( V_T \) can be tuned.
Crossection through the gate of a narrow-channel MOSFET

Figure 17.5 Visualization of surface scattering at the Si–SiO₂ interface.

\[
\bar{\mu}_n = \frac{\int_{x_c(y)}^{x_e(y)} \mu_n(x, y)n(x, y) \, dx}{\int_{0}^{x_e(y)} n(x, y) \, dx}
\]
High Electron Mobility Transistor (HEMT)

Modulation doping is used to create a high mobility channel underneath a gate.
The concept of modulation doping:

We can build a “Quantum Well” in which carriers prefer to stay, and provide these carriers by using “donor” layers which are heavily doped, but do not contribute to large amounts of scattering.
Nowadays, MOSFETs are very common field effect transistors, and are the backbone of digital logic circuits and memory.

**RAM**

( Random Access Memory )

- access any bit of information in a matrix of bits independently.
- two types: static and dynamic.
- SRAM --- retain data stored indefinitely.
- DRAM --- refresh stored data periodically.

**DRAM**

- can simplify structure to one transistor and a capacitor.
- high packing density.
- increase capacitor area and find high \( \varepsilon \) insulator.
- example: Hi-C DRAM, double level DRAM, trench DRAM.
Now, there are two categories of commonly used transistors:

Field effect transistors (source, gate, drain)

Bipolar transistors (emitter, base, collector)
This is where this has lead...