- Intrinsic and doped semiconductors
- Doping methods
- Law of Mass Action
- Chemical potential in semiconductors
- Conductivity of intrinsic and doped semiconductors
**Recall: Semiconductors**

**Semiconductors**: the highest completely filled band (the valence band) is separated from the lowest empty band (the conduction band) by the band gap $E_g < 3.2 \, eV$ ($\hbar \omega \approx 3.2 \, eV$ corresponds to the upper (violet) limit of visible light spectrum).

Electrons and holes can be created by several processes: thermal excitations, photon absorption, field effect, etc. The activation energies can be controlled separately for electrons and holes by *doping*.

Because of their significance for applications, extremely clean semiconductors have been fabricated and methods of controllable addition of doping impurities have been developed. Unintentional doping concentration were reduced below $\sim 10^{16} m^{-3}$, which is comparable to the intrinsic carrier concentration in undoped semiconductors at RT. The atom concentration in $Si$ is $\sim 5 \times 10^{28} m^{-3}$, so the material purity is in the part-per-trillion range!

Moderate doping preserves the band structure of the host material. It provides the charge carriers that otherwise wouldn’t be available. However, doping also produces disorder (carrier scattering by charged donors). In some cases, e.g. in GaAs/AlGaAs heterostructures, carriers and parent charged donors can be spatially separated.
Each material has several valence and conduction bands.

For these materials, the valence bands are similar (though the effective hole mass is different for different bands). The maxima of all valence bands are at the BZ center, where two hole bands are degenerate.

The conduction bands are quite different. The minimum of the conduction band in GaAs is also at the BZ center (direct bandgap semiconductor), for Si and Ge – at a non-zero $k$ (indirect bandgap semiconductors).
Where Electrons and Holes Live in the $k$-space

Even with heavy doping, the carrier concentration rarely exceeds $10^{26} m^{-3}$. The atom concentration in $Si$ is $\sim 5 \times 10^{28} m^{-3}$. Thus, electrons occupy only a very small fraction of the conduction band where their energy is lowest (and holes occupy only a very small fraction of the valence band), and we should be mostly interested in the properties of the energy spectrum in the vicinity of the $\varepsilon(k)$ minimum for the conduction band and in the vicinity of its maximum for the valence band. The effective mass approximation applies, the effective mass is typically a tensor.
Effective Masses in bulk Si, Ge, and GaAs

\[ g_C(\varepsilon) = \frac{m_e^*}{\pi^2 \hbar^3} \sqrt{2m_e^* (\varepsilon - E_C)} \]

Averaging is different for the effective masses that are relevant for the conductivity and DoS.

\[ m_{e,DOS} = M_C^{2/3} \sqrt{m_lm_t m_t} \]

\[ M_C \] - the number of equivalent band minima

\[ m_{e,\sigma} = 3 \left( \frac{1}{m_l} + \frac{1}{m_t} + \frac{1}{m_t} \right)^{-1} \]

When the electron motion is restricted to 2D, the effective masses are different: in [100] n-type Si MOSFETs \( m_{e,\sigma} \approx 0.19m_e \)


- Smallest energy bandgap at 300 K
- Effective mass for density of states calculations
- Effective mass for conductivity calculations
- Free electron mass

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Germanium</th>
<th>Silicon</th>
<th>Gallium Arsenide</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_g ) (eV)</td>
<td>0.66</td>
<td>1.12</td>
<td>1.424</td>
</tr>
<tr>
<td>( m_{e,DOS}/m_0 )</td>
<td>0.56</td>
<td>1.08</td>
<td>0.067</td>
</tr>
<tr>
<td>( m_{h,DOS}/m_0 )</td>
<td>0.29</td>
<td>0.57/0.81 ( \uparrow )</td>
<td>0.47</td>
</tr>
<tr>
<td>( m_{e,\text{cond}}/m_0 )</td>
<td>0.12</td>
<td>0.26</td>
<td>0.067</td>
</tr>
<tr>
<td>( m_{h,\text{cond}}/m_0 )</td>
<td>0.21</td>
<td>0.36/0.386 ( \uparrow )</td>
<td>0.34</td>
</tr>
<tr>
<td>( m_0 ) (kg)</td>
<td>9.11 x 10^{-31}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Shallow impurities**: large-radius “Bohr’s” orbit and small binding energy (singly-ionized ions + small effective mass + large dielectric constant). Energy is either close to the top of the VB, or to the bottom of the CB.

**Deep traps**: small-radius orbit and large binding energy (e.g., multiply-ionized atoms).
Shallow Impurities

Shallow donors (acceptors) can be viewed as Hydrogen-like “atoms” with an effective electron mass $m^*$ in a medium with the effective dielectric constant $\varepsilon_r$. Modified electron energy and Bohr’s radius are:

$$E^\text{eff} = E_0 \left( \frac{m^*}{\varepsilon_r^2 m_e} \right)$$

$$a^\text{eff}_0 = a_0 \left( \frac{\varepsilon_r}{m^*} \right)$$

For Si $m^* \approx 0.2 m_e$, $\varepsilon_r \approx 12 \Rightarrow E^\text{eff} \approx 20\text{meV}$, $a^\text{eff}_0 \approx 3\text{nm} \gg a$

Thus, the energy of electron “bonding” to donors (holes to acceptors) is much smaller than the band gap in Si. The electron states are by $E^\text{eff}$ lower than the bottom of the conduction band, whereas the hole states are by $E^\text{eff}$ higher than the top of the valence band.

Because $E_D (E_A)$ is so small for shallow impurities, $RT \ (= 26\text{meV})$ is sufficient to ionize the dopant atoms and create free charge carriers in the conduction or valence bands.
Doping Techniques

**Bulk Doping**: dopants can be added during the crystal growth process.

**Surface doping**: by diffusion and implantation (dominant over the last 20 years)

In electronic industry, the doping requirements span several orders of magnitudes in energy and dose, for a wide range of dopant masses. For example, heavy doping is used to form the source/drain contacts in MOSFETs (next lecture).

Leading edge CMOS processes which are used to fabricate a modern microprocessor require *up to twenty* different types of doping (by ion implantation) per wafer.
Ion Implantation

The ion energies are in the range of $100eV$ to $3MeV$. A mono-energetic ion beam forms a broad depth distribution of dopants. Under typical conditions the depth ranges between $10nm$ and $1\mu m$. Energies below a few $keV$ at a high dose are used for the formation of ultra-shallow dopant distributions.

Crystal damage created by ion implantation has to be removed by rapid annealing. Annealing – to heal the crystal and eliminate dangling bonds (carrier traps), rapid - to prevent dopant diffusion.
Advantages of Ion Implantation

- Precise control of dose and depth profile. Energy of ions controls the distribution, the ion beam current controls the dose.
- Low-temperature process (photoresist can be used as a self-aligned mask).
- Less sensitive to surface cleaning procedures.

Application example: self-aligned MOSFET source/drain regions
Substitutional doping of semiconductors up to the alloying range is hampered by the solubility of the doping impurity into the host lattice. Above the solubility limit the solid solution becomes thermodynamically unstable, leading to phase separation and formation of aggregates. Among the various out-of-equilibrium techniques developed to overcome this limitation, laser-assisted processes are well adapted to silicon, germanium and their alloys. A laser burst melts the surface of the crystal during a few tens of nanoseconds. After the burst, the speed of re-crystallization is of the order of $10^{10}$ K s$^{-1}$, slow enough to allow crystal reconstruction on the underlying non-melted template, but fast enough to prevent impurity diffusion and precipitation. Doping atoms can be introduced by pre-implantation (laser thermal anneal) or in situ by exposure of the surface to impurity-carrying gas molecules in a vacuum chamber (gas-immersion laser doping or GILD), as illustrated below.

Recently, the GILD technique$^{37}$ has been applied to induce superconductivity in silicon by boron doping$^4$. In-situ real-time monitoring of the transient reflectivity at a 675-nm wavelength made it possible to adjust the power and duration of the pulses for optimal doping and to ensure that the dopant incorporation profile was flat with an abrupt interface. High-resolution X-ray diffraction measurements and secondary ion mass spectroscopy on the doped layers demonstrated a maximum doping level of the order of 10 at% atomic (that is, $\sim 5 \times 10^{21}$ atoms cm$^{-3}$) for typically 500 subsequent laser shots. This substitutional concentration is far above the critical threshold for the MIT (a few $10^{18}$ at cm$^{-3}$) and even above the solubility limit ($\sim 10^{20}$ at cm$^{-3}$). Unfortunately, this technique, well suited to silicon and germanium, cannot be applied to most of the other semiconducting compounds, including diamond.

X. Blasé et al., Nat. Mater. 8, 375 (2009)
Doping concentration for Si may range from \(10^{19}\) to \(10^{26}\). At \(N_{A/D} > 10^{24}\), “impurity bands” are formed (separated from CB and VB). This is a misleading term as the wavefunctions may remain localized. Indeed, though the impurity bands are no more than half-filled, they don’t necessarily support extended electron/hole states (disorder \(\Rightarrow\) Anderson MIT, el.-el. correlations \(\Rightarrow\) Mott MIT).

A semiconductor with delocalized states in donor/acceptor band is referred to as “degenerate” (c.f. “degenerate Fermi gas”). For electrons in Si: B the metal-insulator transition in the donor band occurs at \(n_d \approx 4 \times 10^{24}\) (the atom concentration in Si is \(\sim 5 \times 10^{28}\), so the donor concentration at the MIT is 0.01%).
**Impurity Bands, Degenerate Semiconductors (cont’d)**

(b) At higher concentration, when the inter-impurity distance is small, quantum tunneling tends to delocalize the electron/hole wavefunctions, leading to a metallic behavior at zero temperature with a Fermi level pinned inside the impurity band.

(c) Because of strong on-site correlations, the spin-degenerate half-filled impurity band can split into an empty band and a full band (Mott insulator).

(d)-(e) On further doping, these two bands begin to overlap and a degenerate semiconductor is formed (the metal–insulator transition takes place).

X. Blasé et al., Nat. Mater. 8, 375 (2009)
"Electrons" and "holes" in semiconductors are different from the "electron-like" and "hole-like" quasiparticles in metals.

In metals, the "hole" has no simple "image". Each band can have both "electrons" and "holes".

In semiconductors, this image exists – excitations are related to broken covalent bonds. Electrons occupy the conduction band, holes – the valent band. The number of "electrons" and "holes" is constant at a given $T$, and the current magnitude is determined by the speed of the charge carriers which they acquire in the electric field.

The difference is due to the lattice bonding in these materials:
- metallic bonding with positively charged ionic lattice in the negatively charge sea of electrons;
- covalent bonding in semiconductors.

At a given $T$, there is a **dynamic equilibrium** between the number of broken covalent bonds and the number of recombination processes.
Recall: Chemical Potential in Metals

By definition, the chemical potential is the total energy change (kinetic + potential) in the process of adding one particle to a system.

In metals at $T = 0$, adding one electron results in an increase of the kinetic energy of the degenerate electron gas by $E_F$ (potential energy is ignored). At low $T$, the situation is similar - on average, adding of an electron increases the system energy by $E_F$. At high (comparable to $\mu$) temperatures:

$$
\mu(T) \approx \mu(0) \left[ 1 - \frac{\pi^2}{12} \left( \frac{k_B T}{\mu(0)} \right)^2 \right]
$$

- this means that since the occupancy becomes $<1/2$ at $\mu(0)$, the price to pay for adding an electron on average is less than $\mu(0)$. $\mu$ becomes negative when the occupancy at $\varepsilon = 0$ becomes less than 1/2. This is a non-degenerate case, never realized in metals.

Thus, in metals we used interchangeably the chemical potential and the Fermi energy, and $E_F$ separated filled and empty electron states at $T = 0$.

In semiconductors $\mu$ is usually in the band gap, and its meaning differs from $E_F$ in the “metallic” sense.
In intrinsic (\( n = p \), no dopants) semiconductors, only the “Boltzmann” tails of the Fermi function are in the \( CB \) and \( VB \), there are no electron states near the chemical potential which is in the band gap.

The position of \( \mu \) can be found from the electro-neutrality condition:

\[
\text{total # of } + \text{ = total # of } -
\]

In intrinsic semiconductors the position of \( \mu \) is controlled by the DoS in \( CB \) and \( VB \) - \( \mu \) is closer to the band with lower DoS.

**Fig. 2.11** Fermi function, density of states, and spectral carrier densities \( q(E) (q = n, p) \) in an intrinsic semiconductor with \( m^* = 2m_e^* \). The temperature is \( \Theta = 0.25E_g/k_B \), which corresponds to a chemical potential \( \mu = 0.39E_g/2 \).
**Statistics of Electrons and Holes in Semiconductors**

**Intrinsic semiconductors** contain a negligibly small amount of impurities compared with thermally generated electrons and holes.

\[
n(T) = \int_{E_C}^{\infty} g_C(\varepsilon)f_C(\varepsilon)\,d\varepsilon
\]

- electron concentration, integration over the conduction band

\[
p(T) = \int_{-\infty}^{E_V} g_V(\varepsilon)f_V(\varepsilon)\,d\varepsilon
\]

- hole concentration, integration over the valence band

\[
g_C(\varepsilon) = \frac{m_e^*}{\pi^2\hbar^3} \sqrt{2m_e^*(\varepsilon - E_C)}
\]

\[
g_V(\varepsilon) = \frac{m_h^*}{\pi^2\hbar^3} \sqrt{2m_h^*(E_V - \varepsilon)}
\]

\[
f_C(\varepsilon) = \frac{1}{e^\beta(\varepsilon-\mu) + 1} \approx e^{-\beta(\varepsilon-\mu)}
\]

\[
f_V(\varepsilon) = \frac{1}{e^\beta(\mu-\varepsilon) + 1} \approx e^{-\beta(\mu-\varepsilon)} \quad \beta \equiv 1/k_B T
\]

(e.g., for intrinsic Si \( \frac{1}{2}(E_C - E_V) \approx 0.55eV \approx 21k_B(300K) \), the Boltzmann approximation is applicable).
If $\mu$ is close to the middle of the band gap, the Fermi systems of electrons and holes are non-degenerate (Boltzmann distributions, $\mu - \varepsilon \gg k_B T$):

$$n(T) \approx \int_{E_C}^{\infty} g_C(\varepsilon)e^{-\beta(\varepsilon - \mu)} \, d\varepsilon = \frac{m^*_e \sqrt{2m^*_e}}{\pi^2 \hbar^3} \int_{E_C}^{\infty} \sqrt{\varepsilon - E_C} e^{-\beta(\varepsilon - \mu)} \, d\varepsilon = \frac{1}{4} \left(\frac{2m^*_e k_B T}{\pi \hbar^2}\right)^{3/2} e^{-\beta(E_C - \mu)}$$

$$n(T) = \int_{E_C}^{\infty} g_C(\varepsilon)e^{-\beta(\varepsilon - \mu)} \, d\varepsilon = N_C e^{-\beta(E_C - \mu)}$$

$$p(T) = \int_{-\infty}^{E_V} g_V(\varepsilon)e^{-\beta(\mu - \varepsilon)} \, d\varepsilon = N_V e^{-\beta(\mu - E_V)}$$

Law of Mass Action

$$n(T)p(T) = \frac{1}{2} \left(\frac{k_B T}{\pi \hbar^2}\right)^3 (m^*_e m^*_h)^{3/2} e^{-\beta E_g} \equiv N_C N_V e^{-\beta E_g}$$

$$E_g \equiv E_C - E_V$$

This product is independent of doping provided that $\mu$ is in the band gap and $T$ is not too high.

Intrinsic ($i$) - charge carriers only due to excitation from $VB$ to $CB$ ($n_i = p_i$).
Example: Intrinsic Silicon

\[ n_i(T)p_i(T) = \frac{1}{2} \left( \frac{k_B T}{\pi \hbar^2} \right)^{3/2} e^{-\beta E_g} = \equiv N_C N_V e^{-\beta E_g} \]

At home: estimate \( E_g \) from the plot.
Does it match the expected \( E_g \approx 1.1\,eV \)?

\[ N_C \approx 2.8 \times 10^{25} \left( \frac{T}{300K} \right)^{1.5} \, m^{-3} \quad N_V \approx 1 \times 10^{25} \left( \frac{T}{300K} \right)^{1.5} \, m^{-3} \]

At 300K \( e^{-\beta E_g} = e^{-1.1/0.026} \approx 4 \times 10^{-19} \)

\[ n_i(300K)p_i(300K) \approx 1 \times 10^{32} \, m^{-6} \]

\[ n_i(300K) = p_i(300K) \approx 1 \times 10^{16} \, m^{-3} \]

This provides an order-of-magnitude estimate for “reasonable” purity of intrinsic Si: the concentration of dopants should not exceed \( 10^{15} \, m^{-3} \) to be negligible at \( T = 300K \).
\( \mu \) in Intrinsic Semiconductors \((n_i = p_i)\)

*Intrinsic* - charge carriers are created only by excitation from VB to CB \((n_i = p_i)\).

\[
n_i(T) = p_i(T) = \frac{1}{\sqrt{2}} \left( \frac{k_B T}{\pi \hbar^2} \right)^{3/2} (m_e^* m_h^*)^{3/4} e^{-\beta E_g / 2}
\]

\[
n_i(T)p_i(T) = N_C N_V e^{-\beta E_g}
\]

\[
1 = \frac{N_C}{N_V} e^{-\beta (E_V + E_C - 2\mu)}
\]

\[
\mu = \frac{1}{2} (E_V + E_C) + \frac{1}{2} (k_B T) \ln \left( \frac{N_V}{N_C} \right) = \frac{1}{2} (E_V + E_C) + \frac{3}{4} (k_B T) \ln \left( \frac{m_h^*}{m_e^*} \right)
\]

- satisfies the neutrality requirement, \(n = p\).

\( \mu \) is precisely in the middle of the band gap if \(T = 0\) or \(m_h^* = m_e^*\) \((N_C = N_V)\).

**Example.** How close to the mid-gap is the chemical potential in Si at \(T = 300K\)

(for DoS calculations: \(m_e^* \approx 1.08m_e\), \(m_{lh}^* \approx 0.81m_e\))

\[
\mu = \frac{E_g}{2} + \frac{3}{4} (k_B T) \ln \left( \frac{0.81}{1.08} \right) = 0.55 eV - 0.75 \times 0.026 eV \times 0.29 = 0.550 eV - 0.006 V
\]
Consider a semiconductor with concentrations of acceptors and donors $N_A$ and $N_D$. Assume sufficiently high $T$ that all dopants are ionized. The chemical potential can be found from the charge neutrality requirement:

$$n + N_A = p + N_D$$

Also $np = n_i^2$ (all dopants are ionized)

$$n = \frac{1}{2} \left[ N_D - N_A + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right]$$

$$p = \frac{1}{2} \left[ N_A - N_D + \sqrt{(N_A - N_D)^2 + 4n_i^2} \right]$$

$\approx N_D - N_A$ if $N_D - N_A \gg n_i$

$\approx N_A - N_D$ if $N_A - N_D \gg n_i$

Homework. Check that we still have $np = n_i^2$. 

![Graph showing carrier concentration in doped semiconductors](image)
**μ** in Doped Semiconductors

Strong $n$-doping pins $\mu$ near the bottom of the conduction band.

$$n = N_D$$
$$n = N_C e^{-\beta(E_C - \mu)}$$

$$\mu = E_C - k_B T \ln \left( \frac{N_D}{N_C} \right)$$

The electron density and the chemical potential of a doped semiconductor as a function of inverse temperature.
At sufficiently low $T$, the electrons/holes recombine with their “parents” (charged donors/acceptors) – **carrier freeze-out**.

How large is $E_C - \mu(T = 0)$ (left panel)?
**Problem.** Silicon is doped with donors, \( N_D = 10^{23} \text{m}^{-3} \). This donor concentration is much smaller than \( N_D = 10^{25} \text{m}^{-3} \) that corresponds to the degenerate case. Assume that all donors are ionized at room temperature. It is also given that \( n_i^2 = 10^{32} \text{m}^{-6} \).

What are the electron and hole concentrations at (a) 300K, and (b) 0K? Provide short explanations.

(a) 300K

\[
n = \frac{1}{2} \left[ N_D + \sqrt{(N_D)^2 + 4n_i^2} \right] \approx 10^{23} \text{m}^{-3}
\]

\[
p = \frac{1}{2} \left[ -N_D + \sqrt{(N_D)^2 + 4n_i^2} \right] \approx \frac{n_i^2}{N_D} = 10^9 \text{m}^{-3}
\]

(b) 0K  \( n = p = 0 \)
Conductivity of Doped Semiconductors

\[ \sigma = en\mu \quad \text{In metals, } n(T) = \text{const and } \sigma(T) \propto \mu(T). \]

In non-degenerate semiconductors, \( n(T) \propto e^{-\frac{E_C-\mu}{k_B T}} \) dominates in \( \sigma(T) \).

Electron mobility in doped Si

\[
\frac{1}{\mu} = \frac{1}{\mu_{ph}} + \frac{1}{\mu_{imp}}
\]
Summary

- Intrinsic and doped semiconductors
- Doping methods
- Law of Mass Action
- Chemical potential in semiconductors
- Conductivity of intrinsic and doped semiconductors

Today: Simon Ch. 17, Ibach and Luth 12.1-12.5
Next time: MOSFETs
Effective Mass for DoS Calculations (optional)


The effective mass for density of states calculations equals the mass which provides the density of states using the expression for one isotropic maximum or minimum or:

\[
g_c(E) = \frac{8\pi\sqrt{2}}{h^3} m_e^{3/2} \sqrt{E - E_c} \quad \text{for} \quad E \geq E_c
\]  

(2.3.12)

for the density of states in the conduction band and:

\[
g_v(E) = \frac{8\pi\sqrt{2}}{h^3} m_h^{3/2} \sqrt{E_v - E} \quad \text{for} \quad E \leq E_v
\]  

(2.3.13)

for the density of states in the valence band.

For instance for a single band minimum described by a longitudinal mass and two transverse masses the effective mass for density of states calculations is the geometric mean of the three masses. Including the fact that there are several equivalent minima at the same energy one obtains the effective mass for density of states calculations from:

\[
m_{e,dos}^* = M_c^{3/2} \sqrt[3]{m_l m_l m_t}
\]  

(2.3.14)

where \(M_c\) is the number of equivalent band minima. For silicon one obtains:

\[
m_{e,dos}^* = M_c^{3/2} \sqrt[3]{m_l m_l m_t} = 6^{3/2} \sqrt[3]{0.89 \times 0.19 \times 0.19 m_0} = 1.08 m_0
\]  

(2.3.15)

Effective Masses for Conductivity Calculations (optional)

The effective mass for conductivity calculation is the mass, which is used in conduction related calculations accounting for the detailed structure of the semiconductor. These calculations include mobility and diffusion constant calculations. Another example is the calculation of the shallow impurity levels using a hydrogen-like model.

As the conductivity of a material is inversionally proportional to the effective masses, one finds that the conductivity due to multiple band maxima or minima is proportional to the sum of the inverse of the individual masses, multiplied by the density of carriers in each band, as each maximum or minimum adds to the overall conductivity. For anisotropic minima containing one longitudinal and two transverse effective masses one has to sum over the effective masses in the different minima along the equivalent directions. The resulting effective mass for bands, which have ellipsoidal constant energy surfaces, is given by:

\[
m_{e,\text{cond}}^* = \frac{3}{\frac{1}{m_l} + \frac{1}{m_t} + \frac{1}{m_t}}
\]

provided the material has an isotropic conductivity as is the case for cubic materials. For instance electrons in the X minima of silicon have an effective conductivity mass given by:

\[
m_{e,\text{cond}}^* = \frac{3}{\frac{1}{m_l} + \frac{1}{m_t} + \frac{1}{m_t}} = \frac{3}{\frac{1}{0.89} + \frac{1}{0.19} + \frac{1}{0.19}} = 0.26m_0
\]