Semiconductor Devices

p-*n* Junctions: Part 4

M. B. Patil
mbpatil@ee.iitb.ac.in
www.ee.iitb.ac.in/~sequel

Department of Electrical Engineering
Indian Institute of Technology Bombay

Some implications of the Shockley diode equation

$$I = qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right)$$

Some implications of the Shockley diode equation

$$
\begin{aligned}
I &= qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p \tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n \tau_n}}\right)\left(e^{V_a/V_T} - 1\right)
\end{aligned}
$$

Some implications of the Shockley diode equation

$$
\begin{aligned}
I &= qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p\tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n\tau_n}}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right)\left(e^{V_a/V_T} - 1\right).
\end{aligned}
$$

Some implications of the Shockley diode equation

$$
\begin{aligned}
I &= qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p \tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n \tau_n}}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right)\left(e^{V_a/V_T} - 1\right).
\end{aligned}
$$

Different materials ($T = 300\,\mathrm{K}$):

Some implications of the Shockley diode equation

$$
\begin{aligned}
I &= qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p \tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n \tau_n}}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right)\left(e^{V_a/V_T} - 1\right).
\end{aligned}
$$

Different materials ($T = 300\,\text{K}$):

Semiconductor	N_c (cm^{-3})	N_v (cm^{-3})	E_g (eV)	n_i (cm^{-3})
Ge	1.04×10^{19}	6.0×10^{18}	0.664	2.33×10^{13}
Si	2.8×10^{19}	1.04×10^{19}	1.12	1.02×10^{10}
GaAs	4.7×10^{17}	7.0×10^{18}	1.424	2.1×10^{6}

Some implications of the Shockley diode equation

$$
\begin{aligned}
I &= qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p\tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n\tau_n}}\right)\left(e^{V_a/V_T} - 1\right) \\
&= qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right)\left(e^{V_a/V_T} - 1\right).
\end{aligned}
$$

Different materials ($T = 300\,\mathrm{K}$):

Semiconductor	N_c (cm^{-3})	N_v (cm^{-3})	E_g (eV)	n_i (cm^{-3})
Ge	1.04×10^{19}	6.0×10^{18}	0.664	2.33×10^{13}
Si	2.8×10^{19}	1.04×10^{19}	1.12	1.02×10^{10}
GaAs	4.7×10^{17}	7.0×10^{18}	1.424	2.1×10^{6}

Since $I \propto n_i^2$, the I-V curves (for the same diode area) are substantially different.

Some implications of the Shockley diode equation

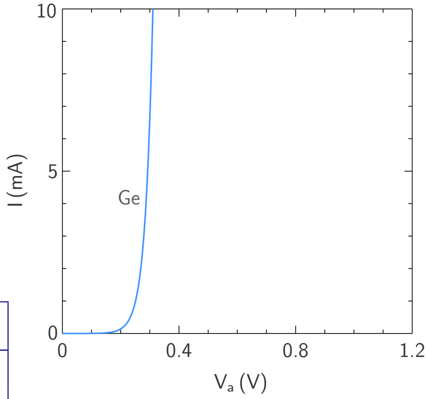
$$I = qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right)$$

$$= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p\tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n\tau_n}}\right)\left(e^{V_a/V_T} - 1\right)$$

$$= qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right)\left(e^{V_a/V_T} - 1\right).$$

Different materials ($T = 300\,\mathrm{K}$):

Semiconductor	N_c (cm^{-3})	N_v (cm^{-3})	E_g (eV)	n_i (cm^{-3})
Ge	1.04×10^{19}	6.0×10^{18}	0.664	2.33×10^{13}
Si	2.8×10^{19}	1.04×10^{19}	1.12	1.02×10^{10}
GaAs	4.7×10^{17}	7.0×10^{18}	1.424	2.1×10^{6}



Since $I \propto n_i^2$, the I-V curves (for the same diode area) are substantially different.

Some implications of the Shockley diode equation

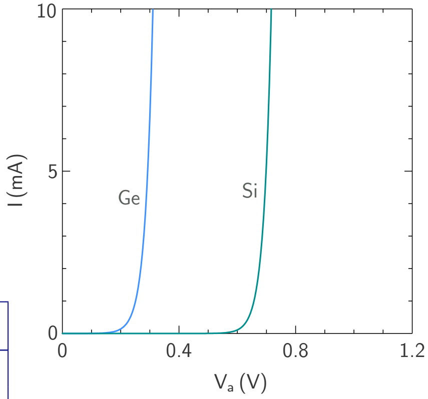
$$I = qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right)$$

$$= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p\tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n\tau_n}}\right)\left(e^{V_a/V_T} - 1\right)$$

$$= qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right)\left(e^{V_a/V_T} - 1\right).$$

Different materials ($T = 300\,\mathrm{K}$):

Semiconductor	N_c (cm^{-3})	N_v (cm^{-3})	E_g (eV)	n_i (cm^{-3})
Ge	1.04×10^{19}	6.0×10^{18}	0.664	2.33×10^{13}
Si	2.8×10^{19}	1.04×10^{19}	1.12	1.02×10^{10}
GaAs	4.7×10^{17}	7.0×10^{18}	1.424	2.1×10^{6}



Since $I \propto n_i^2$, the I-V curves (for the same diode area) are substantially different.

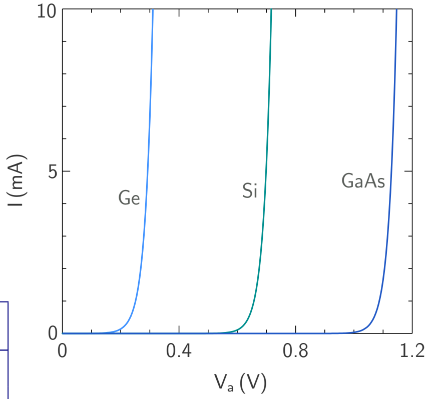
$$I = qA\left(\frac{D_p p_{n0}}{L_p} + \frac{D_n n_{p0}}{L_n}\right)\left(e^{V_a/V_T} - 1\right)$$

$$= qA\left(\frac{D_p n_i^2}{N_d\sqrt{D_p\tau_p}} + \frac{D_n n_i^2}{N_a\sqrt{D_n\tau_n}}\right)\left(e^{V_a/V_T} - 1\right)$$

$$= qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right)\left(e^{V_a/V_T} - 1\right).$$

Different materials ($T = 300\,\mathrm{K}$):

Semiconductor	N_c (cm^{-3})	N_v (cm^{-3})	E_g (eV)	n_i (cm^{-3})
Ge	1.04×10^{19}	6.0×10^{18}	0.664	2.33×10^{13}
Si	2.8×10^{19}	1.04×10^{19}	1.12	1.02×10^{10}
GaAs	4.7×10^{17}	7.0×10^{18}	1.424	2.1×10^{6}



Since $I \propto n_i^2$, the I-V curves (for the same diode area) are substantially different.

Temperature dependence

$I = I_s \left(e^{V_a/V_T} - 1 \right)$, where

$$I_s = qA \left(\frac{n_i^2}{N_d} \sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a} \sqrt{\frac{D_n}{\tau_n}} \right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.

$\boxed{\text{Temperature dependence}}$

$I = I_s \left(e^{V_a/V_T} - 1 \right)$, where

$$I_s = qA \left(\frac{n_i^2}{N_d} \sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a} \sqrt{\frac{D_n}{\tau_n}} \right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.

$$n_i(T) = \sqrt{N_c(T) N_v(T)} \exp\left(-\frac{E_g(T)}{2kT} \right),$$

where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

Temperature dependence

$I = I_s \left(e^{V_a/V_T} - 1 \right)$, where

$$I_s = qA \left(\frac{n_i^2}{N_d} \sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a} \sqrt{\frac{D_n}{\tau_n}} \right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.

$$n_i(T) = \sqrt{N_c(T) N_v(T)} \, \exp \left(- \frac{E_g(T)}{2kT} \right),$$

where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T + \beta}$ (eV),

with $E_g(0) = 1.17\,\text{eV}$, $\alpha = 4.73 \times 10^{-4}$, $\beta = 636$.

Temperature dependence

$I = I_s\left(e^{V_a/V_T} - 1\right)$, where

$$I_s = qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right).$$

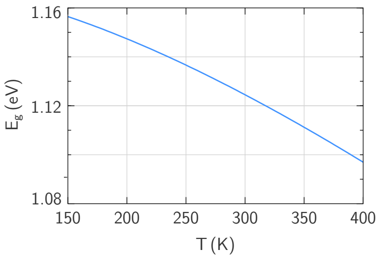
The temperature dependence of I_s comes mainly from $n_i(T)$.

$$n_i(T) = \sqrt{N_c(T)N_v(T)}\,\exp\left(-\frac{E_g(T)}{2kT}\right),$$

where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T+\beta}$ (eV),

with $E_g(0) = 1.17\,\text{eV}$, $\alpha = 4.73\times 10^{-4}$, $\beta = 636$.

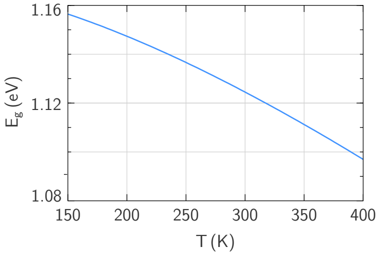


Temperature dependence

$I = I_s \left(e^{V_a/V_T} - 1\right)$, where

$$I_s = qA \left(\frac{n_i^2}{N_d} \sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a} \sqrt{\frac{D_n}{\tau_n}} \right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.



$$n_i(T) = \sqrt{N_c(T) N_v(T)} \exp\left(-\frac{E_g(T)}{2kT} \right),$$

where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T + \beta}$ (eV),

with $E_g(0) = 1.17$ eV, $\alpha = 4.73 \times 10^{-4}$, $\beta = 636$.

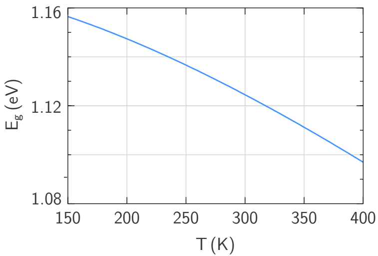
As T increases, $E_g/2kT$ decreases, and n_i increases substantially because of the exponential function $\rightarrow$ I increases.

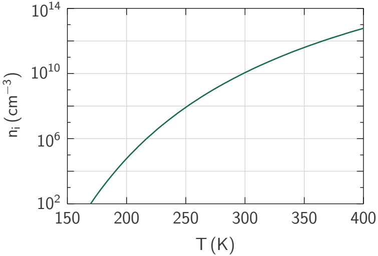
Temperature dependence

$I = I_s\left(e^{V_a/V_T} - 1\right)$, where

$$I_s = qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.





$$n_i(T) = \sqrt{N_c(T)N_v(T)}\,\exp\left(-\frac{E_g(T)}{2kT}\right),$$

where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T+\beta}$ (eV),

with $E_g(0) = 1.17\,\text{eV}$, $\alpha = 4.73\times10^{-4}$, $\beta = 636$.

As T increases, $E_g/2kT$ decreases, and n_i increases substantially because of the exponential function $\rightarrow$ I increases.

Temperature dependence

$I = I_s\left(e^{V_a/V_T} - 1\right)$, where

$$I_s = qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.

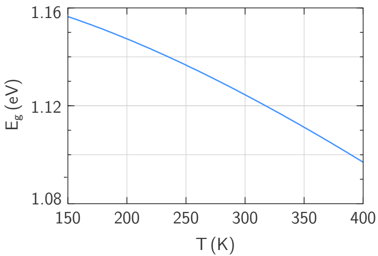
$$n_i(T) = \sqrt{N_c(T)N_v(T)}\,\exp\left(-\frac{E_g(T)}{2kT}\right),$$

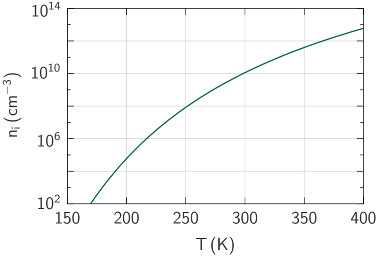
where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

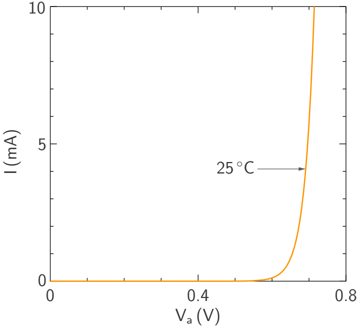
$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T+\beta}$ (eV),

with $E_g(0) = 1.17\,\text{eV}$, $\alpha = 4.73 \times 10^{-4}$, $\beta = 636$.

As T increases, $E_g/2kT$ decreases, and n_i increases substantially because of the exponential function $\rightarrow$ I increases.







Temperature dependence

$I = I_s\left(e^{V_a/V_T} - 1\right)$, where

$$I_s = qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.

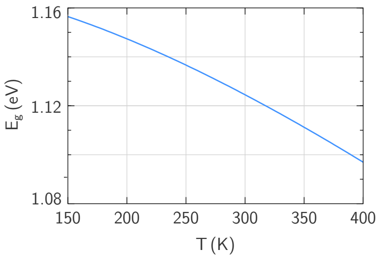
$$n_i(T) = \sqrt{N_c(T)N_v(T)}\,\exp\left(-\frac{E_g(T)}{2kT}\right),$$

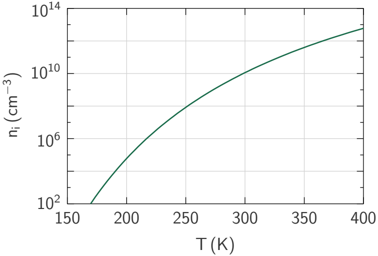
where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

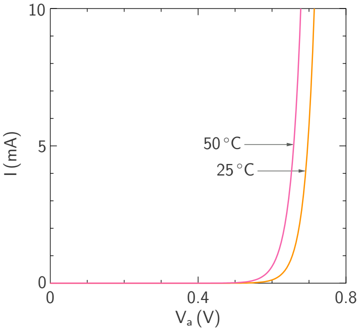
$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T+\beta}$ (eV),

with $E_g(0) = 1.17$ eV, $\alpha = 4.73\times10^{-4}$, $\beta = 636$.

As T increases, $E_g/2kT$ decreases, and n_i increases substantially because of the exponential function $\rightarrow$ I increases.







Temperature dependence

$I = I_s \left(e^{V_a/V_T} - 1\right)$, where

$$I_s = qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.

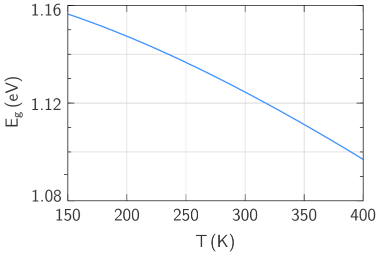
$$n_i(T) = \sqrt{N_c(T)N_v(T)}\,\exp\left(-\frac{E_g(T)}{2kT}\right),$$

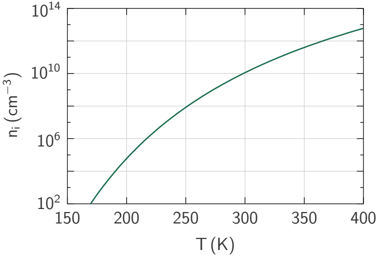
where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

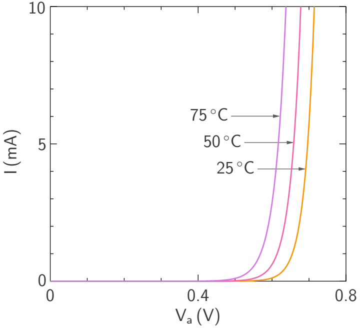
$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T+\beta}$ (eV),

with $E_g(0) = 1.17\,\text{eV}$, $\alpha = 4.73\times 10^{-4}$, $\beta = 636$.

As T increases, $E_g/2kT$ decreases, and n_i increases substantially because of the exponential function $\rightarrow$ I increases.





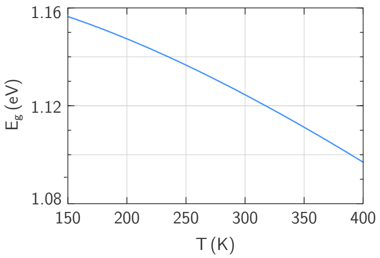


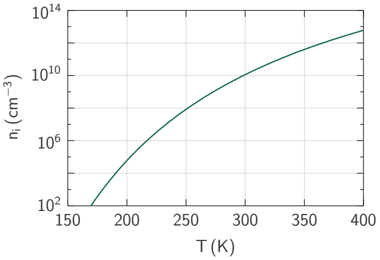
Temperature dependence

$I = I_s\left(e^{V_a/V_T} - 1\right)$, where

$$I_s = qA\left(\frac{n_i^2}{N_d}\sqrt{\frac{D_p}{\tau_p}} + \frac{n_i^2}{N_a}\sqrt{\frac{D_n}{\tau_n}}\right).$$

The temperature dependence of I_s comes mainly from $n_i(T)$.





$$n_i(T) = \sqrt{N_c(T)N_v(T)}\,\exp\left(-\frac{E_g(T)}{2kT}\right),$$

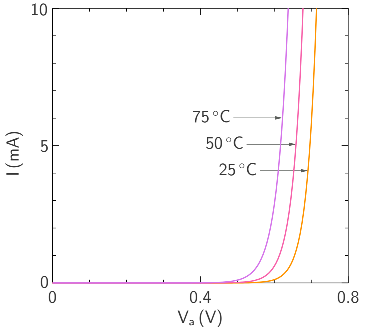
where $N_c \propto T^{3/2}$, $N_v \propto T^{3/2}$.

$E_g(T)$ for silicon is given by $E_g(T) = E_g(0) - \dfrac{\alpha T^2}{T+\beta}$ (eV),

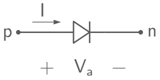
with $E_g(0) = 1.17$ eV, $\alpha = 4.73 \times 10^{-4}$, $\beta = 636$.

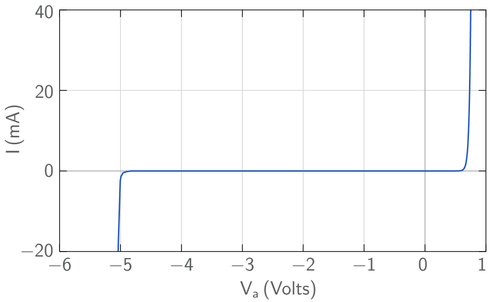
As T increases, $E_g/2kT$ decreases, and n_i increases substantially because of the exponential function $\rightarrow$ I increases.

For silicon, the I-V curve shifts by about $-2\,\text{mV}/^\circ\text{C}$ as the temperature is increased.

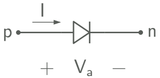


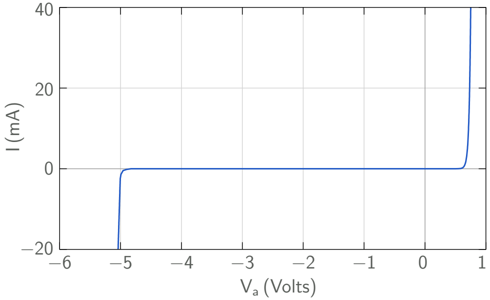
Reverse breakdown





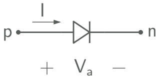
Reverse breakdown

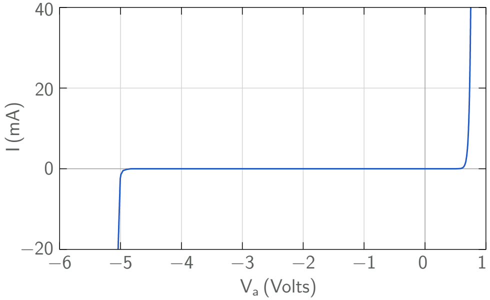




* A real diode cannot withstand indefinitely large reverse voltages and "breaks down" at some point as V_R is increased.

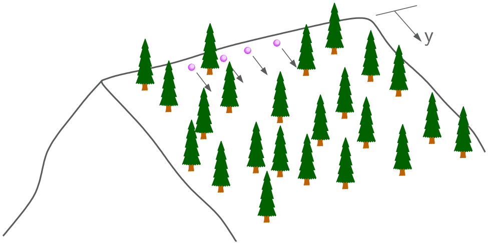
Reverse breakdown

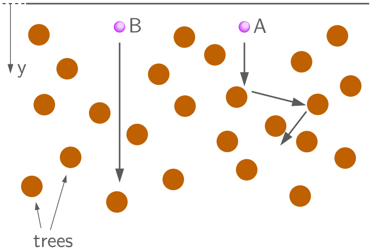




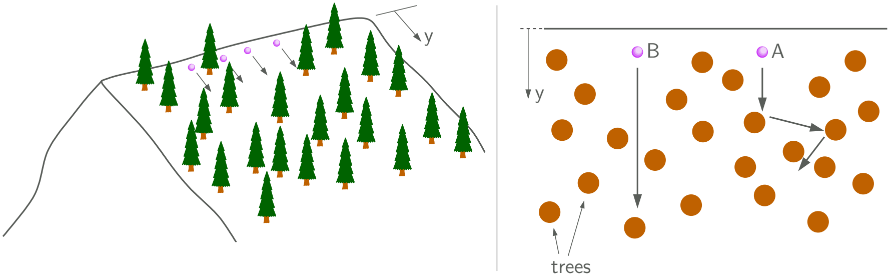
* A real diode cannot withstand indefinitely large reverse voltages and "breaks down" at some point as V_R is increased.
* Reverse breakdown can be due to
  - impact ionisation (avalanche breakdown)
  - tunneling (Zener breakdown)

Avalanche breakdown: impact ionisation



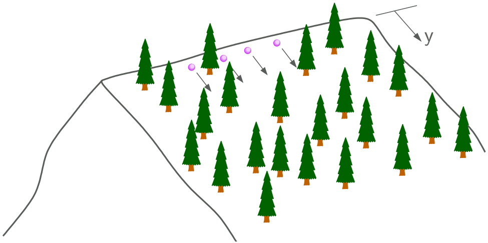


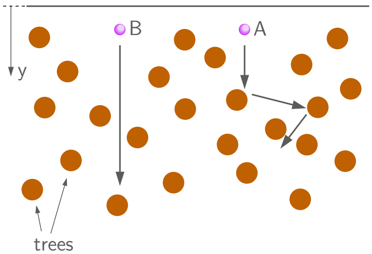
Consider spherical objects starting down from the top.



Consider spherical objects starting down from the top.

* When an objects hits a tree, it loses some of its kinetic energy and starts downhill again with a reduced kinetic energy.

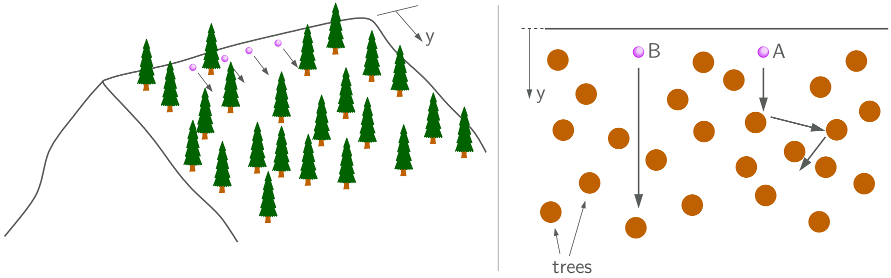




Consider spherical objects starting down from the top.

* When an objects hits a tree, it loses some of its kinetic energy and starts downhill again with a reduced kinetic energy.
* There could be a "lucky" object which does not encounter a collision and therefore gains a higher kinetic energy as compared to other objects.

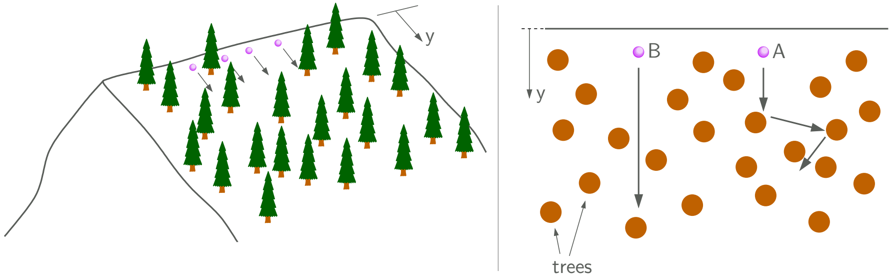
Avalanche breakdown: impact ionisation



Consider spherical objects starting down from the top.

* When an objects hits a tree, it loses some of its kinetic energy and starts downhill again with a reduced kinetic energy.
* There could be a "lucky" object which does not encounter a collision and therefore gains a higher kinetic energy as compared to other objects.
* The slope plays a crucial role. Higher the slope, higher is the probablity that an object will gain a high kinetic energy.

Avalanche breakdown: impact ionisation

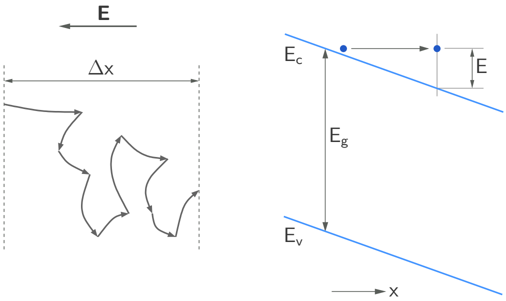


Consider spherical objects starting down from the top.

* When an objects hits a tree, it loses some of its kinetic energy and starts downhill again with a reduced kinetic energy.
* There could be a "lucky" object which does not encounter a collision and therefore gains a higher kinetic energy as compared to other objects.
* The slope plays a crucial role. Higher the slope, higher is the probablity that an object will gain a high kinetic energy.

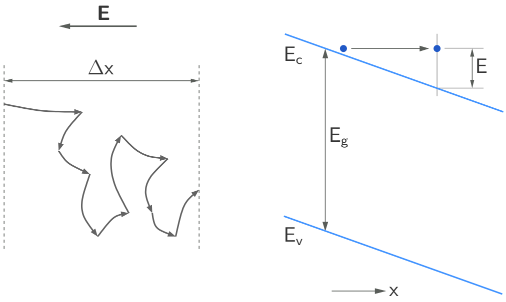
With this situation in mind, let us look at carrier transport in a semiconductor.

Avalanche breakdown: impact ionisation



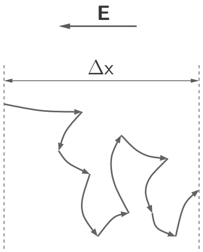
* If an electron moves some distance without scattering, it gains an energy E with respect to E_c at that location. This energy can be thought of as the kinetic energy of the electron.

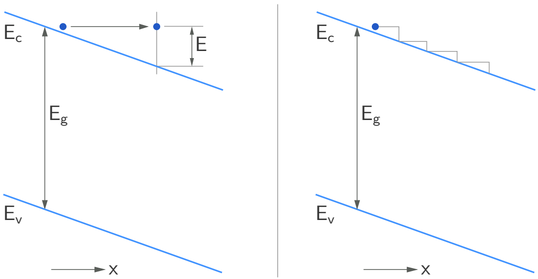
Avalanche breakdown: impact ionisation



* If an electron moves some distance without scattering, it gains an energy E with respect to E_c at that location. This energy can be thought of as the kinetic energy of the electron.
* Because of scattering events (phonons, impurity ions, etc.), an electron keeps losing the energy gained from the electric field, and it is constrained to stay near E_c.

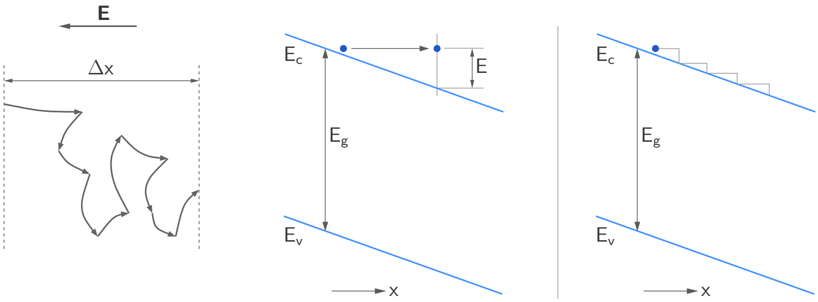
Avalanche breakdown: impact ionisation





* If an electron moves some distance without scattering, it gains an energy E with respect to E_c at that location. This energy can be thought of as the kinetic energy of the electron.
* Because of scattering events (phonons, impurity ions, etc.), an electron keeps losing the energy gained from the electric field, and it is constrained to stay near E_c.

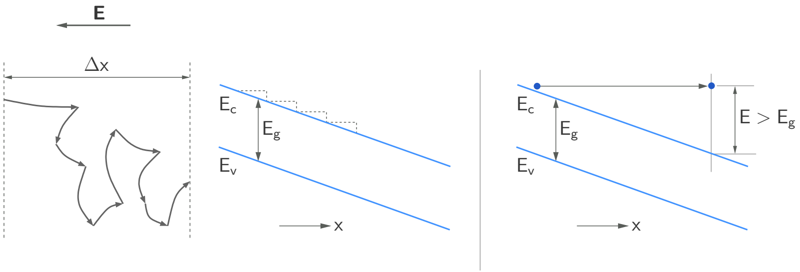
Avalanche breakdown: impact ionisation



* If an electron moves some distance without scattering, it gains an energy E with respect to E_c at that location. This energy can be thought of as the kinetic energy of the electron.
* Because of scattering events (phonons, impurity ions, etc.), an electron keeps losing the energy gained from the electric field, and it is constrained to stay near E_c.

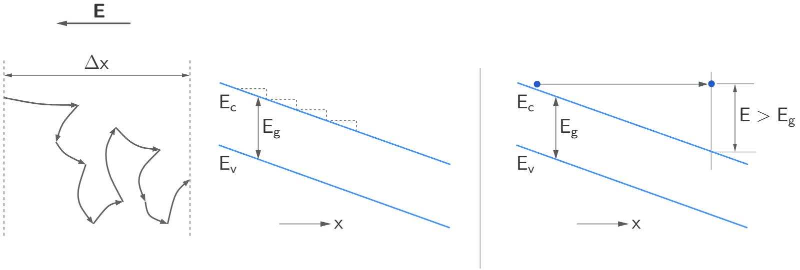
  (Note: For simplicity, we have not discussed the changes in the electron momentum in the other two directions.)

Avalanche breakdown: impact ionisation



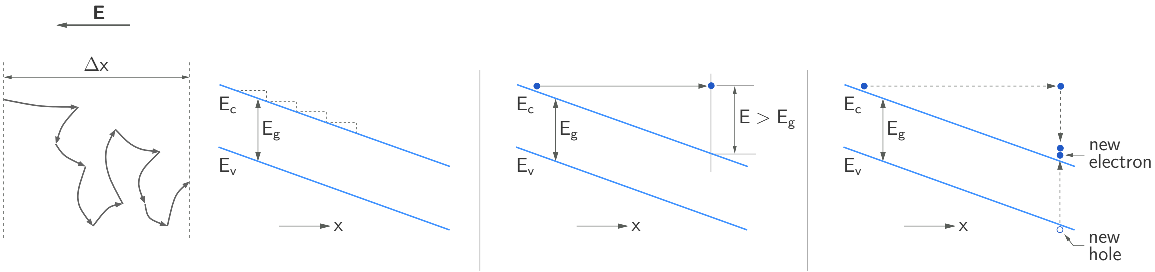
* A "lucky" electron which undergoes fewer scattering events can gain substantial energy.

Avalanche breakdown: impact ionisation



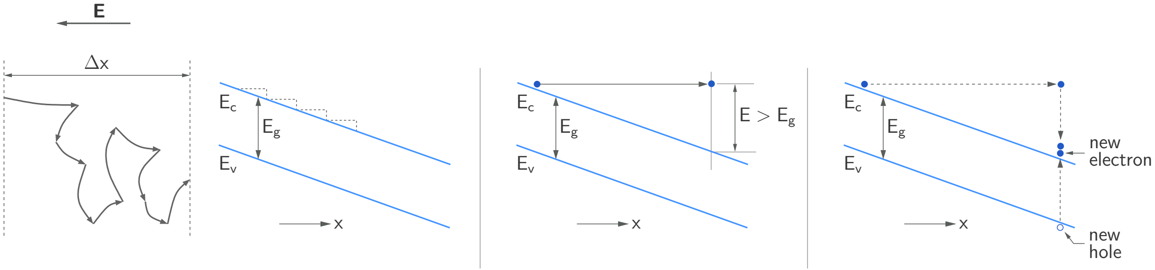
* A "lucky" electron which undergoes fewer scattering events can gain substantial energy.
* When E exceeds E_g (or 1.5 E_g for some semiconductors), impact ionisation can occur.

Avalanche breakdown: impact ionisation



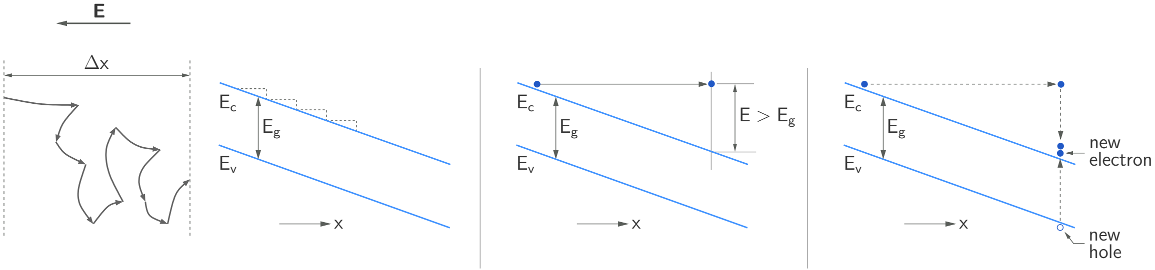
* A "lucky" electron which undergoes fewer scattering events can gain substantial energy.
* When E exceeds E_g (or 1.5 E_g for some semiconductors), impact ionisation can occur.

Avalanche breakdown: impact ionisation



* A "lucky" electron which undergoes fewer scattering events can gain substantial energy.
* When E exceeds E_g (or 1.5 E_g for some semiconductors), impact ionisation can occur.
* Impact ionisation: The energy lost by an electron in the conduction band is used by an electron in the valence band to make a transition to the conduction band.

Avalanche breakdown: impact ionisation

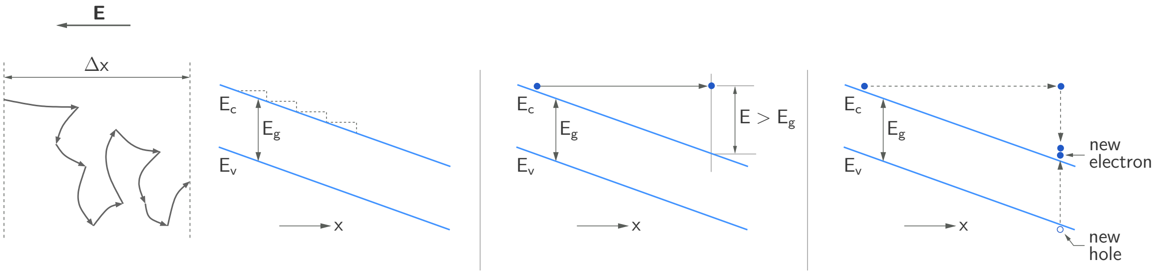


* A "lucky" electron which undergoes fewer scattering events can gain substantial energy.
* When E exceeds E_g (or 1.5 E_g for some semiconductors), impact ionisation can occur.
* Impact ionisation: The energy lost by an electron in the conduction band is used by an electron in the valence band to make a transition to the conduction band.

  Before impact ionisation: one electron

  After impact ionisation: two electrons and one hole

Avalanche breakdown: impact ionisation

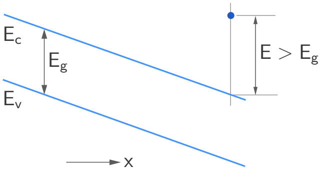


* A "lucky" electron which undergoes fewer scattering events can gain substantial energy.
* When E exceeds E_g (or 1.5 E_g for some semiconductors), impact ionisation can occur.
* Impact ionisation: The energy lost by an electron in the conduction band is used by an electron in the valence band to make a transition to the conduction band.

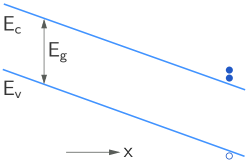
  Before impact ionisation: one electron

  After impact ionisation: two electrons and one hole
* The key requirement for impact ionisation to occur is a high electric field.

Avalanche breakdown: impact ionisation

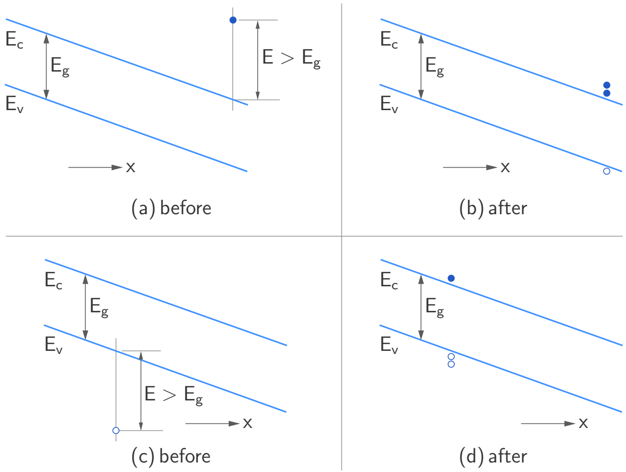


(a) before

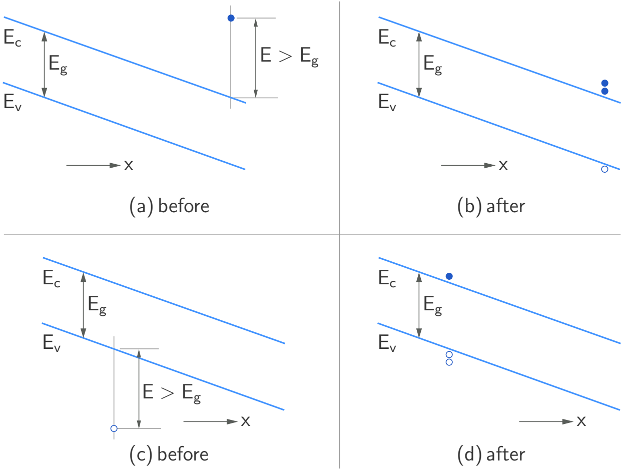


(b) after

Avalanche breakdown: impact ionisation

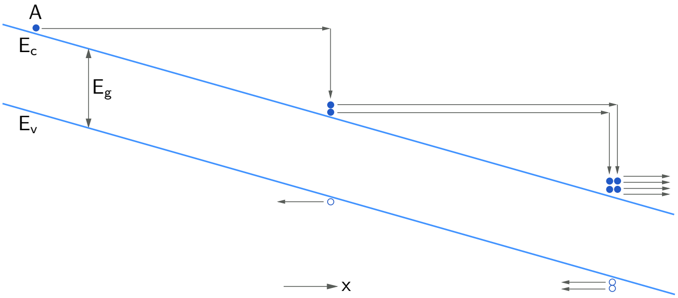


Avalanche breakdown: impact ionisation

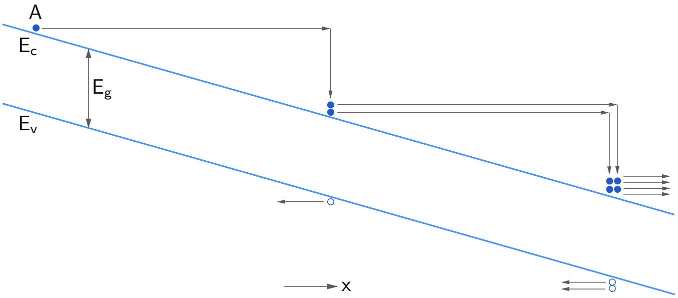


* Impact ionisation can be caused by a high-energy electron or a high-energy hole.

Avalanche breakdown: impact ionisation

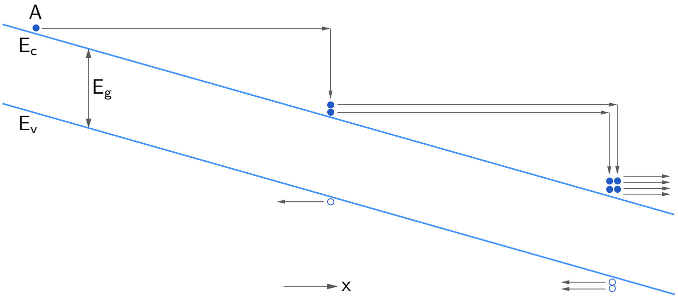


Avalanche breakdown: impact ionisation



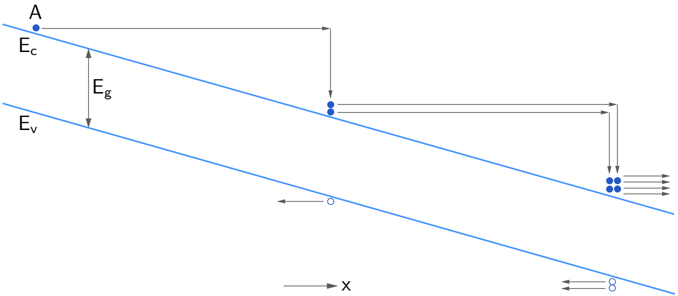
* A lucky electron gains energy, causes impact ionisation and gives rise to two electrons and one hole.

Avalanche breakdown: impact ionisation



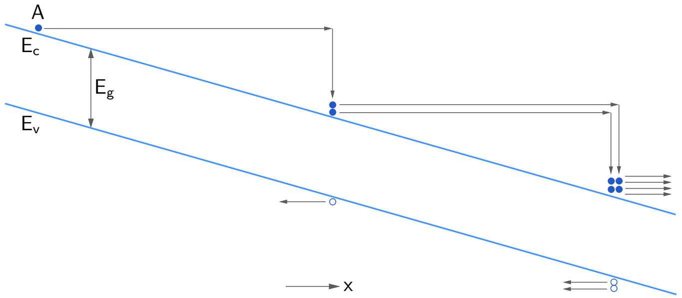
* A lucky electron gains energy, causes impact ionisation and gives rise to two electrons and one hole.
* Each of these carriers can potentially cause impact ionisation because of the high electric field (if they travel without scattering) → "avalanche" multiplication.

Avalanche breakdown: impact ionisation



* A lucky electron gains energy, causes impact ionisation and gives rise to two electrons and one hole.
* Each of these carriers can potentially cause impact ionisation because of the high electric field (if they travel without scattering) $\rightarrow$ "avalanche" multiplication.
* The avalanche multiplication process is obviously very complicated, but it can be characterised in a simple manner by a "critical field" $\mathcal{E}_c$ — if the electric field in the device exceeds $\mathcal{E}_c$, we can expect avalanche multiplication, leading to large currents.

Avalanche breakdown: impact ionisation



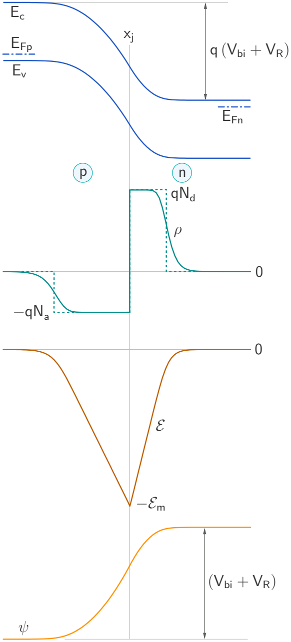
* A lucky electron gains energy, causes impact ionisation and gives rise to two electrons and one hole.
* Each of these carriers can potentially cause impact ionisation because of the high electric field (if they travel without scattering) → "avalanche" multiplication.
* The avalanche multiplication process is obviously very complicated, but it can be characterised in a simple manner by a "critical field" $\mathcal{E}_c$ — if the electric field in the device exceeds $\mathcal{E}_c$, we can expect avalanche multiplication, leading to large currents.
* For Ge, Si, and GaAs, the critical field at room temperature is about 100, 300, and 400 kV/cm, respectively.

Example

For a p^+n silicon diode, estimate the doping density required on the n side for a breakdown voltage of 50 V, given that $\mathcal{E}_c = 250\,\text{kV/cm}$.

Example

For a p^+n silicon diode, estimate the doping density required on the n side for a breakdown voltage of 50 V, given that $\mathcal{E}_c = 250\,\text{kV/cm}$.

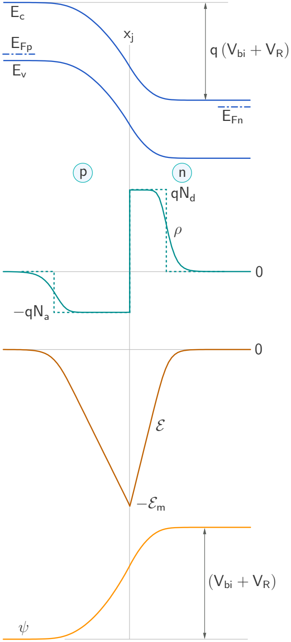


Example

For a p^+n silicon diode, estimate the doping density required on the n side for a breakdown voltage of 50 V, given that $\mathcal{E}_c = 250\,\text{kV/cm}$.

Solution: The maximum electric field occurs at the junction and is given by

$$\mathcal{E}_m = \frac{2(V_{\text{bi}} + V_R)}{W} = \frac{2(V_{\text{bi}} + V_R)}{\sqrt{\dfrac{2\epsilon}{q}\,\dfrac{N_a + N_d}{N_a N_d}\,(V_{\text{bi}} + V_R)}} \approx \sqrt{\frac{2qN_d(V_{\text{bi}} + V_R)}{\epsilon}}.$$



Example

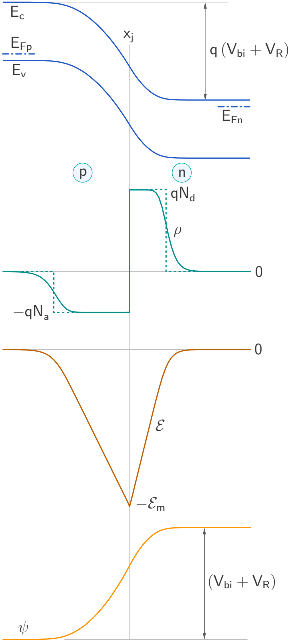
For a p^+n silicon diode, estimate the doping density required on the n side for a breakdown voltage of 50 V, given that $\mathcal{E}_c = 250\,\text{kV/cm}$.

Solution: The maximum electric field occurs at the junction and is given by

$$\mathcal{E}_m = \frac{2(V_{\text{bi}} + V_R)}{W} = \frac{2(V_{\text{bi}} + V_R)}{\sqrt{\dfrac{2\epsilon}{q}\,\dfrac{N_a + N_d}{N_a N_d}\,(V_{\text{bi}} + V_R)}} \approx \sqrt{\frac{2qN_d(V_{\text{bi}} + V_R)}{\epsilon}}.$$

Since $V_R \gg V_{\text{bi}}$ (which is less than 1 V), we get

$$N_d = \frac{\epsilon\,\mathcal{E}_m^2}{2qV_R} = \frac{\epsilon\,\mathcal{E}_c^2}{2qV_R} = \frac{11.7 \times 8.85 \times 10^{-14} \times (2.5 \times 10^5)^2}{2 \times 1.6 \times 10^{-19} \times 50} = 4 \times 10^{15}\,\text{cm}^{-3}.$$



Example

For a p^+n silicon diode, estimate the doping density required on the n side for a breakdown voltage of 50 V, given that $\mathcal{E}_c = 250\,\text{kV/cm}$.

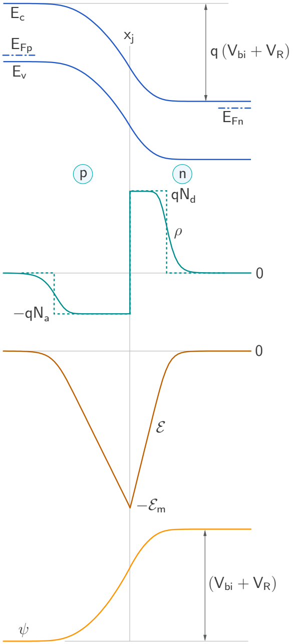
Solution: The maximum electric field occurs at the junction and is given by

$$\mathcal{E}_m = \frac{2(V_{\text{bi}} + V_R)}{W} = \frac{2(V_{\text{bi}} + V_R)}{\sqrt{\dfrac{2\epsilon}{q}\,\dfrac{N_a + N_d}{N_a N_d}\,(V_{\text{bi}} + V_R)}} \approx \sqrt{\frac{2qN_d(V_{\text{bi}} + V_R)}{\epsilon}}.$$

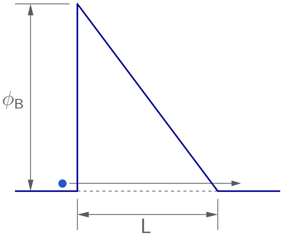
Since $V_R \gg V_{\text{bi}}$ (which is less than 1 V), we get

$$N_d = \frac{\epsilon\,\mathcal{E}_m^2}{2qV_R} = \frac{\epsilon\,\mathcal{E}_c^2}{2qV_R} = \frac{11.7 \times 8.85 \times 10^{-14} \times (2.5 \times 10^5)^2}{2 \times 1.6 \times 10^{-19} \times 50} = 4 \times 10^{15}\,\text{cm}^{-3}.$$

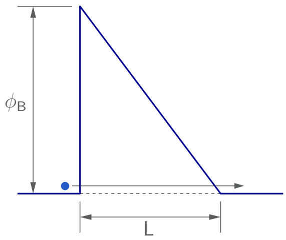
$$\text{Units: } \frac{\text{F}}{\text{cm}}\left(\frac{\text{V}}{\text{cm}}\right)^2 \frac{1}{\text{C}}\,\frac{1}{\text{V}} = \frac{\text{C}}{\text{V}\cdot\text{cm}}\left(\frac{\text{V}}{\text{cm}}\right)^2 \frac{1}{\text{C}}\,\frac{1}{\text{V}} = \frac{1}{\text{cm}^3}.$$



Zener breakdown

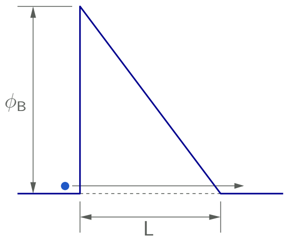


Zener breakdown



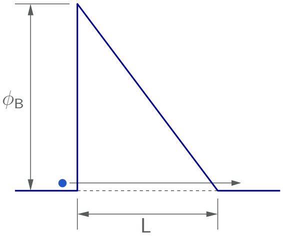
* Another mechanism which can cause reverse breakdown of a *pn* junction is quantum-mechanical tunneling.

Zener breakdown



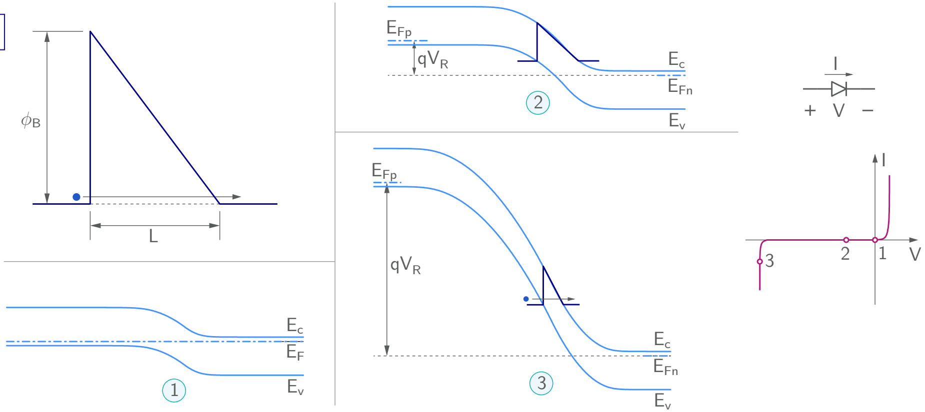
* Another mechanism which can cause reverse breakdown of a *pn* junction is quantum-mechanical tunneling.
* For a significant tunneling current,
  - The barrier width (L) must be small, typically few nanometers.
  - There must be a large number of electrons on one side and a large number of available states on the other side.

Zener breakdown

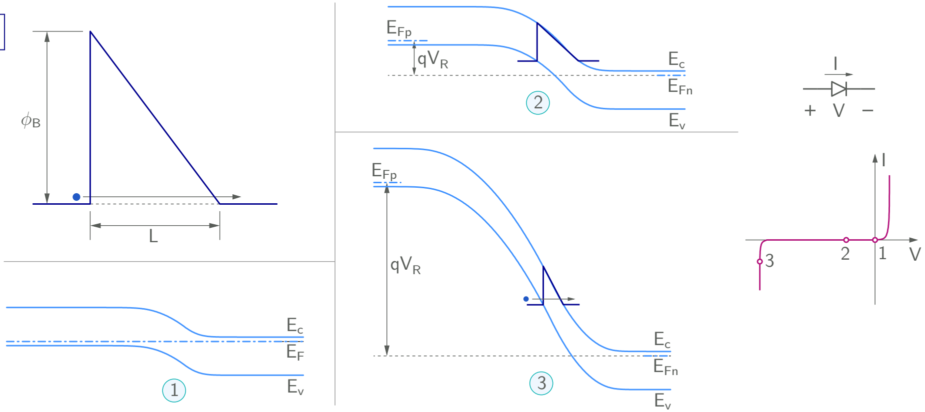


* Another mechanism which can cause reverse breakdown of a *pn* junction is quantum-mechanical tunneling.
* For a significant tunneling current,
  - The barrier width (L) must be small, typically few nanometers.
  - There must be a large number of electrons on one side and a large number of available states on the other side.
* With these points in mind, let us now look at a reverse-biased *pn* junction.

Zener breakdown

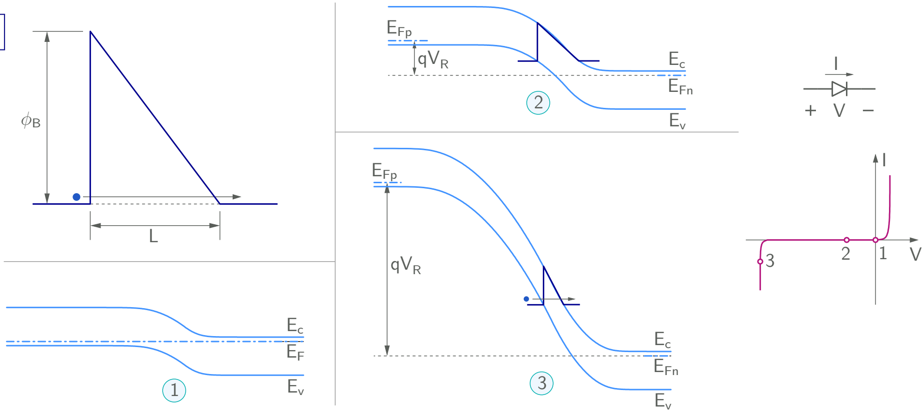


Zener breakdown



* The depletion width is given by $W = \sqrt{\dfrac{2\epsilon}{q}\left(\dfrac{1}{N_a}+\dfrac{1}{N_d}\right)(V_{bi}+V_R)}\,.$

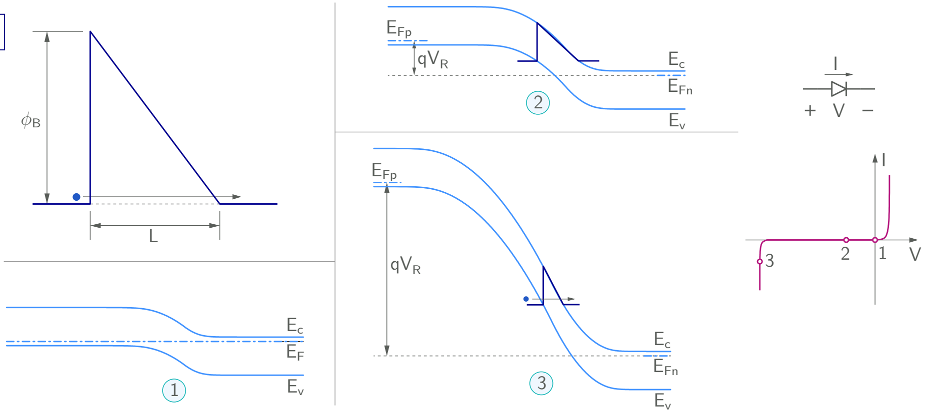
Zener breakdown



* The depletion width is given by $W = \sqrt{\dfrac{2\epsilon}{q}\left(\dfrac{1}{N_a}+\dfrac{1}{N_d}\right)(V_{bi}+V_R)}\,.$

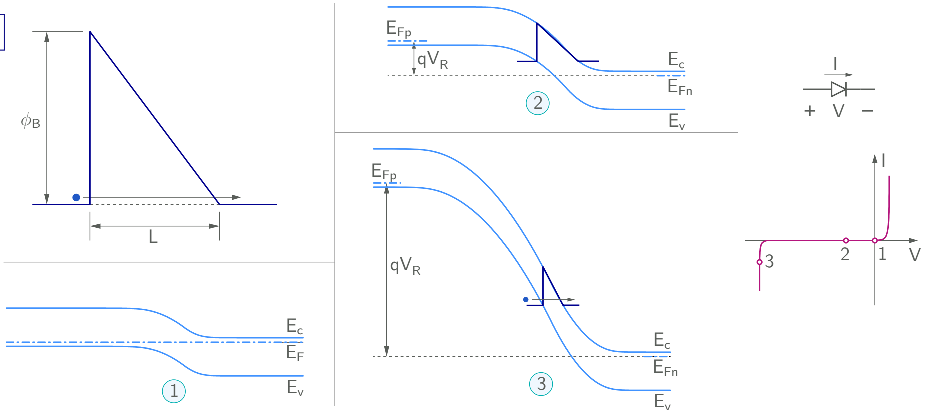
* As the reverse bias is increased, the barrier becomes thinner.

Zener breakdown



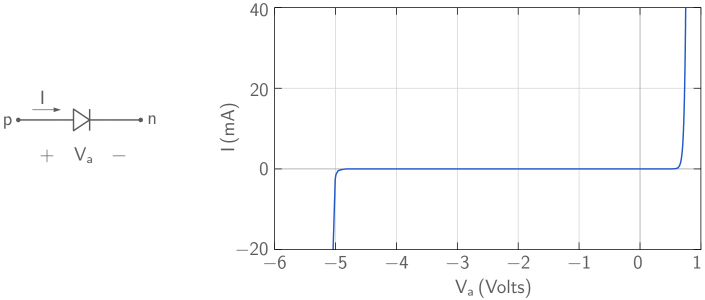
* The depletion width is given by $W = \sqrt{\dfrac{2\epsilon}{q}\left(\dfrac{1}{N_a}+\dfrac{1}{N_d}\right)(V_{bi}+V_R)}$.
* As the reverse bias is increased, the barrier becomes thinner.
* There is a large number of electrons in the valence band on the *p* side and a large number of states (vacancies) in the conduction band on the *n* side.

Zener breakdown

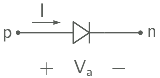


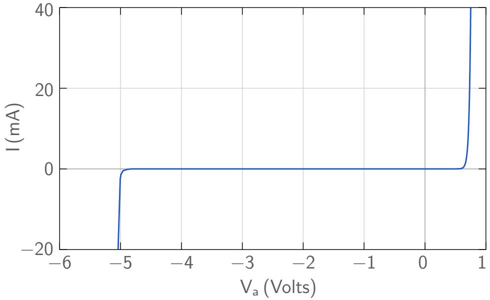
* The depletion width is given by $W = \sqrt{\dfrac{2\epsilon}{q}\left(\dfrac{1}{N_a}+\dfrac{1}{N_d}\right)(V_{bi}+V_R)}\,.$
* As the reverse bias is increased, the barrier becomes thinner.
* There is a large number of electrons in the valence band on the *p* side and a large number of states (vacancies) in the conduction band on the *n* side.
* Relatively large doping densities are required to ensure that the barrier is sufficiently thin for tunneling to occur.

Reverse breakdown

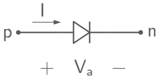


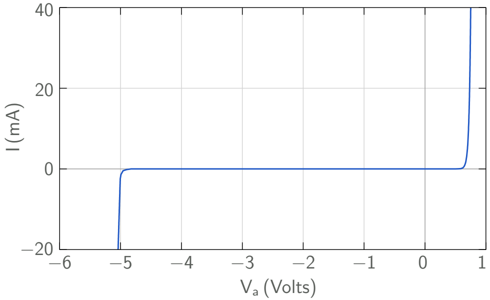
Reverse breakdown





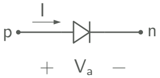
* Silicon diodes with breakdown voltage V_{BR} ranging from a few volts to a thousand volts are commercially available.

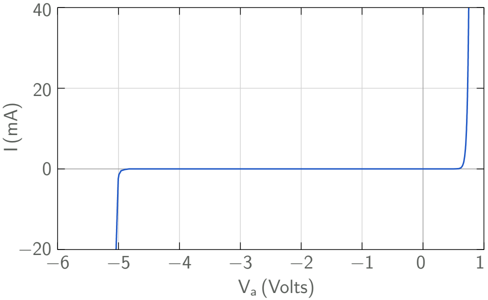




* Silicon diodes with breakdown voltage V_{BR} ranging from a few volts to a thousand volts are commercially available.
* In diodes with $V_{BR} < 5\,V$, the breakdown is generally due to tunnelling; for higher values of V_{BR}, it is due to avalanche multiplication.

Reverse breakdown





* Silicon diodes with breakdown voltage V_{BR} ranging from a few volts to a thousand volts are commercially available.
* In diodes with $V_{BR} < 5\,V$, the breakdown is generally due to tunnelling; for higher values of V_{BR}, it is due to avalanche multiplication.
* In some diodes (with $V_{BR} \simeq 5\,V$), it is possible that both mechanisms are active simultaneously.

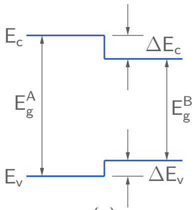
Heterojunctions

* The *pn* junctions we have considered so far are called "homojunctions," i.e., junctions between similar (same) semiconductors on the *p* and *n* sides.

* The *pn* junctions we have considered so far are called "homojunctions," i.e., junctions between similar (same) semiconductors on the *p* and *n* sides.
* In a *pn* "heterojunction," two different semiconductors A and B are involved, where A is doped *p*-type, and B is doped *n*-type.

* The *pn* junctions we have considered so far are called “homojunctions,” i.e., junctions between similar (same) semiconductors on the *p* and *n* sides.
* In a *pn* “heterojunction,” two different semiconductors A and B are involved, where A is doped *p*-type, and B is doped *n*-type.
* The two semiconductors must be lattice-matched, i.e., they must have the same lattice constant to avoid dislocations and device degradation.

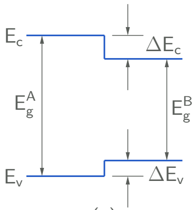
Heterojunctions



(a)

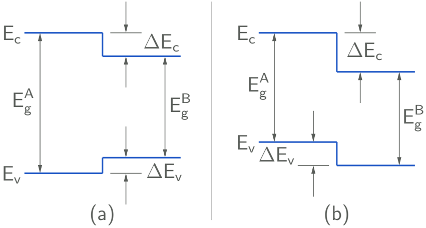
* Because A and B are different semiconductors, there are discontinuities in the band edges (ΔE_c and ΔE_v).

Heterojunctions

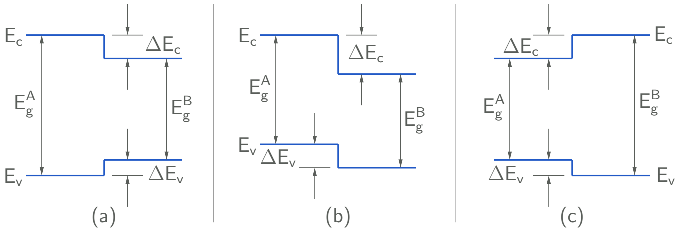


(a)

* Because A and B are different semiconductors, there are discontinuities in the band edges (ΔE_c and ΔE_v).
* Depending on the material properties of A and B, different signs for ΔE_c and ΔE_v are possible.

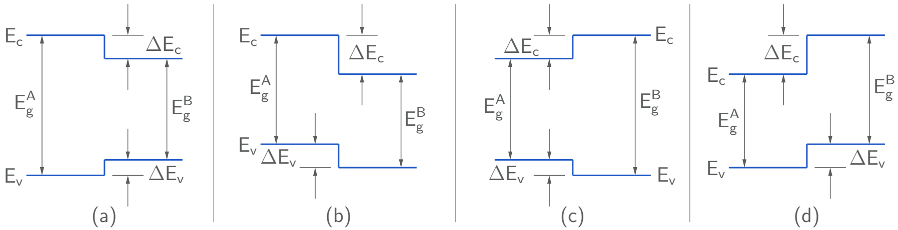


* Because A and B are different semiconductors, there are discontinuities in the band edges (ΔE_c and ΔE_v).
* Depending on the material properties of A and B, different signs for ΔE_c and ΔE_v are possible.

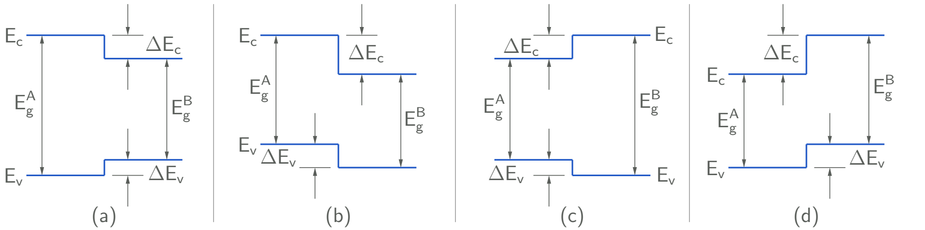


* Because A and B are different semiconductors, there are discontinuities in the band edges (ΔE_c and ΔE_v).
* Depending on the material properties of A and B, different signs for ΔE_c and ΔE_v are possible.

Heterojunctions



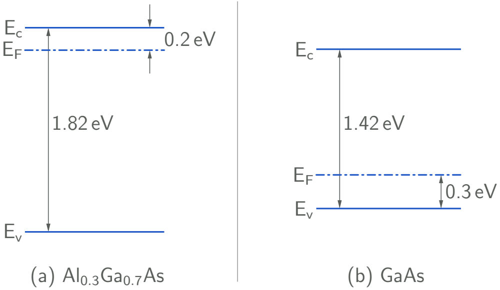
* Because A and B are different semiconductors, there are discontinuities in the band edges (ΔE_c and ΔE_v).
* Depending on the material properties of A and B, different signs for ΔE_c and ΔE_v are possible.



* Because A and B are different semiconductors, there are discontinuities in the band edges (ΔE_c and ΔE_v).
* Depending on the material properties of A and B, different signs for ΔE_c and ΔE_v are possible.

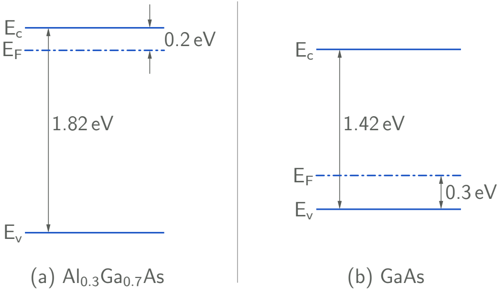
  (The bands are shown to be flat for simplicity. In practice, there will be some band bending due to the presence of an electric field.)

Example



Consider a heterojunction between n-$Al_{0.3}Ga_{0.7}As$ and p-GaAs (which are lattice matched at 300 K).

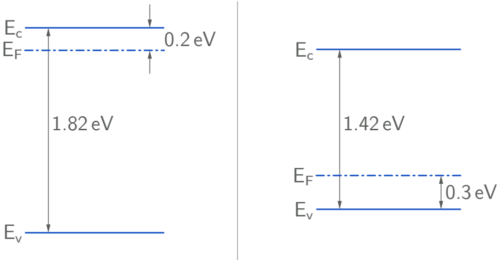
Example



Consider a heterojunction between n-$Al_{0.3}Ga_{0.7}As$ and p-GaAs (which are lattice matched at 300 K).

The doping densities are such that the Fermi levels in the isolated materials are $E_F = E_c - 0.2\,\text{eV}$ and $E_F = E_v + 0.3\,\text{eV}$ on the n and p sides, respectively.

Example



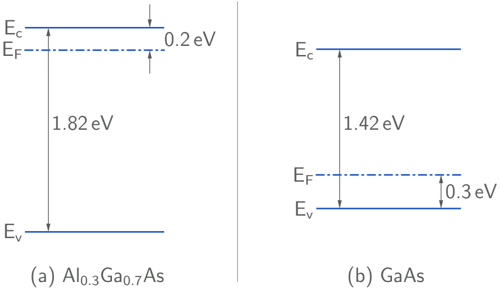
(a) $Al_{0.3}Ga_{0.7}As$ (b) GaAs

Consider a heterojunction between n-$Al_{0.3}Ga_{0.7}As$ and p-GaAs (which are lattice matched at 300 K).

The doping densities are such that the Fermi levels in the isolated materials are $E_F = E_c - 0.2\,\text{eV}$ and $E_F = E_v + 0.3\,\text{eV}$ on the n and p sides, respectively.

At the junction, E_c for $Al_{0.3}Ga_{0.7}As$ is higher than E_c for GaAs by $\Delta E_c = 0.27\,\text{eV}$, and E_v for $Al_{0.3}Ga_{0.7}As$ is lower than E_v for GaAs by $\Delta E_c = 0.13\,\text{eV}$.

Example



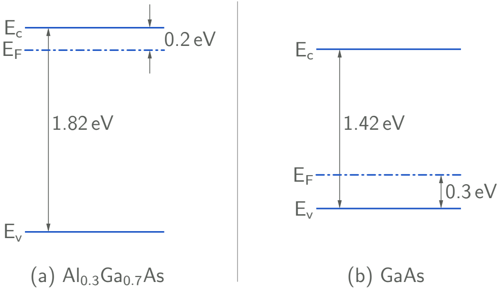
Consider a heterojunction between n-Al$_{0.3}$Ga$_{0.7}$As and p-GaAs (which are lattice matched at 300 K).

The doping densities are such that the Fermi levels in the isolated materials are $E_F = E_c - 0.2\,\text{eV}$ and $E_F = E_v + 0.3\,\text{eV}$ on the n and p sides, respectively.

At the junction, E_c for Al$_{0.3}$Ga$_{0.7}$As is higher than E_c for GaAs by $\Delta E_c = 0.27\,\text{eV}$, and E_v for Al$_{0.3}Ga_{0.7}$As is lower than E_v for GaAs by $\Delta E_c = 0.13\,\text{eV}$.

The energy gaps of Al$_{0.3}$Ga$_{0.7}$As and GaAs are 1.82 eV and 1.42 eV, respectively.

Example



Consider a heterojunction between n-$Al_{0.3}Ga_{0.7}As$ and p-GaAs (which are lattice matched at 300 K).

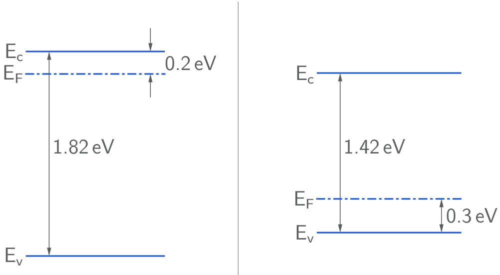
The doping densities are such that the Fermi levels in the isolated materials are $E_F = E_c - 0.2\,\text{eV}$ and $E_F = E_v + 0.3\,\text{eV}$ on the n and p sides, respectively.

At the junction, E_c for $Al_{0.3}Ga_{0.7}As$ is higher than E_c for GaAs by $\Delta E_c = 0.27\,\text{eV}$, and E_v for $Al_{0.3}Ga_{0.7}As$ is lower than E_v for GaAs by $\Delta E_c = 0.13\,\text{eV}$.

The energy gaps of $Al_{0.3}Ga_{0.7}As$ and GaAs are 1.82 eV and 1.42 eV, respectively.

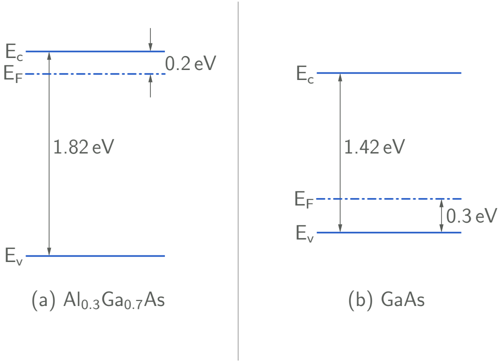
Sketch the band diagram of the pn junction in equilibrium.

Example



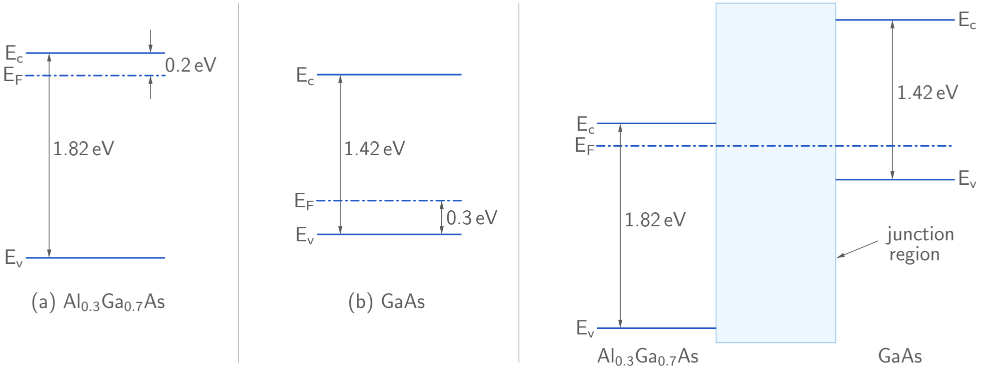
(a) $Al_{0.3}Ga_{0.7}As$ (b) GaAs

Example

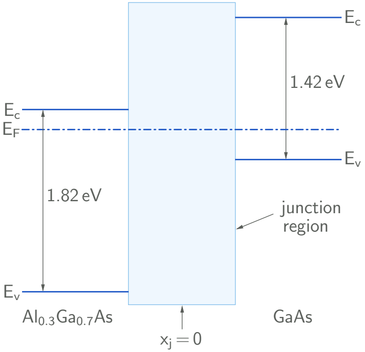


* In equilibrium, the Fermi level must be constant.

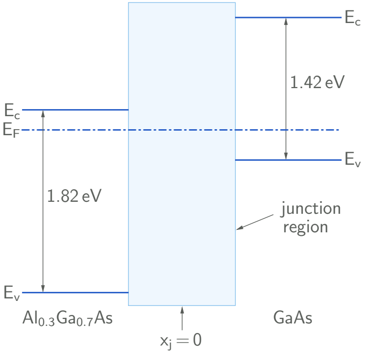
Example



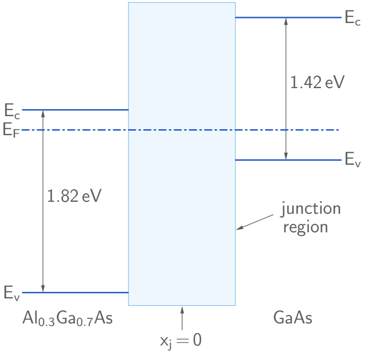
* In equilibrium, the Fermi level must be constant.



E_c
1.42 eV
E_c
E_F
E_v
1.82 eV
junction
region
E_v
$Al_{0.3}Ga_{0.7}As$
GaAs
$x_j = 0$

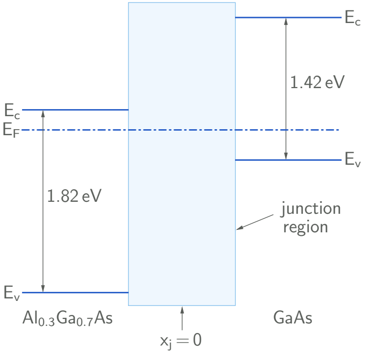


* The electric field must be continuous except at the interface (x_j) [where $\epsilon(0^-)\mathcal{E}(0^-) = \epsilon(0^+)\mathcal{E}(0^+)$]. We assume that there is no surface charge at the interface.

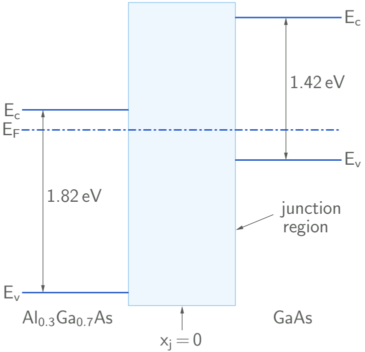


* The electric field must be continuous except at the interface (x_j) [where $\epsilon(0^-)\mathcal{E}(0^-) = \epsilon(0^+)\mathcal{E}(0^+)$]. We assume that there is no surface charge at the interface.

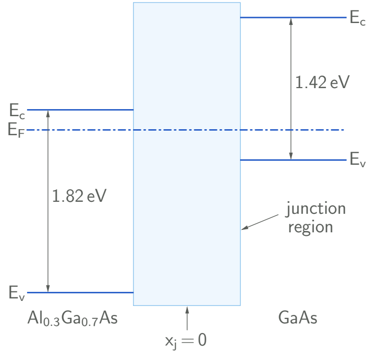
* $E_c(0^-) - E_c(0^+) = 0.27\,\text{eV}$, $E_v(0^+) - E_c(0^-) = 0.13\,\text{eV}$ (given).

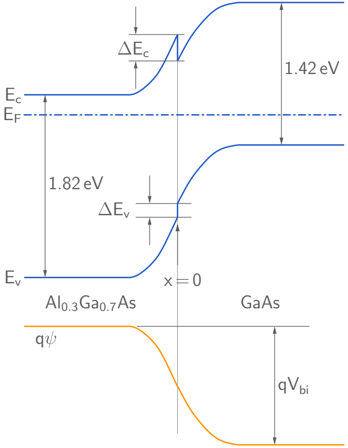


* The electric field must be continuous except at the interface (x_j) [where $\epsilon(0^-)\mathcal{E}(0^-) = \epsilon(0^+)\mathcal{E}(0^+)$]. We assume that there is no surface charge at the interface.
* $E_c(0^-) - E_c(0^+) = 0.27\,\text{eV}$, $E_v(0^+) - E_c(0^-) = 0.13\,\text{eV}$ (given).
* $E_c(x) = -q\psi + \text{constant}$, $E_v(x) = E_c(x) - E_g(x)$, where $E_g(x) = 1.82\,\text{eV}$ for $x < 0$ and $E_g(x) = 1.42\,\text{eV}$ for $x > 0$.



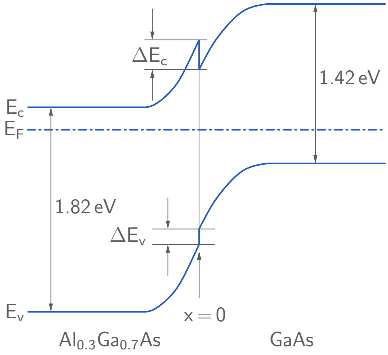
* The electric field must be continuous except at the interface (x_j) [where $\epsilon(0^-)\mathcal{E}(0^-) = \epsilon(0^+)\mathcal{E}(0^+)$]. We assume that there is no surface charge at the interface.
* $E_c(0^-) - E_c(0^+) = 0.27\,\text{eV}$, $E_v(0^+) - E_c(0^-) = 0.13\,\text{eV}$ (given).
* $E_c(x) = -q\psi + \text{constant}$, $E_v(x) = E_c(x) - E_g(x)$, where $E_g(x) = 1.82\,\text{eV}$ for $x < 0$ and $E_g(x) = 1.42\,\text{eV}$ for $x > 0$.
* The charge density variation is similar to that of a *pn* homojunction $\rightarrow$ the potential variation is also similar in the two cases.



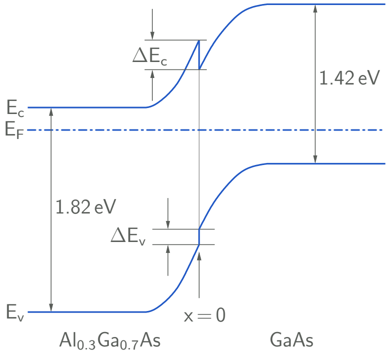


* The electric field must be continuous except at the interface (x_j) [where $\epsilon(0^-)\mathcal{E}(0^-) = \epsilon(0^+)\mathcal{E}(0^+)$]. We assume that there is no surface charge at the interface.
* $E_c(0^-) - E_c(0^+) = 0.27\,\text{eV}$, $E_v(0^+) - E_c(0^-) = 0.13\,\text{eV}$ (given).
* $E_c(x) = -q\psi + \text{constant}$, $E_v(x) = E_c(x) - E_g(x)$, where $E_g(x) = 1.82\,\text{eV}$ for $x < 0$ and $E_g(x) = 1.42\,\text{eV}$ for $x > 0$.
* The charge density variation is similar to that of a *pn* homojunction $\rightarrow$ the potential variation is also similar in the two cases.

Effect of grading

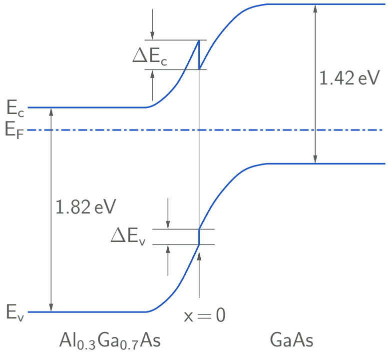


Effect of grading



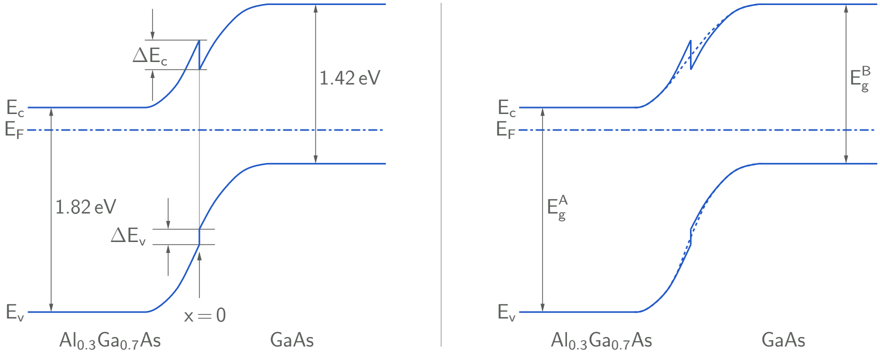
* The spike (discontinuity) ΔE_c is useful for fabricating high electron mobility transistors (HEMT).

Effect of grading



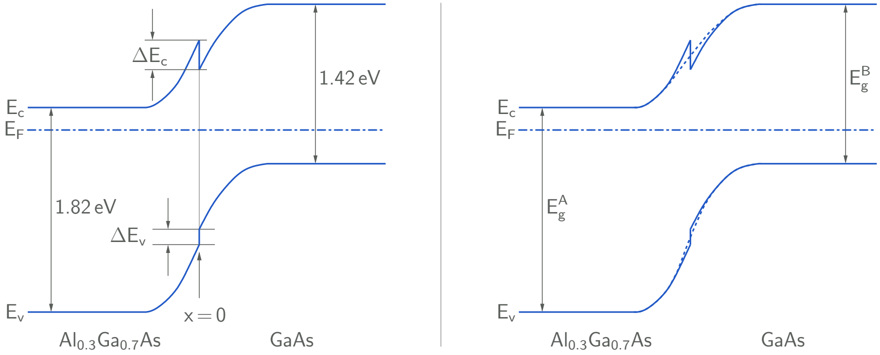
* The spike (discontinuity) ΔE_c is useful for fabricating high electron mobility transistors (HEMT).
* If the heterojunction is *graded*, with semiconductor A changing over to semiconductor B gradually (say, over a few hundred angstroms), the discontinuities in E_c and E_v can be smoothed out.

Effect of grading



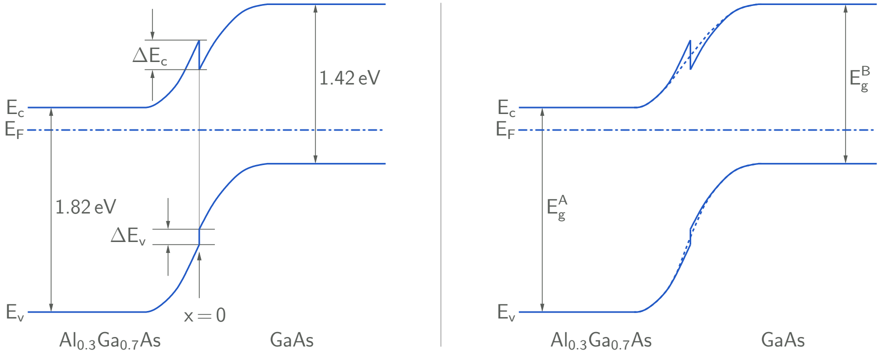
* The spike (discontinuity) ΔE_c is useful for fabricating high electron mobility transistors (HEMT).
* If the heterojunction is *graded*, with semiconductor A changing over to semiconductor B gradually (say, over a few hundred angstroms), the discontinuities in E_c and E_v can be smoothed out.

Effect of grading



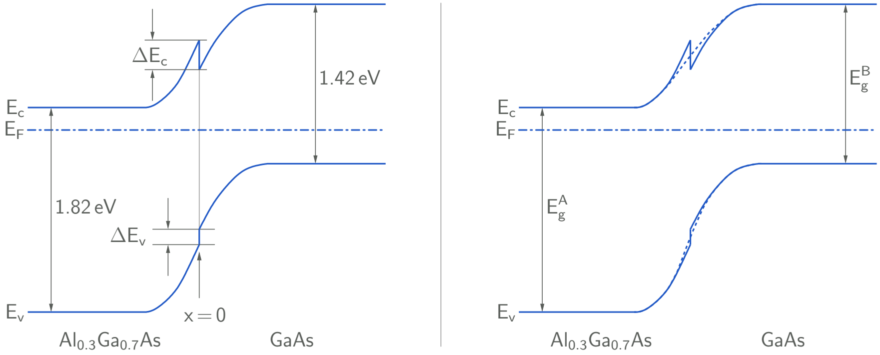
* The spike (discontinuity) ΔE_c is useful for fabricating high electron mobility transistors (HEMT).
* If the heterojunction is *graded*, with semiconductor A changing over to semiconductor B gradually (say, over a few hundred angstroms), the discontinuities in E_c and E_v can be smoothed out.
* The behaviour of a *pn* heterojunction is generally quite similar to that of a *pn* homojunction:

Effect of grading



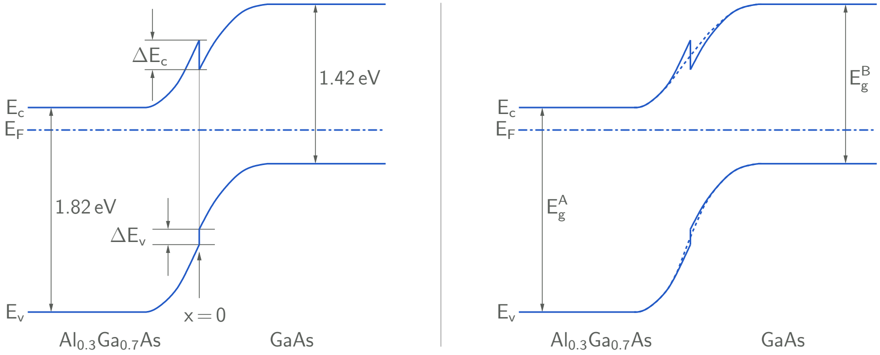
* The spike (discontinuity) ΔE_c is useful for fabricating high electron mobility transistors (HEMT).
* If the heterojunction is *graded*, with semiconductor A changing over to semiconductor B gradually (say, over a few hundred angstroms), the discontinuities in E_c and E_v can be smoothed out.
* The behaviour of a *pn* heterojunction is generally quite similar to that of a *pn* homojunction:
  With forward bias, the depletion region shrinks. The current increases exponentially with forward bias.

Effect of grading



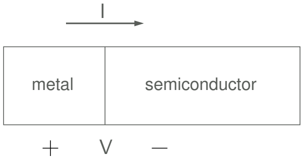
* The spike (discontinuity) ΔE_c is useful for fabricating high electron mobility transistors (HEMT).
* If the heterojunction is *graded*, with semiconductor A changing over to semiconductor B gradually (say, over a few hundred angstroms), the discontinuities in E_c and E_v can be smoothed out.
* The behaviour of a *pn* heterojunction is generally quite similar to that of a *pn* homojunction:
  With forward bias, the depletion region shrinks. The current increases exponentially with forward bias.
  With reverse bias, the depletion region expands. The current is negligibly small.

Effect of grading



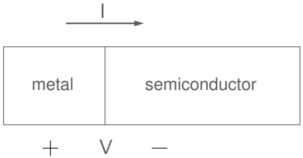
* Graded heterojunctions are used to fabricate heterojunction bipolar transistors (HBT) in which a high current gain and a small device resistance are simultaneously made possible because of different semiconductors used for the emitter and base regoins of the device.

Metal-semiconductor junctions



Metal-semiconductor (M-S) junctions serve two important purposes in semiconductor devices:

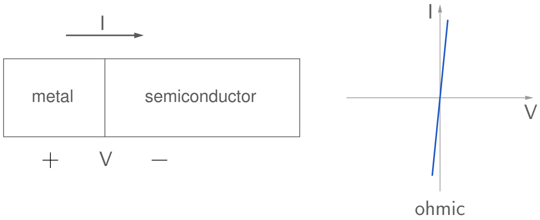
Metal-semiconductor junctions



Metal-semiconductor (M-S) junctions serve two important purposes in semiconductor devices:

* A metallic contact to a semiconductor device serves as the interface of the device with the external circuit. In this case, the M-S junction must be *ohmic*, i.e., it should be able to conduct a reasonably large current *in either direction* with a very small voltage drop across the junction.

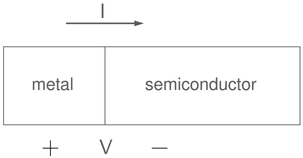
Metal-semiconductor junctions

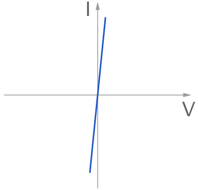


Metal-semiconductor (M-S) junctions serve two important purposes in semiconductor devices:

* A metallic contact to a semiconductor device serves as the interface of the device with the external circuit. In this case, the M-S junction must be *ohmic*, i.e., it should be able to conduct a reasonably large current *in either direction* with a very small voltage drop across the junction.

Metal-semiconductor junctions

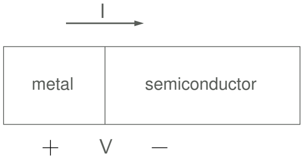


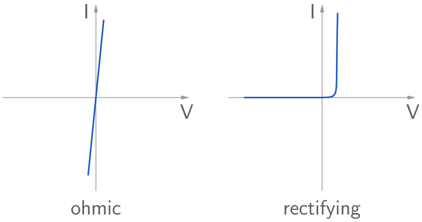


Metal-semiconductor (M-S) junctions serve two important purposes in semiconductor devices:

* A metallic contact to a semiconductor device serves as the interface of the device with the external circuit. In this case, the M-S junction must be *ohmic*, i.e., it should be able to conduct a reasonably large current *in either direction* with a very small voltage drop across the junction.
* A *rectifying* M-S junction plays a key role in the operation of some semiconductor devices like the MESFET (Metal-Semiconductor Field-Effect Transistor). As the name implies, a rectifying M-S junction (also called a "Schottky contact") conducts well in one direction but blocks current in the other direction.

Metal-semiconductor junctions



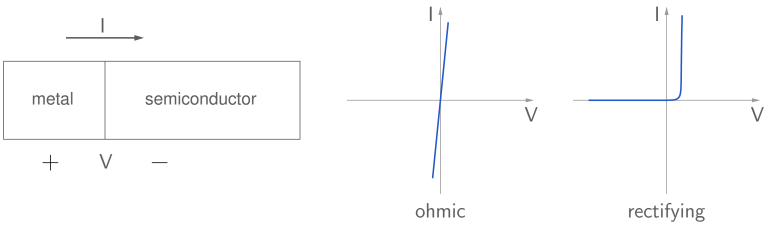


Metal-semiconductor (M-S) junctions serve two important purposes in semiconductor devices:

* A metallic contact to a semiconductor device serves as the interface of the device with the external circuit. In this case, the M-S junction must be *ohmic*, i.e., it should be able to conduct a reasonably large current *in either direction* with a very small voltage drop across the junction.

* A *rectifying* M-S junction plays a key role in the operation of some semiconductor devices like the MESFET (Metal-Semiconductor Field-Effect Transistor). As the name implies, a rectifying M-S junction (also called a "Schottky contact") conducts well in one direction but blocks current in the other direction.

Metal-semiconductor junctions

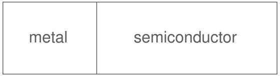


Metal-semiconductor (M-S) junctions serve two important purposes in semiconductor devices:

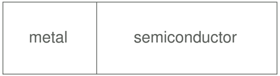
* A metallic contact to a semiconductor device serves as the interface of the device with the external circuit. In this case, the M-S junction must be *ohmic*, i.e., it should be able to conduct a reasonably large current *in either direction* with a very small voltage drop across the junction.

* A *rectifying* M-S junction plays a key role in the operation of some semiconductor devices like the MESFET (Metal-Semiconductor Field-Effect Transistor). As the name implies, a rectifying M-S junction (also called a "Schottky contact") conducts well in one direction but blocks current in the other direction.

What decides whether a given M-S junction is ohmic or rectifying?

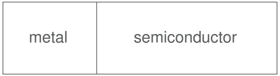


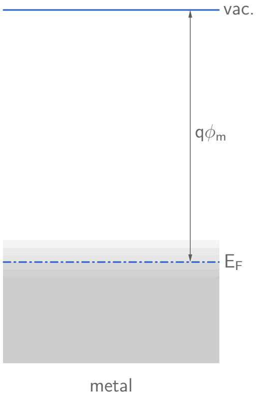
The band diagram of a metal-semiconductor junction is determined by



The band diagram of a metal-semiconductor junction is determined by

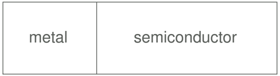
- metal work function ϕ_m (difference between the "vacuum level" and the Fermi level)

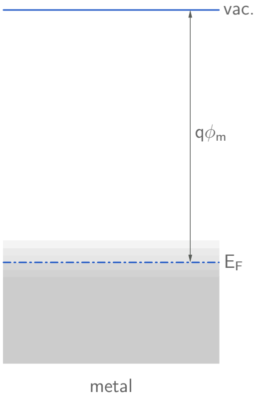




The band diagram of a metal-semiconductor junction is determined by

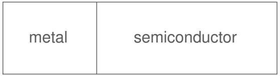
- metal work function ϕ_m (difference between the "vacuum level" and the Fermi level)

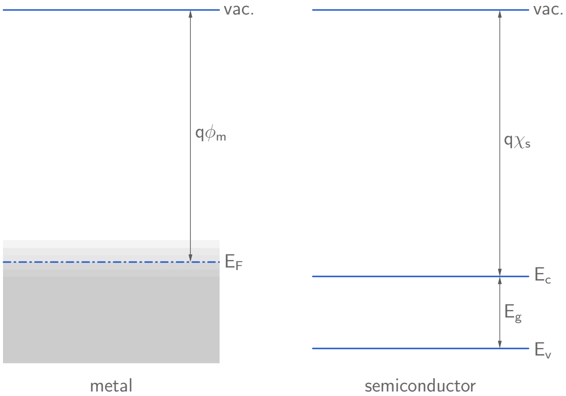




The band diagram of a metal-semiconductor junction is determined by

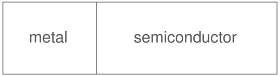
- metal work function ϕ_m (difference between the "vacuum level" and the Fermi level)
- electron affinity χ_s of the semiconductor (difference between the conduction band edge E_c and the vacuum level)

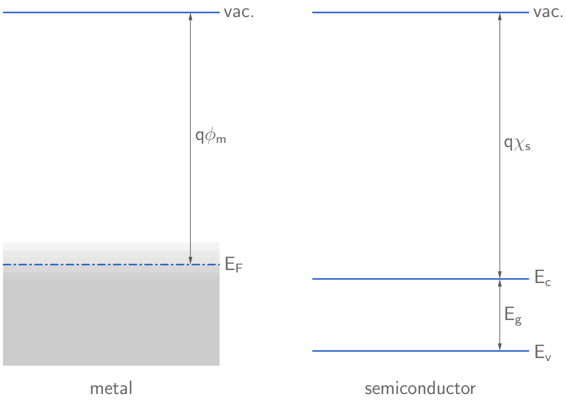




The band diagram of a metal-semiconductor junction is determined by

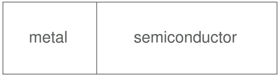
- metal work function ϕ_m (difference between the "vacuum level" and the Fermi level)
- electron affinity χ_s of the semiconductor (difference between the conduction band edge E_c and the vacuum level)

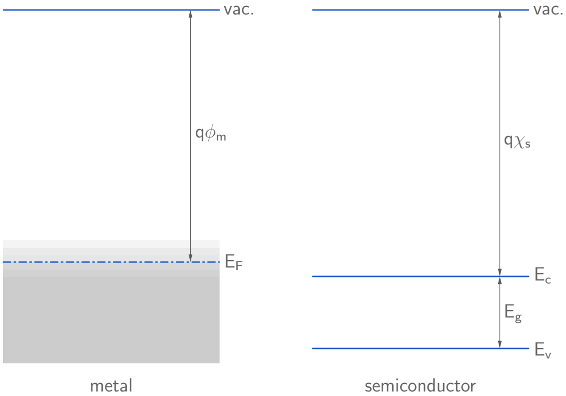




The band diagram of a metal-semiconductor junction is determined by

- metal work function ϕ_m (difference between the "vacuum level" and the Fermi level)
- electron affinity χ_s of the semiconductor (difference between the conduction band edge E_c and the vacuum level)
- doping density in the semiconductor (which sets the equilibrium Fermi level in the semiconductor)



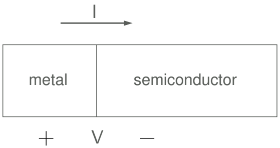


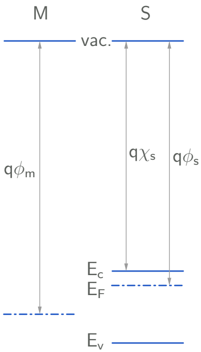
The band diagram of a metal-semiconductor junction is determined by

- metal work function ϕ_m (difference between the "vacuum level" and the Fermi level)
- electron affinity χ_s of the semiconductor (difference between the conduction band edge E_c and the vacuum level)
- doping density in the semiconductor (which sets the equilibrium Fermi level in the semiconductor)
- additional electron states within the band gap at the interface

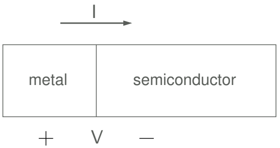
  (We will ignore this effect, i.e., we will assume the M-S interface to be perfect.)

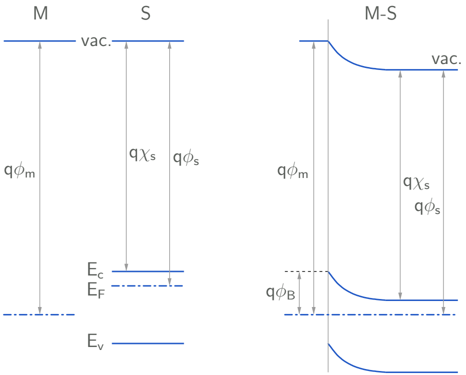
n-type semiconductor, $\phi_m > \phi_s$



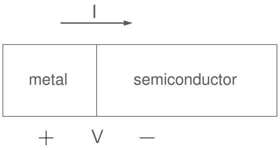


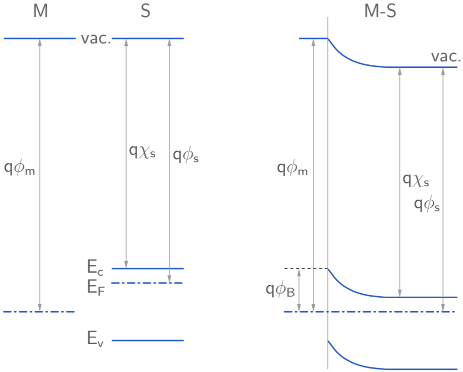
n-type semiconductor, $\phi_m > \phi_s$





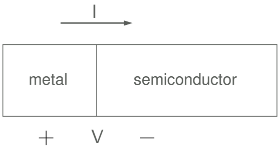
n-type semiconductor, $\phi_m > \phi_s$

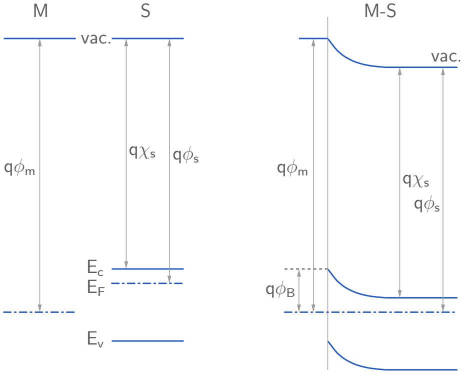




* There is a substantial barrier $qV_{\text{bi}} = q\phi_m - q\phi_s$ to electron flow from S to M.

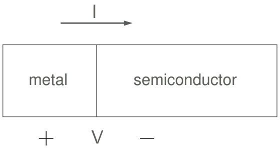
n-type semiconductor, $\phi_m > \phi_s$

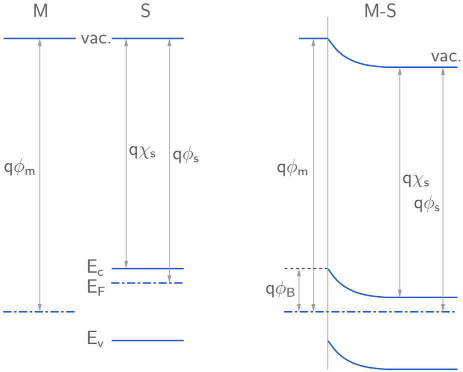




* There is a substantial barrier $qV_{\text{bi}} = q\phi_m - q\phi_s$ to electron flow from S to M.
* In the opposite direction (from M to S), there is also a substantial barrier $q\phi_B = q\phi_m - q\chi_s$.

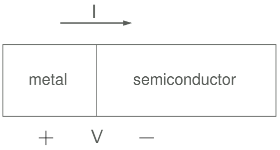
n-type semiconductor, $\phi_m > \phi_s$

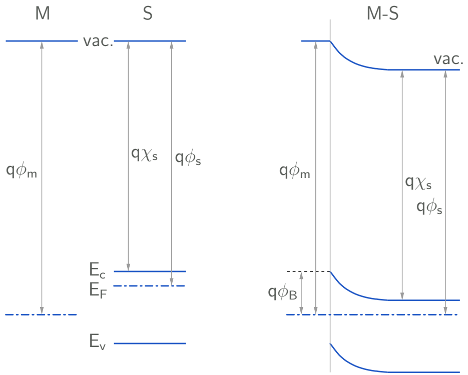




* There is a substantial barrier $qV_{\text{bi}} = q\phi_m - q\phi_s$ to electron flow from S to M.
* In the opposite direction (from M to S), there is also a substantial barrier $q\phi_B = q\phi_m - q\chi_s$.
* If a positive voltage is applied to the metal (w.r.t. the semiconductor), the barrier to electron flow from S to M reduces → current can flow.

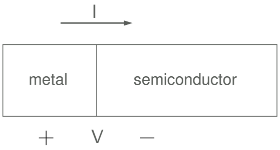
n-type semiconductor, $\phi_m > \phi_s$

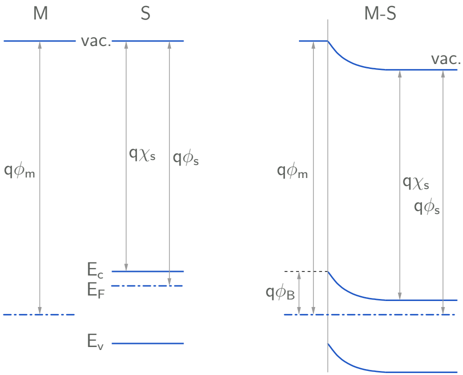




* There is a substantial barrier $qV_{\text{bi}} = q\phi_m - q\phi_s$ to electron flow from S to M.
* In the opposite direction (from M to S), there is also a substantial barrier $q\phi_B = q\phi_m - q\chi_s$.
* If a positive voltage is applied to the metal (w.r.t. the semiconductor), the barrier to electron flow from S to M reduces $\rightarrow$ current can flow.
* If a negative voltage is applied to the metal, the barrier increases $\rightarrow$ negligible current flow.

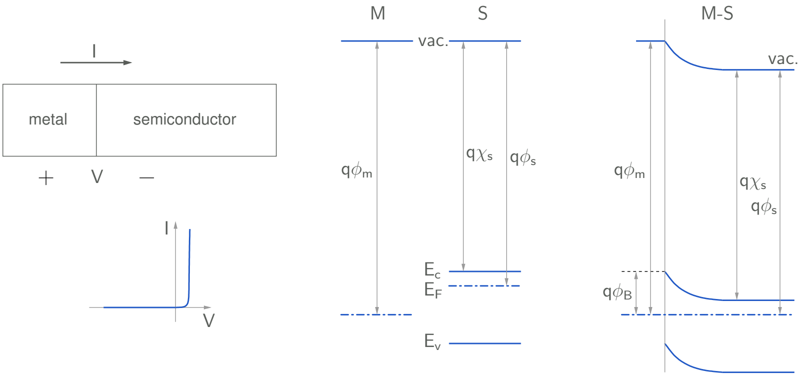
n-type semiconductor, $\phi_m > \phi_s$





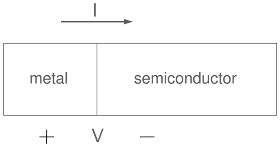
* There is a substantial barrier $qV_{\text{bi}} = q\phi_m - q\phi_s$ to electron flow from S to M.
* In the opposite direction (from M to S), there is also a substantial barrier $q\phi_B = q\phi_m - q\chi_s$.
* If a positive voltage is applied to the metal (w.r.t. the semiconductor), the barrier to electron flow from S to M reduces $\rightarrow$ current can flow.
* If a negative voltage is applied to the metal, the barrier increases $\rightarrow$ negligible current flow.
* This condition leads to a rectifying contact.

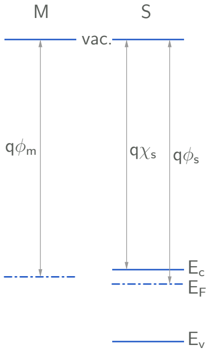
n-type semiconductor, $\phi_m > \phi_s$



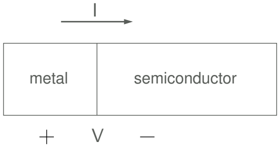
* There is a substantial barrier $qV_{\text{bi}} = q\phi_m - q\phi_s$ to electron flow from S to M.
* In the opposite direction (from M to S), there is also a substantial barrier $q\phi_B = q\phi_m - q\chi_s$.
* If a positive voltage is applied to the metal (w.r.t. the semiconductor), the barrier to electron flow from S to M reduces $\rightarrow$ current can flow.
* If a negative voltage is applied to the metal, the barrier increases $\rightarrow$ negligible current flow.
* This condition leads to a rectifying contact.

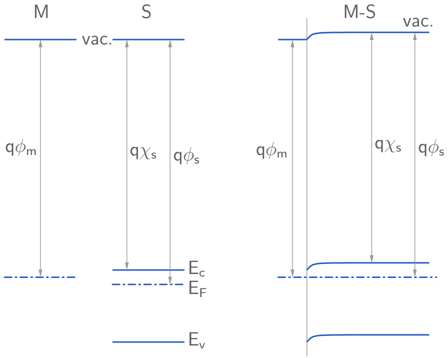
n-type semiconductor, $\phi_m < \phi_s$



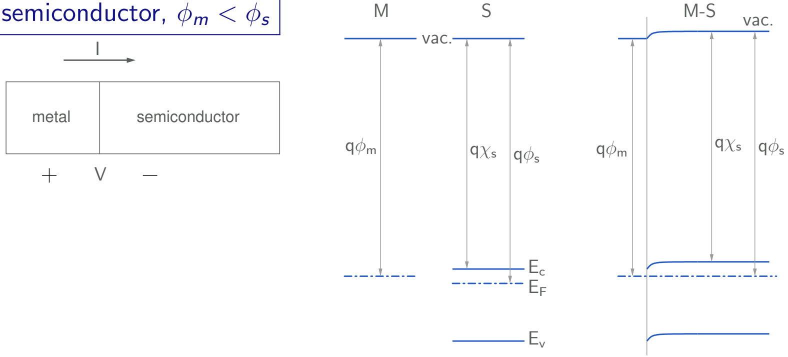


n-type semiconductor, $\phi_m < \phi_s$



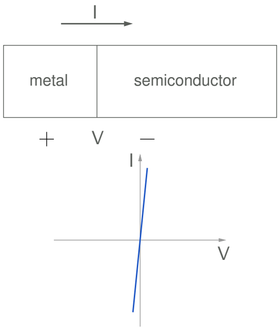


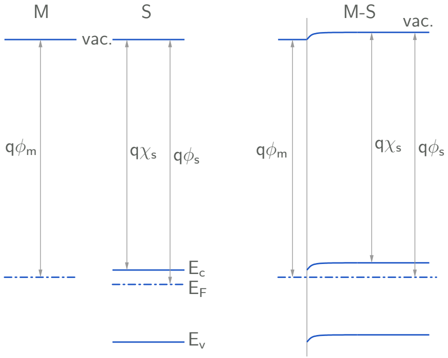
n-type semiconductor, $\phi_m < \phi_s$



* There is a no barrier to electron flow from S to M $\rightarrow$ ohmic contact.

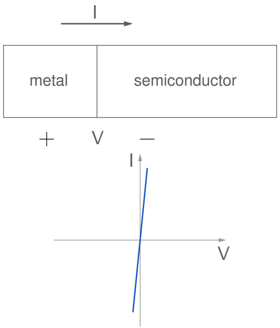
n-type semiconductor, $\phi_m < \phi_s$

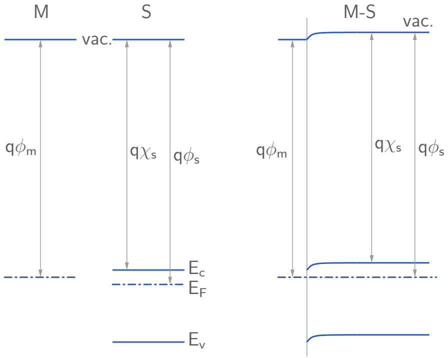




* There is a no barrier to electron flow from S to M $\rightarrow$ ohmic contact.

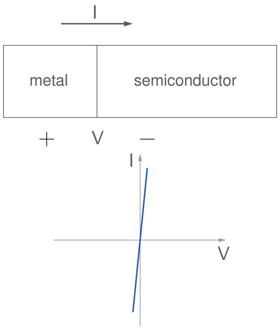
n-type semiconductor, $\phi_m < \phi_s$

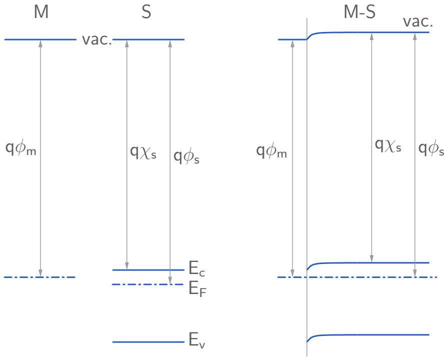




* There is a no barrier to electron flow from S to M $\rightarrow$ ohmic contact.
* In real semiconductors, there are significant departures from the ideal situation we have described.

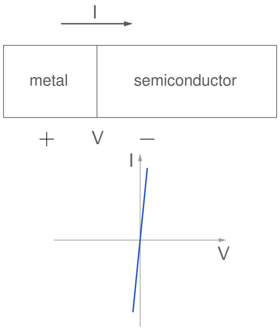
n-type semiconductor, $\phi_m < \phi_s$

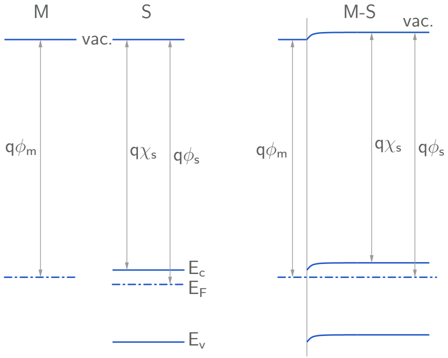




* There is a no barrier to electron flow from S to M $\rightarrow$ ohmic contact.
* In real semiconductors, there are significant departures from the ideal situation we have described.
  - At the M-S interface, the semiconductor crystal structure gets abruptly terminated, which results in incomplete ("dangling") covalent bonds or, in terms of the band picture, electron states within the forbidden gap.

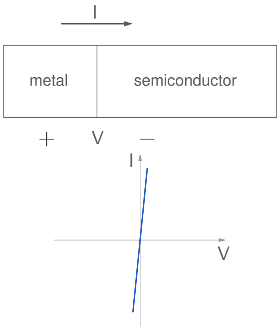
n-type semiconductor, $\phi_m < \phi_s$

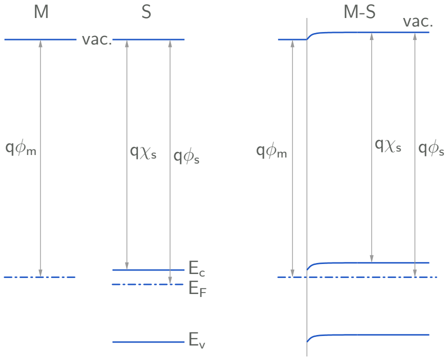




* There is a no barrier to electron flow from S to M → ohmic contact.
* In real semiconductors, there are significant departures from the ideal situation we have described.
  - At the M-S interface, the semiconductor crystal structure gets abruptly terminated, which results in incomplete ("dangling") covalent bonds or, in terms of the band picture, electron states within the forbidden gap.
  - There may be a thin ($\sim 10\,\text{Å}$) oxide layer between the metal and the semiconductor.

n-type semiconductor, $\phi_m < \phi_s$

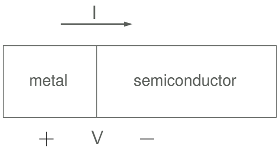


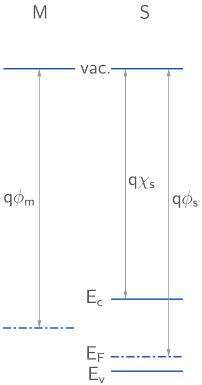


* There is a no barrier to electron flow from S to M $\rightarrow$ ohmic contact.
* In real semiconductors, there are significant departures from the ideal situation we have described.
  - At the M-S interface, the semiconductor crystal structure gets abruptly terminated, which results in incomplete ("dangling") covalent bonds or, in terms of the band picture, electron states within the forbidden gap.
  - There may be a thin ($\sim 10\,\text{Å}$) oxide layer between the metal and the semiconductor.

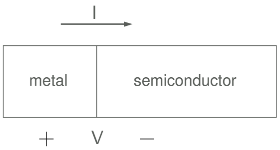
Because of these complications, the barrier heights get modified. However, the qualitative picture remains valid as long as the actual experimentally measured barrier heights are used.

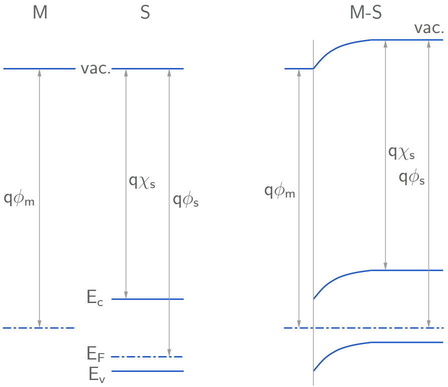
p-type semiconductor, $\phi_m < \phi_s$



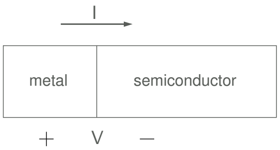


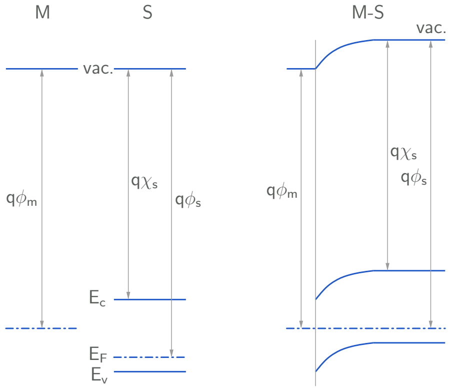
p-type semiconductor, $\phi_m < \phi_s$





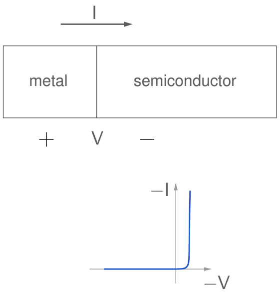
p-type semiconductor, $\phi_m < \phi_s$

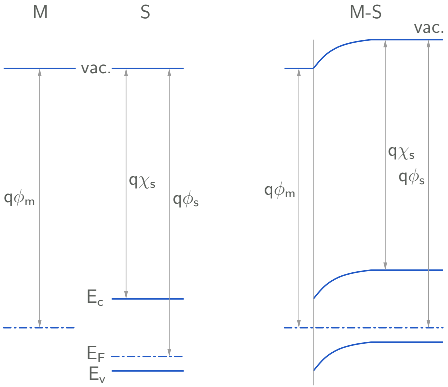




* There is a substantial barrier $qV_{\text{bi}} = q\phi_s - q\phi_m$ to hole flow from S to M.
  In the opposite direction (from M to S), there is also a substantial barrier $q\phi_B = q\chi_s + E_g - q\phi_m$, and the contact is therefore rectifying.

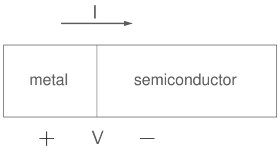
p-type semiconductor, $\phi_m < \phi_s$

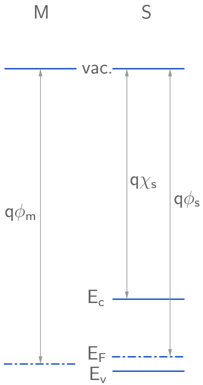




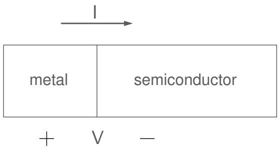
* There is a substantial barrier $qV_{\rm bi} = q\phi_s - q\phi_m$ to hole flow from S to M.
  In the opposite direction (from M to S), there is also a substantial barrier $q\phi_B = q\chi_s + E_g - q\phi_m$, and the contact is therefore rectifying.

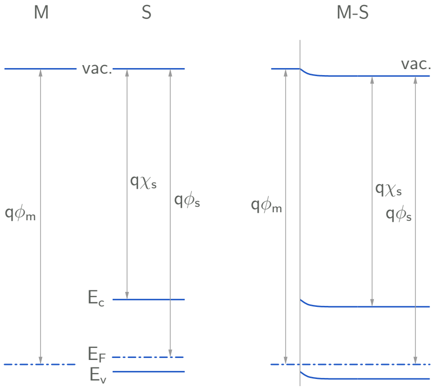
p-type semiconductor, $\phi_m > \phi_s$



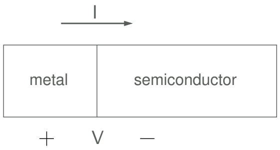


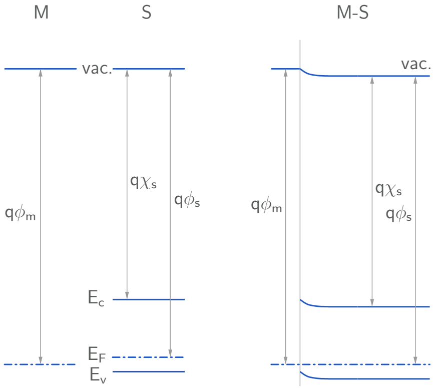
p-type semiconductor, $\phi_m > \phi_s$





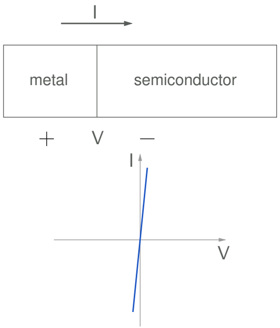
p-type semiconductor, $\phi_m > \phi_s$

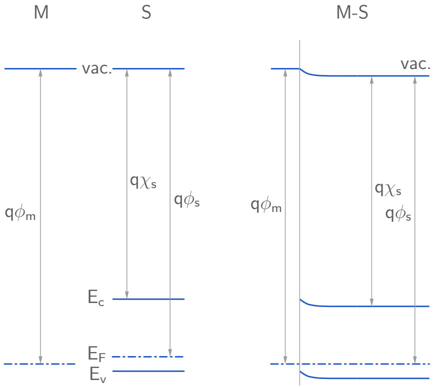




* There is a no barrier to hole flow from S to M → ohmic contact.

p-type semiconductor, $\phi_m > \phi_s$





* There is a no barrier to hole flow from S to M → ohmic contact.

Current-voltage relationship for a rectifying M-S junction

* The *I*-*V* behaviour of a rectifying M-S junction is similar to that of a *pn* junction:
  Under forward bias, a large current flows with a small applied voltage.
  Under reverse bias, there is virtually no current flow.

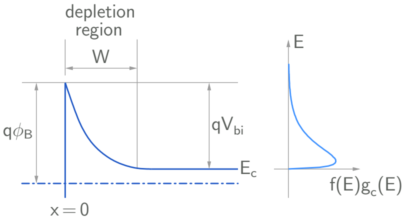
Current-voltage relationship for a rectifying M-S junction

* The *I*-*V* behaviour of a rectifying M-S junction is similar to that of a *pn* junction:
  Under forward bias, a large current flows with a small applied voltage.
  Under reverse bias, there is virtually no current flow.
* The mechanisms of current conduction in a rectifying M-S junction is very different.

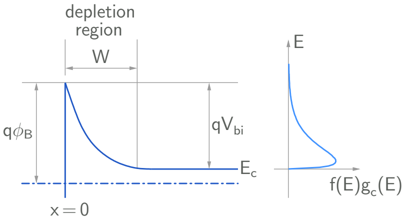
Current-voltage relationship for a rectifying M-S junction

* The I-V behaviour of a rectifying M-S junction is similar to that of a *pn* junction:
  Under forward bias, a large current flows with a small applied voltage.
  Under reverse bias, there is virtually no current flow.
* The mechanisms of current conduction in a rectifying M-S junction is very different.
* In a *pn* junction, the current is due to injection of minority carriers (e.g., injection of electrons from the n side to the p side).

* The *I*-*V* behaviour of a rectifying M-S junction is similar to that of a *pn* junction:
  Under forward bias, a large current flows with a small applied voltage.
  Under reverse bias, there is virtually no current flow.
* The mechanisms of current conduction in a rectifying M-S junction is very different.
* In a *pn* junction, the current is due to injection of minority carriers (e.g., injection of electrons from the *n* side to the *p* side).
* In an M-S junction, minority carriers play no role in current conduction, and it is the injection of the majority carriers from semiconductor to metal which determines the current.

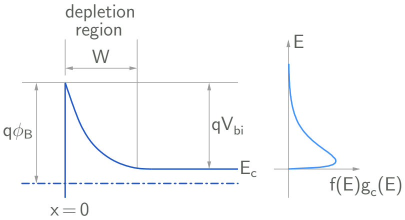


Consider a M-S junction in equilibrium.



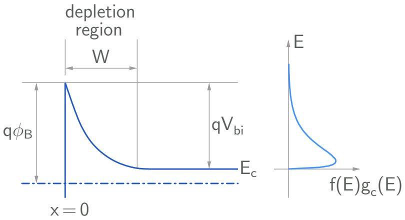
Consider a M-S junction in equilibrium.

* Close to the M-S interface, there is a depletion region, and for $x > W$, there is a charge-neutral region with a large number of electrons (majority carriers).



Consider a M-S junction in equilibrium.

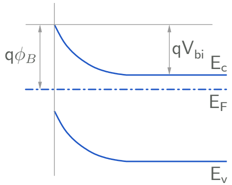
* Close to the M-S interface, there is a depletion region, and for $x > W$, there is a charge-neutral region with a large number of electrons (majority carriers).
* The electron flow from S to M is determined by how many electrons can surmount the barrier qV_{bi}. This current is called the "thermionic emission" current and is given by $J_{S \rightarrow M} = A^* T^2 e^{-\phi_B/V_T}$,
  where A^* is the Richardson's constant (with units of $\mathrm{A/cm^2K^2}$).



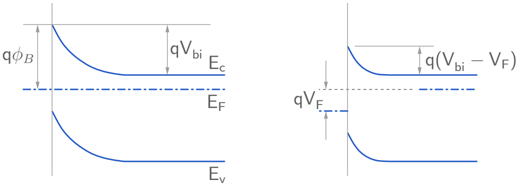
Consider a M-S junction in equilibrium.

* Close to the M-S interface, there is a depletion region, and for $x > W$, there is a charge-neutral region with a large number of electrons (majority carriers).
* The electron flow from S to M is determined by how many electrons can surmount the barrier $qV_{\rm bi}$. This current is called the "thermionic emission" current and is given by $J_{S \rightarrow M} = A^* T^2 \, e^{-\phi_B/V_T}$, where A^* is the Richardson's constant (with units of $\mathrm{A/cm^2K^2}$).
* In equilibrium, there is an equal and opposite current density, $J_{M \rightarrow S} = -J_{S \rightarrow M} = -A^* T^2 \, e^{-\phi_B/V_T}$, resulting in a net current density $J=0$.

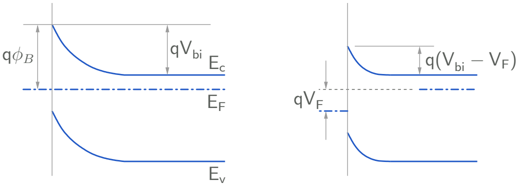
Current-voltage relationship for a rectifying M-S junction



Current-voltage relationship for a rectifying M-S junction



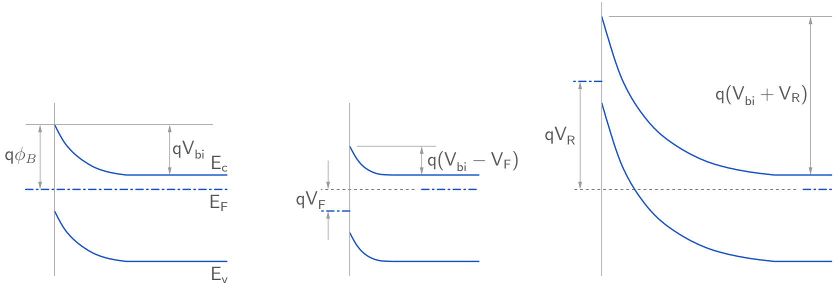
Current-voltage relationship for a rectifying M-S junction



With a forward voltage, the barrier to electron flow from S to M decreases by V_F while that for M to S remains the same, and the net current density is

$$J = J_{S\to M} - J_{M\to S} = A^* T^2 \left[e^{-(\phi_B - V_F)/V_T} - e^{-\phi_B/V_T}\right]$$

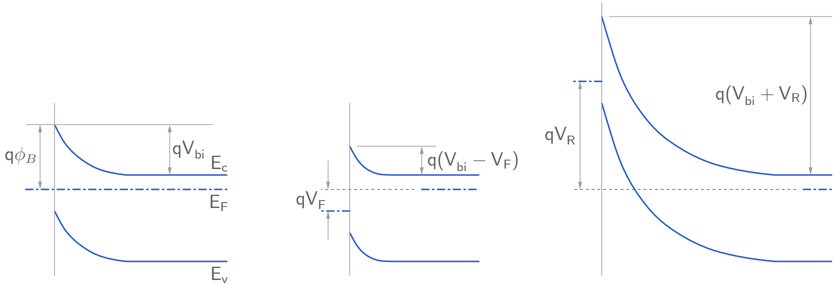
$$= A^* T^2 e^{-\phi_B/V_T}\left[e^{V_F/V_T} - 1\right].$$



With a forward voltage, the barrier to electron flow from S to M decreases by V_F while that for M to S remains the same, and the net current density is

$$J = J_{S\rightarrow M} - J_{M\rightarrow S} = A^* T^2 \left[e^{-(\phi_B - V_F)/V_T} - e^{-\phi_B/V_T} \right]$$
$$= A^* T^2 e^{-\phi_B/V_T} \left[e^{V_F/V_T} - 1 \right].$$

Current-voltage relationship for a rectifying M-S junction

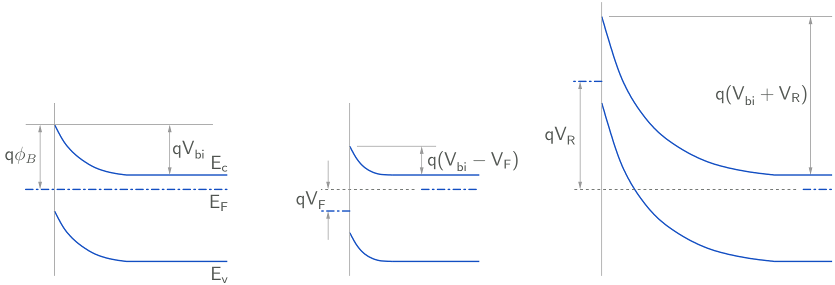


With a forward voltage, the barrier to electron flow from S to M decreases by V_F while that for M to S remains the same, and the net current density is

$$J = J_{S\to M} - J_{M\to S} = A^* T^2 \left[e^{-(\phi_B - V_F)/V_T} - e^{-\phi_B/V_T} \right]$$
$$= A^* T^2 e^{-\phi_B/V_T} \left[e^{V_F/V_T} - 1 \right].$$

With a reverse bias, the S→M barrier increases by V_R, and the above equation holds with, $V_F \to -V_R$.

Current-voltage relationship for a rectifying M-S junction



With a forward voltage, the barrier to electron flow from S to M decreases by V_F while that for M to S remains the same, and the net current density is

$$J = J_{S\to M} - J_{M\to S} = A^* T^2 \left[e^{-(\phi_B - V_F)/V_T} - e^{-\phi_B/V_T} \right]$$

$$= A^* T^2 e^{-\phi_B/V_T} \left[e^{V_F/V_T} - 1 \right].$$

With a reverse bias, the S→M barrier increases by V_R, and the above equation holds with, $V_F \to -V_R$.

In summary, $J = J_s \left[e^{V/V_T} - 1 \right]$, where $J_s = A^* T^2 \, e^{-\phi_B/V_T}$.

* For both *pn* and rectifying M-S junctions, we have $J = J_s\left(e^{V_a/V_T} - 1\right)$.

* For both *pn* and rectifying M-S junctions, we have $J = J_s \left(e^{V_a/V_T} - 1 \right)$.
* For a *pn* junction, $J_s = q \left(\dfrac{D_n n_{p0}}{L_n} + \dfrac{D_p p_{n0}}{L_p} \right)$ involves minority carrier densities and lifetimes.

* For both *pn* and rectifying M-S junctions, we have $J = J_s \left(e^{V_a/V_T} - 1\right)$.
* For a *pn* junction, $J_s = q \left(\dfrac{D_n n_{p0}}{L_n} + \dfrac{D_p p_{n0}}{L_p}\right)$ involves minority carrier densities and lifetimes.
* For a M-S junction, current conduction is dominated by thermionic emission of majority carriers. (In addition, there can be a tunnelling component of the current called "thermionic field emission."[1])

[1] M. Shur, *Physics of Semiconductor Devices*. New Delhi: Prentice-Hall India, 1990.

- For both *pn* and rectifying M-S junctions, we have $J = J_s\left(e^{V_a/V_T} - 1\right)$.
- For a *pn* junction, $J_s = q\left(\dfrac{D_n n_{p0}}{L_n} + \dfrac{D_p p_{n0}}{L_p}\right)$ involves minority carrier densities and lifetimes.
- For a M-S junction, current conduction is dominated by thermionic emission of majority carriers. (In addition, there can be a tunnelling component of the current called "thermionic field emission."[1])
- As we will see, turn-off of a *pn* diode involves removal of the excess minority carrier charge, a slow process governed by the lifetime of the minority carriers. In Schottky diodes, there is no such requirement, and therefore they can be turned off much faster.

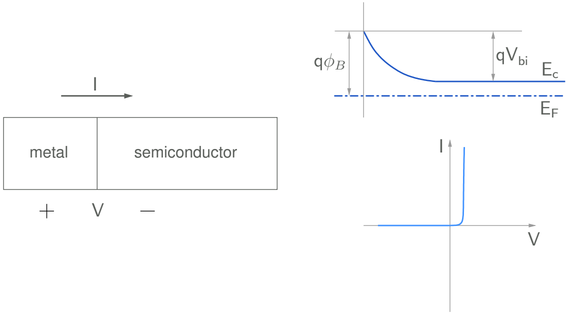
[1] M. Shur, *Physics of Semiconductor Devices*. New Delhi: Prentice-Hall India, 1990.

Rectifying M-S junction

* For both *pn* and rectifying M-S junctions, we have $J = J_s\left(e^{V_a/V_T} - 1\right)$.
* For a *pn* junction, $J_s = q\left(\dfrac{D_n n_{p0}}{L_n} + \dfrac{D_p p_{n0}}{L_p}\right)$ involves minority carrier densities and lifetimes.
* For a M-S junction, current conduction is dominated by thermionic emission of majority carriers. (In addition, there can be a tunnelling component of the current called "thermionic field emission."[1])
* As we will see, turn-off of a *pn* diode involves removal of the excess minority carrier charge, a slow process governed by the lifetime of the minority carriers. In Schottky diodes, there is no such requirement, and therefore they can be turned off much faster.
* Remark: The process of thermionic emission also takes place in a *p-n* junction, but it can be ignored.

[1] M. Shur, *Physics of Semiconductor Devices*. New Delhi: Prentice-Hall India, 1990.

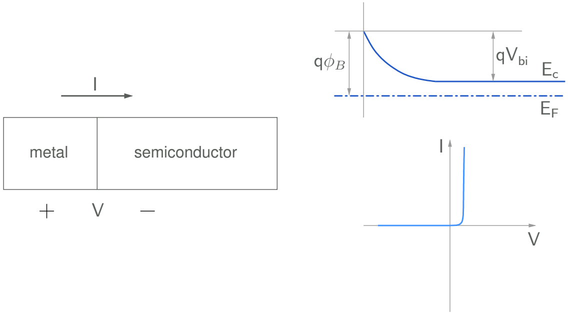
Effect of high doping density



(a)

* The contact in (a) is rectifying because of the potential barrier.

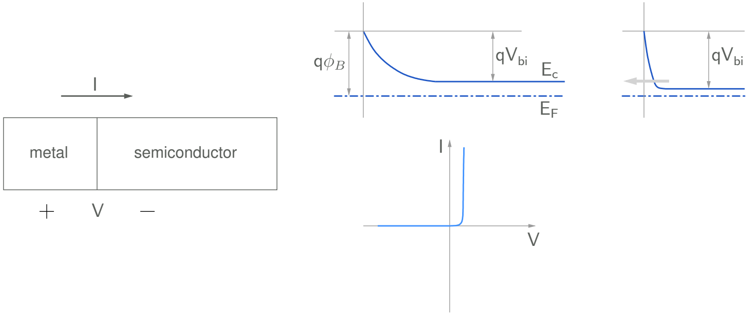
Effect of high doping density



(a)

* The contact in (a) is rectifying because of the potential barrier.
* If the doping density is increased, the barrier width (depletion region width) decreases, and tunneling of electrons becomes possible even with a small applied voltage (of either polarity)

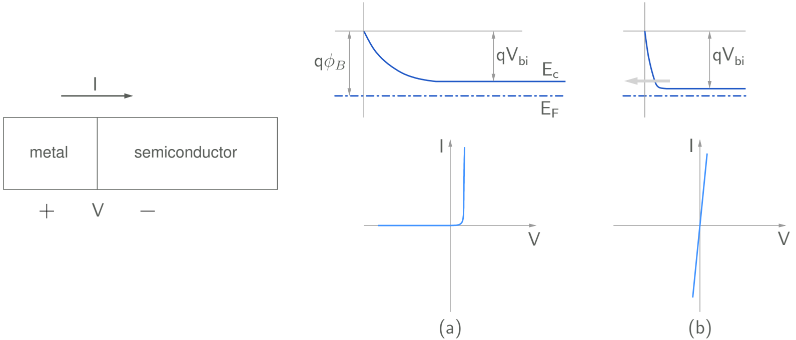
Effect of high doping density



(a)

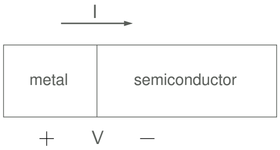
* The contact in (a) is rectifying because of the potential barrier.
* If the doping density is increased, the barrier width (depletion region width) decreases, and tunneling of electrons becomes possible even with a small applied voltage (of either polarity)

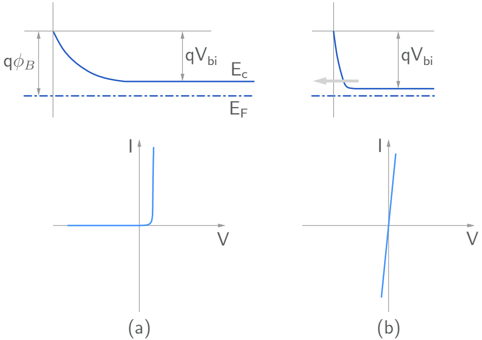
Effect of high doping density



* The contact in (a) is rectifying because of the potential barrier.
* If the doping density is increased, the barrier width (depletion region width) decreases, and tunneling of electrons becomes possible even with a small applied voltage (of either polarity)

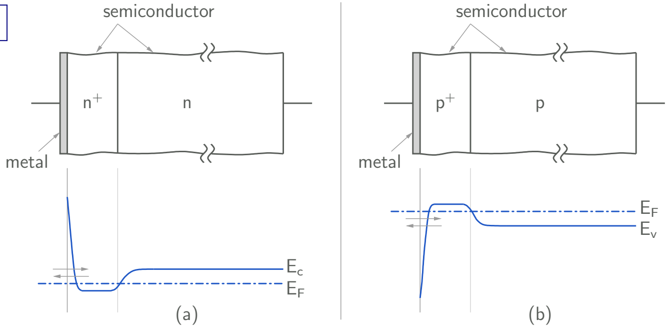
Effect of high doping density



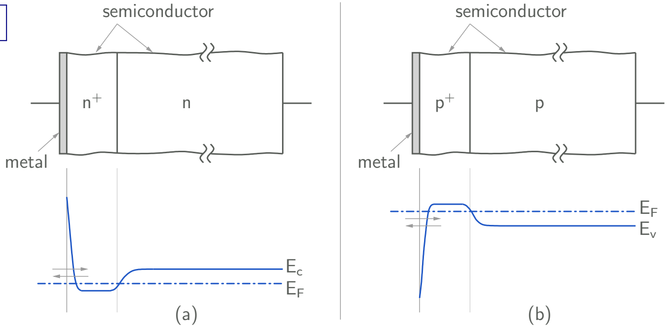


* The contact in (a) is rectifying because of the potential barrier.
* If the doping density is increased, the barrier width (depletion region width) decreases, and tunneling of electrons becomes possible even with a small applied voltage (of either polarity) → an ohmic contact.

Practical ohmic contacts

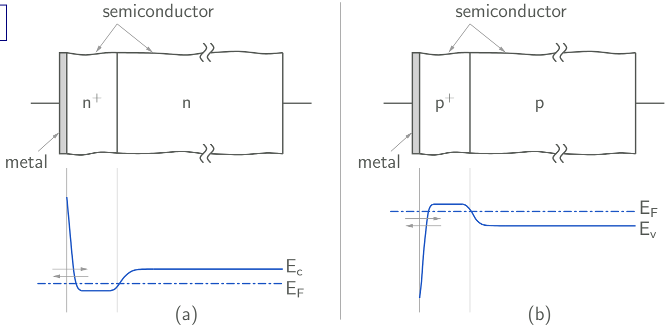


Practical ohmic contacts



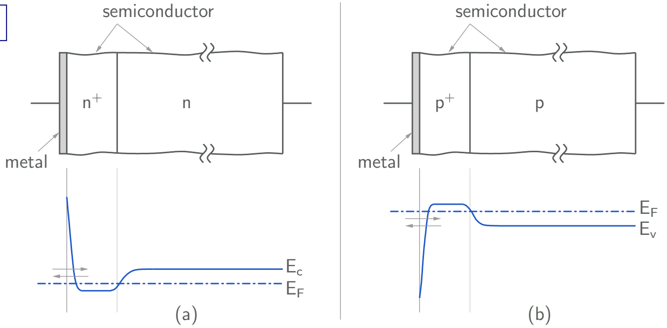
* In many practical situations, ohmic contacts are required to be made to an *n*-type or *p*-type semiconductor region with a moderate doping density, and a metal which will form an ohmic contact is either not available or is not technologically convenient.

Practical ohmic contacts



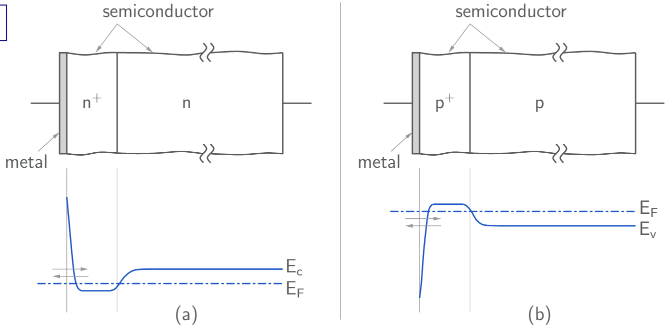
* In many practical situations, ohmic contacts are required to be made to an *n*-type or *p*-type semiconductor region with a moderate doping density, and a metal which will form an ohmic contact is either not available or is not technologically convenient.
* In such cases, a heavily doped region (of the same type) is first created, forming an n^+n or p^+p junction.

Practical ohmic contacts



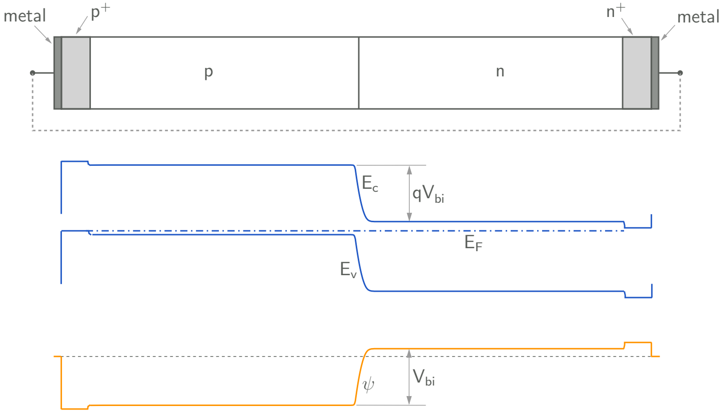
* In many practical situations, ohmic contacts are required to be made to an *n*-type or *p*-type semiconductor region with a moderate doping density, and a metal which will form an ohmic contact is either not available or is not technologically convenient.
* In such cases, a heavily doped region (of the same type) is first created, forming an n^+n or p^+p junction.
* These n^+n or p^+p junctions are essentially ohmic since a large number of majority carriers (of the same type) are available for conduction on both sides of the junction.

Practical ohmic contacts

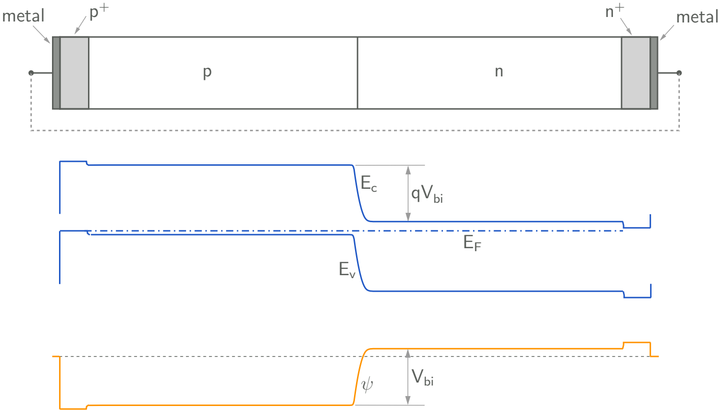


* In many practical situations, ohmic contacts are required to be made to an *n*-type or *p*-type semiconductor region with a moderate doping density, and a metal which will form an ohmic contact is either not available or is not technologically convenient.
* In such cases, a heavily doped region (of the same type) is first created, forming an n^+n or p^+p junction.
* These n^+n or p^+p junctions are essentially ohmic since a large number of majority carriers (of the same type) are available for conduction on both sides of the junction.
* Next, metal is deposited to make a metal-n^+ or metal-p^+ junction, which is ohmic — irrespective of the barrier ϕ_B — because of tunnelling. In this manner, the objective of making a low-resistance metallic contact is achieved. (In practice, metallic contacts also need to be "alloyed" by subjecting them to temperatures of $\sim 450\,^\circ$C for a few minutes.)

pn junction: band diagram with contact regions

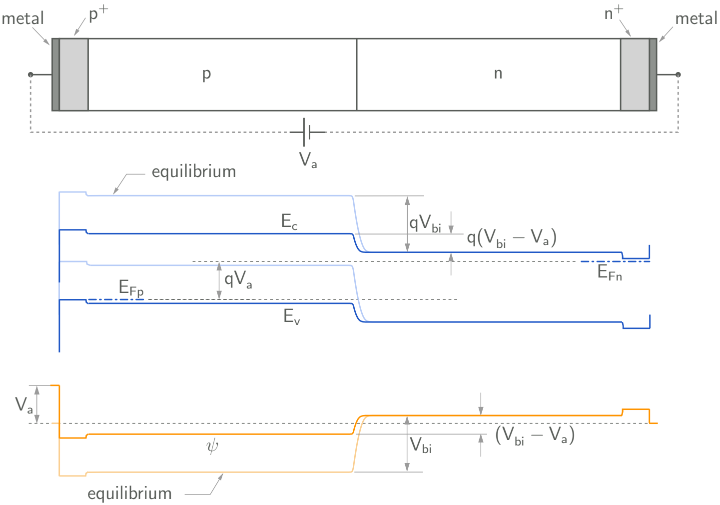


pn junction: band diagram with contact regions

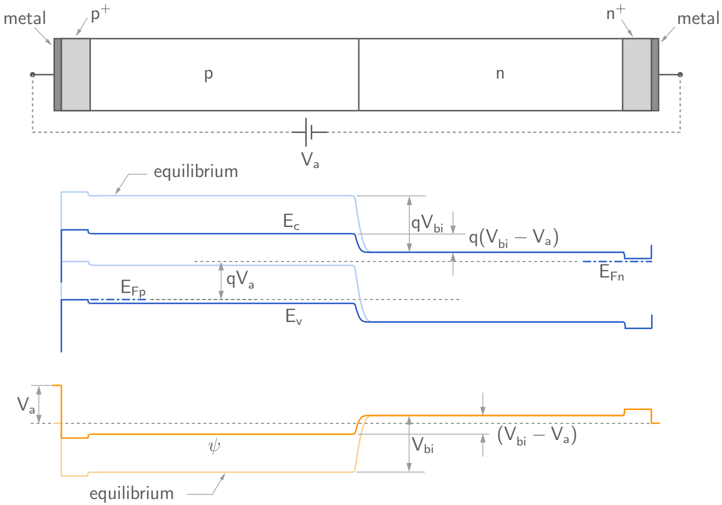


* Equilibrium: The net voltage drop is zero; the voltage drop (V_{bi}) across the depletion region is equal and opposite to the sum of the other voltage drops.

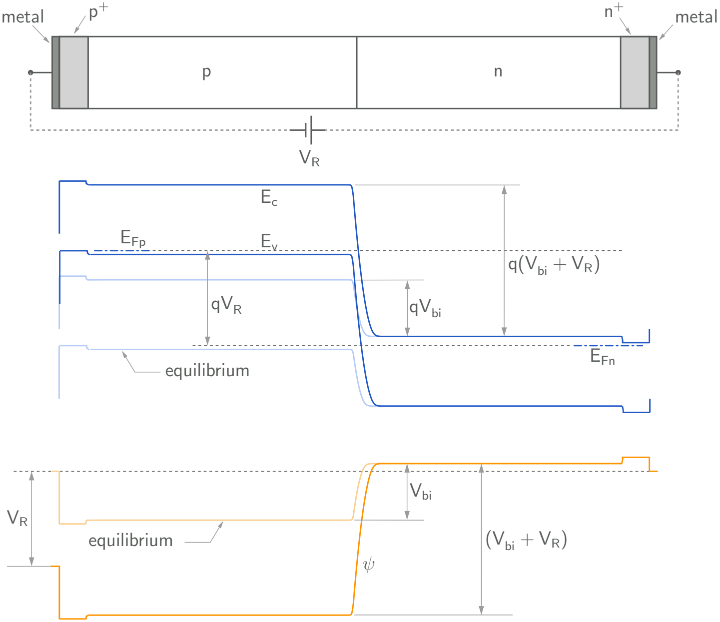
pn junction: band diagram with contact regions

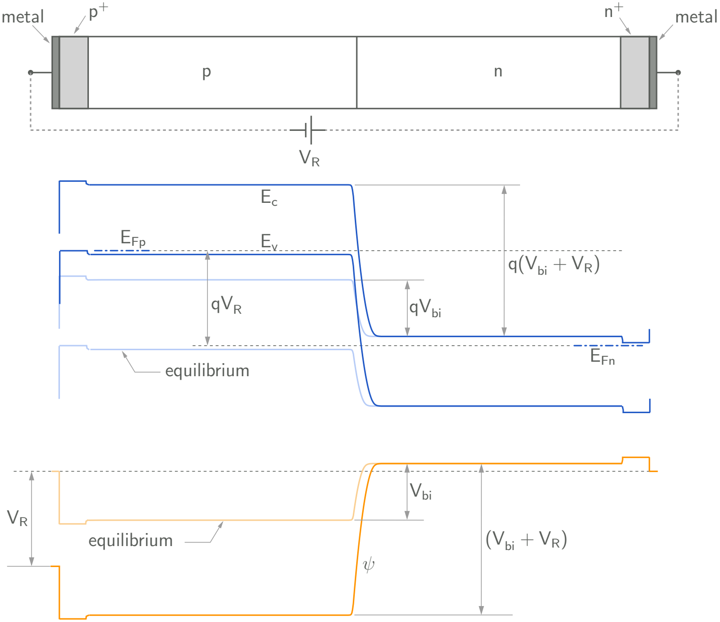


pn junction: band diagram with contact regions



* Forward bias: The voltage drops across the M-S junctions, the n^+-n junction, and the p^+-p junction remain the same as in equilibrium; the applied forward voltage appears across the depletion region ($V_{\mathrm{bi}} \rightarrow V_{\mathrm{bi}} - V_a$).


metal
p+
n+
metal
p
n
V_R
E_c
E_Fp
E_v
q(V_bi + V_R)
qV_R
qV_bi
E_Fn
equilibrium
V_R
V_bi
equilibrium
(V_bi + V_R)
ψ



* Reverse bias: The voltage drops across the M-S junctions, the n^+-n junction, and the p^+-p junction remain the same as in equilibrium; the applied reverse voltage appears across the depletion region ($V_{\text{bi}} \rightarrow V_{\text{bi}} + V_R$).