Semiconductor Optoelectronic Devices

Cheng Wang

Phone: 20685263
Office: SIST 401E
wangcheng1@shanghaitech.edu.cn
About the course

- Content:
  - Part I: Semiconductor electrical and optical properties
  - Part II: Semiconductor Lasers
  - Part III: Photodetectors
  - Part IV: Laser modulators

- 4 credits, 64 credit hours, 16 weeks
- 60% Exame + 40% Projects & Experiments

- Reference:
Background: Fiber-optic communications
# Electromagnetic spectrum

<table>
<thead>
<tr>
<th>Name</th>
<th>Wavelength</th>
<th>Frequency (Hz)</th>
<th>Photon energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma ray</td>
<td>&lt; 0.02 nm</td>
<td>&gt; 15 EHz</td>
<td>&gt; 62.1 keV</td>
</tr>
<tr>
<td>X-ray</td>
<td>0.01 nm – 10 nm</td>
<td>30 EHz – 30 PHz</td>
<td>124 keV – 124 eV</td>
</tr>
<tr>
<td>Ultraviolet</td>
<td>10 nm – 400 nm</td>
<td>30 PHz – 750 THz</td>
<td>124 eV – 3 eV</td>
</tr>
<tr>
<td>Visible light</td>
<td>390 nm – 750 nm</td>
<td>770 THz – 400 THz</td>
<td>3.2 eV – 1.7 eV</td>
</tr>
<tr>
<td>Infrared</td>
<td>750 nm – 1 mm</td>
<td>400 THz – 300 GHz</td>
<td>1.7 eV – 1.24 meV</td>
</tr>
<tr>
<td>Microwave</td>
<td>1 mm – 1 m</td>
<td>300 GHz – 300 MHz</td>
<td>1.24 meV – 1.24 μeV</td>
</tr>
<tr>
<td>Radio</td>
<td>1 mm – 100,000 km</td>
<td>300 GHz – 3 Hz</td>
<td>1.24 meV – 12.4 feV</td>
</tr>
</tbody>
</table>
Optical network architectures

- Long-Haul Networks
  - Optical WDM cross-connect mesh
  - Multiple regional and backbone providers
- Metropolitan Networks
  - Intra-metro connections
  - Inter-metro connections
  - Access Networks/Clients
    - Residential DSL, Cable or FTTH
    - Fixed voice, cellular
    - ISPs
    - Corporate enterprise clients
- Long-haul/metro
  - Frame Relay
  - ESCON Fiber Channel
  - Gigabit Ethernet
Basic optical communication components

- **Optical Transmitters:**
  Convert electrical data into an optical bit stream suitable for transmission.

- **Communication Channel:**
  Optical fibers are used for transmitting optical bit streams in most terrestrial networks.

- **Optical Receivers:**
  Convert optical bit stream into the original electrical form.

Optical transmitters

- **External modulation scheme**

  - Optical source (laser or LED) provides the optical carrier
  - Modulator creates the optical bit stream

- **Direct modulation scheme**

  - Lasers are directly modulated by the current to produce the bit stream, no external modulators required.

---

<table>
<thead>
<tr>
<th>Band</th>
<th>Description</th>
<th>Wavelength Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>O band</td>
<td>original</td>
<td>1260 to 1360 nm</td>
</tr>
<tr>
<td>E band</td>
<td>extended</td>
<td>1360 to 1460 nm</td>
</tr>
<tr>
<td>S band</td>
<td>short wavelengths</td>
<td>1460 to 1530 nm</td>
</tr>
<tr>
<td>C band</td>
<td>conventional (&quot;erbium window&quot;)</td>
<td>1530 to 1565 nm</td>
</tr>
<tr>
<td>L band</td>
<td>long wavelengths</td>
<td>1565 to 1625 nm</td>
</tr>
<tr>
<td>U band</td>
<td>ultralong wavelengths</td>
<td>1625 to 1675 nm</td>
</tr>
</tbody>
</table>
Optical receivers

- Photodetector used for optical-to-electrical conversion.
- Demodulator re-creates the electrical bit stream.
- Noise added during transmission and at receiver leads to errors.
- Bit-error rate (BER) is required to be $< 10^{-9}$.
- All receivers need a certain minimum power to operate reliably.
- This power level is known as the receiver sensitivity.
Optical signal amplification

- Single-mode fibers with low losses (0.2 dB/km near 1550 nm) act as a communication channel.
- Transmission distance is still limited by fiber losses.
- Losses compensated periodically using regenerators or amplifiers.
- Dispersive and nonlinear effects then limit the total distance.
Wavelength-division multiplexing (WDM)

- Each channel is assigned a unique carrier frequency (ITU grid).
- An optical source at a precise wavelength is employed.
- Channel spacing 50 GHz or less for dense WDM.
## Record speeds in one single fiber

<table>
<thead>
<tr>
<th>Year</th>
<th>Organization</th>
<th>Effective speed</th>
<th>WDM channels</th>
<th>Per channel speed</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>NTT[16]</td>
<td>69.1 Tbit/s</td>
<td>432</td>
<td>171 Gbit/s</td>
<td>240 km</td>
</tr>
<tr>
<td>2011</td>
<td>KIT[17]</td>
<td>26 Tbit/s</td>
<td>1</td>
<td>26 Tbit/s</td>
<td>50 km</td>
</tr>
<tr>
<td>2011</td>
<td>NEC[18]</td>
<td>101 Tbit/s</td>
<td>370</td>
<td>273 Gbit/s</td>
<td>165 km</td>
</tr>
<tr>
<td>2012</td>
<td>NEC, Corning[19]</td>
<td>1.05 Petabit/s</td>
<td>12 core fiber</td>
<td></td>
<td>52.4 km</td>
</tr>
</tbody>
</table>
Part I
Semiconductor electrical and optical properties

---- Electrical properties

---- Optical properties
Basics of semiconductors
Valence and conduction bands

- In semiconductors, the energy gap is on the order of 1-2 eV, so some electrons have enough energy to reach the conduction band, leaving holes in the valence band.

- In pure (intrinsic 本征) semiconductors, charge transport is bipolar (electrons and holes), the conductivity is low, and dependent on the gap.

- Dopants can be added to provide large numbers of electrons to the conduction band (donors 施主, n-type) or holes to the valence band (acceptors 受主, p-type).

- PN junction is formed by n- and p-type doped semiconductors.
Semiconductor materials

- Elemental semiconductors (元素半导体)
  - Group IV --- Silicon or Germanium (Si or Ge)

- Compound semiconductors (化合物半导体)
  - Group IV compounds --- Silicon Carbide (SiC), Silicon germanium (SiGe)
  - III-V compounds --- Gallium Arsenide (GaAs), Indium Phosphide (InP), Gallium Nitride (GaN)
  - II-VI compounds --- Zinc oxide (ZnS), Cadmium telluride (CdTe), Mercury telluride (HgTe)
Semiconductor crystal structure

- In crystals, the point of the atomic position is defined as \( r = k\mathbf{a}_1 + l\mathbf{a}_2 + m\mathbf{a}_3 \), where \( k, l, m \) are integer numbers and \( \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \) are the primitive vectors (原矢) denoting the primitive cell (原胞, 晶胞).
The crystal is made up of a periodic arrangement of Bravais lattice, which has 14 possible groups. In semiconductors, only cubic and hexagonal lattices are important. Most semiconductors are cubic (Si, Ge, GaAs, InP), while some are hexagonal (SiC, GaN).

Cubic lattice has three kinds: simple cubic, body-centered cubic and face-centered cubic. Cubic semiconductor crystal structure can be interpreted as two shifted and compenetrated face-centered Bravais lattices.
Miller index

- The miller index (米勒指数) is used to denote planes and reference directions within a lattice.
- Plane \((h,k,l)\) denotes the set of parallel planes that intercepts the three points \(a_1/h, a_2/k,\) and \(a_3/l\). In case, \(h\) is a negative value, it is written as \((\bar{h},k,l)\).
- Direction \([h,k,l]\) denotes the direction orthogonal to the plane \((h,k,l)\).

![Examples of planes and directions according to the Miller notation.](image)

Example 1.1
Electrons in a crystal are characterized by an energy-momentum relation $E(k)$, where the wavevector $k$ is related to the electron momentum $p$ as $p = \hbar k$.

The dispersion relation $E(k)$ is defined in the momentum ($k$, reciprocal) space, periodic in the reciprocal space, whose fundamental period is called the first Brillouin zone (FBZ).

Important points in the FBZ are the center (Γ point), the X point (center of the square face), and the L point (center of the hexagonal face).
The full details of the dispersion relation are not essential for understanding the physics, and attentions can be restricted to the branches of low-energy electrons in the conduction band, and low-energy holes in the valence band. Because most carriers are close the band edges, according to the Boltzmann energy distribution.

Simplified dispersion relation for GaAs.
The diagram shows the lowest branch of the conduction band, the heavy hole band, the light hole band, and the split-off (分裂) band.

The minimum E of conduction band and the minimum E of the valence band has the same momentum k=0 at the Gamma point. This is the case of direct band gap.

The semiconductor bandgap is 1.42 eV at the Gamma point, 1.72 eV at L point, and 1.90 at X point.

The split-off band usually has little contribution to the optical processes, and thus can be neglected to the first-order treatment.
Around the Gamma point, the E-k relation can be approximated by a parabolic function, using the effective mass (a mass considering the impact of material potential field). This is known as the parabolic band approximation.

\[
E_n = E_c + \frac{\hbar^2 k^2}{2m_n^*}, \quad E_h = E_v - \frac{\hbar^2 k^2}{2m_h^*}
\]

The electron’s effective mass is usually smaller than the hole’s effective mass.

For GaAs, \(m_n^* = 0.067m_0\), \(m_h^* = 0.47m_0\).
Conduction and valence band wavefunctions

- In the position space of the crystal, the wavefunctions near the Gamma point (k~0) are single-atom-like functions.
- The wavefunctions are periodic functions in the crystal space.
- The conduction band wavefunctions are *s*-type, that is, they have a probability distribution with spherical constant-probability surfaces.
- The valence band wavefunctions are *p*-type, that is, they are even with respect to two orthogonal directions, and off with respect to the third.

Conduction band (*s*-type) and valence band (*p*-type) wavefunctions: probability distribution and wavefunction sign (for *p*-type).
Direct & indirect bandgap

- Direct bandgap (GaAs): the conduction band minimum and the valence band maximum has the same momentum at the Gamma point.
- Carriers interact directly with photons without momentum change.
- It is able to absorb and emit light efficiently.
Direct & indirect bandgap

- Indirect bandgap (Si): the conduction band minimum and the valence band maximum has the different momentum.
- Carriers interact indirectly with photons with the help of phonons to maintain the momentum conservation.
- It is able to absorb less efficiently but unable to emit light.
Ge is an indirect bandgap semiconductor because the lowest conduction band minimum is at L point.

But optical properties are influenced by the fact that high-energy photons can excite electrons directly from the valence band to the direct minimum. Therefore, Ge’s absorption properties exhibit both indirect- and direct-bandgap semiconductor features, depending on the photon energy.
Effective mass and bandgap

Table 1.4 Main band properties of Si and GaAs. The electron mass $m_0$ is $9.11 \times 10^{-34}$ kg.

<table>
<thead>
<tr>
<th>Property</th>
<th>Si</th>
<th>GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron effective masses</td>
<td>$m_{nl}^* = 0.98m_0$</td>
<td>$m_n^* = 0.067m_0$</td>
</tr>
<tr>
<td></td>
<td>$m_{nt}^* = 0.19m_0$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$m_{n,D}^* = 1.08m_0$</td>
<td>$m_{n,D}^* = 0.067m_0$</td>
</tr>
<tr>
<td></td>
<td>$m_{n,tr}^* = 0.26m_0$</td>
<td>$m_{n,tr}^* = 0.067m_0$</td>
</tr>
<tr>
<td>Hole effective masses</td>
<td>$m_{hh}^* = 0.49m_0$</td>
<td>$m_{hh}^* = 0.45m_0$</td>
</tr>
<tr>
<td></td>
<td>$m_{lh}^* = 0.16m_0$</td>
<td>$m_{lh}^* = 0.08m_0$</td>
</tr>
<tr>
<td></td>
<td>$m_{h,D}^* = 0.55m_0$</td>
<td>$m_{h,D}^* = 0.47m_0$</td>
</tr>
<tr>
<td></td>
<td>$m_{h,tr}^* = 0.37m_0$</td>
<td>$m_{h,tr}^* = 0.34m_0$</td>
</tr>
<tr>
<td>Energy gap $E_g(T), T$ (K)</td>
<td>$1.17 - \frac{4.37 \times 10^{-4} T^2}{636 + T}$</td>
<td>$1.52 - \frac{5.4 \times 10^{-4} T^2}{204 + T}$</td>
</tr>
<tr>
<td>Electron affinity $q\chi$ (eV)</td>
<td>4.01</td>
<td>4.07</td>
</tr>
</tbody>
</table>

Note that the bandgap shrinks with the increased temperature.
III-V semiconductor bandstructures

(a) InP
(b) InAs
(c) AlAs
Carrier densities in semiconductors
The electron and hole populations \( n \) and \( p \) depend on the number of electron and hole states per unit volume in the two bands, and on how those states are populated as a function of the energy.

The density of states (DOS) describes the number of states per unit volume. In the effective mass approximation, the DOS of bulk semiconductors is

\[
g_c(E) = \frac{4\pi}{h^3} \left(2m_n^*\right)^{3/2} \sqrt{E - E_c} \\
g_v(E) = \frac{4\pi}{h^3} \left(2m_h^*\right)^{3/2} \sqrt{E_v - E}
\]

Note that the DOS of the valence band is larger than the conduction band, due to the heavier effective mass.
Fermi-Dirac distributions

- Under the **thermal equilibrium condition**, electrons and holes follow the Fermi-Dirac distribution.

\[
f_n(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{k_B T}\right)}
\]

\[
f_h(E) = \frac{1}{1 + \exp\left(\frac{E_F - E}{k_B T}\right)} = 1 - f_n(E)
\]

- When the Fermi level is within the band gap, it approximately follows the **Boltzmann distribution**:

\[
f_n(E) \approx \exp\left(-\frac{E - E_F}{k_B T}\right); \quad f_h(E) \approx \exp\left(\frac{E - E_F}{k_B T}\right);
\]
The carrier density is determined by the density of states and the occupation probability.

\[
n = \int_{E_c}^{\infty} N_c(E) f_n(E) \, dE = N_c \exp\left(\frac{E_F - E_C}{k_B T}\right)
\]

\[
p = \int_{-\infty}^{E_v} N_v(E) f_h(E) \, dE = N_v \exp\left(\frac{E_v - E_F}{k_B T}\right)
\]

with \(N_c, N_v\) the effective density of states

\[
N_c = 2 \frac{(2\pi m_n^* k_B T)^{3/2}}{h^3}, \quad N_v = 2 \frac{(2\pi m_h^* k_B T)^{3/2}}{h^3}
\]
For intrinsic semiconductor $p=n=n_i$, thus

$$N_c \exp\left(\frac{E_{Fi} - E_C}{k_B T}\right) = N_v \exp\left(\frac{E_V - E_{Fi}}{k_B T}\right)$$

⇒ The intrinsic Fermi level is

$$E_{Fi} = k_B T \log \left(\frac{m_n^*}{m_h^*}\right)^{3/4} + \frac{E_C + E_v}{2}$$

The intrinsic carrier concentration is

$$n_i^2 = n_i n_p = N_c N_v \exp\left(-\frac{E_g}{k_B T}\right)$$

High temperature increases the intrinsic concentration, which can make doping ineffective.
Intrinsic semiconductor

Carrier concentration
In equilibrium conditions, the product of carrier densities $n$ and $p$ does not depend on the position of the Fermi level (while the Fermi level can be changed by the doping). This is coined the mass action law (质量作用定律), which also holds for doped semiconductors.

$$np = n_i^2$$
N-type semiconductor is doped by a donor (density $N_D$) from, say group V, which provides extra electrons in the conduction band. The Fermi level becomes closer to the conduction band.

$$n \approx N_D; \quad p \approx \frac{n_i^2}{N_D}$$
P-type semiconductor is doped by an acceptor (density $N_A$) from, say group III, which provides extra holes in the valence band. The Fermi level becomes closer to the valence band.

$$p \approx N_A; \quad n \approx n_i^2 / N_A$$
Non-equilibrium carrier densities

- The carrier population can be out of equilibrium, with photon or current injection. This non-equilibrium condition is described by the quasi-Fermi distribution. This perturbed condition is described by a quasi-equilibrium distribution using two quasi-Fermi levels $E_{Fn}$ and $E_{Fh}$:

$$f_n(E, E_{Fn}) = \frac{1}{1 + \exp\left(\frac{E - E_{Fn}}{k_B T}\right)}; \quad f_p(E, E_{Fh}) = \frac{1}{1 + \exp\left(-\frac{E - E_{Fh}}{k_B T}\right)}$$

- Within the Boltzmann approximation, the carrier densities become

$$n = N_c \exp\left(\frac{E_{Fn} - E_c}{k_B T}\right); \quad p = N_v \exp\left(\frac{E_v - E_{Fh}}{k_B T}\right)$$

$$np = n_i^2 \exp\left(\frac{E_{Fn} - E_{Fh}}{k_B T}\right)$$

- In case of carrier injection $E_{Fn} > E_{Fh}; \quad np > n_i^2$

- In case of carrier depletion $E_{Fn} < E_{Fh}; \quad np < n_i^2$
Heterostructures & bandstructure engineering
Heterostructures

- Crystals with different lattice constants epitaxially grown together is called heterostructure/heterojunction.
- The material discontinuity arising in the heterojunction leads to improvement of electronic and optical properties, such as confinement of carriers owing to band gap difference, and confinement of photons owing to the refractive index difference.
- Heterostructures can be lattice-matched (same lattice constant) or affected by a slight mismatch (maximum on the order of 1%), which induces tensile or compressive strain. This is named strained heterostructures.
- A double heterostructure made with a thin semiconductor layer (on the order of 100 nm), sandwiched between two layers creates a quantum well.
- A succession of weakly interacting Qwells is called multi quantum wells (MQW).
- If the MQW has many layers, with significant overlapping between the wavefunctions of adjacent wells, we obtain a superlattice.
Heterostructures

Pseudomorphic or strained growth. Above, the epilayer lattice constant is larger than that of the substrate: nonepitaxial growth with interface misfit dislocations and strained epitaxy. Below, the epilayer lattice constant is smaller than that of the substrate.
Heterostructures are largely based on semiconductor alloys. One motivation of alloys is for obtaining certain bandgap, by tailoring the lattice constant, which approximately follow a linear law with respect to the individual component parameters. The other motivation is to achieve lattice matching to the substrate.

Alloys can be made of two (binary), three (ternary), or four (quarternary) elements. AlGaAs is alloy of GaAs and AlAs; InGaAsP is alloy of InAs, InP, GaAs and GaP. By proper selection of the alloy composition, semiconductor alloys emitting the right wavelength and matched to the right substrate can be generated.

Varying the composition of the alloy changes the lattice constant, the bandgap, and the effective mass, and finally the bandstructure and related properties.

Note that GaAs is direct bandgap, and AlAs is indirect. For a large Al fraction, the alloy AlGaAs becomes indirect bandgap.
The lines represent ternary compounds
The dashed lines are indirect gap
The areas enclosed by lines are quaternaries.
The enclosed area provides enough degrees of freedom to adjust the bandgap without changing the lattice constant.
Common substrates: Si, GaAs, InP, SiC, GaSb, GaP, CdTe

2, 4, 6, 8 inch wafers (standard size from 1inch 25.4 mm to 11.8 inch 300mm)
Bandstructure engineering

- Although the bandstructure can be varied through changing the lattice constant, significant variations in the bandstructure cannot be obtained.
- Another way is to introduce quantized effects through heterojunctions, which significantly changes the density of states, and hence optical properties.
- Moreover, strain allows for controlling the degeneracy between heavy and light hole subbands.
- According to the band disalignments, bandstructures are classified into 4 types, and type I is the most commonly used.

Classification of heterostructures according to band alignment: \( \Delta E_j = E_{jB} - E_{jA} \).
Double heterostructure creates potential wells in the conduction and valence band, which can confine carriers to achieve high density, and are able to recombine radiatively. In addition, the emitted radiation is confined by the refractive index step associated with the heterostructure (the refractive index is larger in narrow gap materials). This NiP structure in direct bias that may operate like the active region of a light-emitting diode or a semiconductor laser.

- The carrier confinement is achieved through different band gaps.
- The photon confinement is achieved through refractive index difference.

Example of carrier and light confinement in a \(NiP\) double heterostructure in direct bias.
The 2000 Nobel Prize in Physics

“For basic work on information and communication technology”

“For developing semiconductor heterostructures used in high-speed opto-electronics”

Zhores I. Alferov
Iaffe Physico-Technical Institute, St. Petersburg Russia

Herbert Kroemer
University of California USA
In case the potential well is very narrow, the allowed energy levels of the confined electrons and holes will be quantized, leading to quantum well structure. It has a different bandstructure vs. bulk, where sets of energy subbands appear, and density of states is strongly affected.

The QW bandstructure can be calculated by the Schrodinger equation by a 3D effective mass approximation.
Quantum well

1. Confinement in one direction ($x$): particles are confined along $x$ by a potential well but are free to move along $y$ and $z$ (quantum well).
2. Confinement in two directions ($x$ and $y$): particles are confined along $x$ and $y$, but they are free to move along $z$ (quantum wire).
3. Confinement in three directions ($x$, $y$, $z$): particles are entirely confined and cannot move (quantum dot).
For all allowed energies $E_i$ in a Qwell, the total particle energy is

$$E = E_i + E_k = E_i + \frac{\hbar^2 (k_x^2 + k_y^2)}{2m^*}$$
For the \( l \)-th energy level in the QW, the DOS (per unit area) is

\[
g_l(E) = \begin{cases} 
\frac{m^*}{\pi \hbar^2} & \text{for } E > E_l \\
0 & \text{for } E > E_l 
\end{cases} 
\]

The total DOS for all the energy levels is given by the sum of \( g_l(E) \)

\[
g_{QW}(E) = \sum_l g_l(E)
\]

The carrier density is still given by the DOS and the Fermi distribution

\[
n = \int_{E_c}^\infty g_{QW}(E) f_n(E) dE = \frac{2\pi m^*}{\hbar^2} \sum_l \int_{E_l}^\infty \frac{1}{1 + \exp \left( \frac{E - E_F}{k_B T} \right)} dE
\]
Quantum wires & quantum dots

\[ E = E_c + \frac{\hbar^2}{2m^*}(k_x^2 + k_y^2 + k_z^2) \]

\[ D_x(E) = \frac{1}{2\pi^2} \left( \frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_c} \]

\[ D_x(E) = \sum_{i=1}^{\infty} \frac{m^*}{\pi \hbar^2} \sigma(E - E_i) \]

\[ \sigma \text{ is the Heaviside step function} \]

\[ E = E_i + \frac{\hbar^2}{2m^*}(k_y^2 + k_z^2) \]

\[ E = E_{ij} + \frac{\hbar^2}{2m^*}(k_z^2) \]

\[ D_c(E) = \delta(E - E_{ij}) \]

\[ \delta \text{ is the delta function} \]

\[ E = E_{ijk} \]
Quantum wires & quantum dots

Quantum wires

Quantum dots
Molecular beam epitaxy (MBE) allows several, almost monoatomic, layers to be grown in a controlled and orderly way. The resulting structure is a MQW or, for a large number of (coupled) wells, a superlattice. From the electronic standpoint, N coupled QWs cause the N-fold splitting of the system energy levels, finally leading to subbands, in much the same way as coupled atoms merge their individual energy levels into crystal energy bands. Superlattices are therefore a kind of artificial 3D medium allowing for new features – e.g., low-energy transitions between subbands can be exploited to absorb (emit) long-wavelength IR. Superlattices can also be obtained by a periodic arrangement of QWires or QDs.
Effects of strain on bandstructure

For bulk semiconductors, the strain changes the bandstructure as follows:

- Lattice constant is smaller
- The bandgap is larger, inversely proportional to the lattice constant
- The degeneracy of HH and LH bands is eliminated

For compressive strain:

- Lattice constant is smaller
- The bandgap is larger, inversely proportional to the lattice constant
- The degeneracy of HH and LH bands is eliminated
Carrier generation and recombination
Generation & recombination

- Generation–recombination (GR) of carriers is described by generation and recombination rates for electrons and holes (G_n for the number of electrons generated per unit time and volume, R_n for the number of electrons recombining per unit time and volume, and similarly for holes) and by the electron and hole net recombination rates:

\[ U_n = R_n - G_n; \quad U_h = R_h - G_h; \]

\[ U_n = U_h \quad \text{for interband transitions} \]

- The recombination rate is usually characterized by the \textit{carrier lifetime}, which is inversely proportional to the carrier population. However, \textit{carrier lifetime is constant for minority carrier, determined by the doping density}.

\[ U_n = r_n (p n - n_i^2) \]

\[ \tau_n = (1 / r_n p) \]

For p doping, \( p = N_A \)
GR mechanisms can be phonon-assisted or thermal, photon-assisted or radiative (optical), and, finally, assisted by other electrons or holes. Moreover, generation and recombination can occur through interband transitions (direct mechanisms), or through indirect mechanism assisted by intermediate trap levels in the forbidden band. In direct-bandgap semiconductors, direct optical GR is typically the dominant mechanism, whereas in indirect-bandgap semiconductors trap-assisted GR can be a stronger competitor to the weaker optical GR.
Consider a semiconductor with a trap density \( N_t \), and suppose that traps introduce, in the forbidden gap, a discrete energy level \( E_t \). Thermal carrier transitions from the valence to the conduction bands are made easier by the trap level.

The trap-assisted GR is called Schockley-Read-Hall (SRH) GR. The trap-assisted recombination rate is

\[
U_{SRH} = \frac{pn - n_i^2}{\tau_h^{SRH} (n + n_1) + \tau_n^{SRH} (p + p_1)}
\]

with

\[
\tau_h^{SRH} = 1 / (\tau_{ch}^{SRH} N_t) ; \tau_n^{SRH} = 1 / (\tau_{cn}^{SRH} N_t) ;
\]

\( \tau_{ch,n}^{SRH} \) are trap capture coefficients

\[
n_1 = \frac{n_i}{g} \exp \left(- \frac{E_{Fi} - E_t}{k_B T_0} \right) ; p_1 = n_i g \exp \left( \frac{E_{Fi} - E_t}{k_B T_0} \right);
\]

Note that the SRH lifetime is inversely proportional to the trap density.
The SRH recombination is a unwanted competitor to radiative recombination, in high-injection conditions, especially for indirection-gap semiconductors, which has a long radiative lifetime. In direct-gap semiconductor lasers, the radiative lifetime is shorter, and thus the situation is better.

For Si:

\[
\tau_{SRH}^\text{SRH} = \frac{1}{\tau_c^{\text{SRH}} N_t} = \frac{1}{\left(10^{-6} \text{ cm}^3 / \text{s}\right)\left(10^{14} / \text{cm}^3\right)} = 10\text{ns}
\]
The electron- or hole-assisted recombination is called **Auger recombination**, and the related rate is proportional to $p^2n$ or $pn^2$. The Auger recombination is important (unwanted competitor of radiative recombination) in high-injection devices like semiconductor lasers.

**Impact ionization** (碰撞电离) is an inverse process of Auger recombination. In high-field conditions (i.e., for fields of the order of 100 kV/cm), electrons and holes gather enough energy from the electric field between two successive scattering events (i.e., collisions with phonons, impurities) to be able to interact with another electron and promote it to the conduction band. Each electron or hole is therefore able to generate, over a certain length, a number of electron–hole pairs. The resulting chain can lead to diverging current, i.e., to **avalanche breakdown** in the semiconductor. This effect is used to get **avalanche photodiode** (APD).
Avalanche breakdown occurs for electric fields of the order of the breakdown field, which increases exponentially with the material gap;

Avalanche breakdown fields for some important semiconductors. Insulators (such as C, SiO₂, Si₃N₄) have breakdown fields in excess of $10^4$ kV/cm.