

SEMICONDUCTOR FUNDAMENTALS

Intrinsic carrier concentration in a semiconductor:

- n_i is the concentration of electrons and holes that exist in a semiconductor material due to thermal generation. (An electron-hole pair, EHP, is created whenever an electron escapes from a covalent bond between Si atoms, *i.e.* when an electron moves from the valence band to the conduction band of allowed energy levels.)
- The probability that an electron can escape from a covalent bond increases exponentially with its thermal energy, so that n_i increases rapidly with increasing temperature. In other words, **n_i is a sensitive function of temperature.** For example, if T increases by only 10%, from 300K to 330K, n_i increases by a factor of 10 (from 10^{10} cm^{-3} to 10^{11} cm^{-3})

Bond model vs. band model:

- When an electron escapes from a covalent bond between Si atoms, it becomes a “conduction electron” that is free to move about the Si lattice to conduct current (since it is a charged particle).
 - The resultant half-filled covalent bond (with an associated $+q$ net charge) can also move from one atomic position to another to conduct current, and is called a “hole.”
- The band gap energy is the minimum amount of energy needed to move an electron from a state in the valence band to a state in the conduction band, *i.e.* it is the minimum energy required to remove an electron from a covalent bond so that it is free to roam about the semiconductor lattice. Note that this energy is smaller than the energy needed to break the covalent bond.
 - The conduction-band electron can move easily about the Si lattice because there are many empty allowed energy states within the conduction band.
 - The resultant unoccupied energy state in the valence band (called a “hole”) can also move easily about the Si lattice because an electron can readily move into the unoccupied state from a nearby filled energy state, because there are many filled allowed energy states within the valence band.

Doping:

- Dopant atoms have small ionization energies (relative to the band gap energy), so that we can assume that they are all ionized at typical IC operation temperatures (at or slightly above room temperature).
- Typical dopant concentrations are in the range from $1\text{E}15/\text{cm}^3$ to $1\text{E}20/\text{cm}^3$, which is many orders of magnitude higher than the intrinsic concentration of carriers due to thermal generation (freeing of electrons from covalent bonds due to lattice vibrations).
→ At typical IC operation temperatures, the concentration of majority carriers does not change significantly with increasing/decreasing temperature.

Carrier concentrations in compensated semiconductor material:

- Acceptor atoms in a semiconductor lattice contribute holes. Donor atoms in a semiconductor lattice contribute electrons. If a semiconductor contains both acceptor atoms and donor atoms, the contributed holes and electrons will recombine such that $pn = n_i^2$ in equilibrium, and the majority carrier concentration will be determined by the net doping concentration:
 - If $N_A > N_D$ then the material is **p-type** with hole concentration $p = N_A - N_D$ and electron concentration $n = n_i^2 / (N_A - N_D)$ if $N_A - N_D$ is much greater than n_i . If n_i is comparable to the net dopant concentration, however, then the exact equations must be used to calculate the carrier concentrations accurately.
 - If $N_D > N_A$ then the material is **n-type** with electron concentration $n = N_D - N_A$ and hole concentration $p = n_i^2 / (N_D - N_A)$ if $N_D - N_A$ is much greater than n_i . If n_i is comparable to the net dopant concentration, however, then the exact equations must be used to calculate the carrier concentrations.

Quick determination of carrier concentrations and Fermi level relative to intrinsic Fermi level:

- The position of the Fermi level relative to the intrinsic Fermi level can be quickly deduced by remembering that $kT \cdot \ln(10) = 60 \text{ meV}$:
 - $E_F - E_i = kT \cdot \ln(n/n_i)$. If $n = 10^M \text{ cm}^{-3}$ then $\log(n/n_i) = M - 10$ so $E_F - E_i = (M - 10) \times 60 \text{ meV}$
 - $E_i - E_F = kT \cdot \ln(p/n_i)$. If $p = 10^M \text{ cm}^{-3}$ then $\log(p/n_i) = M - 10$ so $E_i - E_F = (M - 10) \times 60 \text{ meV}$
- If $|E_F - E_i|$ is a multiple of 60 meV, then the carrier concentrations can be quickly deduced:
 - If $E_F - E_i = M \times 60 \text{ meV}$ then $n = 10^{M+10} \text{ cm}^{-3}$
 - If $E_i - E_F = M \times 60 \text{ meV}$ then $p = 10^{M+10} \text{ cm}^{-3}$

Visualization of carrier kinetic energy and potential energy in an energy-band diagram:

- The conduction band edge energy (E_c) corresponds to the potential energy of an electron in the conduction band. The kinetic energy of an electron in the conduction band corresponds to the difference between its total energy (E) and its potential energy, *i.e.* $E - E_c$.
- The valence band edge energy (E_v) corresponds to the potential energy of a hole in the valence band. The kinetic energy of a hole in the valence band corresponds to the difference between its total energy (E) and its potential energy, *i.e.* $E_v - E$.

Visualization of electric field and carrier drift in an energy-band diagram:

- ➔ If there is band bending (*i.e.* E_c and E_v change with distance) then there is a non-zero electric field, since $\mathcal{E} = (1/q)(dE_c/dx)$
 - Electrons “fall” downhill (drift in the direction of lower E_c), while holes “float” (drift in the direction of higher E_v)

Conductivity of a semiconductor:

- The conductivity (σ) of a semiconductor material is dependent on **both** the concentrations of mobile charge carriers (n and p) and the mobilities of the mobile charge carriers (μ_n and μ_p , respectively):

$$\sigma = q\mu_n n + q\mu_p p$$

- Carrier mobility can be limited by phonon scattering or ionized impurity scattering:
 - If mobility is limited by phonon scattering (*i.e.* if the mean time between phonon scattering events is much smaller than the mean time between ionized impurity scattering events), then mobility will decrease with an increase in temperature.
 - If mobility is limited by ionized impurity scattering (*i.e.* if the mean time between ionized impurity scattering events is much smaller than the mean time between phonon scattering events), then mobility will increase with an increase in temperature.
 - Ionized impurity scattering increases with increasing total dopant concentration ($N_A + N_D$), except at very concentrations ($>10^{19} \text{ cm}^{-3}$) due to the screening effect.
 - **Carrier mobility is proportional to the average time between scattering events, not the minority-carrier recombination lifetime!**

Recombination-generation (R-G) of carriers in silicon:

- The minority-carrier recombination lifetime is the average amount of time that it takes for a minority carrier to recombine with a majority carrier.
 - In Si, majority carriers fill trap states that are located energetically near to the middle of the energy band gap. Recombination occurs at the instant that a *minority carrier* is trapped by the trap state. (This is why the recombination lifetime is referred to as the *minority-carrier* recombination lifetime.) The higher the density of trap states, the more quickly a minority carrier is trapped, *i.e.* the smaller the recombination lifetime.
- Note that trap states (R-G centers) have energy levels near the middle of the semiconductor energy band gap, in contrast to dopants which have energy levels near to the band edges. **Therefore, increased dopant concentration does not result in shorter recombination lifetime, *i.e.* the minority-carrier lifetime is not dependent on dopant concentration!**
 - Trap states are associated with crystalline defects and “deep-level” impurities such as gold. To shorten the minority-carrier lifetime, such impurities are added to the semiconductor.
- In general, the net rate of recombination is proportional to $(np - n_i^2)$. R-G always acts to restore a system to equilibrium:
 - If the excess carrier concentrations (Δn and Δp) are greater than zero (so that $np > n_i^2$), then there will be net recombination of carriers.
 - If the excess carrier concentrations are less than zero (so that $np < n_i^2$), then there will be net generation of carriers, *i.e.* the net recombination rate is negative.
 - In equilibrium, the carrier recombination rate equals the carrier generation rate, so that the net rate of recombination is zero.
- Under low-level injection conditions, the net rate of electron-hole recombination is dependent on the excess carrier concentration.

Illumination of a semiconductor with light:

- Electron-hole pairs (EHPs) can be generated within a semiconductor sample by irradiating it with light. The rate of EHP generation, G_L , is dependent on the intensity of the light.
- The rate at which holes and electrons recombine is proportional to the excess carrier concentration. (The more excess carriers there are, the less time it takes for an electron and hole to meet and annihilate each other.) The recombination rate is inversely proportional to the minority carrier lifetime τ . (The lifetime is the average time that a minority carrier “survives” before it is annihilated by recombination; so the smaller the lifetime the higher the rate of recombination.)
- If the electric field within a semiconductor sample is negligible, then the carrier concentrations within it can be increased by irradiation. As the carrier concentrations (and hence the excess carrier concentrations) increase, the rate of recombination increases. Eventually, the rate of generation (G_L) will equal the rate of recombination ($\Delta p/\tau_p$ for n-type semiconductor, $\Delta n/\tau_n$ for p-type semiconductor) and a steady state will be reached (when $\Delta p = G_L\tau_p$ for n-type semiconductor, $\Delta n = G_L\tau_n$ for p-type semiconductor).

METAL-SEMICONDUCTOR CONTACTS

Schottky contact:

- The magnitude of the built-in potential (V_{bi}) is related to the Schottky barrier height (Φ_B) and location of the Fermi level (E_F) in the flatband region of the semiconductor:
 - For a metal contact to n-type semiconductor: $|qV_{bi}| = \Phi_{Bn} - (E_c - E_{FS})_{FB}$
 - For a metal contact to p-type semiconductor: $|qV_{bi}| = \Phi_{Bp} - (E_F - E_v)_{FB}$
 Ideally, this is equal to the magnitude of the difference between the metal work function and the semiconductor work function. In practice, however, this is almost never the case due to trapped charge at the metal-semiconductor interface.
- Only the majority carriers within the semiconductor which have sufficiently high kinetic energy to surmount the potential barrier (which is modulated by the applied voltage) can diffuse into the metal. Only the majority carriers within the metal which have sufficiently high kinetic energy to surmount the Schottky barrier (which does not change with the applied voltage) can diffuse into the metal. At equilibrium (with zero applied voltage), these two components of current flow across the junction balance each other out (so that $I = 0$).
 - Note that the energetic distribution of carriers within the conduction band (and valence band) depends on temperature: the higher the temperature, the more the conduction electrons are “spread out” to higher energy levels associated with higher electron kinetic energy. (Similarly, the higher the temperature, the more the valence holes are “spread out” to lower energy levels associated with higher hole kinetic energy.)
 - Under forward bias, the barrier to majority carrier flow into the metal is decreased, so that current due to majority carrier flow into the metal increases exponentially with the applied voltage and is dominant.
 - Under reverse bias, the barrier to majority carrier flow into the metal is increased, so that current due to majority carrier flow into the metal decreases exponentially; the dominant current is due to the relatively small number of carriers which have sufficient kinetic energy to overcome the Schottky barrier height to flow into the semiconductor.
- The total amount of band-bending in the semiconductor is $q(V_{bi} + V_A)$ for metal contact to p-type semiconductor. (The total amount of band-bending in the semiconductor is $q(V_{bi} - V_A)$ for metal contact to n-type semiconductor.)
 - The depletion width increases with the total amount of band bending.

Ohmic Contact:

- A true ohmic contact ($\Phi_M < \Phi_S$ for contact to n-type semiconductor, or $\Phi_M > \Phi_S$ for contact to p-type semiconductor) is never seen in practice. A Schottky contact can be made to be practically ohmic (allowing current to easily flow in both directions) by doping the semiconductor very heavily so that the width of the depletion region is very narrow (<10 nm) to allow electrons and holes to tunnel easily through it. Since current flows relatively easily through an ohmic contact, the applied bias across an ohmic contact is very small in practice
- ➔ $E_{FM} \cong E_{FS} \cong E_c$ so the Schottky barrier height Φ_M is approximately equal to qV_{bi} .

pn JUNCTION DIODES

Electrostatics:

- When a pn junction is initially formed, mobile charge carriers diffuse across the junction (due to the concentration gradients), so that the semiconductor becomes depleted of majority carriers near the junction. Within the depletion region, the charge density is non-zero ($-qN_A$ on the p-side and qN_D on the n-side); this results in an electric field within the depleted region and hence a built-in electrostatic potential across the depleted region, which result in carrier drift which counteracts carrier diffusion; under thermal equilibrium conditions the drift current components balance the diffusion current components so that the net current is zero.
- The derivative of the electric field ε is directly proportional to the charge density (Poisson's Equation). Thus, the slope of the ε vs. x plot in a depletion region (where the mobile carrier concentrations are approximated to be zero) should be proportional to the dopant concentration in the semiconductor.
 - As the depletion width in a M-S junction or p-n junction changes with the applied bias, the slope of the ε vs. x plot does not change, because the dopant concentration does not change with applied bias.
- The electric field ε distribution is obtained by integrating the charge density distribution. The electric potential distribution $V(x)$ is obtained by integrating the electric field ε distribution. Thus, the larger the potential drop across a semiconductor region, the more integrated charge there must be in the semiconductor.
 - **If the total potential drop across the semiconductor is increased due to an applied bias, then the depletion width is increased** (to satisfy the requirement that the integrated charge must be greater, given that charge density cannot change because it is determined by the dopant concentration, which does not change with an applied bias).

Current flow:

- Charged mobile carriers contribute to current flow via diffusion and drift mechanisms. At various locations along the diode (from the p-side contact to the n-side contact) the total current is the same, but the components of current (hole drift, hole diffusion, electron drift, electron diffusion) can be different.
- Due to the large carrier concentration gradients, there is a strong tendency for holes to diffuse to the n-type side and for electrons to diffuse to the p-type side. The built-in potential counteracts this. The only carriers that diffuse across the junction (to become minority carriers on the other side) are those that have sufficient kinetic energy to overcome the potential barrier.
- In equilibrium, a very small amount of majority carriers diffuse across the junction; this diffusion current is exactly cancelled out by the very small amount of minority carriers that drift across the junction. (One can say that the diffusion current is balanced by the drift current in the opposite direction.)
- A forward bias ($V_A > 0$) reduces the potential barrier to carrier diffusion. The number of carriers with sufficient kinetic energy to overcome the potential barrier increases exponentially with decreasing barrier height - due to the exponential distribution of electrons and holes within the conduction and valence bands - and hence the diffusion current flowing across the junction increases exponentially with increasing forward bias. The drift current flowing in the opposite direction does not depend on the potential barrier height. Therefore, **under forward bias the diode current is dominated by majority carrier diffusion across the junction**, increasing exponentially with increasing forward bias voltage. Generally, the more heavily doped side contributes more to carrier diffusion (hence current flow) across the metallurgical junction under forward bias. The majority carriers which diffuse across the metallurgical junction become minority carriers diffusing within the quasi-neutral regions on the other side. This is referred to as "minority carrier injection" under forward bias.
 - The minority-carrier diffusion length is the average distance that a minority carrier "survives" before it is annihilated by recombination. The shorter the minority-carrier

diffusion length, the more rapidly the minority-carrier concentration goes to zero within the quasi-neutral region, *i.e.* the steeper the minority-carrier concentration profile and hence the larger the minority-carrier diffusion current. This is why each component of the diode current is inversely dependent on the minority-carrier diffusion length.

- A reverse bias ($V_A < 0$) increases the potential barrier to carrier diffusion, so that the number of carriers with sufficient kinetic energy to overcome the potential barrier decreases exponentially with decreasing barrier height and becomes negligible as compared to the drift current flowing in the opposite direction. Therefore, **under reverse bias the diode current is dominated by minority carrier drift across the junction**, which is a weak function of reverse bias voltage. Since the minority carrier concentration is greater on the more lightly doped side, it contributes more to current flow across the junction under reverse bias. The minority carriers which drift across the metallurgical junction become majority carriers within the quasi-neutral regions on the other side. This is referred to as “minority carrier collection” under reverse bias.

Carrier concentrations within the depletion region of a pn junction diode:

- Under forward bias, the product of the carrier concentrations (not each individual carrier concentration) within the depletion region is greater than in equilibrium. Each of n and p can be much less (more than $10\times$ smaller) than the ionized dopant concentration within the depletion region, which is why the depletion approximation is valid.
- Within the depletion region (which is not in equilibrium when $V_A \neq 0$ due to the significant electric field within this region), the electron and hole concentrations can be tracked using the quasi-Fermi levels. It can be shown that F_P and F_N do not change significantly within the depletion region; due to energy-band bending within the depletion region, however, $F_P - E_v$ and $E_c - F_N$ change monotonically within the depletion region and the carrier concentrations decay \sim exponentially (from p_p to p_n for holes; from n_n to n_p for electrons) across the depletion region.

pn junction diode with narrow quasi-neutral region:

- If the quasi-neutral region is much shorter than the minority-carrier diffusion length, then \sim all of the injected minority carriers will diffuse to the metal contact before recombining with majority carriers there. In other words, \sim none of the minority carriers will be annihilated by recombining with majority carriers within the quasi-neutral region and hence the minority-carrier diffusion current will not decay with distance, *i.e.* it will be \sim constant throughout the quasi-neutral region. This means that the gradient (slope) of the minority-carrier concentration profile is constant, *i.e.* the minority-carrier concentration is a linear function of distance, going from its value at the depletion-region edge to zero at the metal contact.
 - The gradient of the minority-carrier concentration profile becomes steeper as the quasi-neutral region becomes narrower. Therefore, the minority-carrier diffusion current increases with decreasing quasi-neutral region width if it is much shorter than the minority-carrier diffusion length.

THE BIPOLAR JUNCTION TRANSISTOR (BJT)

- The common-emitter d.c. current gain is *defined* to be the ratio of collector current to base current (*i.e.* $\beta_{dc} \equiv I_C / I_B$) **when the BJT is biased in the active mode**. When the BJT is biased in saturation mode, $I_C < \beta_{dc} I_B$ because the collector is injecting carriers into the base region – in the direction opposite to carriers (injected from the emitter) being collected into the collector region – so that the net collector current is smaller than in active mode.
- To achieve high emitter injection efficiency in a homojunction BJT, the emitter should be much more heavily doped than the base (so that many more carriers are injected from the emitter into the base, as compared with carriers injected from the base into the emitter). The emitter does NOT necessarily have to be degenerately doped, to achieve high emitter injection efficiency and therefore high common-emitter d.c. current gain.
- **The charge control model (Equation 12.14 in Pierret) is only valid for active mode operation and only if recombination is the dominant component of base current.** (Thus, $\tau_B / \tau_t = \beta_{dc}$ only if recombination is the dominant component of base current.) In modern BJTs, the dominant component of base current is actually injection of carriers into the emitter.
- In saturation mode, both the emitter junction and the collector junction are forward biased. The forward bias on the collector junction is always less than the forward bias on the emitter junction. (Otherwise, the emitter current would not be positive!)
- In active mode, the collector current is essentially equal to the minority carrier diffusion current in the base. (This is because the collector junction is reverse biased, so that the only current which flows is the reverse current due to minority carriers diffusing into the depletion region.)
 - The minority carrier diffusion current is proportional to the gradient (slope) of the minority-carrier concentration profile as well as the minority carrier diffusion constant.
- The base transit time τ_t is the time that it takes a minority carrier to traverse the quasi-neutral base.
 - The smaller the minority carrier diffusion constant, the more slowly minority carriers diffuse across the base and hence the longer the base transit time.
 - The shorter the quasi-neutral base width, the shorter the base transit time.
 - The minority carrier recombination lifetime in the base, τ_B , is not related to τ_t unless recombination current is the dominant component of base current (which is rarely the case in modern BJTs).

THE METAL-OXIDE-SEMICONDUCTOR FIELD-EFFECT TRANSISTOR (MOSFET)

- In the ON-state (when $|V_{GS}| > |V_T|$), the MOSFET operates as a voltage-controlled resistor, since current flow between the source and drain is limited by the rate at which carriers in the inversion layer drift across the channel.
- In the OFF-state (*i.e.* in the sub-threshold regime, when $V_{GS} < |V_T|$), the MOSFET operates similarly to a BJT, since current flow is limited by thermionic injection of carriers over the source-channel junction potential barrier. The rate at which carriers diffuse into the channel (from the source) determines the transistor current.
- For amplifier applications, the MOSFET is biased in the saturation region of ON-state operation. Thus, its operation as an amplifier cannot be compared directly to that of a BJT since the dominant mechanism of MOSFET current flow is carrier drift (not diffusion as in a BJT).
- The field-effect mobility (μ_{eff}) is a strong function of the effective vertical electric field seen by the carriers in the inversion layer. This is because the carriers are attracted to the gate-dielectric interface by the vertical electric field, in the ON state for an enhancement-mode MOSFET. Thus, the carriers scatter more frequently as they drift from the source to the drain, as compared to carriers drifting through a uniformly doped semiconductor resistor with the same lateral electric field. The vertical electric field seen by the inversion layer (in the MOSFET ON state) is dependent on the following parameters:
 - Gate voltage: The higher V_{GS} is, the larger the voltage dropped across the gate dielectric and hence the larger the vertical electric field (in the gate oxide as well as at the channel surface).
 - Gate-dielectric thickness: For a fixed gate voltage (*i.e.* fixed voltage dropped across the gate dielectric), the vertical electric field in the gate oxide -- and hence in the semiconductor channel at the channel surface -- increases with decreasing gate T_{oxe} .
 - Threshold voltage:
 - V_T is larger for a more heavily doped channel. In order to invert the surface in this case, more energy-band bending across a smaller distance (W_T) is needed; therefore, the vertical electric field at the threshold condition is higher.
 - V_T is larger if the body-source pn junction is reverse biased. Recall that the peak electric field in the depletion region of a pn junction increases with increasing reverse bias (due to increasing depletion width, hence increasing integrated charge density on either side of the junction). Similarly, the electric field seen by the inversion layer will increase with increasing reverse body bias.
- The bulk charge factor m accounts for the fact that the areal depletion charge density Q_{dep} (units: C/cm^2) is not constant along the channel; it increases in magnitude toward the drain (*i.e.* with increasing y). Since the total areal charge density in the semiconductor must be constant along the channel, this means that the areal inversion charge density (Q_{inv}) decreases in magnitude toward the drain. The simplistic “square law” I - V equations for the MOSFET are derived based on the assumption that Q_{dep} (hence Q_{inv}) is constant along the channel and therefore they overestimate the transistor current I_{DS} . The “bulk-charge model” I - V equations properly account for the variation in Q_{dep} (hence Q_{inv}) with y , by incorporating the parameter m . Note that since Q_{inv} decreases in magnitude toward the drain, the drain voltage at which the inversion layer becomes pinched off at the drain end, V_{Dsat} , is reduced due to the bulk charge effect.
 - m increases with channel/body doping because the channel depletion depth W_T decreases, so that the areal capacitance between the channel surface and the quasi-neutral body, C_{dep} , increases relative to the areal gate oxide capacitance, C_{ox} .
- The body effect factor γ accounts for the fact that the threshold voltage is dependent on the bias between the body and source. If the body-source junction is reverse-biased, the voltage dropped between the channel surface and quasi-neutral body is increased so that Q_{dep} is increased. Thus, Q_{inv} is decreased, *i.e.* the threshold voltage is effectively increased (since $Q_{inv} \propto V_{GS} - V_T$).