Semiconductor Fundamentals:

- n_i is the electron concentration and the hole concentration in undoped semiconductor material. ($n = p = n_i$ in an undoped semiconductor.)
    - $n_i = 10^{10}$ cm^{-3} in silicon at room temperature (T = 300K)
- A semiconductor can be doped with donor atoms and/or acceptor atoms.
    - If the concentration of donor atoms (N_D) is greater than the concentration of acceptor atoms (N_A), then the semiconductor material is <u>n-type</u>, because the concentration of electrons (n) is greater than the concentration of holes (p):
        - $n = N_D - N_A$
        - $p = n_i^2/n$
    - If $N_A > N_D$, then the semiconductor material is <u>p-type</u>, because $p > n$:
        - $p = N_A - N_D$
        - $n = n_i^2/p$
- The units of ***resistivity*** are ohm-cm.
- The electron and hole mobilities (μ_n and μ_p, respectively) each depend on the <u>total</u> dopant concentration.
    - As $N_D + N_A$ increases, Coulombic carrier scattering occurs more frequently → mobility is lowered.

PN Junctions:

- The ***built-in voltage V_0*** is the potential dropped across the depletion region under zero bias ($V_D = 0$):

$$V_0 = \frac{kT}{q}\ln\left(\frac{N_D N_A}{n_i^2}\right)$$

  where kT/q is the thermal voltage (V_T = 26 mV at room temperature), N_D is the net n-type dopant concentration (N_D-N_A) on the n-type side and N_A is the net p-type dopant concentration (N_A-N_D) on the p-side.

- The current that flows across the metallurgical junction under forward bias ($V_D > 0$, which lowers the potential drop across the depletion region to V_0-V_D) is predominantly due to diffusion of electrons into the p-side and diffusion of holes into the n-side:
  - Current density (A/cm^2) due to **electrons diffusing across the junction**: $J_{n,diff} = \frac{qD_n n_i^2}{L_n N_A}\left(e^{V_D/V_T} - 1\right)$

    where $D_n = V_T \cdot \mu_n$ is the electron diffusion constant within the quasi-neutral p-type region, and L_n is the electron diffusion length within the quasi-neutral p-type region.
    (Note: If the quasi-neutral p-type region is much shorter than L_n – as in the case of the quasi-neutral base region of an NPN bipolar junction transistor – then L_n should be replaced by the length of the quasi-neutral p-type region in the equation above.)
  - Current density (A/cm^2) due to **holes diffusing across the junction**: $J_{p,diff} = \frac{qD_p n_i^2}{L_p N_D}\left(e^{V_D/V_T} - 1\right)$

    where $D_p = V_T \cdot \mu_p$ is the hole diffusion constant within the quasi-neutral n-type region, and L_p is the hole diffusion length within the quasi-neutral n-type region.
    (Note: If the quasi-neutral n-type region is much shorter than L_p – as in the case of the quasi-neutral base region of a PNP bipolar junction transistor – then L_p should be replaced by the length of the quasi-neutral n-type region in the equation above.)

  By adding the currents due to electron and hole flow across the junction, we obtain the total diode current:

$$I_D = AJ_{p,diff} + AJ_{n,diff} = Aqn_i^2\left[\frac{D_n}{L_n N_A} + \frac{D_p}{L_p N_D}\right]\left(e^{V_D/V_T} - 1\right) \equiv I_S\left(e^{V_D/V_T} - 1\right)$$

  where A is the junction area and I_S is the diode reverse saturation current.

  Note that the dominant current component flowing across the metallurgical junction is that associated with minority-carrier diffusion in the more lightly doped side, *i.e.* most of the current flowing across the junction is due to minority carriers injected from the more heavily doped side.

  Note that if the dopant concentration is increased on one side, the minority-carrier diffusion current on that side is reduced.
  - Example: If the dopant concentration on the n-type side is increased by a factor of 10 (*i.e.* if N_D is increased by 10×), the hole diffusion current is reduced by a factor of 10 (*i.e.* $J_{p,diff}$ is reduced by 10×). As a result, $J_{n,diff}/J_{p,diff}$ is increased by 10×.

- The current that flows across the metallurgical junction under reverse bias ($V_D < 0$, which raises the potential drop across the depletion region to $V_0 + |V_D|$) is predominantly due to drift of electrons from the p-type side into the n-side and drift of holes from the n-type side into the p-side. The reverse-bias current (I_S) is limited by the rate at which minority carriers diffuse from the quasi-neutral regions into the depletion region (*i.e.* the rate at which electrons on the p-type side diffuse to the edge of the depletion region and the rate at which holes on the n-type side diffuse to the edge of the depletion region).
  - Since $I_D \cong -I_S$ for $V_D < 0$, we can deduce from the formula for I_S that $Aqn_i^2\left[\frac{D_n}{L_n N_A}\right]$ is the reverse-bias current component due to electrons (which diffuse in the quasi-neutral p-type region to the depletion

region and are then swept across the junction into the quasi-neutral n-type region by the electric field) and $Aqn_i^2\left[\frac{D_p}{L_p N_D}\right]$ is the reverse-bias current component due to holes (which diffuse in the quasi-neutral n-type region to the depletion region and are then swept across the junction into the quasi-neutral p-type region by the electric field).

Bipolar Junction Transistors:
In the forward active mode of operation:

- The collector current consists primarily of minority carriers collected from the base (which then proceed to drift in the quasi-neutral collector region and out through the collector contact). The rate at which these minority carriers are collected is limited by the rate at which they diffuse across the quasi-neutral **base** region and enter the collector-junction depletion region.

$$I_C = \frac{A_E q D_B n_i^2}{N_B W_B}\left(e^{|V_{BE}|/V_T} - 1\right) \cong I_S\left(e^{|V_{BE}|/V_T} - 1\right)$$

  where A_E is the area (in cm^2) of the emitter-base junction, D_B is the minority-carrier diffusion constant within the quasi-neutral base region, N_B is the net dopant concentration in the base, and W_B is the width of the quasi-neutral base region. Note that the pre-exponential factor is approximately I_S. This is because the reverse saturation current of the base-emitter PN junction is $I_S = A_E q n_i^2\left[\frac{D_B}{N_B W_B} + \frac{D_E}{N_E W_E}\right] \cong \frac{A_E q D_B n_i^2}{N_B W_B}$ since the emitter dopant concentration (N_E) is typically much higher than the base dopant concentration (N_B).

- The base current is primarily comprised of carrier flow from the base into the emitter:

$$I_B = \frac{A_E q D_E n_i^2}{N_E W_E}\left(e^{|V_{BE}|/V_T} - 1\right) \equiv \frac{I_C}{\beta}$$

  where A_E is the area (in cm^2) of the emitter-base junction, D_E is the minority-carrier diffusion constant within the quasi-neutral emitter region, N_E is the net dopant concentration in the base, and W_E is the width of the quasi-neutral emitter region. (In modern BJTs, the emitter is "short", *i.e.* much shorter than the minority-carrier diffusion length.)

  Note that the equation for I_B looks like that of a PN-junction diode, except that the pre-exponential factor contains only one (small) term, associated with diffusion of minority carriers across the quasi-neutral **emitter** region. This is why the base-emitter junction is modeled as a diode in the large-signal model of a BJT.

- Note that only I_C depends on W_B, so that only I_C shows the "Early effect" (increasing current with increasing reverse bias on the collector junction). The common-emitter current gain β is thus a function of $|V_{CE}|$ due to the Early effect.

Circuit Analysis:

- If two impedances Z_1 and Z_2 are connected in parallel, their combined impedance is

$$Z_1 \| Z_2 = \frac{Z_1 Z_2}{Z_1 + Z_2}$$

  which is smaller than either Z_1 or Z_2.

  Note that if $Z_2 >> Z_1$, then $Z_1 \| Z_2 = \frac{Z_1 Z_2}{Z_1 + Z_2} \cong \frac{Z_1 Z_2}{Z_2} = Z_1$.

  For example, in the small-signal analysis of a BJT circuit, $r_\pi \| \frac{1}{g_m} = \frac{\beta}{g_m} \| \frac{1}{g_m} \cong \frac{1}{g_m}$

MOSFETs:

- **Threshold Voltage:**
  A MOSFET is in the on state (*i.e.* mobile charge carriers can flow from the source to the drain under the influence of a lateral electric field) when an inversion layer is formed in the channel region.
  - In the equations for MOSFET current, the source voltage is used as the reference voltage. Thus, the threshold voltage (V_{TH}) of a MOSFET is defined to be the gate-to-source voltage (rather than the gate-to-body voltage) required to form an inversion layer in the channel region.
    - Recall that the voltage dropped within the semiconductor at the threshold condition is $|2\phi_B|$, so that

      $$V_{TH} = V_{FB} + 2\phi_B + \frac{\sqrt{2qN_A\varepsilon_{Si}(2\phi_B + V_{SB})}}{C_{ox}}$$

      for an n-channel MOSFET. (The second term on the right-hand side represents the voltage dropped within the semiconductor.) The last term in this equation represents the voltage dropped within the gate oxide, V_{ox}, which is proportional to the total areal charge density (units: C/cm^2) in the semiconductor (from Gauss' Law – ref. Lecture 15, Slide 17), which is approximately equal to the areal depletion charge density (since the inversion charge density at the threshold condition is negligible in comparison):
      $Q_{dep} = -\sqrt{2qN_A\varepsilon_{Si}(2\phi_B + V_{SB})}$.
      (For a p-channel MOSFET, the depletion charge in the semiconductor is of opposite sign:
      $Q_{dep} = +\sqrt{2qN_D\varepsilon_{Si}(2\phi_B + V_{SB})}$. Thus, the voltage dropped across the gate oxide is negative. This is why the last term in the V_{TH} equation has a negative sign, for a p-channel MOSFET – ref. Lecture 16, Slide 8. Note that $2\phi_B$ is negative for a p-channel MOSFET, as well.)
  - If the body voltage is not equal to the source voltage, (*i.e.* if $V_{SB} \neq 0$) then V_{TH} will be altered according to the following equation (for an n-channel MOSFET):

    $$V_{TH} = V_{FB} + 2\phi_B + \frac{\sqrt{2qN_A\varepsilon_{Si}(2\phi_B + V_{SB})}}{C_{ox}} = V_{TH0} + \gamma\left(\sqrt{2\phi_B + V_{SB}} - \sqrt{2\phi_B}\right)$$

    $$\text{where} \quad \phi_B \equiv \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) \quad \text{and} \quad \gamma \equiv \frac{\sqrt{2qN_A\varepsilon_{Si}}}{C_{ox}}$$

    (V_{TH0} is defined to be the threshold voltage for V_{SB} = 0V.)
    - **Note that V_{SB} does not directly affect V_{TH}, *i.e.* the change in V_{TH} with V_{SB} is not simply equal to V_{SB}!** Rather, V_{SB} affects the areal depletion charge density in the semiconductor (Q_{dep}) and thereby the voltage drop across the gate oxide (V_{ox}=Q_{dep}/C_{ox}).
      - If the **body-source PN junction is reverse biased** (*i.e.* if $V_{SB} > 0$), then the **depletion depth increases** and hence the **depletion charge in the body is increased**; thus, the **magnitude of V_{TH} is increased.**
      - From the equation for V_{TH}, we can see that the more heavily the body is doped (*i.e.* the larger N_A is), the more influence the body bias has on the channel potential and therefore on V_{TH}. This is because the areal depletion capacitance (C_{dep}) increases with N_A so that the capacitive coupling between the channel potential and the body voltage is increased.

- **Current flow in the on state ($V_{GS} > V_{TH}$, for an n-channel MOSFET; $V_{GS} < V_{TH}$, for a p-channel MOSFET):**
  - The drain current I_D is proportional to the amount of areal inversion charge density (units: C/cm^2), Q_{inv}, in the channel region at the source end, which in turn depends on V_{GS}:

$$|Q_{inv}| = |C_{ox}(V_{GS} - V_{TH})|$$

→ As $|V_{GS}-V_{TH}|$ increases, $|Q_{inv}|$ increases hence I_D increases, for a fixed drain-to-source voltage difference, V_{DS}.

Long-channel n-MOSFETs:

- For a fixed value of $V_{GS} > V_{TH}$:
  - o $V_{DS} < V_{GS}-V_{TH}$: ("Triode" region of operation)
    - As V_{DS} increases, the voltage applied across the inversion layer increases; hence the lateral electric field increases and so the carrier drift velocity increases → I_D increases.
      - The rate at which I_D increases with increasing V_{DS} gets smaller as V_{DS} increases, however, because the average potential in the inversion layer increases with increasing V_{DS}, so that the average value of Q_{inv} (from the source end to the drain end) decreases with increasing V_{DS}. Eventually, I_D reaches a "maximum" value when $V_{DS} = V_{GS}-V_{TH}$.
  - o $V_{DS} = V_{GS}-V_{TH} \equiv V_{Dsat}$: ("Pinch-off" or "Edge of Saturation")
    - The channel potential at the drain end is too high to allow an inversion layer to form there. (The voltage difference between the gate and the channel at the drain end is equal to V_{TH}, so $Q_{inv} = 0$ there.) Thus, the inversion layer is "pinched off" just at the drain end of the channel.
  - o $V_{DS} > V_{Dsat}$: ("Saturation" region of operation)
    - The channel potential near to the drain end is too high to allow an inversion layer to form there. (The voltage difference between the gate and the channel near to the drain is smaller than V_{TH}.) Thus, there is no inversion layer in the region near the drain end of the channel, called the "pinch-off region."
    - **As V_{DS} increases beyond $V_{GS}-V_{TH} \equiv V_{Dsat}$,** more voltage ($V_{DS}-V_{Dsat}$) is dropped across the pinch-off region; whereas **the same amount of voltage ($V_{GS}-V_{TH}$) is dropped across the inversion layer.**
    - **Current flow is limited by the rate at which mobile charge carriers reach the pinch-off region** (where they are quickly swept across the pinch-off region – which is the depletion region of the reverse-biased body-drain PN junction – into the drain). **This in turn is determined by the average lateral electric field in the inversion layer, ~$(V_{GS}-V_{TH})/L_1$, which is ~independent of V_{DS}.** (L_1 is the length of the inversion layer.) **Thus, I_D "saturates."** (I_D does not increase rapidly with increasing V_{DS} in the saturation region of operation.)
    - **If the length of the pinch-off region (L-L_1) is a substantial fraction of the channel length**, then I_D increases noticeably with increasing V_{DS}. This is because the size of the pinch-off region grows as the voltage dropped across it ($V_{DS}-V_{Dsat}$) increases, so **the length of the inversion-layer (L_1) decreases substantially with increasing V_{DS}**; thus, the lateral electric field in the inversion layer (~$(V_{GS}-V_{TH})/L_1$) increases **so that the drift velocity of the mobile charge carriers (hence I_D) increases substantially with increasing V_{DS}.**
      - The fractional increase in I_D with V_{DS} in the saturation region of operation is $\lambda(V_{DS}-V_{Dsat})$. λ is the "channel length modulation coefficient."
      - **The "channel-length modulation" effect becomes more significant as *L* is decreased, since the size of the pinch-off region relative to the channel length becomes larger.** This is why λ is approximately proportional to the inverse of L.
      - MOSFET small-signal model: The incremental increase in I_D with an incremental increase in V_{DS} is modeled as an extra resistance (r_o), in parallel with the voltage-dependent current source ($g_m v_{gs}$) that models the *change* in I_D due to a *change* in V_{GS}, between the drain and source terminals. If r_o is large, this means that the incremental change in I_D with an incremental increase in V_{DS} is small, *i.e.* λ is small. ($r_o \propto 1/\lambda$)

Short-channel n-MOSFETs:

A short-channel MOSFET is one in which I_D "saturates" with increasing V_{DS}, due to velocity saturation: As V_{DS} increases above the saturation voltage $E_{sat}L$ – where E_{sat} is the electric-field strength at which the saturation velocity is reached – the carrier velocity does not increase (even though the average lateral electric field in the inversion layer increases) so I_D does not increase.

- For a fixed value of $V_{GS} > V_{TH}$:
    - $V_{DS} < E_{sat}L$: ("Triode" region of operation)
        - As V_{DS} increases, the voltage applied across the inversion layer increases; hence the lateral electric field increases and so the carrier drift velocity increases → I_D increases.
            - The rate at which I_D increases with increasing V_{DS} gets smaller as V_{DS} increases, however, because the average potential in the inversion layer increases with increasing V_{DS}, so that the average value of Q_{inv} (from the source end to the drain end) decreases with increasing V_{DS}.
    - $V_{DS} = E_{sat}L \equiv V_{Dsat}$: ("Edge of Saturation")
        - The mobile charge carriers in the inversion layer reach maximum drift velocity.
    - $V_{DS} > E_{sat}L$: ("Saturation" region of operation)
        - I_D does not increase rapidly with increasing V_{DS} because the velocity of the mobile charge carriers is limited to a maximum of the saturation velocity, v_{sat}. I_D is not strongly dependent on V_{DS}/L (the average lateral electric field in the channel); it is simply proportional to the width (W) of the transistor channel region and the areal inversion charge density at the source end:

$$I_D = Wv_{sat}C_{ox}(V_{GS}\text{-}V_{TH})$$

        - As V_{DS} increases, the reverse bias on the body-drain PN-junction increases and so the depletion region in the body at the drain junction increases. This helps to deplete the channel region underneath the gate electrode near to the drain, so that less charge is needed on the gate in order to form an inversion layer in the channel region, *i.e.* the threshold voltage (V_{TH}) is reduced. Since $I_D \propto (V_{GS}\text{-}V_{TH})$, this means that I_D increases with increasing V_{DS}. This effect is called "drain-induced barrier lowering" (DIBL).
            - The fractional increase in I_D with V_{DS} in the saturation region of operation is $\lambda(V_{DS}\text{-}V_{Dsat})$. (Note that the DIBL effect is modeled in the same way as channel-length modulation for a long-channel MOSFET.)
            - DIBL becomes more significant as L is decreased, since the size of the depletion region in the body at the drain junction relative to the channel length becomes larger. This is why λ is approximately proportional to the inverse of L, for a short-channel MOSFET.
            - MOSFET small-signal model:
                - The incremental increase in I_D with an incremental increase in V_{DS} is modeled as an extra resistance (r_o), in parallel with the voltage-dependent current source ($g_m v_{gs}$) that models the change in I_D due to a change in V_{GS}, between the drain and source terminals. If r_o is large, this means that the incremental change in I_D with an incremental increase in V_{DS} is small, *i.e.* λ should be small. ($r_o \propto 1/\lambda$)
                - **Note that the transconductance of a MOSFET operating in the velocity-saturation-limited regime is independent of V_{GS}: $g_m = Wv_{sat}C_{ox}$**
            - To reduce DIBL (*i.e.* to **increase r_o**), we need to reduce the lateral extent of the depletion region in the body at the drain junction. This can be done **by increasing the doping in the channel region near to the drain junction**. (In the semiconductor industry, this type of doping is referred to as "halo doping.")