Microwave Semiconductor Devices

First division: a) active, and b) pasive. Active devices used in amplifiers and oscillators and pasive in mixers.

Two types of metal contacts to semiconductor: ohmic and rectifying. Ideal ohmic contacts conduct current in both directions, unlike the rectifying one. The only difference is in metal used and/or type of material used for semiconductor.

Schottky diode: simple, easy to fabricate and usefull all the way to THz region! Also forms a gate in MESFETs. Used in Si as well as GaAs (or III-V compounds in general). It is a **unipolar** device, *i.e.*only electrons (holes) are involved in conduction and its control.

Energy Band Diagram: Unlike the p-n junction, here we have contact between two different materials. A few definitions:

Fermi Level: energy level at which probability of finding an electron is 0.5

Electron afinity: energy required to excite electron from the conduction band edge to the vacuum level (*i.e.*outside of the semiconductor)

Work function: energy required to excite electron from the Fermi level to the vacuum level

Fermi levels must line up in equilibrium! In addition, vacuum level must be continuous (see diagrams).

Barrier height $q\phi_{Bn} = q(\phi_m - \chi)$ in n-type semic., or $q\phi_{Bp} = E_g - q(\phi_m - \chi)$

Measured barrier heights are different from the above equation. Reasons: surface states in the forbidden bandgap.

Analysis is similar to p-n junction and we need to determine the barrier height: $V_{bi} = \phi_{Bn} - V_n$, where V_n is the potential difference between Fermi level E_F and conduction band edge E_C. Standard approximation: **abrupt junction**, which means that electric field and space charge are zero outside the depletion layer.

$$W = \sqrt{\frac{2\epsilon_s}{qN_D}(V_{bi} - V)} \tag{1}$$

$$|\mathcal{E}(x)| = \frac{qN_D}{\epsilon_s}(W - x) = \mathcal{E}_m - \frac{qN_D}{\epsilon_s}x \tag{2}$$

$$\psi(x) = \frac{qN_D}{\epsilon_s}\left[Wx - \frac{1}{2}x^2\right] - \phi_{Bn} \tag{3}$$

where V is applied voltage (> 0 for forward bias, < 0 for reverse bias), and $\mathcal{E}_m$ is the maximum field $\mathcal{E}_m = 2(V_{bi} - V)/W$.

Another important characteristic of Schottky diode is its capacitance, which is defined as change of charge per change of voltage:

$$C = \left|\frac{\partial Q_{sc}}{\partial V}\right| = \sqrt{\frac{q\epsilon_s N_D}{2(V_{bi} - V)}} = \frac{\epsilon_s}{W} \qquad F/cm^2 \tag{4}$$

This can be further rewritten as:

$$\frac{1}{C^2} = \frac{2(V_{bi} - V)}{q\epsilon_s N_D} \qquad \frac{-d(1/C^2)}{dV} = \frac{2}{q\epsilon_s N_D} \tag{5}$$

The measurements of the capacitance C per unit area as a function of voltage provides the impurity distribution and the built-in potential directly.

I-V Characteristics

Unlike p-n junction, in Schottky diode only the majority carriers contribute to the current. The mechanism of electron transport is different than in p-n junction where we have drift and diffusion. Here another mechanism takes over: thermionic emission.

Thermionic emission describes those electrons that are energetic enough to overcome the potential barrier and either go from metal into semiconductor ($m \rightarrow s$) or from semiconductor to metal ($s \rightarrow m$). Remember that **electron temperature** describes average energy of an ensemble of electrons.

In equilibrium ($V_{applied} = 0$) two electron flows must cancel each other, $i.e. J_{m\to s} = -J_{s\to m}$. Both of these components are proportional to the electron concentration at the semiconductor surface n_s. Remember that electron concentration is proportional to the density of states N_C and exponentially dependent on difference between Fermi level and conduction band edge, $i.e. n = N_C \exp\left(-\frac{E_C - E_F}{kT}\right)$. At the surface of Schottky diode, we know that $E_C - E_F = q\phi_{Bn}$ and can, therefore, write:

$$|J_{m\to s}| = |J_{s\to m}| = C_1 N_C \exp\left(-\frac{q\phi_{Bn}}{kT}\right) \tag{6}$$

where C_1 is a proportionality constant.

Outside of equilibrium ($V_{applied} \neq 0$) we must have some dependence of n_s on applied voltage. Not surprisingly, this dependence is again exponential

$$n(s) = N_C \exp\left(-\frac{q(\phi_{Bn} - V_F)}{kT}\right) \tag{7}$$

When forward (reverse) bias is applied, the potential difference between two sides is reduced (increased); however, the barrier for electron flow is very different, depending on whether the electron is coming from metal or from semiconductor. If the electron is coming from the metal, the barrier does not change which implies that $J_{m\to s}$ does not change relative to equilibrium. On the other hand, $J_{s\to m}$ changes quite dramatically, in proportion to the above equation for n_s.

The net current across the junction is

$$J = J_{s\to m} - J_{m\to s} = C_1 N_C \exp\left(-\frac{q\phi_{Bn}}{kT}\right)\left[\exp\left(\frac{qV_F}{kT}\right) - 1\right] \tag{8}$$

For reverse bias, V_F is replaced by $-V_R$. Furthermore, the coefficient $C_1 N_C$ is equal to $A^* T^2$, where A^* is called the effective Richardson constant and T is temperature. The constant terms can be collected in one constant resulting in

$$J = J_s\left[\exp\left(\frac{qV}{kT}\right) - 1\right] \text{ and } J_s \equiv A^* T^2 \exp\left(-\frac{q\phi_{Bn}}{kT}\right) \tag{9}$$

In Schottky diodes there is a hole current which is due to hole injection from the metal to the semiconductor. This injection is the same as in $p^+ - n$ junction. In normal operation conditions it is much smaller than the electron component (why?) resulting in a unipolar operation.

Ideal ohmic contact should conduct in either direction and should not have any voltage drop across the metal-semiconductor interface. This is only approximately true in real ohmic contacts. Two techniques are used: 1. utilizing a metal with low barrier height (ϕ_{Bn}), and 2. high doping. The conduction can be dominated by either thermionic emission or tunneling.

The JFET

Why bother with JFET? It is good introduction to MESFETs which are much more common and the principle of operation is very similar.

JFET is basically a resistor that has a control gate. Voltage applied at the control gate shrinks or opens up the **channel** for electrons to flow through. The wider the channel, the smaller the resistance and vice versa. See simplified picture. The output electrode is called drain, the name "gate" is used for control gate, and the other contact, which is usually grounded, is called source.

Under normal operating conditions $V_G \leq 0$ and $V_D \geq 0$ (why?). For p-channel device, polarities are reversed, but p-channel is very rarely used. What is the resistance of a channel in the equilibrium?

$$R = \rho \frac{L}{A} = \frac{L}{q\mu_n N_D A} = \frac{L}{2q\mu_n N_D Z(a-W)} \tag{10}$$

where N_D is donor concentration, A is the cross-sectional area for current flow ($= 2Z(a-W)$), and W is the depletion layer width of upper and lower p-n junctions.

Case a); keep $V_G \approx 0$ and V_D small. Then $I_D = V_D/R$. Output current I_D varies linearly with output voltage V_D

Case b): $V_G \approx 0$, V_D is allowed to change up to "pinch-off." The following must be taken into account: the voltage difference between drain and source must occur along the channel. This results in higher potential near the drain contact and lower potential near the source contact. Since we now allow V_D to increase at will, the potential difference between channel and gate is going to vary: near the drain the gate-channel junction is more reverse biased than it is at the source side. This can be simply seen from evaluating $V_{G-channel}$: at the drain side it is equal to V_D (> 0) but on the source side it is exactly equal to zero.

Therefore Case b) results in variable resistance R, and we can expect that the rate of increase of current should decrease as drain voltage is increased, i.e. R increases with V_D

Case c): the two depletion regions touch at the drain side. This is called pinch-off, and the device enters the saturation region. Voltage at which this happens is called saturation voltage $V_{D,sat}$ and is calculated from

$$V_{D,sat} = \frac{qN_D a^2}{2\epsilon_s} - V_{bi} \qquad \text{for } V_G = 0 \tag{11}$$

At pinch-off a large drain current flows, called saturation current $I_{D,sat}$ and drain and source are completely (almost) isolated from each other by a depletion region. This is similar to the situation in BJT where electrons, after traversing the base region, are injected into the collector depletion region.

Beyond the pinch-off increase in V_D cannot increase I_D any more. This can be explained as follows: Any increase of V_D above saturation voltage is used up only to extend the deplation region away from the drain and into the channel region. The point P at which the potential is equal to $V_{D,sat}$ is pushed away from drain into the channel. So long as the change of the channel length is small compared to the channel length (so called long channel approximation), the current is still going to be $I_{D,sat}$ since the potential drop is from point P to source is still the same and the resistance (to the first order) is unchanged.

If the gate is reverse biased, that only makes the starting resistance R larger, but the same reasoning as above applies. To obtain $V_{D,sat}$ the value of V_G must be subtracted, $i.e.$saturation voltage decreases with increase in magnitude of V_G ($V_G < 0$ for reverse bias in n-channel device). Important: gate voltage is actually negative with respect to source, but from now on when V_G is used it means the **magnitude** of the applied gate voltage.

I-V Characteristics

Derivation of I_D vs. V_D, V_G proceedes as follows:

1. find the voltage drop across a small cross-section of the channel

$$dV = I_D dR = \frac{I_D dy}{2q\mu_n N_D Z[a - W(y)]} \text{ or} \tag{12}$$

$$I_D dy = 2q\mu_n N_D Z[a - W(y)]dV \tag{13}$$

   (recall the expression for R).

2. for abrup p-n junction the deplation layer width is

$$W(y) = \sqrt{\frac{2\epsilon_s[V(y) + V_G + V_{bi}]}{qN_D}} \tag{14}$$

3. variations of voltage and depletion layer width are related through (take square of above, take total differential of both sides)

$$dV = \frac{qN_D}{\epsilon_s} W dW \tag{15}$$

4. substitute dV into eq. for current and integrate (or, thihk of it as summation) with respect to y from y=0 to y=L. This results in

$$I_D = \frac{1}{L}\int_{W_1}^{W_2} 2q\mu_n N_D Z(a-W)\frac{qN_D}{\epsilon_s}W\,dW$$
$$\ldots$$
$$= I_P\left[\frac{V_D}{V_P} - \frac{2}{3}\left[\frac{V_D+V_G+V_{bi}}{V_P}\right]^{3/2} + \frac{2}{3}\left[\frac{V_G+V_{bi}}{V_P}\right]^{3/2}\right]$$

where $I_P = Z\mu_n q^2 N_D^2 a^3/(\epsilon_s L)$ and $V_P = qN_D a^2/(2\epsilon_s)$. V_P is called pinch-off voltage, *i.e.*it is the voltage across the junction at the drain end when $W_2 = a$.

Two distinct regions: a) Linear region, and b) Saturation region.

Linear region: assumption is $V_D \ll V_G + V_{bi}$ and the full equation can be approximated by

$$I_D \simeq \frac{I_P}{V_P}\left[1 - \sqrt{\frac{V_G+V_{bi}}{V_P}}\right] \tag{16}$$

- One interesting parameter: output (channel) conductance, defined as $\partial I_D/\partial V_D$ with V_G kept constant

$$g_D = \frac{I_P}{V_P}\left[1 - \sqrt{\frac{V_G + V_{bi}}{V_P}}\right] \tag{17}$$

What is the value of g_D beyond pinch-off?

Saturation region: evaluate drain current at pinch-off point — it stays constant for any V_D above saturation volgate $V_{D,sat} = V_P - V_G - V_{bi}$

$$I_{D,sat} = I_P\left[\frac{1}{3} - \frac{V_G + V_{bi}}{V_P} + \frac{2}{3}\left[\frac{V_G + V_{bi}}{V_P}\right]^{3/2}\right] \tag{18}$$

Another important parameter is transconductance $g_m = \partial I_D/\partial V_G$ with V_D kept constant.

$$g_m = \frac{I_P}{V_P}\left[1 - \sqrt{\frac{V_G + V_{bi}}{V_P}}\right] \tag{19}$$

Parasitics, Equivalent Circuit etc

First we have to find out a.c. currents and voltages based on our knowledge of d.c. currents and voltages. Assuming that a.c. amplitudes are small (satisfied by definition) it seems appropriate to expand the d.c. quantities into Taylor series around a particular bias point. We will treat the currents as unknown variables and voltages as known. From this we conclude that

$$\tilde{i}_D = \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G} \tilde{v}_D + \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D} \tilde{v}_G = g_D \tilde{v}_D + g_m \tilde{v} v_G \tag{20}$$

If the series resistances are included, by inspection we write:

$$\tilde{v}_{DS} = \tilde{v}_{D'S'} - (R_S + R_D)\tilde{i}_D \qquad \tilde{v}_{GS} = \tilde{v}_{G'S'} - R_S \tilde{i}_D \tag{21}$$

which yields output current

$$\tilde{i}_D = \left[\frac{g_m}{1 + R_S g_m + (R_S + R_D) g_D}\right] \tilde{v}_{G'S'} \tag{22}$$

$$+ \left[\frac{g_D}{1 + R_S g_m + (R_S + R_D) g_D}\right] \tilde{v}_{D'S'} \tag{23}$$

Check out the equivalent circuits: idealized and more realistic (low/high frequency)

Unity current gain frequency (f_T) — name describes it all! It is, quite generally, expressed as:

$$f_T = \frac{g_m}{2\pi C_{in}} \tag{24}$$

where C_{in} is total input capacitance as seen by signal at gate contact. In our, simplified, model, we can easily see that $C_{in} = C_{GS} + C_{GD}$. For power amplification we will use f_{max}.

Equivalent Circuit

See fig for simple eq. circuit. In addition to intrinsics, some parasitic elements: R_G= gate resistance, R_S= source resistance. G_π is a measure of leakage current on the gate; it should be neglegible, but if the diode is near turn-on, that may not hold.

Gate-source capacitance $C_{GS} = C_G - C_{DG}$. Total input capacitance on the gate = C_G. Capacitance between gate and drain = C_{DG}. They can be obtained from

$$C_G = \left.\frac{\partial Q_G}{\partial V_{GS}}\right|_{V_{DS}} \text{ and } C_{DG} = -\left.\frac{\partial Q_G}{\partial V_{DS}}\right|_{V_{GS}} \tag{25}$$

C_{DG} contains both intrinsic and parasitic components: intrinsic for the part that is across depletion layer, parasitic that is not.

Intrinsic components: g_{mi}= intrinsic transconductance

$$g_{mi} = \left.\frac{\partial I_{DS}}{\partial V_{GS}}\right|_{V_{DS}} = \left.\frac{\partial I_{DS}}{\partial Q_G}\right|_{V_{DS}} \left.\frac{\partial Q_G}{\partial V_{GS}}\right|_{V_{DS}} = \frac{C_G}{\tau} \tag{26}$$

g_{mi} measures transistor's gain. For max. g_m we need to minimize τ (sounds familiar?). Note: cannot avoid capacitance; it is an intrinsic part of the device.

Further components: output conductance and capacitance.

f_T as usual; for max. frequency of oscillation we have:

$$f_{max} = \frac{f_T}{2\sqrt{G_0(R_G + R_S) + 2\pi f_T C_{DG} R_G}} \tag{27}$$

Basic structure: see fig. Effective gate length definition and channel thickness. Requirement: eff. length should not be significantly different from physical length. Gate design calls for restricted get-finger width. Problem: phase delay along the gate finger. Rule of thumb: Z/2 $< \lambda/16$. 5mm at 1 GHz down to 50μm at 100 GHz.

For microwaves: impedance matching imposes some limits (connecting MESFETs in parallel). Furthermore, temperature must not be allowed to rise above certian level, which reduces the number of cells. Final resort: use circuit combining.

DC characteristics

As for JFETs, pinchoff voltage $V_P = qN_D a^2/(2\epsilon_s)$ = voltage needed to deplete the channel of thickness a.

Threshold voltage: onset of current conduction. $V_T = V_{bi} - V_P$. (all of this for depletion mode devices).

Depletion vs. enhancement mode (normally on and normally off). For enahncement devices, V_T is slightly above 0V (they are depleted in equilibrium).

The max. gate voltage for both cases is limited by forward biasing the gate: cca. 0.7V. Depletion devices have thicker channel regions and/or higher doping. They also provide more currents at highest gate voltage. Breakdown - avalanching between gate and drain.

Saturation region in fig. Saturation current for $V_{GS} = 0$ is called I_{DSS}.

Different model for small devices, based on velocity saturation (fig) . Here the current saturates due to velocity saturation, not as in long channel case. For this to occur we need only 5-10 kV/cm in the channel. Also, dipole formation.

$$I_{DS} = qv_s Z N_D (a - h) = qav_s Z N_D \left(1 - \sqrt{\frac{V_{bi} - V_{GS}}{V_P}}\right) \tag{28}$$

which gives expressions for transconductance:

$$g_{mi} = \frac{v_s \epsilon_s Z}{a} \sqrt{\frac{V_P}{V_{bi} - V_{GS}}} = v_s Z \sqrt{\frac{qN_D \epsilon_s}{2(V_{bi} - V_{GS})}} \tag{29}$$