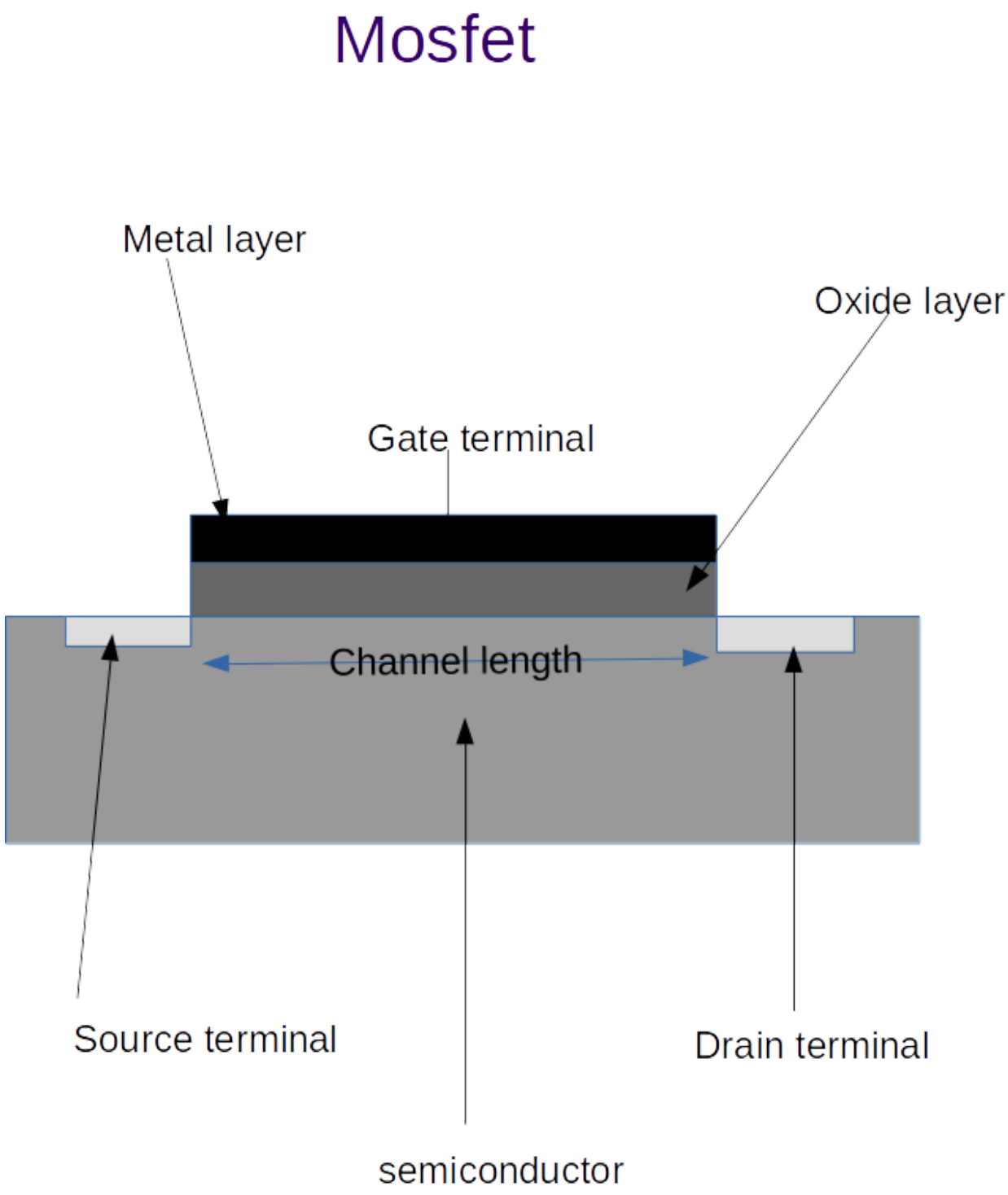
Modelling the fringing effects and other parasitic effects and their impact on High Electron Mobility Transistor

-Pranav Sankhe

In 2016 summers I had an opportunity to work under Prof. Dipankar Saha on the topic of modelling the high frequency transistors considering the effects of fringing effect of field lines on the performance and get their I-V characteristics accordingly. So here I have penned(typed) down a brief expert of the whole process and I got to say, this whole process may seem all tedious and geeky at the first glance but you know it's actually cool. I bet you will really enjoy exploring a particular field a little out of you regular courses and much more in detail.

Watch out for every electronic item around you. You will find no device without a mosfet in it. Your cell phones contain billions and trillions of them and these guys are the ones who actually behind the execution of everything you do on a smartphone.

So let's see what's so great about this baby out there. Fundamentally mosfets (metal oxide semiconductor field effect transistor) are electronic devices used to amplify or switch electrical signals. You can design logic gates by using mosfets and this opens the whole digital world for you.



MOSFET showing gate (G), body (B), source (S) and drain (D) terminals. The gate is separated from the body by an insulating layer (white)

Composition :

- The gate terminal is essentially made up of a highly conducting materials. Earlier they used metals like aluminium for it but nowadays highly doped semiconductor is used.
- The semiconductor used in the substrate usually is silicon. Better semiconductors like gallium arsenide are not preferred since they fail to provide good insulating interfaces.
- The oxide is used for its insulating properties. Generally silicon dioxide is used an oxide.

Working:

When a voltage is applied between gate and body terminal, electrons accumulate at the interface of oxide and substrate. Thus an electron channel is formed which serves the purpose of conducting charge carriers from source to drain terminals. By controlling the voltage we can control the current flowing between the source and drain. By changing the voltage what we change is the conductivity of the channel and thus changing the current. In JFETS enhanced mode and depletion mode are two major types of transistor which are classified based on the fact that whether the transistor is in ON state or OFF state. By ‘On’ state I simply mean that the channel at the semiconductor-oxide interface is conducting some current.

Modes of operation: Enhancement and depletion Mode

In enhancement mode the device is OFF when the gate voltage is zero. The gate voltage needs to be changed to turn the device on. Enhancement mode mosfets are generally used in switching the electrical signals.

In depletion mode the mosfet is On even when the gate voltage is zero. The reason for this is that a thin conducting layer is introduced at the semiconductor-oxide interface. Thus a gate voltage is not needed for the channel to form since it is already there. Depletion mode mosfets are generally used to amplify the signals.

Short channel effects in mosfets

Mosfets are said to have short channels when the channel length is of same order as the width of the depletion layer. The channel length is reduced to increase the operational speed and to reduce the components per chip. But by reducing the channel length we observe some effects called as short-channel effects**.**

The short-channel effects we observe are attributed to two facts:

1. The limitations on electron characteristics when they flow through the channel.
2. The threshold voltage of the mosfet changes.

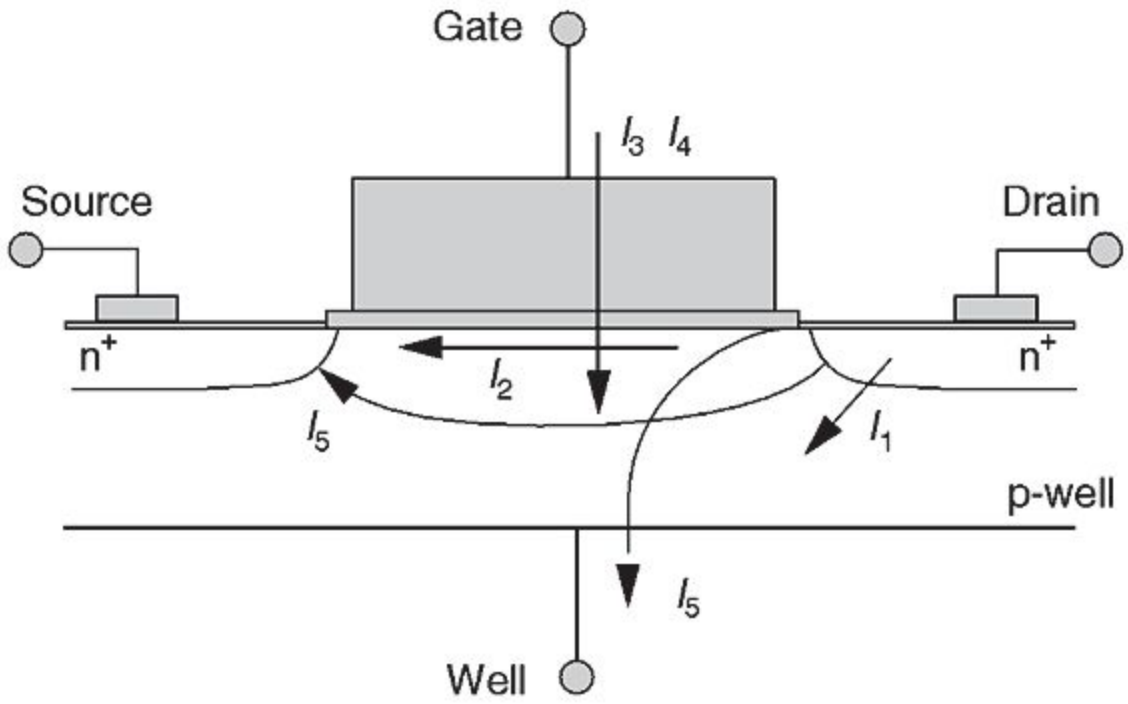
Note: threshold voltage is the bare minimum voltage needed to just start the conduction through the channel.

The short-channel effects are named as follows:

1. drain-induced barrier lowering and punchthrough
2. surface scattering
3. velocity saturation
4. impact ionization
5. hot electrons

Drain-induced barrier lowering and punchthrough:

Let’s understand the punchthrough effect. When the depletion layer surrounding the source terminal extends and merges with the depletion layer of drain terminal the charge carriers get a route to travel through. This is called as punchthrough effect. Punchthrough can be minimized with thinner oxides, larger substrate doping, shallower junctions, and obviously with longer channels.



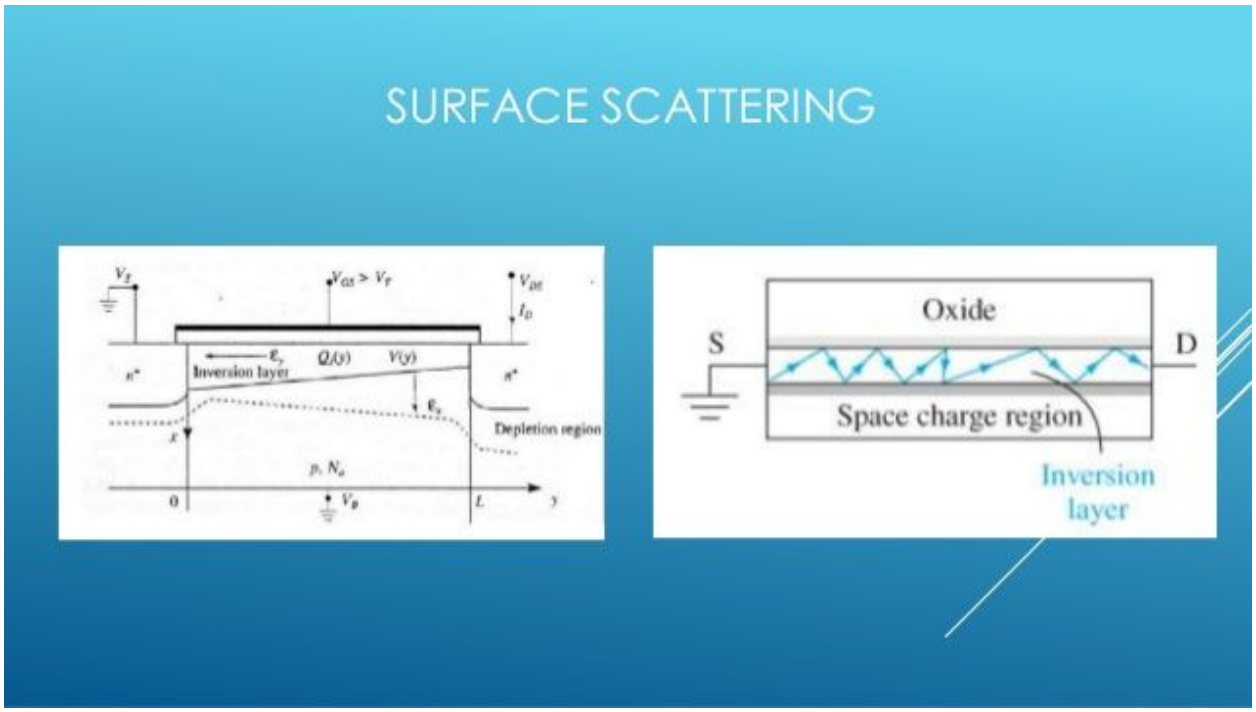
Drain induced barrier lowering: we know that when when the gate voltage is zero there exists a potential barrier which prevents the electrons to flow. When the gate voltage is increased the barrier potential lowers enabling the current to flow. Since the in normal mosfets the channel length is sufficiently large the source-drain voltage doesn't affect the potential barrier. But in short-channel mosfets the drain barrier potential gets affected by both gate voltage and source-drain voltage.

If you increase the source-drain voltage the potential barrier decreases allowing the electrons to flow. This is called drain induced barrier lowering . The channel current that flows under this conditions (Vgs<Vt) is called the sub-threshold current.

Surface scattering :

Since the the depletion layer extends into the channel region of the mosfet the electric field component (directed from drain to source) increases. The surface mobility therefore becomes field-dependent. Since the carrier transport in a MOSFET is confined within the narrow inversion layer, and the surface scattering (that is the collisions suffered by the electrons that are

accelerated toward the interface by Ex) causes reduction of the mobility, the electrons move with great difficulty parallel to the interface.



Velocity Saturation:

At low Ey, the electron drift velocity vde in the channel varies linearly with the electric field intensity. However, as ey increases above 104 V/cm, the drift velocity tends to increase more slowly, and approaches a saturation value of Vde(sat)=107 cm/s around ey =105 V/cm at 300 K.

Note : the current here is limited due to velocity saturation not due to pinchoff.

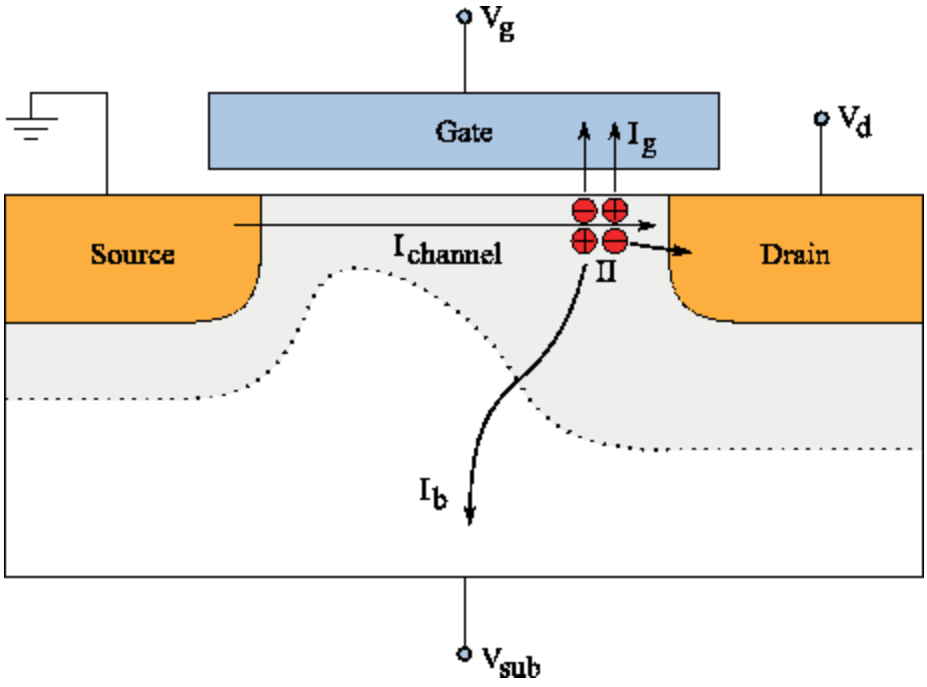
Note: pinchoff : To form the channel the potential difference between gate and oxide-semiconductor interface must exceed the bare minimum needed voltage i.e threshold voltage.

When the drain voltage Vd > Vg + Vt, the current becomes saturated. The reason is that when you increase the drain voltage more than a certain value which is defined by the above equation, the channel thickness at the drain terminal becomes zero. Thus the current through the channel becomes saturated.

Electrically, the effect of pinch off is that the channel no longer acts like a simple resistor. The current becomes fixed (saturated) at the value just prior to pinch off.

Impact Ionization:

Another undesirable short channel effect is caused due to the high velocity of electrons in the presence of high strength longitudinal electric fields that can generate electron-hole pairs due to impact ionization(by impacting electrons on silicon atoms and ionizing them). In general Impact ionization is the process in a material by which one energetic charge carrier can lose energy by the creation of other charge carriers.



It happens as follows: electrons are normally attracted towards the drain while the holes enter the substrate material. The region between the source and drain can act as base of npn transistor, drain acts as a collector and source as an emitter. If the above mentioned holes formed due to impact ionization are collected by the source the corresponding hole current can create a voltage drop of about 0.6 volts and the normally reverse biased pn junction.

When electrons are injected into the base from the emitter(source) they travel through the channel towards the drain and in the process gain more energy. This creates more electron-hole pairs making the situation worse.

Hot electrons:

Another problem related to the high electric fields is that the high energy electrons can enter in the oxide layer where they get trapped thus charging the oxide layers which increases the threshold voltage and simultaneously adversely affecting the gate control over the drain current.

To get the expression of the threshold voltage of the short channel mosfet click here.

The analysis of how mosfets work is largely based on the analysis of the mos capacitance. So let's have a look at it.

After all the analysis and stuff we get the mosfet current and voltage relation.

$$i_D = 0$$
$$v_{GS} \le V_t, \qquad \textit{cut-off} \qquad (1a)$$

$$i_D = \frac{k}{2}\left[2(v_{GS} - V_t)v_{DS} - v_{DS}^2\right]$$
$$v_{GS} > V_t, v_{DS} \le (v_{GS} - V_t), \quad \textit{ohmic} \qquad (1b)$$

$$i_D = \frac{k}{2}\left[2(v_{GS} - V_t)^2\right]$$
$$v_{GS} > V_t, v_{DS} > (v_{GS} - V_t), \quad \textit{saturation} \quad (1c)$$

Where, Vgs = voltage applied at the gate terminal Ids = Current flowing from source and drain terminal Vds = voltage difference between source and drain terminal Vt = threshold voltage K depends on mos capacitance, permittivities and dimensions of mosfet.

Mos Capacitance

Basically, mosfet works on the charge build up by the mos capacitance. Whenever we talk about the mos capacitance we are referring to the capacitance between the body(semiconductor) and the gate electrode(metal).

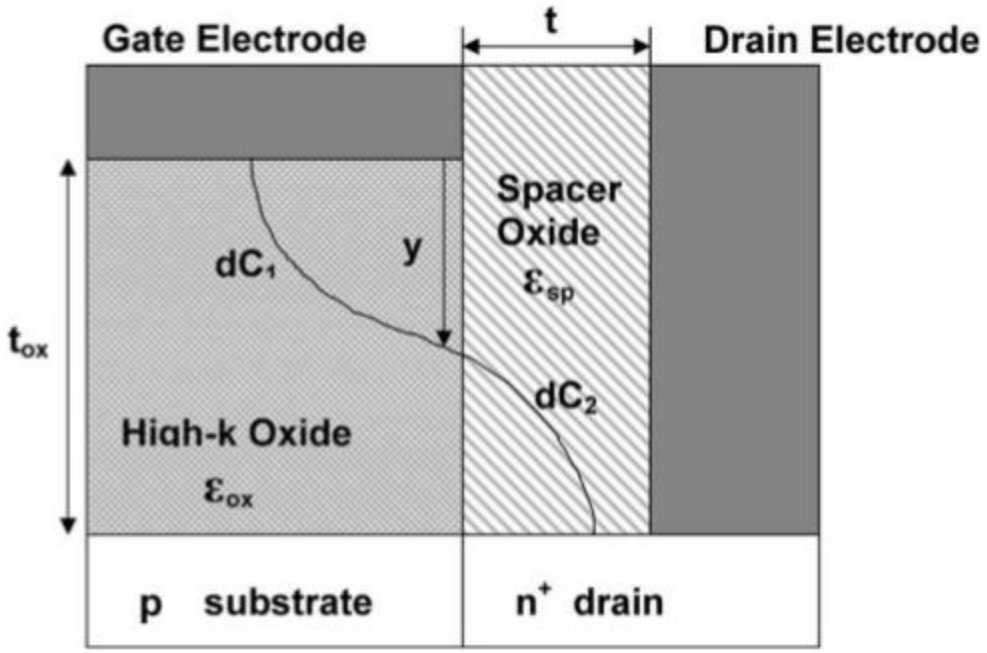
The conventional formulae of current and voltage relation considers just the parallel capacitance. When we go to nano scale the fringing of field lines which is ignored at normal dimensions, come to picture and significantly affect the performance of these devices. So the non-trivial part is to model this fringing effect.

A more deeper insight : When I need to see how the electron channel in the mosfet is formed, I need to analyse how the associated regions of mosfet interact with each other. This interaction leads to the build up of some charges. We are looking for this interaction which turns out to be the capacitance. Therefore to study the overall capacitance we need to account every single electric field line. And that's what we are gonna do.

Capacitance calculation in normal mosfets

The infinitesimal capacitance in the high k and spacer regions can be written as follows:

Consider a fringing field line intersecting the interface between the spacer oxide and high k oxide at a distance y from gate electrode.



High k region(C1) Spacer region(C2)

$$dC_1 = \frac{\varepsilon_{ox} W\, dy}{\left(\frac{\pi}{2}\right) y} \quad \text{and} \quad dC_2 = \frac{\varepsilon_{sp} W\, dy}{\left(\frac{\pi}{2}\right) (t_{ox} - y)}$$

……..(1)

As you can see from the diagram the capacitances dC1 and dC2 are in series the net infinitesimal capacitance will be a series equivalent of the above mentioned capacitances.

Finally after obtaining the infinitesimal capacitance, to obtain the net capacitance we need to integrate it by setting the limits of ‘t’ from ‘0’ to ‘tox’.

The expression we get is :

$$C_{bottom} = \int_0^{t_{ox}} \frac{2\varepsilon_{ox}\varepsilon_{sp} W\, dy}{\pi(\varepsilon_{ox} t_{ox} + \varepsilon_{sp} y - \varepsilon_{ox} y)}$$

$$= \frac{2\varepsilon_{ox}\varepsilon_{sp} W}{\pi(\varepsilon_{ox} - \varepsilon_{sp})} \ln\left(\frac{\varepsilon_{ox}}{\varepsilon_{sp}}\right).$$

…….(2)

Now during the derivation of this expression we assumed the fringe lines to be circular. We will eliminate that now.

The model given by Kamchouchi and Zaky of parasitic capacitance gives the parasitic capacitance per unit length considering the electric field lines fringing from the entire perimeter of the bottom edge of the gate electrode.

$$C = \frac{(2 - \ln 4)\varepsilon_{ox}}{2\pi}$$

……..(3)

The total fringe capacitance can be obtained by multiplying Cbottom by the perimeter of the bottom edge of the gate electrode.

But since here we are accounting only the electric field lines fringing from the bottom edge of the gate electrode to either the source or the drain region only, the internal fringe capacitance can be written as:

$$C_{\text{bottom}} = \frac{(2-\ln 4)\varepsilon_{\text{ox}} W}{2\pi} \cong \frac{(0.3)\varepsilon_{\text{ox}} W}{\pi}$$

$$\text{for} \quad \varepsilon_{\text{ox}} = \varepsilon_{\text{sp}} \quad \text{and} \quad \frac{t_{\text{ox}}}{L_g} \ll 1.$$

……..(4)

If you carefully observe you can see that eq(2) reduces to eq(4) if you substitute $\varepsilon_{\text{ox}} = \varepsilon_{\text{sp}}$ but for a constant term. This constant term accounts for the fact that the fringe lines are not circular. Therefore we will multiply eq(2) with 0.15. The final expression we get :

$$C_{\text{bottom}} = \frac{(0.3)\varepsilon_{\text{ox}}\varepsilon_{\text{sp}} W}{\pi(\varepsilon_{\text{ox}} - \varepsilon_{\text{sp}})} \ln\left(\frac{\varepsilon_{\text{ox}}}{\varepsilon_{\text{sp}}}\right).$$

…….(5)

The Kamochouchi and Zaky model assumes that the separation between the electrodes is very small in comparison to the length of the elctrodes which is obviously not the case in short channel mosfets.

Now as tox/Lg increases ,there is more crowding of the field lines in the spacer region. The fringing field from the gate to source/drain regions increases as the function of tox/Lg.

To account all this we modify the dielectric of spacer dielectric constant.

$$\varepsilon'_{\text{sp}} = \left(1 + \frac{t_{\text{ox}}}{L_g}\right)\varepsilon_{\text{sp}}.$$

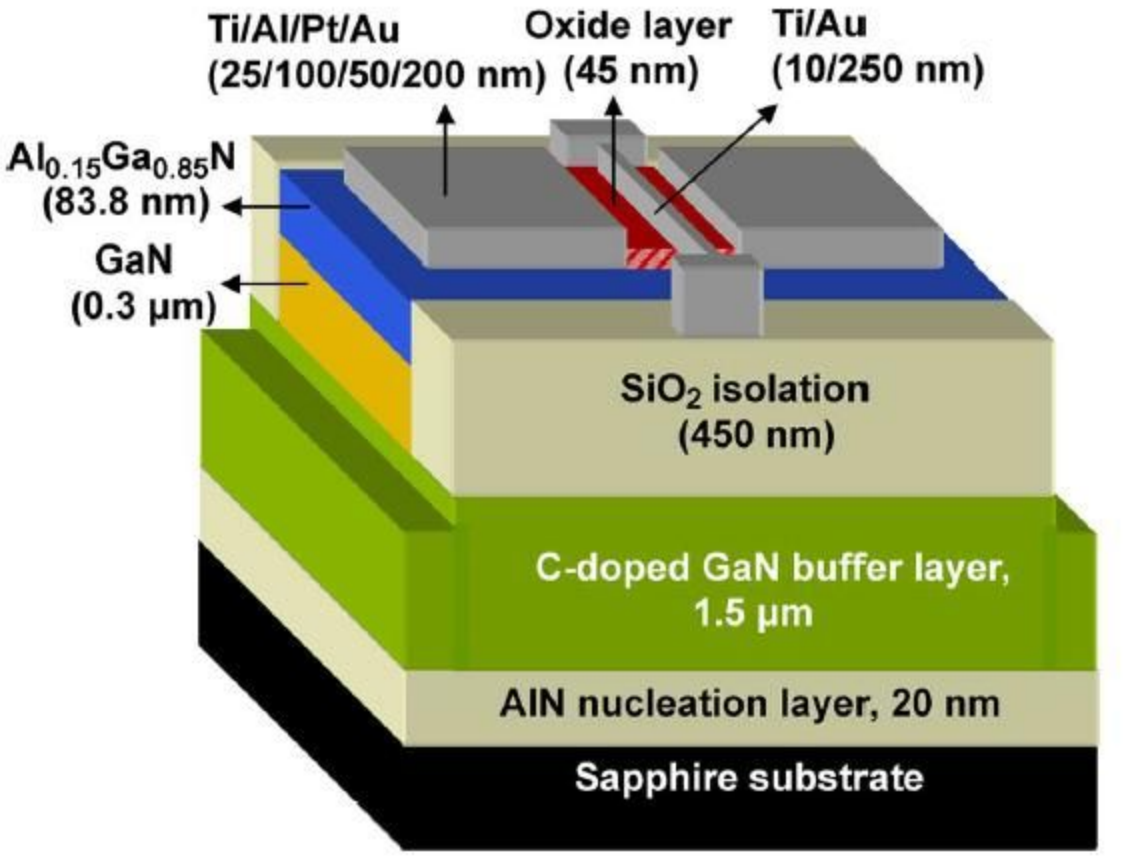
Substitute this in equation(5) . So finally we have our expression :

$$C_{\text{bottom}} = \frac{(0.3)\varepsilon_{\text{eff}} W}{\pi}$$

Where

$$\varepsilon_{\text{eff}} = \frac{\varepsilon_{\text{ox}}\varepsilon'_{\text{sp}}}{\varepsilon_{\text{ox}} - \varepsilon'_{\text{sp}}} \ln\left(\frac{\varepsilon_{\text{ox}}}{\varepsilon'_{\text{sp}}}\right).$$

My job was to analyse a different type of mosfets called as HEMTs (High Electron Mobility Transistors)



Schottky Contact : This is a contact between a metal and a lightly doped semiconductor.

[Note : Ohmic contact is a contact between a metal and a doped semiconductor.]

In the above diagram of hemt, source and drains are in ohmic contacts and gate has a schottky contact.

These heterojunction have high carrier mobility.

Such high speeds are due to the following reasons:

One layer is heavily doped with donor atoms thus it has excess of electrons in its conduction band. This electrons will diffuse to the adjacent non-doped layer due to availibility of lower energy states there. This creates an electric field. This diffusion process continues until electron diffusion and drift balance each other, creating a junction. The undoped region now has excess

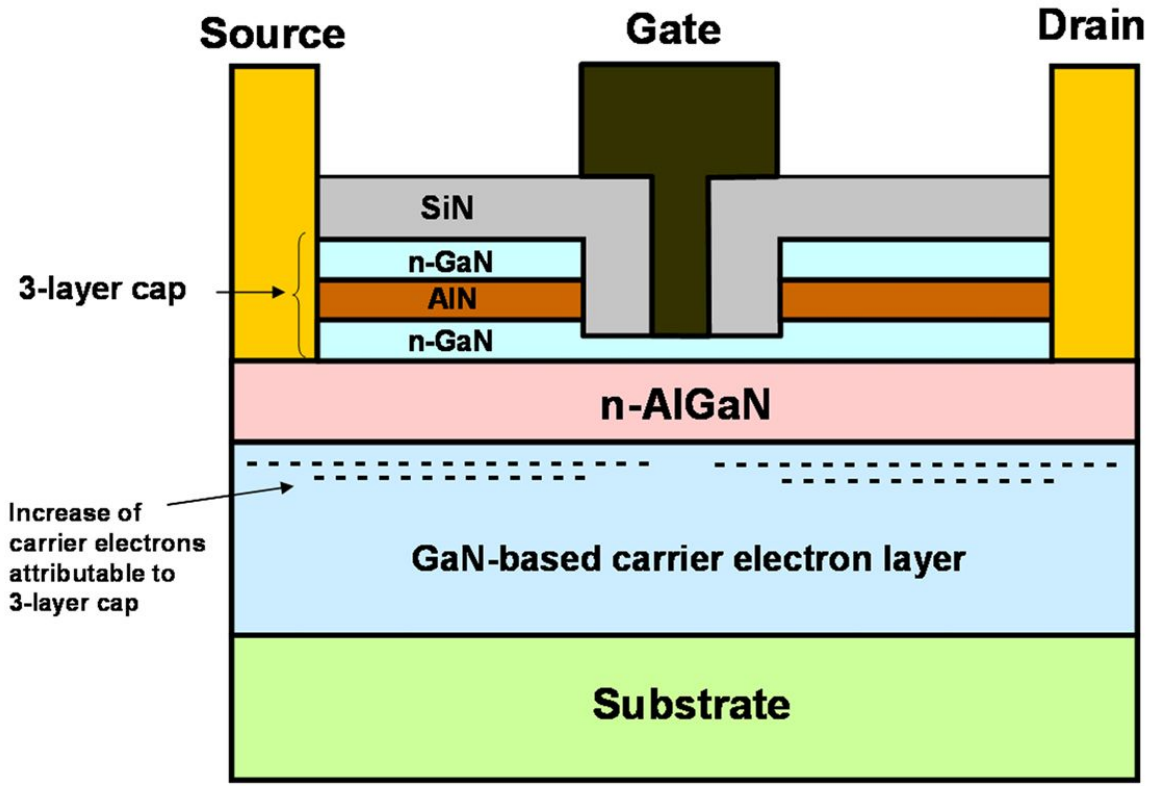
majority charge carriers and also since the undoped layer didn't had any donor atoms to prove as an obstacle for electron movement.

The band discontinuity across the conduction and valence band among the different layers can be modified separately.

The diffusion process leads to the accumulation of the electron along the boundary of these layers(doped and non doped). The accumulation of electrons leads to a very high current in the heterojunction. The accumulated electrons are also known as 2 DEG or 2 D electron gas.

Note: 2DEG is an electron gas that can is free to move in 2 dimensions but tightly confined in the 3rd dimension. this confinement leads to quantized energy levels for motion in 3rd dimension. Thus it appears to be 2-D sheet embedded 3D.

So to summarize we can say that hemt is a field effect transistor which incorporates a junction between 2 materials with different band gaps as the channel.



It is found out that the gate extrinsic capacitance contributes significantly to the parasitic delay -approx 50% of the delay in these highly scaled devices.

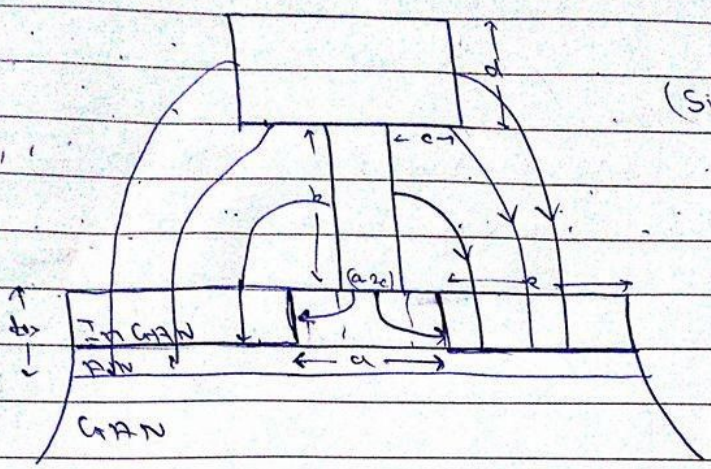
Now this gate capacitance comprises of 2 components:

1. Parallel plate capacitance between the Tgate and the surrounding electrodes.

2. The fringing capacitance between the gate stem and the access regions. </br> When the gate length is reduced below 100nm the parasitic RC charging delay caused by source drain resistance and parasitic capacitance can account for a significant fraction in the delay of the device. </br>

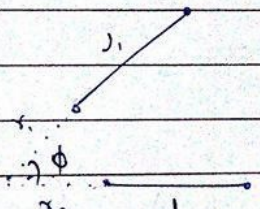
Now we will proceed to the actual calculations involved in obtaining the net capacitance of hemt. Sorry for not using mathjax :p

We would need a analytical formula for capacitance between 2 plates of specified dimensions and a particular angle between them.

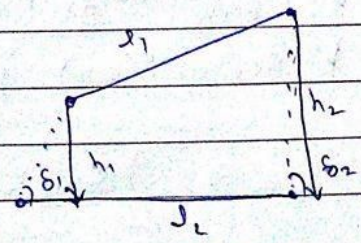


We will use the following result for calculating capacitance between inclined plates obtained after applying conformal transforms.

$$C = \frac{\epsilon_0 k'(k_{in})}{k(k_{in})}$$



$$k_{in} = \sqrt{\frac{\left(r_1^{\pi/\phi} + r_2^{\pi/\phi}\right)\left((r_1+d_1)^{\pi/\phi} + (r_2+d_2)^{\pi/\phi}\right)}{\left(r_1^{\pi/\phi} + (r_2+d_2)^{\pi/\phi}\right)\left(r_2^{\pi/\phi} + (r_1+d_1)^{\pi/\phi}\right)}}$$



for $\phi \to 0$

$$C = \frac{\epsilon_0 k'(\operatorname{sech}(\pi\phi/2d_1))}{k(\operatorname{sech}(\pi\phi/2d_2))}$$

$$k_{in} = \frac{1+\sin(\delta_1-\phi)}{1+\sin\delta_1}$$

$$k_{in} = \left[\frac{\left(1+\left(\frac{\sin(\delta_1-\phi)}{\sin\delta_1}\right)^{\pi/\phi}\right)\left(1+\left(\frac{\sin(\delta_2-\phi)}{\sin\delta_2}\right)^{\pi/\phi}\right)}{\left(1+\left(\frac{h_2+d_1\sin\phi)/\sin(\delta_2-\phi)}{h_1\sin\delta_2}\right)^{\pi/\phi}\right)\left(1+\left(\frac{h_1\sin(\delta_1-\phi)}{(h_1+d_1\sin\phi)\sin\delta_1}\right)^{\pi/\phi}\right)}\right]^{1/2}$$

$$\rightarrow (C_{de})_1 = \varepsilon_{SiCON}\left[\frac{k'(k_{1m})}{k(k_{1m})} + \frac{k'(k_{2m})}{k(k_{2m})}\right] w$$

$$= \varepsilon_{SiCON}\frac{k'(k_{1m})\,w}{k(k_{1m})}$$

$$k_{1m}\left((d+b),\ b,\ \tan^{-1}(b/c),\ \pi-\tan^{-1}\left(\frac{b+d}{e+c}\right),\ \pi/2,\ d\right)$$

where $k_{1m} = k_{1m}(h_1, h_2, \delta_1, \delta_2, \theta, d_1)$

$$\therefore C_{de} = (C_{de})_2 = \varepsilon_{SiCON}\frac{k'\left(\operatorname{sech}\left(\frac{\pi b}{2d}\right)\right) w}{k\left(\operatorname{sech}\left(\frac{\pi b}{2h}\right)\right)}$$

$$C_{de} = \frac{(C_{de})_1 \times (C_{de})_2}{(C_{de})_1 + (C_{de})_2}$$ [capacitances in series]

$$\rightarrow (C_{ce})_1 = \varepsilon_{SiCON}\frac{k'(k_{1m})\,w}{k(k_{1m})}$$

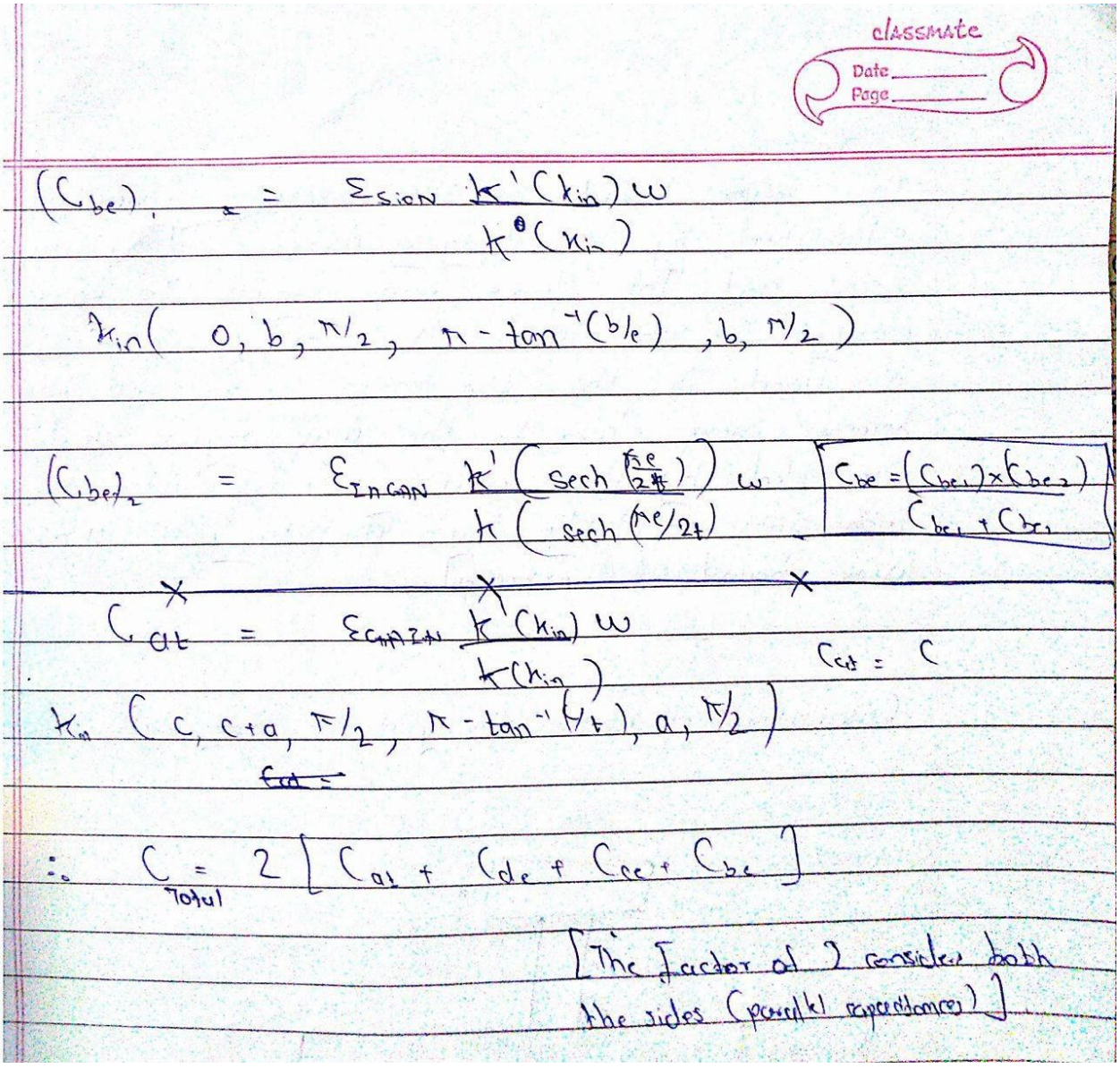
for $\alpha > (\theta+\pi)$

$$k_{1m}\left(b,\ b,\ \pi/2,\ \tan^{-1}\left(\frac{b}{e-c}\right),\ 0,\ c\right) \quad e \rightarrow e+a+x$$

$$(C_{ce})_2 = \varepsilon_{SiCON}\frac{k'\left(\operatorname{sech}\left(\pi e/2t\right)\right) w}{k\left(\operatorname{sech}\left(\pi e/2b\right)\right)}$$

$$C_{ce} = \frac{(C_{ce})_1 \times (C_{ce})_2}{(C_{ce})_1 + (C_{ce})_2}$$

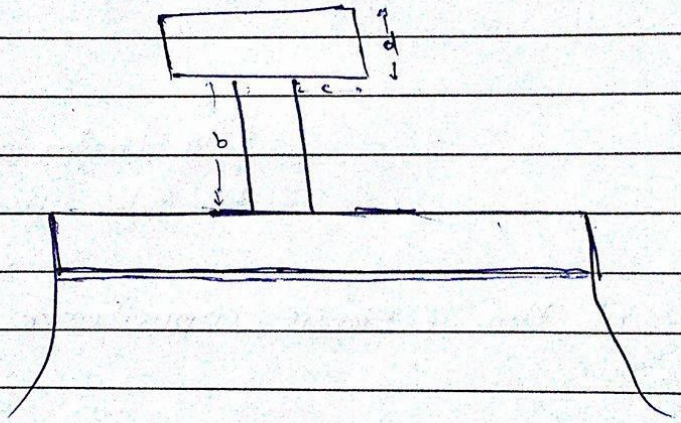
**



So now we have the expression of the net capacitance of HEMTS.

Having the capacitance expression we proceed to find the charge density of the channel and then get the current expression.

Calculation of electron density of the ~~charge~~ electron channel



$$V(x) = \frac{V_D x}{L} \quad \text{[assumed linear]}$$

$$Q(x) = C(x)\,[V_{GS} - V(x)]$$

$$I = \frac{Q \cdot dV(x)}{dx} = \frac{Q V_{DS}}{L}$$

$$\therefore I = C(x)\,[V_{GS} - V(x)] \cdot \frac{V_{DS}}{L}$$

**

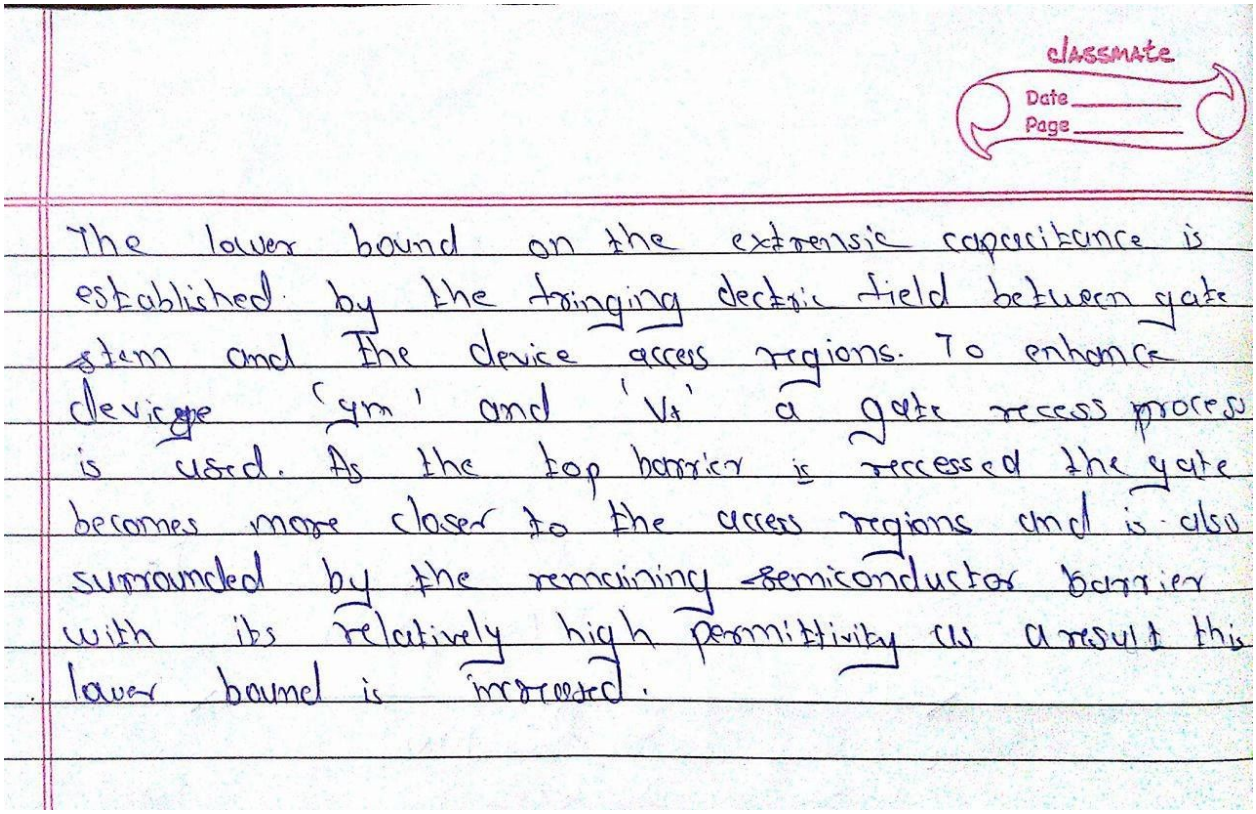
$$\therefore I = 2\left[C_{ad} + C_{de} + C_{ce} + C_{bc}\right]\left[V_{GS} - V(x)\right]\cdot\frac{V_{DS}}{L}$$

$$\therefore I = \frac{2V_{DS}}{L}\left[V_{GS} - V(x)\right]\left[C_{ad} + C_{de} + C_{ce} + C_{bc}\right]$$

Some of methods proposed to reduce the parasitic capacitance and its contribution in the delay of HEMTS.

Optimization of the extrensic capacitance.

The parasitic capacitance accounts for a substantial part in the delay of high electron mobility transistors. Reduction of extrensic capacitance can be ~~selected~~ achieved by raising the T-gate stem height, dielectric etchback, ultrathin passivation schemes and even elimination of T-gate cap.



Credits to some papers refered :

- The electrostatic capacitance of an inclined plate capacitor by Yumin Xiang
- Effect of Fringing Capacitances on the RF Performance of GaN HEMTs With T-Gates