LECTURE 21

MOSFET

Overview

- Solar cell fundamentals
- Novel solar cell structures
- Thin film solar cells
- Next generation solar cell

Solid State Electronics

Topics we'll be covering

- JFETs, MESFETs, MOSFETs and IGFETs.
- Qualitative explanation of the operation of the MOSFET
- Fabrication of MOSFET, depletion or enhancement
- MOSFET equations and models.
- Small signal model and parasitic correspondence.
- Imperfections such as short channel effects, hot electron effect , latch-up
- Sub threshold operation, punch-through, pinch-off

History

Historically, field-effect transistors (FETs) were proposed prior to bi-polar transistors, in 1925 by Lilienfeld. In particular MOSFETs were the first type of transistor to be ever suggested.

What actually happened is that the manufacturing technology was not good enough at the time to make a FET and Bardeen and Brattain (Bell Labs) made the first BJT using Germanium when trying to make a FET by trying almost everything they could think off.

Shockley (Bell Labs) understood what happened and suggested an improved design. Thus entered the age of the BJT (1950). It took to 1960 to make the first MOSFET (Kahng and Atalla, Bell Labs). Since then Silicon MOSFETs have gradually replaced BJTs in most applications.

History

For interest, Schockley left Bell Labs to set up his own company, Schockley Semiconductors in California, near San Francisco. His company had the best engineers going but he had a pet project, 4 layer diodes which dominated the direction of the company.

Eight of his best engineers left because of Schockley's personality and set up Fairchild Semiconductor, to concentrate on transistors. Later two more left to set up Intel. Intel was the first company to make MOS RAM, a 1K chip in 1970's. Fairchild is currently worth USD 2 Billion and Intel is worth USD 168 Billion. This was the start of Silicon Valley.

Schockley's company went bankrupt, he ended up lecturing where after some racial comments on intelligence and genetics, he lost that too. Basically he ended up penniless and disgraced.

JFETs, MESFETs, MOSFETs and IGFETs

- **JFET**
  - Junction Field Effect Transistor
- **MESFET**
  - Metal-Semiconductor Field Effect Transistor
- **MOSFET**
  - Metal-Oxide-Semiconductor Field Effect Transistor
- **IGFET**
  - Insulated Gate Field Effect Transistor
- **MISFET**
  - Metal-Insulator-Semiconductor Field Effect Transistor

FETs and BJTs

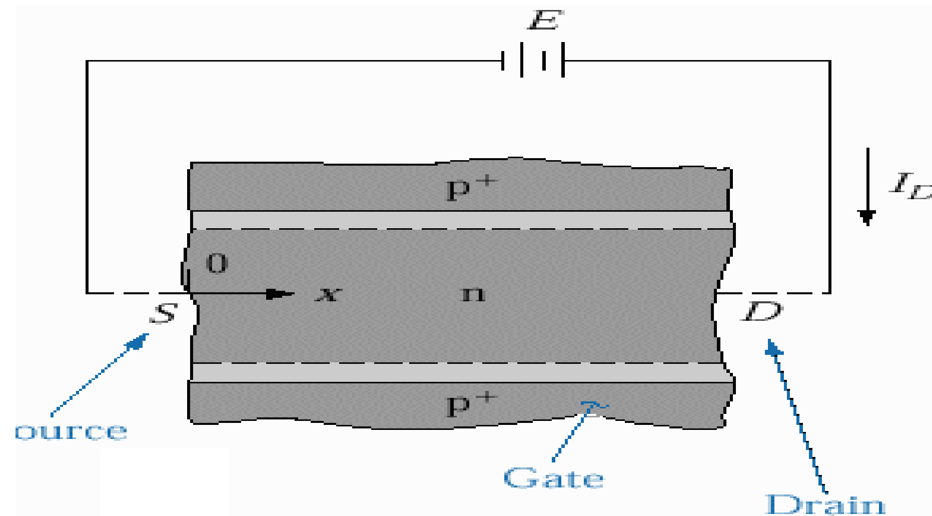
In PN junctions we had minority and majority current carriers on both P side and N side of the junction.

In FETs only the majority carriers are used. For this reason FETs are often called *Unipolar Transitors.* The current in a FET does not cross any PN junction.

In BJTs, there are two PN junctions and current is carried by both minority and majority carriers and hence are called *Bipolar Transistors*, hence the B in BJT.

JFET

In very simple terms the JFET can be considered as a sandwich of 3 layers, the top and bottom being one type with the centre the other type.



The top and bottom layers are connected together and are called the GATE of the transistor.

The current flows through the centre layer

JFET

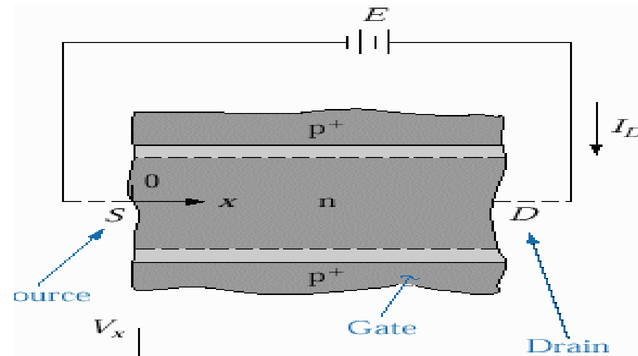
As can be seen from the diagram, there is actually no difference between the left and right hand side of a FET in theory. Therefore we can call either side DRAIN or SOURCE.

It's more common to say that current flows from DRAIN to SOURCE.

The names come from electron flow directions, the SOURCE to DRAIN is the direction of electron flow

In practice there are some minor manufacturing issues.

JFET



The JFET operation is dependent on the variation of the depletion region.

The depletion region is defined as being a region where all the carriers have been depleted due to recombination at the junction.

Thus normally current would flow through a depletion region unless it was diffusing across the

JFET

However the depletion region width is dependent on the bias voltage. If we make the depletion region bigger then the conducting centre of the sandwich gets smaller, looks more resistive, carries less current.

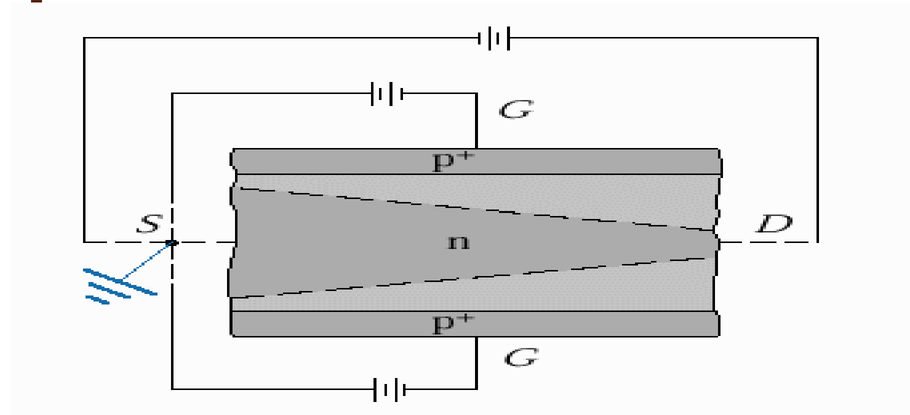
If we make it smaller, the conductive region increases, carries more current.

JFET

So to make a JFET work,

- Set the gate source voltage to reverse bias the junction. This prevents current flowing into gate, a reversed bias diode.
- Changing this reverse bias voltage changes the depletion width.
- Changing the depletion width changes the resistance of the conducting channel
- Increase it too much and you can close the channel off.

JFET

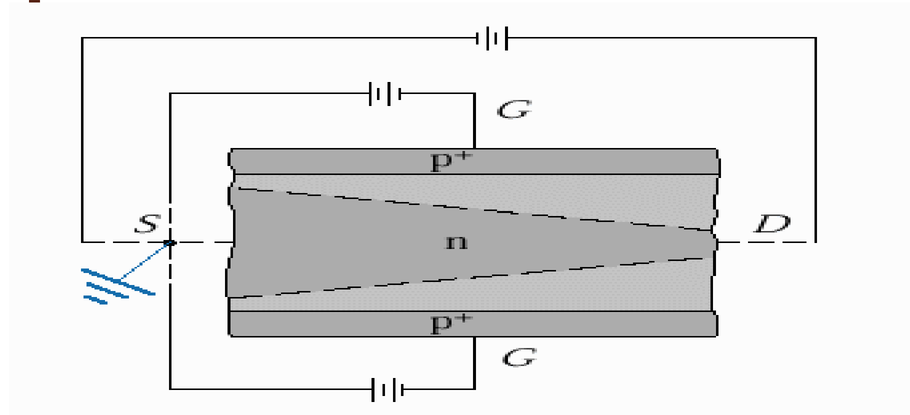


The conducting channel of the JFET is normal semiconductor bulk material.

However conducting semiconductors are very resistive. This resistance we can calculate, from earlier, from mobilities and doping.

However if current flows there'll be a voltage drop across transistor. This has an effect on the depletion region width

JFET



The drain is the current input to the device, therefore at the drain there is no voltage drop compared to drain voltage.

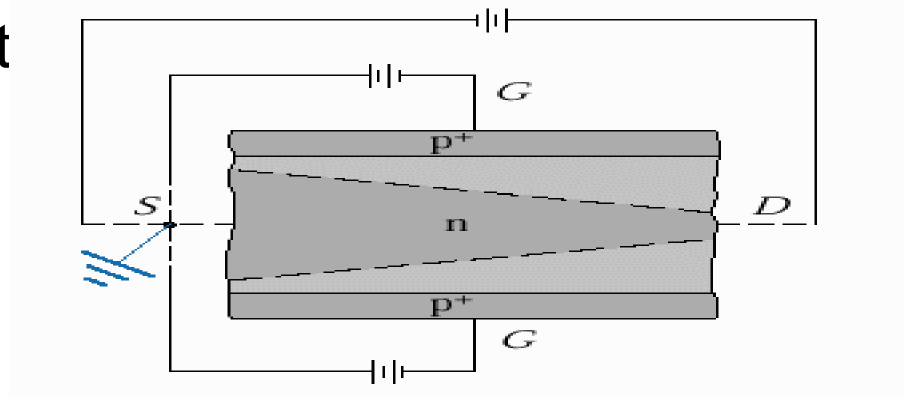
At the source, the current has passed through the channel and the source voltage is now the drain voltage less the current * channel resistance

$$V_S = V_D - I_{fet}R_{fet}$$

JFET

Thus there is a gradient in the voltage across the transistor. This means that between the channel (say N type) and the gate (P type) there is a varying voltage.

This means that the depletion layer will vary across the junction, thicker at the drain than the source. Remember the drain needs to be at a positive voltage to t



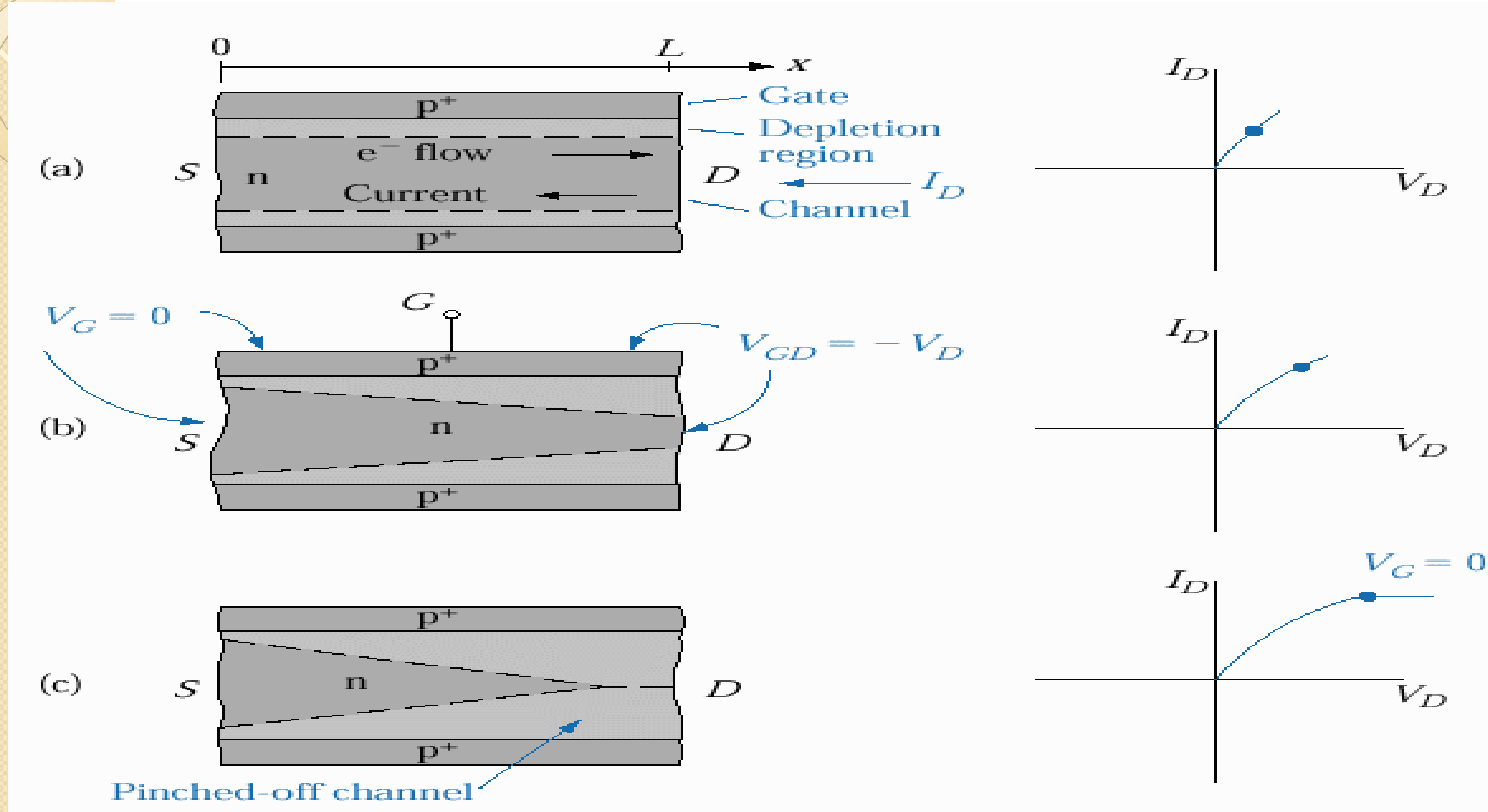
JFET - Pinch-Off

Now it's quite easy to see that if there is sufficient reverse voltage across the junction then the depletion layer will close of the channel.

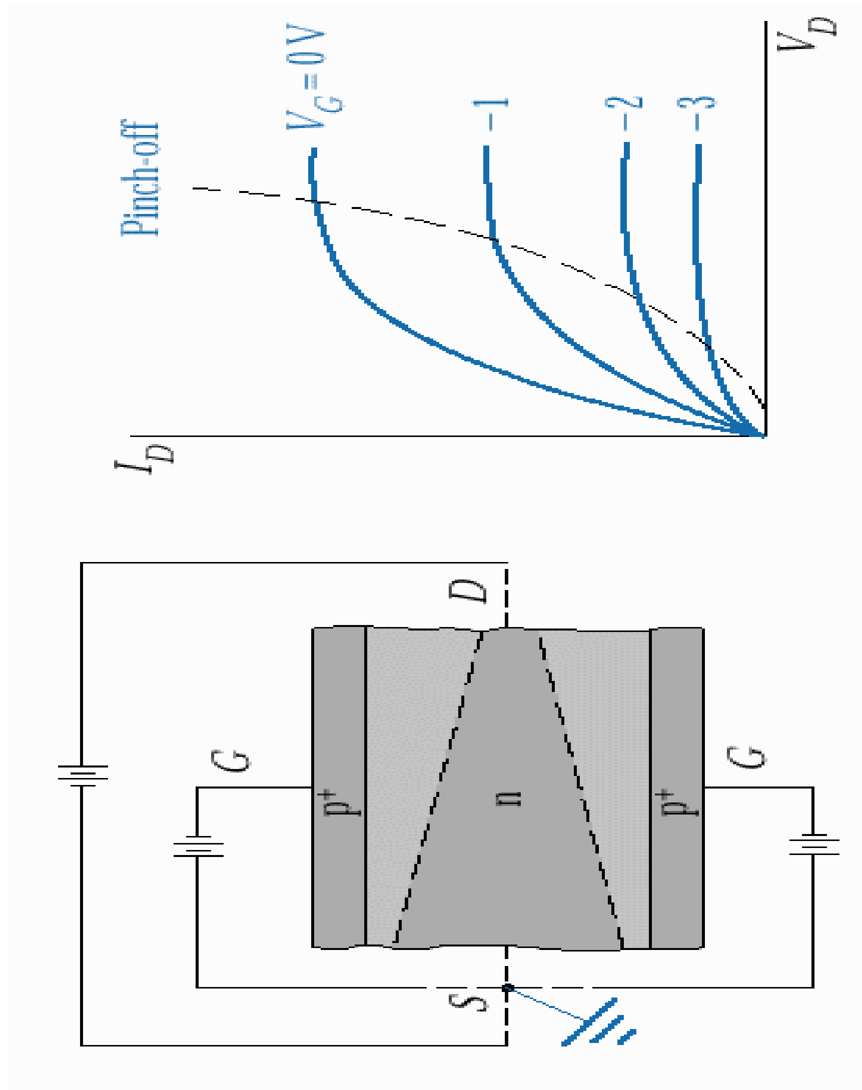
This is called pinch-off. For a given voltage, if you drive sufficient current, ie increase the drain voltage sufficiently, you can force the channel close.

At this point you can't drive any more current through the system, it levels off. Any more current would cut the channel, current begin to drop but then current would flow again… equilibrium position is just at pinch-off.

JFET - Pinch-Off



JFET - Pinch-Off



JFET - Pinch-Off

Now we know from before that the width of the depletion region is

$$W = \sqrt{\frac{2\varepsilon_i \Phi}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)}$$

We also said that if we heavily dope one side of the junction compared to the other then effectively all the depletion region is the side of least doping.

For good conductivity in the gate, to make it metal-like, we heavily dope the gate and so we can assume all the depletion region is in the centre channel.

JFET - Pinch-Off

Now we know from before that the width of the depletion region is

$$W = \sqrt{\frac{2\varepsilon_i(\Phi - V_{drain})}{q}\left(\frac{1}{N_D}\right)}$$

N type channel

We also said that if we heavily dope one side of the junction compared to the other then effectively all the depletion region is the side of least doping.

For good conductivity in the gate, to make it metal-like, we heavily dope the gate and so we can assume all the depletion region is in the centre channel.

JFET - Pinch-Off

Now pinch-off occurs when the depletion width is half that of the channel width. There are two junctions, on each side of the channel.

So defining the channel width to be "**T**", so

$$\frac{T}{2} = \sqrt{\frac{2\varepsilon_i(\Phi + V_{GD})}{q}\left(\frac{1}{N_D}\right)}$$

$$\frac{T^2 N_D}{4} = \frac{2\varepsilon_i(V_{Pinch})}{q}$$

$$V_{Pinch} = \frac{qT^2 N_D}{8\varepsilon_i}$$

JFET - Channel Current

We won’t work out the current equation. It is messy mathematically but not difficult. The steps are as follows

- The current at all points in the channel must be equal.
- In any part of the channel the resistance of that part can be calculated by the width of the available channel and the conductivity. The resistance goes up as the channel narrows.
- So you can set the current to be a function of the rate of change of depletion width which is dependent on the current and resistance.

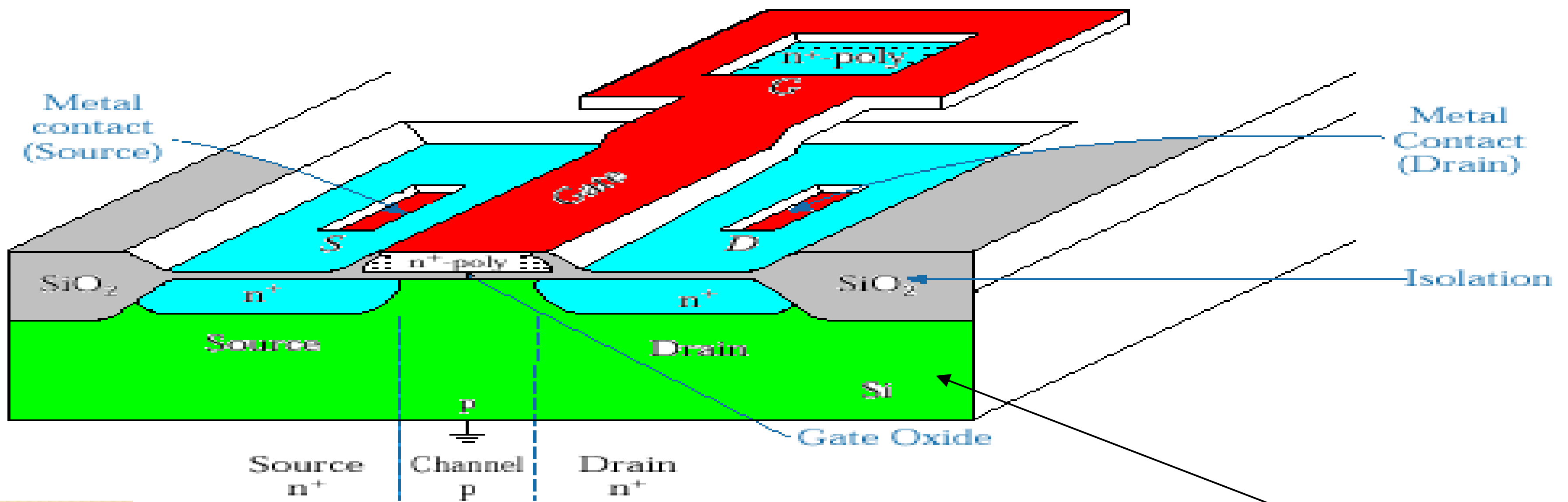
MOSFET /IGFET

This is probably the most widely used transistor. It is formed by a three layer sandwich.

- The top layer is either metal or very heavily doped semiconductor.
- The middle layer is an oxide, SiO_2. This is a very good insulator.
- The third layer is the bulk of the transistor and it is this bulk material that states whether it is N or P type.

On either side of the bulk of the MOSFET there are two contact regions, opposite type to the bulk and these form the DRAIN and SOURCE.

MOSFET /IGFET



MOSFET

The MOSFET device operates by generating (or inhibiting) a conductive layer between the source and drain.

In an enhancement MOSFET, the most common type, a channel is formed by the collection of suitable charge carriers just below the gate. This comes from the gate and the bulk material appearing similar to a capacitor.

In depletion MOSFETs, a channel naturally exists between the the source and drain but with sufficient reverse bias voltage, a sufficient collection of opposite charge carriers can form and this disrupts

MOSFET

For the rest of this discussion, only enhancement MOSFETs will be considered.

As there is an insulator between the gate and the channel, a later of SiO_2, there is **NO** conductive path between the gate and either the drain or source.

Any path that does exist will develop from parasitic capacitors and are only an issue for high frequency operation.

There is no DC path though these capacitors can be charged up and this is a difficulty with discrete components.

MOSFET

The main function of the MOSFET is effectively that of a capacitor.

In a JFET, the reverse bias voltage was such that a depletion region was created. No free carriers exist close to the junction because there is both diffusion and drift current across the junction.

In the MOSFET with the insulating layer there is no current and hence no depletion region. However the charge on the gate attracts carriers of the opposite sign to the region just under the channel. The greatest concentration of these lie closest to the gate.

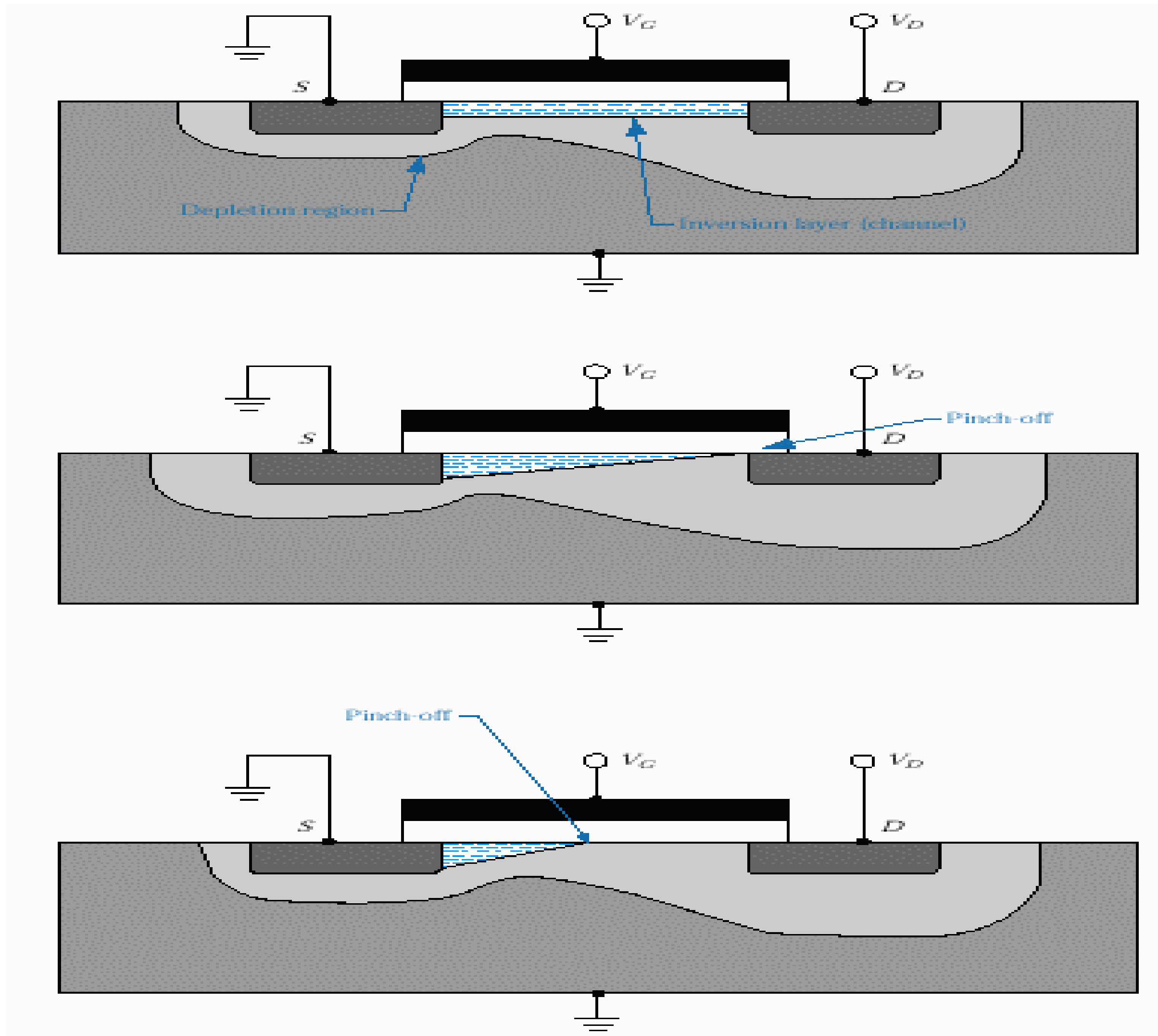
MOSFET

If sufficient number of free carriers are available it is possible to overcome the lightly doped bulk material and change the material type, ie P type to N type.

If this occurs and forms a suitable channel between drain and source then current will flow.

The width of the channel is dependent on the gate bulk voltage. However as in JFET current flow in bulk semiconductor is resistive and the voltage will drop across it, changing the width of the channel. Hence pinch-off can occur in MOSFETs as well.

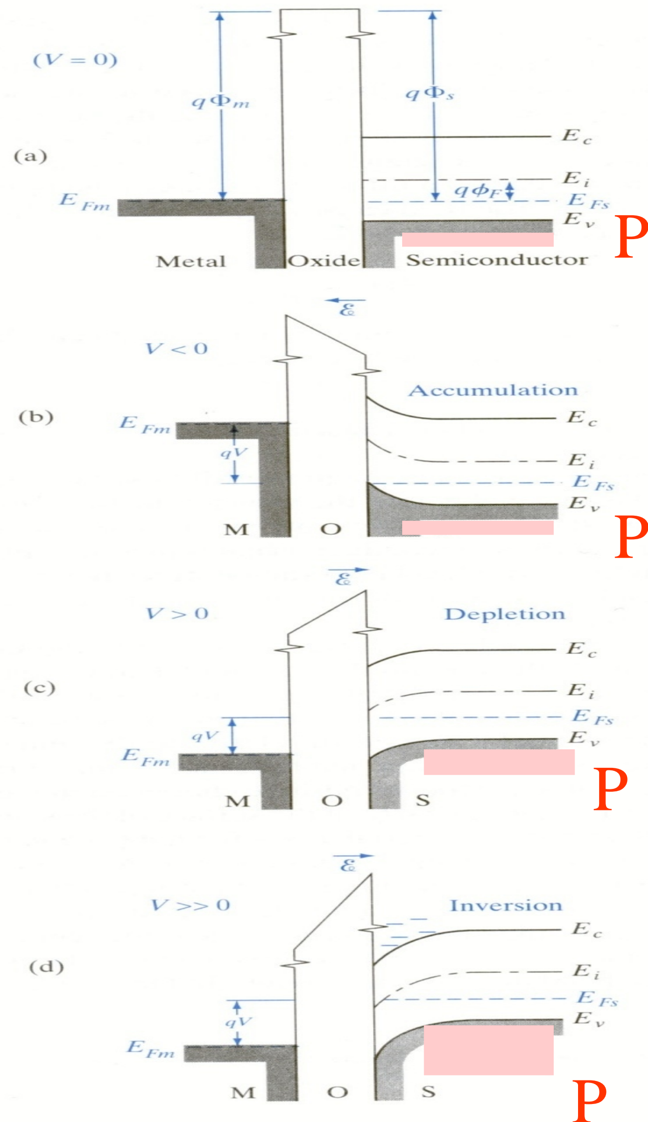
MOSFET



MOS Capacitors

As the MOSFET is in many ways just a capacitor we'll spend some time investigating MOS capacitors and the voltage required to create a region of inversion in the bulk material.

MOS Capacitors



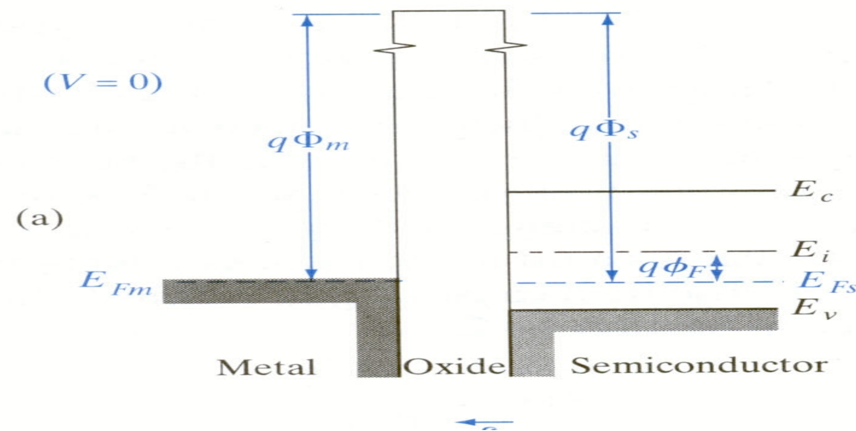
Equilibrium

Accumulation
(negative voltage)

Depletion
(positive voltage)

Inversion
(large positive voltage)

MOS Capacitors

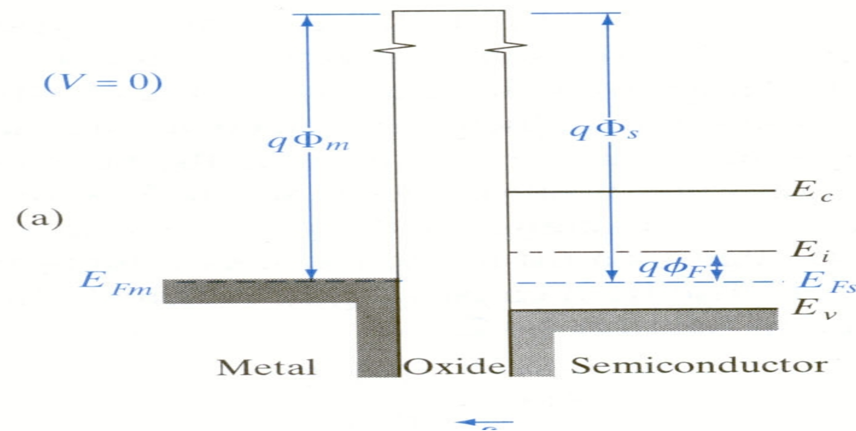


A material's work function is a measure of the energy required to move an electron from a material to outside of it.

$q\Phi_M$ = Energy required to move electron from metal's Fermi Level to insulators conduction band.

Metals conduction band and valence band are the same

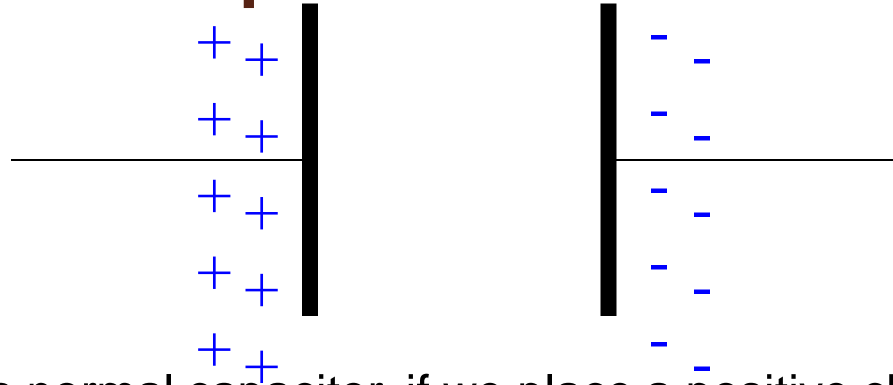
MOS Capacitors



$q\Phi_S$ = Energy required to move electron from semiconductor Fermi-Level to insulator's conduction band.

$q\Phi_F$ = Energy required to move electron from semiconductor Fermi-Level to the semiconductors intrinsic Fermi energy level.

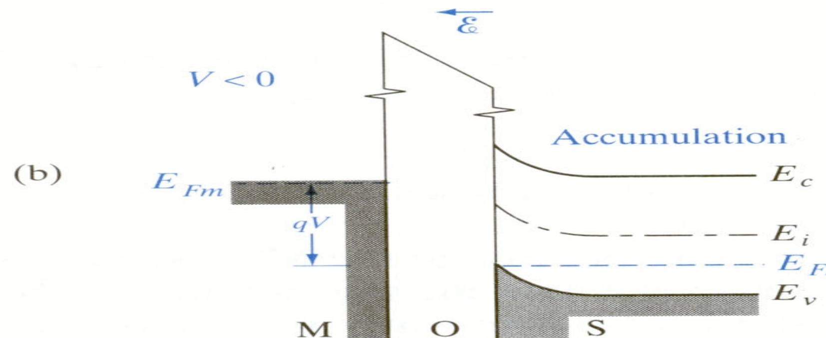
MOS Capacitors



Consider a normal capacitor, if we place a positive charge on one side of then negative charge accumulates on the other side.

In the MOS capacitor, the same happens. We charge up one side of the plate (ie the gate) to some voltage, and this accumulates the opposite charge in the bulk semiconductor.

MOS Capacitors



Now let's assume that we have P type semiconductor and place a negative voltage on the metal.

If we do this then we pull additional holes to the surface of the semiconductor near the insulator.

But the hole concentration in a semiconductor is defined by the following equation

$$p_o \approx p_i \exp\left(\frac{E_i - E_F}{kT}\right)$$

NOTE: Ei defined as halfway between conduction and valence bands.

MOS Capacitors

$$p_o \approx p_i \exp\left(\frac{E_i - E_F}{kT}\right)$$

In equilibrium in our MOS junction, the Fermi Energy level on both sides would be equal, however the bias applied across the junction will shift it.

Apart from this bias, E_F will remain constant in the material.

So if p_o increases, and p_i and E_F, then something has to change, and it's E_i. Changes in this value indicates that the behaviour of the material is changing.

MOS Capacitors

If E_i is moving then the position of the conduction and valence also move with it. Basically, for our example, if more hole carriers are available, then the probably of free electron carriers decreases.

So graphically, the conduction band goes up, indicating the higher energy required for a free electron carrier. Similarly for the valence band.

When the number of dominant free carriers increases due to the MOS capacitance effect, the effect is called ACCUMULATION

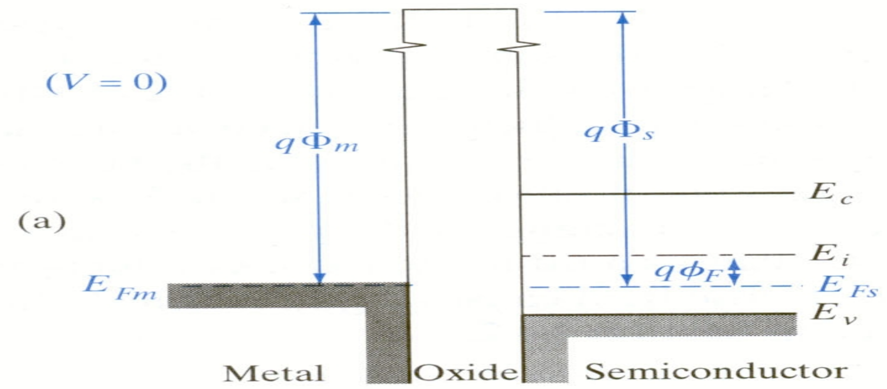
MOS Capacitors

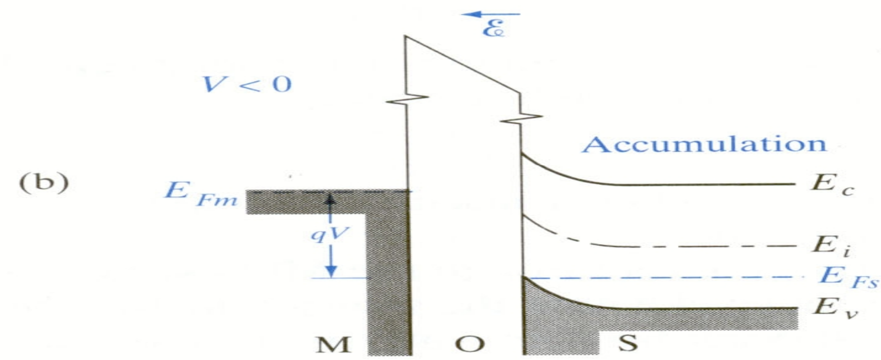
Now imagine we place a positive charge on the metal gate. This would mean the in the P-type semiconductor negative charge would be attracted to the surface of the semiconductor.

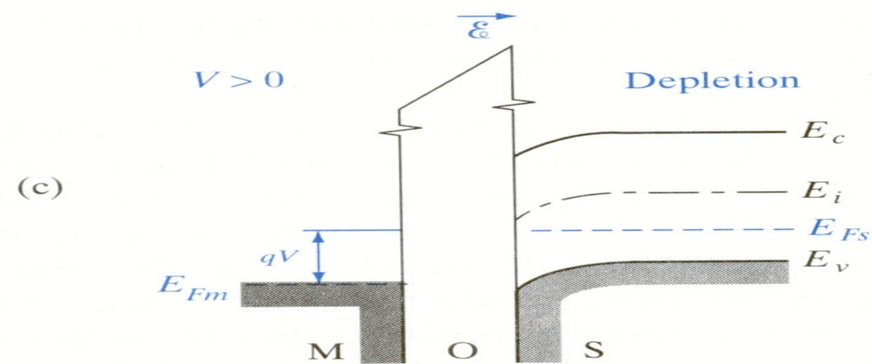
From before, we can predict that this will force the energy bands in the opposite direction from before, the conduction band dipping lower near the edge of the semiconductor.

When the number of dominant free carriers decreases due to the MOS capacitance effect, the effect is called DEPLETION.

MOS Capacitors







MOS Capacitors

If a sufficiently positive voltage is applied then the conduction band can drop below the Fermi Energy Level of the material. Now if we look at the following equation we can see an important effect.

$$n_o \approx n_i \exp\left(\frac{E_F - E_i}{kT}\right)$$

Normally E_i is greater than E_F in a P-type semiconductor, but where E_i is less than E_F, this equation indicates that there will be a large number of free electron carriers in the conduction band.

MOS Capacitors

Similarly if the equation predicts a large number of electron carriers then the equivalent hole equation indicates a lack of hole carriers.

Therefore the material close to the semiconductor edge is acting like an N type semiconductor.

When the semiconductor is forced to change dominant carrier types, we call the effect INVERSION.

MOS Capacitors

What is important is to determine the applied voltage required to cause inversion in the bulk semiconductor.

This is what we'll be considering for the next couple of lectures.

MOS Capacitors

We know inversion is obtained when E_i is reduced below E_F. Basically the midpoint between the conduction and valence band, E_i , needs to move by Φ_F.

However for STRONG INVERSION, it's normally stated that strong inversion occurs when the N-type region caused by inversion is as much N type and the region was origionally P type, so if E_i was originally Φ_F above E_F it now needs to be Φ_F below it. Thus E_i needs to move by $2*\Phi_F$.

MOS Capacitors

Working on this we can say that at the surface of the semiconductor, if we say the change in E_i is Φ_S, we can say that

$$\Phi_S \geq 2\Phi_F$$

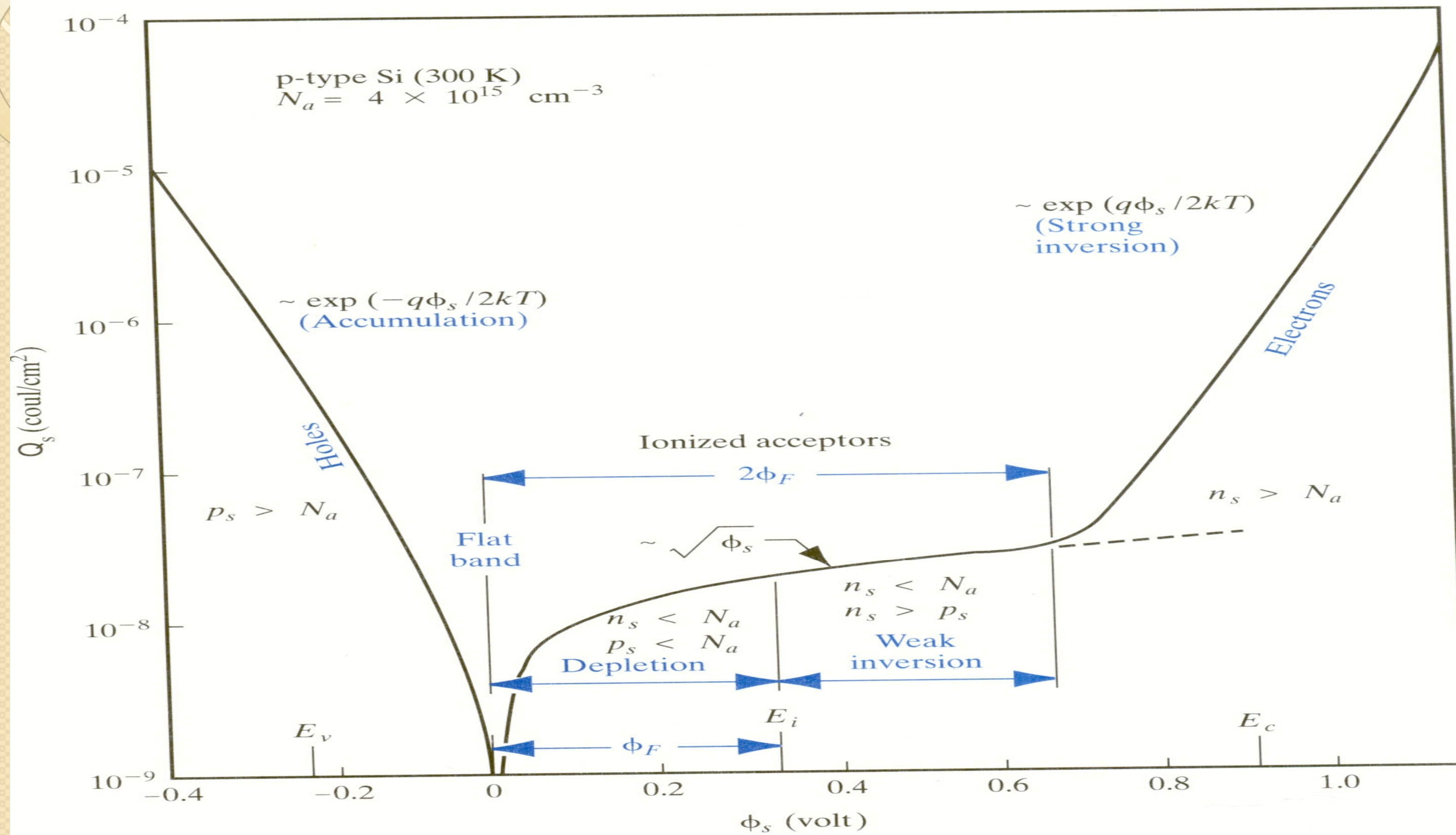
But as Φ_F is the difference between E_F and E_i, we can say that (assuming P type material)

$$p_o = n_i \exp\left(\frac{E_i - E_F}{kT}\right)$$

$$N_A = n_i \exp\left(\frac{q\Phi_F}{kT}\right)$$

$$\Phi_F = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right)$$

MOS Capacitors



Threshold Voltage V_T

On the following page there is a diagram. It shows the charge, electric field and voltage across the MOS device.

The charge on each side of the insulator must be equal. The charge in the metal is equally distributed across the width of metal.

The charge in the semiconductor is greatest at the edge, where there is an inversion region, and then there is some charge in a depletion region, insufficient to invert the type but we'll assume sufficient to deplete all available carriers.

Threshold Voltage V_T

The electric field in the metal is zero but once you enter the insulator there is a large constant field due to the distance from the electric charge in the metal.

Once you enter the semiconductor, first the inversion region causes a rapid decline in electric field and the remainder of the electric field is dissipated in the depletion region in a linear fashion.

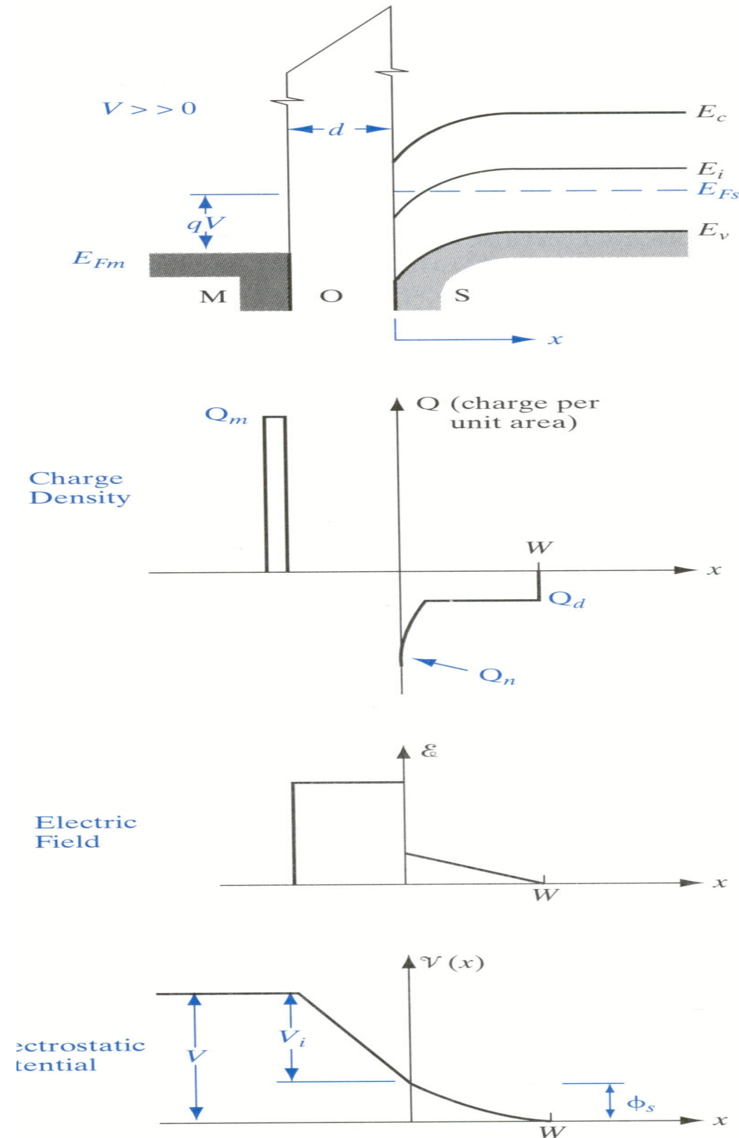
Threshold Voltage V_T

In terms of voltage, a certain amount of the voltage is dropped across the insulator and the rest in the semiconductor itself.

The voltage dropped across the insulator is the voltage required for the capacitor formed by the MOS.

The rest of the voltage is dropped across the depletion region caused by the collection of charge at the junction (assuming we are going for depletion). If this voltage is sufficiently large inversion will occur.

Threshold Voltage V_T



Threshold Voltage V_T

Now to get the required applied voltage, we need consider the voltage across the capacitor.

The voltage across the capacitor is given by

$$V_C = \frac{Q_C}{C_i}$$

where

V_C = is the voltage across the capacitor

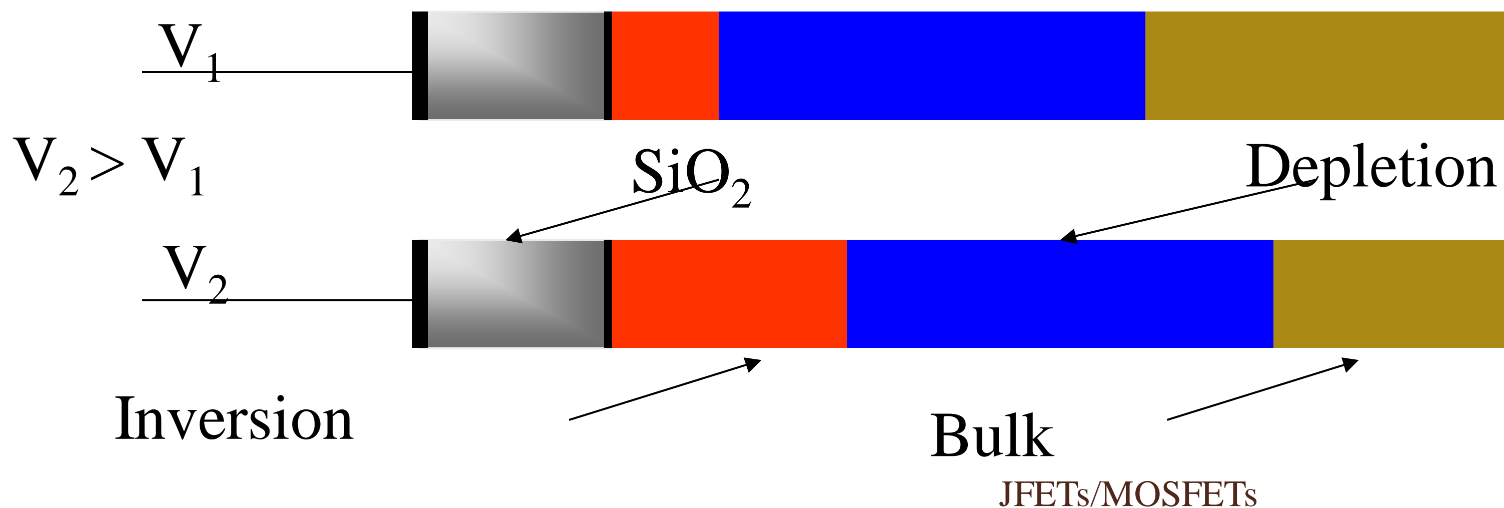
Q_C = the charge stored on either side of capacitor

C_i = the capacitance per unit area of the insulator

Threshold Voltage V_T

To calculate the charge of the capacitor, we can work out the charge in the depletion region as in the PN junction.

In this case, increased voltage will only make the depletion region grow UNTIL strong inversion is caused. After that, increased voltage will only strengthen the inversion. Consider it as a moving barrier. (This is a simplification as ever)



Threshold Voltage V_T

So assuming no inversion, the depletion region can be calculated as in the PN junction, just up to the point where strong inversion develops. Also we'll use the fact that there is no depletion region in metal, thus

$$W = \sqrt{\frac{2\varepsilon_i V}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)}$$

$$= \sqrt{\frac{2\varepsilon_i \Phi_S}{qN_A}}$$

Now the maximum depletion region occurs just as strong inversion occurs, giving

$$W_{\max} = \sqrt{\frac{4\varepsilon_{semi}\Phi_F}{qN_A}}$$

$$= \sqrt{\frac{4\varepsilon_{semi} kT \ln\left(N_A / n_i\right)}{q^2 N_A}}$$

Threshold Voltage V_T

$$W_{max} = \sqrt{\frac{4\varepsilon_{semi}\Phi_F}{qN_A}}$$

Now with W, we can say the charge in the depletion region, just as strong inversion is about to form, is given by

$$\begin{aligned} Q_D &= -qN_A W_{max} \\ &= -2\sqrt{qN_A \varepsilon \Phi_F} \end{aligned}$$

Therefore the voltage required to maintain this charge in a capacitor is given by

$$V_C = \frac{Q_D}{C_i}$$

Threshold Voltage V_T

So to obtain depletion, we need $\Phi_s = 2\Phi_F$ to obtain strong inversion and we also need V_C to support the charge in the capacitor also required at strong depletion.

$$V_T = \Phi_S - \frac{Q_D}{C_i}$$
$$= 2\Phi_F - \frac{Q_D}{C_i}$$

We are now nearly at the final threshold voltage that the gate needs to obtain in reverse bias for strong inversion to occur. This voltage would be sufficient if the original Fermi Energy Levels were equal in equilibrium.

Flat Band Voltage

In practice the metal's Fermi Energy Level and that of the semiconductor are not equal and as they are forced equal a voltage is set up across the junction.

This bends the energy bands close to the insulator. The effect of this is to always approach inversion. A metal's Fermi Energy level is always lower.

The difference between the metal and semiconductor Fermi Energy Levels is called Φ_{ms}. It always is of a value that reduces the required threshold voltage. It can be so large as to cause inversion by itself.

The Flat-Band Voltage is the voltage that would need to be applied to get the bands flat

Real Surfaces

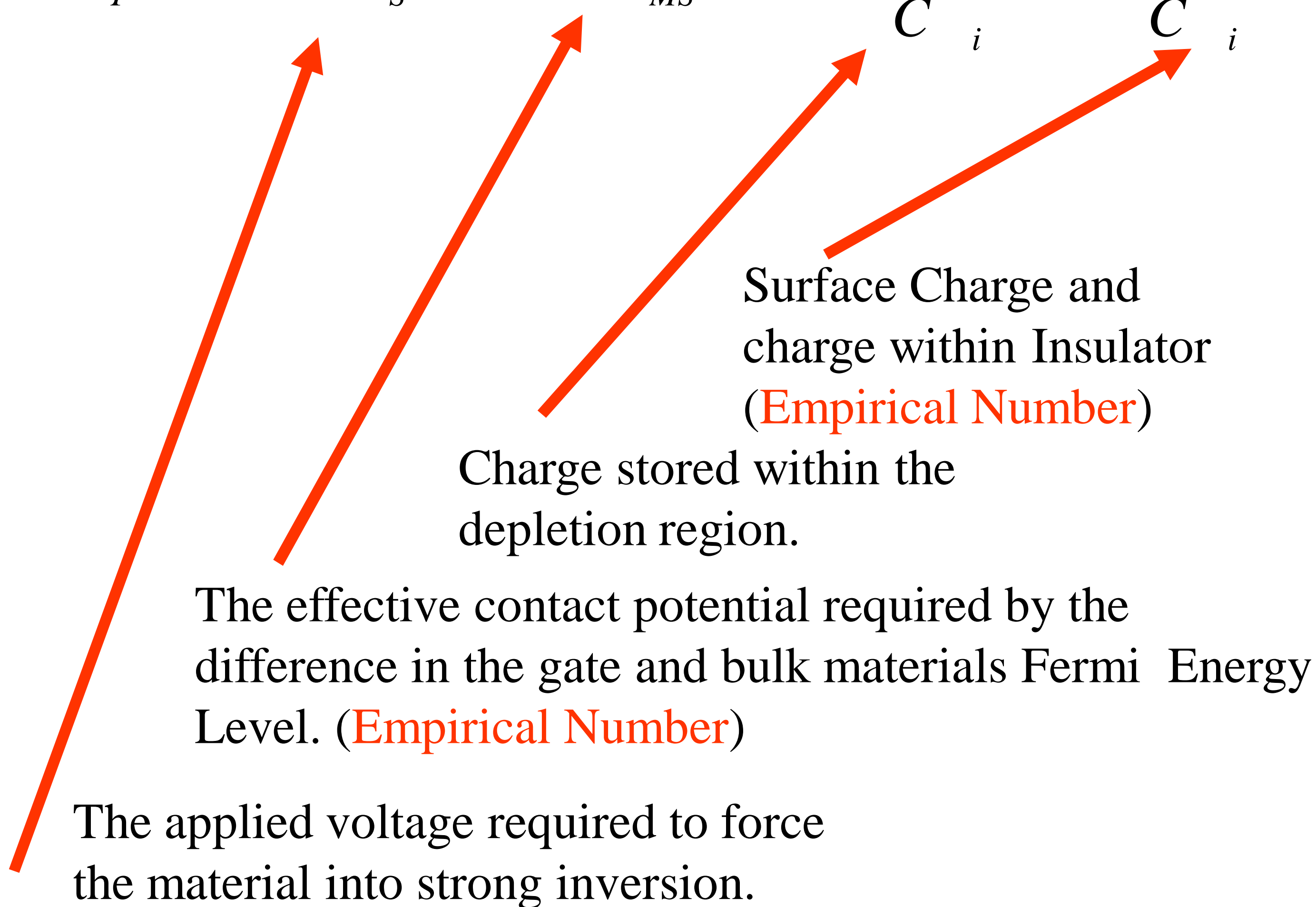
In real surfaces and insulators there is always the possibility of trapped charges, either from impurities entering the manufacturing process or lattice defects at the junction between the semi-conductor and the oxide.

These act as additional charge on the insulator capacitor that the applied voltage needs to overcome.

They also can be considered to be band-bending artifacts and are often factored into the flat-band voltage.

This additional charge is often designated C_i in the literature.

Threshold Voltage



Depletion Region Capacitance

Another capacitor that needs to be considered is the depletion region capacitance. This has a charge with-respect-to the metal plate which we worked out earlier.

In strong accumulation or strong inversion, the insulator capacitor is dominant but when it's just undergoing depletion, this is the dominant junction capacitance and it is much lower than the insulator capacitance.

Review of MOS-C

We have determined that it is possible to use a MOS Capacitor structure to invert the type of a semiconductor through the application of a voltage.

We have determined the minimum required voltage and the factors that contribute to this voltage.

It is easy to see how in a MOSFET, this inversion layer allows current to flow.

Example

Calculate the inversion voltage V_T for a MOS capacitor structure that has inherent band mismatch. The bulk material is P type, the insulator is SiO_2 and the gate is polysilicon.

N_A = 10^{16} /cm^3

n_i^2 = $2.25.10^{20}$ /cm^3

$\varepsilon_r(SiO_2)$ = 3.9

$\varepsilon_r(Si)$ = 11.8

oxide thickness

= 0.01 um

Φ_{MS} = -0.35V

Interface charge

= 5.10^{10} electrons /cm^2

Example

Now V_T is given by this equation, so all we need do is work out each term.

$$V_T = \Phi_S + \Phi_{MS} - \frac{Q_D}{C_i} - \frac{Q_i}{C_i}$$

$$\Phi_S = 2\Phi_F$$

So first let's calculate Φ_F

$$\Phi_F = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right)$$

$$= (0.026)\ln\left(\frac{10^{16}}{\sqrt{2.25 .10^{20}}}\right)$$

$$= 0.35\ V$$

Example

Now we've been given Q_i and Φ_{MS} so all we need calculate is Q_D and C_i

$$Q_D = -2\sqrt{qN_A \varepsilon \Phi_F}$$

$$= -2\sqrt{(1.6 . 10^{-19})(10^{16})(11.8 * 8.85 . 10^{-14})(0.35)}$$

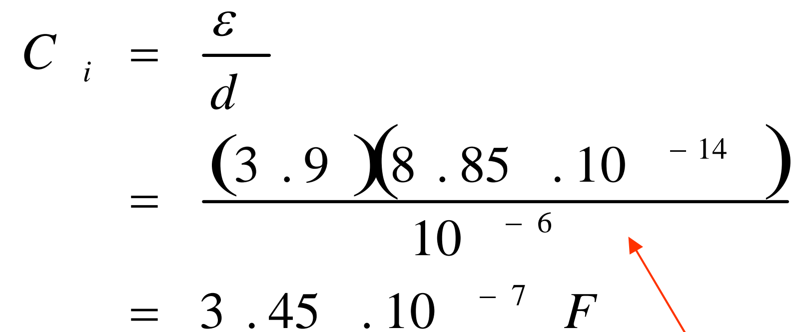
$$= -2\sqrt{5.848 . 10^{-16}}$$

$$= -4.84 . 10^{-8}\ C$$

As to be expected, the charge in the depletion region is negative, acceptors have recombined with electrons to become negatively charged.

Example

The last term C_i



$$C_i = \frac{\varepsilon}{d}$$

$$= \frac{(3.9)(8.85 . 10^{-14})}{10^{-6}}$$

$$= 3.45 . 10^{-7}\ F$$

Convert to cm as everything else is in cm's in this equation.

Example

So bringing this all together

$$V_T = \Phi_S + \Phi_{MS} - \frac{Q_D}{C_i} - \frac{Q_i}{C_i}$$

$$= 2(0.35) - 0.35$$

$$- \frac{-4.84 . 10^{-8}}{3.45 . 10^{-7}}$$

$$- \frac{-(5 . 10^{10} * 1.6 . 10^{-19})}{3.45 . 10^{-7}}$$

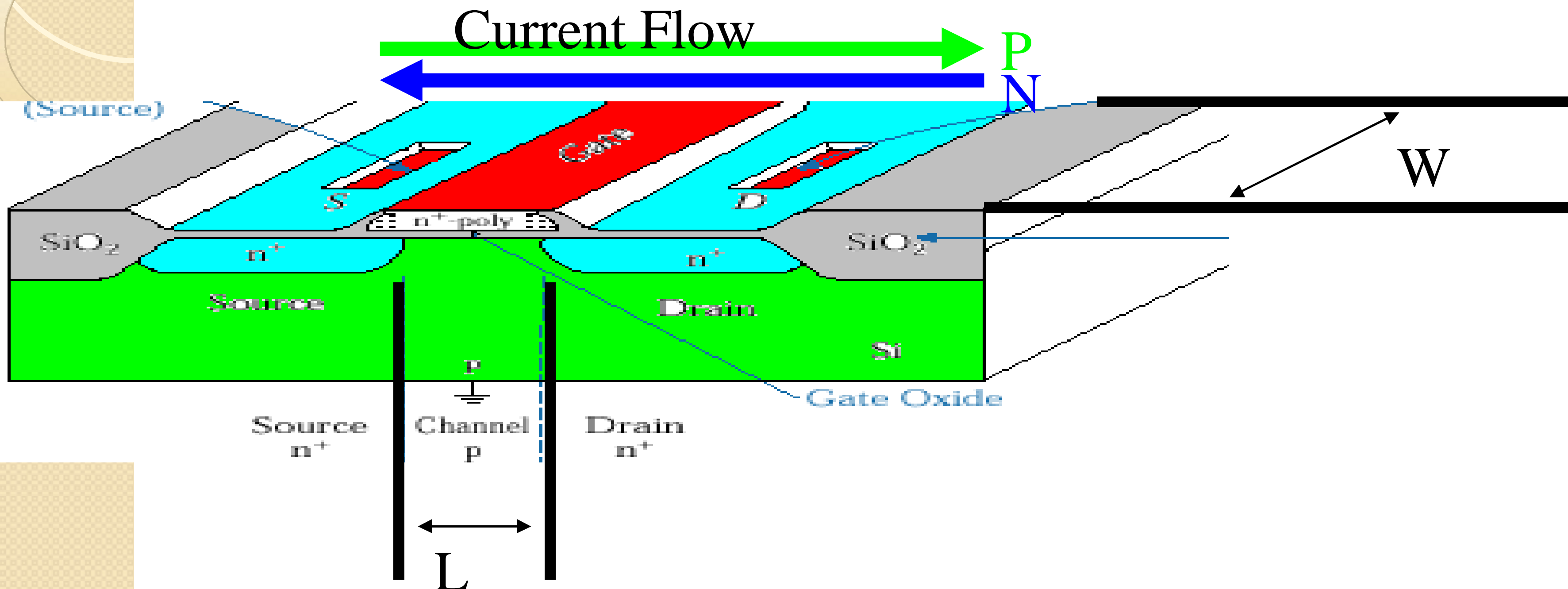
$$= 0.69 - 0.35 + 0.14 + 0.023$$

$$= 0.5 V$$

If Φ_{MS} were much larger then the channel could be naturally inverted.

Interface charge is a small contributer

MOSFET



In a MOSFET there are two important physical parameters,

W	Width of the channel
L	Length of the channel

MOSFET

Given the MOSFET, the important characteristic we need to determine is the relationship between drain-source current and gate voltage.

This is a straightforward task as the current never crosses a junction.

As the current only flows through one type of material, the voltage-current relationship is resistive.

The resistance presented to the current is dependent on the mobility and the available number of carriers.

The carriers available in the inverted region depends on the charge induced.

MOSFET

The gate voltage, V_G, induces an inversion layer, charge in the depletion region and also has a component required to obtain a flat-band at equilibrium.

$$V_G = \Phi_S + V_{FB} - \frac{Q_S}{C_{ox}}$$

Insulator (oxide) capacitance

where

- V_{FB} is the required voltage to obtain flat bands in the semiconductor prior to application of a gate voltage. It includes Φ_{MS} and any other contributions from trapped surface charges or charges in the insulator oxide, SiO_2
- Φ_S is the charge induced in the semiconductor forming the substrate of the device, it includes free charge carriers in the inversion region and the charge in the depletion region.

MOSFET

Φ_S is the charge induced in the semiconductor forming the substrate of the device.

It includes free charge carriers in the inversion region and the charge in the depletion region.

$$Q_S = Q_D + Q_{inv}$$

We are interested in the charge in the inversion region as this is the free charge available for carrying current.

$$Q_{inv} = -C_{ox}\left(V_G - \left(V_{FB} + \Phi_S - \frac{Q_D}{C_{ox}}\right)\right)$$

MOSFET

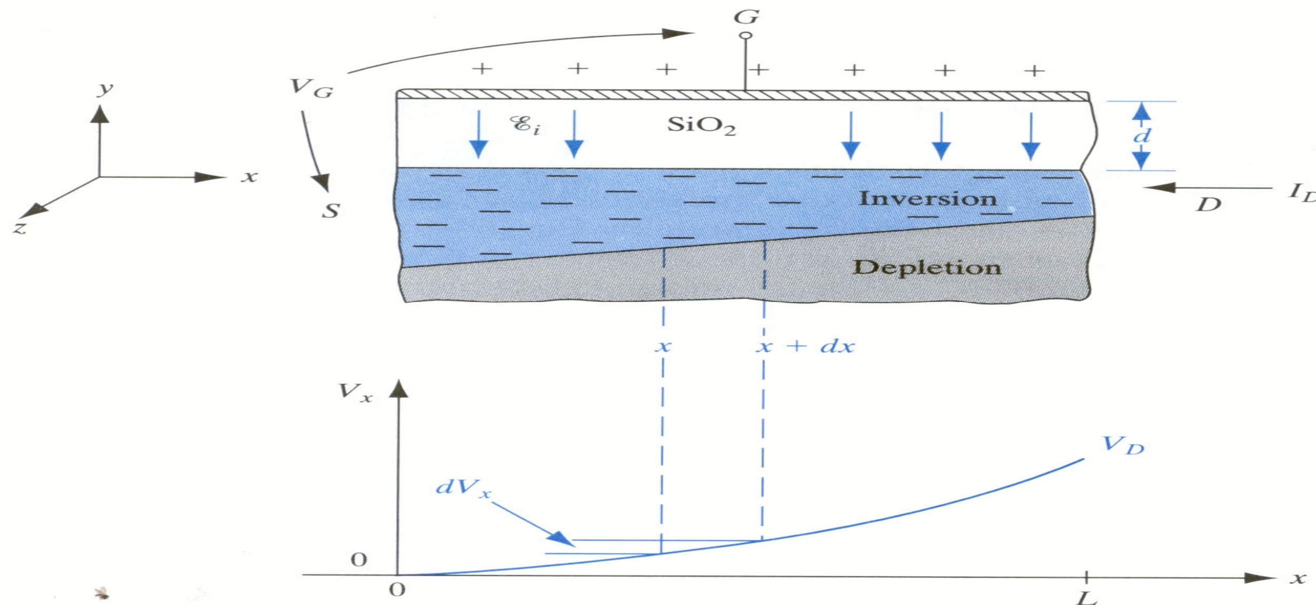
$$Q_{inv} = -C_{ox}\left(V_G - \left(V_{FB} + \Phi_S - \frac{Q_D}{C_{ox}}\right)\right)$$

At threshold, just as the inversion region is formed, the term inside the double brackets is the threshold voltage, V_T, so from here we'll simplify this to V_T

$$Q_{inv} = -C_{ox}\left(V_G - \left(V_T\right)\right)$$

MOSFET Current I_D

As in the JFET, current travelling through the channel will cause a voltage drop and this voltage drop will cause a variation in the inversion region.



MOSFET

Now the voltage at either end of the channel is defined by the drain and source voltages. So using the source as a reference, define a voltage V(x) to indicate the change in voltage across the channel.

This changes the equation from the previous page to

$$Q_{inv} = -C_{ox}\left(V_G - \left(V_{FB} + \Phi_S - \frac{Q_D(x)}{C_{ox}}\right) - V(x)\right)$$

$$= -C_{ox}\left(V_G - V_{FB} - 2\Phi_F - V(x) + \frac{Q_D(x)}{C_{ox}}\right)$$

Note that $Q_D(x)$ has a dependency on the applied voltage.

MOSFET

Now we know the charge stored in the depletion region, and in the case of a P-type bulk and an N-type inversion region, the charge is negative.

$$Q_{inv} = -C_{ox}\left(V_G - V_{FB} - 2\Phi_F - V(x) + \frac{Q_D(x)}{C_{ox}}\right)$$

We have an equation for the charge within the depletion region, but from earlier we know that it maxes out when the inversion region forms and thereafter does not grow. (simplification)

If it does not grow, then the charge within the region stays the same, so we can say that if there is an inversion region, the depletion region charge is constant.

MOSFET

If we can assume that the charge in the depletion region is constant, a fairly reasonable assumption, then the charge in the inversion region can be simplified to

$$Q_{inv} = -C_{ox}\left(V_G - V_T - V(x)\right)$$

MOSFET Current Eqn

Once we have the charge in the inversion layer at any point for a given drain voltage, we can use this to calculate the current in the MOSFET channel.

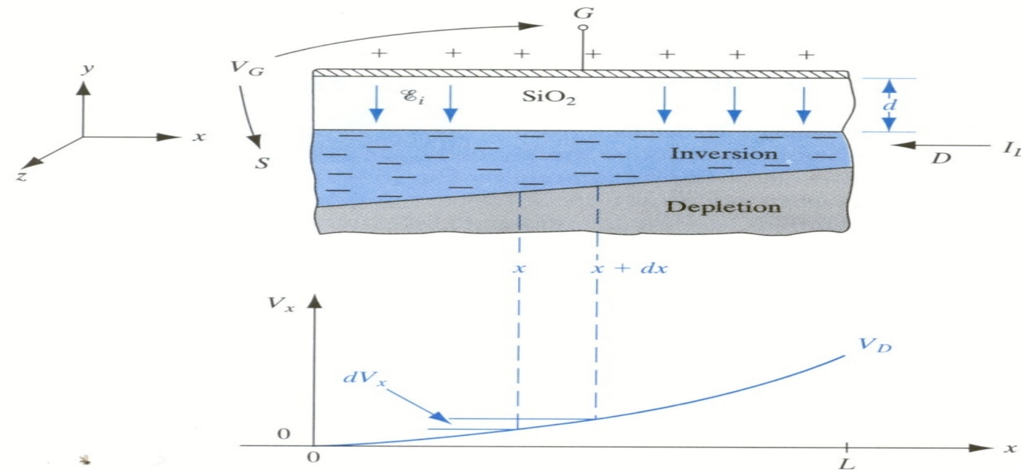
The channel acts like a resistor and we know the equation for conductivity

$$\sigma_{channel} = q\,\overline{\mu}\,N$$

where μ is the surface mobility in the inverted channel and qN represents the available charge in the channel.

So we are going to look at the conductivity of a slice of the channel and from that estimate the current.

MOSFET Current Eqn



Take an infinitesimal slice, Δx along the length of the channel. The crossectional area of that slice is length (Δx) by width (W).

$$\sigma_{slice} = Q(x)\overline{\mu}_n\left(\frac{W}{dx}\right)$$

Note, in conductivity,
Width increases,
length decreases.
Opposite to resistivity

MOSFET Current Eqn

$$\sigma_{slice} = Q(x)\overline{\mu}_n\left(\frac{W}{dx}\right)$$

Now

$$I_D(x) = \sigma_{slice}\, dV_x$$
$$= Q(x)\overline{\mu}_n\left(\frac{W}{dx}\right)dV_x$$

Therefore

$$I_D(x)\,dx = Q(x)\overline{\mu}_n W dV_x$$
$$\int_0^L I_D(x)\,dx = \int_0^{V_{DS}} Q(x)\overline{\mu}_n W dV_x$$
$$I_D L = \overline{\mu}_n W C_{ox} \int_0^{V_{DS}} (V_{GS} - V_T - V_x)\,dV_x$$
$$I_D = \frac{W}{L}\overline{\mu}_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

MOSFET Current Eqn

$$I_D = \frac{W}{L}\overline{\mu}_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

This is the large signal equation that determines the current in the MOSFET, with the assumption that the charge in each slice of the channel remains constant.

This is not true, and the voltage dependency of Q(x) should be added to the equations and also integrated. If this is done, the resulting equation is

$$I_D = \frac{W}{L}\overline{\mu}_n C_{ox}\begin{pmatrix} (V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2 \\ -\frac{2}{3}\frac{\sqrt{2\varepsilon qN_A}}{C_{ox}}\left[(V_D + 2\Phi_F)^{\frac{3}{2}} - (2\Phi_F)^{\frac{3}{2}}\right] \end{pmatrix}$$

MOSFET Current Eqn

$$I_D = \frac{W}{L}\overline{\mu}_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

$$I_D = \frac{W}{L}\overline{\mu}_n C_{ox}\left(\begin{array}{l}(V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2 \\ -\dfrac{2}{3}\dfrac{\sqrt{2\varepsilon qN_A}}{C_{ox}}\left[(V_{DS} + 2\Phi_F)^{\frac{3}{2}} - (2\Phi_F)^{\frac{3}{2}}\right]\end{array}\right)$$

The second equation is a much more complete and accurate model and is used when drawing the transfer characteristics.

However for most design applications, the first equation is sufficiently accurate to provide a good first iteration solution without requiring the use of computer solutions.

MOSFET Characteristics

$$I_D = \frac{W}{L}\mu_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

If we take this simplified equation and look at if for small values of V_{DS}, then presuming that an inversion layer exists ($V_{GS} > V_T$) then the device looks like a resistor.

$$I_D = \frac{W}{L}\mu_n C_{ox}(V_{GS} - V_T)(V_{DS})$$

$$R = \frac{V_{DS}}{I_D} = \frac{L}{W}\frac{1}{\mu_n C_{ox}(V_{GS} - V_T)}$$

So low V_{DS} means small currents, so the MOSFET looks like a resistor when the variation in the width of the inversion region is not large.

MOSFET Characteristics

$$I_D = \frac{W}{L}\mu_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

Now if the current increases then eventually we'll hit pinch-off. The inversion region is smallest at the drain and the current saturates.

Now we can approximately say that the saturation voltage for V_{DS} occurs when it is as positive as $(V_{GS}-V_T)$. As the applied voltage is actually between the gate and the channel and if the channel rises in voltage the effective gate voltage decreases. Saturation occurs about when the excess of V_{GS} over V_T has been removed. (Very simplified)

$$V_{DS}(sat) \cong V_{GS} - V_T$$

MOSFET Characteristics

$$I_D = \frac{W}{L}\mu_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

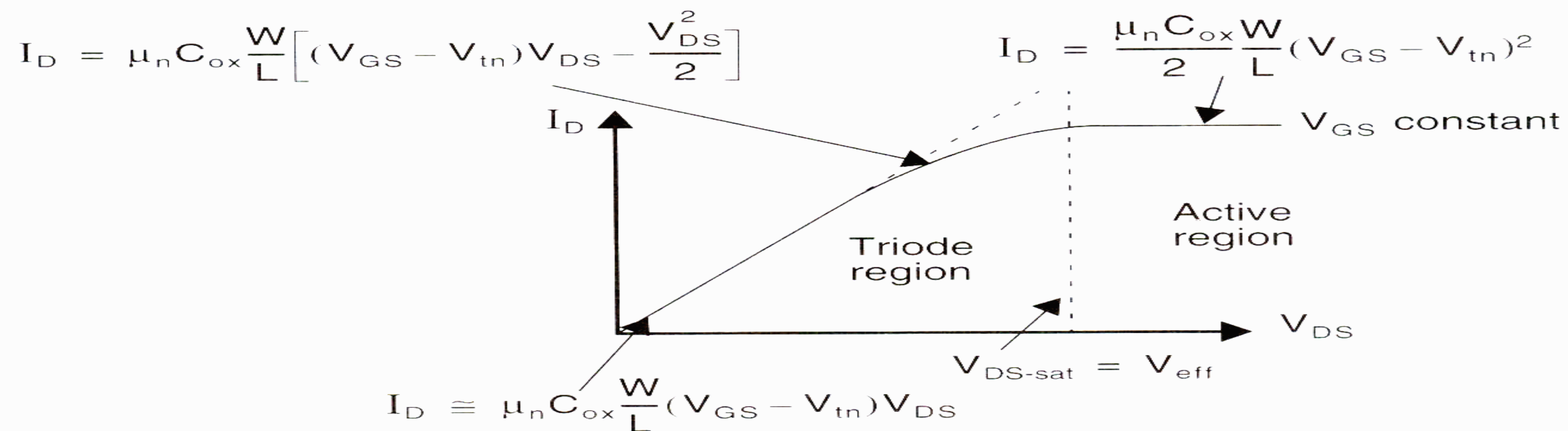
$$V_{DS}(sat) \cong V_{GS} - V_T$$

If we take $V_{DS(sat)}$ and place it into the equation for I_D, we can find the saturation current at pinch-off.

$$I_D(sat) = \frac{W}{L}\mu_n C_{ox}\left(\tfrac{1}{2}(V_{GS} - V_T)^2\right)$$

$$= \frac{W}{L}\frac{\mu_n C_{ox}}{2}(V_{GS} - V_T)^2$$

Large Signal Characteristics



$$I_D = \frac{W}{L}\mu_n C_{ox}\left((V_{GS} - V_T)V_{DS}\right)$$

$$I_D = \frac{W}{L}\mu_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

$$I_D(sat) = \frac{W}{L}\frac{\mu_n C_{ox}}{2}(V_{GS} - V_T)^2$$

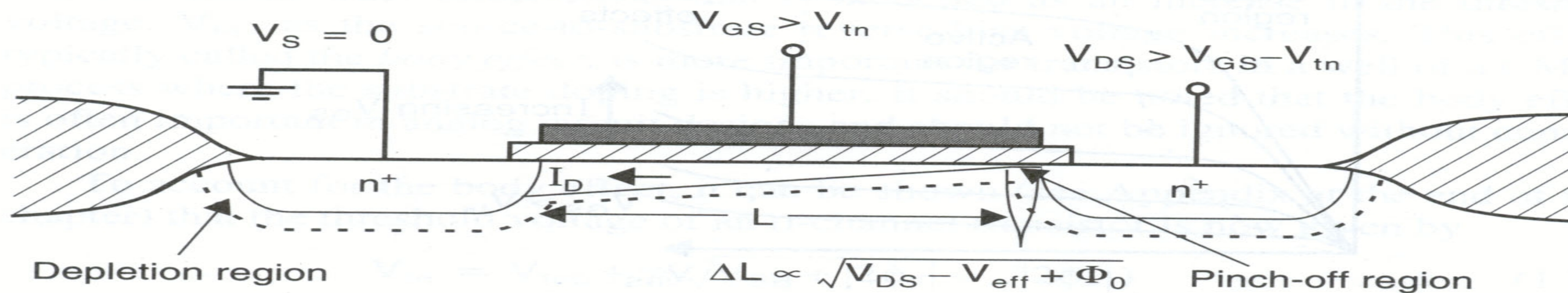
Channel Length Modulation

We have now calculated the saturation current for the transistor however it has not taken into account all the effects. It has been simplified.

The major assumption that we made was that after pinch-off, there was no variation in I_D for changes in the drain-source voltage, V_{DS}. However this is not true.

This is because we made a simplification earlier that the depletion region did not grow once there was an inversion region, however at pinch-off there isn't really an inversion region so the depletion region changes with applied voltage.

Channel Length Modulation



The voltage at the end of the channel, during pinch-off, is fixed at

$$V_{channel}(pinch - off) \cong V_{GS} - V_T$$

Now the end of the channel and the drain are separated by a depletion region through which the current still flows.

Channel Length Modulation

The voltage difference between the section of the channel at pinch-off and the voltage on the drain is dropped across this depletion region.

This depletion region grows with the voltage difference, according to the equations from before.

$$W = \sqrt{\frac{2\varepsilon_i V}{q}\left(\frac{1}{N_A}\right)}$$

Use only P type region as it's safe to assume that the drain (N-type) is very heavily doped.

$$= \left(\sqrt{\frac{2\varepsilon_i}{qN_A}}\right)\sqrt{\Phi_0 + V_{drain-channel}}$$

In future, we'll simplify the constant to

$$k_{ds} = \sqrt{\frac{2\varepsilon_i}{qN_A}}$$

Channel Length Modulation

Now accept that the current will flow across this depletion region, an equilibrium is set up. All injected carriers are swept to the drain.

However if the depletion region around the drain has increased, the distance in which the current has to traverse the inversion region has gotten shorter.

The shorter the length of the region, the less resistance it presents to the current. The less resistance, the more current.

Channel Length Modulation

Now we need to express I_D taking into account the channel modulation factor. Remember L is being modulated.

A good way of doing this, it produces a nice answer, is to take a Taylor approximation for I_D around its operating value of

$$V_{DS} = V_{GS} - V_T$$

This gives an expression for current

$$I_D = I_{Dsat} + \frac{\partial I_D}{\partial L} \frac{\partial L}{\partial V_{DS}} \Delta V_{DS}$$

Channel Length Modulation

$$I_D = I_{Dsat} + \frac{\partial I_D}{\partial L}\frac{\partial L}{\partial V_{DS}}\Delta V_{DS}$$

$$= I_{Dsat}\left(1 + \frac{k_{ds}\left(V_{DS} - \left(V_{GS} - V_T\right)\right)}{2L\sqrt{V_{DS} - \left(V_{GS} - V_T\right) + \Phi_0}}\right)$$

$$= \frac{\mu_n C_{ox}}{2}\frac{W}{L}\left(V_{GS} - V_T\right)^2\left(1 + \frac{k_{ds}\left(V_{DS} - \left(V_{GS} - V_T\right)\right)}{2L\sqrt{V_{DS} - \left(V_{GS} - V_T\right) + \Phi_0}}\right)$$

And this is commonly reduced to

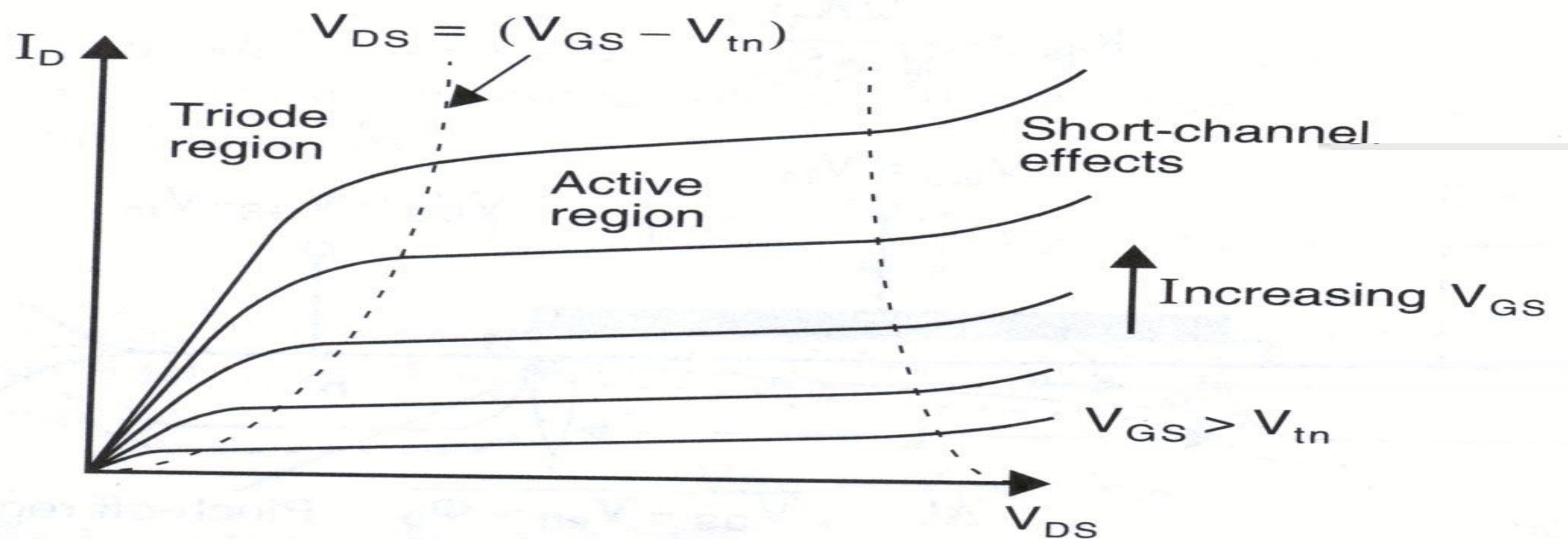
$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}\left(V_{GS} - V_T\right)^2\left(1 + \lambda\left(V_{DS} - \left(V_{GS} - V_T\right)\right)\right)$$

Where the λ, the output impedance constant is

$$\lambda = \frac{k_{ds}}{2L\sqrt{V_{DS} - \left(V_{GS} - V_T\right) + \Phi_0}}$$

Channel Length Modulation

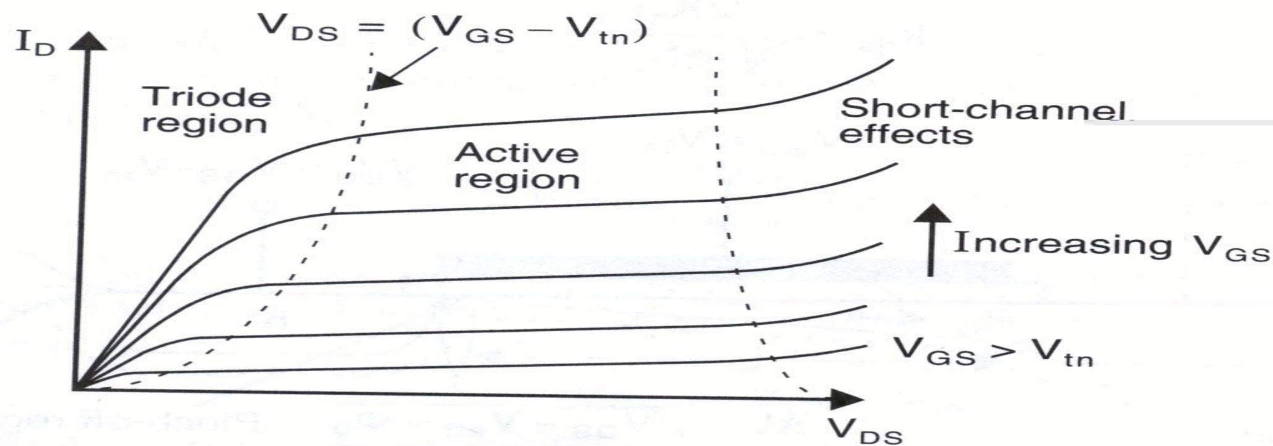
Now we have considered one major second order effect. If we were to draw the characteristics now for the transfer function, in saturation the curves would have a small slope to them.



Channel Length Modulation

This is accurate until the voltages get to high and then we have something that is called a short-channel effect but is different in nature to the effects that we'll encounter next.

At high voltages the carriers are being accelerated and after a certain V_{DS} voltage the current in the channel beings to increase faster than expected in saturation.



Short Channel Effects

Another major effect, and one that is becoming much more dominant with modern devices are short-channel effects.

In brief, none of our analyses so far has considered that carriers have a maximum speed within a lattice. So if we increase the electric field that accelerates the electrons, at some point the electrons velocity saturate. In short devices, even for low voltages, velocity saturation can occur.

In short channel length devices this means that over most of the channel the electrons are travelling at the same speed.

Short Channel Effects

In this case the drain current is given by the width by channel charge/unit area and the saturation velocity.

$$I_D = WC_{ox}(V_{GS} - V_T)velocity$$

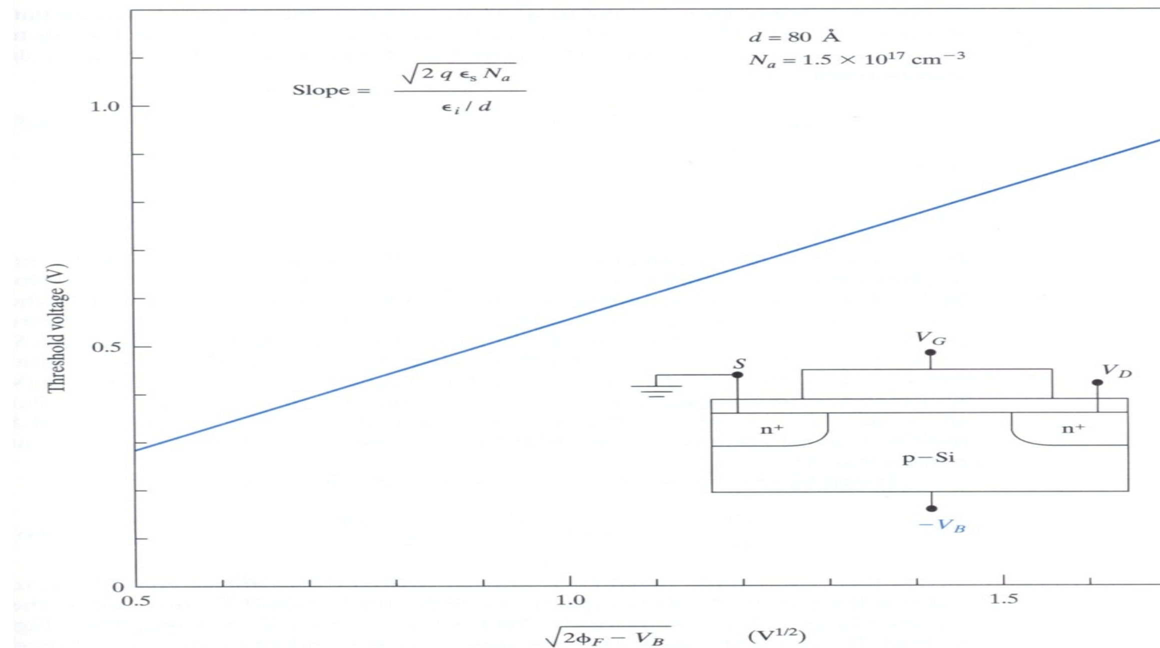
Now if we look back to our original equation for saturation current:

$$I_{D(sat)} = \frac{W}{L}\frac{\mu_n C_{ox}}{2}(V_{GS} - V_T)^2$$

So now we can see that in short channel devices the relationship is no longer quadratic (x^2) but linear. In current modern devices it's actually somewhere in between, about $x^{1.5}$ for minimum geometry devices. If you decide to make your devices longer, you get back to original.

Body Effect

Another very important second order effect is the *body effect*. The effect of this is to vary the threshold voltage.



The body effect occurs when the substrate of the transistor (in which the channel forms) is not at the same voltage as the source.

Body Effect

We've assumed so far that the source and substrate have been connected. If they are not connected then they must be reverse biased.

If they are not reversed bias then the PN junction caused by the substrate and source regions will be forward biased and the transistor will not operate.

If they are reversed biased then a larger depletion region will form between the two regions. This will introduce additional depletion region charge which the applied gate voltage will need to overcome to achieve inversion.

Body Effect

A simplified approach would be to consider that the substrate (bulk) voltage, V_{SB} (source-bulk voltage), has the same effect across the entire inversion channel, so basically if acts to reduce the applied voltage.

The depletion region charge is now given by

$$Q_D' = -\sqrt{2qN_A\varepsilon\left(2\Phi_F + V_{SB}\right)}$$

Assuming Pchannel

The change in substrate voltage due to the source substrate bias (V_{SB})

$$\Delta V_T = \frac{Q_D - Q_D'}{C_{ox}}$$

$$= \frac{-\sqrt{2qN_A\varepsilon}\left(\sqrt{2\Phi_F - V_{SB}} - \sqrt{2\Phi_F}\right)}{C_{ox}}$$

Body Effect

$$\Delta V_T = \frac{\sqrt{2qN_A\varepsilon}\left(\sqrt{2\Phi_F + V_{SB}} - \sqrt{2\Phi_F}\right)}{C_{ox}}$$

Now it is clear that if we want the PN junction reverse biased, we need V_{SB} positive and hence this will mean that the threshold voltage will increase.

If we assume V_{SB} to be much larger than Φ_F we can make a rough approximation

$$\Delta V_T \approx \frac{\sqrt{2qN_A\varepsilon V_{SB}}}{C_{ox}}$$

Thus if the source rises above the bulk voltage in a NMOS, the threshold voltage rises. In an NMOS, the P channel bulk is normally tied to ground.

Body Effect

Correction to handed out slides.

Now we said that in an NMOS, the P channel bulk is normally tied to ground.

In a PMOS, the N channel bulk is normally tied to the positive voltage supply and the source if it drops below the positive voltage supply voltage will begin to affect the threshold voltage.

This is to ensure that all substrate-to-channel junctions are always reverse biased or at least unbiased. It helps prevent latch up and unplanned current paths.

If this approach is taken, it avoids having to check continually to ensure that all PN junctions in the MOSFET are correctly biased. A good design tactic.

Body Effect

Almost always in a design the bulk and source of the transistors are always tied to the same voltage, and normally the most positive (PMOS) or negative voltage (NMOS).

If this were not the case then the source would vary according to the signal on the gate, which would then vary the source-bulk voltage which would affect the threshold voltage which would affect the biasing of the transistor which would affect the gain of the device. This could lead to very poor performance.

Sub-Threshold Operation

Now the triode region current equation is given by

$$I_D = \frac{W}{L}\mu_n C_{ox}\left((V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right)$$

Now this equation is only valid when there is strong inversion, if $V_{GS}<V_T$, it does not mean there is negative current or no current.

In practice we are now in weak inversion. Remembering back to the MOS Capacitor, the channel goes from depletion to weak-inversion to strong inversion. We assumed strong inversion when the inverted N-type region was as doped as the original P-type.

Sub-Threshold Operation

However it is possible for current to flow during weak inversion, it's just not as strong a current flow as strong inversion. Operation with only weak inversion is called sub-threshold operation.

The current in subthreshold can easily be calculated. The charge in the depletion region and the inversion region is dependent on the applied voltages.

$$I_D = \frac{W}{L}\mu_n (C)\left(\frac{kT}{q}\right)^2\left(1 - e^{\frac{-qV_{DS}}{kT}}\right)\left(e^{\frac{q(V_{GS}-V_T)}{nkT}}\right)$$

where

$$n = \frac{C_{depl} + C_{OX}}{C_{OX}} \approx 1.5$$

Modern Devices

Sub-Threshold Operation

$$I_D = \frac{W}{L}\mu_n (C\ )\left(\frac{kT}{q}\right)^2\left(1 - e^{\frac{-qV_{DS}}{kT}}\right)\left(e^{\frac{q(V_{GS} - V_T)}{nkT}}\right)$$

In subthreshold operation, the current is exponentially dependent on the gate voltage exceeding the threshold voltage and also on the drain-source voltage.

Once the drain source voltage is over a few kT/q then it's no longer relevant, but the current is highly dependent on the gate voltage, exponentially so.

This relationship can be used for specific applications but it is difficult to design for as you are matching V_T's which are highly process dependent. Subthreshold operation is normally only used for low frequency or ultra-low-power applications.

Sub-Threshold Operation

$$I_D = I_{D0}\left(e^{\frac{q(V_{GS}-V_T)}{nkT}}\right)$$

One aspect of sub-threshold operation is the leakage current of a MOSFET. From this simplified form above, where I_{D0} in modern devices is about 20nA, the leakage current is dependent on the difference between V_T and V_{GS}. Normally V_T has been designed for 0.7V, so that if V_{GS}=0, then the exponential result is a tiny fraction of 20nA.

If V_T is decreased then the leakage current goes up exponentially.

Sub-Threshold Operation

In portable applications saving power is important. The aim is to ever decrease the supply voltage, the target is to get down to a single battery, ie less than 1.5 V.

The easiest way to do this is to reduce V_T. That will allow V_{GS} to be smaller or to be biased such that a bigger swing is available. However if V_T is reduced the current goes up.

If the leakage current was even 1nA, a million transistors, easily obtainable, will take 1mA in leakage current alone. This will quickly drain a battery reducing battery lifetime. The tendency is to reduce V_T but also to reduce the leakage current through better designed devices.

Sub-Threshold Operation

One aspect of sub-threshold operation is the leakage current of a MOSFET. From this simplified form above, where I_{D0} in modern devices is about 20nA, the leakage current is dependent on the difference between V_T and V_{GS}. Normally V_T has been designed for 0.7V, so that if V_{GS}=0, then the exponential result is a tiny fraction of 20nA.

Example

Calculate the DC current that will flow for a gate voltage of -1 volt and a V_{DS} of -500 mV. First calculate assuming no channel length modulator and then with this taken into account. The bulk material is N type, the insulator is SiO_2 and the gate is polysilicon. The width of the channel is 20μm with a length of 1.8 μm. Interface and insulator charge can be considered to be negligible.

Bulk material in N type so all are equations with voltages will have minus signs in front of every voltage.

N_D	$= 10^{21}$ /cm^3
n_i^2	$= 2.25.10^{20}$ /cm^3
$\varepsilon_r(SiO_2)$	= 3.9
$\varepsilon_r(Si)$	= 11.8
μ_n(channel)	= 1000 cm^2/Vs

oxide thickness

= 0.001 μm (0.1 10^{-6} cm)

Φ_{MS} = 6.0 V

Example

Example

Example

Example

Example

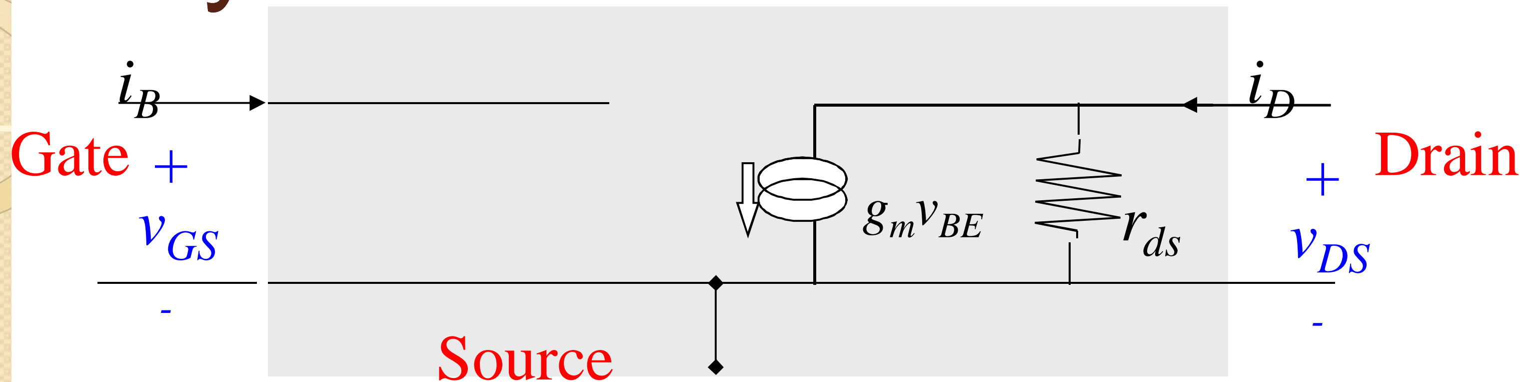
Example

Example

Example

Example

Hybrid π Model



This is a simplified version of the Hybrid- π model presented in GE204: Analog Electronics. The most obvious lack is any input resistance, this is because we have an insulated gate and there is no direct current path.

The two remaining important parameters are $\mathbf{g_m}$ and $\mathbf{r_{ds}}$. These can be obtained from the large signal current equation for I_D

Small Signal Modelling

Now let's consider the small signal parameter g_m from the hybrid π model that was covered in Analog Electronics.

g_m is defined as the small signal partial derivative of output current with respect to input voltage, ie

Therefore

Small Signal Modelling

Now for the saturated, or active region, as before it is possible to determine g_m. Except now that there is no drain voltage term, we need to do it with respect to V_{GS}

Small Signal Modelling

Parameters, where possible, are commonly defined in MOSFETs in terms of the DC drain current. This is because we normally have control over the drain current.

So

but

therefore

Small Signal Modelling

Another useful variation of g_m, I_D and V_{GS}
is quickly found again from the original equation for g_m.

but

therefore

Small Signal Modelling

The remaining parameter in this simple model is r_{ds}. Now r_{ds} is defined as the partial derivative of output voltage to output current (as it's small signal).

If we look at the saturated (active) region where the FET is most commonly used.

therefore

Small Signal Modelling

Now, if λ is small, ie a small second order relationship between I_{Dsat} and V_{DS}, then we can make a simplifying assumption, that the I_{Dsat} current is constant and equal to I_{dsat}.

where

and

Small Signal Modelling

Before leaving this, it is worth noting that if λ is small, and it is normally so, the r_{ds} is the reciprocal of a small current by a very small number.

I_D is normally sub-milliAmp, and λ is normally below 0.1, thus r_{ds} is at least tens of thousands of ohms.

However in modern devices, as the minimum length of the devices decrease, then λ will increase and this means that the variation in current on V_{DS} in saturation will **increase** and the small signal output resistance **decreases**. The transistor gets less and less ideal the smaller we compact it.

Note on Notation

It is quite common to express the voltage V_{GS}-V_T as the *effective* voltage or the *overdrive* voltage. The difference is the voltage that you are using to drive the gate once the gate has been turned on.

On the following pages, the equations will be expressed with V_{eff} rather than the more cumbersome V_{GS}-V_T.

Note on Notation

$$V_T = \Phi_S + \Phi_{MS} - \frac{Q_D}{C_i} - \frac{Q_i}{C_i}$$

$$\Phi_F = \frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right)$$

$$Q_D = -2\sqrt{qN_A \varepsilon \Phi_F}$$

$$W_{\max} = \sqrt{\frac{4\varepsilon_{semi}\Phi_F}{qN_A}}$$

Note on Notation

$$V_{eff} = (V_{GS} - V_T)$$

$$I_D = \frac{W}{L}\mu_n C_{ox}(V_{eff} V_{DS})$$ Very low V_{ds}

$$I_D = \frac{W}{L}\mu_n C_{ox}\left(V_{eff} V_{DS} - \tfrac{1}{2} V_{DS}^2\right)$$ Triode region

$$I_{D(sat)} = \frac{W}{L}\frac{\mu_n C_{ox}}{2}(V_{eff})^2$$ Saturation Region

$$k_{ds} = \sqrt{\frac{2\varepsilon_i}{qN_A}}$$

$$\lambda = \frac{k_{ds}}{2L\sqrt{-[V_{DS} - (V_{GS} - V_T) + \Phi_0]}}$$

$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda(V_{DS} - (V_{GS} - V_T)))$$

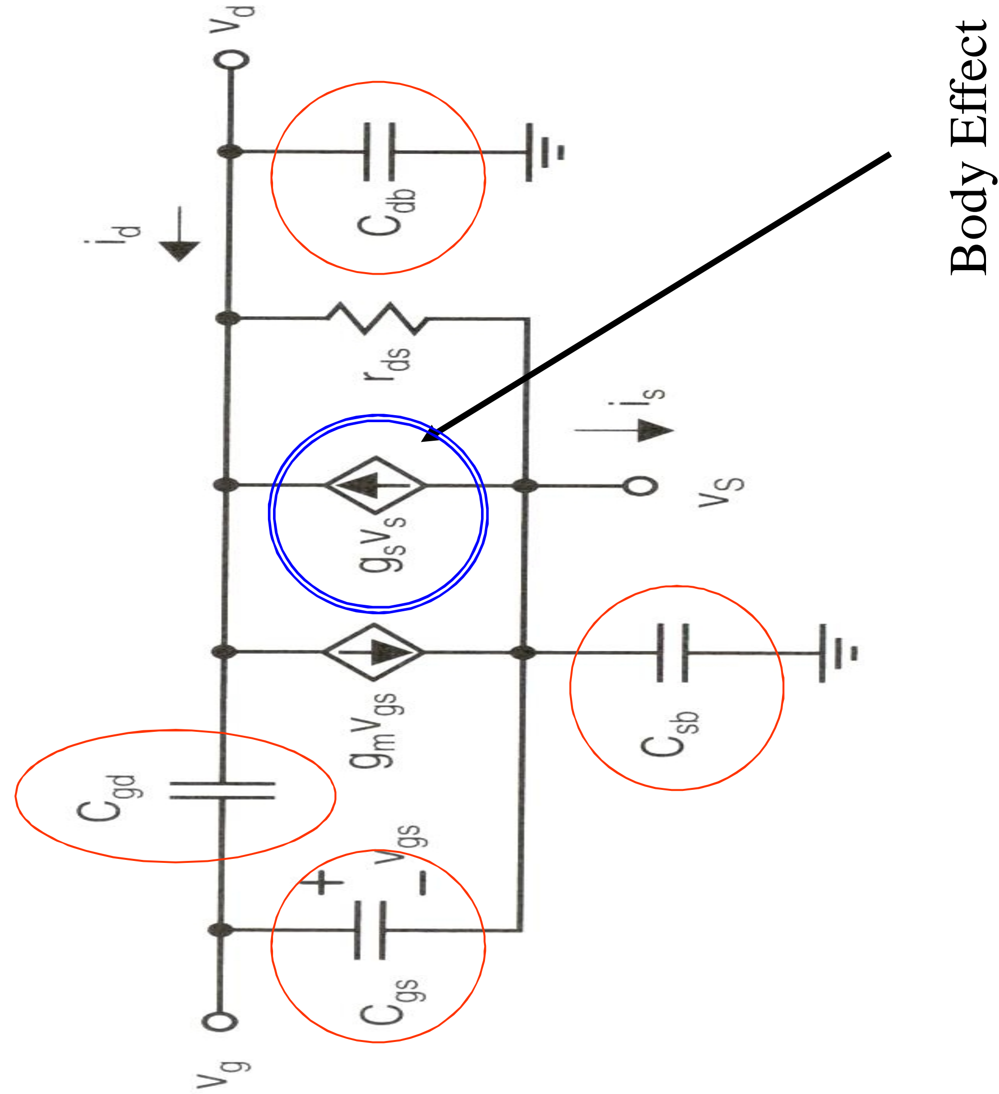
Note on Notation

$$g_m = \frac{W}{L}\mu_n C_{ox}\left(V_{GS} - V_T\right)$$

$$g_{ds} = \lambda \frac{\mu_n C_{ox}}{2}\frac{W}{L}\left(V_{GS} - V_T\right)^2$$

$$r_{ds} = \frac{1}{g_{ds}} \approx \frac{1}{\lambda I_{Dsat}}$$

Extended Small Signal Model

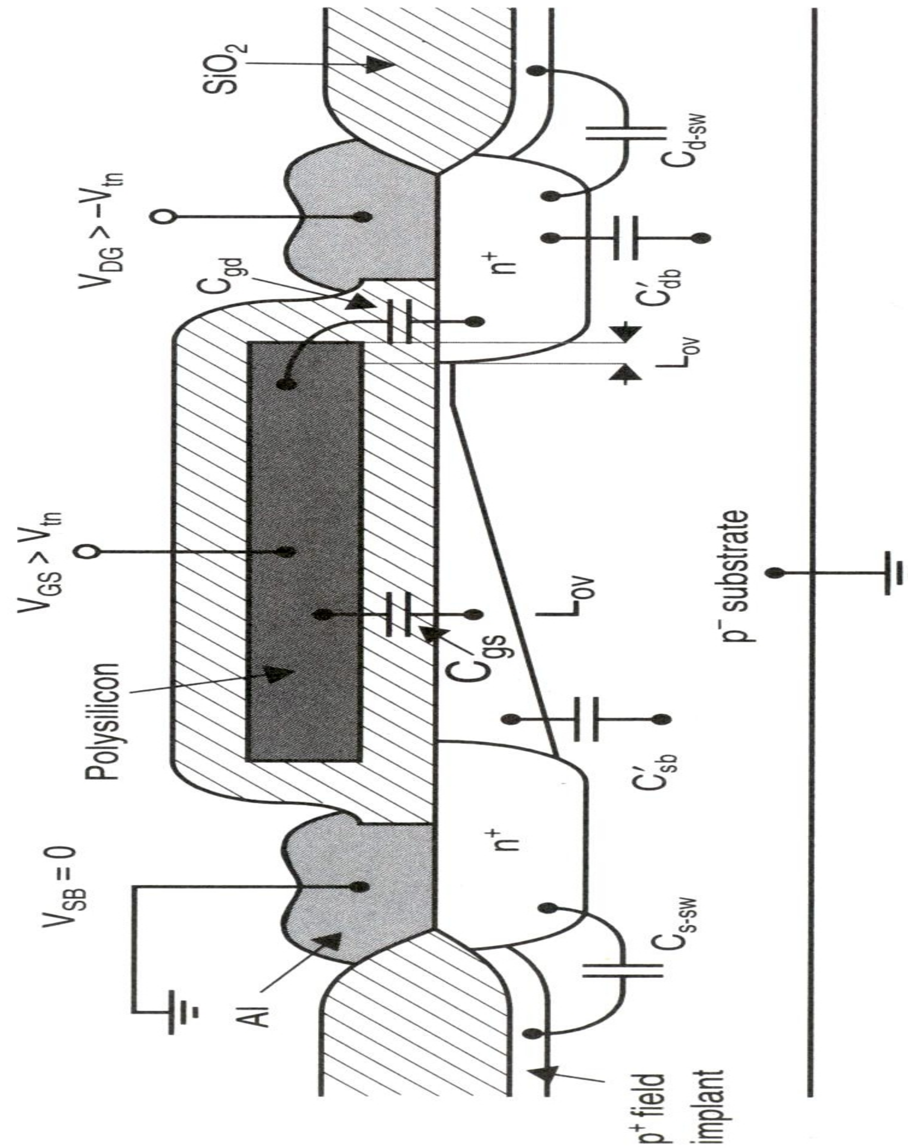


MOSFET Capacitors

V_s - this is the substrate voltage and it has an effect both in the small signal and large signal. However as we generally type the substrate to AC earth, this term is normally excluded

However the capacitors in the MOSFET generally arise from the construction of the MOSFET. There are capacitances due to the depletion regions and inversion regions but these are not dominant.

MOSFET Capacitors



MOSFET Capacitors

C_{GS}

Gate-Source capacitance, connecting the gate to effectively the ground.

At high frequencies this is a major factor in loss of gain. It is due to the change in channel charge with change in gate voltage. This is the largest capacitor in the MOSFET and is empirically defined as

MOSFET Capacitors

C_{GD}

Gate-drain capacitance, connecting the gate to the drain. This is an important capacitance as in high voltage gain applications the size of the capacitor is scaled by the gain of the system. It is often also called the *Miller Capacitance*. This capacitance is primarily due to the overlap between the gate and the drain and any fringing capacitance.

MOSFET Capacitors

C_{SB}

Source-Bulk(substrate) capacitance, connecting the gate to the bulk.

This is due to the reverse biased PN junction caused by the source-bulk PN junction depletion region. It is the second largest capacitance but normally is not critically as both the source and bulk are *normally* connected to AC earth. This capacitance is normally stated including the channel’s depletion region capacitance.

MOSFET Capacitors

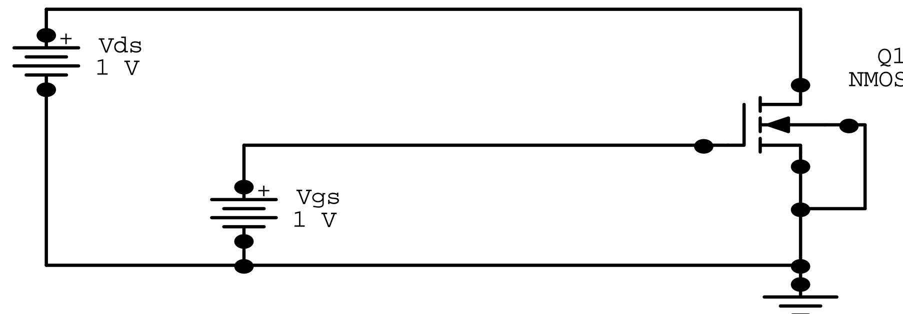
C_{DB}

Drain-Bulk(substrate) capacitance, connecting the gate to the bulk.

This is due to the reverse biased PN junction caused by the drain-bulk PN junction depletion region. It is smaller than the source-bulk capacitance as it does not include the channel area. Again not normally an issue.

Example

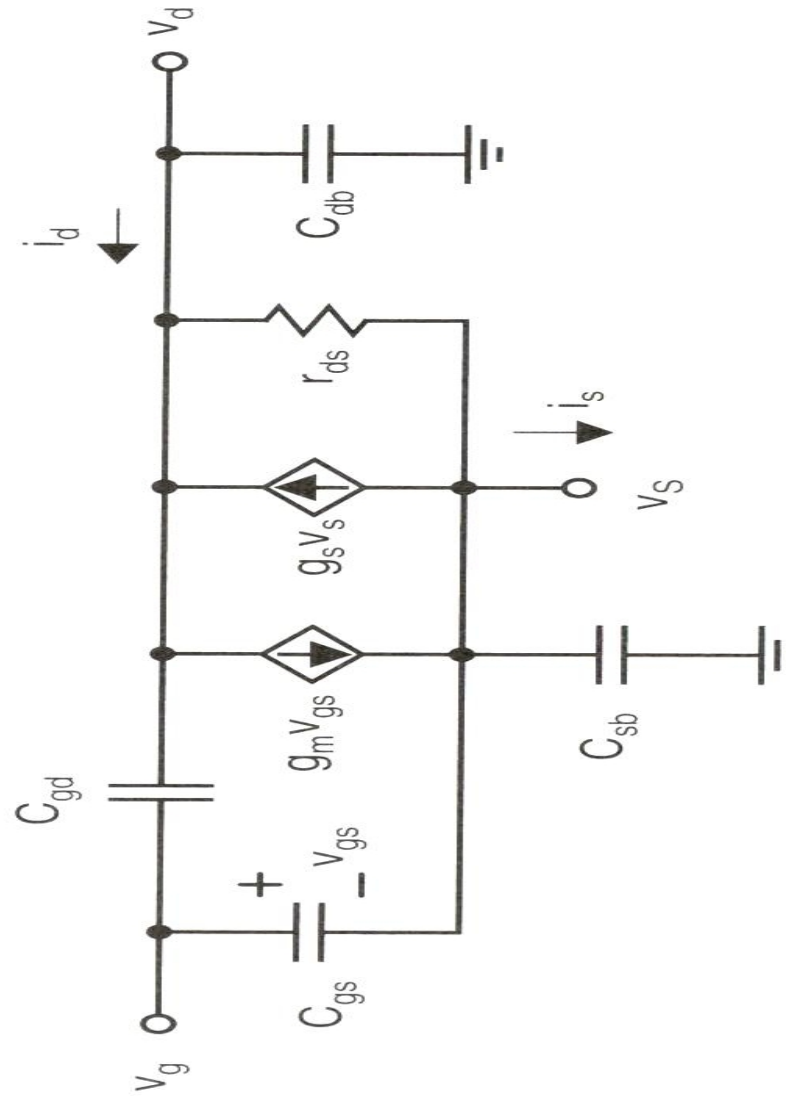
A N-Channel (P type substrate) MOSFET (NMOS) is wired up as shown in the diagram. Draw the small signal equivalent circuit, give values for the values of the key parameters given the following conditions, and commenting on any that you decide can be ignored.



N_A	$= 10^{17}$ /cm^3	V_T	= 0.8 V
W	= 180 μm	L	= 1.8 μm
$\mu_n C_{ox}$	= 92 μA/V^2	λ	= 0.05
C_{ox}	= 0.01 μF	I_D	= 100 μA

Drain-gate L_{ov} = 0.5 μm   Source-gate L_{ov} = 0.9 μm

Example



Example

The equations for the other terms are as follows

$$g_m = \frac{W}{L}\mu_n C_{ox}(V_{GS} - V_T)$$

$$g_{ds} = \lambda \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{GS} - V_T)^2$$

$$r_{ds} = \frac{1}{g_{ds}}$$

$$C_{GD} = WL_{ov}C_{ox}$$

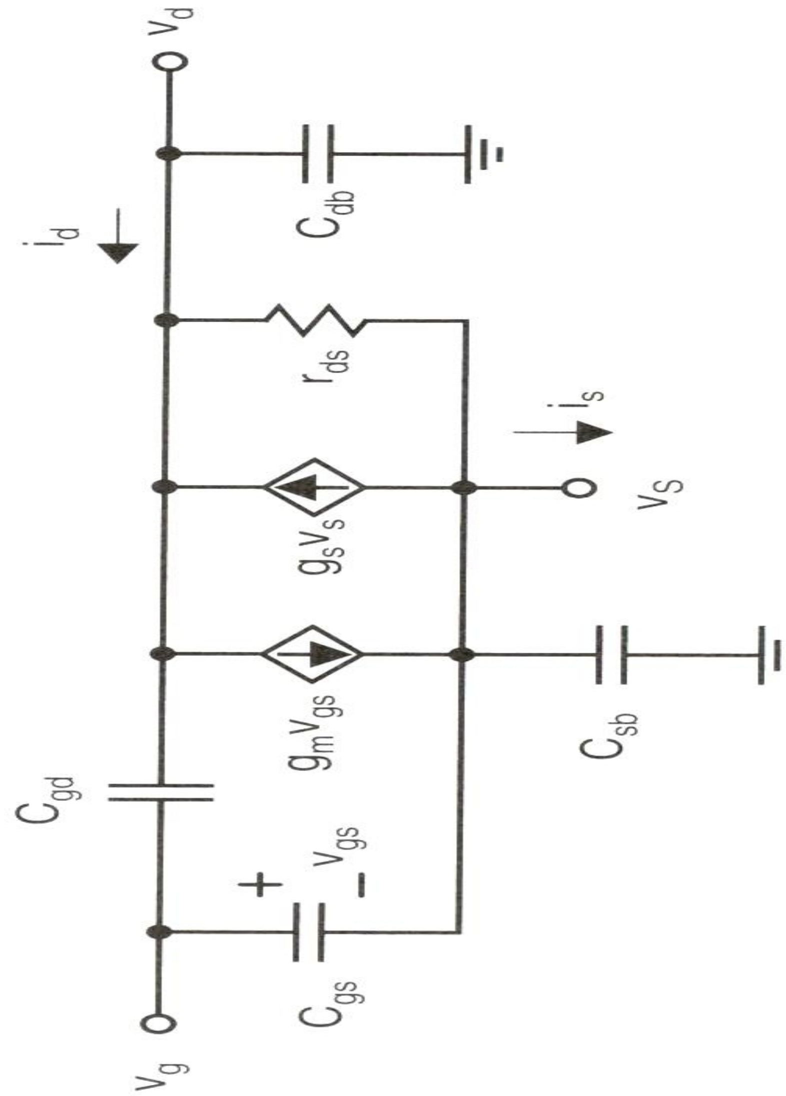
$$C_{GS} = \frac{2}{3}WLC_{ox} + WL_{ov}C_{ox}$$

The equations for the other terms are as follows

The equations for the other terms are as follows

The equations for the other terms are as follows

Example



Example

Given the following diagram of the construction of a MOSFET, indicate the major, if any, capacitances, diodes, resistances and inductances that may exist in the MOSFET. Indicate what polarity the voltages on the drain, gate, source and bulk connections should be.

Example