Part 5: MOSFET

MOSFET Operating principles

MOSFET Characteristics

MOSFET equivalent circuits

MOSFET

The **Metal Oxide Semiconductor Field Effect Transistor** is now the most common device for amplifying or switching signals.

It has largely replaced bipolar junction transistors, which were once favoured for these applications. However bipolar still plays an important, if minor role, in audio and radio-frequency amplification

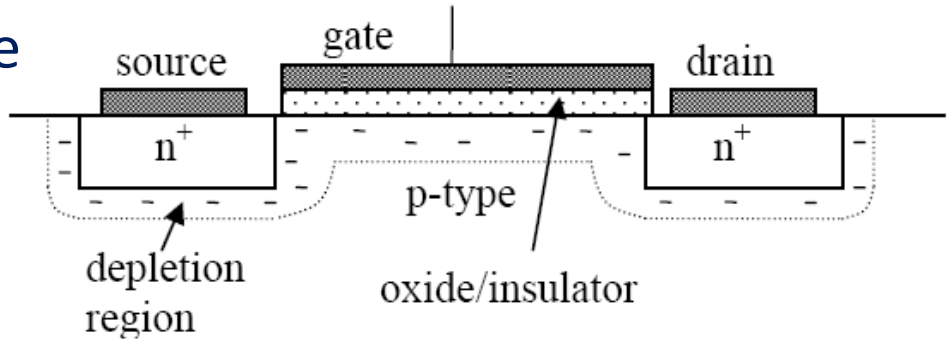
As the name suggests, it relies on a metal-oxide-semiconductor gate to control the flow of electrons (or holes).

One of the great successes of MOSFET technology has been its scalability to smaller and smaller device (gate) sizes.

The combination of n- and p-type MOSFETs is used in **Complementary Metal Oxide Semiconductor** FET (CMOS) technology, the dominant approach for integrated circuits

MOSFET Principles

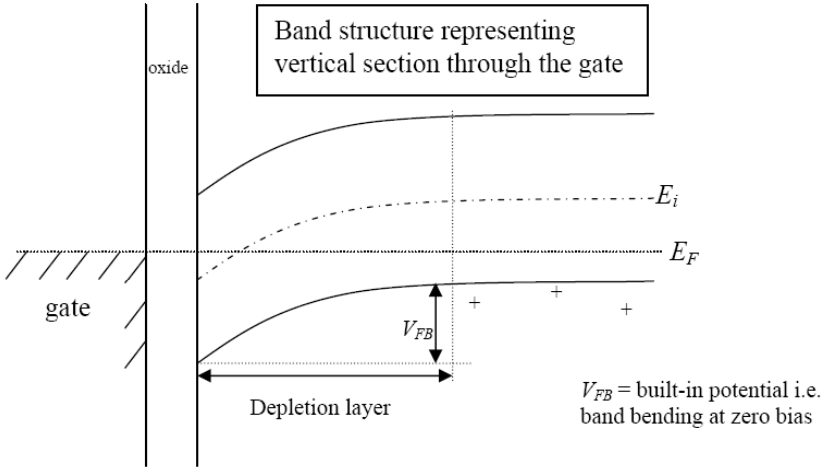
Consider p-substrate with n-type source and drain implants and an oxide insulator on the gate



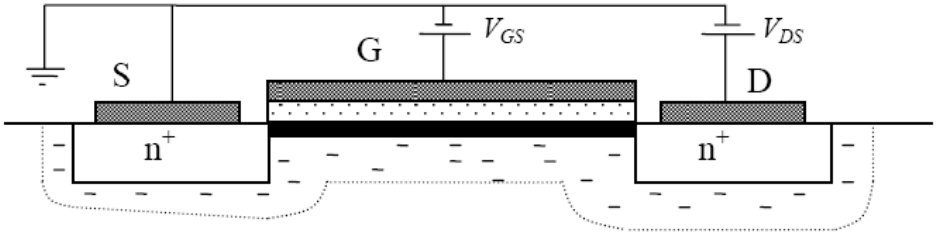
We define V_{GS} (or just V_G) as the gate source voltage and V_{DS} (or V_D) as the drain source voltage

<u>$V_{GS} \geq 0$ (no channel)</u>

A small positive or zero gate bias induces a depletion region in p-type material. There is no mobile charge between the source and drain therefore device is 'off'.



MOSFET Principles

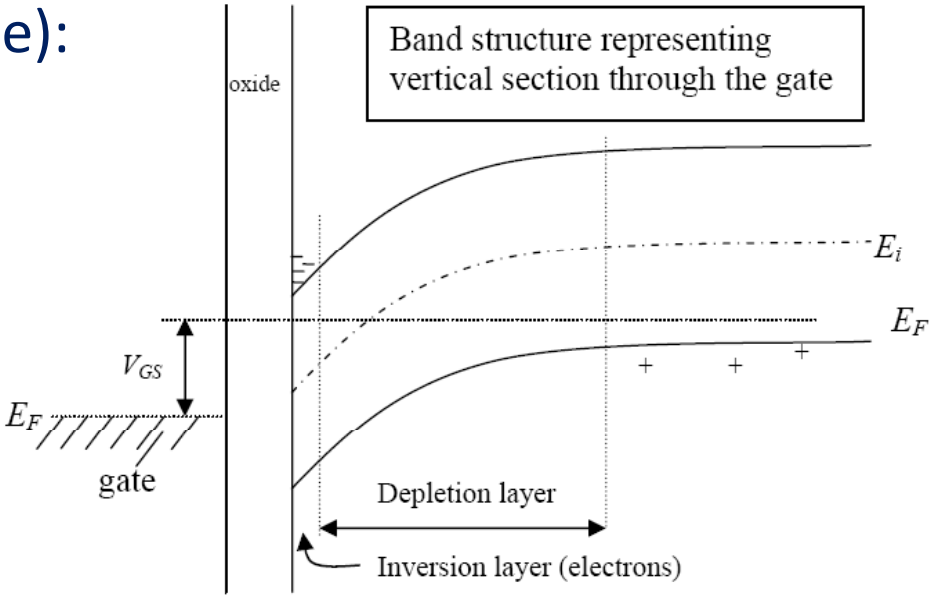


$V_{DS} << V_{GS} < V_T$

As more positive charge is placed on the gate the depletion into the p-type region increases. Eventually, when $V_{GS} \cong V_T$ (the threshold voltage) electrons are formed under the gate

The Fermi level moves towards the conduction band (looks like n-type): this is called an **inversion layer**.

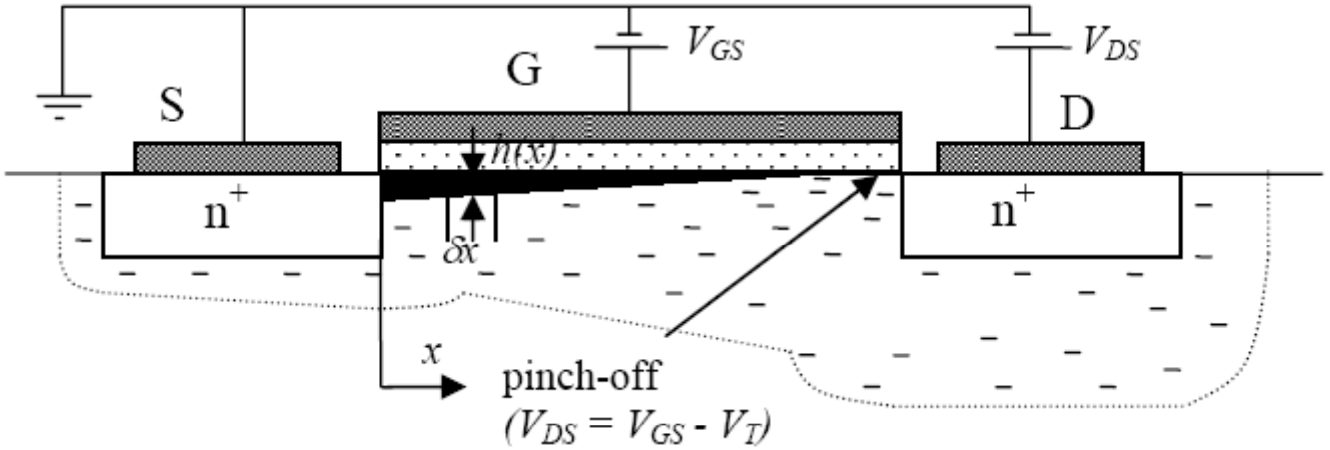
The source and drain are now connected by a mobile electron channel and the device turns on.



MOSFET Principles

$V_{DS} \cong V_{GS} - V_T$

As V_{DS} approaches $V_{GS} - V_T$ the net voltage at the drain end of the channel falls below V_T and the channel 'pinches-off' at the drain end.



h(x) – channel thickness at x
δx – distance element at position x

Can do an analysis similar to that for the MESFET

MOSFET Principles

Resistance of volume element δx $$R_{dx} = \frac{\rho \partial x}{Zh(x)}$$ **(1)**

where Z = gate width and ρ = resistivity of channel

Drain current, I_D, is continuous and is related to the voltage drop, ∂V_x, across the element $$I_d = \frac{\partial V_x}{R_{\partial x}} = \frac{Zh(x)}{\rho}\frac{\partial V_x}{\partial x}$$ **(2)**

Inverse resistivity is just the conductivity and conductivity is given by the mobile charge multiplies by the mobility

$$\frac{1}{\rho} = \sigma = n_{dx} q \mu$$ (from basic semiconductor notes)

where n_{dx} is the no of electrons in a element δx and q is the electronic charge

MOSFET Principles

Substitute for resistivity $I_d = \frac{\partial V_x}{R_{\partial x}} = \frac{Zh(x)}{\rho}\frac{\partial V_x}{\partial x} \Rightarrow Zh(x)n_{dx}q\mu\frac{\partial V_x}{\partial x} \Rightarrow ZQ_n\mu\frac{\partial V_x}{\partial x}$

where Q_n is the charge per unit area

This charge gives rise to a capacitance $C = Q/V$ $Q_n = C_{OX}[V_{GS} - V_T - V_x]$

where C_{OX} = gate capacitance per unit area

$V_x = 0$ at the source = V_{DS} at the drain. Set up to integrate this over the whole channel length (to X=L and V_x=V_{DS})

$$I_d.\partial x = Z\mu Q_n.\partial V_x = Z\mu C_{ox}[V_{GS} - V_T - V_x].\partial V_x$$

$$I_D\int_0^L \partial x = Z\mu\int_0^{V_{DS}} Q_n\partial V_x = Z\mu C_{OX}\int_0^{V_{DS}}((V_{GS} - V_T) - V_x)\partial V_x \qquad \textbf{(3)}$$

MOSFET Principles

gives

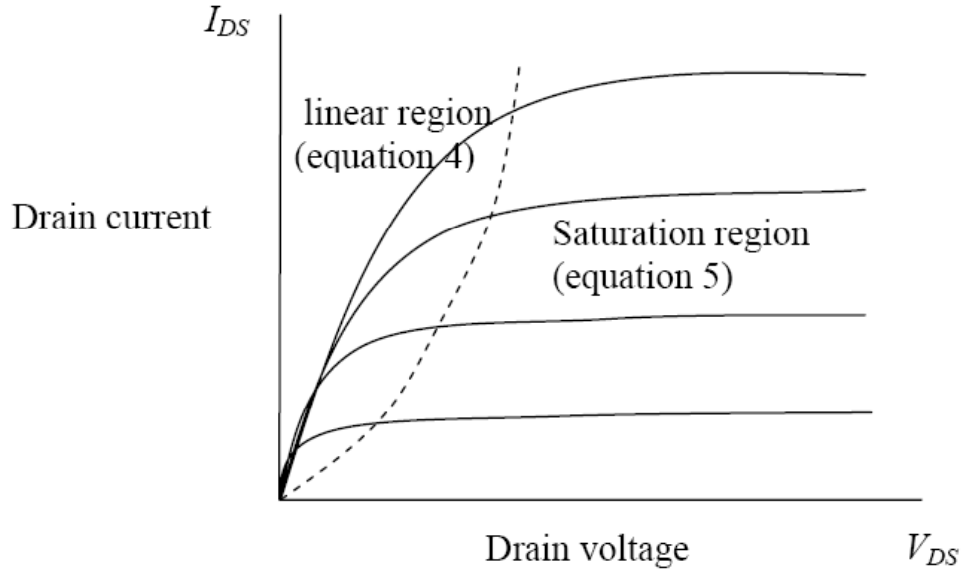
$$I_D = \frac{Z\mu C_{OX}}{L}\left[(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right] \qquad \textbf{(4)}$$

This equation is valid only up to the pinch off voltage.

At pinch off $V_{x=L} = V_{DS} = V_{GS} - V_T$ ($Q_n = 0$ at $x = L$)

$$I_D = \frac{Z\mu C_{OX}}{2L}\left[V_{GS} - V_T\right]^2 \qquad \textbf{(5)}$$

Plotting equation (4) gives the output characteristics of the MOSFET (notice- similar to the MESFET)



MOSFET Principles

In saturation (normal operating range) the transconductance (gain) is given from equation (5)

$$g_m = \frac{\partial I_{DS}(sat)}{\partial V_{GS}} = \frac{Z\mu C_{OX}}{L}\left[V_{GS} - V_T\right]$$

This implies that short gate lengths and high mobility will result in high g_m *and* I_{DS}.

As before, the current gain = output current/input current

Output current = $g_m V_{GS}$

Input current = CV, but this is small signal ac so its ωCV
= $C_{OX} V_{GS} ZL$, since C_{OX} is the capacitance per unit area

MOSFET Principles

Current gain = $\frac{g_m V_{GS}}{V_{GS}\omega C_{OX} ZL}$

Gain =1 for f = f_T (from definition of f_T). Substitute for g_m from above

$$f_T = \frac{g_m}{2\pi C_{OX} ZL} = \frac{Z\mu C_{OX}}{2\pi C_{OX} ZL}(V_{GS} - V_T) = \frac{\mu}{2\pi L^2}(V_{GS} - V_T)$$

So short gate length and high mobility are important for high speed devices. Velocity saturation will reduce g_m and f_T expectations from the above equations.

The n-channel MOSFET as described up to now is a so-called an **enhancement mode** MOSFET i.e. Device is normally off, and an applied positive gate bias switches the device on.

MOSFET

There are also **depletion mode** devices which are normally on and in this case an applied negative gate voltage switches the device off. This needs a special design- will be discussed later

In addition, there are **p-channel devices** (holes form the conducting channel) which, when combined with n-channel devices form CMOS (complementary MOS). The p-channel devices require the opposite gate bias (negative) to switch them on.

Threshold Voltage, V_T, for MOSFETs

This is an important parameter for the design of integrated circuits; will determine the power supply needs, power dissipation etc

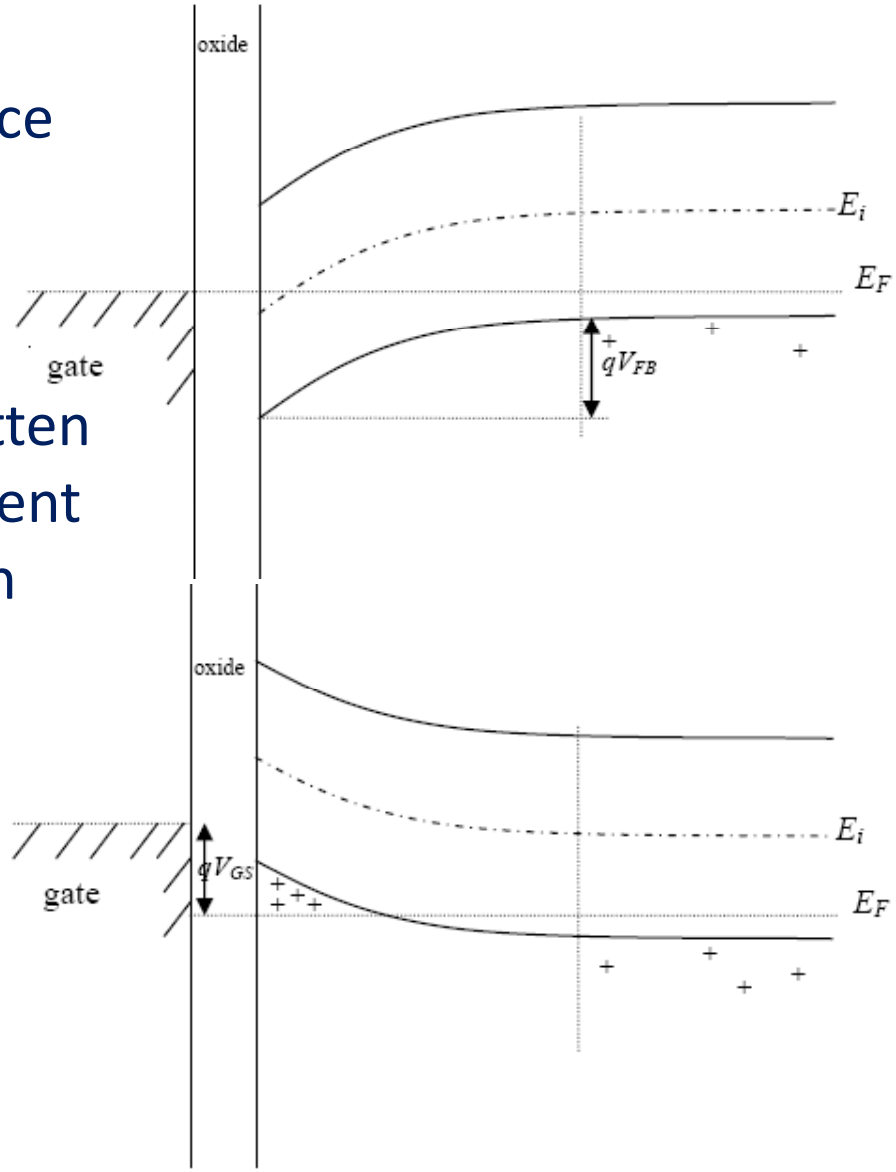
Consider first the situation in equilibrium (zero bias)

MOSFET

Flat-band voltage V_{FB} is the difference in work functions = $q(\phi_M - \phi_S)$

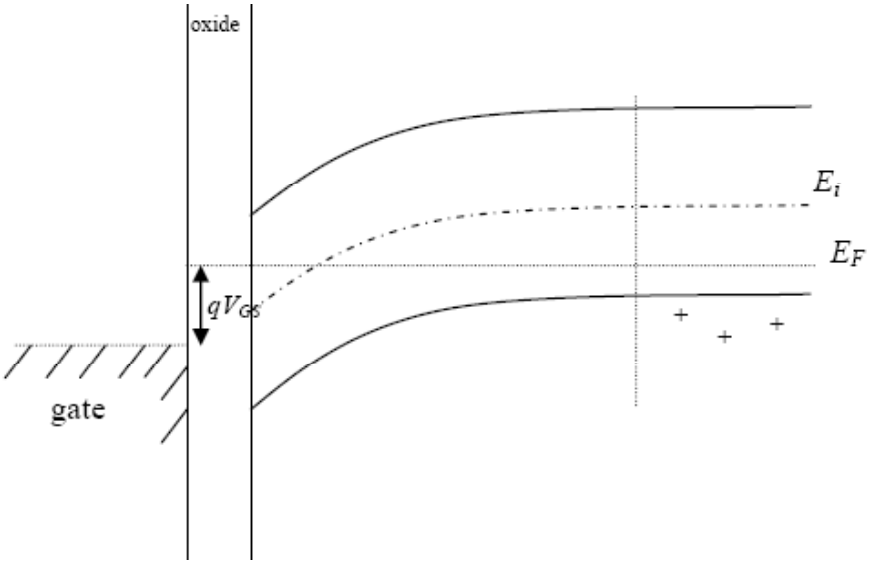
This is the potential required to flatten the bands at the interface – equivalent to the built-in potential in a junction device

Accumulation
(V_{GS} negative)



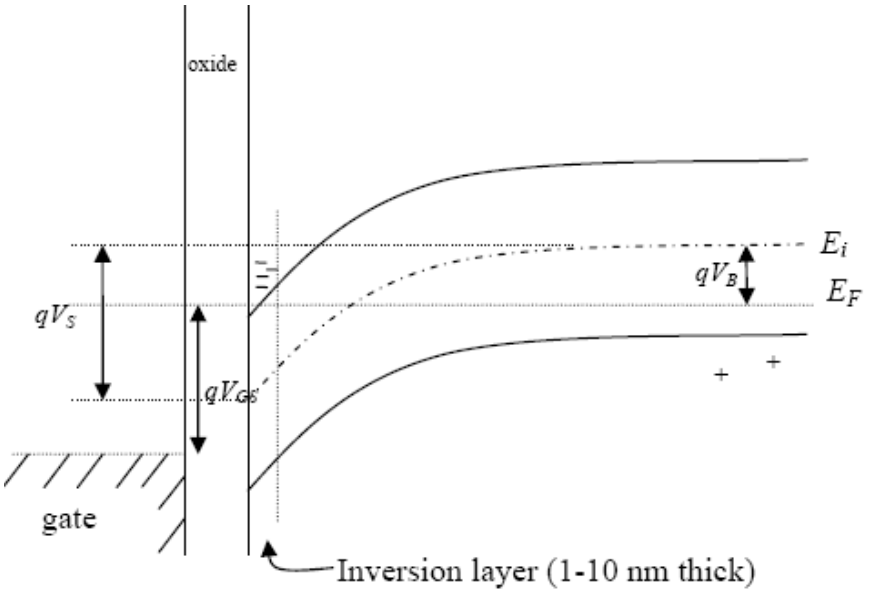
MOSFET

Depletion
(V_{GS} positive)



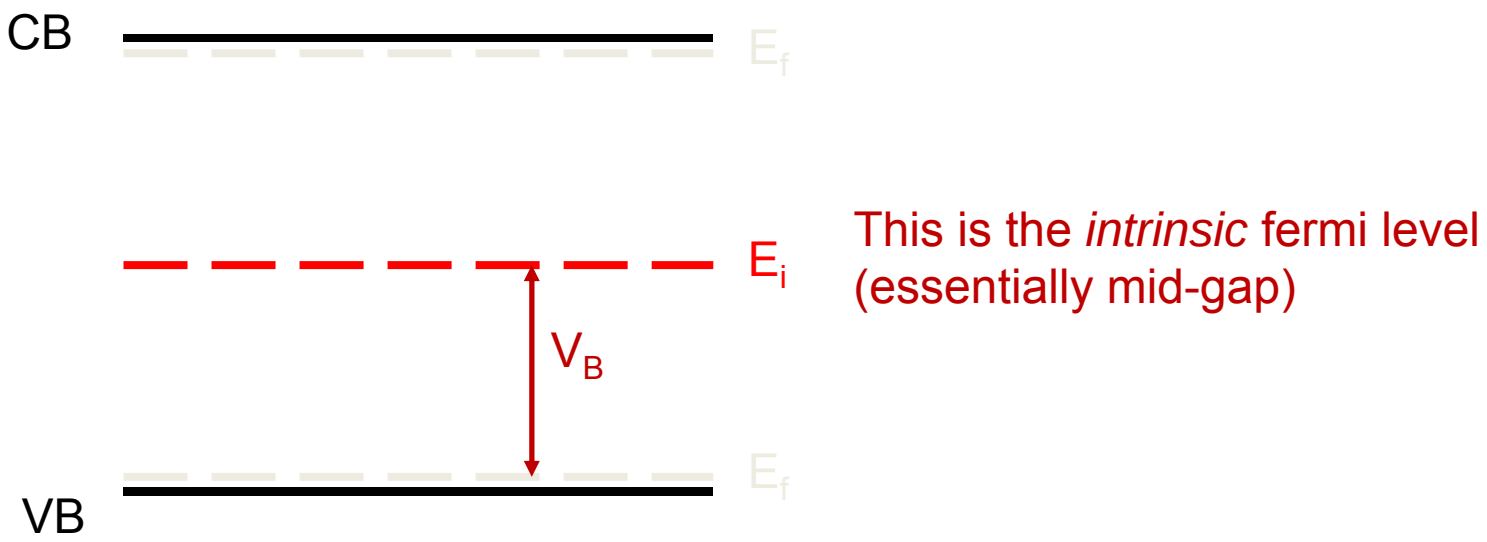
Inversion
($V_{GS} > V_T$)

V_s is the surface potential describing how much the intrinsic Fermi level is bent at the surface



MOSFET

What we are doing here is changing the surface type original p-type material, making it appear n- type the surface



Started with p-type material - Fermi level is down towards the valance band. To 'invert this' (convert this to n-type) we need to move the Fermi level to be close to the conduction band (a voltage of 2x V_B). So V_s needs to be $\geq -2V_B$

MOSFET

The gate voltage (usually measured with respect to the source) = V_{GS} will have to equal this voltage to get inversion.

There are however some other components

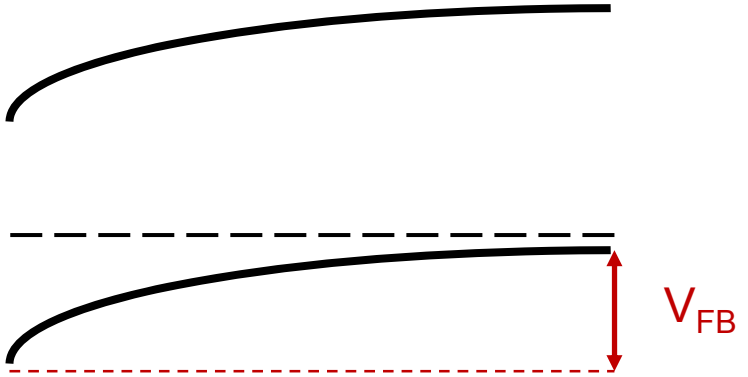
$V_{GS} = V_S + V_{OX} - V_{FB}$

When the device is at threshold
V_{GS}= the threshold voltage (V_T)

V_{ox} is voltage drop caused by the oxide (the oxide has charge on it and a capacitance C= Q/V so V_{ox}= Q/C. This increases V_T

V_{FB} is the existing band bending at equilibrium. This helps lower V_T

Initial band bending aids inversion



MOSFET

As V_{GS} increases (more positive) the depletion region increases to uncover more fixed negative acceptors – eventually inversion occurs, at which point the depletion region is fixed.

The inversion condition can be defined as when the surface electron concentration equals the background hole concentration i.e. it looks n-type (n_s=p).

Inversion condition $V_S = -2V_B$

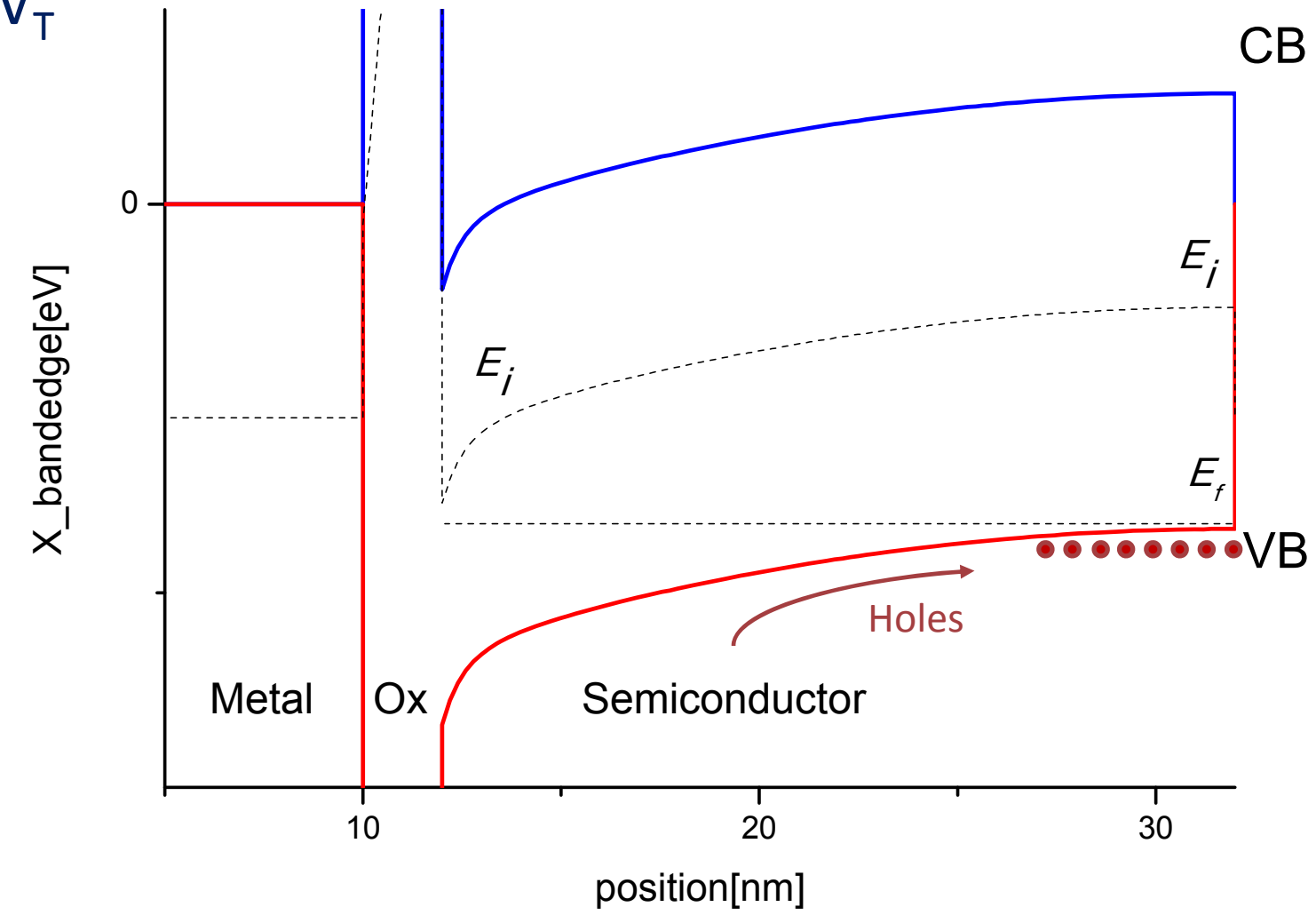
$$V_B = -\frac{kT}{q}\ln\frac{p}{n_i}$$

From basic semiconductor theory, e.g: see Streetman

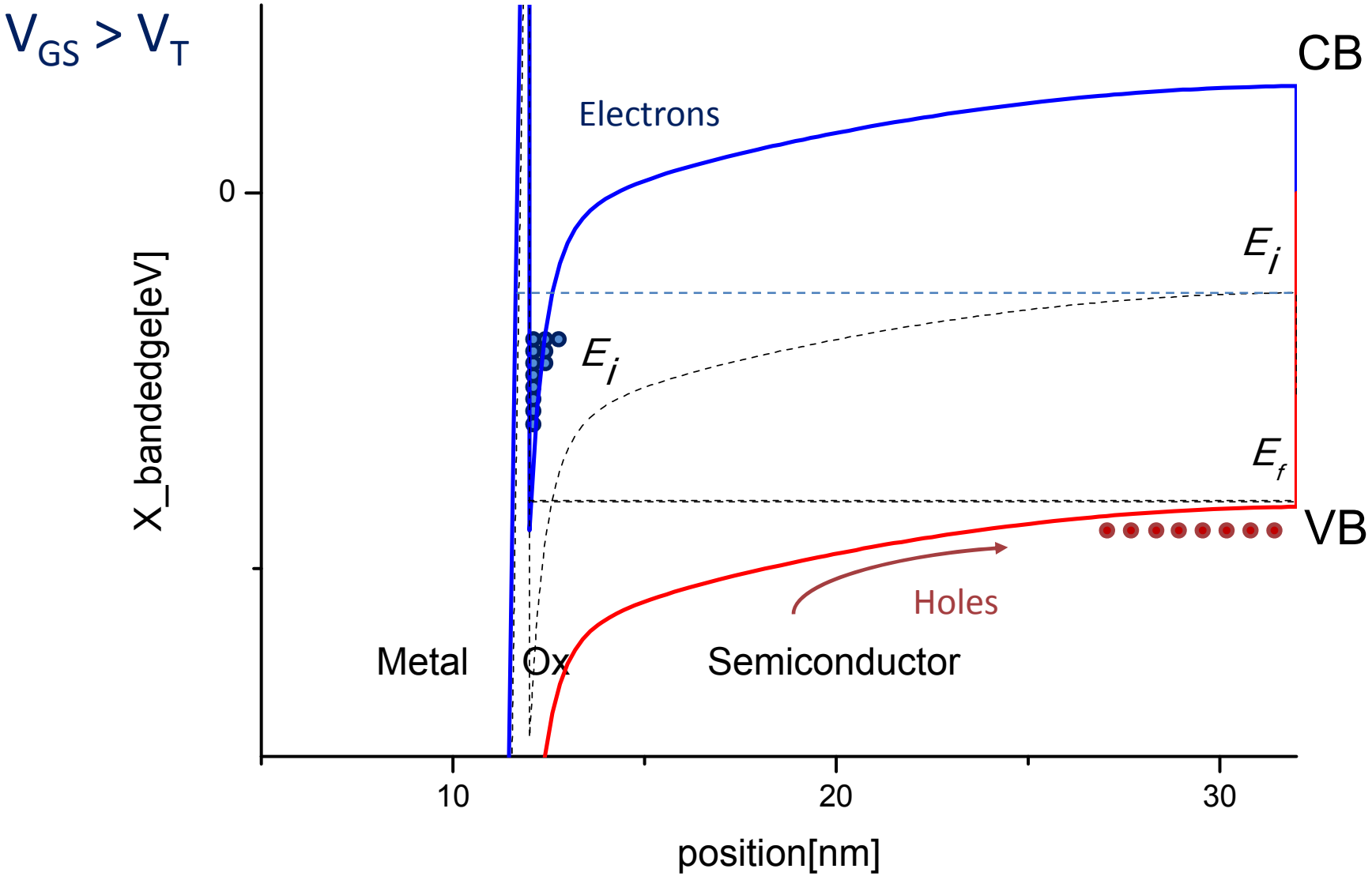
k = Boltzmann const. T = absolute temp. n_i = intrinsic carrier concentration N_A = channel acceptor concentration ($p \cong N_A$ at room temperature and above)

MOSFET

$V_{GS} < V_T$



MOSFET


$V_{GS} > V_T$
CB
Electrons
E_i
E_i
E_f
VB
Holes
Metal
Ox
Semiconductor
0
X_bandedge[eV]
10
20
30
position[nm]

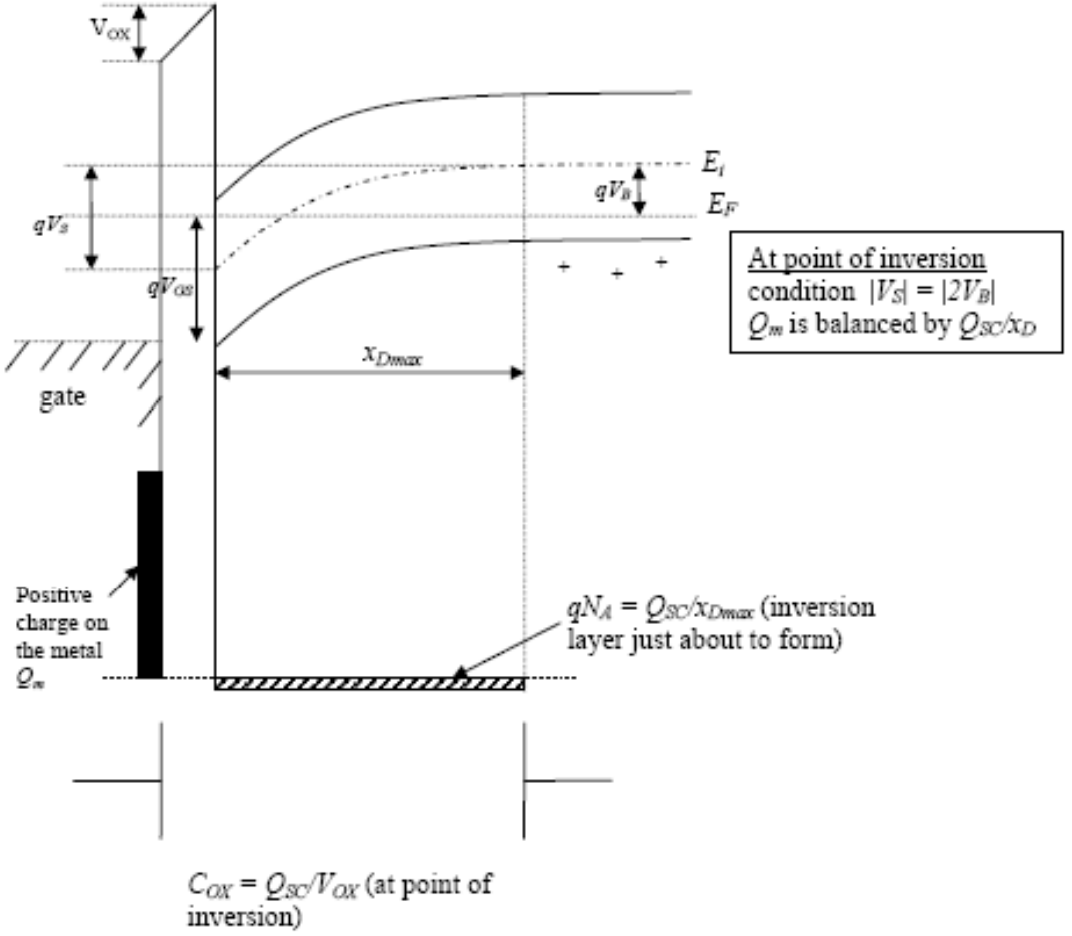
MOSFET

So $V_B \approx \frac{kT}{q} \ln \frac{N_A}{n_i}$ $V_S \text{ at inversion} = \frac{2kT}{q} \ln \frac{N_A}{n_i}$

The inversion condition depends on the amount of p-type doping

Can also look at this from a charge point of view . Consider the situation just before inversion

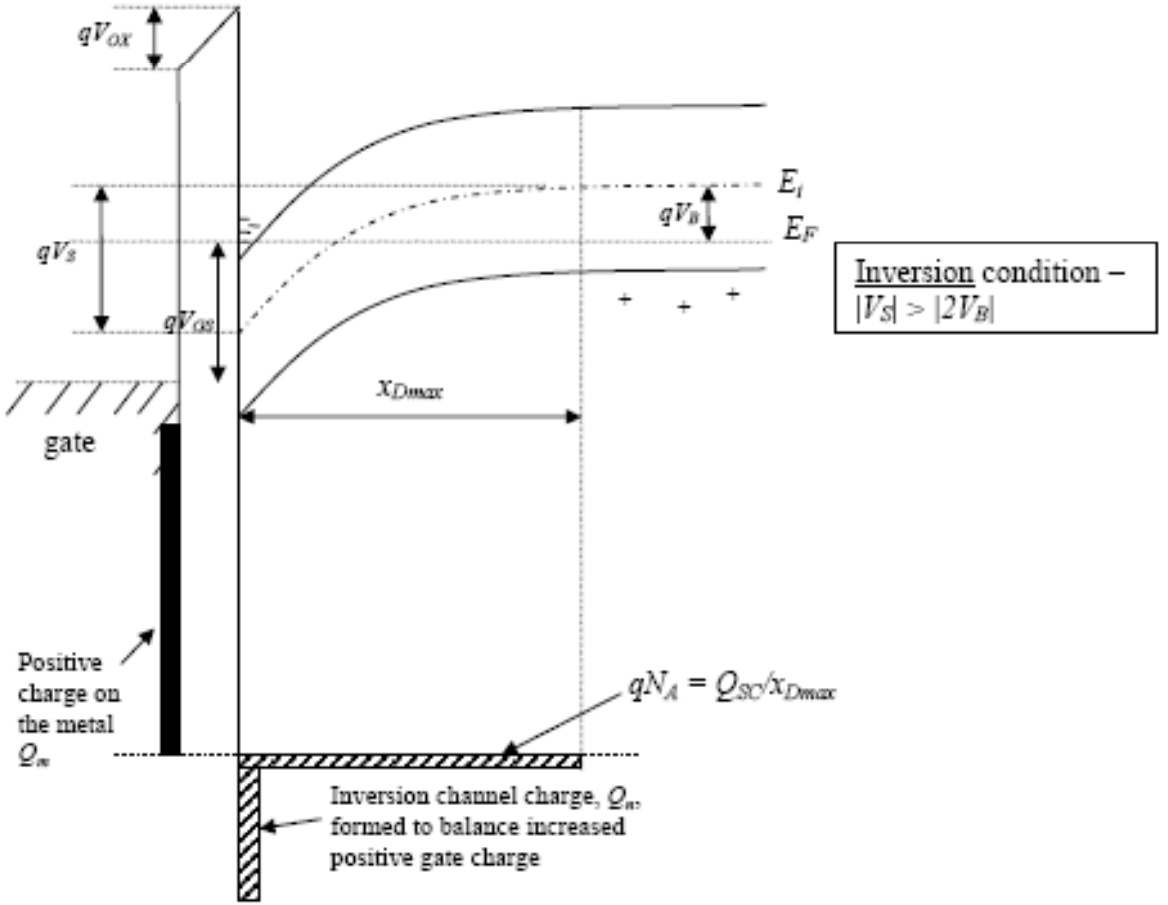
Here we have a positively charged metal gate (charge Q_n) on one side opposed by a negatively charged depletion of holes in the p-type semiconductor (called a space charge, (Q_{SC})



MOSFET

With more charge on the gate then inversion can occur. A second region of negative charge appears associated with electrons in the inversion layer.

This actually prevents the depletion region extending any further, so at the threshold X_D reaches a maximum



MOSFET

Now $Q_{SC} = qN_A x_{Dmax}$ coulomb s per unit area

(at point of inversion, x_D does not increase further)

$V_{OX} = Q_{SC}/C_{OX}$ C_{OX} = *oxide capacitance per unit area*

For X_{Dmax} calculate depletion width (see books, or previous courses)

$$x_{D\max} = \left(\frac{2\varepsilon_S V_S}{qN_A}\right)^{1/2} = \left(\frac{4\varepsilon_S V_B}{qN_A}\right)^{1/2}$$

($V_S = 2V_B$)

ε_S =semiconductor dielectric constant (Si)

$$\therefore V_{OX} = \frac{(4qN_A\varepsilon_S V_B)^{1/2}}{C_{OX}}$$

(substitute for x_{Dmax})

Threshold Voltage

$$V_T = -|V_{FB}| + 2|V_B| + \frac{(2q\varepsilon_S N_A |2V_B|)^{1/2}}{C_{OX}}$$

MOSFET

Substrate bias

Operation in a real circuit normally requires the substrate (p-type) to be negatively biased with respect to the source contact by an amount V_{SB}

In this case $V_S=-2V_B$ changes to $V_S= -2V_B -V_{SB}$

Threshold voltage changes to

$$V_T = -|V_{FB}| + 2|V_B| + \frac{\left(2q\varepsilon_S N_A\{-V_{SB} + 2|V_B|\}\right)^{1/2}}{C_{OX}}$$

This is useful because now V_T can now be also controlled by the substrate bias (useful for example if V_B large)

MOSFET

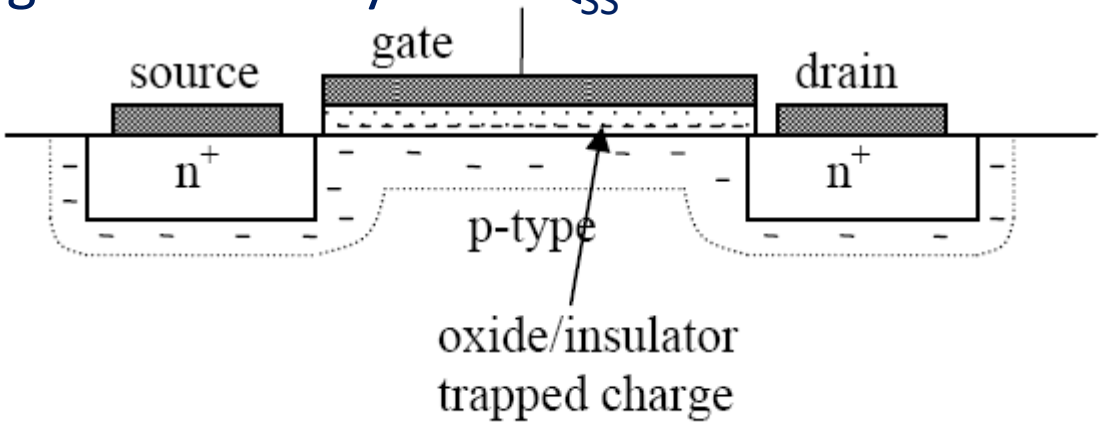
Positive V_{SB} means that the effective V_T is increased

Negative V_{SB} that the effective V_T is reduced

Interface (semiconductor surface)Charge

Any charge at the interface between the oxide and the semiconductor can influence the MOSFET properties (particularly V_T)

We normally give this the symbol Q_{SS}



Oxide creates an added potential

(Q_{SS} is the surface charge per unit area)

$$\Delta V_{OX} = \frac{-Q_{SS}}{C_{OX}}$$

MOSFET

Revise threshold voltage to take into account of these factors

$$V_T = -|V_{FB}| + 2|V_B| + \frac{\left(2q\varepsilon_S N_A \{-V_{SB} + 2|V_B|\}\right)^{1/2}}{C_{OX}} + \left|\frac{Q_{SS}}{C_{OX}}\right|$$

To achieve consistent and low V_T, controlling the interfacial oxide charge is important . Early MOS technologies showed large interfacial charges. The negative trapped electron charge led to large values of V_T implying large drive voltages

Reducing V_T is important for MOSFET technology. A low V_T means lower power dissipation. Power dissipation is a major factor in high density MOSFET circuits.

MOSFET

If V_T is low, we can use low gate voltages for normal operation. This is important because the gate oxide is somewhat 'leaky'- has a leakage current which scales with voltage. Low operating voltage gives lower gate leakage

CMOS processor

Gate length 45nm, Gate width 500nm.

Gate voltage= 2V Gate current =10^{-3}A.cm^{-2}

= 225 fA per gate

But you have 1x10^9 gates, so total = 225µA

There has been a progressive reduction in V_T and therefore the operating gate voltage

Latest Ultra low power (ULP) devices - V_T~ 0.2-0.3V

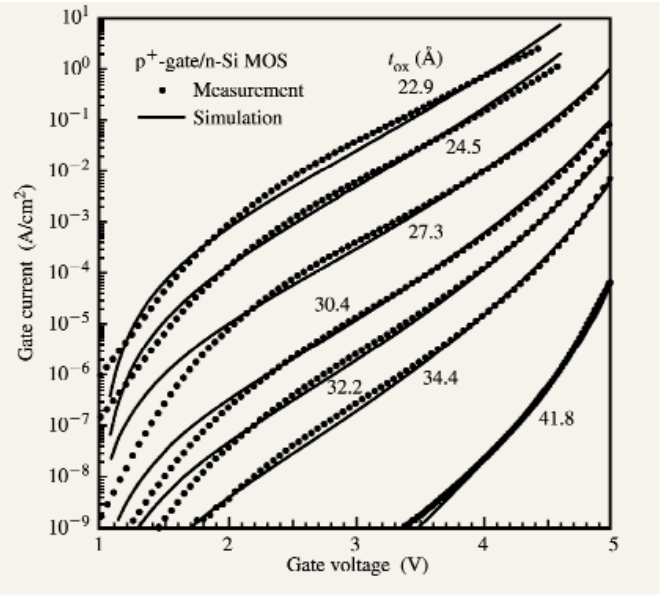
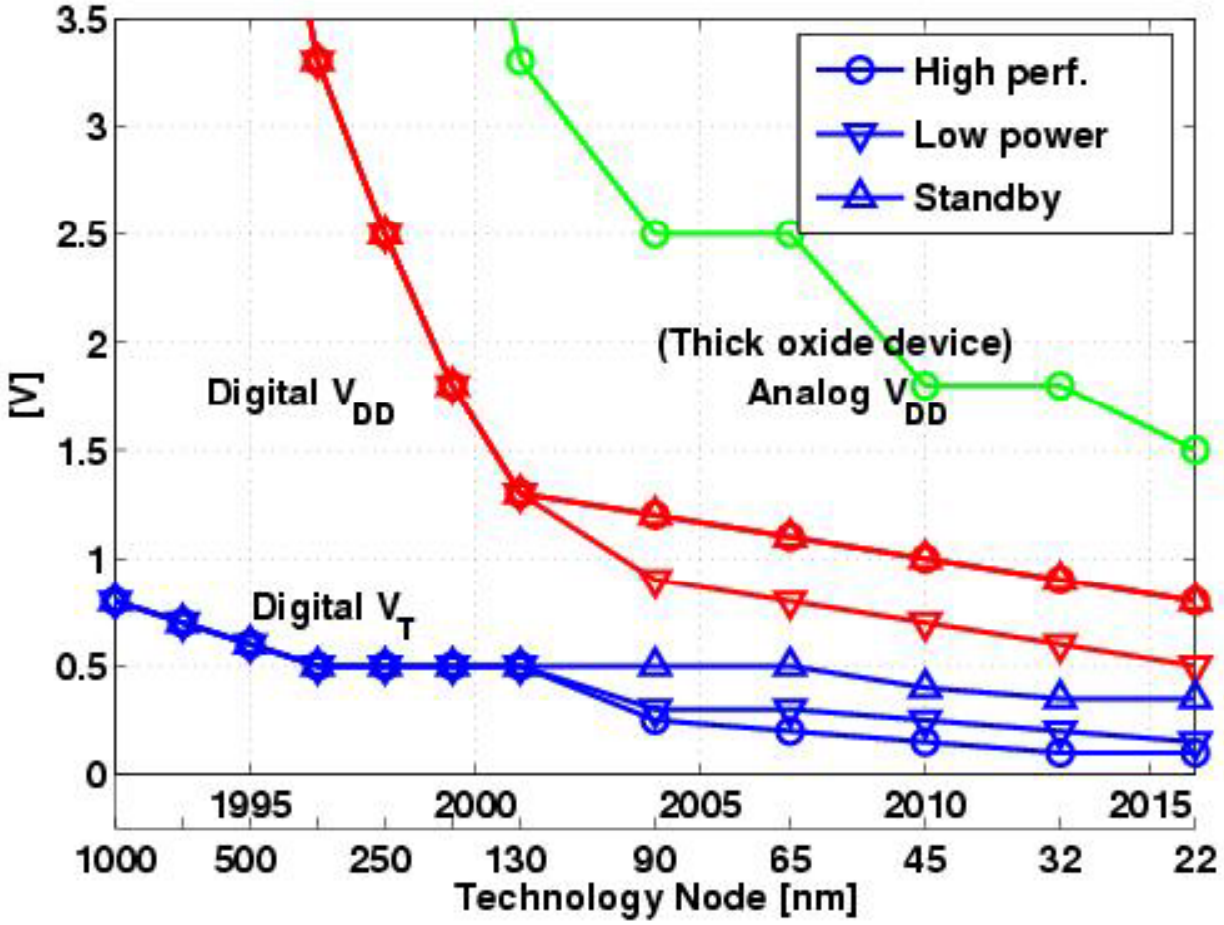


Figure 14

Measured and simulated I_G–V_G characteristics under accumulation conditions of p$^+$-gate/n-Si MOS devices with oxides ranging from 22.9 to 41.8 Å. The thickness is determined using the QM scheme.

MOSFET

The threshold voltage, V_T. , has dropped considerably over the years as consequence of device scaling, allowing a reduction in the supply voltage V_{DD}

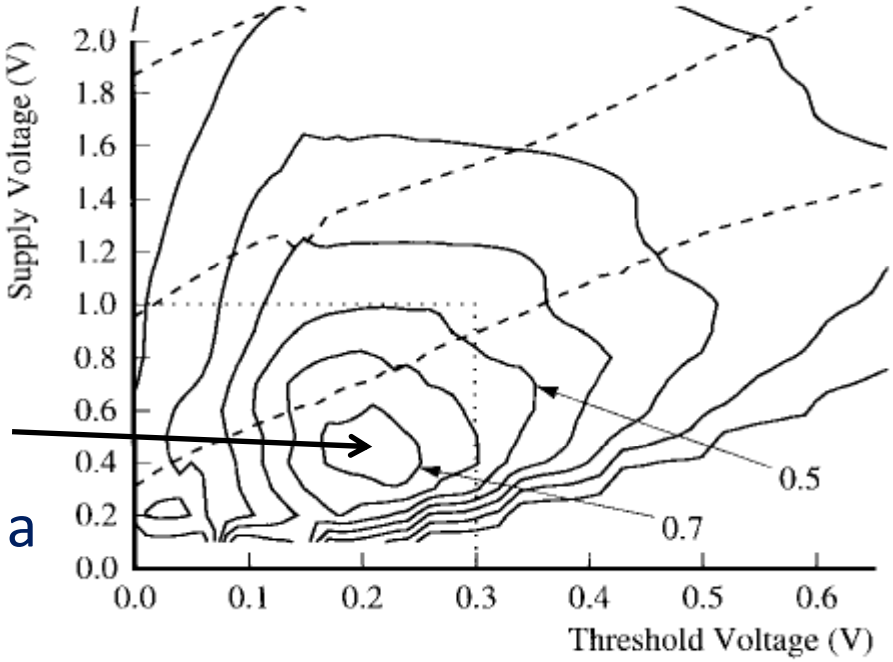


MOSFET

Note the supply voltage (V_{DD}) need to be 2-3x that of V_T

Actually interplay is quite complex: need to optimise

Optimum

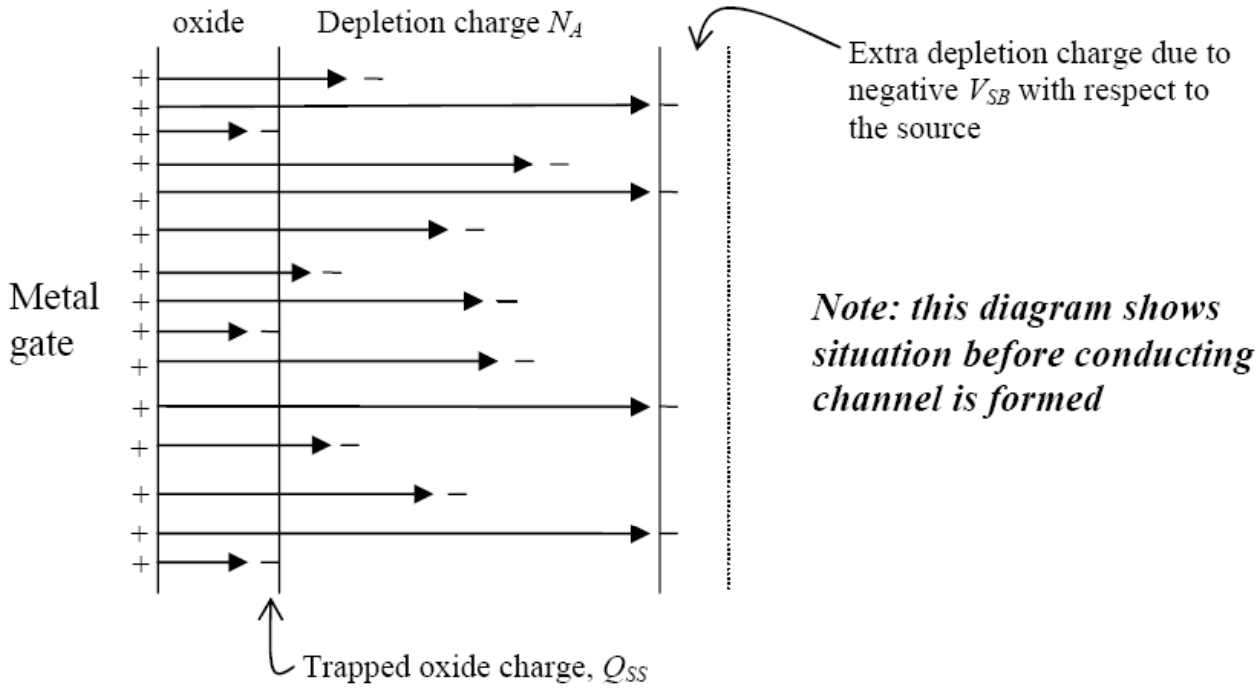


Possible to operate a device with a V_T of 0.2V with a 0.5V supply.

These low voltages are very attractive for low power operation. However they are very sensitive to manufacturing variations which can change the operating point. If there is a need to provide margins so that certain speed and power requirements are met, then the advantages disappear

MOSFET

Charge description of the threshold voltage

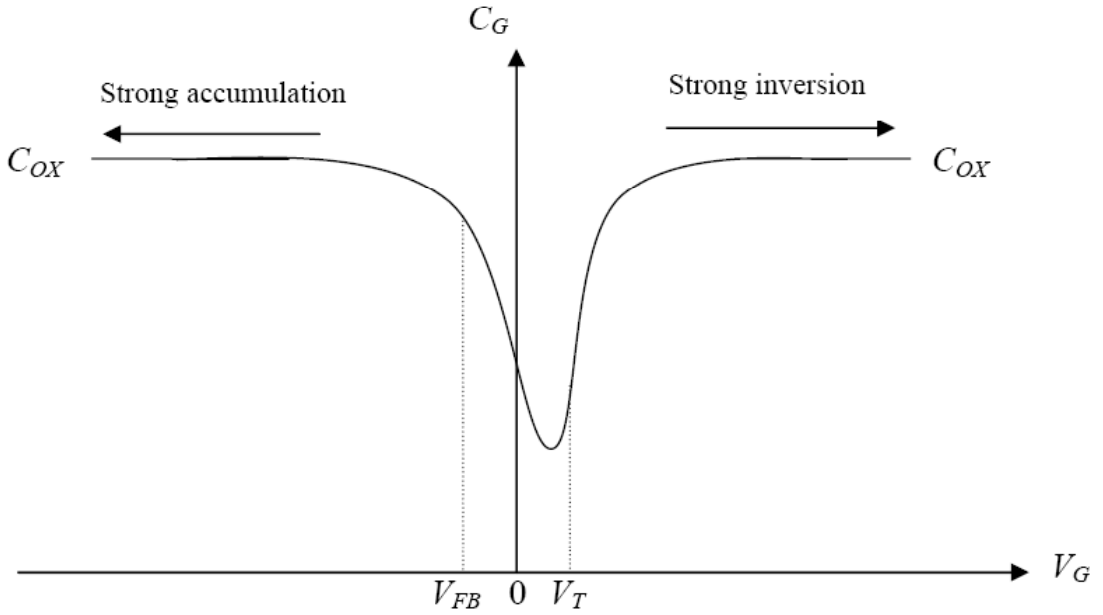


Negative ionized acceptors in the channel region contribute to the band bending and hence the production of the inversion layer. However other negative charges in the oxide or in the substrate soak up the positive gate charge

MOSFET

In this diagram, the overall charge must be neutral in moving from the gate to the substrate. At some point a conducting channel will form to balance an increasing positive gate voltage

Capacitance-Voltage relationship for the MOSFET



The gate capacitance is composed of the fixed oxide capacitance, C_{ox} with a negative and voltage dependent semiconductor capacitance ($C_{S)}$

MOSFET

$$C_{ox} = \frac{\varepsilon}{d}$$

Where ε is the dielectic constant and d is the thickness of the oxide

For the semiconductor, we can use the general expression

$$C_s = \frac{\partial Q_{SC}}{\partial V_G}$$

Where Q_{SC} is the net charge on the semiconductor

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_s}$$

Variation of capacitance

V_G large and negative: Accumulation of holes at the surface of the semiconductor. C_s is very large because Q is changing rapidly. C -> C_{ox}

V_G large and positive: Strong inversion occurs and Q is increased rapidly. C -> C_{ox}

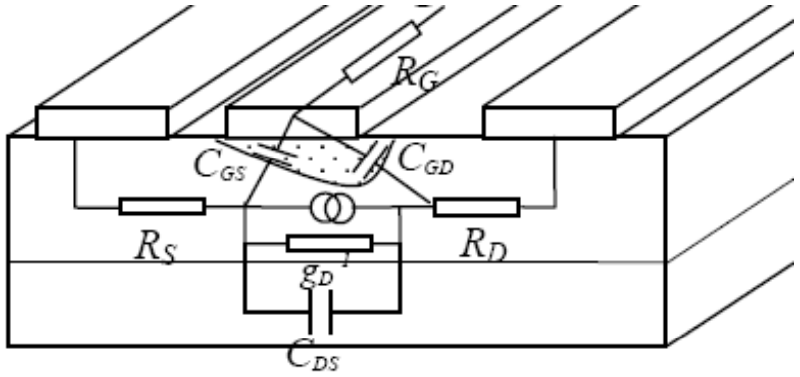
V_G small and positive ($V_{FB}<V_G < V_T$): C_{ox} and C_s have a similar value. Changes rapidly with V_G

FET Equivalent circuits

Previously we have considered only *intrinsic* delays in FETs.

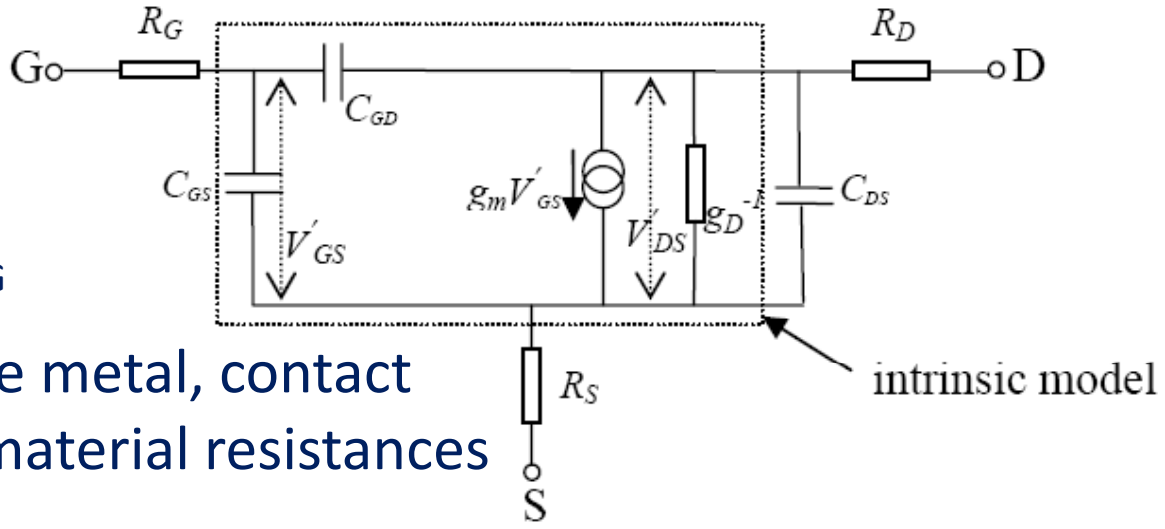
In real devices there are *extrinsic* parasitic RC components which can influence the speed of response.

Equivalent circuits are used to consider these effects



MESFET

Firstly we need to take into account the source, gate and drain parasitic resistances R_S, R_D, and R_G



These take account of the metal, contact and the semiconductor material resistances

FET Equivalent circuits

R_G can be quite significant due to the use of narrow gate lengths.

The current source in this circuit represents the output of the FET

$$I_{DS} = g_m V_{GS}$$

A further g_D^{-1} resistance is added to take into account of any changes in the output conductance

$$g_D = \frac{\partial I_D}{\partial V_{GS}}$$

This is due to channel length modulation and will be discussed later.

Finally there are capacitance components between all the inputs, C_{GS},C_{GD}, and C_{DS}

Need to consider small signal a.c. voltages- use lower case $V \rightarrow v$

Distinguish between extrinsic and intrinsic using a 'tick' $v \rightarrow v'$

FET Equivalent circuits

Extrinsic output voltage

$$v'_{DS} = v_{DS} + I_D(R_S + R_D)$$

extrinsic — intrinsic

Extrinsic input voltage

$$v'_{GS} = v_{GS} + R_S i_D + R_G i_G$$

$$\cong v_{GS} + R_S i_D \qquad \text{Since } i_G \ll i_D$$

Drain Current

$$i_D = g_m v_{GS} + g_D v_{DS}$$

but $g_m >> g_D$, so we can neglect the second term

Substitute v_{GS} *for* v'_{GS}

$$i_D = g_m v_{GS} = g_m(v'_{GS} - R_S i_D)$$

Solve for i_D

$$i_D = \frac{g_m v'_{GS}}{1 + R_S g_m}$$

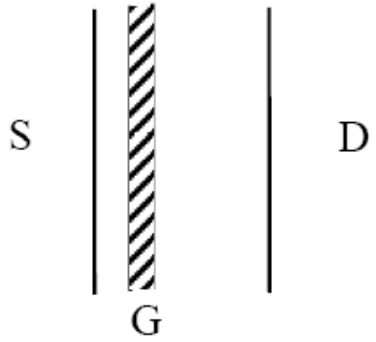
FET Equivalent circuits

Extrinsic Transconductance

$$g_m' = \frac{\partial i_D}{\partial v_{GS}} = \frac{g_m}{1 + R_S g_m}$$

Note this is the previous *intrinsic* transconductance reduced by a factor R_S. Also note it is not affected by R_D

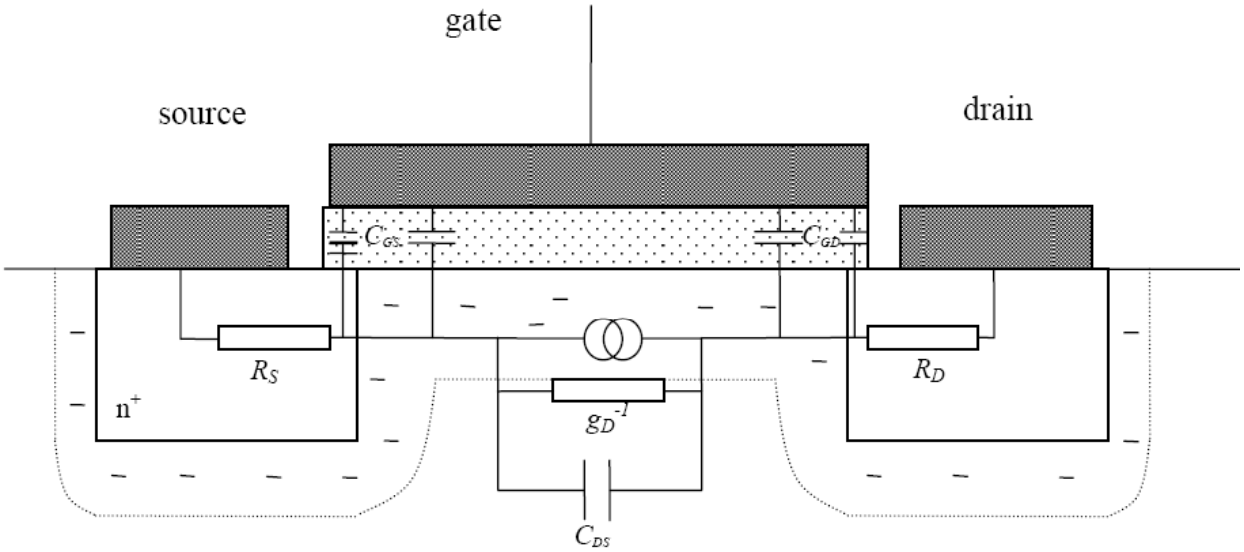
The gate is often positioned offset and closer to the source than the drain because of this issue.



The geometry has the added advantage of increasing the gate-drain breakdown voltage, increasing the power output

FET Equivalent circuits

MOSFET Equivalent circuit



Note C_{GS} and C_{GD} contain an element of gate overlap capacitance

Now lets consider the effect of the output resistance g_D^{-1}

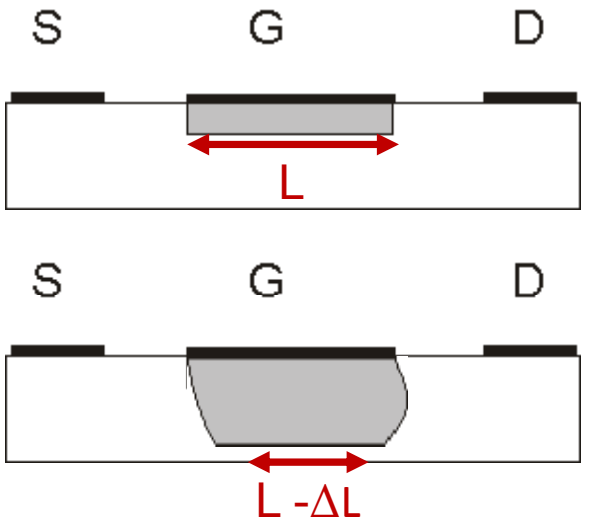
This occurs as a consequence of channel width modulation

For short gate lengths, as V_D increases under saturation the effective gate length decreases

FET Equivalent circuits

The cause of this is different in different devices

In the MESFET, as V_{DS} increases, the depletion layer gets dragged towards the drain. This movement is not symmetric however and results in a reduction in the length of the channel under the influence of the gate

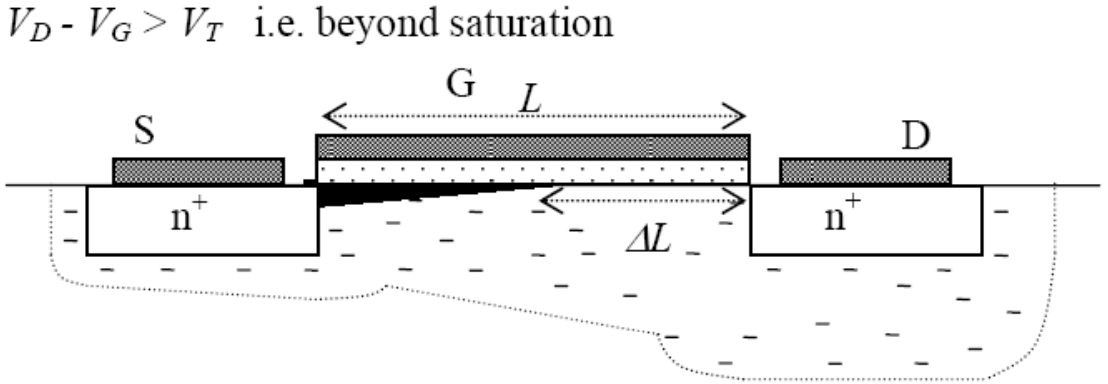


The effective gate width reduces by ΔL

FET Equivalent circuits

In the MOSFET the channel length will shorten with increased drain voltage

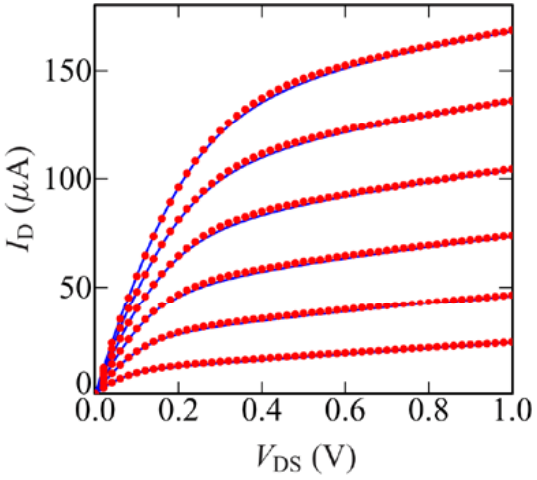
$L_{eff} = L - \Delta L$



Since $I_D(sat)$ is proportional to $1/L_{eff}$ for both the MESFET and the MOSFET then as V_{DS} increases, L_{eff} decreases and $I_D(sat)$ increases

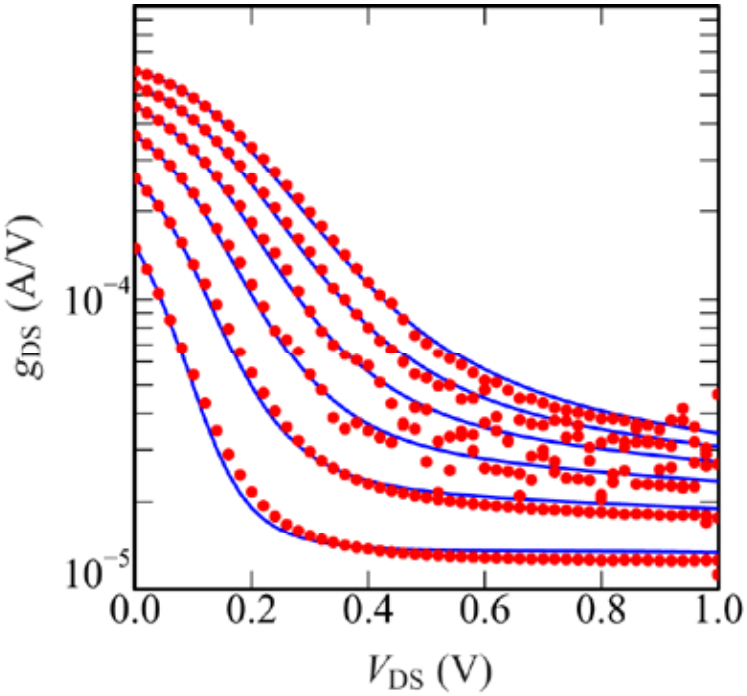
The output conductance is finite and not zero, unlike the idealised characteristics shown before

$$g_D = \frac{\partial I_D}{\partial V_{DS}}$$



FET Equivalent circuits

Output conductance g_{DS} calculated from the previous figure. If these values become significant, they cannot be neglected as in the previous argument



Calculation of g_D

Again MESFET is easier to consider, but MOSFET results are similar

Assume a simple depletion model $$\Delta L = \sqrt{\frac{2\varepsilon(V_{DS} - V_P + V_G)}{qN_D}}$$

This is the depletion width above the pinch off condition, which is characterised by a voltage V_P.

FET Equivalent circuits

Since $I_D(sat) \propto 1/L_{eff}$ we can write $I_D' = I_D(sat)\frac{L}{L-\Delta L}$

Using the first two terms of a binomial expansion $I_D' = I_D(sat)\left[1+\frac{\Delta L}{L}\right]$

$$\Delta I_D = I_D' - I_D(sat) = \left[1+\frac{\Delta L}{L}\right]$$

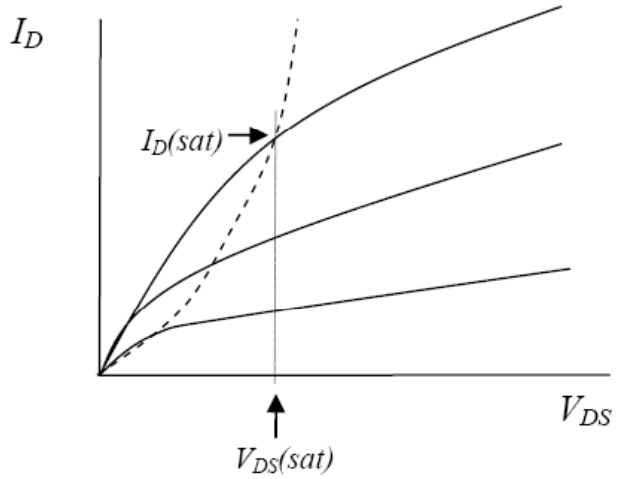
$$g_D = \frac{\Delta I_D}{\Delta V_{DS}} = \frac{I_D(sat)\Delta L}{\Delta V_{DS} L} \propto I_D(sat)\left[1+\frac{\Delta L}{L}\right]$$

So g_D increases as $I_D(sat)$ increases (different V_G)

Also g_D is larger for shorter gate lengths

FET Equivalent circuits

Therefore we end up with the typical 'fan' type characteristics reflecting changes in DL as I_{sat} increases



Example: Consider a V_{DS} of 1V, beyond the pinch off

i.e. $V_{DS} - V_{DS}(sat) = V_{DS} - V_P + V_G = 1V$

$N_D = 5x10^{22}\ m^{-3}$, $\varepsilon(GaAs) = 13.2 \times 8.85x10^{-12}\ Fm^{-1}$.

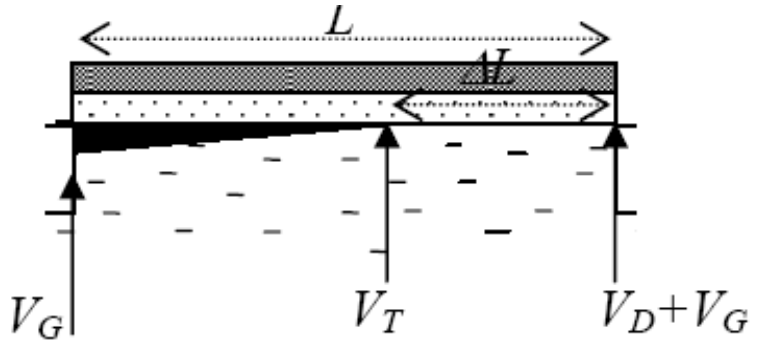
$$\Delta L = \sqrt{\frac{2\varepsilon(V_{DS} - V_P + V_G)}{qN_D}}$$ = 0.17μm

This is very significant reduction if the gate length is < 1μm

FET Equivalent circuits

MOSFET case (a lot more complex)

Consider the MOSFET beyond V_T



$$\Delta I_D = I'_D - I_D(sat) = \frac{I_D(sat)\Delta L}{L} \qquad g_D = \frac{I_D(sat)\Delta L}{\Delta V_{DS} L}$$

ΔL not so easy to express analytically as in the MESFET case, but the relationship is similar.