Part 3: Transistor basics

The Bipolar Junction Transistor (BJT)

The Metal-Semiconductor Field Effect Transistor (MESFET)

Semiconductor Transistor

The basic building block of analogue and digital circuits is the semiconductor transistor.

This can be either:

A bipolar junction transistor (BJT)

(voltage controlled diffusion of charged carriers across a semiconductor junction)

Or a **Field effect transistor (FET)**

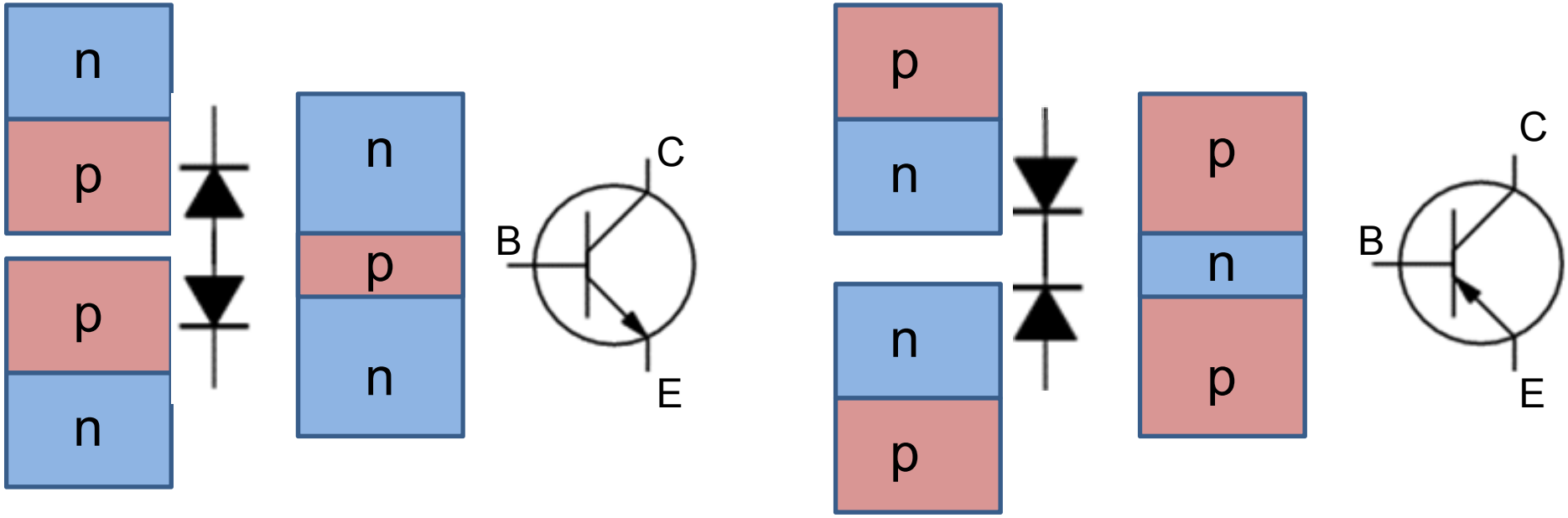
(voltage controlled channel size/conductivity)

or some combination of the two; e.g.: IGBJT

Depending on the circuit the transistor acts as an amplifier, a buffer or switch.

Bipolar Junction Transistor (BJT)

Bipolar junction transistor –two back-to-back pn diodes

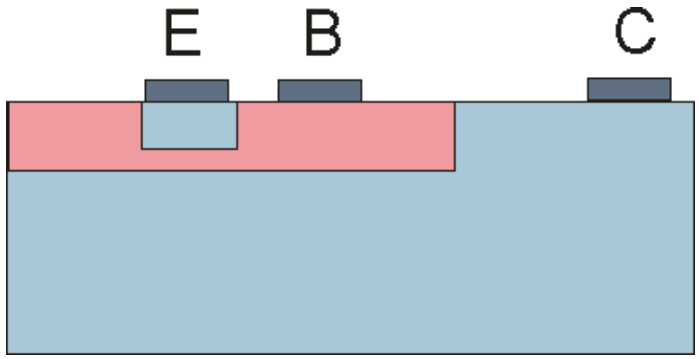


Carriers injected (emitted) from the emitter into the base

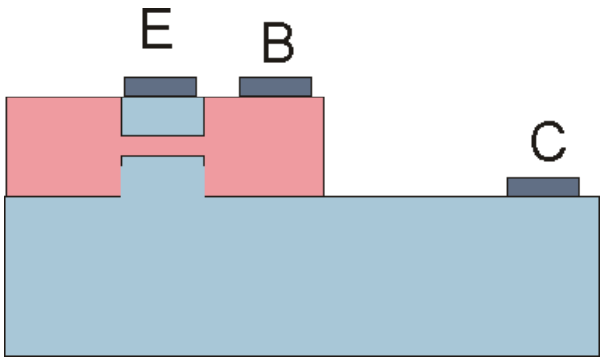
Base very thin (< diffusion length), allowing most of the carriers to be transported to the collector

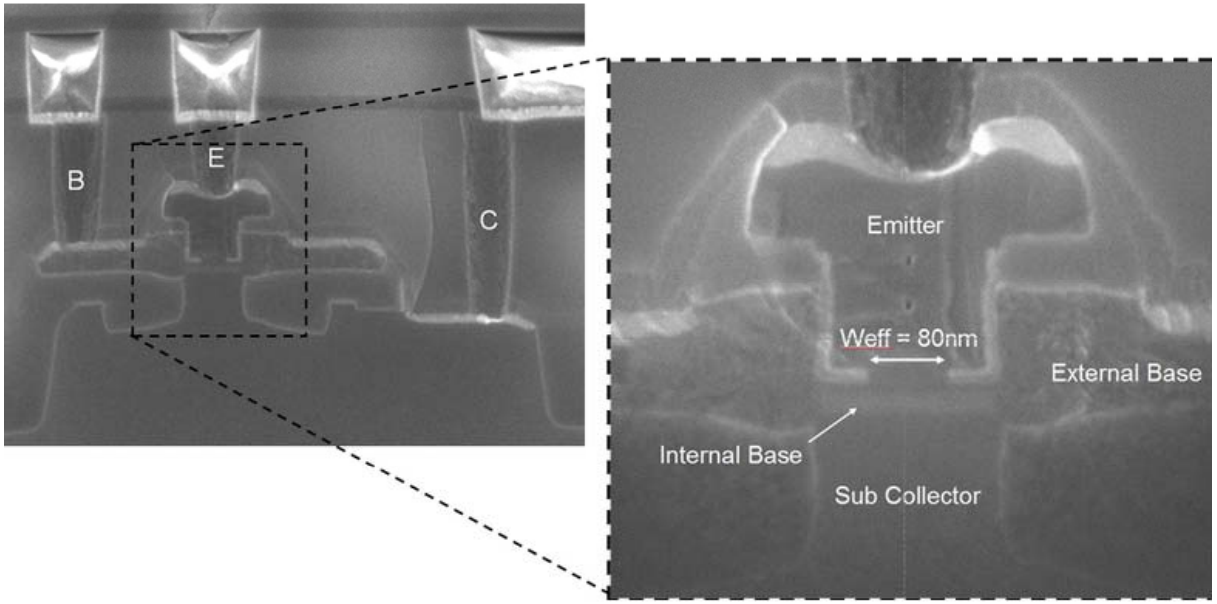
A small change in the base current results in a large change in collector current

Bipolar Junction Transistor



Basic construction, using ion implantation and a high speed version (HBT) below with a H shaped base T shaped emitter and a sub-collector





Transistor basics (BJT)

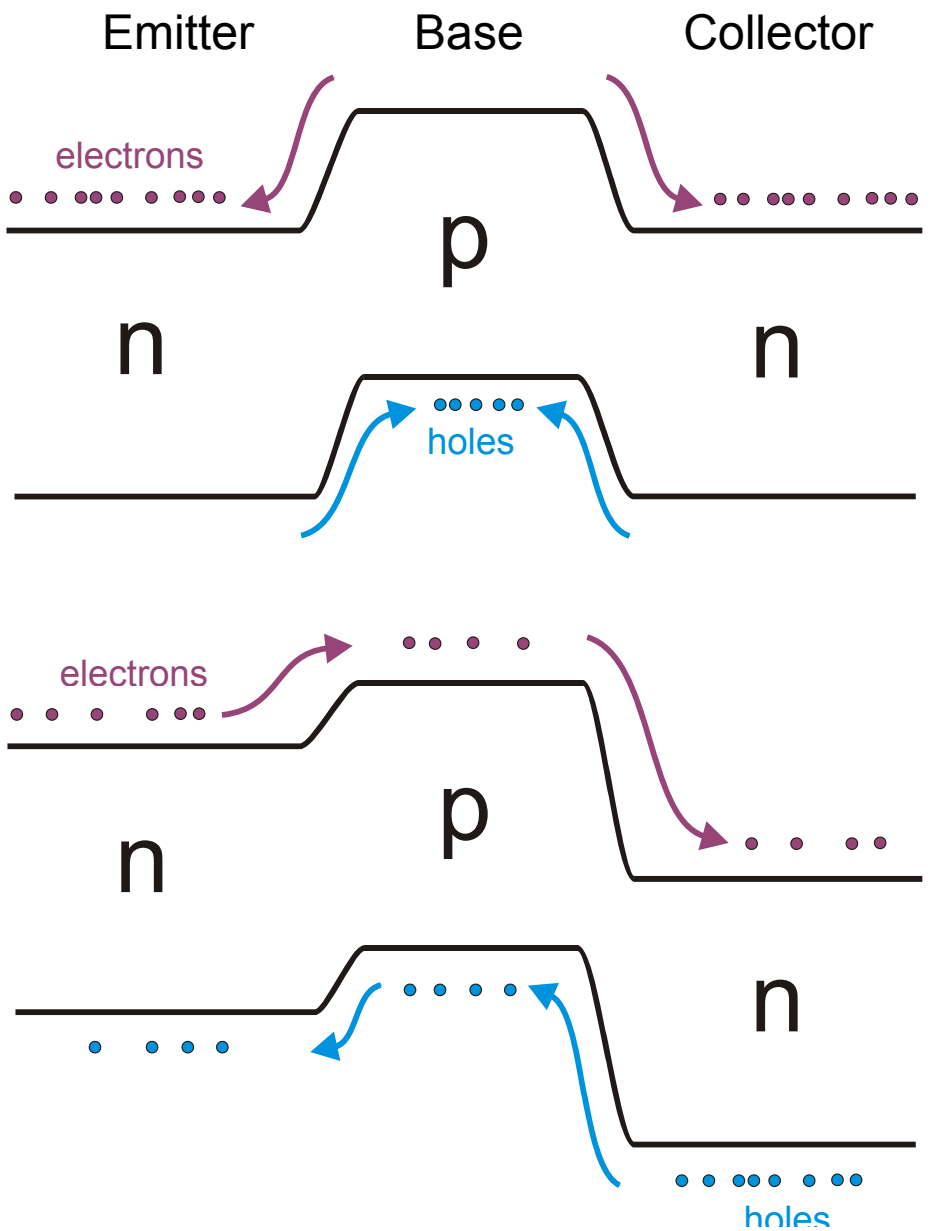
Consider the common-base configuration

At equilibrium, no current flows through the device due to the built in voltage (~0.7V for Si)

Under forward bias conditions the emitter-base barrier is reduced. As V_{BE} approaches 0.7V a base current I_B appears

The small base current I_B effectively controls a large collector current I_c

Note h_{FE} is the small signal current gain (DC current gain -α_B)

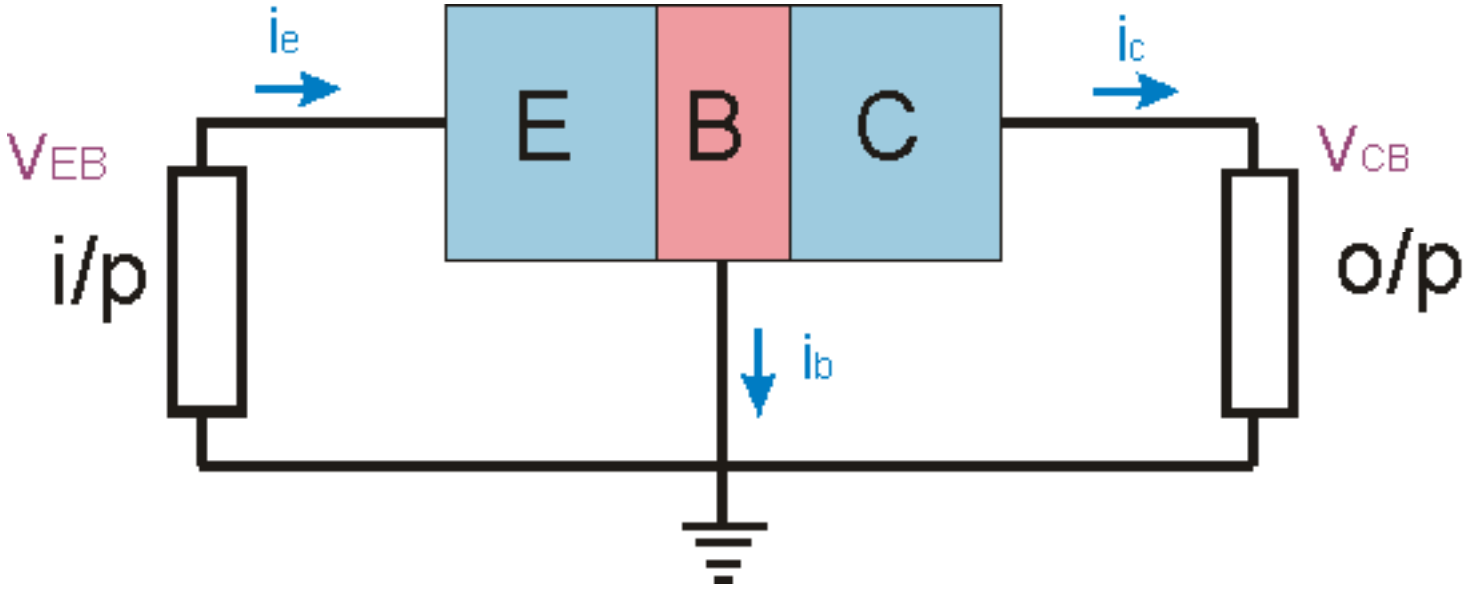


Bipolar Junction Transistor

Common Base: Base connected to ground. The emitter acts as the input and the collector as the output

Characteristics: gain: $h_{FE} = -\frac{\Delta I_C}{\Delta I_E} \approx 1$ (unity gain), but low input impedance

Applications: voltage amplifier, current follower

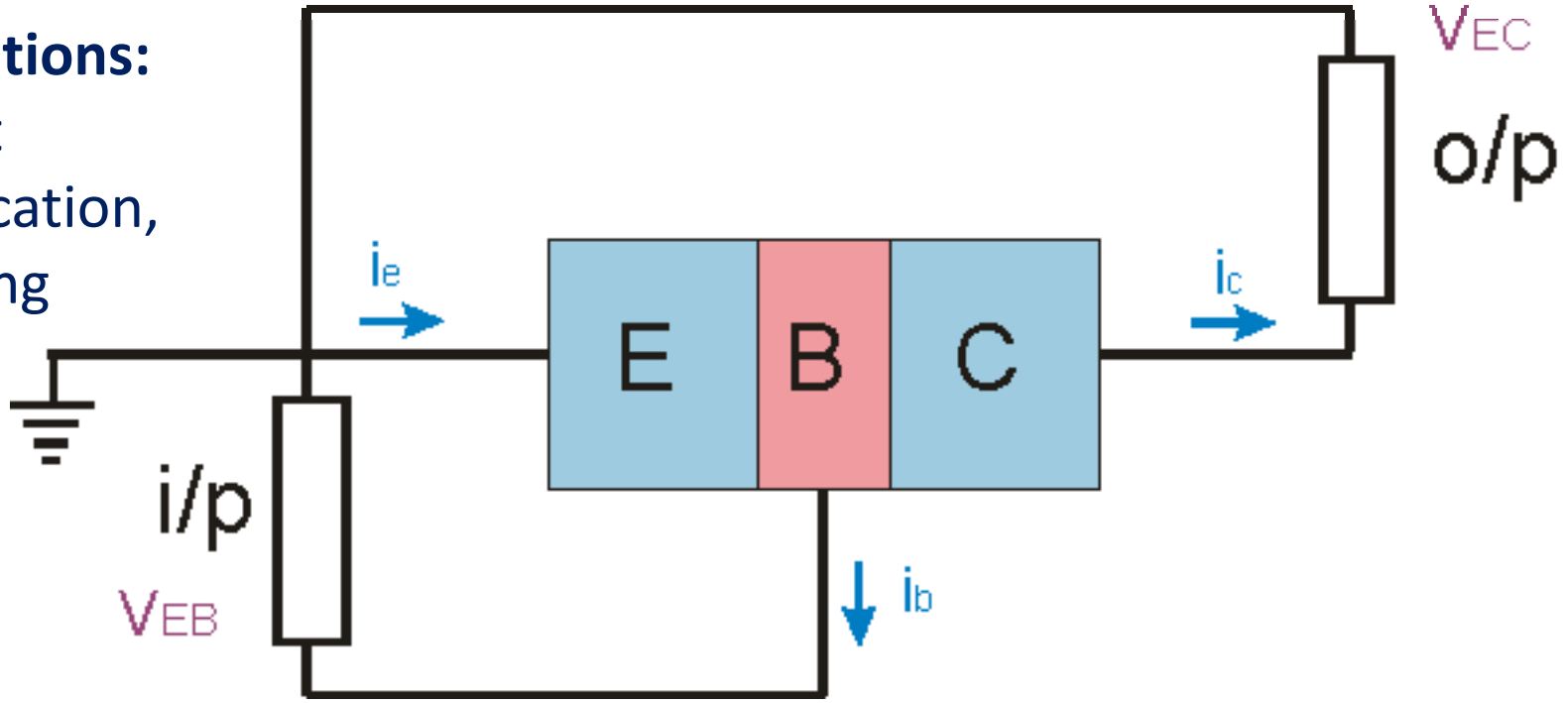


Bipolar Junction Transistor

Common emitter: Emitter connected to ground. The base acts as the input and the collector is used as the output

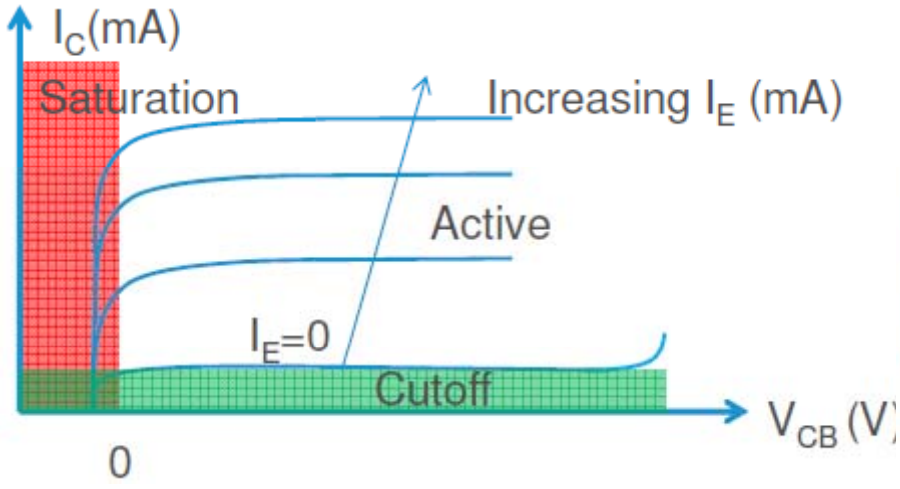
Characteristics: gain: $h_{FE} = -\frac{\Delta I_C}{\Delta I_B} \approx 20 - 500$ with high input impedance

Applications:
Current amplification, switching

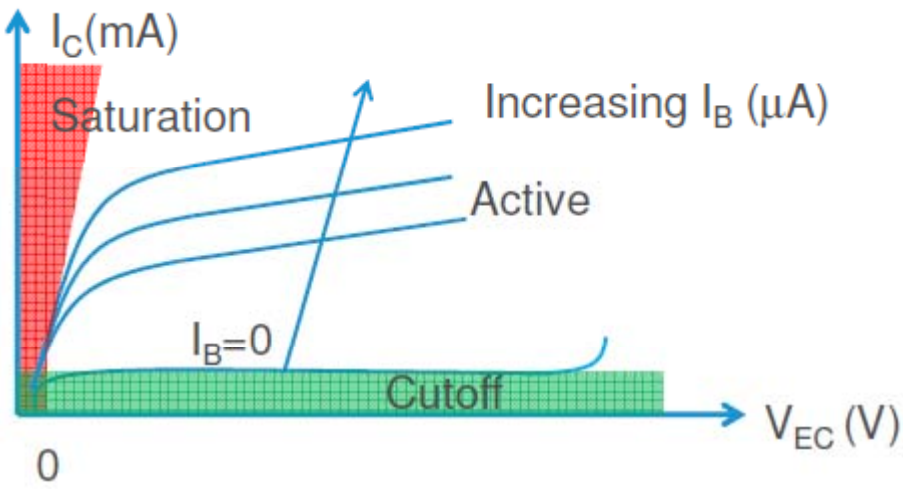


Bipolar Junction Transistor

Common Base



Common Emitter

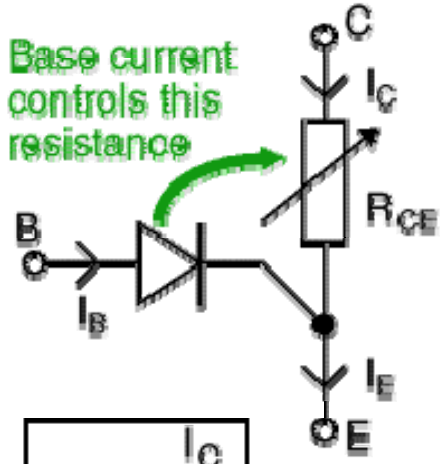


Saturation: Both junctions are forward biased and device functions as a permanently on switch. I_C rises to a maximum and is not well controlled by I_B

Active: Base-emitter forward biased, base collector reverse biased. I_c controlled by I_B

Cut-off: Both junctions reverse biased, no current flow

Bipolar Junction Transistor



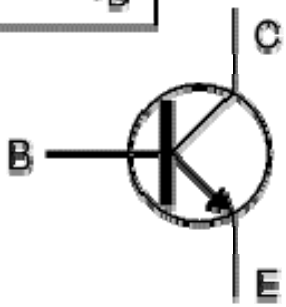
May also think of this as a variable resistor controlled by the base terminal (actually this is the origin of the term transistor (***transfer- resistor***)

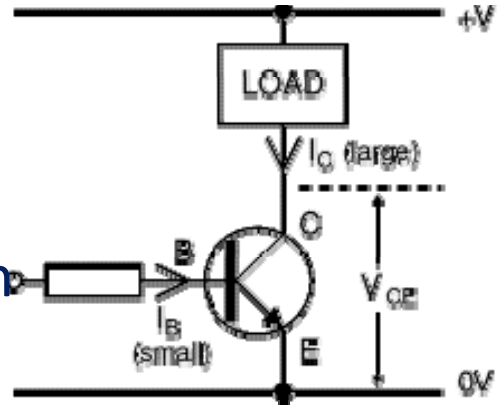
In the OFF state: power = $I_C \times V_{CE}$, but $I_C \rightarrow 0$, so the power is zero.

In the full ON state: power = $I_C \times V_{CE}$, but $V_{CE} \rightarrow 0$ so the power is very small

We say that I_C and $V_{CE} \rightarrow 0$ but these are not = 0. A leakage current in I_C may exist of typically 10-100µA whilst V_{CE} may never be zero due to a finite resistance (few Ω).
This gives a finite power dissipation, which takes on great significance for ICs.

$$h_{FE} = \frac{I_C}{I_B}$$





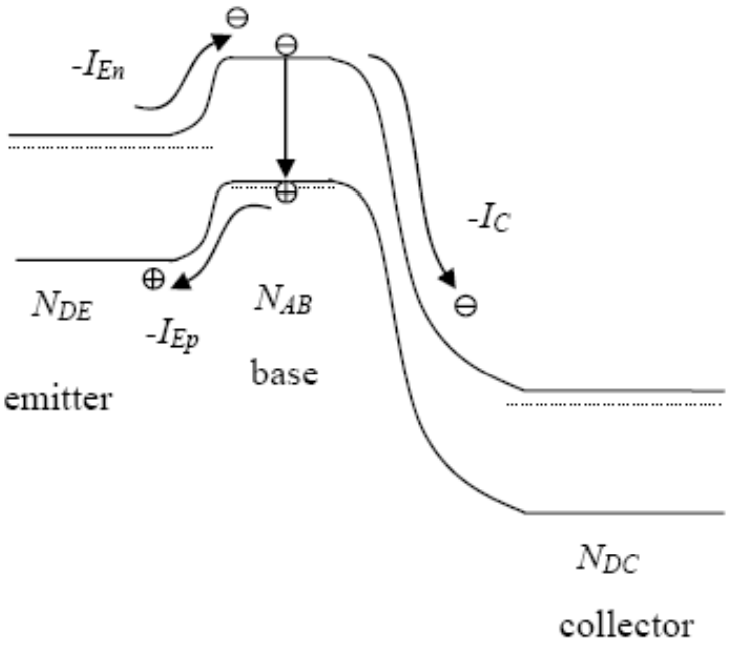
Bipolar Junction Transistor

BJT Design

Consider an NPN transistor

No of electrons in the emitter needs to be much more than the number of holes in the base $N_{DE} >> N_{AB}$

Therefore $I_{En} >> I_{Ep}$ Emitter current dominated by electrons $I_E = I_{Ep} + I_{En}$



Emitter injection efficiency **γ** (proportion of emitter current due to electrons)

$$\gamma = \frac{I_{En}}{I_{En} + I_{Ep}} \approx 1$$

Some proportion of I_E will pass through the base if it is thin enough. So $I_C = \alpha I_{En}$ where **α** is the base transport factor .
α ≈ 1 for a good transistor (thin base).

Bipolar Junction Transistor

$$\frac{I_C}{I_E} = \frac{\alpha I_{En}}{I_{En} + I_{Ep}} = \alpha\gamma$$

So we end up with the basic equations of operation of the BJT

$$I_B = I_{Ep} + I_{En}(1-\alpha)$$

And the current gain **β**

$$\beta = \frac{I_C}{I_B} = \frac{\alpha I_{En}}{I_{Ep} + I_{En}(1-\alpha)} = \frac{\alpha\gamma}{1-\alpha\gamma}$$

Note α, γ ≈ 1, so β can be high (typically 50-1000)

Remember we need I_{EP} small compared to I_{EN}

- Low base doping (but this has consequences in terms of increasing the base resistance)
- Reduce the electron barrier, increase the hole barrier (this needs heterojunctions, see the HBT later)
- Reduce recombination

Field Effect Transistor

The BJT and its variants use an applied field to change the built-in band structure of pn junctions

In a field effect transistor there are no built-in junctions and instead an electric field is used to create a band structure that can attract or repel electrons. The field is used to create or destroy a conducting channel.

The field effect transistor is in many ways a simpler device and was envisaged before the bipolar transistor. However it is by nature a near surface device and in the early days of semiconductors surfaces were of poor quality (impurities, defects)

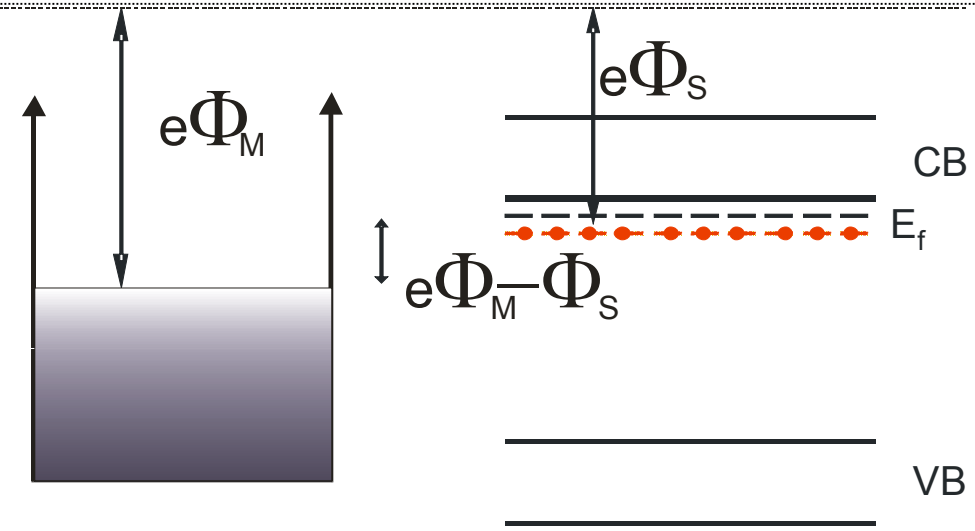
Metal-Semiconductor FET (MESFET)

A number of FET types exist, but we start with the MESFET which is in many ways the most simplest device. **Also known as a JFET.**

To understand MESFETS we have to revise metal-semi-conductor junctions, of which there are two types; ohmic and rectifying, depending on the metal and semiconductor work functions.

Case 1

$\emptyset_m > \emptyset_s$ n-type



Barrier, Φ_B

$= e(\Phi_M - \Phi_S)$

Electrons spill over from the semiconductor to the empty levels in the metal , leaving a depletion layer in the semiconductor.

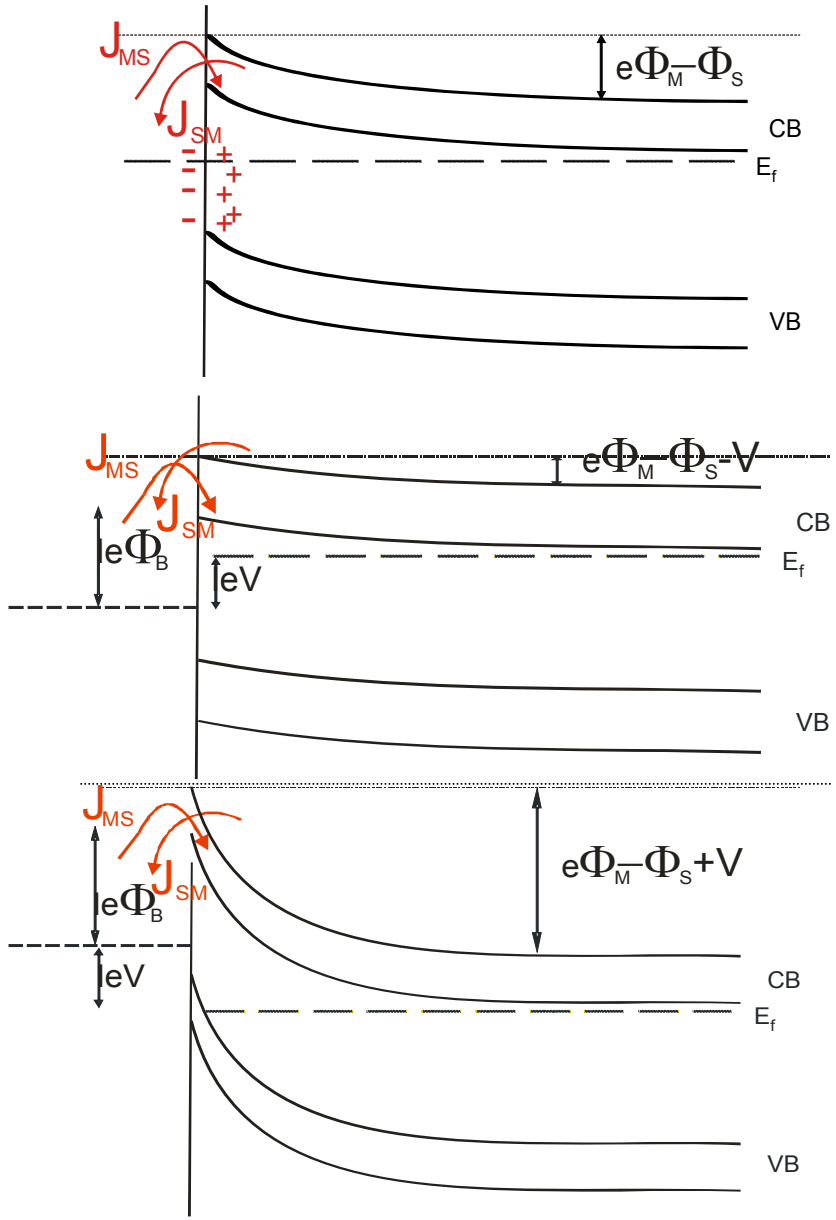
MESFET

Due to electron depletion the band edges bend-up to form a potential energy barrier = e ($Ø_m$ – $Ø_s$), opposing further current flow

Forward bias: Semiconductor levels rises and the barrier height decreases for electrons in the semiconductor. A current J_{SM} flows

Reverse bias: Semiconductor levels fall. Barrier height increases and $J_{SM} \rightarrow 0$

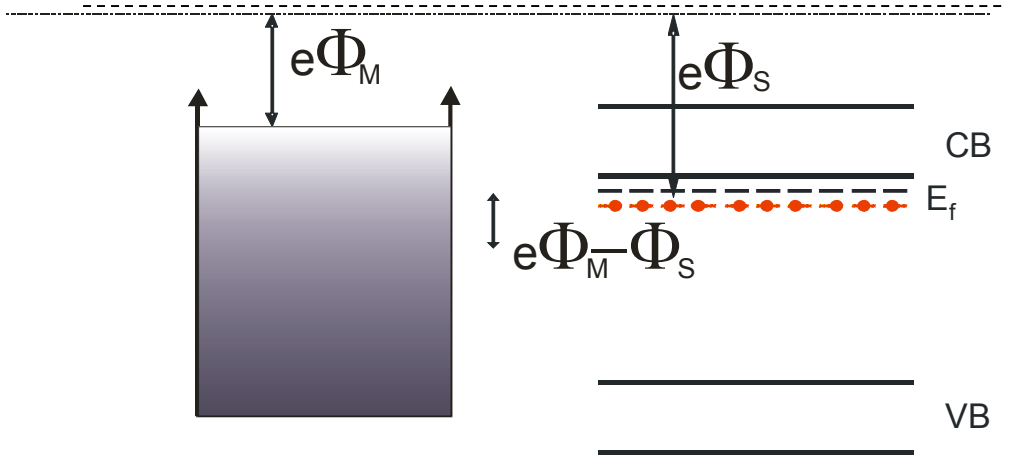
This configuration has Schottky (rectifying) characteristics



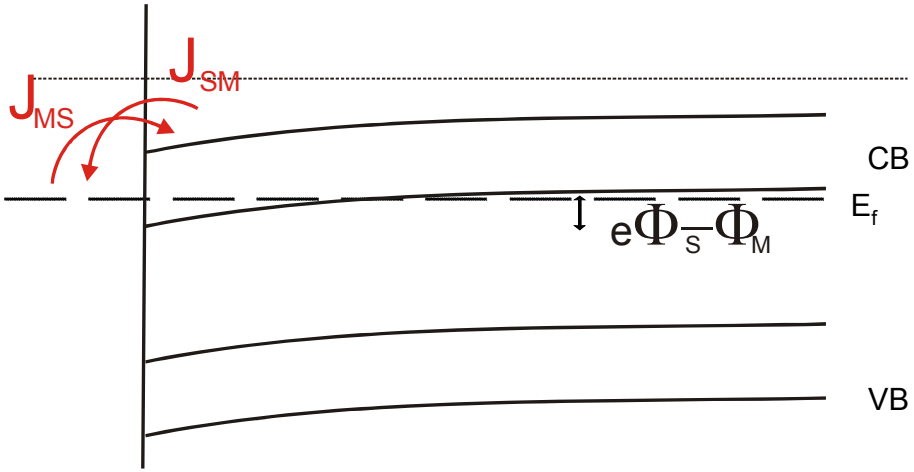
MESFET

Case 2

$Ø_s > Ø_m$ n-type



On contact electrons flow from metal to semiconductor leading to electron accumulation at the interface
The bands bend down



MESFET

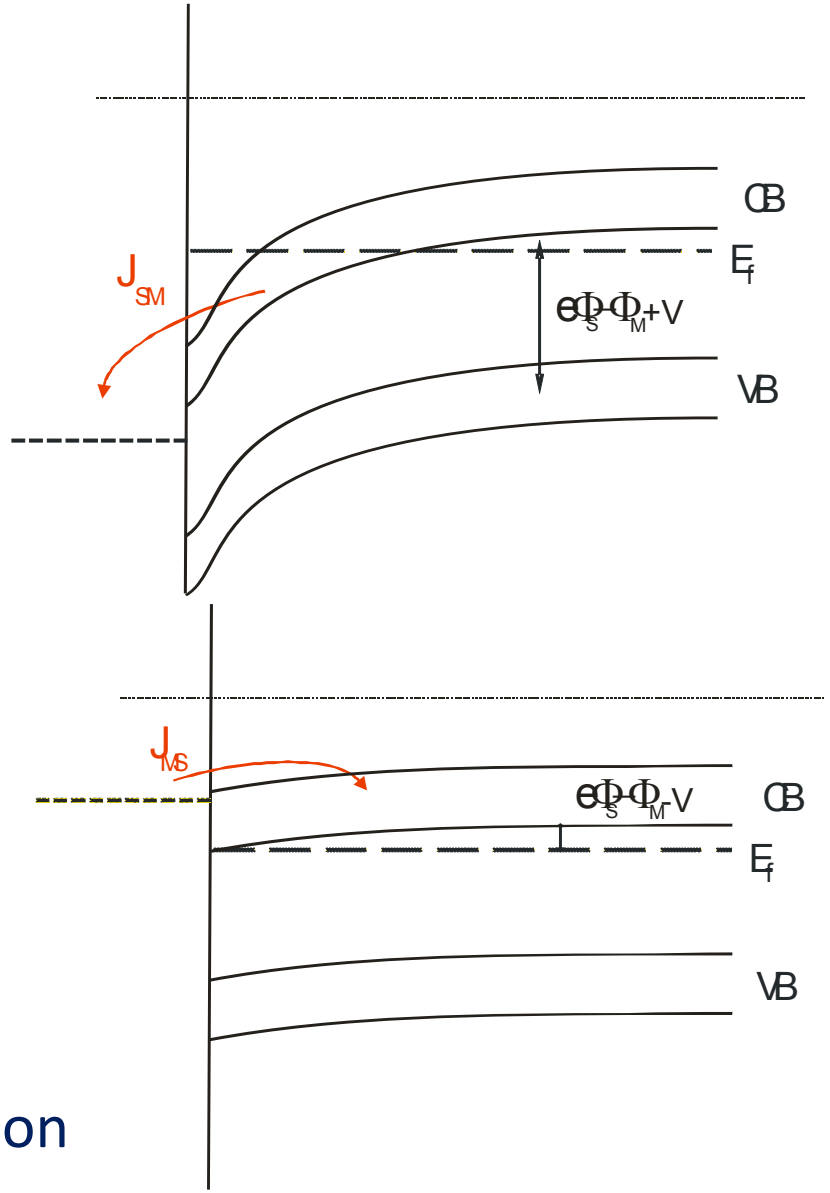
Forward bias: semiconductor levels raised, J_{SM} flows

Reverse bias: semiconductor levels fall, little or no barrier for electrons from the metal, J_{MS} rises

Current flows in both directions

Behaviour is ohmic

Note we always need good ohmic contacts, whatever the device function



MESFET

So two types of behaviour, depending on the relative work functions Q_M and Q_S

Table shows the barrier height of various metal-semiconductor systems. Plenty of good Schottky contacts, not so many good Ohmic.

	Ag	Al	Au	Cr	Ni	Pt	W
Φ_M (in vacuum)	4.3	4.25	4.8	4.5	4.5	5.3	4.6
n-Ge	0.54	0.48	0.59		0.49		0.48
p-Ge	0.5		0.3				
n-Si	0.78	0.72	0.8	0.61	0.61	0.9	0.67
p-Si	0.54	0.58	0.34	0.5	0.51		0.45
n-GaAs	0.88	0.8	0.9			0.84	0.8
p-GaAs	0.63		0.42				

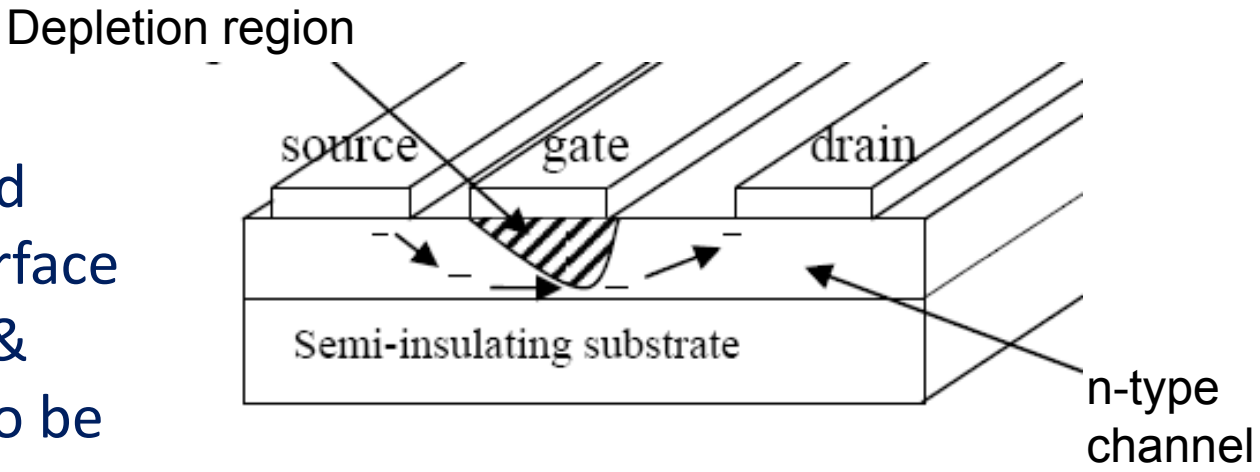
Methods for Ohmic contacts

- Alloying- inter-diffuse metal and semiconductor at high T. Can structurally and chemically lower the barrier height. (examples: formation of metal silicides, alloying of Zn in GaAs)
- Deposition of a low band gap semiconductor, since naturally these have high workfunctions

MESFET

MESFET Structure

The MESFET is created from a conducting surface layer . Ohmic source & drain contacts need to be made and a Schottky 'gate' in the middle

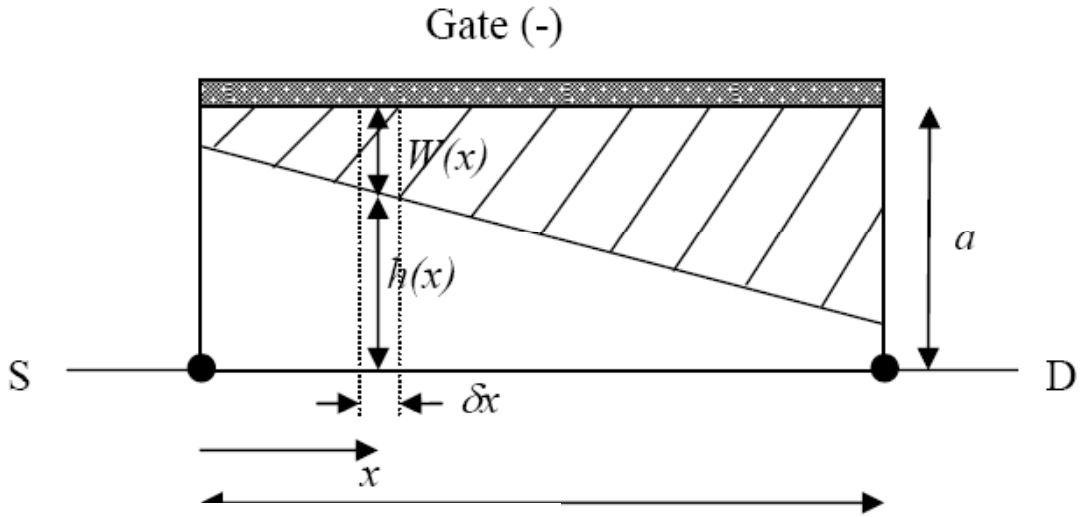


As the gate is reverse biased, the Schottky depletion region extends until it 'pinches off' the channel.

Variations in the gate voltage therefore control the depletion width which then control the current flow under the gate.

MESFET

Consider a simplified section of the MESFET channel under the gate



Depletion width

$$W(x) = \left[\frac{2\varepsilon(V_x + V_G)}{qN_D}\right]^{1/2} \quad (1)$$

where ε = dielectric constant of semiconductor, V_X= potential at x (varies with x due to combination of constant V_G and D-S bias, V_D). V_G = gate bias including built-in voltage of Schottky gate and N_D= doping density in the channel

MESFET

Resistance of volume element, δ*x*

$$R = \frac{\rho l}{A} \quad R_{\partial x} = \frac{\rho \partial x}{Zh(x)}$$

where Z = gate width and ρ = resistivity of channel

Drain current, I_D, related to the voltage drop, ∂V_x, across the element

$$I_D = \frac{\partial V_x}{R_{\delta x}} = \frac{Zh(x)}{\rho}\frac{\partial V_x}{\partial x} \quad (2)$$

Now *h(x) = a-W(x). Using W(x) from eq.(1), gives:*

$$h(x) = a - \left(\frac{2\varepsilon(V_G + V_x)}{qN_D}\right)^{1/2} = a\left[1 - \left(\frac{(V_G + V_x)}{V_P}\right)^{1/2}\right] \quad (3)$$

MESFET

Where $V_P = \frac{qa^2 N_D}{2\varepsilon}$ = pinch off voltage i.e. the voltage at which the channel is fully depleted.

Substituting for h(x) above into eq.2 for I_D gives:

$$I_D \partial x = \frac{Za}{\rho}\left[1 - \left(\frac{V_x + V_G}{V_P}\right)^{1/2}\right]\partial V_x$$

Integrate this over the whole channel length (L)

$$I_D \int_0^L \partial x = \frac{Za}{\rho}\int_0^{V_D}\left[1 - \left(\frac{V_x + V_G}{V_P}\right)^{1/2}\right]\partial V_x$$

Gives the drain current $I_D = \frac{Za}{\rho L}V_P\left[\frac{V_D}{V_P} + \frac{2}{3}\left(\frac{V_G}{V_P}\right)^{3/2} - \frac{2}{3}\left(\frac{V_D + V_G}{V_P}\right)^{3/2}\right]$ (4)

MESFET

Where V_D = *drain voltage*

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

At pinch off *the drain current saturates.*

$$I_D(sat) = \frac{Za}{\rho L} V_P \left[\frac{V_D}{V_P} + \frac{2}{3}\left(\frac{V_G}{V_P}\right)^{3/2} - \frac{2}{3} \right]$$

At pinch off $V_D + V_G = V_P$ V_D is positive with respect to the source whilst V_G is negative with respect to the source.

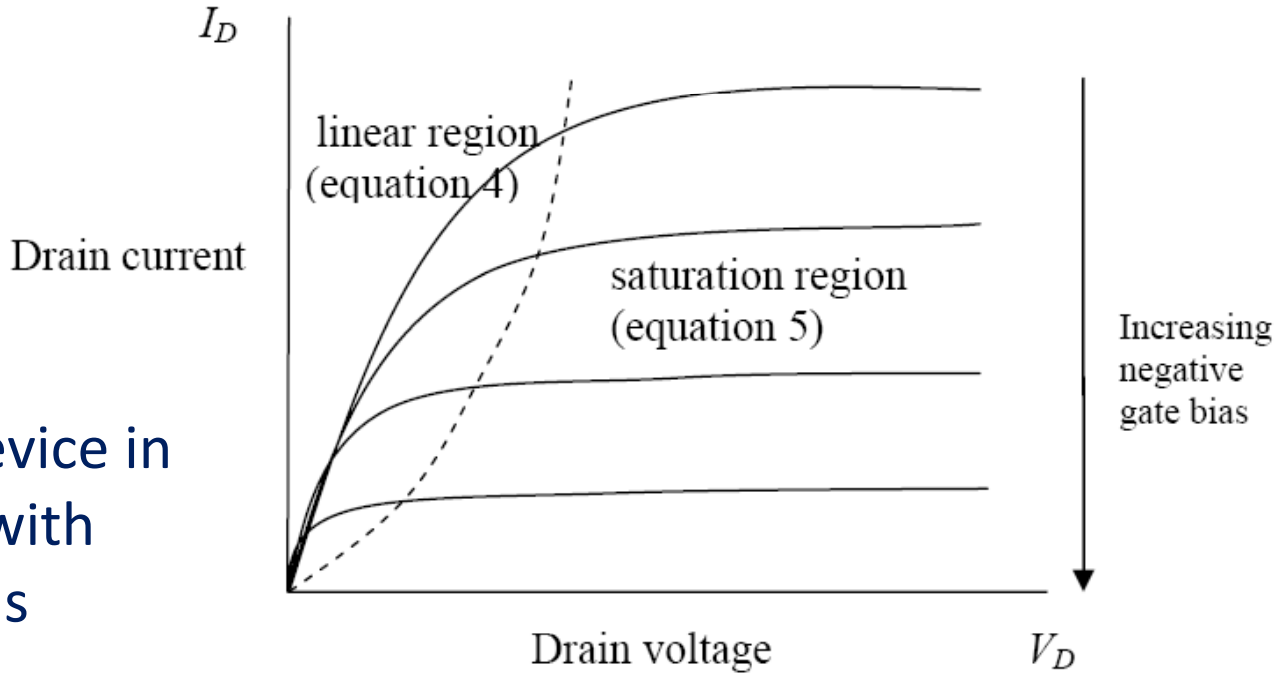
Substitute in for V_D

$$I_D(sat) = \frac{Za}{\rho L} V_P \left[\frac{1}{3} - \frac{V_G}{V_P} + \frac{2}{3}\left(\frac{V_G}{V_P}\right)^{3/2} \right]$$

Note this is independent of V_D

MESFET

Typical output characteristics of a MESFET



Note that this is a *depletion mode* device in which ID reduces with increasing gate bias

Normal to operate the device in the saturation region with I_D solely controlled by the gate voltage.

Above the pinch-off point (i.e: for $V_D + V_G > V_P$), the lateral fields under the drain end of the gate become very large such that the electron velocity reaches saturation.

MESFET

If we differentiate I_D (sat) with respect to V_G then we get the small signal conductance (actually not a direct conductance but a transfer-conductance or **transconductance)**

The transconductance is a measure of how well the FET controls the drain current, I_D, by changing the gate voltage, V_G.

$$g_m = \frac{\partial I_D(sat)}{\partial V_G} = -\frac{Za}{\rho L}\left[1-\left(\frac{V_G}{V_P}\right)^{1/2}\right]$$

The negative sign means I_D decreases as the negative gate bias increases towards V_P

Experimentally this is written as

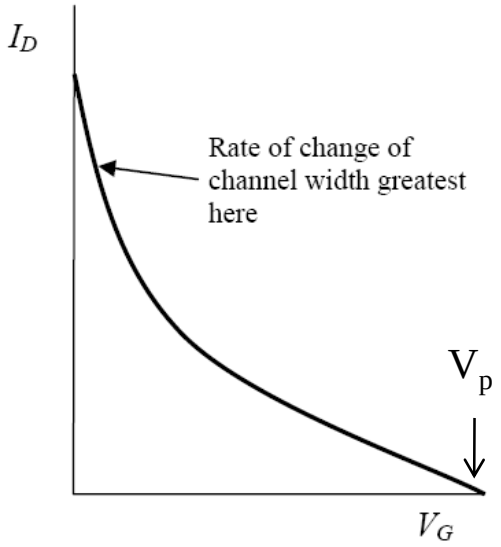
$$g_m \approx I_{DSS}\left[\left(1-\frac{V_G}{V_P}\right)^{1/2}\right]$$

where $I_{DSS} = I_D(sat)$ when $V_G = 0$

MESFET

I_D and therefore g_m is dependent on the carrier mobility, μ_e, μ_h.

g_m decreases with increasing V_G since depletion width $\propto V_G^{1/2}$. The gate voltage has its greatest influence on the channel width near the gate i.e. at low gate voltage.



When the channel is reduces towards pinch off, intuitively I_D should drop to zero. However, it never completely reaches zero– why?

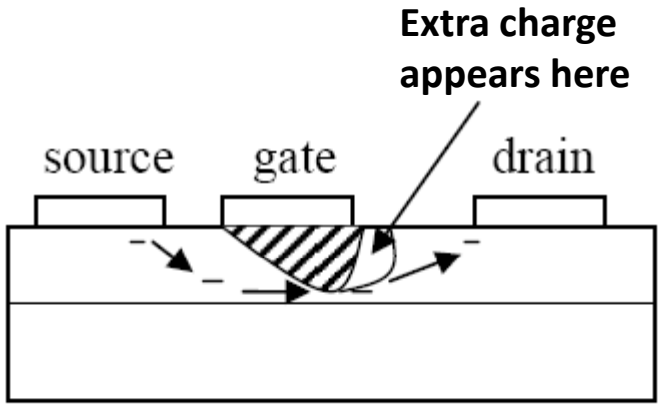
Under the gate, just before the pinch off, the drain current I_D (= channel area . n(x).q.v) must be **continuous**

If it where not continuous, there would be an accumulation of charge on one side of the device, which cannot happen

MESFET

As the gate bias increases the depletion width increases, the channel width reduces and charge begins to build up on the source side.

Rather the pinching-off completely the depletion begins to extend towards the drain end of the device



A very high field now exists in the channel region forcing electrons through

Further increases in V_G accumulate more charge towards the drain
The depletion width never closes the channel. Instead a small current flow remains (typically pA-nA)

MESFET

Advantages: The MESFET has a buried channel, unlike the MOSFET which has a surface channel. Carriers have higher 'bulk like' mobility as opposed to the MOSFET in which they strongly interact with the surface and the oxide layer.

This leads to a high g_m and better high frequency performance which is favourable for microwave applications.

The buried channel also yields a better noise performance as the trapping and release of carriers into and from surface states and defects is eliminated.

Disadvantages: The device is a normally on depletion mode device. This function is not very attractive for digital applications as it would lead to a high static power dissipation.

MESFET

The issues with this are due to the use of a Schottky gate. The forward bias that can be applied is limited to below the turn-on voltage of the Schottky diode (typically 0.6-0.7V).

Because of this the MESFET normally operates in depletion mode (reverse bias gate) and enhancement mode devices are virtually impossible.

By virtue of its insulating gate, voltages of either direction can be applied to the MOSFET, allowing both depletion and enhancement mode devices.

As a result the MOSFET (see later) has largely replaced the MESFET in most application areas

MESFET

Frequency Limitations of the MESFET (neglecting parasitic effects)

The gain of all transistors degrades as the operating frequency increases towards its limit.

The **cut off frequency, f_T**, is an important parameter to characterise high frequency performance.

It is defined as the frequency at which the current gain reaches 1 (unity gain). f_T is an important figure of merit for the speed of response of a transistor.

Consider an input small signal a.c.current (gate-source)

$$I_{GS} = \frac{V_g}{Z_G} = \frac{V_g}{\frac{1}{\omega C_G}} = 2\pi f C_G V_G$$

MESFET

where V_G = input (gate) voltage.

C_G = gate capacitance per unit gate width

Note: capacitance, g_m, and drain current are often quoted per unit gate width: this helps when comparing the performance of devices

Output a.c. current (drain-source) = $g_m v_G$, since $g_m = \frac{\partial I_D}{\partial V_G}$

Current gain $= \frac{g_m V_G}{2\pi f C_G V_G} = 1$ when f = f_T (definition of f_T)

So $f_T = \frac{g_m}{2\pi C_G}$ (1) (with g_m in units of S per unit gate length)

Now $g_m = \frac{\partial i_D}{\partial v_G}$ $i_D = n_S q v_{sat}$ (per unit gate width)

$n_S q$ = sheet charge in the channel (per unit area)

MESFET

Therefore $g_m = \frac{\partial(n_S q)}{\partial v_G}.v_{sat} = \frac{C_G}{L}.v_{sat}$ (2)

where L is the gate ~~width~~ length

Capacitance per unit gate area C_G/L

N.B. C_G = gate capacitance per unit gate width

Combining (1) and (2) $f_T = \frac{C_G}{L} v_{sat} \frac{1}{2\pi C_G} = \frac{v_{sat}}{2\pi L} \Rightarrow \boxed{f_T = \frac{1}{2\pi\tau}}$

where τ = transit time under the gate (i.e. the intrinsic speed limit is due to the transit time under the gate)

⇒ **Need a short gate length and fast electron transport**

Rules of thumb for circuit designers for operation at a frequency *f*

$f_T \sim 10 \times f$ - for low noise amp $f_T \sim 4 \times f$ - for power amp

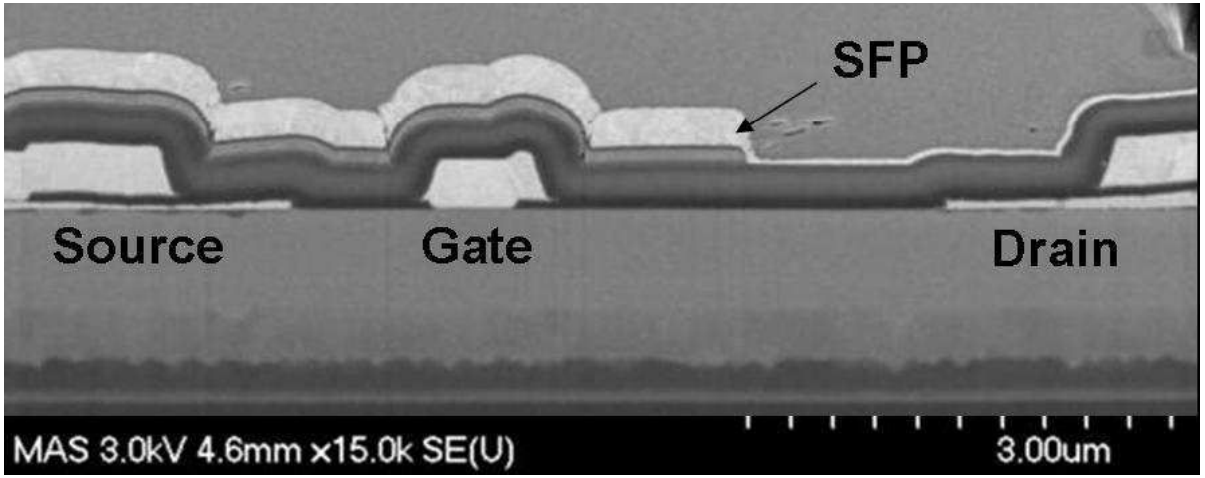
MESFET

So what is the MESFET good for?

We have already said that it has a higher mobility channel leading to a higher transconductance and transit frequency of the device. It also has very low noise.

This is enhanced even further if we use a materials such as GaAs with a higher electron mobility (x5) and saturation velocity (x2)than Si. It also can be used with semi-insulating GaAs substrates to reduce substrate parasitics.

Applications:
low noise
microwave
amplification

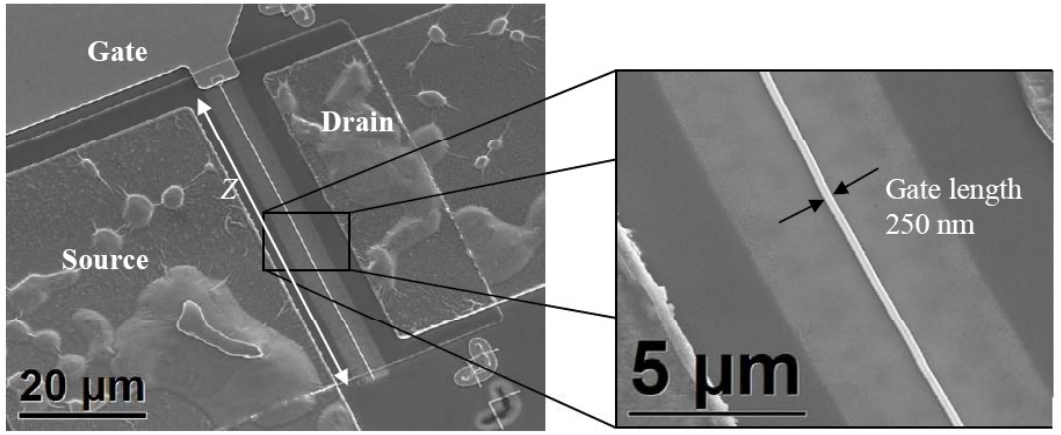


MESFET

Also good for power FET applications, as the power capability scales well with gate width. Particularly useful if based on a wide-band gap semiconductor such as SiC or GaN.

- High Breakdown Voltages
- High Electrical and Thermal Conductivity
- High Saturated Electron Velocity
- High stability and melting point

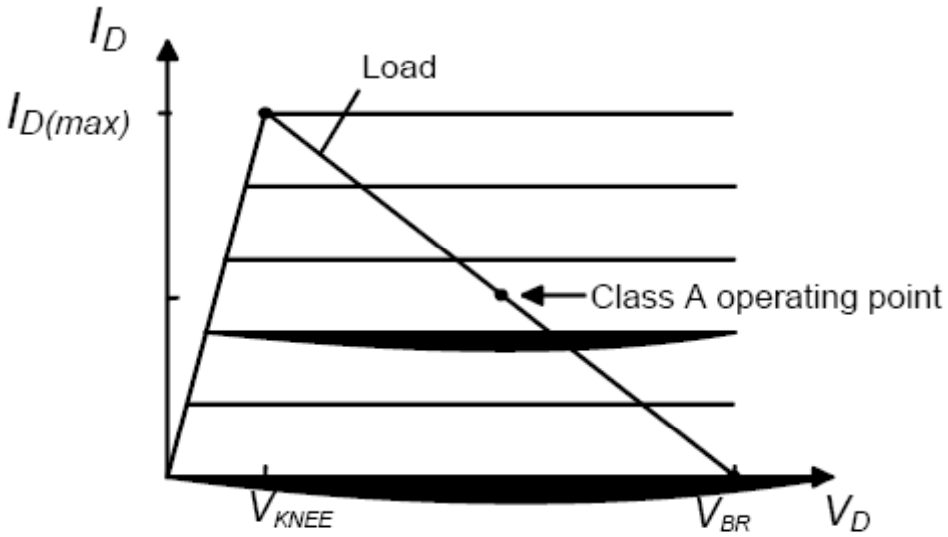
Example: GaN power FET (Sheffield). Short gate length for high speed but long gate width to increase drain current



Power MESFET

A large drain current I_D and high drain voltage V_D are needed for high power operation. V_D needs to be higher that V_{KNEE} and is limited by the breakdown voltage V_{BR}

Maximum output power $$P_{out} = \frac{(V_{Br} - V_{Knee}).I_{D(Max)}}{8}$$



Increase Power:

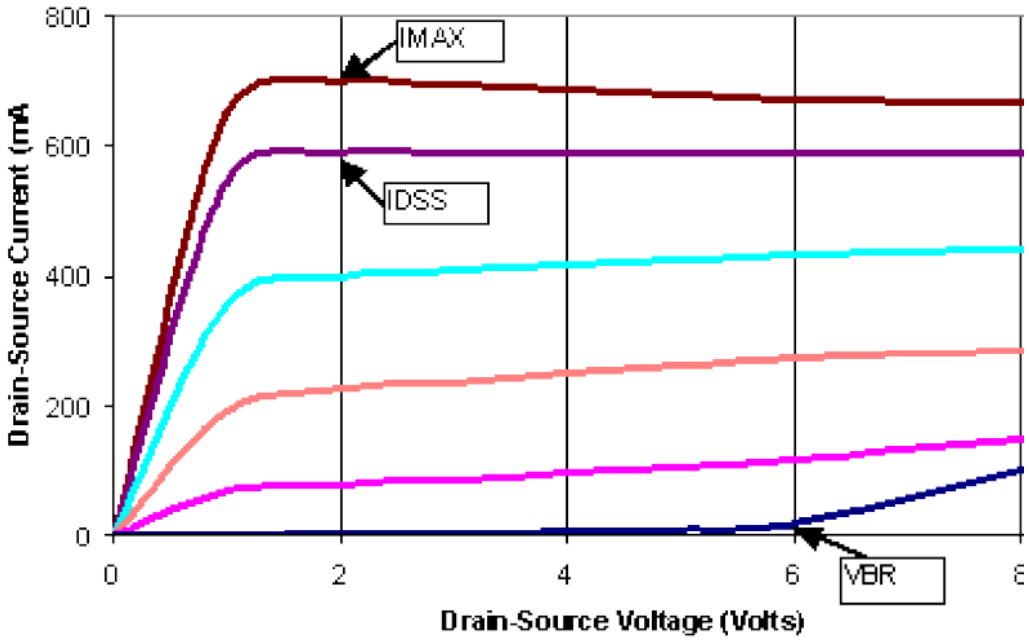
Increase I_D →High mobility, large gate widths)

Increase V_{Br} → Wide gap materials

Reduce V_{Knee} → not trivial (device scaling)

Power MESFET

V_{knee} : Knee voltage: the voltage at which the curves transition from "linear" to "saturation i.e: where the current reaches I_{DSS}



V_{BR}: Gate-drain breakdown voltage

At high drain-source potential and near pinch-off, the IV curves tend to bend up. A pinch-off you have a high gate voltage and a high drain source voltage (in the figure it is 10 volts (VGS=-4 volts and VDS=6 volts. The E-field is sufficiently high to accelerate electrons to an energy at which they will undergo impact ionisation

Power MESFET

Long linear gates are not practical → better to use interdigital gate

A typical power MESFET may have several gate 'fingers' (as well as source-drain fingers) to scale up the drain current