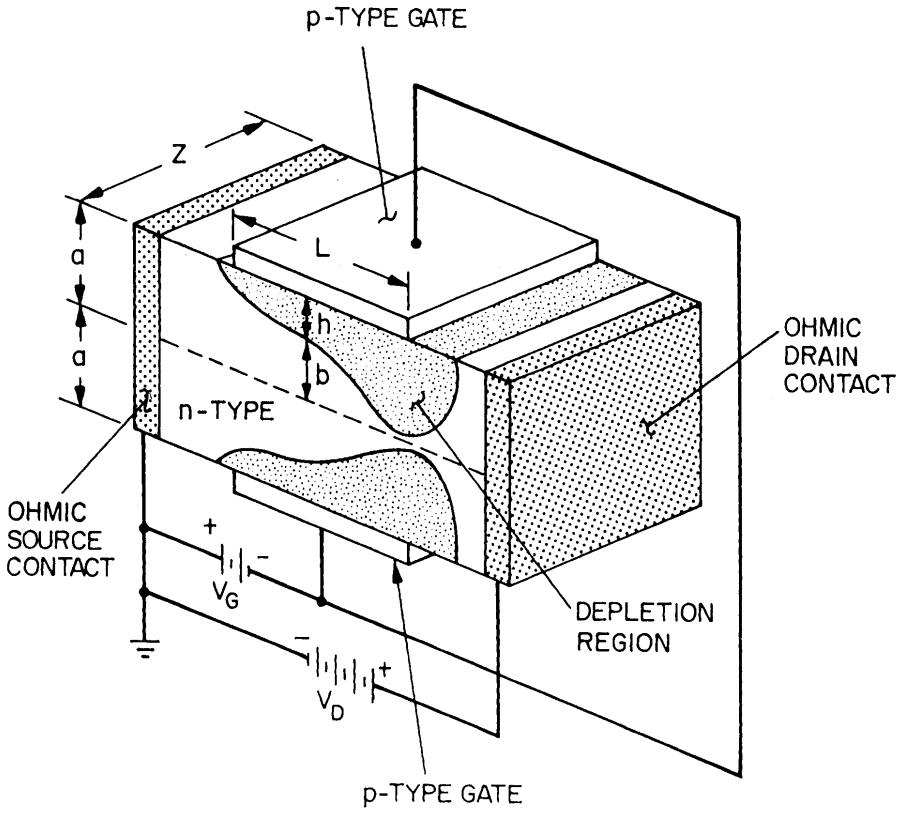
IX.4. Field Effect Transistors

Field Effect Transistors (FETs) utilize a conductive channel whose resistance is controlled by an applied potential.

1. Junction Field Effect Transistor (JFET)

In JFETs a conducting channel is formed of *n* or *p*-type semiconductor (GaAs, Ge or Si).
Connections are made to each end of the channel, the Drain and Source.

In the implementation shown below a pair of gate electrodes of opposite doping with respect to the channel are placed at opposite sides of the channel. Applying a reverse bias forms a depletion region that reduces the cross section of the conducting channel.



(from Sze)

Changing the magnitude of the reverse bias on the gate modulates the cross section of the channel.

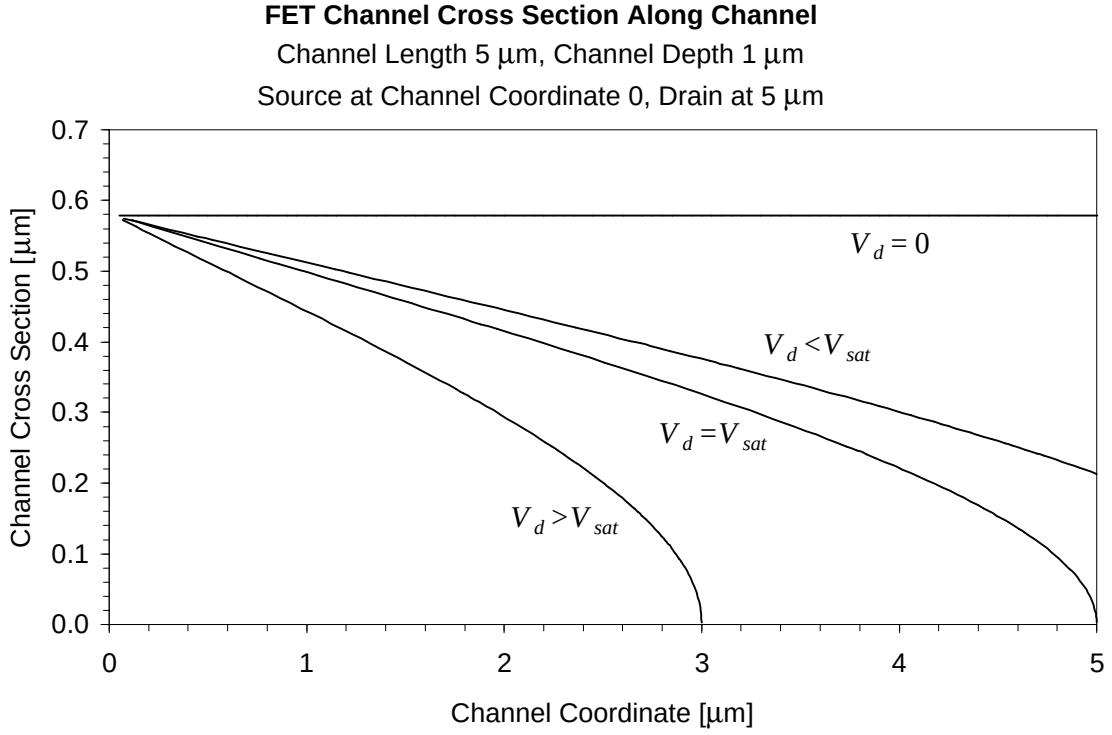
First assume that the drain voltage is 0.
Increasing the reverse gate potential will increase the depletion width, i.e. reduce the cross section of the conducting channel, until the channel is completely depleted. The gate voltage where this obtains is the “pinch-off voltage” V_P.

Now set both the gate and drain voltages to 0. The channel will be partially depleted due to the “built-in” junction voltage.

Now apply a drain voltage. Since the drain is at a higher potential than the source, the effective depletion voltage increases in proximity to the drain, so the width of the depletion region will increase as it approaches the drain.

If the sum of the gate and drain voltage is sufficient to fully deplete the channel, the device is said to be “pinched off”.

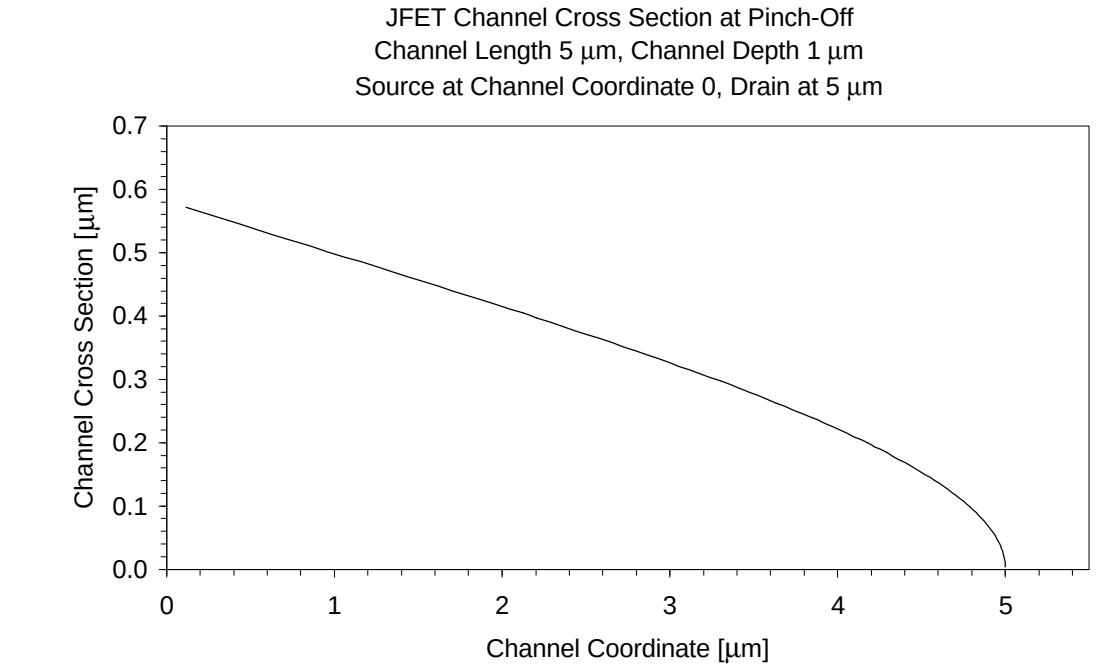
Increasing the drain voltage beyond this point moves the pinch-off point towards to the source.

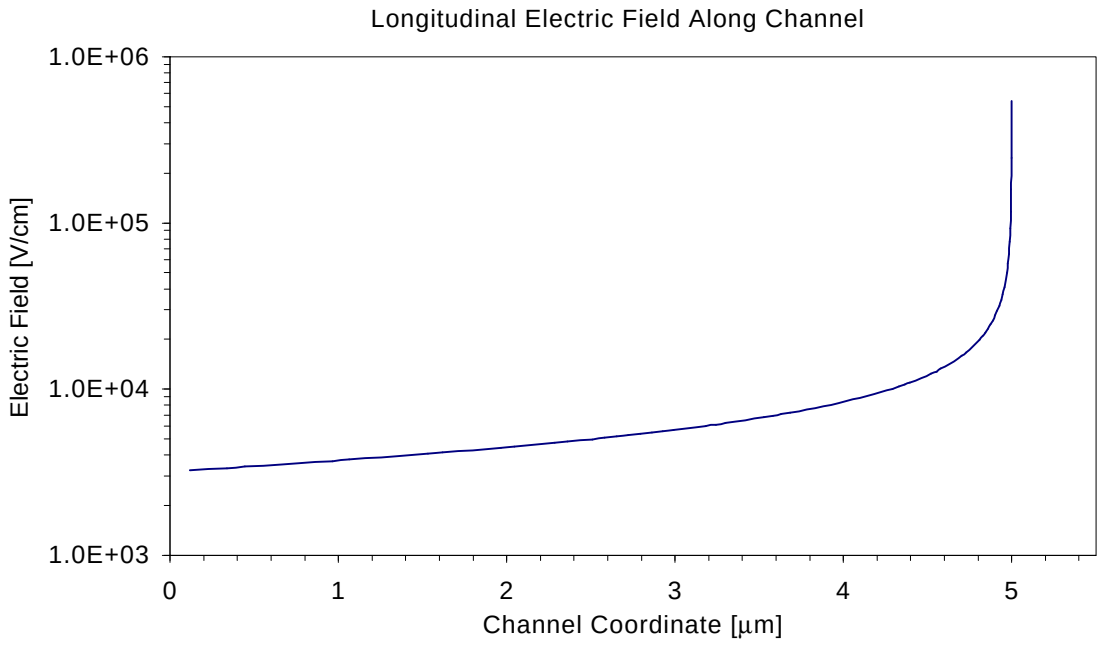


Pinching off the channel does not interrupt current flow. All thermally excited carriers have been removed from the depleted region, but carriers from the channel can still move through the potential drop to the drain.

The profile of the depletion region is not determined by the static potentials alone.

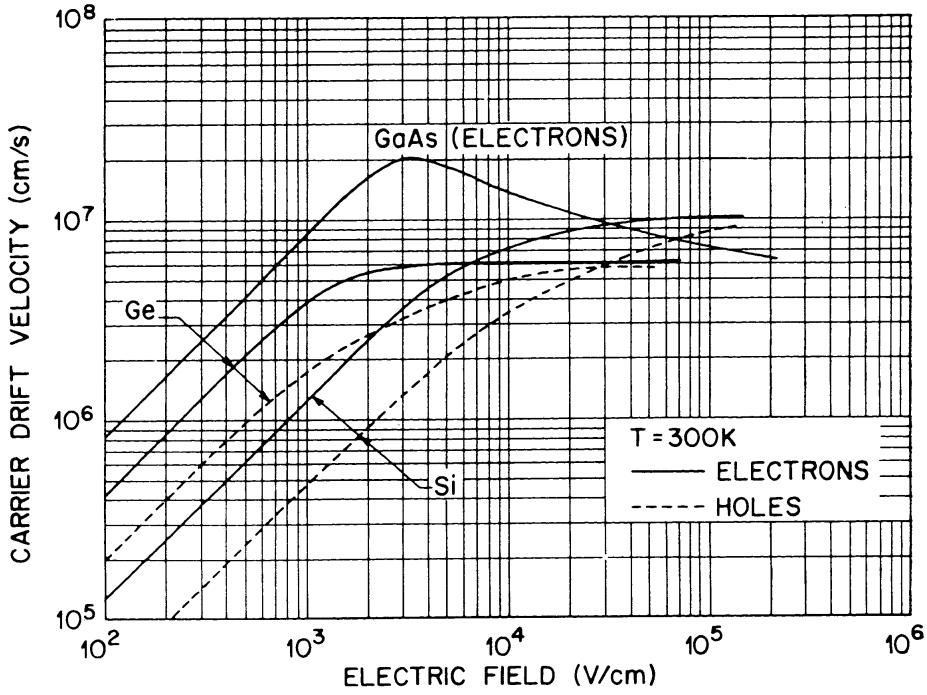
Current flow along the channel changes the local potential.
As the channel cross section decreases, the incremental voltage drop increases, i.e. the longitudinal drift field that determines the carrier velocity increases.





At high electric fields the mobility decreases.

This comes about because as the carriers' velocity increases they begin to excite optical phonons. At fields above 10^5 V/cm practically all of the energy imparted by an increased field goes into phonon emission.



(from Sze)

Since the velocity saturates at high fields, the current

$$I = N_C q_e v$$

also saturates, since the number of carriers N_C remains constant, i.e. at high fields silicon acts as an incremental insulator ($dI/dV = 0$).

As the drain voltage is increased beyond pinch-off, the additional voltage decreases the length of the resistive channel, but also increases the potential drop in the drain depletion region.

$\Rightarrow$ The current increases only gradually with drain voltage.

Current Voltage Characteristics

Low drain and gate voltages:

The resistive channel extends from the source to the drain.

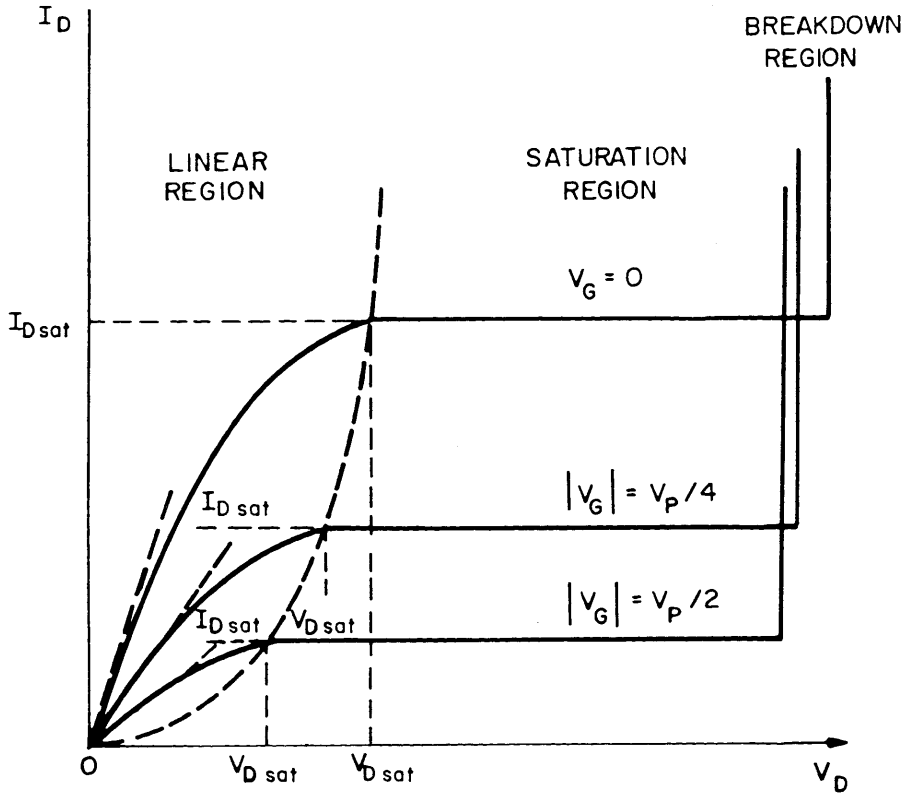
The drain current increases linearly with drain voltage.

"Linear Region"

Gate and drain voltages sufficiently high to pinch off the resistive channel:

The drain current remains constant with increasing drain voltage.

"Saturation Region"



(from Sze)

The drain saturation voltage V_{Dsat} increases as the gate voltage is changed from the static pinch off voltage V_p towards 0.

For use in amplifiers the characteristics in the saturation region are of the most interest.

For a uniform dopant distribution in the channel the dependence of drain current on gate voltage is

$$I_D = I_{DSS}\left(1 - 3\left(\frac{V_G + V_{bi}}{V_P}\right) + 2\left(\frac{V_G + V_{bi}}{V_P}\right)^{3/2}\right)$$

where the drain saturation current

$$I_{DSS} = \frac{1}{6\varepsilon}\mu(q_e N_{ch})^2 d^3 \frac{W}{L}$$

is determined by the carrier mobility μ, the doping level in the channel N_{ch} and the channel depth d, width W and length L. ε is the dielectric constant.

In reality the dopant distribution is not uniform, but the I_D-V_G dependence does not change much with dopant distribution.
To a good approximation

$$I_D = I_{DSS}\left(1 - \left(\frac{V_G + V_{bi}}{V_P}\right)\right)^2$$

The transconductance

$$g_m = \left|\frac{dI_D}{dV_G}\right| = \frac{2I_{DSS}}{V_P}\left(1 - \frac{V_G + V_{bi}}{V_P}\right)$$

The transconductance is maximum for V_G= 0, i.e. maximum drain current, and for small pinch off voltages. Then

$$g_m\big|_{V_G=0} \approx \frac{2I_{DSS}}{V_P}$$

Since the pinch off voltage is simply the depletion voltage of a diode with thickness d, the pinch-off voltage is

$$V_P = \frac{q_e}{2\varepsilon} N_{ch} d^2 - V_{bi}$$

Since we want to see how device parameters affect the transconductance, we'll ignore the built-in voltage since it varies only weakly with doping $(k_B T/q_e)\log(N_{ch}/n_i)$.

With this approximation

$$g_m\big|_{V_G=0} \approx \frac{2I_{DSS}}{V_P} = \frac{W}{L}\frac{\mu(q_e N_{ch})^2 d^3}{3q_e N_{ch} d^2} \propto \frac{W}{L}\mu N_{ch} d$$

Obviously, a high carrier mobility will increase the transconductance, since for a given carrier concentration this will increase the magnitude of the current.

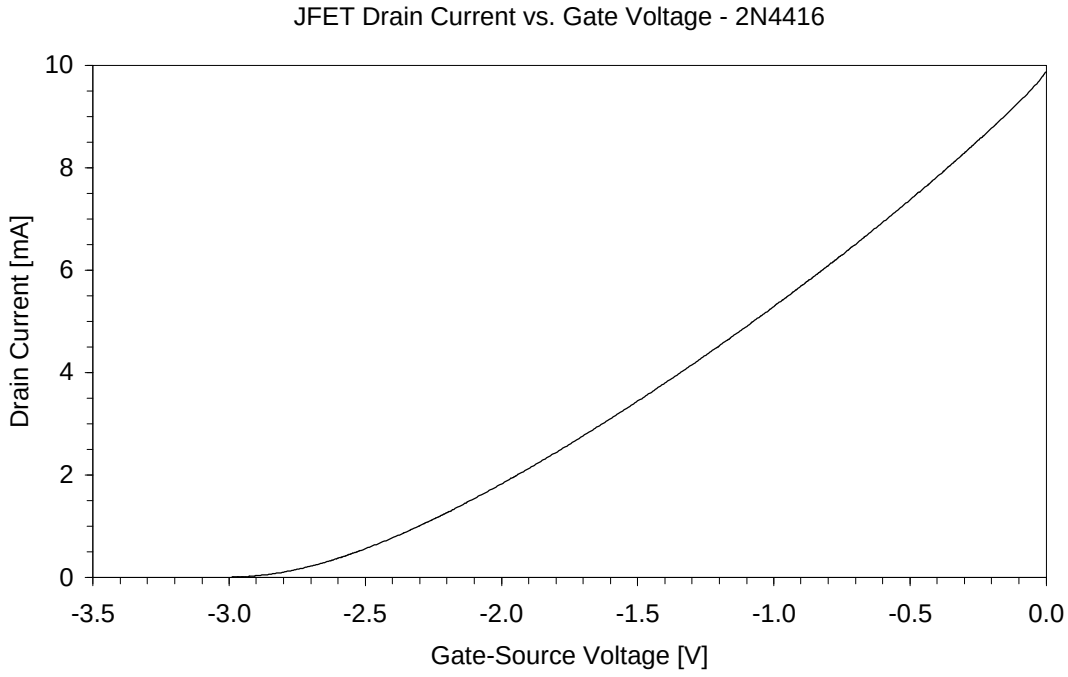
1. The proportionality of transconductance to width W is trivial, since it is equivalent to merely connecting device in parallel, so the normalized transconductance g_m/W is used to compare technologies.
2. The normalized transconductance increases with the number of carriers per unit length $N_{ch}d$ and decreasing channel length L.
3. Transconductance incraeses with drain current

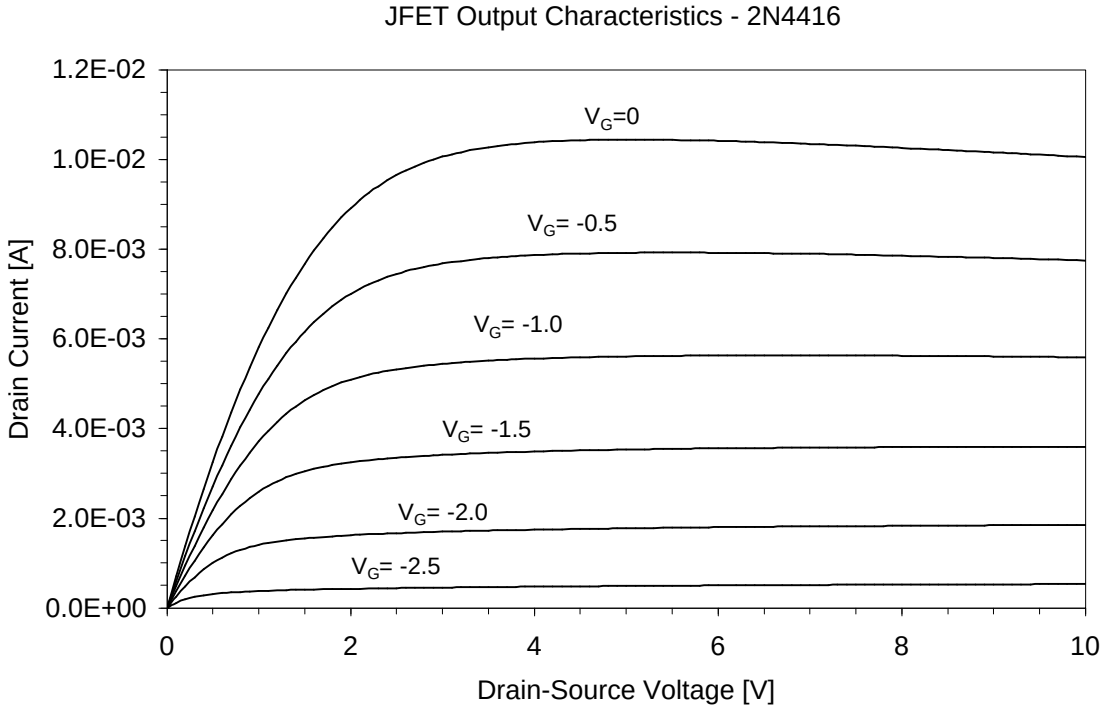
$$g_m = \left|\frac{dI_D}{dV_G}\right| = \frac{2I_{DSS}}{V_P}\left(1 - \frac{V_G + V_{bi}}{V_P}\right) = \frac{2\sqrt{I_{DSS}}}{V_P}\sqrt{I_D}$$

   i.e. drain current is the primary parameter; the applied voltages are only the means to establish I_D.

All of these optimizations also increase the power dissipation. For low power systems optimization is more involved.

Measured JFET Characteristics



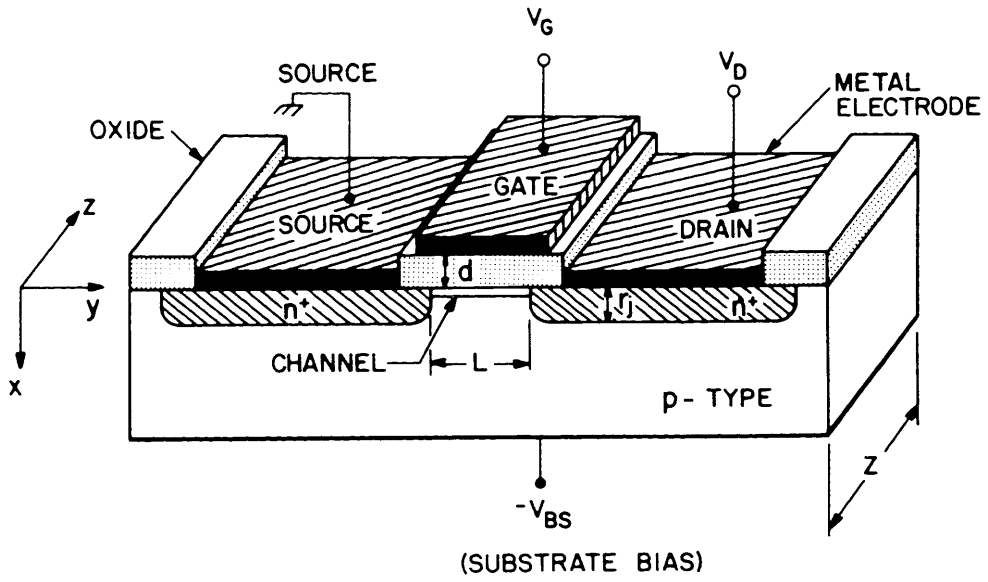


The slight drop in drain current for V_G=0 is due to self-heating.

Metal Oxide Field Effect Transistors (MOSFETs)

Both JFETs and MOSFETs are conductivity modulated devices, utilizing only one type of charge carrier. Thus they are called unipolar devices, unlike bipolar transistors, for which both electrons and holes are crucial.

Unlike a JFET, where a conducting channel is formed by doping and its geometry modulated by the applied voltages, the MOSFET changes the carrier concentration in the channel.



(from Sze)

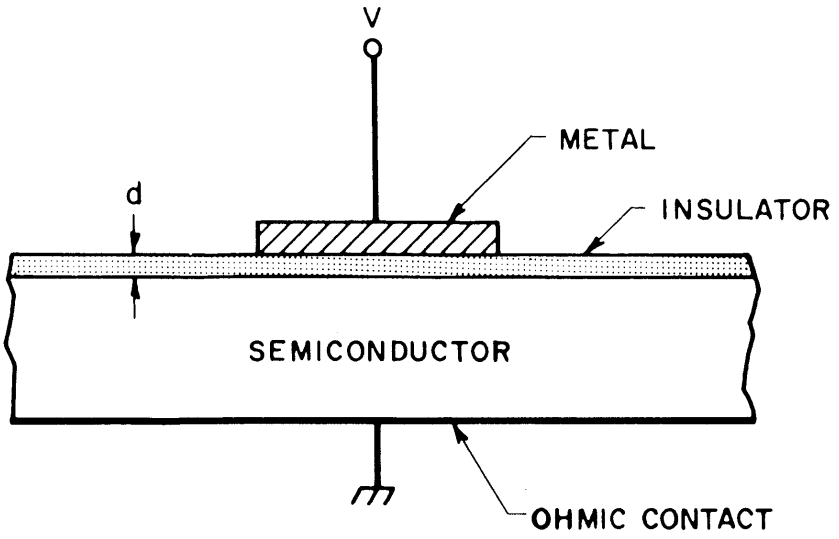
The source and drain are n^+ regions in a p-substrate.

The gate is capacitively coupled to the channel region through an insulating layer, typically SiO_2.

Applying a positive voltage to the gate increases the electron concentration at the silicon surface beneath the gate.

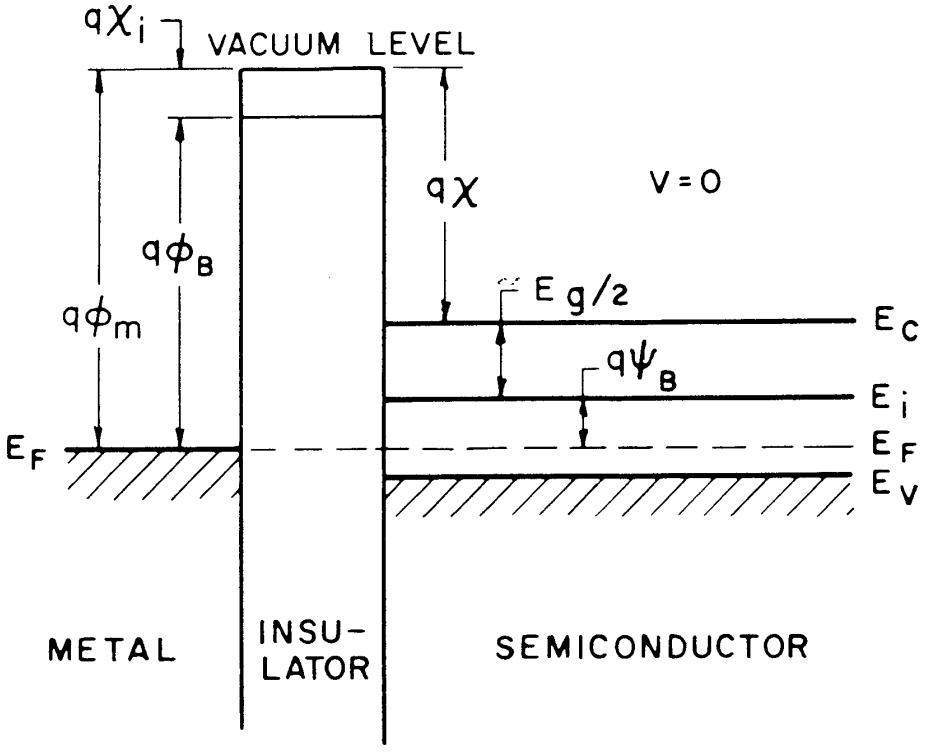
As in a JFET the combination of gate and drain voltages control the conductivity of the channel.

Formation of the Channel - The MOS Capacitor



(from Sze)

Band structure in an ideal MOS capacitor on a *p*-Substrate

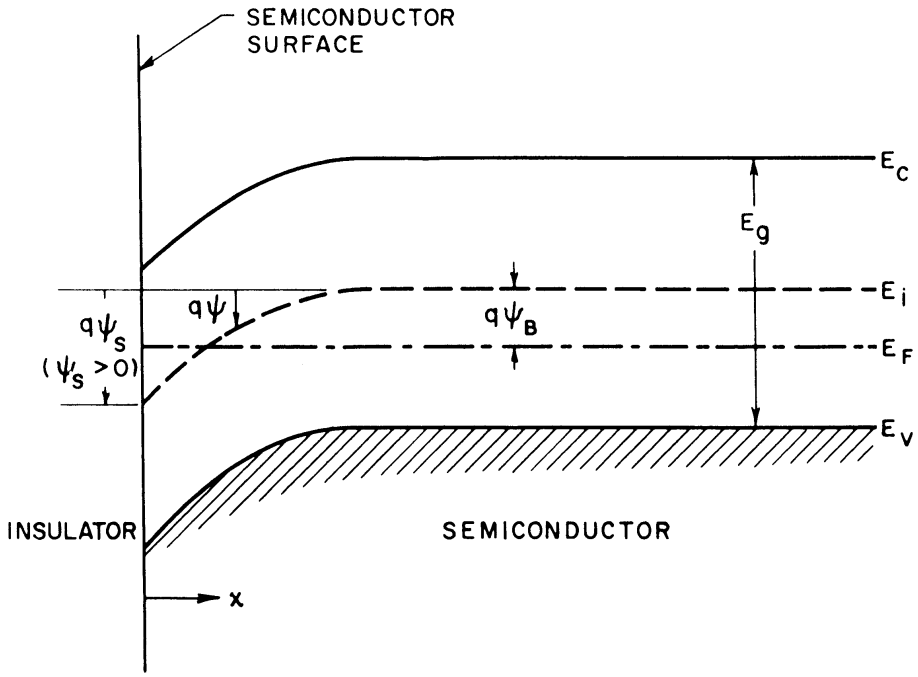


(from Sze)

In its natural state, however, the band structure is not flat as just shown.

The discontinuity in the crystal structure and charge trapped at the surface change the potential at the surface, so the bands bend.

Energy band diagram of a *p*-type semiconductor



(from Sze)

As shown above the surface potential Ψ_s is positive.

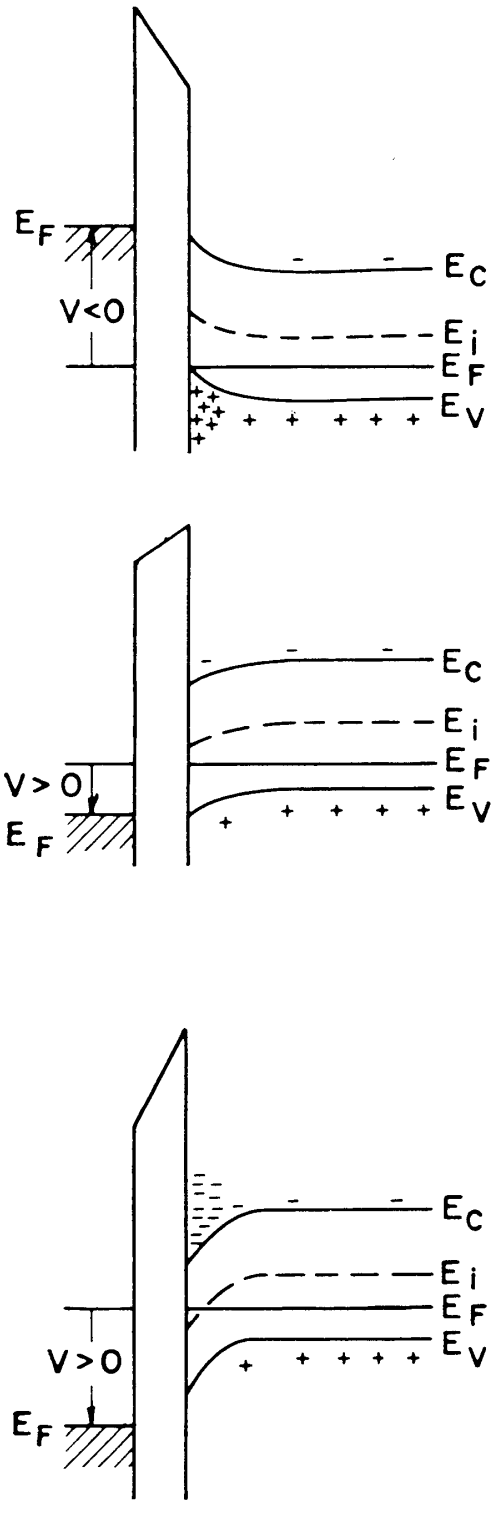
$\Psi_S < 0$ the bands bend upwards, increasing the hole concentration at the surface.

$\Psi_B > \Psi_S > 0$ the bands bend slightly downwards, reducing the concentration of holes at the surface (depletion)

$\Psi_S > \Psi_B$ the conduction band edge dips below the Fermi level, leading to an accumulation of electrons at the surface (inversion)

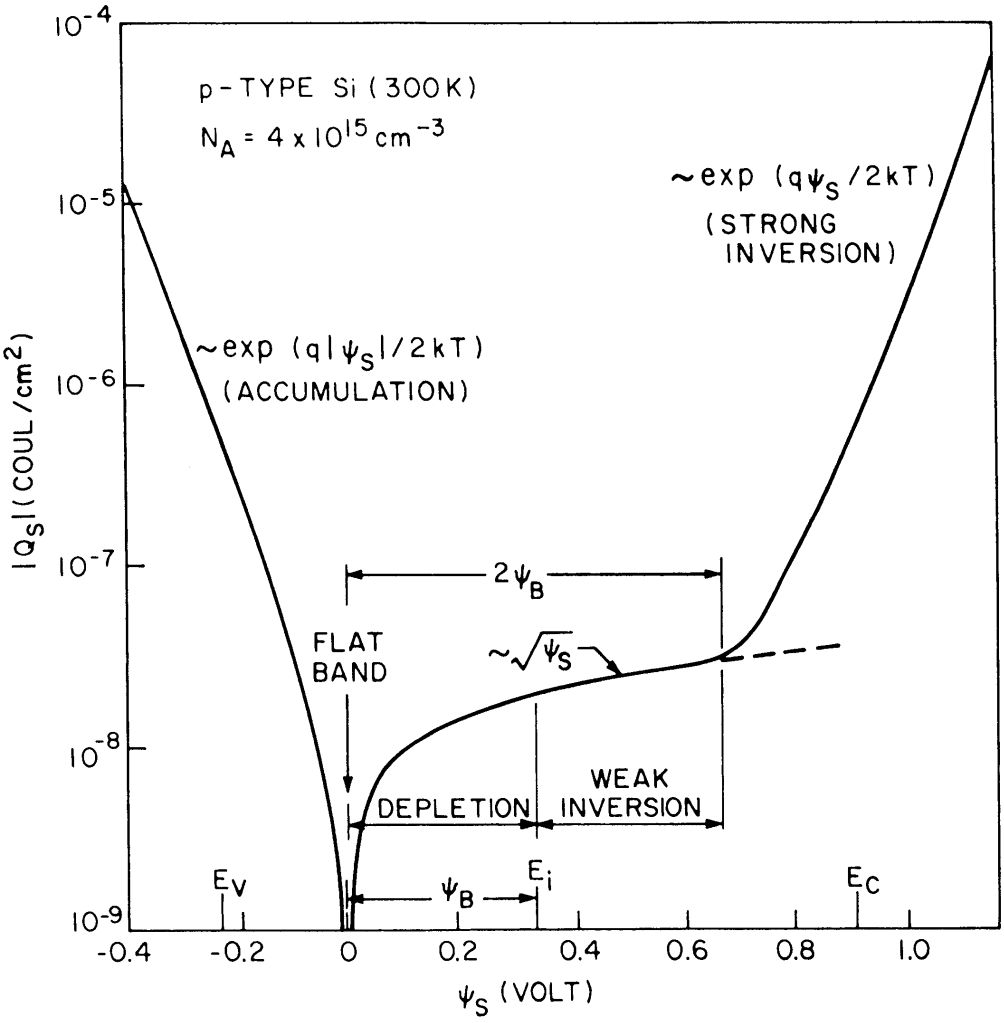
In the absence of any special surface preparation the surface of silicon is *n*-type, i.e. *p*-type silicon inverts at the surface.

Surface concentration vs. band-bending in *p*-type material



(from Sze)

Variation of space-charge density vs. surface potential



(from Sze)

When Ψ_S =0, the bands are flat.

As Ψ_S is made more negative the surface concentration of holes increases exponentially (“accumulation”).

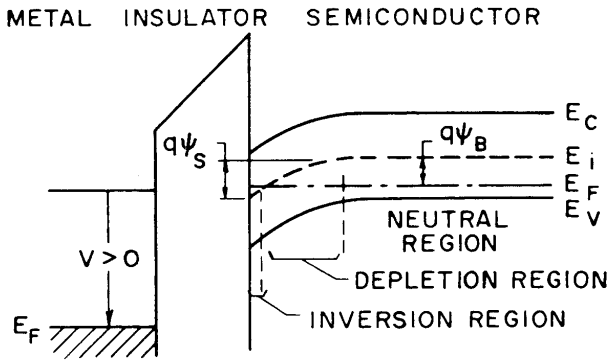
As Ψ_S is increased from the flat-band value the hole concentration decreases - the surface depletes.

When $\Psi_S > \Psi_B$ the surface begins to invert, but the electron concentration is less thann or of the order of the bulk hole concentration (“weak inversion”)

For $\Psi_S > 2\Psi_B$ the electron concentration much exceeds the bulk hole concentration (“strong inversion”).

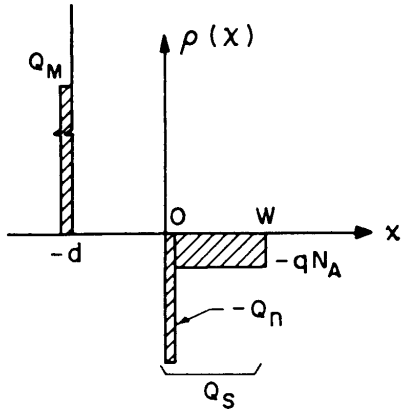
Application of a voltage to the gate deposits a charge on the electrode, which must be balanced by an equal, but opposite, charge in the bulk.

Immediately at the Si-SiO_2 interface the surface is inverted, but farther into the bulk the hole concentration is depleted before attaining the bulk equilibrium concentration.

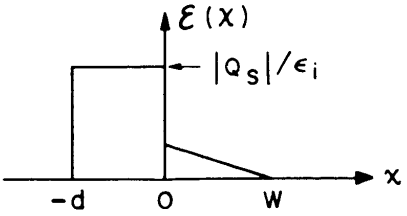


The bulk charge that balances the applied charge is made up of

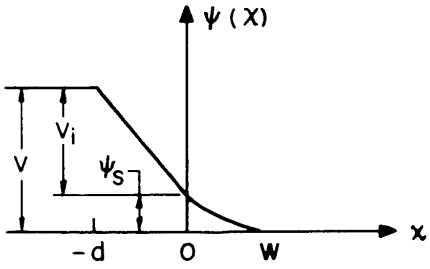
a) the (negative) electron charge in the inversion layer
b) the (negative) acceptor charge in the depletion region



The electric field is uniform in the insulator and falls linearly with depth in the depletion region

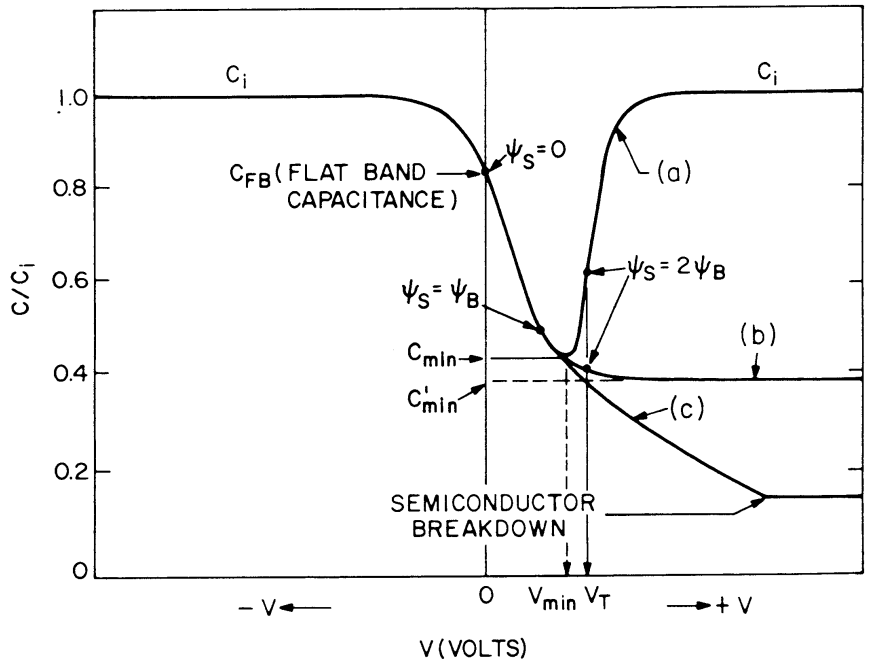


Correspondingly, the potential drops linearly in the insulator and quadratically in the depletion zone.



(from Sze)

Capacitance vs. Voltage of an MOS Diode



(from Sze)

Capacitance is measured by superimposing a small AC signal on the DC voltage and measuring the alternating current. The capacitance is measured at low and high frequencies.

At large negative voltages holes accumulate at the surface, so the capacitance is determined by the thickness of the insulator and remains constant.

With increasing voltage the surface concentration decreases and a depletion layer forms, leading to a decrease in capacitance.

Increasing the DC voltage beyond V_{min} yields a capacitance that depends on frequency:

a) at low frequencies the capacitance increases as an inversion layer forms at the oxide-silicon interface, so the capacitance approaches the oxide capacitance.

b) at high frequencies the recombination-generation rates are not fast enough for the charge to respond, so the capacitance remains at the depletion value.

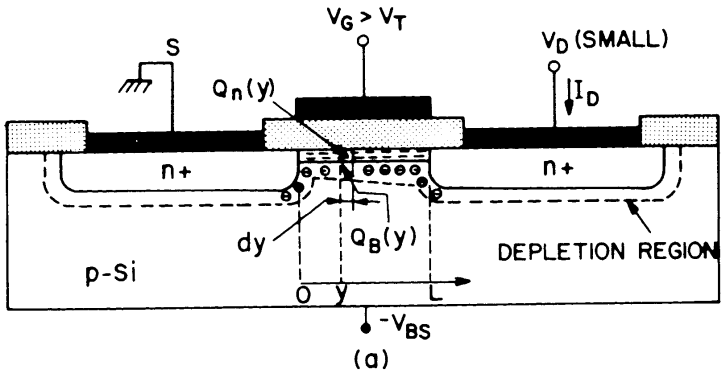
c) For very short pulses equilibration doesn't take place even in depletion, so the capacitance decreases further.

An *n*-channel MOSFET utilizes an *n*-channel in a *p*-substrate, so application of a positive potential to the gate forms the inversion layer needed for the channel.

As in the JFET, the combination of current flow in the channel and the applied potentials forms a depletion region that is greatest near the drain. At a sufficiently large drain potential the channel “pinches off”.

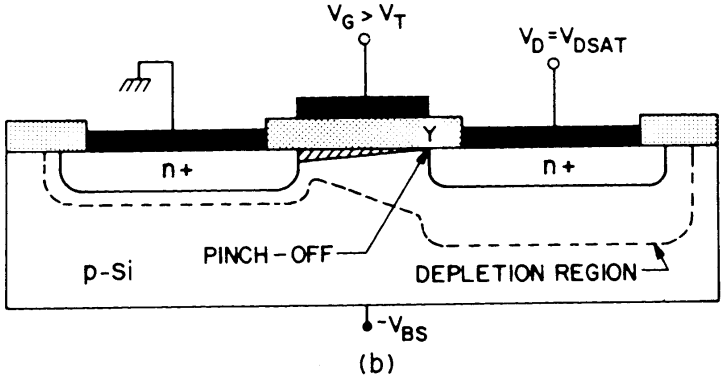
Low drain voltage
$V_D < V_{sat}$

Resistive channel



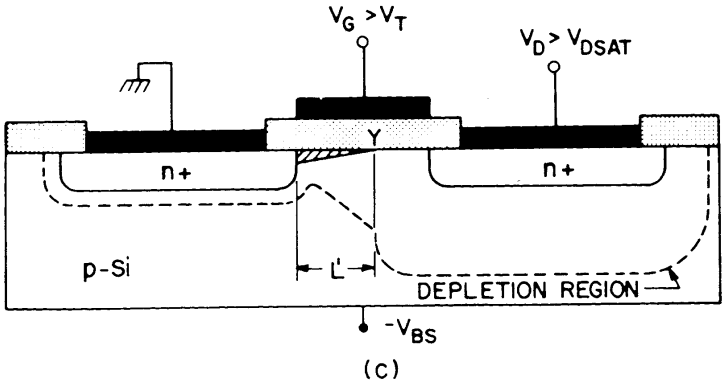
$V_D = V_{sat}$

Onset of current saturation



$V_D > V_{sat}$

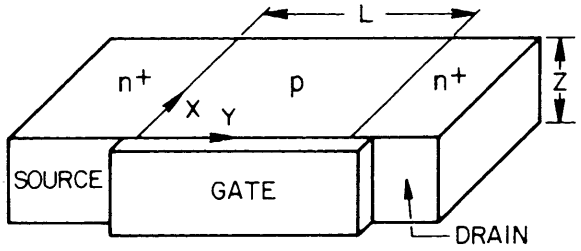
Output current saturated



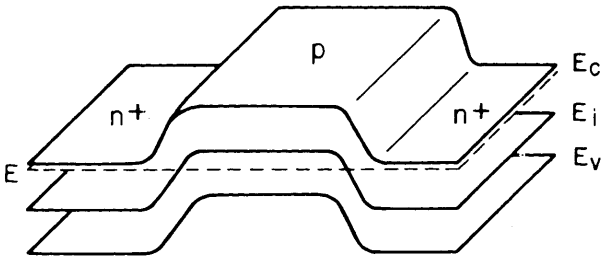
(from Sze)

Three-Dimensional Potential Diagram of an *n*-Channel MOSFET

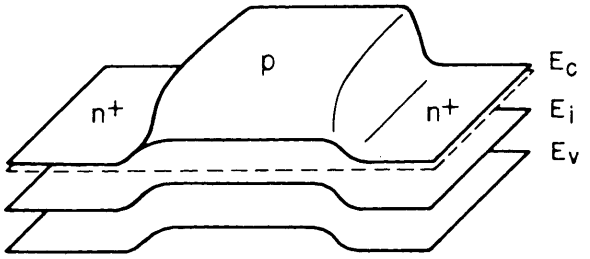
Device orientation



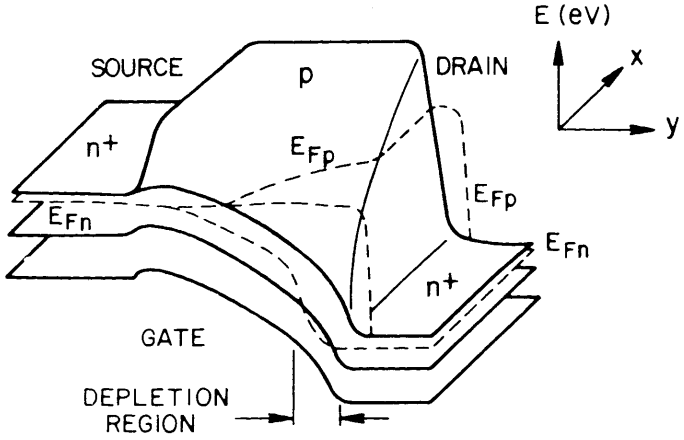
Flat-band condition



Formation of inversion layer



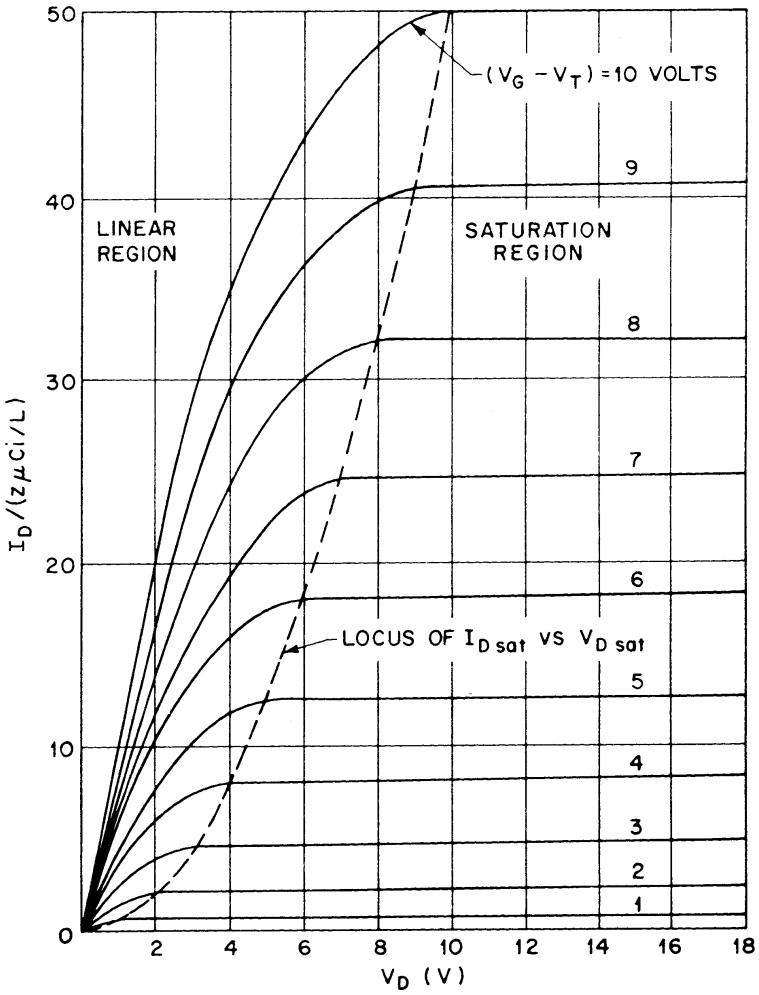
Drain voltage beyond saturation



(from Sze)

The output curves of a MOSFET are similar to a JFET.

The drain voltage required to attain saturation increases with the operating current.



(from Sze)

In saturation

$$I_D = \frac{W}{L}\frac{\mu\, C_i}{2}(V_G - V_T)^2$$

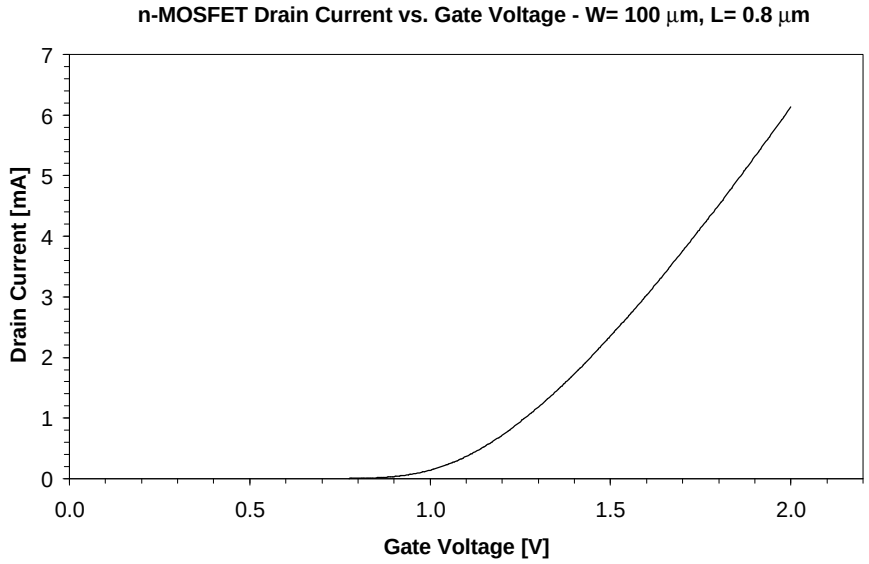
where C_i is the gate capacitance per unit area ε_{ox}/d_{ox} and V_T is the gate voltage corresponding to the onset of strong inversion ("threshold voltage")

From this the transconductance is

$$g_m = \frac{W}{L} C_i \mu\,(V_G - V_T) = \frac{W}{L}\frac{\varepsilon_{ox}}{d_{ox}}\mu\,(V_G - V_T) = \sqrt{\frac{W}{L}\cdot\frac{\varepsilon_{ox}}{d_{ox}}\mu \cdot I_D}$$

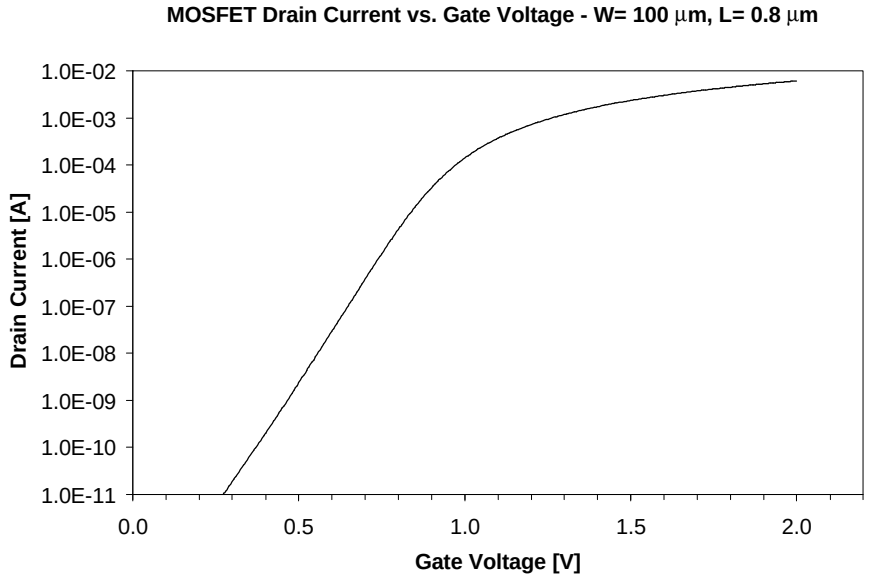
For a given width W and drain current I_D the transconductance is increased by decreasing the channel length L and the thickness of the gate oxide d_{ox}.

Measured characteristics of an *n*-channel MOSFET with 0.8 μm channel length and 20 nm gate oxide thickness



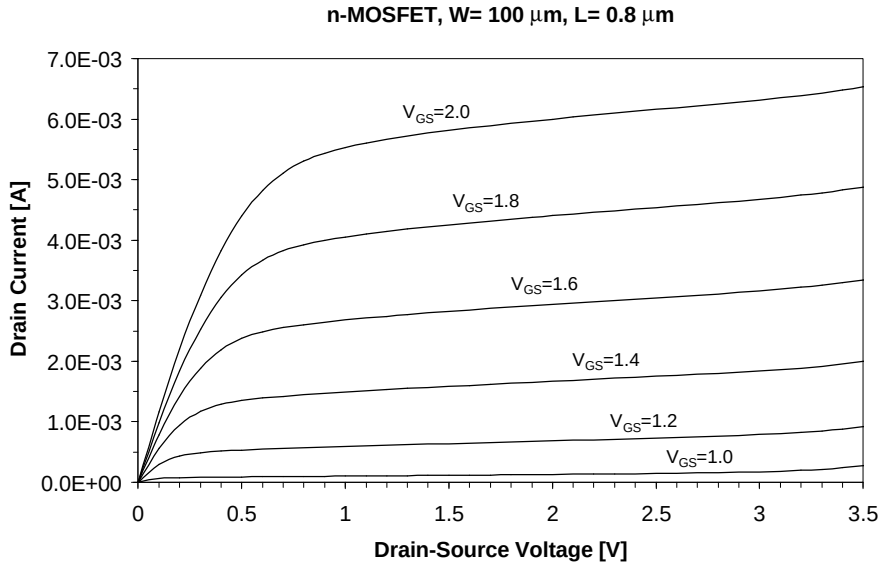
For this device the threshold voltage V_T is about 1.2 V.

The transition from weak to strong inversion becomes more apparent in a logarithmic plot.



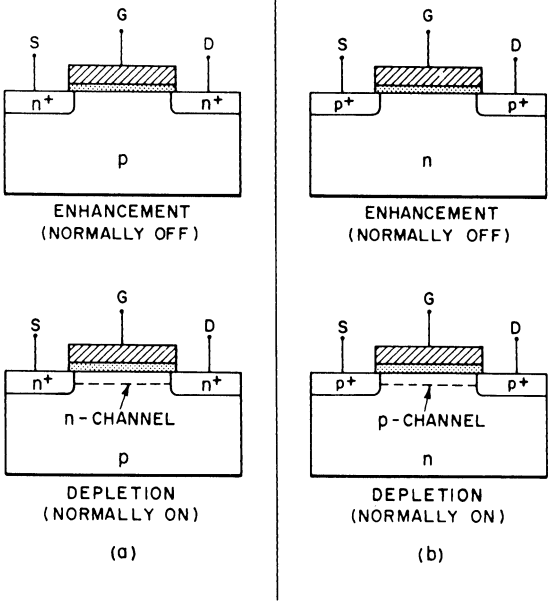
In the subthreshold regime the drain current is proportional to the inversion carrier concentration, i.e. it increases exponentially with gate voltage.

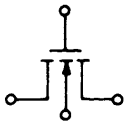
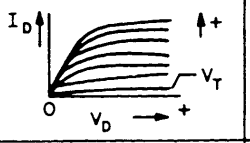
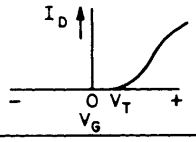
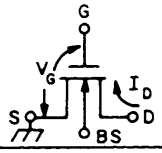
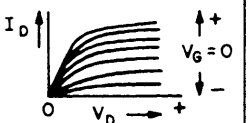
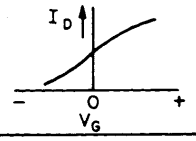
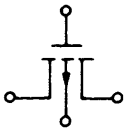
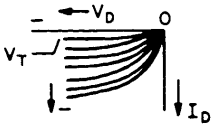
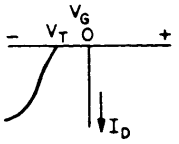
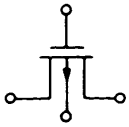
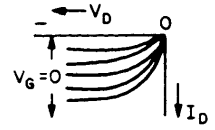
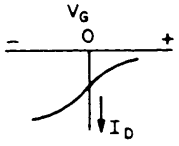
The increase in output saturation voltage with gate voltage can be seen clearly in the output curves.



Depending on the substrate doping MOSFETs can be implemented with either *n* or *p*-channels.

A thin surface layer can be implanted to adjust the threshold voltage. With this devices can be normally on at zero gate voltage (depletion mode) or normally off, i.e. require additional voltage to form the inversion layer (enhancement mode), as illustrated below

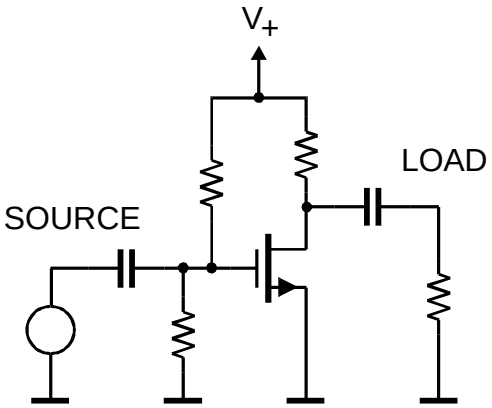


TYPE	ELECTRICAL SYMBOL	OUTPUT CHARACTERISTIC	TRANSFER CHARACTERISTIC
N-CHANNEL ENHANCEMENT (NORMALLY OFF)			
N-CHANNEL DEPLETION (NORMALLY ON)			
P-CHANNEL ENHANCEMENT (NORMALLY OFF)			
P-CHANNEL DEPLETION (NORMALLY ON)			

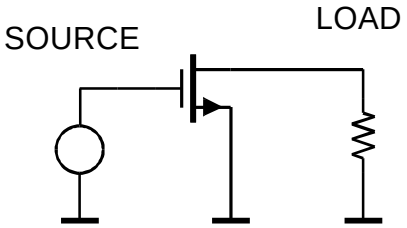
MOS Transistors in Amplifiers

As shown for BJTs, three different circuit configurations are possible:

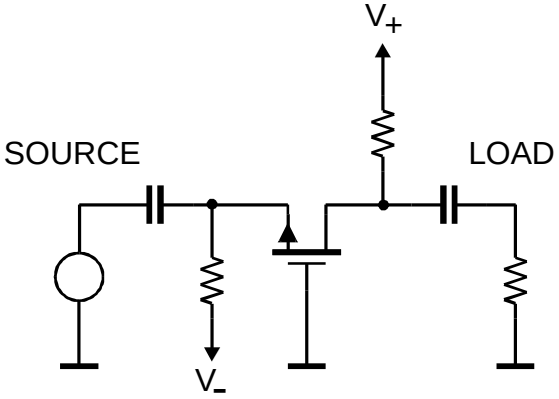
1. Common Source

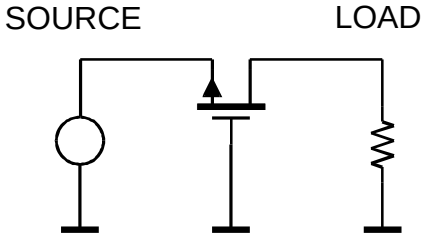


Equivalent Circuit

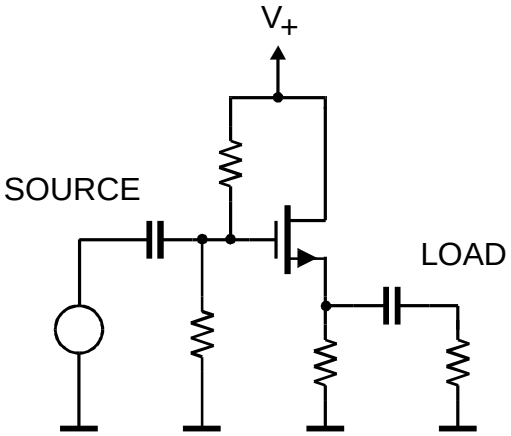


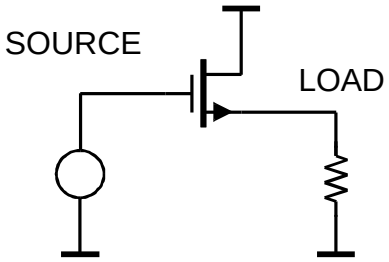
2. Common Gate





3. Common Drain (Source Follower)

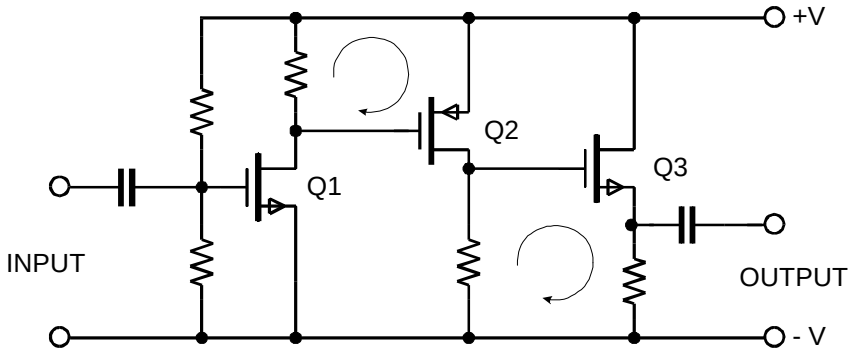




Advantages of MOS/CMOS

- high input resistance (capacitive gate)
- complementary devices (NMOS, PMOS)
- magnitude of gate voltages allows simple direct coupling
- tailoring of device characteristicschoice by choice of geometry (W, L)
- low-power, high-density logic circuitry (CMOS)

Amplifier cascade using NMOS and PMOS devices



Drawbacks

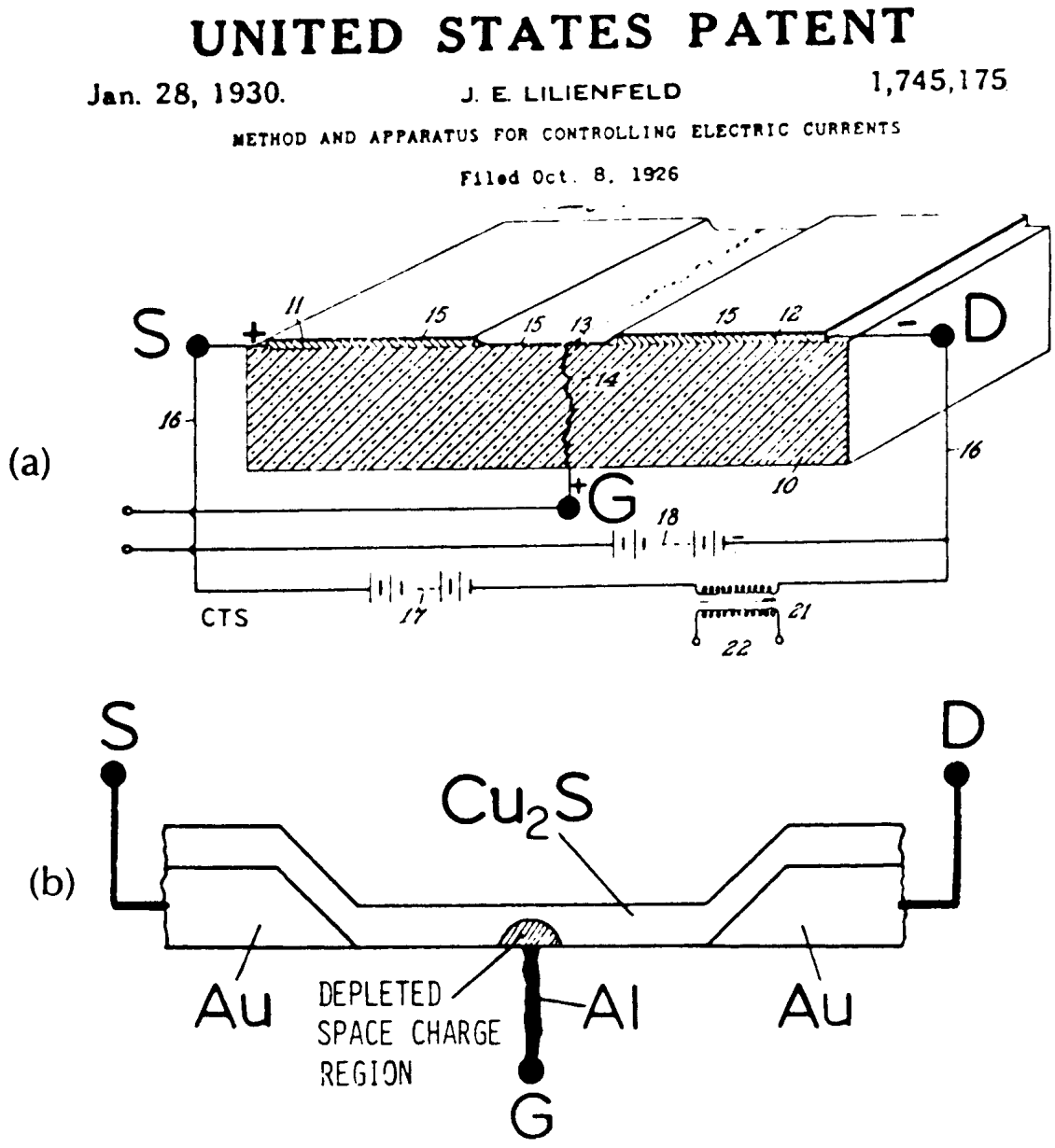
relative to bipolar transistors …

- more current for given transconductance (speed)
- more current for given noise level in detector circuits using short shaping times
- inferior device matching
- analog characteristics less predictable (more difficult to model)

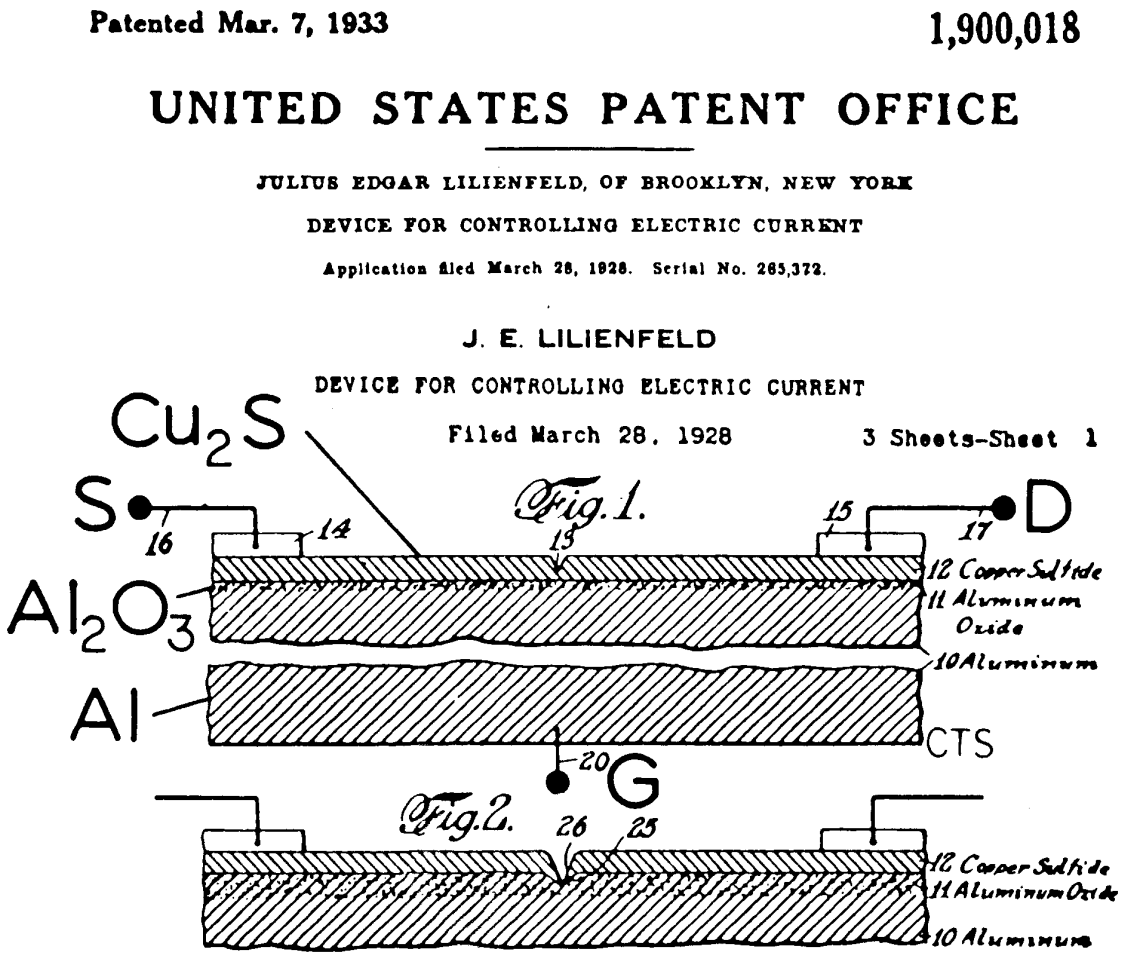
Modern fabrication processes combine bipolar transistor and MOS technology to exploit best features of both (BiCMOS).

… So what else is new?

The first patent awarded for a junction field effect transistor was submitted in 1926



In 1928 Lilienfeld submitted a patent application for a MOSFET



Patented Mar. 7, 1933 1,900,018

UNITED STATES PATENT OFFICE

JULIUS EDGAR LILIENFELD, OF BROOKLYN, NEW YORK

DEVICE FOR CONTROLLING ELECTRIC CURRENT

Application filed March 28, 1928. Serial No. 265,372.

J. E. LILIENFELD

DEVICE FOR CONTROLLING ELECTRIC CURRENT

Filed March 28, 1928 3 Sheets-Sheet 1

Although Lilienfeld appears to have fabricated prototypes, the results were not reproducible, because surface states and impurity levels could not be controlled.

Furthermore, unknown to everyone at the time, the dynamics of electrons and holes and practically all of semiconductor physics had yet to be understood.

Nevertheless, these concepts provided the impetus for the research that lead to the first

bipolar transistor in 1947,
JFET in 1953 and
practical Si MOSFETs in 1960.