PHYS: The current versus voltage equations for an enhancement mode, n-type Metal-Oxide Semiconductor (NMOS) Field Effect Transistor (FET) are derived here.

Fig. 1 shows the circuit symbol for an NMOS FET.

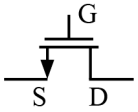


Fig. 1. Schematic symbol for NMOS FET.

The FET has three terminals:

1) The Source (S), which is the source of electrons that flow through the FET,
2) The Drain (D), which passes electrons from the source out of the FET, and
3) The Gate (G), which controls the conductivity of the channel between the Source and Drain via its voltage.

A key property of the FET is that the gate is isolated by a dielectric from the channel between the Source and Drain. The gate and the channel form a small capacitor and, once the capacitor is charged, no more current flows into the Gate. Consequently, the FET is ideal for circuits that require a high input impedance.

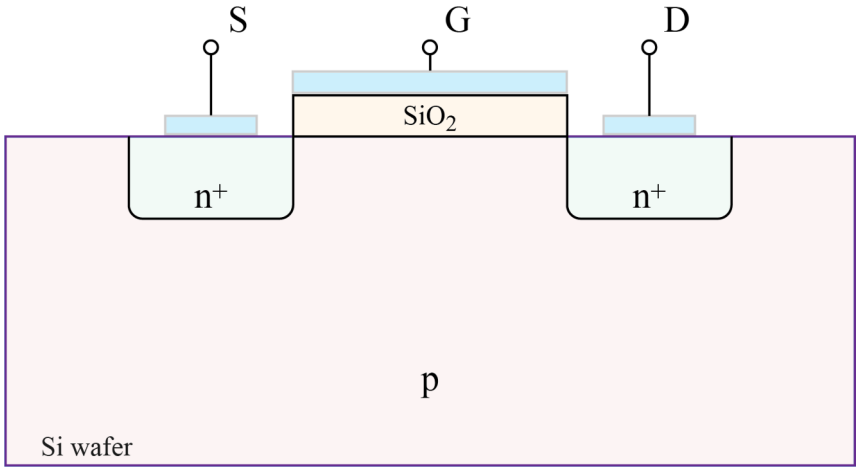
The FET is usually used as an electrically controlled switch. When the Gate voltage is high enough and the Drain to Source voltage is high enough, the FET passes a constant current that may be quite large. This means the FET can translate a low-current input at the gate to a high-current output.

Because the relationship between the current in the channel between Drain and Source is a nonlinear function of the Gate voltage, the FET is challenging to use in simple analog circuits. This property is another reason why FETs, especially discrete ones, are well suited as switches that are on or off.

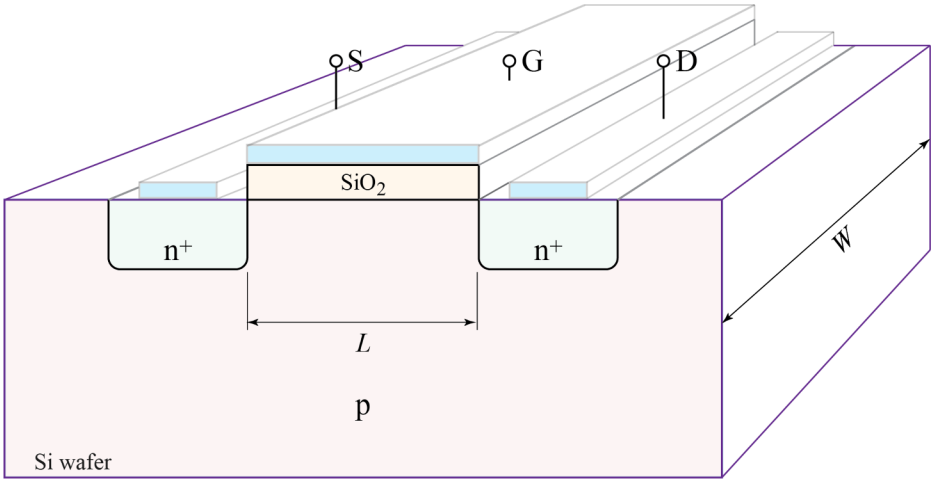
Nevertheless, we require the equations describing the FET current and voltage. We turn now to the derivation of the nonlinear equations governing the FET behavior.

Fig. 2a shows the cross-section of a FET. Note that the device extends into the page with the same cross-section to a depth of *W* for "Width", as shown in Fig. 2b.

The NMOS FET is made on a *p*-doped Silicon wafer (or in a large, *p*-doped well). The *p*-type material has embedded atoms of materials like boron or gallium that readily accept an electron to complete their outer shell.



(a)



(a)

Fig. 2. (a) FET cross-section, (b) FET 3-D view.

Two heavily-doped, *n*-type wells form the Source and Drain that are connected by metal (aluminum, copper, etc.) contacts to pins on the FET. The Gate of the FET is a pad of metal on top of an insulating silcon-dioxide (SiO_2) layer. The "channel" is

the area below the SiO_2 layer and between the n^+ regions (doped with atoms of phosphorus, arsenic, etc. that have one electron in the outer shell) for the Source and Drain.

Fig. 3 shows a representation of positive and negative carriers in the FET. The carriers arise from the random behavior of atoms. Some of the *p*-type dopant atoms may be thought of as looking like a positive charge that can absorb an electron. Some of the *n*-type dopant atoms release an electron, leaving them with a complete outer-shell, and these electrons are free to move and conduct current. Note that the regions in the FET are charge neutral, and the + and - charges in Fig. 3 are available to move (or recombine) but will be replaced by other charges when they move.

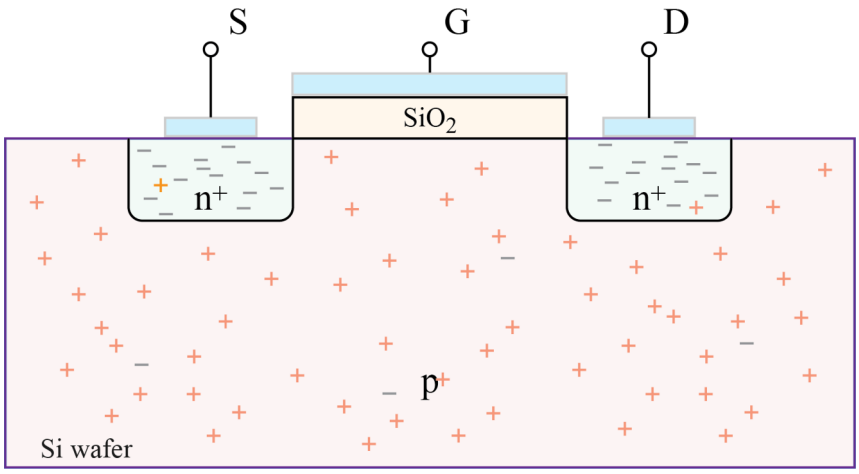


Fig. 3. Carriers in FET with no voltage applied.

To get the FET to conduct current from Drain to Source, i_D, positive voltages are applied to the Gate and Drain relative to the Source, as shown in Fig. 4.

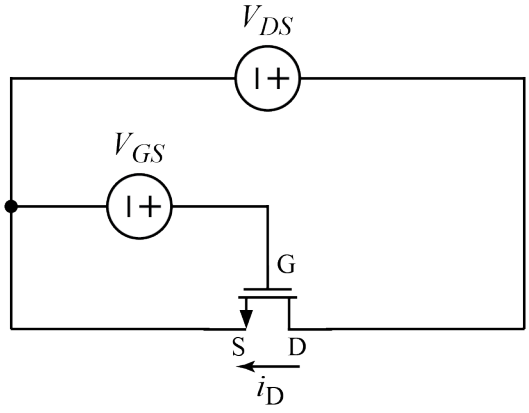


Fig. 4. Connection of voltage sources to FET to create current flow.

Fig. 5, shows the voltage sources connected to the FET. To achieve conduction from Drain to the Source (conventional current moves opposite to the flow of electrons), the applied voltages must alter the carrier densities in a way that causes electrons to move into the channel and form a continuous path between the Source and Drain.

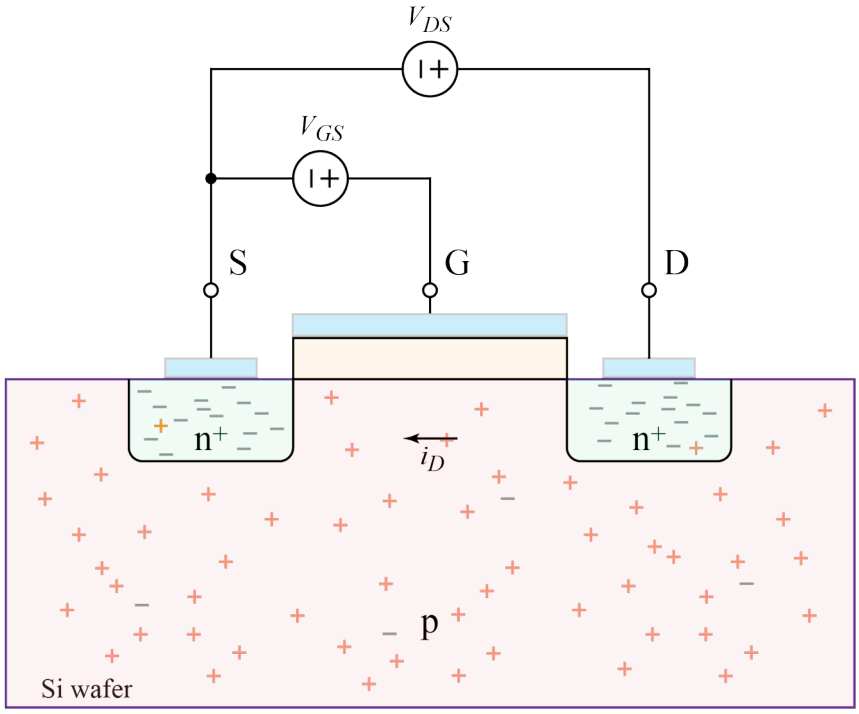


Fig. 5. Voltage sources connected to FET. When the sources are turned on, the carrier densities in the FET will change.

If the Gate-to-Source voltage, V_{GS}, is zero, as in Fig. 6, the lack of electrons in the channel results in no current, i_D, from source to drain. The FET looks like a very high resistance even if a significant V_{DS} is applied.

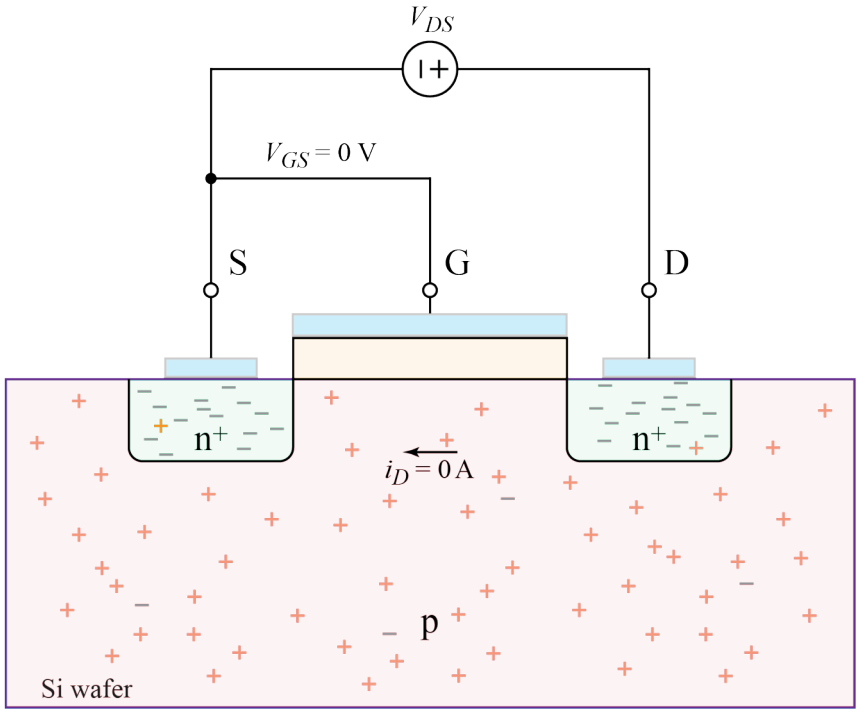


Fig. 6. Zero gate voltage makes the FET high resistance.

Fig. 7 shows that, as we start to increase the value of V_{GS}, positive charges move onto the Gate and push positive charges away from the *p*-type atoms in the other plate of the capacitor, which is the channel. This means that the p-type atoms in the channel are absorbing electrons to complete their outer shell, but there are still no extra electrons that are free to move and conduct current through the channel.

If we increase V_{GS} increases sufficiently, (i.e. above the "Threshold voltage", V_{TH}, a characteristic of the FET and listed on its data sheet), the electric filed in the channel becomes strong enough to draw electrons from the n^+ regions into the channel. Fig. 8 shows the "Triode" case where V_{DS} is small and V_{GS} is above threshold.

$V_{GS} > V_{TH}$ and V_{DS} small $\Rightarrow$ Triode region

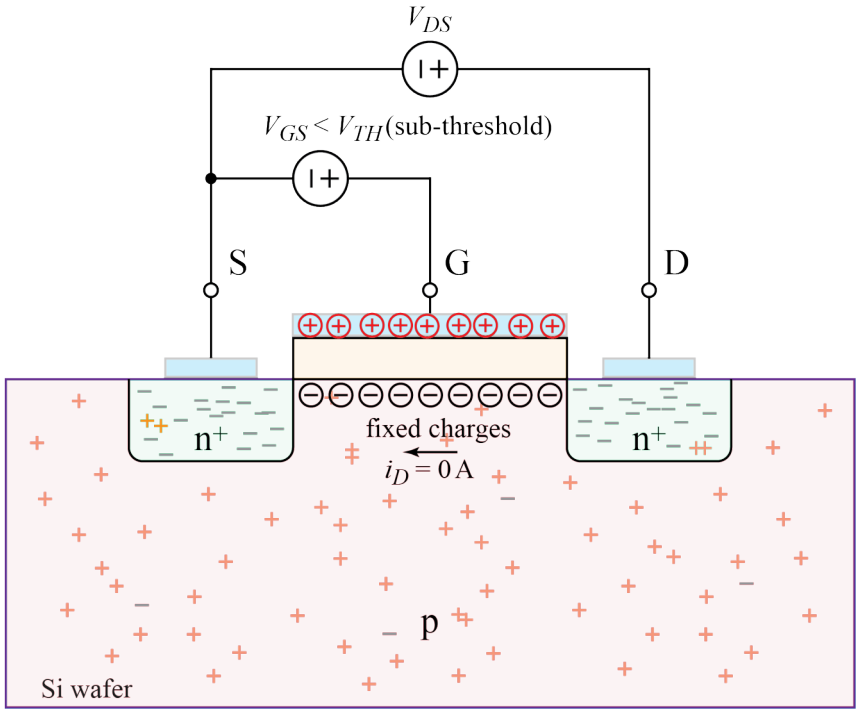


Fig. 7. Low V_{GS} causes fixed negative charges to accumulate under the gate.

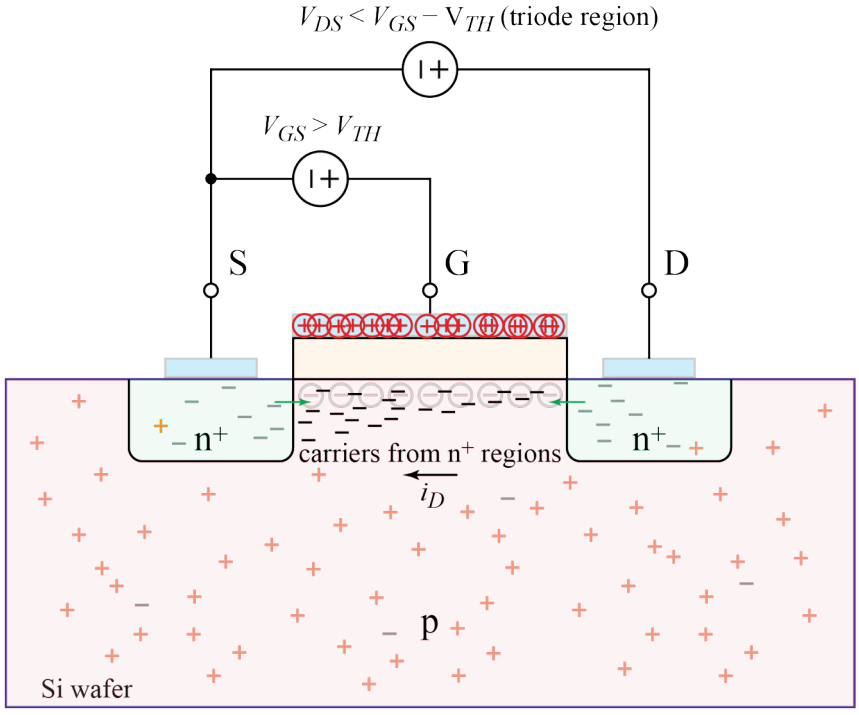


Fig. 8. V_{GS} greater than threshold pulls electrons from Source and Drain into the channel so current can flow from Drain to Source.

To derive the equation for i_D, we calculate the density of carriers in the channel, and then we determine how fast they are moving. The first step in the calculation is to define coordinates of position in the channel. We have channel length, L, shown in Fig. 9, and we define position in the channel as x.

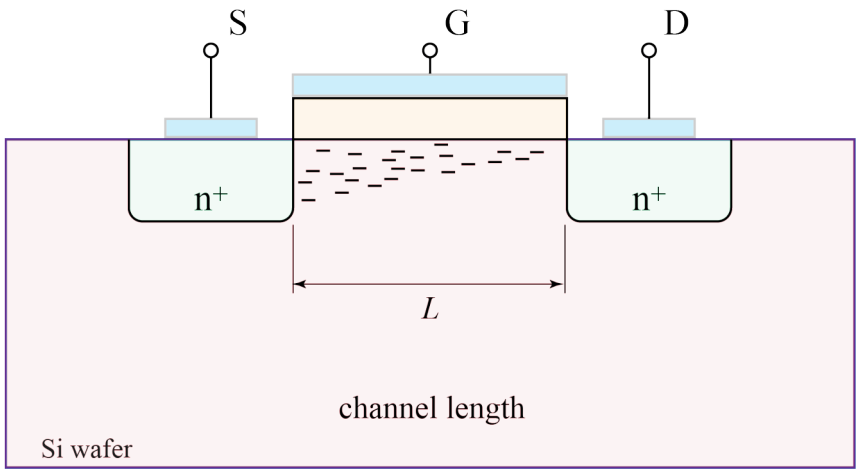


Fig. 9. Channel length, L, is the distance between the Source and Drain, which is on the order of nanometers for today's integrated circuits.

Fig. 10 shows the variation of voltage in the channel, $V(x)$, as we move from position $x = 0$ at the Source to position $x = L$ at the Drain.

We will derive the precise form of the $V(x)$ curve later. For now, it is sufficient to note that voltage is measured relative to the Source. So $V(x = 0) = 0\text{V}$, and $V(x = L) = V_{DS}$. If the channel were a resistor, the voltage would vary linearly from 0 to L, but the presence of the Gate voltage changes things.

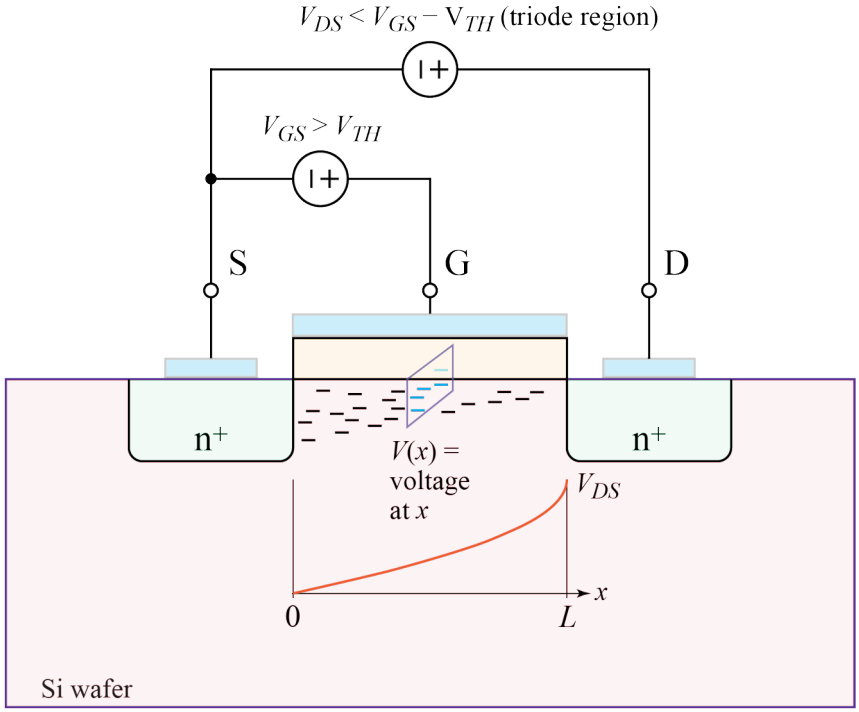


Fig. 10. The voltage in the channel varies along its length since it must rise from $V(x) = 0$V at the Source to $V(x) = V_{DS}$ at the Drain.

To understand the influence of the Gate voltage, we consider the voltage loop indicated in Fig. 11. Since the carriers in the channel are effectively on one plate of the capacitor formed by the Gate and channel, the concentration of carriers in the channel will be a function of the capacitance of the Gate and channel and of the net voltage, V_{GS} - $V(x)$, across the capacitor plates at position x. That is, we have a capacitor where where more charge is stored under the Gate where $V(x)$ is small (i.e., at the Source) and less charge is store under the Gate where $V(x)$ is larger (i.e., at the Drain). The negative charges shown in Fig. 11 in gray, indicate this changing density of charge along the length of the channel.

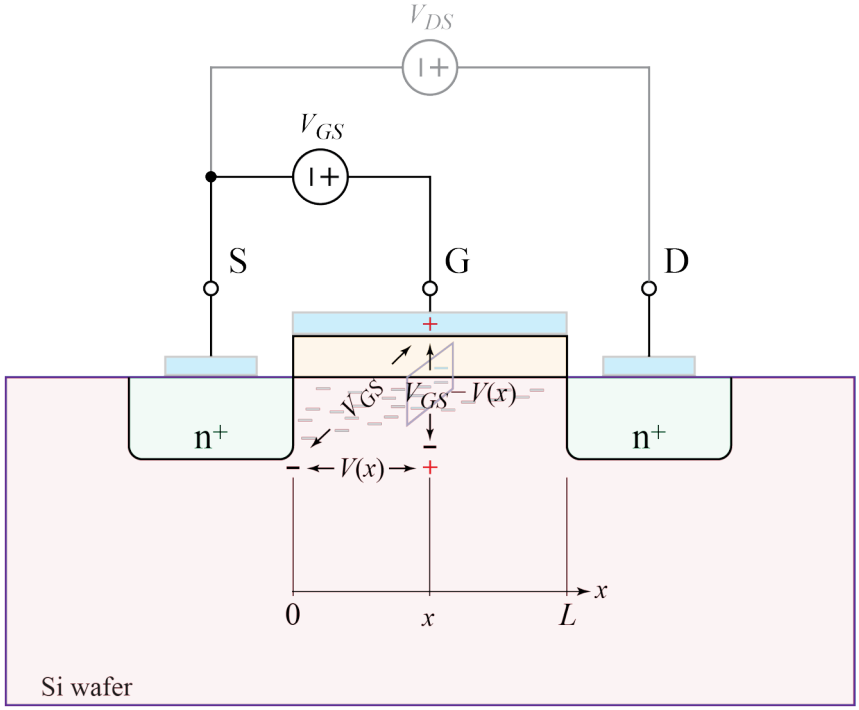


Fig. 11. Voltage loop illustrating the effective voltage for creating carriers at x in the channel is $V_{GS} - V(x)$.

Because we need a voltage higher than the threshold voltage, V_{TH}, to draw charges into the channel, we can say that the net voltage, V_n(x), determining carrier density in the channel is

$$V_n(x) = [V_{GS} - V(x)] - V_{TH}. \tag{1}$$

To find the charge density at x in the channel, we use capacitance times voltage:

$$Q(x) = CV_n(x) \tag{2}$$

where

$C = WC_{\text{ox}}$ is the capacitance of the Gate per unit length.

Fig. 12 shows C_{ox}. This value of the Gate-to-Source capacitance per area is determined by the dielectric constant of the SiO_2 insulating layer, the thickness of the SiO_2 layer (thinner makes C_{ox} higher), and the area of the gate.

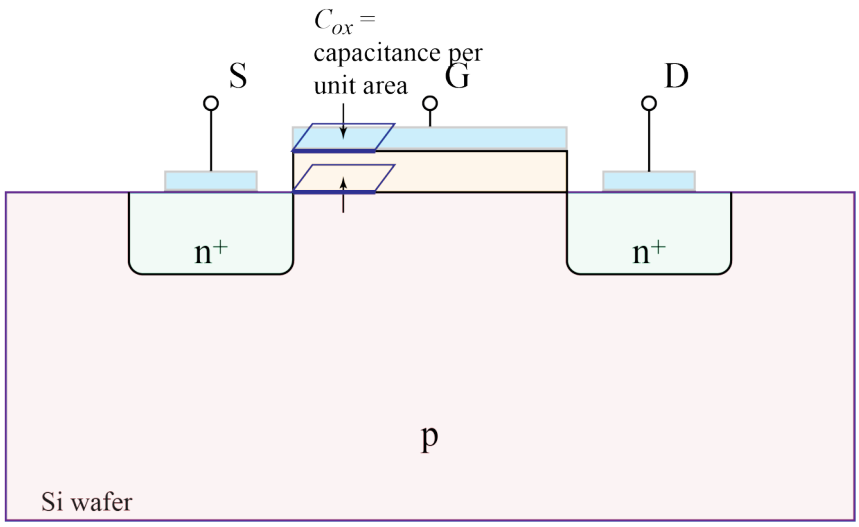


Fig. 12. Gate to Source capacitance is determined by the dielectric constant of the SiO_2 insulating layer, the thickness of the SiO_2 layer, and the area of the gate.

Recall that W is the width of the FET (measured into the paper).

Substituting for C and $V_n(x)$, we have

$$Q(x) = WC([V_{GS} - V(x)] - V_{TH}). \tag{3}$$

This equation tells us that, because the voltage difference between the Gate and the channel decreases as x moves closer to the Drain, the charge density in the channel gets smaller as x moves closer to the Drain.

An example plot of the charge density, $Q(x)$, is shown in Fig. 13.

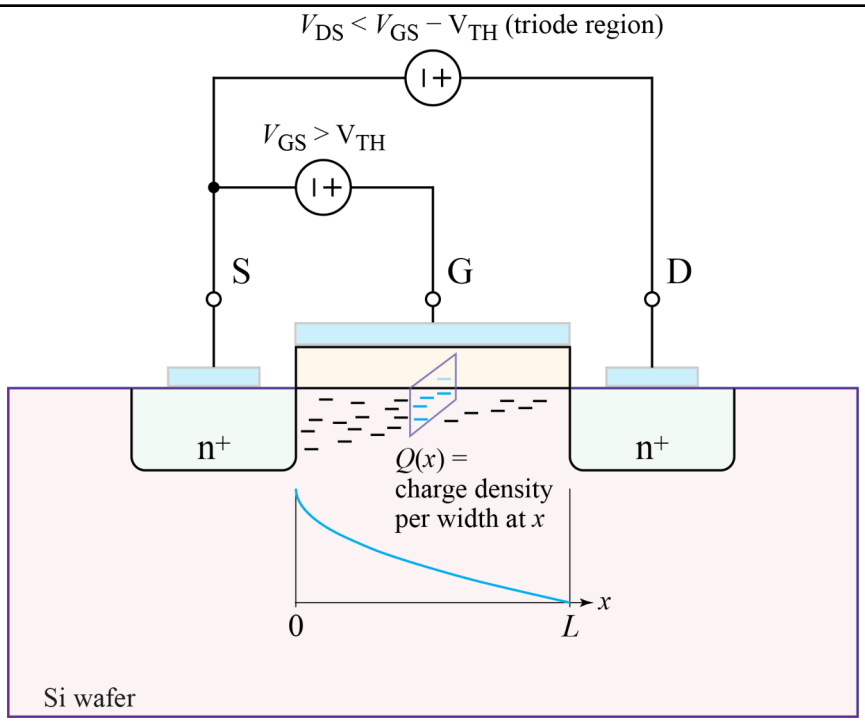


Fig. 13. Charge distribution in channel. The charge is highest where the Gate-to-channel voltage is highest, (i.e., at the Source).

Now that we know the charge density, we may calculate current i_D as charge times velocity.

$$i_D = Q(x)v(x) \tag{4}$$

where

$v(x) \equiv$ velocity at x.

To find the velocity of carriers in the channel, we use the derivative of the voltage $V(x)$, (i.e., the electric field), as a function of x.

$$v(x) = \mu_n \frac{dV(x)}{dx} \tag{5}$$

where

$\mu_n \equiv$ electron mobility in p-region

Fig. 14 shows a plot of velocity versus x, and Fig. 15 shows the derivative of $V(x)$, on which the velocity depends.

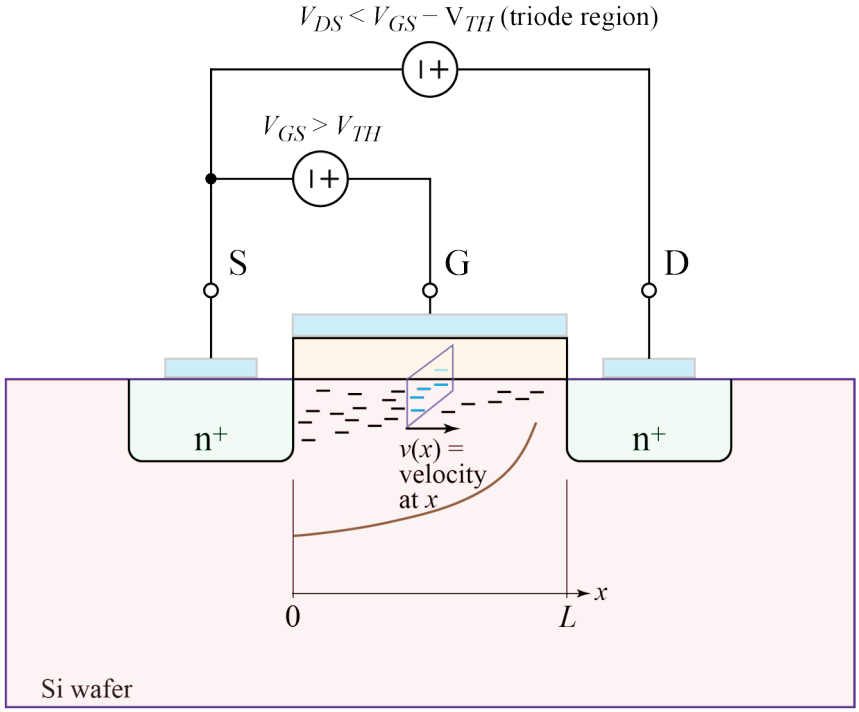


Fig. 14. Velocity as a function of position in the channel.

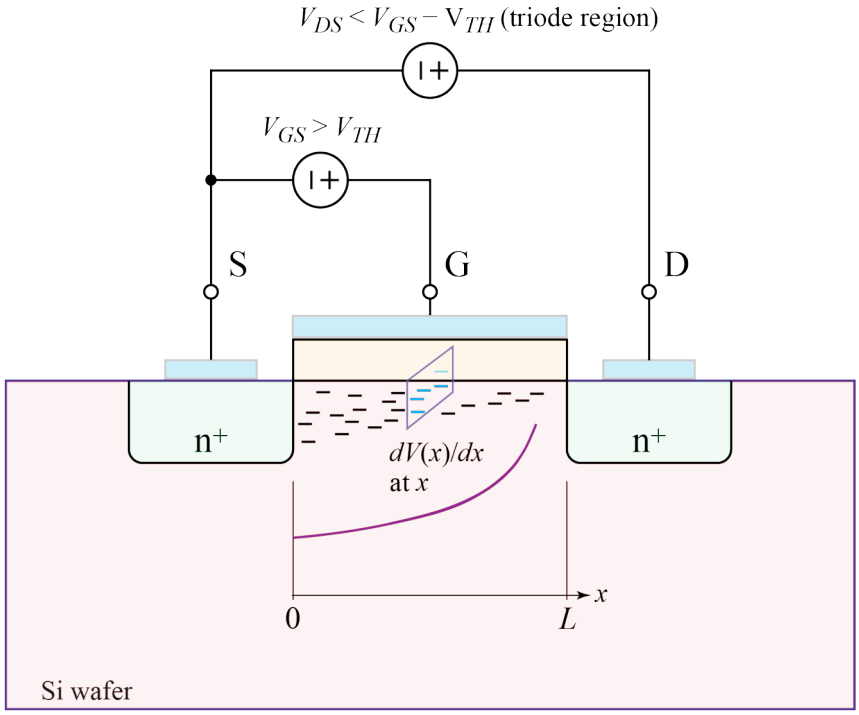


Fig. 15. The derivative of channel voltage (the electric field) versus position.

Note that we are finding the electric field (the derivative of the voltage) by seeing how voltage changes in the x direction, so we use $V(x)$. Curiously, we could use $V_n(x)$ and get the same value for $v(x)$, since $V(x)$ and $V_n(x)$ only differ by a constant and consequently have the same derivative.

Substituting terms gives us the formula for channel current, i_D.

$$i_D = W\mu_n C_{ox}([V_{GS} - V(x)] - V_{TH})\frac{dV(x)}{dx} \tag{6}$$

This is a differential equation we now solve in order to find $V(x)$ and i_D. We have the following boundary conditions:

$$V(x=0) = 0 \text{ V}$$
$$V(x=L) = V_{DS}$$

By Kirchhoff 's current law, i_D must be the same for all x once steady-state is reached. Otherwise, charge concentrations would accumulate indefinitely in some part of the channel.

Curiously, we can find the value of i_D without knowing the value of $V(x)$ or its derivative. (See Appendix for solution to V(x).) We take the integral of the constant i_D over the length of the channel to obtain an integral of $V(x)\,dV(x)$:

$$\int_{x=0}^{x=L} I_D(x)dx = \int_{x=0}^{x=L} \mu_n C_{ox} W[V_{GS} - V(x) - V_{TH}]\frac{dV(x)}{dx}dx$$

or

$$\int_{x=0}^{x=L} I_D(x)dx = \int_{V(x=0)}^{V(x=L)} \mu_n C_{ox} W[V_{GS} - V(x) - V_{TH}]dV(x). \tag{7}$$

The left side is the integral of a constant, and the right side is an integral of the linear form $ax + b$, which is a quadratic function.

$$I_D L = -\frac{1}{2}\mu_n C_{ox} W[V_{GS} - V(x) - V_{TH}]^2\Big|_{V(x)=0}^{V(x)=V_{DS}} \tag{8}$$

or

$$I_D = -\frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left([V_{GS}-V_{DS}-V_{TH}]^2-[V_{GS}-V_{TH}]^2\right) \tag{9}$$

or, after some cancellation,

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left[2(V_{GS}-V_{TH})V_{DS}-V_{DS}^2\right]. \tag{10}$$

TRIODE: With the definition of a convenient term, k_n, for physical constants, we have a more succinct version of (10):

$$I_D = \frac{k_n}{2}\left[2(V_{GS}-V_{TH})V_{DS}-V_{DS}^2\right] \tag{11}$$

where

$$k_n \equiv \mu_n C_{ox}\frac{W}{L}. \tag{12}$$

This final equation gives the current for the FET operating in the so-called "triode region" where $V_{DS} < V_{GS} - V_{TH}$. (Presumably, the term "triode" refers to similar behavior seen in vacuum tubes.) As Fig. 16 shows for example parameters, using (11) to plot I_D versus V_{DS} for various fixed values of V_{GS} yields upside-down parabolas.

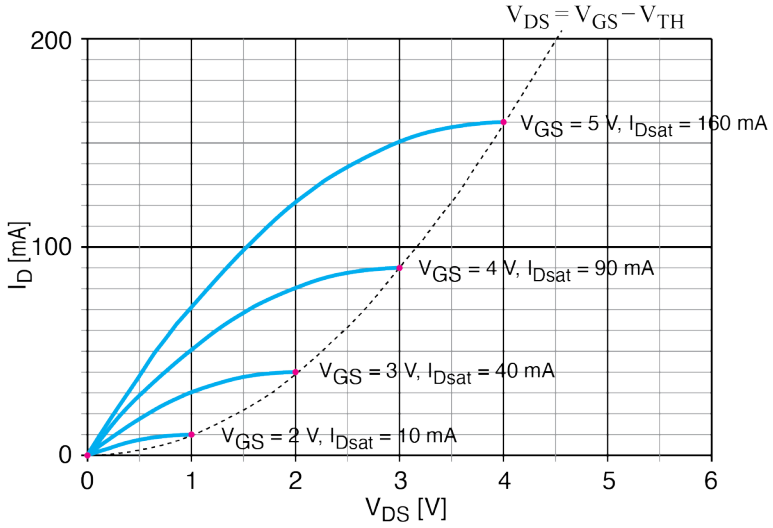


Fig. 16. Characteristic curves of FET with $k_n = 20$ mA/V^2 and $V_{TH} = 1$V.

The current reaches its peak value when $V_{DS} = V_{GS} - V_{TH}$.

LINEAR: The curves in Fig. 16 are approximately straight lines for small values of V_{DS}. Thus, we may identify a "Linear" region where the I_D curves differ from linear by no more than a specified percentage, x.

To derive the boundary of the linear region, we write I_D as a linear term plus a quadratic term.

$$I_D = I_{Dlin} + I_{Dquad} \tag{13}$$

where

$$I_{Dlin} = k_n(V_{GS} - V_{TH})V_{DS} \tag{14}$$

$$I_{Dquad} = -\frac{k_n}{2}V_{DS}^2 \tag{15}$$

Since $I_{Dlin} > I_{Dquad}$, we set up the error equation as $-I_{Dquad}/I_D$.

$$-\frac{x}{100} = \frac{-\frac{k_n}{2}V_{DS}^2}{-\frac{k_n}{2}V_{DS}^2 + k_n(V_{GS} - V_{TH})V_{DS}} \tag{16}$$

or

$$-\frac{x}{100} = \frac{1}{1 - \frac{2}{V_{DS}}(V_{GS} - V_{TH})} \tag{17}$$

Solving for V_{DS} in terms of $x/100$, we obtain an equation for V_{DS} that is a function of which V_{GS} curve we are considering.

$$V_{DS} = \frac{2(V_{GS} - V_{TH})}{1 + \frac{100}{x}} \tag{18}$$

Substituting (18) into (14) gives the boundary of the linear region as a fraction of the saturation I_D.

$$I_{Dx} = k_n(V_{GS} - V_{TH})\frac{2}{1 + \frac{100}{x}}(V_{GS} - V_{TH})^2 = \frac{k_n}{2}(V_{GS} - V_{TH})^2\frac{4}{1 + \frac{100}{x}} \tag{19}$$

or

$$I_{Dx} = I_{Dsat}\frac{4}{1 + \frac{100}{x}} \tag{20}$$

Fig. 17 shows the linear region of the characteristic curves for x = 10%.

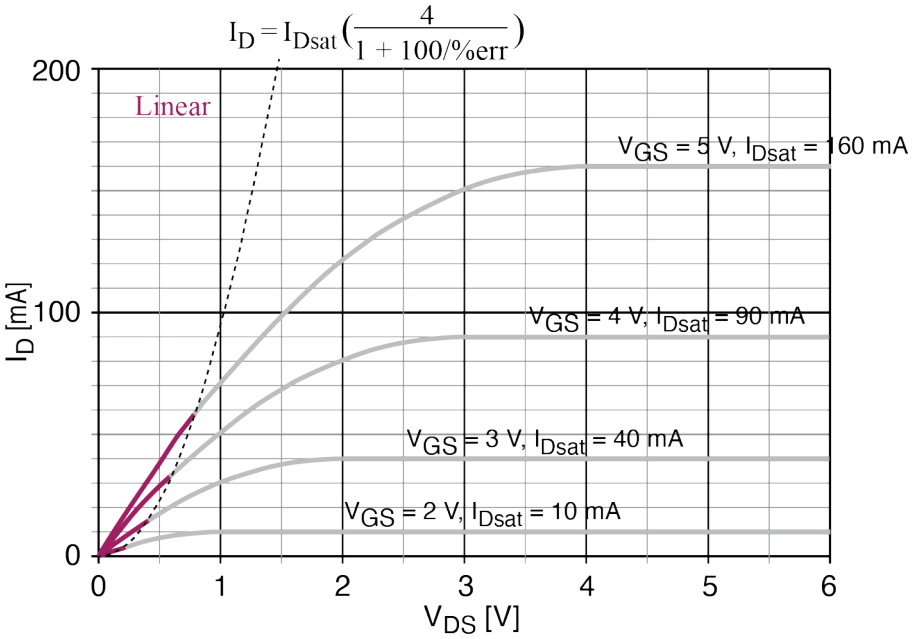


Fig. 17. Linear region within 10% of true I_D for FET with k_n = 20 mA/V^2 and V_{TH} = 1V.

SAT: At the right end of the triode region, a different physical behavior occurs. According to (11), increasing V_{DS} beyond $V_{DS} = V_{GS} - V_{TH}$ would cause the parabolic I_D curve to drop. As one might expect, however, I_D will remain at least as high as the peak value if V_{DS} increases further.

For $V_{DS} > V_{GS} - V_{TH}$, the voltage from the gate to the drain becomes negative, but current continues to flow in the channel. The concentration of electrons in the channel is said to "pinch-off" near the drain, as shown in Fig. 18. Despite the pinch-off, current continues to flow from the drain to the source.

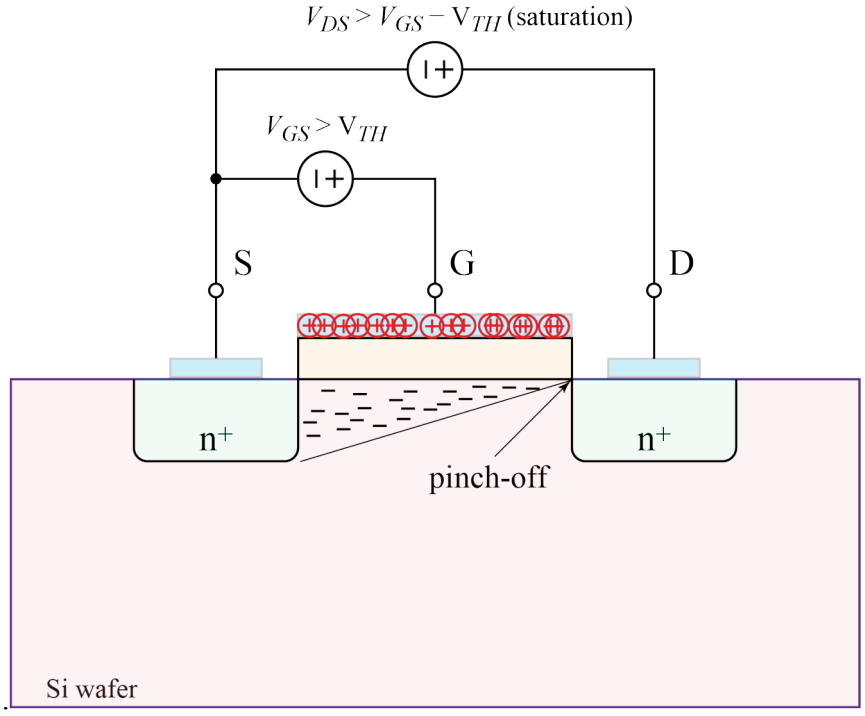


Fig. 18. Pinch-off for $V_{DS} > V_{GS} - V_{TH}$.

To a first approximation, the current stays constant for $V_{DS} > V_{GS} - V_{TH}$, and the FET is said to be in "saturation". If the current stays constant in saturation, the FET looks like an ideal current source.

The value of saturation current is the peak value of I_D from the triode region formula.

$$I_D = \frac{k_n}{2}(V_{GS} - V_{TH})^2 \qquad (21)$$

Fig. 19 shows the complete characteristic curves for our example FET.

The constant-current behavior in saturation simplifies circuit design, especially for analog applications. Consequently, FETs are usually operated in saturation for analog designs.

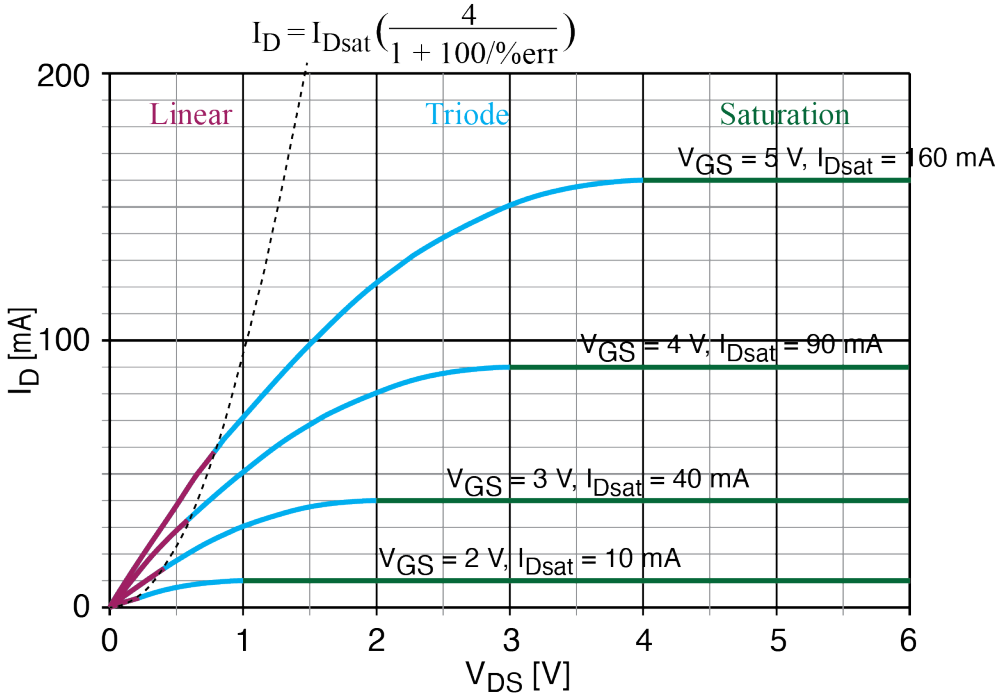


Fig. 19. Complete characteristic curves for FET example with k_n = 20 mA/V^2 and V_{TH} = 1V.

The characteristic curves are now divided into three regions: linear, triode, and saturation.

CHANNEL MOD: In most FETs, the characteristic curves in saturation continue to increase approximately linearly for $V_{DS} > V_{GS} - V_{TH}$ rather than being flat. This behavior is due to channel modulation, meaning the length of the channel effectively shortens as V_{DS} increases. In FETs used as power switches, the slope of the curves may be dramatic. Fig. 20 shows the channel shortening in such a way that I_D will increase linearly with V_{DS}, giving us a new equation for I_D:

$$I_D = \frac{1}{2}k_n(V_{GS} - V_{TH})^2(1 + \lambda V_{DS}) \tag{22}$$

where

$$k_n \equiv \mu_n C_{ox}\frac{W}{L}. \tag{23}$$

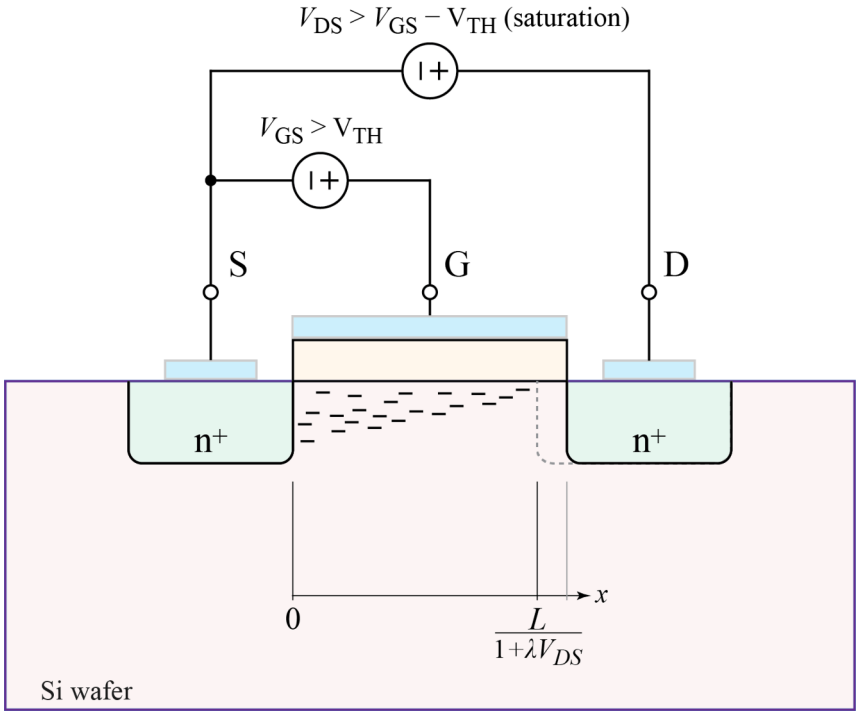


Fig. 20. Channel modulation occurs in saturation where $V_{DS} > V_{GS} - V_{TH}$. The channel length shortens as V_{DS} increases.

Fig. 21 shows the effect of channel modulation on characteristic curves. It should be noted that the correction term in (31), namely $(1 + \lambda V_{DS})$, starts at $V_{DS} = 0$V rather than at the start of saturation. For the curves in Fig. 21, the correction factor was applied in the triode region as well as the saturation regions. Otherwise, the characteristic curve would be discontinuous, suddenly jumping up at the start of

saturation. This observation suggests that channel modulation starts well before saturation. Also, the view of channel modulation in Fig. 20 suggests that the pinch-off region will lengthen as V_{DS} increases, since the n^+ region does not grow. The true physics of pinch-off and channel modulation are undoubtedly complex.

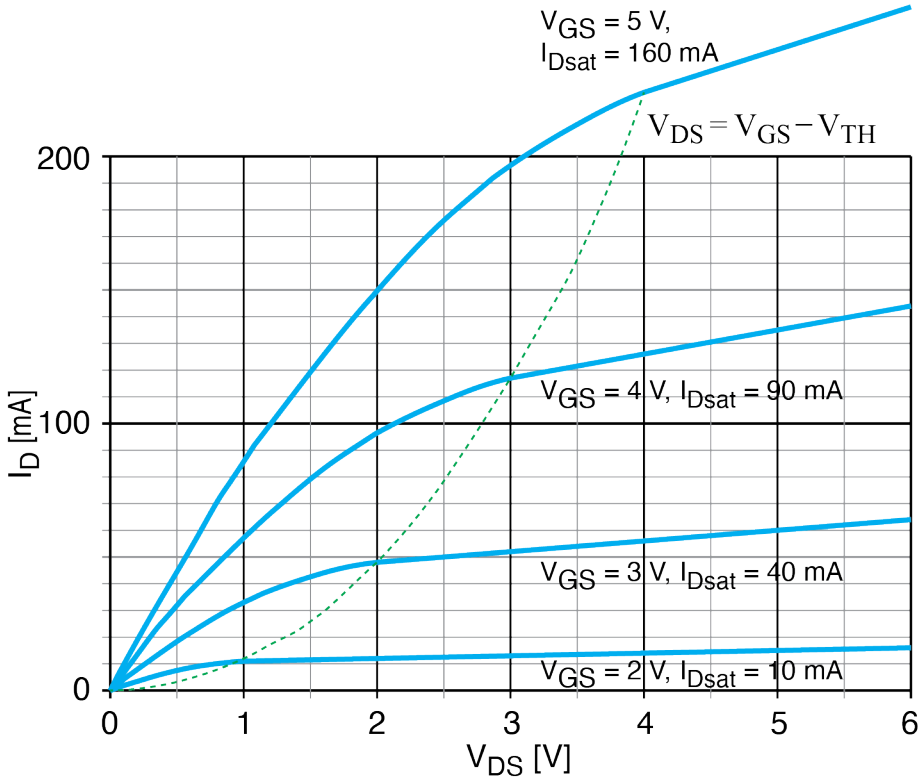


Fig. 21. Characteristic curves for previous FET example with channel modulation of $\lambda = 0.1/V$.

Fortunately, the correction term in (31) may be modeled for small signals as a resistor, r_o, from the source to the drain. To derive r_o, we write the drain-source voltage as a DC value, V_{DS}, plus a small signal or AC part, v_{ds}:

$$I_D = \frac{k_n}{2}(V_{GS} - V_{TH})^2[1 + \lambda(V_{DS} + vds)] = I_{DC} + I_{DCmod} + i_{dc} \tag{24}$$

where $I_{DC} \equiv \frac{k_n}{2}(V_{GS} - V_{TH})^2$ (25)

$$I_{DCmod} \equiv \frac{k_n}{2}(V_{GS} - V_{TH})^2 \lambda V_{DS}$$

$$i_{ac} \equiv \left[\frac{k_n}{2}(V_{GS} - V_{TH})^2 \lambda\right] v_{ds} = [I_{DC}\lambda] vds. \tag{26}$$

We see that (26) is Ohm's law with resistance r_o:

$$r_o = \frac{1}{I_{DC}\lambda} \tag{27}$$

Note that an accurate calculation of r_o requires the value of I_D *without* channel modulation. Furthermore, calculation of the DC current for I_D when there is channel modulation may require solution of a cubic equation, since the value of V_{GS} may affect the value of V_{DS}. In such cases, graphical solutions for I_D using characteristic curves are a good option.

SUMMARY: Without channel modulation

Linear $$I_D = k_n(V_{GS} - V_{TH})V_{DS} \tag{14}$$

Triode $$I_D = \frac{k_n}{2}\left[2(V_{GS} - V_{TH})V_{DS} - V_{DS}^2\right] \tag{11}$$

Saturation $$I_D = \frac{k_n}{2}(V_{GS} - V_{TH})^2 \tag{21}$$

where $$k_n \equiv \mu_n C_{ox}\frac{W}{L} \tag{12}$$

With channel modulation

Linear $$I_D = k_n(V_{GS} - V_{TH})V_{DS}(1 + \lambda V_{DS}) \tag{28}$$

Triode $$I_D = \frac{k_n}{2}\left[2(V_{GS} - V_{TH})V_{DS} - V_{DS}^2\right](1 + \lambda V_{DS}) \tag{29}$$

Saturation $$I_D = \frac{k_n}{2}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS}) \tag{22}$$

APPENDIX: The solution of the differential equation in (6) is

$$V(x) = (V_{GS} - V_{TH})(1 - \sqrt{\kappa x/L}) \tag{30}$$

where

$$\kappa \equiv 1 - \left(\frac{V_{GS} - V_{TH} - V_{DS}}{V_{GS} - V_{TH}}\right)^2.$$

REF: Behzad Razavi, *Fundamentals of Microelectronics with Robotics and Bioengineering Applications, 3rd Ed., Enhanced eText,* New York, NY: Wiley, 2021. ISBN: 978111969439