MOSFET Models

- The basic MOSFET model consist of:
  - junction capacitances CBS and CBD between source (S) to body (B) and drain to B, respectively.
  - overlap capacitances CGDO and CGSO due to gate (G) to S and G to D overlap, respectively.
  - S to B and D to B diodes.
- We will calculate dc current $I_D$ for different applied voltages.
MOSFET DC Models

Let us apply a large potential at the gate of an MOS capacitor to cause inversion, i.e. $V_G > V_{th}$.

The conductance, $1/R_I$ of the inversion layer is:

$$g_I = \frac{1}{R_I} = \frac{W}{L} \int_0^{x_l} q \mu_n n_I(x) dx$$

(1)

Where,

$n_I(x) = \text{e- density in the inversion layer}$

$x_I = \text{depth of the inversion layer}$

$\mu_n = \text{e- mobility in the inversion layer}$
Inversion Layer Conductance

The inversion layer mobility is, also, called the surface mobility and is \( \approx 1/2 \) bulk mobility.

Now, the inversion layer charge per unit area is given by:

\[
Q_I = -\int_{0}^{x_i} q n_1(x) dx
\]  
(2)

Assuming \( \mu_n = \text{constant} \), we get from (1) and (2):

\[
g_I = \frac{W}{L} \mu_n Q_I \quad (- \Rightarrow e^-)
\]  
(3)

and the inversion layer resistance of an elemental length \( dy \) is:

\[
dR_I = \frac{dy}{W \mu_n Q_I}
\]  
(4)

Eq. (4) for MOS capacitors can be directly used to derive drain current for MOSFETs under different biasing conditions.
Level-1 MOSFET Model

Let us consider the following MOSFET structure. The gate bias $V_G$ provides the control of surface carrier densities.

If $V_G > V_{th}$, inversion layer exists. \(\therefore\) a conducting channel exists $D \rightarrow S$ and current $I_D$ will flow.

$V_{th}$ is determined by the properties of the structure and is given by the equation we derived for MOS capacitors. That is,

$$V_{th} = \phi_{MS} - \frac{Q_f}{C_o} + 2\phi_F + \frac{\sqrt{2qN_AK_s\varepsilon_o(2\phi_F)}}{C_o} \quad (5)$$

For $V_G < V_{th}$, the structure consists of two diodes back to back and only leakage current flows ($\approx I_o$ of PN junctions); i.e., $I_D \sim 0$. 
I – V Characteristics

Note:

- The depletion region is wider around the drain because of the applied drain voltage $V_D$.
- The potential along the channel varies from $V_D$ at $y = L$ to 0 at $y = 0$ between the drain and source.
- The channel charge $Q_I$ and the bulk charge $Q_B$ will in general be $f(y)$ because of the influence of $V_D$, i.e. potential varies along the channel length.
I – V Characteristics

Using (4), the voltage drop across an elemental length $dy$ of MOSFET channel is:

$$dV = I_D dR_f = \frac{I_D dy}{W \mu_n Q_I(y)}$$  \hspace{1cm} (6)

Now, at any point in silicon the induced charge due to $V_G$ is:

$$Q_s(y) = Q_I(y) + Q_B(y)$$  \hspace{1cm} (7)

Again,

$$V_G = V_{FB} - \frac{Q_s}{C_{OX}} + \phi_s$$  \hspace{1cm} (8)

Note that $V_{FB}$ includes $\phi_{MS}$ and $Q_f$. Combining (7) and (8), we get:

$$Q_I(y) = -\{V_G - V_{FB} - \phi_s(y)\}C_{OX} - Q_B(y)$$  \hspace{1cm} (9)

Since the surface is inverted, $\phi_s \equiv 2\phi_F$ plus any reverse bias that exists between the channel and substrate (due to $V_D$ or $V_{Sub}$).
I – V Characteristics

\[ \therefore \phi_s(y) = V(y) + 2\phi_F \quad (10) \]

Also, we know from MOS capacitor analysis that

\[ Q_B(y) = -qN_A x_{d_{\text{max}}}(y) \]
\[ = -\sqrt{2qK_s\varepsilon_o N_A [V(y) + 2\phi_F]} \quad (11) \]

Note:
As we move from S → D, V(y)↑ due to IR drop in the channel.
\[ \therefore x_{d_{\text{max}}} \uparrow \text{ as we move toward the drain} \]
\[ Q_B \uparrow \text{ as we move toward the drain.} \]

Substituting (10) and (11) into (9), we have:

\[ Q_I(y) = -\{V_G - V_{FB} - V(y) - 2\phi_F\} C_{OX} \]
\[ + \sqrt{2qK_s\varepsilon_o N_A [V(y) + 2\phi_F]} \quad (12) \]
I – V Characteristics

Substituting (12) in (6), we get:

\[ I_D dy = -W \mu_n \left\{ \left[ V_G - V_{FB} - V(y) - 2\phi_F \right] C_{OX} - \sqrt{2qK_s \varepsilon_o N_A \left[ V(y) + 2\phi_F \right]} \right\} dV \]

If we assume \( Q_B(y) \) = constant, i.e. neglect the influence of channel voltage on \( Q_B \), then we can write:

\[ I_D dy = -W \mu_n C_{OX} \left[ V_G - V_{FB} - V(y) - 2\phi_F - \frac{2qK_s \varepsilon_o N_A \left[ V(y) + 2\phi_F \right]}{C_{OX}} \right] dV \]

\[ = -W \mu_n C_{OX} \left[ V_G - V_{th} - V(y) \right] dV \]

(13)

In Eq. (13) \( V_{th} \) includes the effects of \( \phi_F, \phi_{MS} \), and \( Q_B(y=0) \).

Integrating within the limits: \( \begin{cases} y = 0 \\ V = 0 \end{cases} \) \ to \ \( \begin{cases} y = L \\ V = V_D \end{cases} \) \( \Rightarrow \)

\[ I_D = \frac{W}{L} \mu_n C_{OX} \left[ V_G - V_{th} - \frac{V_D}{2} \right] V_D \]

(14)
I – V Characteristics: Linear Region

Eq. (14) is used in Level 1 MOSFET SPICE model to calculate $I_D$. If we define:

$\kappa \equiv \mu_n C_{OX} =$ process transconductance

$\beta \equiv \mu_n C_{OX}(W/L) = \kappa(W/L) =$ gain factor of the device

$$I_D = \beta [V_G - V_{th} - \frac{V_D}{2}] V_D$$

(15)

If $V_D < 0.1$ V, we can simplify (15) to:

$$I_D \equiv \beta (V_G - V_{th}) V_D$$

(16)

Eq. (16) shows that current varies linearly with $V_D$. This region of MOSFET operation is called the *linear region* of operation. From (16), the effective resistance between the source and drain is:

$$R_{ch} = \frac{V_D}{I_D} \cong \frac{1}{\beta (V_G - V_{th})}$$

(17)
I – V Characteristics

- $I_D - V_D$ plots for different $V_G$ from Eq. (15) show $I_D \downarrow$ for higher $V_D$. While the measured $I_D$ saturates at higher $V_D$. This discrepancy is due to the breakdown of gradual channel approximation near the drain–end of the channel at high $V_D$.

- Eq. (15) is valid only as long as an inversion layer exists all the way from $S \rightarrow D$.

- The maximum value of $I_D$ vs. $V_D$ plots occur at
  \[ V_D = V_G - V_{th} \equiv V_{DSAT} = \text{drain saturation voltage} \quad (18) \]

- For $V_D > V_{DSAT}$, a channel will not exist all the way to the drain.
I – V Characteristics: Saturation

What happens for $V_D > V_{DSAT}$?

• e- travelling in the inversion layer are injected into the depletion region. There the high ε-field pulls e- into the drain.

• Further increase of $V_D$ do not change $I_D$ (to a first order).

∴ $I_D \approx$ constant for $V_D > V_{DSAT}$.

• The boundary between the linear and the saturation regions is described by:

$$V_G - V_{th} = V_{DSAT} \quad (18)$$
I – V Characteristics: Saturation

Substituting (18) in (16), we can show that the saturation drain current is given by:

\[ I_{DSAT} \equiv \frac{W}{2L} \mu_n C_{OX} (V_G - V_{th})^2 \]  \hspace{1cm} (19)

⇒ square law theory of MOSFET devices.

At \( V_D > V_{DSAT} \), channel is pinched-off and the effective channel length is given by: \( L_{\text{eff}} = L - l_d = L(1 - l_d/L) \)

\[ \therefore I_D \equiv \frac{W}{2(L-l_d)} \mu_n C_{OX} (V_G - V_{th})^2 = \frac{I_{DSAT}}{1 - \frac{l_d}{L}} \]  \hspace{1cm} (20)

Typically, \( l_d \ll L \), then from (20):

\[ I_D \equiv I_{DSAT} \left( 1 + \frac{l_d}{L} \right) \]  \hspace{1cm} (21)
I – V Characteristics: Saturation

Since \( l_d = f(V_D) \), we can write:

\[
1 + \frac{l_d}{L} \equiv 1 + \lambda V_D
\]  
(22)

Here, \( \lambda = f(V_D) = \) channel length modulation model parameter.

\[ \therefore I_D \equiv I_{DSAT} \left(1 + \lambda V_D \right) \]  
(23)

From (23), \( I_D = 0 \) at \( V_D = -1/\lambda \).

\[ \therefore \lambda \text{ can be extracted from the } V_D \text{-intercept of } I_D \text{ vs. } V_D \text{ plots at } I_D = 0 \text{ as shown in Fig.} \]
Level 1 MOSFET Model: Summary

- Current Eq:

\[
I_D = \begin{cases} 
0 & V_G \leq V_{th} \\
\beta [V_G - V_{th} - \frac{V_D}{2}]V_D & V_G \geq V_{th}, V_D \leq V_{DSAT} \\
\frac{\beta}{2} (V_G - V_{th})^2 (1 + \lambda V_D) & V_G \geq V_{th}, V_D \geq V_{DSAT}
\end{cases}
\] (24)

- Model Parameters:
  - \( VTO \) = threshold voltage at \( V_B = 0 \)
  - \( KP \) = process transconductance
  - \( GAMMA \) = body factor
  - \( LAMBDA \) = channel length modulation factor
  - \( PHI = 2|\phi_F| \) = bulk Fermi-potential.
Level 1 MOSFET Model: Summary

- Assumptions:
  - gradual channel approximation (GCA) is valid
  - majority carrier current is negligibly (such as hole current for nMOSFETs is neglected)
  - recombination and generation are negligible
  - current flows in the y-direction (along the length of the channel) only
  - carrier mobility $\mu_x$ in the inversion layer is constant in the y-direction
  - current flow is due to drift only (no diffusion current)
  - bulk charge $Q_B$ is constant at any point in the y-direction.

The accuracy of Level 1 model is poor even for long devices.
Level 2 Model: Bulk-Charge Model

In reality, $Q_B$ varies along the channel form source at $y = 0$ to drain at $y = L$ because of applied bias $V_D$. Then from (11) with back bias $V_B$ we have:

$$Q_B(y) = -qN_Ax_{d_{max}}(y) = -\sqrt{2qK_s\varepsilon_o N_A[V(y) + 2\phi_F + V_B]}$$

$$= -C_{OX}\gamma\sqrt{2\phi_F + V_B + V(y)}$$

$$\therefore Q_I(y) = -C_{OX}\left(V_G - V_{FB} - 2\phi_F - V(y) - \gamma\sqrt{2\phi_F + V_B + V(y)}\right)$$

(25)

Using (26) in (6) and integrating form $y = 0$ to $y = L$ we get:

$$I_D = \frac{W\mu_n C_{OX}}{L}\left[(V_G - V_{FB} - 2\phi_F - \frac{V_D}{2})V_D - \frac{2}{3}\gamma\{(V_D + 2\phi_F + V_B)^{3/2} - (2\phi_F + V_B)^{3/2}\}\right]$$

(27)

Eq (27) is the current used by SPICE Level 2 MOSFET model.
Level 1 vs. Level 2 \( I - V \) Characteristics

We see that for a typical device the approximate Eq. (14) works well at low \( V_D \) but predicts higher current at large \( V_D \). This is because some of the gate charge supports \( Q_B \) at higher \( V_D \) which is ignored by (14).

- Both Eqn. (14) and (27) are valid only as long as an inversion layer exists all the way from \( S \rightarrow D \).
- The expression for saturation drain voltage for Level 2 model is given by:

\[
V_{DSAT} = V_G - V_{FB} - 2\Phi_F + \gamma^2 - \gamma \sqrt{V_G - V_{FB} + V_B + \frac{\gamma^2}{4}}
\]  

(28)
Level 2 MOSFET Model Current

- Consideration of $Q_B$ variation along the channel offers more accurate $I_D$ modeling in the linear and saturation region.
- The current given by (27) is accurate, however, complex.

∴ $I_D$ is simplified using a new model parameters:

$$\delta = \frac{0.5}{\sqrt{2\phi_F + V_B}} \quad (29)$$

- If $\alpha \equiv 1 + \delta \gamma$, then

$$I_D = \begin{cases} 
0 & V_G \leq V_{th} \quad \text{(cut-off region)} \\
\beta [V_G - V_{th} - \frac{\alpha V_D}{2}]V_D & V_G \geq V_{th}, V_D \leq V_{DSAT} \quad \text{(linear region)} \\
\frac{\beta}{2\alpha} (V_G - V_{th})^2 (1 + \lambda V_D) & V_G \geq V_{th}, V_D \geq V_{DSAT} \quad \text{(saturation region)} 
\end{cases} \quad (31)$$
Sub-threshold Region Model

Simple long channel device equations assume: \( I_{DS} = 0 \) for \( V_{GS} < V_{th} \). In reality, \( I_{DS} \neq 0 \) and varies exponentially with \( V_{GS} \) in a manner similar to a bipolar transistor.

In order to develop a theory of sub-threshold conduction, let us consider MOS band diagram with applied source (S) and gate (G) bias measured with respect to the substrate (Sub), that is:

- \( V_{S_{Sub}} \)
- \( V_{G_{Sub}} \)
Sub-threshold Region Model

In the sub-threshold (weak inversion) region, we know the following:

1. As $E_F$ is pulled above $E_i$, the number of minority carrier $e^-$ at the surface increases exponentially with $E_F - E_i$.

2. When $\phi_s = 2\phi_F$, strong inversion ($n_{surf} = p_{bulk}$) is achieved and we obtain $V_{th}$.

3. For $\phi_s < 2\phi_F$, the dominant charges present near the surface are ionized acceptor atoms, i.e. $n_{surf} << N_A^-$.
   - Thus, there is no $\varepsilon$-field laterally along the surface since Poisson’s equation is the same everywhere.
   - Thus, any current flow must be due to diffusion only.
Sub-threshold Region Model

\[ I_D = -AqD \frac{dn}{dx} \quad (32) \]

4. The e- gradient along the channel (dn/dx) must be constant in order to maintain constant current.

\[ I_D = -AqD[\{n(0) - n(L)\}/L] \quad \cdots (33) \]

Now, we can now use carrier statistics to calculate n(0) and n(L). Referring band diagram on page 20:

\[ n(0) = n_t e \quad \frac{q(\phi_s - V_{SSub} - \phi_F)}{kT} \quad (34) \]

\[ n(L) = n_t e \quad \frac{q(\phi_s - V_{DSub} - \phi_F)}{kT} \quad (35) \]

Where \( \phi_s \) is the surface potential with respect to the substrate. Also, note that \( \phi_s = \) constant along the channel [\( \varepsilon(x) = 0 \)].
Sub-threshold Region Model

Since the charge in the substrate is assumed uniform \( (N_A^-) \), then from Poisson’s equation:

\[
\frac{d^2 \phi}{dx^2} = \frac{qN_A}{\varepsilon_s} = -\frac{dE}{dx}
\]  \hspace{1cm} (36)

∴ \( \phi \) varies parabolically and \( E \) varies linearly with distance.

Since the e- concentration falls off as \( e^{-q\phi/kT} \) away from the surface, essentially, all of the minority carrier e- are contained in a region in which the potential drops by \( kT/q \).

∴ The depth of the inversion layer, \( X_{inv} = \Delta \phi/E_s \).

where \( \Delta \phi = kT/q \), and \( E_s \) is the \( \varepsilon \)-field at the surface.

Again, from Gauss’ Law: \( \varepsilon_s E_s = -Q_B = \sqrt{2\varepsilon_s qN_A \phi_s} \)  \hspace{1cm} (37)

\[
\therefore X_{inv} = \frac{\varepsilon_s kT/q}{\sqrt{2\varepsilon_s qN_A \phi_s}} = \frac{kT}{q} \sqrt{\frac{\varepsilon_s}{2qN_A \phi_s}}
\]  \hspace{1cm} (38)
Sub-threshold Region Model

Using (34), (35), and (38) in (33), we get:

\[ I_D = q \frac{W}{L} Dn_i \left\{ e^{\frac{q\left(\phi_s - V_{SSub} - \phi_F\right)}{kT}} - e^{\frac{q\left(\phi_s - V_{DSsub} - \phi_F\right)}{kT}} \right\} \frac{kT}{q} \sqrt{\frac{\epsilon_s}{2qN_A\phi_s}} \]  \hspace{1cm} (39)

Here \( A = WX_{inv} \).

Using \( V_{DSub} = V_{DS} + V_{SSub} \),

\[ I_D = \frac{W}{L} kTDn_i \sqrt{\frac{\epsilon_s}{2qN_A\phi_s}} \left\{ e^{\frac{q\left(\phi_s - V_{SSub} - \phi_F\right)}{kT}} \right\} \left\{ 1 - e^{\frac{-qV_{DS}}{kT}} \right\} \]  \hspace{1cm} (40)

To make use of this equation, we need to know how \( \phi_s \) varies with the externally applied potential \( V_G \).

\[ V_{GSub} = V_{FB} + \phi_s + \sqrt{\frac{2q\epsilon_s N_A\phi_s}{C_o}} \]  \hspace{1cm} (41)
Sub-threshold Region Model

Generally, it is more common to use the source potential as the reference so that:

\[ V_{Gsub} = V_{GS} + V_{SSub} \]
\[ \phi_s = \psi_s + V_{SSub} \]

\[ \therefore I_D = \frac{W}{L} kT D n_i \sqrt{\frac{\varepsilon_s}{2qN_A \phi_s}} \left\{ e^{\frac{q(\psi_s-\phi_F)}{kT}} \right\} \left\{ 1 - e^{-\frac{qV_{DS}}{kT}} \right\} \] (42)

\[ V_{GS} = V_{FB} + \psi_s + \frac{\sqrt{2q\varepsilon_s N_A (\psi_s + V_{SSub})}}{C_o} \] (43)

The depletion layer capacitance is:

\[ C_D = \sqrt{\frac{q\varepsilon_s N_A}{2(\psi_s + V_{SSub})}} \] (44)

Therefore, from (43):

\[ \frac{dV_{GS}}{d\psi_s} = 1 + \frac{1}{C_o} \sqrt{\frac{q\varepsilon_s N_A}{2(\psi_s + V_{SSub})}} = 1 + \frac{C_D}{C_o} \equiv n \] (45)
Sub-threshold Region Model

In order to eliminate $\psi_s$ from (42), we expand $V_{GS}$ in a series about the point $\psi_s = 1.5\phi_F$ (weak inversion corresponds to $\phi_F \leq \psi_s \leq 2\phi_F$).

\[ \therefore V_{GS} \cong V_{GS}\big|_{\psi_s = 1.5\phi_F} + n(\psi_s - 1.5\phi_F) \quad (46) \]

Where $V^*_{GS} \equiv V_{GS}\big|_{\psi_s = 1.5\phi_F}$ and is obtained from (43).

Combining (46) with (42) to eliminate $\psi_s$ in the exponential and using (44) to eliminate the square root of $\psi_s$ in (42), we obtain:

\[ I_D = \frac{W}{L} \left(\frac{kT}{q}\right)^2 \mu C_D n_i \left\{ e^{\left[ \frac{q(V_{GS} - V^*_{GS})}{nkT} \frac{q\phi_F}{2kT} \right]} \right\} \left\{ 1 - e^{-\frac{qV_{DS}}{kT}} \right\} \quad (47) \]
Sub-threshold Region Model

Thus, the sub-threshold current is given by:

\[
I_D = \frac{W}{L} \left( \frac{kT}{q} \right)^2 \mu C_D n_i \left\{ e^{-\left[ \frac{q(V_{GS} - V_{GS}^*)}{nkT} \right] \frac{q\Phi_F}{2kT}} \right\} \left\{ 1 - e^{-\frac{qV_{DS}}{kT}} \right\}
\]

(47)

From (47) we note that:

- \( I_D \) depends on \( V_{DS} \) only for small \( V_{DS} \), i.e. \( V_{DS} \leq 3kT/q \), since \( \exp[-qV_{DS}/kT] \rightarrow 0 \) for larger \( V_{DS} \).

- \( I_D \) depends exponentially on \( V_{GS} \) but with an “ideality factor” \( n > 1 \). Thus, the slope is poorer than a BJT but approaches to that of a BJT in the limit \( n \rightarrow 1 \).

- \( N_A \) and \( V_{SSub} \) enter through \( C_D \).
Sub-threshold Slope (S-factor)

In order to change $I_D$ by one decade, we get from (47):

\[
\frac{q \partial V_{GS}}{nkT} = \ln 10 \Rightarrow S = \frac{kT}{q} \ln 10(n) = \frac{kT}{q} \ln 10 \left(1 + \frac{C_D}{C_o}\right)
\]

\[
\therefore S \cong 60 \frac{mV}{\text{decade}} I \left(1 + \frac{C_D}{C_o}\right) \quad (@ \text{ room } T) \quad \ldots (48)
\]
Sub-threshold Model - Final Note

- In weak inversion or subthreshold region, MOS devices have exponential characteristics but are less “efficient” than BJTs because $n > 1$.

- Subthreshold slope $S$ does not scale and is $\approx$ constant. Therefore, $V_{th}$ cannot be scaled as required by the ideal scaling laws.

- $V_{DS}$ affects $V_{th}$ as well as subthreshold currents.

- In order to optimize $S$, the desirable parameters are:
  - thin oxide
  - low $N_A$
  - high $V_{Sub}$. 
Level 2/3 MOSFET Model Current

- Using sub-threshold region model, for \( V_G < V_{on} \), we can show:

\[
I_D = I_{on} e^{q(V_G - V_{on})/nkT}
\]

(49)

Here \( V_{on} = V_G \) at the point of intersection between the strong and weak inversion while \( I_{on} = I_D @ V_G = V_{on} \)

- Thus, the complete long-channel DC MOSFET model is given by:

\[
I_D = \begin{cases} 
I_{on} \exp \left[ \frac{q(V_G - V_{on})}{nkT} \right] & V_G \leq V_{on} \\
\beta [V_G - V_{th} - \frac{\alpha V_D}{2}] V_D & V_G \geq V_{th}, V_D \leq V_{DSAT} \\
\frac{\beta}{2\alpha} (V_G - V_{th})^2 (1 + \lambda V_D) & V_G \geq V_{th}, V_D \geq V_{DSAT}
\end{cases}
\]

(50)
Inversion Layer Mobility

- In our derivation of $I_D$, we have considered $\mu_s = \text{constant}$ which is not true under applied $V_G$ and $V_D$.
- As the vertical field $E_x$ and lateral field $E_y$ increases with increasing $V_G$ and $V_D$ respectively, carriers undergo increased scattering.
  \[
  \therefore \mu_s = f(E_x, E_y)
  \]
- For simplicity of $I_D$ calculation an effective mobility defined as the average mobility of carriers is used in MOSFETs:

$$
\mu_{eff} = \frac{\int_{x_{inv}}^{0} \mu_s (x, y)n(x, y)dx}{\int_{0}^{x_{inv}} n(x, y)dx}
$$

(51)
Mobility Degradation due to $V_G$

- In reality, $\mu$ is significantly reduced by large vertical e-fields.
- The vertical e-field, pulls the inversion layer e- towards the surface causing more surface scattering and interaction of e- with oxide charges ($Q_f, N_{it}$) causing coulomb scattering.
- Effective mobility vs. effective normal e-field show:
  - at high fields: *universal* behavior independent of doping density.
  - at low fields: dependence on 1) doping density and 2) interface charge.
Mobility Degradation due to $V_G$

- Mobility versus effective normal electric field curve is modeled using three components:
  - coulomb scattering (due to ionized impurity)
  - phonon scattering - *almost constant*
  - surface roughness scattering (at Si/SiO$_2$ interface).

- At low normal fields:
  - less inversion charge density
  - ionized impurity scattering dominates and $\mu_{eff} = f(N_A)$.

- At high normal fields:
  - higher inversion charge density close to the interface
  - surface roughness scattering dominates.
Mobility Degradation due to $V_G$

- The mobility correction has several flavors depending on the level of MOSFET model.

- In SPICE MOSFETs Levels 3 and 4 uses a simple form such as:

$$\mu_{eff} = \frac{\mu_0}{1 + \theta (V_G - V_{th})} \quad (52)$$

Where the model parameters are:

$\theta =$ mobility degradation parameter due to $V_G$.

$\mu_0 =$ low field mobility.

- The effective mobility due to $V_G$ is called low-field surface mobility.
Mobility Degradation due to $V_D$

An additional complication arises because of the high lateral e-fields $E(y)$.

We assumed, $v_d = \mu E$.

However, for e- in silicon $v_d$ saturates near $E \sim 10^4$ V/cm.

Average e-field for short channel devices $> 10^4$ V/cm.

Therefore, small geometry MOS devices will operate with:

$$v_d = v_{sat} \approx 10^7 \text{ cm/sec}.$$  

Since $\mu \neq$ constant, we need to account for high lateral field effect in the expression for $I_D$ derived from simple theory.
Mobility Degradation due to $V_D$

- The velocity saturation of inversion carriers due to increased lateral field causes:
  - current saturation sooner than predicted by $V_{DSAT} = V_G - V_{th}$
  - lower $I_{DSAT}$ than the predicted value by simple theory.

- The drift velocity due to high field effect is given by:
  \[
  v = \frac{\mu_s E_y}{1 + \left[\left(\frac{E_y}{E_C}\right)^\beta\right]^\beta}
  \]

- Where $E_C \equiv \frac{\nu_{sat}}{\mu_s}$ = critical lateral field for velocity saturation.

- For simplicity of numerical solution, $\beta = 1$ is used for modeling.

  \[
  \therefore v = \frac{\mu_s E_y}{1 + \mu_s \frac{E_y}{\nu_{sat}}} \tag{53}
  \]
Modeling Mobility Degradation

- The effective mobility due to combine effect of $V_G$ and $V_D$ is given by:

$$\mu_{\text{eff}} = \frac{\mu_0}{1 + \theta (V_G - V_{th}) + \theta_b V_B + \theta_c V_D}$$

(54)

Where $\theta_C \equiv 1/LE_C$

$L = \text{channel length of MOSFETs.}$

- Thus, $\mu_{\text{eff}}$ is modeled by the parameter set:
  - $\{(\mu_0, \theta, \theta_b, \text{and } v_{\text{sat}} (E_C))\}$.

- The model parameters are obtained by curve fitting the experimental data to the model equation.
Modeling $V_{DSAT}$ at High e-Fields

Assume a piecewise linear model, i.e., $v_d$ saturates abruptly at $E = E_c$:

$$v_d = \frac{\mu_{eff} E}{1 + \frac{E}{E_c}}, (E \leq E_c)$$

$$= v_{sat}, (E \geq E_c) \quad (55)$$

Where $E = \text{lateral e-field}$. $E_c = \text{critical e-field at which carriers are velocity saturated, i.e.}$

$$v = v_{sat}.$$ 

Now, the current at any point $x$ in the channel is given by:

$$I_D = I(x) = WC_o[V_G - V_{th} - V(x)]v(x) \quad (56)$$

Where

$V(x) = \text{potential difference between the drain and channel at } x.$

$v(x) = \text{carrier velocity at } x.$
Modeling $V_{DSAT}$ at High e-Fields

Substituting (55) in (56) we get:

$$E(x) = \frac{I_D}{W\mu_{eff} C_o [V_G - V_{th} - V(x)] - I_D/E_c} = \frac{dV(x)}{dx}$$ (57)

Integrating (57) from $x = 0$ to $x = L$ and $V(x) = 0$ to $V(x) = V_D$, we get in the linear region ($V_{DS} \leq V_{DSAT}$):

$$I_D = \frac{W\mu_{eff} C_o \left( V_G - V_{th} - \frac{1}{2} V_{DS} \right) V_{DS}}{L \left( 1 + \frac{V_{DS}}{E_c L} \right)}$$ (58)

Let us define, $V_{DSAT}$ = the drain saturation voltage due to $v_{sat}$, i.e. at $E = E_c$.

Using this condition in (57) we get in the saturation region ($V_{DS} \geq V_{DSAT}$):

$$I_D = \frac{W\mu_{eff} C_o \left( V_G - V_{th} - V_{DSAT} \right) E_c}{2}$$ (59)
Modeling $V_{DSAT}$ at High e-Fields

$I_D$ given by (58) and (59) must be continuous @ $V_{DS} = V_{DSAT}$. Thus, Equating (58) and (59) we can solve for $V_{DSAT}$:

$$V_{DSAT} = \frac{E_C L (V_G - V_{th})}{E_C L + (V_G - V_{th})} = \frac{V_G - V_{th}}{1 + \frac{V_G - V_{th}}{E_C L}}$$

(60)

From (60), if $L$ is large, $V_{DSAT} \rightarrow (V_G - V_{th})$ used in the simple theory.

Eq (60) is used to calculate drain current $I_D$ for $V_D > V_{DSAT}$ in BSIM3 and BSIM4 MOSFET models.
Hot-Carrier Effects

For \( V_{DS} > V_{DSAT} \), lateral e-field increases beyond \( E_c \) and reaches a maximum value at the drain end. From pseudo-2d analysis it can be shown that maximum e-filed is:

\[
E_m = \sqrt{\left( \frac{V_{DS} - V_{DSAT}}{l_p} \right)^2 + E_c^2} \quad (61)
\]

Where, \( l_p \) depends on:
- gate oxide thickness \( t_{ox} \)
- junction depth \( x_j \)

\[
l_p = 1.73 t_{ox} x_j \quad (62)
\]

If \( (V_{DS} - V_{DSAT}) / l_p \gg E_c \), then:

\[
E_m \approx \frac{V_{DS} - V_{DSAT}}{l_p} \quad (63)
\]
Hot-Carrier Effects

- Channel electron traveling through high electric field near the drain end can:
  - become highly energetic, i.e. hot
  - cause impact ionization and generate $e^-$ and holes
    - holes go into the substrate creating substrate current, $I_{sub}$.
- Some channel $e^-$ have enough energy to overcome the SiO$_2$-Si energy barrier generating gate current, $I_g$.
- The maximum e-field, $E_m$ near the drain has the greatest control of hot carrier effects.
Substrate Current Modeling

Impact ionization coefficient, $\alpha = A_i \exp(-B_i/E)$. Then the substrate current due to impact ionization is:

$$I_{sub} = \int_{l_p}^{l_p} I_D \alpha dx$$

$$= \int_{l_p}^{l_p} I_D A_i e^{-\frac{B_i}{E(x)}} dx$$

(64)

Where

$x = o \Rightarrow$ start of impact ionization region with $E = E_c$.

$x = l_p \Rightarrow$ length of impact ionization region with $E = E_m$.

Changing the limits of integration to $E$ in (19) and after integration we get:

$$I_{sub} = \left( \frac{A_i}{B_i} \right) l_p E_m I_D \exp \left( -\frac{B_i}{E_m} \right)$$

(65)

$\therefore$ As $E_m \uparrow$, $I_{sub} \uparrow$. 