nMOSFET Schematic

- Four structural masks: Field, Gate, Contact, Metal.
- Reverse doping polarities for pMOSFET in N-well.
- Source terminal: Ground potential.
- Gate voltage: $V_g$
- Drain voltage: $V_{ds}$
- Substrate bias voltage: $-V_{bs}$

- $\psi(x,y)$: Band bending at any point $(x,y)$.
- $V(y)$: Quasi-Fermi potential along the channel.
- $V(y=0) = 0$, $V(y=L) = V_{ds}$. 

### nMOSFET Schematic

- Polysilicon gate
- Gate oxide
- Depletion region
- Inversion channel
- p-type substrate
- n$^+$ source
- n$^+$ drain
- Source terminal: Ground potential
- Gate voltage: $V_g$
- Drain voltage: $V_{ds}$
- Substrate bias voltage: $-V_{bs}$
Drain Current Model

Electron concentration: \( n(x, y) = \frac{n_i^2}{N_a} e^{q(\psi - V)kT} \)

Electric field:
\[
\varepsilon^2(x, y) = \left( \frac{d\psi}{dx} \right)^2 = \frac{2kTN_a}{\varepsilon_{si}} \left[ \left( e^{-q\psi/kT} + \frac{q\psi}{kT} - 1 \right) + \frac{n_i^2}{N_a} \left( e^{-qV/kT} \left( e^{q\psi/kT} - 1 \right) - \frac{q\psi}{kT} \right) \right]
\]

Condition for surface inversion:
\( \psi(0, y) = V(y) + 2\psi_B \)

Maximum depletion layer width at inversion:
\[
W_{dm}(y) = \sqrt{\frac{2\varepsilon_{si}[V(y) + 2\psi_B]}{qN_a}}
\]
Gradual Channel Approximation

Assumes that vertical field is stronger than lateral field in the channel region, thus 2-D Poisson’s eq. can be solved in terms of 1-D vertical slices.

Current density eq. (both drift and diffusion):

\[ J_n(x, y) = -q\mu_n n(x, y) \frac{dV(y)}{dy} \]

Integrate in \( x \)- and \( z \)-directions,

\[ I_{ds}(y) = -\mu_{eff} W \frac{dV}{dy} Q_i(y) = -\mu_{eff} W \frac{dV}{dy} Q_i(V) \]

where \( Q_i(y) = -q\int_0^{x_i} n(x, y)dx \) is the inversion charge/area.

Current continuity requires \( I_{ds} \) independent of \( y \), integration with respect to \( y \) from 0 to \( L \) yields

\[ I_{ds} = \mu_{eff} \frac{W}{L} \int_0^{V_{ds}} (-Q_i(V))dV \]
Pao-Sah’s Double Integral

Change variable from \((x,y)\) to \((\psi,V)\),

\[ n(x, y) = n(\psi, V) = \frac{n_i^2}{N_a} e^{q(\psi - V)/kT} \]

\[ Q_i(V) = -q \int_{\psi_s}^{\psi_B} n(\psi, V) \frac{dx}{d\psi} d\psi = -q \int_{\psi_s}^{\psi_B} (\frac{n_i^2}{N_a}) e^{q(\psi - V)/kT} \frac{d\psi}{\mathcal{E}(\psi, V)} \]

Substituting into the current expression,

\[ I_{ds} = q \mu_{eff} \frac{W}{L} \int_0^{V_{ds}} \left[ \int_{\psi_s}^{\psi_B} (\frac{n_i^2}{N_a}) e^{q(\psi - V)/kT} \frac{d\psi}{\mathcal{E}(\psi, V)} \right] dV \]

where \(\psi_s(V)\) is solved by the gate voltage eq. for a vertical slice of the MOSFET:

\[ V_g = V_{fb} + \psi_s - \frac{Q_s}{C_{ox}} = V_{fb} + \psi_s + \frac{\sqrt{2e_s kTN_a}}{C_{ox}} \left[ \frac{q \psi_s}{kT} + \frac{n_i^2}{N_a^2} e^{q(\psi_s - V)/kT} \right]^{1/2} \]
Charge Sheet Approximation

Assumes that all the inversion charges are located at the silicon surface like a sheet of charge and that there is no potential drop across the inversion layer.

After the onset of inversion, the surface potential is pinned at $\psi_s = 2\psi_B + V(y)$.

- **Depletion charge:** $Q_d = -qN_aW_{dm} = -\sqrt{2\varepsilon_{si}qN_a(2\psi_B + V)}$
- **Total charge:** $Q_s = -C_{ox}(V_g - V_{fb} - \psi_s) = -C_{ox}(V_g - V_{fb} - 2\psi_B - V)$
- **Inv. charge:** $Q_i = Q_s - Q_d = -C_{ox}(V_g - V_{fb} - 2\psi_B - V) + \sqrt{2\varepsilon_{si}qN_a(2\psi_B + V)}$

Substituting in $I_{ds} = \mu_{eff} \frac{W}{L} \int_0^{V_{ds}} (-Q_i(V))dV$ and integrate:

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} \left\{ \left(V_g - V_{fb} - 2\psi_B - \frac{V_{ds}}{2}\right)V_{ds} - \frac{2\sqrt{2\varepsilon_{si}qN_a}}{3C_{ox}} \left[ (2\psi_B + V_{ds})^{3/2} - (2\psi_B)^{3/2} \right] \right\}$$
Linear Region I-V Characteristics

For $V_{ds} \ll V_g$,

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} \left( V_g - V_{fb} - 2\psi_B - \frac{\sqrt{4\varepsilon_{si} q N_a \psi_B}}{C_{ox}} \right)$$

$$V_{ds} = \mu_{eff} C_{ox} \frac{W}{L} (V_g - V_t) V_{ds}$$

where $V_t = V_{fb} + 2\psi_B + \frac{\sqrt{4\varepsilon_{si} q N_a \psi_B}}{C_{ox}}$ is the MOSFET threshold voltage.
Saturation Region I-V Characteristics

Keeping the 2nd order terms in $V_{ds}$:

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} \left[ (V_g - V_t)V_{ds} - \frac{m}{2} V_{ds}^2 \right]$$

where $m = 1 + \frac{\sqrt{\varepsilon_s q N_a}}{4 \psi_B C_{ox}} = 1 + \frac{C_{dm}}{C_{ox}} = 1 + \frac{3t_{ox}}{W_{dm}}$ is the body-effect coefficient.

$I_{ds} = I_{dsat} = \mu_{eff} C_{ox} \frac{W}{L} \frac{(V_g - V_t)^2}{2m}$

when

$V_{ds} = V_{dsat} = (V_g - V_t)/m$.

Typically, $m \approx 1.2$. 

\[ 
\begin{align*} 
(V_{ds}, I_{ds}) & \quad \text{(Saturation Region)} \\
\text{Drain Voltage} & \quad \text{Drain Current} \\
V_{g1} & \\
V_{g2} & \\
V_{g3} & \\
V_{g4} & 
\end{align*} 
\]
Pinch-off Condition

From inversion charge density point of view,

\[ Q_i(V) = -C_{ox}(V_g - V_I - mV) \]

while \( I_{ds} = \mu_{eff} \frac{W}{L} \int_0^{V_{ds}} (-Q_i(V))dV \)

At \( V_{ds} = V_{dsat} = (V_g - V_I)/m \), \( Q_i = 0 \) and \( I_{ds} = \) max.
Pinch-off from Potential Point of View

\[ V(y) = \frac{V_g - V_t}{m} - \sqrt{\left(\frac{V_g - V_t}{m}\right)^2 - 2 \frac{y}{L} \left(\frac{V_g - V_t}{m}\right) V_{ds} + \frac{y}{L} V_{ds}^2} \]

At the pinch-off point, \( dV/dy \rightarrow \infty \)

\( \Rightarrow \) Gradual channel approximation breaks down.

Current is injected into the bulk depletion region.
Beyond Pinch-off
Subthreshold Region

$V_{ds} = \frac{V_T - V_i}{m}$

Diffusion Component

Drift Component

Drain Current (arbitrary units)

Gate Voltage (V)
Subthreshold Currents

\[-Q_s = \varepsilon_{si} \varepsilon_s = \sqrt{2\varepsilon_{si}kTN_a} \left[ \frac{q\psi_s}{kT} + \frac{n_i^2}{N_a^2} e^{q(\psi_s - V)/kT} \right]^{1/2} \]

Power series expansion: 1st term \(Q_d\), 2nd term \(Q_i\),

\[-Q_i = \sqrt{\frac{\varepsilon_{si}qN_a}{2\psi_s}} \left( \frac{kT}{q} \right) \left( \frac{n_i}{N_a} \right)^2 e^{q(\psi_s - V)/kT} \]

\[\Rightarrow I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\varepsilon_{si}qN_a}{2\psi_s}} \left( \frac{kT}{q} \right)^2 \left( \frac{n_i}{N_a} \right)^2 e^{q\psi_s/kT} \left( 1 - e^{-qV_{ds}/kT} \right) \]

or,

\[I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} (m - 1) \left( \frac{kT}{q} \right)^2 e^{q(V_g - V_i)/mkT} \left( 1 - e^{-qV_{ds}/kT} \right) \]

Inverse subthreshold slope:

\[S = \left( \frac{d(\log I_{ds})}{dV_g} \right)^{-1} = 2.3 \frac{mkT}{q} \frac{kT}{q} \left( 1 + \frac{C_{dm}}{C_{ox}} \right) \]
Body Effect: Dependence of Threshold Voltage on Substrate Bias

If $V_{bs} \neq 0$,

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} \left[ \left( V_g - V_{fb} - 2\psi_B - \frac{V_{ds}}{2} \right) V_{ds} - \frac{2\sqrt{2\varepsilon_s q N_a}}{3C_{ox}} \left( (2\psi_B + V_{bs})^3/2 - (2\psi_B + V_{bs})^{3/2} \right) \right]$$

Substrate Bias Voltage, $V_{bs}$ (V)

Threshold Voltage, $V_t$ (V)

$dV_t/dV_{bs} = \frac{\sqrt{\varepsilon_s q N_a}}{2(2\psi_B + V_{bs})} C_{ox}$

$t_{ox}=200 \text{ Å}$

$V_{fb}=0$

$N_a=10^{16} \text{ cm}^{-3}$

$N_a=3 \times 10^{15} \text{ cm}^{-3}$
Dependence of Threshold Voltage on Temperature

For n⁺ poly gated nMOSFET, \( V_{fb} = -(E_g/2q) - \psi_B \)

\[
V_t = -\frac{E_g}{2q} + \psi_B + \frac{\sqrt{4\varepsilon_{si} q N_a \psi_B}}{C_{ox}}
\]

\[
\frac{dV_t}{dT} = -\frac{1}{2q} \frac{dE_g}{dT} + \left(1 + \frac{\sqrt{\varepsilon_{si} q N_a / \psi_B}}{C_{ox}}\right) \frac{d\psi_B}{dT} = -\frac{1}{2q} \frac{dE_g}{dT} + (2m-1) \frac{d\psi_B}{dT}
\]

\[
\Rightarrow \frac{dV_t}{dT} = -(2m-1) \frac{k}{q} \left[ \ln \left( \sqrt{\frac{N_c N_v}{N_a}} \right) + \frac{3}{2} \right] + \frac{m-1}{q} \frac{dE_g}{dT}
\]

From Table 2.1, \( dE_g/dT \approx -2.7 \times 10^{-4} \) eV/K and \( (N_c N_v)^{1/2} \approx 2.4 \times 10^{19} \) cm⁻³.

For \( N_a \sim 10^{16} \) cm⁻³ and \( m \sim 1.1 \),

\( dV_t/dT \) is typically \(-1\) mV/K.
MOSFET Channel Mobility

\[
\mu_{\text{eff}} = \frac{\int_{0}^{x_i} \mu_n n(x) \, dx}{\int_{0}^{x_i} n(x) \, dx}
\]

It was empirically found that when \(\mu_{\text{eff}}\) is plotted against an effective normal field \(\varepsilon_{\text{eff}}\), there exists a “universal relationship” independent of the substrate bias, doping concentration, and gate oxide thickness (Sabnis and Clemens, 1979).

Here

\[
\varepsilon_{\text{eff}} = \frac{1}{\varepsilon_{\text{si}}} \left( |Q_d| + \frac{1}{2} |Q_i| \right)
\]

Since \(|Q_d| = \sqrt{4\varepsilon_{\text{si}} qN_A \psi_B} = C_{\text{ox}} (V_t - V_{fb} - 2\psi_B)\) and \(|Q_i| \approx C_{\text{ox}} (V_g - V_t)\),

\[
\varepsilon_{\text{eff}} = \frac{V_t - V_{fb} - 2\psi_B}{3t_{\text{ox}}} + \frac{V_g - V_t}{6t_{\text{ox}}}
\]

For n+ poly gated nMOSFET,

\[
\varepsilon_{\text{eff}} = \frac{V_t + 0.2}{3t_{\text{ox}}} + \frac{V_g - V_t}{6t_{\text{ox}}}
\]
N-channel MOSFET Mobility

- Low field region (low electron density): Limited by impurity or Coulomb scattering (screened at high electron densities).

- Intermediate field region: Limited by phonon scattering,
  \[ \mu_{\text{eff}} \approx 32500 \times \varepsilon^{-1/3} \]

- High field region (> 1 MV/cm): Limited by surface roughness scattering (less temp. dependence).
Temperature Dependence of MOSFET Current
In general, pMOSFET mobility does not exhibit as "universal" behavior as nMOSFET.

\[ \mathcal{E}_{\text{eff}} = \frac{1}{\mathcal{E}_{\text{si}}} \left( |Q_d| + \frac{1}{3} |Q_i| \right) \]
Electron and Hole Mobilities vs. Field

![Graph showing electron and hole mobilities vs. effective electric field. The graph has lines for different oxide thicknesses (35 Å and 70 Å) and different CMOS technologies (1 µm and 0.1 µm). The y-axis represents effective mobility in cm²/V·s, and the x-axis represents effective electric field in MV/cm.](image-url)
Intrinsic MOSFET Capacitance

- **Subthreshold region:** \[ C_g = WL \left( \frac{1}{C_{ox}} + \frac{1}{C_d} \right)^{-1} \approx WLC_d \]

- **Linear region:** \[ C_g = WLC_{ox} \]

- **Saturation region:**

  \[ Q_i(y) = -C_{ox}(V_g - V_t)\sqrt{1 - \frac{y}{L}} \]

  \[ \Rightarrow C_g = \frac{2}{3} WLC_{ox} \]
In the charge-sheet model, $C_i = \infty$ and $Q_i = C_{ox}(V_g - V_t)$.

In reality, inversion layer has a finite thickness and finite capacitance.

$$\frac{d(-Q_i)}{dV_g} = \frac{C_{ox}C_i}{C_{ox} + C_i + C_d} \approx C_{ox}\left(1 - \frac{1}{1 + C_i / C_{ox}}\right)$$
Inversion Layer Capacitance

1st order approximation, \( C_i \approx |Q_i|/(2kT/q) \) and \( |Q_i| \approx C_{ox}(V_g - V_t) \), therefore, \( C_i/C_{ox} = (V_g - V_t)/(2kT/q) \).

\[
-Q_i = C_{ox} \left[ (V_g - V_t) - \frac{2kT}{q} \ln \left( 1 + \frac{q(V_g - V_t)}{2kT} \right) \right]
\]

Note:
Linearly extrapolated threshold voltage is typically \((2-4)kT/q\) higher than the threshold voltage \(V_t\) at \(\psi_s(\text{inv.}) = 2\psi_B\).
Short-Channel Effect

If \( L \downarrow \), \( I_{ds} = I_{dsat} = \mu_{\text{eff}} C_{ox} \frac{W (V_g - V_t)^2}{L} \frac{2m}{2m} \uparrow \)

And \( C_g = \frac{2}{3} W L C_{ox} \downarrow \). But ……

Threshold voltage becomes sensitive to channel length and drain bias.
Short-Channel $V_t$ Roll-off
2-D Potential Contours
(Same gate voltage)

Long channel

Short channel
Drain-Induced Barrier Lowering

Surface Potential

Drain current (A/cm)

Gate voltage (V)

$V_{ds} = 3.0 \text{ V}$

$V_{ds} = 50 \text{ mV}$

$L = 0.2 \mu m$

$L = 0.35 \mu m$

$L = 2.0 \mu m$

$N_a = 3 \times 10^{16} \text{ cm}^{-3}$

$t_{ox} = 100 \text{ Å}$

$L = 6.25 \mu m$

$L = 1.25 \mu m$

$V_{ds} = 0.5 \text{ V}$

$V_{ds} = 5 \text{ V}$
Lateral Field Penetration

2-D Poisson’s Eq.:

\[ \frac{\partial \varepsilon_x}{\partial x} + \frac{\partial \varepsilon_y}{\partial y} = \rho = \frac{-qN_a}{\varepsilon_{si}} \]

- \( \varepsilon_{si} \frac{\partial \varepsilon_x}{\partial x} \): gate controlled depletion charge.
- \( \varepsilon_{si} \frac{\partial \varepsilon_y}{\partial y} \): S/D controlled depletion charge.

Note that the characteristic length of exponential decay is independent of channel length.
2-D Analysis in a Simplified MOSFET Geometry

A 2-D boundary-value problem with Poisson’s equation
In AFGH (oxide), \[ \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0 \]

In ABEF (silicon), \[ \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{qN_a}{\varepsilon_{si}} \]

Boundary conditions:
- \( \psi(-3t_{ox}, y) = V_g - V_{fb} \) along GH,
- \( \psi(x, 0) = \psi_{bi} \) along AB,
- \( \psi(x, L) = \psi_{bi} + V_{ds} \) along EF,
- \( \psi(W_d, y) = 0 \) along CD.

To eliminate the boundary condition at the Si/oxide interface, the oxide region is replaced by an equivalent Si region \( (\varepsilon_{si}/\varepsilon_{ox})t_{ox} \approx 3t_{ox} \) thick.
Let: \[ \psi(x, y) = v(x, y) + u_L(x, y) + u_R(x, y) + u_B(x, y) \]

\( v(x,y) \) is a solution to the inhomogeneous equation and satisfies the top boundary condition.

\( u_L, u_R, u_B \) are solutions to the homogeneous equation such that \( \psi(x,y) \) satisfies the other B.C.’s.
For satisfying the Boundary conditions:

\[
\begin{align*}
  u_L(x, y) &= \sum_{n=1}^{\infty} b_n^* \frac{\sinh \left( \frac{n \pi (L - y)}{W_d + 3t_{ox}} \right)}{\sinh \left( \frac{n \pi L}{W_d + 3t_{ox}} \right)} \sin \left( \frac{n \pi (x + 3t_{ox})}{W_d + 3t_{ox}} \right) \\
  u_R(x, y) &= \sum_{n=1}^{\infty} c_n^* \frac{\sinh \left( \frac{n \pi y}{W_d + 3t_{ox}} \right)}{\sinh \left( \frac{n \pi L}{W_d + 3t_{ox}} \right)} \sin \left( \frac{n \pi (x + 3t_{ox})}{W_d + 3t_{ox}} \right) \\
  u_B(x, y) &= \sum_{n=1}^{\infty} d_n^* \frac{\sinh \left( \frac{n \pi (x + 3t_{ox})}{L} \right)}{\sinh \left( \frac{n \pi (W_d + 3t_{ox})}{L} \right)} \sin \left( \frac{n \pi y}{L} \right)
\end{align*}
\]

Note that for \( u = \sin(kx) \), \( d^2 u/dx^2 = -k^2 u \);
And that for \( u = \sinh(ky) \), \( d^2 u/dy^2 = k^2 u \).
Define scale length,

\[ \lambda = W_{dm} + (\varepsilon_{si}/\varepsilon_{ox})t_{ox} \]

To keep short-channel effect under control, \( L_{min} \) should be kept larger than about 2\( \lambda \).

\[ m = \Delta V_g/\Delta \psi_s = 1 + 3t_{ox}/W_{dm} \]
Depletion Width Scaling

\[ W_{dm}^0 = \sqrt{\frac{4 \varepsilon_{si} kT \ln\left(\frac{N_a}{n_i}\right)}{q^2 N_a}} \]

Substrate Doping Concentration (cm\(^3\))

Maximum Depletion Width (µm)
In the one-region model, the eigenvalues are:

\[ k_n = \frac{n\pi}{W_d + 3t_{ox}} \]

For two regions, assume eigenvalues:

\[ k_n = \frac{\pi}{\lambda_n} \]

B.C. at \( x = 0 \): \( u_{L1} = u_{L2} \)

\( (du_{L1}/dy = du_{L2}/dy) \)

and \( \varepsilon_i du_{L1}/dx = \varepsilon_{si} du_{L2}/dx \)

\[ \Rightarrow \varepsilon_{si} \tan(\pi t_i/\lambda_n) + \varepsilon_i \tan(\pi W_d/\lambda_n) = 0 \]
Generalized Scale Length

Lowest eigenvalue: \[ \varepsilon_{si} \tan(\pi \frac{t_i}{\lambda_1}) + \varepsilon_i \tan(\pi \frac{W_d}{\lambda_1}) = 0 \]

\[ \Delta \psi_{SCE} \propto \exp\left(-\frac{\pi L}{2 \lambda_1}\right) \]

\[ L_{\min} \sim 2 \lambda_1 \]

Note that:
- \( \lambda_1 > W_d \), and \( \lambda_1 > t_i \)
- \( \lambda_1 = 2W_d = 2t_i \) is always a point of symmetry regardless of \( \varepsilon_i, \varepsilon_{si} \).
- If \( \varepsilon_i = \varepsilon_{si} \), \( \lambda_1 = W_d + t_i \)
MOSFET Body Effect

\[ m = \frac{\Delta V_g}{\Delta \psi_s} = 1 + \frac{3t_{ox}}{W_{dm}} \]
To obtain a good subthreshold slope, the body-effect coefficient,

\[ m = \frac{\Delta V_g}{\Delta \psi_s} = 1 + \frac{3t_{ox}}{W_{dm}} \]

should be kept close to unity.

In the intercept region, \( \lambda = W_d + 3t_{ox} \) is a good approximation.

\[ L_{\text{min}} \sim 1.5 \lambda \sim 20t_{ox} \]
High-k Gate Insulator

High-k gate insulator is an active area of Si research because it may replace SiO$_2$ thereby circumventing the tunneling problem.

But

$$\lambda \sim W_d + (\varepsilon_{Si}/\varepsilon_i)t_i$$

is valid only when $t_i << \lambda$.

In general, requires $t_i < \lambda/2$, regardless of $\varepsilon_i$. 

In the figure:

- The graph shows the normalized gate insulator thickness ($t_i/\lambda$) as a function of the normalized Si depletion depth ($W_d/\lambda$).
- The curves represent different values of $\varepsilon_i/\varepsilon_{Si}$.
- The lines indicate the validity of the approximation $t_i < \theta/2$. 

[Graph showing the relationship between normalized gate insulator thickness and normalized Si depletion depth with different curves for various $\varepsilon_i/\varepsilon_{Si}$ values.]
Because of velocity saturation, the saturation of drain current in a short-channel device occurs at a much lower voltage than $V_{dsat} = (V_g - V_t)/m$ for long channel devices.

This causes the saturation current, $I_{dsat}$, to deviate from the $\propto (V_g - V_t)^2$ behavior and from the $1/L$ dependence.
Velocity-Field Relationship

\[ v = \frac{\mu_{\text{eff}} \mathcal{E}}{1 + \left( \frac{\mathcal{E}}{\mathcal{E}_c} \right)^n}^{1/n} \]

- At low fields, \( v = \mu_{\text{eff}} \mathcal{E} \): Ohm’s law.
- As \( \mathcal{E} \to \infty \), \( v = \nu_{\text{sat}} = \mu_{\text{eff}} \mathcal{E}_c \).

Critical Field: \( \mathcal{E}_c = \frac{\nu_{\text{sat}}}{\mu_{\text{eff}}} \)

It is commonly believed that:
- \( n = 2 \) for electrons, \( n = 1 \) for holes.
- \( \nu_{\text{sat}} \) is independent of \( \mu_{\text{eff}} \) (vertical field), but \( \mathcal{E}_c \) depends on \( \mu_{\text{eff}} \).

Only the \( n = 1 \) case can be solved analytically.
Analytical Solution for n=1

\[ I_{ds} = -WQ_i V = -WQ_i(V) \frac{\mu_{\text{eff}} (dV / dy)}{1 + (\mu_{\text{eff}} / \nu_{\text{sat}})(dV / dy)} \]

Current continuity requires that \( I_{ds} \) be a constant, independent of \( y \).

\[ \Rightarrow \quad I_{ds} = -\left( \mu_{\text{eff}} WQ_i(V) + \frac{\mu_{\text{eff}} I_{ds}}{\nu_{\text{sat}}} \right) \frac{dV}{dy} \]

Multiplying \( dy \) on both sides and integrating from \( y = 0 \) to \( L \) and from \( V = 0 \) to \( V_{ds} \), one solves for \( I_{ds} \):

\[ I_{ds} = \frac{-\mu_{\text{eff}} (W / L) \int_0^{V_{ds}} Q_i(V) dV}{1 + (\mu_{\text{eff}} V_{ds} / \nu_{\text{sat}} L)} \]

Charge-sheet model:
\[ Q_i(V) = -C_{ox} (V_g - V_t - mV) \]

Therefore,
\[ I_{ds} = \frac{\mu_{\text{eff}} C_{ox} (W / L) \left[(V_g - V_t)V_{ds} - (m / 2)V_{ds}^2\right]}{1 + (\mu_{\text{eff}} V_{ds} / \nu_{\text{sat}} L)} \]
The saturation voltage, $V_{dsat}$, can be found by solving $dI_{ds}/dV_{ds} = 0$:

$$ V_{dsat} = \frac{2(V_g - V_t)}{1 + \sqrt{1 + 2\mu_{eff}(V_g - V_t) / (mv_{sat}L)}} $$

And the saturation current is:

$$ I_{dsat} = C_{ox} W_{sat}(V_g - V_t) \sqrt{\frac{1 + 2\mu_{eff}(V_g - V_t) / (mv_{sat}L) - 1}{1 + 2\mu_{eff}(V_g - V_t) / (mv_{sat}L) + 1}} $$

(Dashed: long-ch. model, solid: velocity sat. model)
Velocity-Saturation-Limited Current

At the drain end of the channel when $V_{ds} = V_{dsat}$,

$$Q_i(y = L) = -C_{ox}(V_g - V_t - mV_{dsat})$$

and $I_{dsat} = -Wv_{sat}Q_i(y = L)$,

i.e., carriers move at the saturation velocity.

This implies that $dV/dy \to \infty$ at the drain.

Therefore, the gradual channel approximation breaks down and the carriers are no longer confined to the surface channel.

$$I_{dsat} = C_{ox}Wv_{sat}(V_g - V_t)\sqrt{\frac{1 + 2\mu_{eff}(V_g - V_t)/ (m v_{sat}L) - 1}{1 + 2\mu_{eff}(V_g - V_t)/ (m v_{sat}L) + 1}}$$

When $(V_g - V_t) << mv_{sat}L/2\mu_{eff}$,

$$I_{dsat} = \mu_{eff}C_{ox}\frac{W}{L} \frac{(V_g - V_t)^2}{2m}$$

Long channel limit.

In the limit of $L \to 0$,

$$I_{dsat} = C_{ox}Wv_{sat}(V_g - V_t)$$

Velocity saturation limited current.
Velocity saturation is derived from the drift and diffusion model which assumes that carriers are always in thermal equilibrium with the silicon lattice.

But if the MOSFET is only a few mean free path (~10 nm) long, carriers do not travel enough distance to establish equilibrium ⇒ velocity overshoot, i.e., carrier velocity at the high field region near the drain can exceed the saturation velocity.
The velocity at the source does not greatly exceed $10^7$ cm/s; therefore, current does not greatly exceed $I_{\text{dsat}} = C_{\text{ox}} W V_{\text{sat}} (V_g - V_t)$.

Monte-Carlo simulation:

Even in a 30 nm device, nFET/pFET velocity and therefore current ratio is still $\approx 2$ because of the difference in effective masses.
Distribution Function

Fermi-Dirac distribution under equilibrium:

\[ f(E) = \frac{1}{1 + e^{(E-E_f)/kT}} \]

The standard semi-classical transport theory is based on the Boltzmann transport equation (BTE):

\[
\frac{\partial f}{\partial t} + v \cdot \nabla_r f + \frac{e \mathcal{E}}{\hbar} \cdot \nabla_k f = \sum_{k'} \left\{ S(k', k)f(r, k', t)[1 - f(r, k, t)] - S(k, k')f(r, k, t)[1 - f(r, k', t)] \right\}
\]

where \( r \) is the position, \( k \) is the momentum, \( f(r, k, t) \) is the distribution function, \( v \) is the group velocity, \( \mathcal{E} \) is the electric field, \( S(k, k') \) is the transition probability between the momentum states \( k \) and \( k' \).

The summation on the right hand side is the collision term, which accounts for all the scattering events. The terms on the left hand side indicate, respectively, the dependence of the distribution function on time, space (explicitly related to velocity), and momentum (explicitly related to electric field).
Velocities at the Source and at the Drain

\[ I_{ds} = WQ_i \nu \]

\[ I_{ds} = WC_{ox} (V_g - V_t) \nu_s \]

\[ I_{ds} = WC_{ox} (V_g - V_t - V_{dsat}) \nu_d \]

- Inversion charge density at the source is given by \( C_{ox}(V_g - V_t) \).
- Inversion charge density at the drain is much lower because of the drain bias.
- Current continuity is maintained consistent with band bending.
Scattering theory


At high drain bias, $T'=0$, 

$$I_{ds} = TI^+$$

Let $r=n_s^-/n_s^+$, the backscattering coefficient.

Then

$$I_{ds} / W = C_{ox} (V_g - V_t) v_T \left( \frac{1-r}{1+r} \right)$$

Note that $r$ depends on the low-field mobility near the source.

In the ballistic limit, no collisions in the channel, i.e., $r = 0$, and

$$I_{ds} / W = C_{ox} (V_g - V_t) v_T$$
Carrier Thermal Injection Velocity

For 2-D nondegenerate carriers,

\[ \langle E \rangle = \frac{\int_0^\infty EN(E)f(E)dE}{\int_0^\infty N(E)f(E)dE} = kT \]

so

\[ v_{rms} = \sqrt{\frac{2kT}{m}} \]

For uni-directional injection,

\[ v_T = \frac{\int_0^\infty v_x \exp(-mv_x^2/2kT)dv_x}{\int_0^\infty \exp(-mv_x^2/2kT)dv_x} = \sqrt{\frac{2kT}{\pi m}} \]
Injection Velocity in the Degenerate Case

At 0 K, all states below the Fermi energy are filled. In 2-D, define a Fermi circle with velocity $v_F$.

$$\langle v_x \rangle = \frac{\int_{v_x>0} \int_{v_y>0} v_x \, dv_x \, dv_y}{\int_{v_x>0} \int_{v_y>0} dv_x \, dv_y} = \frac{4}{3\pi} v_F$$

Since

$$\frac{1}{2} N(E) E_F = \frac{m}{\pi \hbar^2} \frac{1}{2} m v_F^2 = n_s = \frac{C_{ox} (V_g - V_t)}{q}$$

$$v_T = \frac{4\hbar}{3m} \sqrt{\frac{2C_{ox} (V_g - V_t)}{q \pi}}$$
I-V Curves of a Ballistic MOSFET


\[ I_{ds}/W = \frac{4\hbar}{3m} \sqrt{\frac{2C_{ox}}{q\pi} C_{ox} (V_g - V_t)^{3/2}} \quad (T=0 \text{ K}) \]

Independent of L!