Review - MOS Capacitor

The "Delta-Depletion Approximation"

Flat-band voltage: \( V_{FB} = V_{GB} \) such that \( \phi(0) = \phi_{p-Si} \):
\[ V_{FB} = \phi_{p-Si} - \phi_m \]

Threshold voltage: \( V_T = V_{GB} \) such that \( \phi(0) = -\phi_{p-Si} \):
\[ V_T = V_{FB} - 2\phi_{p-Si} + [2\varepsilon_{Si} qN_A|2\phi_{p-Si}|]^{1/2}/C_{ox}^* \]

Inversion layer sheet charge density:
\[ q_N^* = -C_{ox}^*[v_{GC} - V_T] \]

Charge stores - \( q_G(v_{GB}) \) from below \( V_{FB} \) to above \( V_T \)

Gate Charge:
\( q_G(v_{GB}) \) from below \( V_{FB} \) to above \( V_T \)

Gate Capacitance:
\( C_{gb}(V_{GB}) \)

Sub-threshold charge:
\( q_N(v_{GB}) \) below \( V_T \)

3-Terminal MOS Capacitors - Bias between B and C

Impact is on \( V_T(v_{BC}) \):
\[ |2\phi_{p-Si}| \rightarrow (|2\phi_{p-Si}| - v_{BC}) \]

MOS Field Effect Transistors - Basics of model

Gradual Channel Model: electrostatics problem normal to channel
drift problem in the plane of the channel
The n-MOS capacitor

Right: Basic device with $v_{BC} = 0$

Below: One-dimensional structure for depletion approximation analysis*

* Note: We can't forget the n+ region is there; we will need electrons, and they will come from there.
MOS Capacitors: Where do the electrons in the inversion layer come from?

Diffusion from the p-type substrate?
If we relied on diffusion of minority carrier electrons from the p-type substrate it would take a long time to build up the inversion layer charge. The current density of electrons flowing to the interface is just the current across a reverse biased junction (the p-substrate to the inversion layer in this case):

\[ J_e = q n_i^2 \frac{D_e}{N_A W_{p,\text{eff}}} \] [Coul/cm\(^2\) - s]

The time, \( \tau \), it takes this flux to build up an inversion charge is the the increase in the charge, \( \Delta q_n \), divided by \( J_e \):

\[ \Delta q_N^* = \frac{\varepsilon_{ox}}{t_{ox}} \Delta (v_{GB} - V_T) \]

so \( \tau \) is

\[ \tau = \frac{\Delta q_N^*}{J_e} = \frac{\varepsilon_{ox} N_A W_{p,\text{eff}}}{q n_i^2 D_e t_{ox}} \Delta (v_{GB} - V_T) \]
Diffusion from the p-type substrate?, cont.
Using $N_A = 10^{18}$ cm$^{-3}$, $t_{ox} = 3$ nm, $w_{p,eff} = 10$ µm, $D_e = 40$ cm$^2$/V
and $\Delta(v_{GB}-V_T) = 0.5$ V in the preceding expression for $\tau$ we find $\tau \approx 50$ hr!

Flow from the adjacent n+-region?
As the surface potential is increased, the potential energy barrier between the adjacent n+ region and the region under the gate is reduced for electrons and they readily flow (diffuse in weak inversion, and drift and diffuse in strong inversion) into the channel; that's why the n+ region is put there:

There are many electrons here and they don't have far to go once the barrier is lowered.
Electrostatic potential and net charge profiles - *regions and boundaries*

Accumulation
\[ V_{GB} < V_{FB} \]

Depletion (Weak Inversion when \( \phi(0) > 0 \))
\[ V_{FB} < V_{GB} < V_T \]

Strong Inversion
\[ V_T < V_{GB} \]

Flat Band Voltage
\[ V_{FB} = \phi_p - \phi_m \]

Threshold Voltage
\[ V_T = V_{FB} + (2\varepsilon_{Si}|2\phi_p|q_NA)^{1/2}/C_{ox}^* \]
**MOS Capacitors**: the gate charge as $v_{GB}$ is varied

\[
q_G^* = \frac{\varepsilon_S q N_A}{C_{ox}^*} \left( \sqrt{1 + \frac{2 C_{ox}^* (v_{GB} - V_{FB})}{\varepsilon_S q N_A}} - 1 \right)
\]

\[
q_G^* = C_{ox}^* (v_{GB} - V_T) + q N_{AP} X_{DT}
\]

The charge expressions:

\[
q_G^* (v_{GB}) = \begin{cases} 
C_{ox}^* (v_{GB} - V_{FB}) \\
\frac{\varepsilon_S q N_A}{C_{ox}^*} \left( \sqrt{1 + \frac{2 C_{ox}^* (v_{GB} - V_{FB})}{\varepsilon_S q N_A}} - 1 \right) \\
C_{ox}^* (v_{GB} - V_T) + q N_A X_{DT}
\end{cases}
\]

for $V_T \leq v_{GB}$
MOS Capacitors: the small signal linear gate capacitance, $C_{gb}(V_{GB})$

$$C_{gb}(V_{GB}) \equiv A \left| \frac{\partial q^*_G}{\partial V_{GB}} \right|_{V_{GB} = V_{BG}}$$

$$C_{gb}(V_{GB}) = \left\{ \begin{array}{ll}
A C_{ox}^* & \text{for } V_{GB} \leq V_{FB} \\
A C_{ox}^* \sqrt{1 + \frac{2C_{ox}^* (V_{GB} - V_{FB})}{\varepsilon_{Si} q N_A}} & \text{for } V_{FB} \leq V_{GB} \leq V_{T} \\
A C_{ox}^* & \text{for } V_{T} \leq V_{GB}
\end{array} \right.$$
**An n-channel MOSFET**

Gradual Channel Approximation: There are two parts to the problem: the vertical electrostatics problem of relating the channel charge to the voltages, and the horizontal drift problem in the channel of relating the channel charge drift to the voltages. We will assume they can be worked independently and in sequence.
Gradual Channel Approximation:
- We first solve a one-dimensional electrostatics problem in the x direction to find the channel charge, $q_{N}^{*}(y)$.
- Then we solve a one-dimensional drift problem in the y direction to find the channel current, $i_{D}$, as a function of $v_{GS}$, $v_{DS}$, and $v_{BS}$. 
Gradual Channel Approximation i-v Modeling

*(n-channel MOS used as the example)*

We restrict voltages to the following ranges:

\[ v_{BS} \leq 0, \quad v_{DS} \geq 0 \]

This means that the source-substrate and drain-substrate junctions are always reverse biased and thus that:

\[ i_B(v_{GS}, v_{DS}, v_{BS}) \approx 0 \]

The gate oxide is insulating so we also have:

\[ i_G(v_{GS}, v_{DS}, v_{BS}) \approx 0 \]

With the back current, \( i_B \), zero, and the gate current, \( i_G \), zero, the only current that is not trivial to model is \( i_D \).

The drain current, \( i_D \), is also zero except when when \( v_{GS} > V_T \).

The **in-plane** problem: (for just a minute so we can see where we’re going)

Looking at electron drift in the channel we write \( i_D \) as

\[
  i_D = -W s_{ey}(y) q_n^*(y) = -W \mu_e q_n^*(v_{GS}, v_{BS}, v_{CS}(y)) \frac{dv_{CS}}{dy} \quad \text{(at moderate E - fields)}
\]

This can be integrated from \( y = 0 \) to \( y = L \) (\( v_{CS} = 0 \) to \( v_{CS} = v_{DS} \)) to get \( i_D(v_{GS}, v_{DS}, v_{BS}) \), but first we need \( q_n^*(y) \).

We get \( q_n^*(y) \) from our x-direction electrostatics analysis.
Gradual Channel Approximation i-v Modeling

(n-channel MOS used as the example)

The Gradual Channel Approximation is the approach typically used to model the drain current in field effect transistors.*

It assumes that the drain current, \( i_D \), consists entirely of carriers flowing in the channel of the device, and is thus proportional to the sheet density of carriers at any point and their net average velocity. It is not a function of \( y \), but its components in general are:

\[
i_D = -W \cdot -q \cdot n_{ch}^*(y) \cdot \bar{s}_e(y)
\]

In this expression, \( W \) is the width of the device, \(-q\) is the charge on each electron, \( n_{ch}^*(y) \) is sheet electron concentration in the channel (i.e. electrons/cm\(^2\)) at \( y \), and \( \bar{s}_e(y) \) is the net electron velocity in the \( y \)-direction.

If the electric field is not too large, \( \bar{s}_e(y) = -\mu_e E_y(y) \), and

\[
i_D = -W \cdot q \cdot n_{ch}^*(y) \cdot \mu_e E_y(y) = W \cdot q \cdot n_{ch}^*(y) \cdot \mu_e \frac{dv_{CS}(y)}{dy}
\]

* Junction FETs (JFETs), MEtal Semiconductor FETs (MESFETs\(^1\)), and Hetero junction FETs (HJFETs\(^2\)), as well as Metal Oxide Semiconductor FETs (MOSFETs).

1. Also called Shottky Barrier FETs (SBFETs).  2. Includes HEMTs, TEGFETs, MODFETs, SDFETs, HFETs, PHEMTs, MHEMTs, etc.
To eliminate the derivative from this equation we integrate both sides with respect to $y$ from the source ($y = 0$) to the drain ($y = L$). This corresponds to integrating the right hand side with respect to $v_{CS}$ from $0$ to $v_{DS}$, because $v_{CS}(0) = 0$ to $v_{CS}(L) = v_{DS}$:

$$
\int_{0}^{L} i_D \, dy = W \cdot \mu_e \cdot q \cdot \int_{0}^{L} n_{ch}^*(y) \frac{dv_{CS}(y)}{dy} \, dy = W \cdot \mu_e \cdot q \cdot \int_{0}^{v_{DS}} n_{ch}(v_{CS}) \, dv_{CS}
$$

The left hand integral is easy to evaluate; it is simply $i_D L$. Thus we have:

$$
\int_{0}^{L} i_D \, dy = i_D L \quad \Rightarrow \quad i_D = \frac{W}{L} \cdot \mu_e \cdot q \cdot \int_{0}^{v_{DS}} n_{ch}(v_{CS}) \, dv_{CS}
$$
The various FETs differ primarily in the nature of their channels and thereby, the expressions for $n_{ch}^*(y)$.

For a MOSFET we speak in terms of the inversion layer charge, $q_n^*(y)$, which is equivalent to $-q \cdot n_{ch}^*(y)$. Thus we have:

$$i_D = - \frac{W}{L} \mu_e \int_0^{v_{DS}} q_n^*(v_{GS}, v_{CS}, v_{BS}) \, dv_{CS}$$

We derived $q_n^*$ earlier by solving the vertical electrostatics problem, and found:

$$q_n^*(v_{GS}, v_{CS}, v_{BS}) = - C_{ox}^* \left[ v_{GS} - v_{CS} - V_T(v_{CS}, v_{BS}) \right]$$

with

$$V_T(v_{CS}, v_{BS}) = V_{FB} - 2\phi_{p-Si} + \left\{ 2\varepsilon_{Si} qN_A \left[ 2\phi_{p-Si} - v_{BS} + v_{CS} \right] \right\}^{1/2} / C_{ox}^*$$

Using this in the equation for $i_D$, we obtain:

$$i_D(v_{GS}, v_{DS}, v_{BS}) = \frac{W}{L} \mu_e \int_0^{v_{DS}} \left\{ C_{ox}^* \left[ v_{GS} - v_{CS} - V_T(v_{CS}, v_{BS}) \right] \right\} \, dv_{CS}$$

At this point we can do the integral, but it is common to simplify the expression of $V_T(v_{CS}, v_{BS})$ first, to get a more useful result.
GCA - dealing with the non-linear dependence of $V_T$ on $v_{CS}$

**Approach #1 - Live with it**

Even though $V_T(v_{CS}, v_{BS})$ is a non-linear function of $v_{CS}$, we can still put it in this last equation for $i_D$:

$$i_D = \frac{W}{L} \mu_e \int_0^{v_{DS}} \left\{ C_{ox}^* \left[ v_{GS} - v_{CS} - V_{FB} + 2\phi_{p-Si} - \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{2\varepsilon_{Si}qN_A \left[ 2\phi_{p-Si} - v_{BS} + v_{CS} \right]} \right] \right\} dv_{CS}$$

and do the integral, obtaining:

$$i_D(v_{DS}, v_{GS}, v_{BS}) = \frac{W}{L} \mu_e C_{ox}^* \left( v_{GS} - \left| 2\phi_p \right| - V_{FB} - \frac{v_{DS}}{2} \right) v_{DS}$$

$$+ \frac{3}{2} \sqrt{2\varepsilon_{Si}qN_A} \left[ \left( 2\phi_p + v_{DS} - v_{BS} \right)^{3/2} - \left( 2\phi_p - v_{BS} \right)^{3/2} \right]$$

The problem is that this result is very unwieldy, and difficult to work with. More to the point, we don't have to live with it because it is easy to get very good, approximate solutions that are much simpler to work with.
GCA - dealing with the non-linear dependence of \( V_T \) on \( v_{CS} \)

**Approach #2 - Ignore it**

Early on researchers noticed that the difference between \( V_T \) at 0 and at \( y \), i.e. \( V_T(0,v_{BS}) \) and \( V_T(v_{DS},v_{BS}) \), is small, and that using \( V_T(0,v_{BS}) \) alone gives a result that is still quite accurate and is very easy to use:

\[
\begin{align*}
    i_D(v_{GS},v_{DS},v_{BS}) &= \frac{W}{L} \mu_e \int_0^{v_{DS}} \left\{ C_{ox}^* [v_{GS} - v_{CS} - V_T(0,v_{BS})] \right\} dv_{CS} \\
    &= \frac{W}{L} \mu_e C_{ox}^* \left\{ [v_{GS} - V_T(0,v_{BS})] v_{DS} - \frac{v_{DS}^2}{2} \right\}
\end{align*}
\]

This result looks much simpler than the result of Approach #1, and it is much easier to use in hand calculations. It is, in fact, the one most commonly used by the vast majority of engineers. At the same time, the fact that it was obtained by ignoring the dependence of \( V_T \) on \( v_{CS} \) is cause for concern, unless we have a way to judge the validity of our approximation. We can get the necessary metric through Approach #3.
**Approach #3** - Linearize it (i.e. expand it, keep first order term)
In this approach we leave the variation of $V_T$ with $v_{CS}$ in, but linearize it by doing a Taylor's series expansion about $v_{CS} = 0$:

$$V_T[v_{CS}, v_{BS}] \approx V_T(0, v_{BS}) + \frac{\partial V_T}{\partial v_{CS}}|_{v_{CS}=0} \cdot v_{CS}$$

Taking the derivative and evaluating it at $v_{CS} = 0$ yields:

$$V_T[v_{CS}, v_{BS}] \approx V_T(0, v_{BS}) + \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{\frac{\varepsilon_S q N_A}{2(2\phi_p - v_{BS})}} \cdot v_{CS}$$

With this $q_n^*$ is

$$q_n^*(v_{GS}, v_{CS}, v_{BS}) \approx -C_{ox}^* \left[ v_{GS} - v_{CS} + V_T(0, v_{BS}) - \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{\frac{\varepsilon_S q N_A}{2(2\phi_p - v_{BS})}} \cdot v_{CS} \right]$$

$$= -C_{ox}^* \left[ v_{GS} - \alpha v_{CS} + V_T(v_{BS}) \right]$$

where

$$\alpha \equiv 1 + \frac{t_{ox}}{\varepsilon_{ox}} \sqrt{\frac{\varepsilon_S q N_A}{2(2\phi_p - v_{BS})}}$$

and $V_T(v_{BS}) \equiv V_T(0, v_{BS})$
GCA - dealing with the non-linear dependence of $V_T$ on $v_{CS}$

Using this result in the integral in the expression for $i_D$ gives:

$$i_D(v_{GS},v_{DS},v_{BS}) = \frac{W}{L} \mu_e \int_0^{v_{DS}} \{ C_{ox}^* [v_{GS} - \alpha v_{CS} - V_T(0,v_{BS})] \} \, dv_{CS}$$

$$= \frac{W}{L} \mu_e C_{ox}^* \left\{ [v_{GS} - V_T(v_{BS})] v_{DS} - \alpha \frac{v_{DS}^2}{2} \right\}$$

Plotting this equation for increasing values of $v_{GS}$ we see that it traces inverted parabolas as shown below.

Note: $i_D$ saturates after its peak value (solid lines), rather than decreasing (dashed lines).
Gradual Channel Approximation, cont.

The drain current expression, cont:

The point at which $i_D$ reaches its peak value and saturates is easily found. Taking the derivative and setting it equal to zero we find:

$$\frac{\partial i_D}{\partial v_{DS}} = 0 \quad \text{when} \quad v_{DS} = \frac{1}{\alpha} \left[ v_{GS} - V_T(v_{BS}) \right]$$

What happens physically at this voltage is that the channel (inversion) at the drain end of the channel disappears:

$$q_n^*(L) \approx -C_{ox}^* \left\{ v_{GS} - V_T(v_{BS}) - \alpha v_{DS} \right\}$$

$$= 0 \quad \text{when} \quad v_{DS} = \frac{1}{\alpha} \left[ v_{GS} - V_T(v_{BS}) \right]$$

For $v_{DS} > \left[ v_{GS} - V_T(v_{BS}) \right]/\alpha$, all the additional drain-to-source voltage appears across the high resistance region at the drain end of the channel where the mobile charge density is very small, and $i_D$ remains constant independent of $v_{DS}$:

$$i_D(v_{GS}, v_{DS}, v_{BS}) = \frac{1}{2\alpha} \frac{W}{L} \mu_e C_{ox}^* \left[ v_{GS} - V_T(v_{BS}) \right]^2 \quad \text{for} \quad v_{DS} > \frac{1}{\alpha} \left[ v_{GS} - V_T(v_{BS}) \right]$$
Gradual Channel Approximation, cont.

The full model, cont:

\[ i_G(v_{GS}, v_{DS}, v_{BS}) = 0 \quad i_B(v_{GS}, v_{DS}, v_{BS}) = 0 \]

\[ i_D(v_{GS}, v_{DS}, v_{BS}) = \begin{cases} 
0 & \text{for } [v_{GS} - V_T(v_{BS})] < 0 < \alpha v_{DS} \\
\frac{K}{2}[v_{GS} - V_T(v_{BS})]^2 & \text{for } 0 < [v_{GS} - V_T(v_{BS})] < \alpha v_{DS} \\
K \left\{v_{GS} - V_T(v_{BS}) - \frac{\alpha v_{DS}}{2}\right\} \alpha v_{DS} & \text{for } 0 < \alpha v_{DS} < [v_{GS} - V_T(v_{BS})] 
\end{cases} \]

Saturation or Forward Active Region

Linear or Triode Region

Cutoff Region

with \( K = \frac{W}{\alpha L} \mu_e C_{ox}^{*} \)
The operating regions of MOSFETs and BJTs:
Comparing an n-channel MOSFET and an npn BJT

MOSFET

\[ i_D \approx K [v_{GS} - V_T(v_{BS})]^2/2\alpha \]

BJT

\[ i_B \approx I_{BS} e^{V_{BE}/kT} \]

\[ v_{CE} > 0.2 \text{ V} \]
p-channel MOSFET's: The other "flavor" of MOSFET

Structure:

The voltage progression:

Gradual channel model*:

Valid for $v_{SB} \leq 0$, and $v_{SD} \geq 0$: 

$$-i_D(v_{SG},v_{SD},v_{SB}) = \begin{cases} 0 & \text{for } [v_{SG} - |V_T(v_{SB})|] < 0 < \alpha v_{SD} \\ \frac{1}{2 \alpha L} \, \mu_e \, C_{ox}^* \left[ v_{SG} - |V_T(v_{SB})| \right]^2 & \text{for } 0 < [v_{SG} - |V_T(v_{SB})|] < \alpha v_{SD} \\ \frac{W}{\alpha L} \, \mu_e \, C_{ox}^* \left\{ v_{SG} - [V_T(v_{SB})] - \frac{\alpha v_{SD}}{2} \right\} \alpha v_{SD} & \text{for } 0 < \alpha v_{SD} < [v_{SG} - |V_T(v_{SB})|] \end{cases}$$

$$V_T(v_{SB}) = V_{FB} - 2 \phi_{n-Si} - \gamma [2 \phi_{n-Si} - v_{SB}]^{1/2} \quad \text{with} \quad \gamma = \frac{1}{C_{ox}^*} \left[ 2 \varepsilon_{Si} q N_D \right]^{1/2}$$

* Enhancement mode only, $V_T$ (i.e. $v_{GS}$ at threshold) < 0.
p-channel MOSFET's: cont.

**p-channel**

Symbol and FAR model:
Oriented with source down like n-channel:

![Symbol and FAR model](image)

Oriented as found in circuits:

![Symbol and FAR model](image)

Structure:

![Structure diagram](image)

Symbol:

![Symbol diagram](image)

FAR model:

\[ v_{GS} < V_T \]
\[ v_{BS} > 0 \]
\[ v_{DS} < 0 \]

FAR model:

\[ v_{SG} > -V_T \]
\[ v_{SB} < 0 \]
\[ v_{SD} > 0 \]
Depletion mode MOSFET's: The very last MOSFET variant

It is possible to have n-channel MOSFETs with $V_T < 0$.
In this situation the channel exists with $v_{GS} = 0$, and a negative bias must be applied to turn it off.
This type of device is called a "depletion mode" MOSFET.
Devices with $V_T > 0$ are "enhancement mode."

For a p-channel depletion mode MOSFET, $V_T > 0$.
The expressions for $i_D(v_{GS}, v_{DS}, v_{BS})$ are exactly the same
for enhancement mode and depletion mode MOSFETs.
**6.012 - Microelectronic Devices and Circuits**

**Lecture 10 - MOSFET Basics - Summary**

- **Qualitative operation** - the MOSFET as a switch and transistor

- **Quantitative modeling** - the Gradual-Channel Approximation

Restrict voltage ranges: \( v_{BS} \leq 0; v_{DS} \geq 0 \) (n-channel MOS used as the example)

No gate and substrate currents: \( i_G(v_{GS}, v_{DS}, v_{BS}) \approx 0, i_B(v_{GS}, v_{DS}, v_{BS}) \approx 0 \)

The drain current: \( i_D(v_{GS}, v_{DS}, v_{BS}) \)  

1. **The in-plane problem**: \( i_D = -W \mu_e q_n^* \int v_{CS} \, dy \); this is integrated from \( y = 0 \) to \( y = L \), and \( v_{CS} = 0 \) to \( v_{CS} = v_{DS} \) to get \( i_D(v_{GS}, v_{DS}, v_{BS}) \)
2. **The normal problem**: \( q_n^*(y) \approx -C_{ox}^* [v_{GS} - V_T(v_{BS}) - \alpha v_{CS}(y)] \), where  
   \( V_T(v_{BS}) = V_{FB} - 2\phi_{p-Si} + \gamma(2\phi_{p-Si} - v_{BS})^{1/2} \), \( \gamma = (2e, qN_A)^{1/2} / C_{ox}^* \),  
   and \( \alpha = 1 + \gamma / 2(2\phi_{p-Si} - v_{BS})^{1/2} \) (many texts set \( \alpha = 1 \))
3. **The full drain current expressions**:

\[
  i_D \approx \begin{cases} 
  0 & \text{for } (v_{GS} - V_T) \leq 0 \leq \alpha v_{DS} \\
  (W/L)\mu_e C_{ox}^* (v_{GS} - V_T)^2 / 2\alpha & \text{for } 0 \leq (v_{GS} - V_T) \leq \alpha v_{DS} \\
  (W/L)\mu_e C_{ox}^* (v_{GS} - V_T - \alpha v_{DS} / 2)v_{DS} & \text{for } 0 \leq \alpha v_{DS} \leq (v_{GS} - V_T) 
  \end{cases}
\]

with \( V_T = V_{FB} - 2\phi_{p-Si} + [2\varepsilon_{Si} qN_A(2\phi_{p-Si} - v_{BS})]^{1/2} / C_{ox}^* \) and \( \alpha = 1 + [(\varepsilon_{Si} qN_A / 2(2\phi_{p-Si} - v_{BS}))^{1/2} / C_{ox}^*] \) (often \( \alpha \approx 1 \))