2026-02-09

Weak Inversion vs. Strong Inversion

A MOSFET can conduct in both weak inversion (subthreshold) and strong inversion. The key difference is the transport mechanism:

- **Weak inversion:** no strong inversion layer; current is dominated by diffusion, and I_D depends exponentially on V_{GS}.
- **Strong inversion:** a well-formed inversion layer exists; current is dominated by drift, and long-channel devices follow the square-law model.

We will use the standard notation:

$$V_{TH} \text{ (threshold voltage)}, \qquad V_T = \frac{kT}{q} \text{ (thermal voltage)}.$$

Transconductance g_m

Transconductance measures how effectively a device converts a small gate-voltage change into a drain-current change. In other words, it's the slope on V_{GS} vs. I_D plot under constant V_{DS}.

$$\boxed{g_m \equiv \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{V_{DS}\text{ const}}.} \tag{1}$$

Strong inversion (long-channel, saturation)

Ignoring channel-length modulation for the moment, the long-channel saturation current is

$$\boxed{\begin{gathered} I_D \approx \frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}\left(V_{GS}-V_{TH}\right)^2 \\ (\text{When } V_{DS} \geq V_{GS}-V_{TH}). \end{gathered}} \tag{2}$$

Differentiating w.r.t. V_{GS} gives

$$\boxed{\begin{aligned} g_m &= \mu_n C_{\text{ox}} \frac{W}{L}\left(V_{GS}-V_{TH}\right) \\ &= \sqrt{2\mu_n C_{\text{ox}} \frac{W}{L} I_D} \\ &= \frac{2I_D}{V_{GS}-V_{TH}} \equiv \frac{2I_D}{V_{OV}}, \end{aligned}} \tag{3}$$

where $V_{OV} \equiv V_{GS}-V_{TH}$ is the overdrive voltage.

Weak inversion (subthreshold)

In weak inversion, the drain current can be modeled as

$$I_D \propto \exp\left(\frac{V_{GS}}{nV_T}\right),$$

where n is the subthreshold slope factor. Differentiation yields the well-known result

$$\boxed{g_m = \frac{I_D}{nV_T}.} \tag{4}$$

BJT comparison (useful benchmark)

For a bipolar transistor,

$$\boxed{\begin{aligned} I_C &= I_S e^{V_{BE}/V_T}, \\ g_m &= \frac{\partial I_C}{\partial V_{BE}} = \frac{I_C}{V_T}. \end{aligned}} \tag{5}$$

Thus, at the same current level, a MOSFET in weak inversion has g_m smaller than a BJT by approximately the factor $n > 1$.

Output conductance g_o and output resistance r_o (channel-length modulation)

Channel-length modulation (CLM) introduces a dependence of I_D on V_{DS} in saturation:

$$\boxed{I_D \approx \frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}\left(V_{GS}-V_{TH}\right)^2\left(1+\lambda V_{DS}\right),} \tag{6}$$

where λ is the CLM coefficient.

The small-signal output conductance is

$$\boxed{g_o \equiv \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{GS}} \approx \lambda I_D,} \tag{7}$$

and hence

$$\boxed{r_o \equiv \frac{1}{g_o} \approx \frac{1}{\lambda I_D}.} \tag{8}$$

A key scaling trend is that longer channels exhibit weaker CLM:

$$\lambda \propto \frac{1}{L} \quad \Rightarrow \quad r_o \text{ increases with } L.$$

Back-gate (body-effect) transconductance g_{mb}

When the source-to-bulk voltage V_{SB} changes, the threshold voltage changes (body effect), which in turn modulates I_D. Define

$$\boxed{g_{mb} \equiv \left.\frac{\partial I_D}{\partial V_{SB}}\right|_{V_{GS},V_{DS}}.} \tag{9}$$

Using the chain rule,

$$g_{mb} = \frac{\partial I_D}{\partial V_{TH}}\frac{\partial V_{TH}}{\partial V_{SB}} \tag{10}$$

In saturation (strong inversion), I_D depends on $V_{GS}-V_{TH}$, so

$$\frac{\partial I_D}{\partial V_{TH}} = -\frac{\partial I_D}{\partial V_{GS}} = -g_m,$$

and therefore

$$\boxed{g_{mb} = -g_m \frac{\partial V_{TH}}{\partial V_{SB}}.} \tag{11}$$

Since $\partial V_{TH}/\partial V_{SB} > 0$, g_{mb} is positive in magnitude (i.e., increasing V_{SB} increases V_{TH} and reduces I_D).

Threshold voltage with body effect

A common long-channel model is

$$V_{TH0} = V_{FB} + 2|\phi_F| + \frac{\sqrt{2q\varepsilon_{\mathrm{si}} N_A \left(2|\phi_F|\right)}}{C_{\mathrm{ox}}}, \tag{12}$$

and for $V_{SB} \geq 0$,

$$\boxed{V_{TH}(V_{SB}) = V_{TH0} + \gamma\left(\sqrt{2|\phi_F| + V_{SB}} - \sqrt{2|\phi_F|}\right)} \tag{13}$$

where γ is the body-effect coefficient.

Differentiating gives

$$\boxed{\frac{\partial V_{TH}}{\partial V_{SB}} = \frac{\gamma}{2\sqrt{2|\phi_F| + V_{SB}}}.} \tag{14}$$

Combining with $\partial I_D/\partial V_{TH} = -g_m$ yields the standard relationship

$$\boxed{\begin{aligned} g_{mb} &= g_m \frac{\gamma}{2\sqrt{2|\phi_F| + V_{SB}}} \\ \rightarrow \frac{g_{mb}}{g_m} &= \frac{\gamma}{2\sqrt{2|\phi_F| + V_{SB}}}. \end{aligned}} \tag{15}$$

Small-signal picture (reminder)

In the MOSFET small-signal model, body effect appears as an additional controlled current source $g_{mb}v_{bs}$ in parallel with the main $g_m v_{gs}$ source, both between drain and source, together with r_o accounting for CLM.

MOSFET Gate Capacitances and Charge-Based Modeling

In analog and high-frequency analysis, MOSFET capacitances are most accurately derived using a charge-based formulation. Rather than assigning fixed lumped capacitors, we begin with the distributed inversion charge along the channel and obtain terminal capacitances by differentiation with respect to terminal voltages.

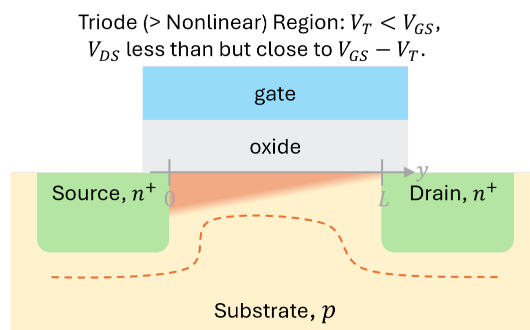

Figure 1: Cross-sectional device diagram of an n-channel MOSFET operating in the triode (linear) region, illustrating the inversion charge distribution along the channel from source ($y = 0$) to drain ($y = L$).

Figure 1 illustrates an nMOS transistor operating in the triode (linear) region. The channel coordinate y is defined from source ($y = 0$) to drain ($y = L$), and the inversion charge density varies along the channel due to the lateral potential $V(y)$.

Definition of gate capacitance

The intrinsic gate-to-source capacitance is defined as

$$C_{GS,i} \equiv \frac{\partial Q_T}{\partial V_{GS}}, \tag{16}$$

where Q_T is the total inversion charge in the channel.

Similarly, intrinsic gate-to-drain and gate-to-bulk capacitances are obtained by differentiating Q_T with respect to V_{GD} and V_{GB}, respectively.

Overlap and fringing capacitances

In addition to channel charge, parasitic capacitances arise from gate overlap and fringing fields. These are typically modeled as geometry-dependent constants:

$$\begin{aligned} C_{GS0} &= C_{ov,GS}\, W, \\ C_{GD0} &= C_{ov,GD}\, W, \\ C_{GB0} &= A_0\, C_{F0}, \end{aligned} \tag{17}$$

where C_{ov} denotes overlap capacitance per unit width, C_{F0} is the fringing capacitance per unit area, and A_0 is the effective fringing area.

Channel charge distribution

In strong inversion, the local inversion charge density per unit area at position y is

$$Q_m(y) = -\,C_{ox}\big(V_{GS} - V_T - V(y)\big), \tag{18}$$

where C_{ox} is the oxide capacitance per unit area and $V(y)$ is the channel potential referenced to the source. The negative sign reflects electron charge.

The total inversion charge is obtained by integrating over the channel area:

$$Q_T = \iint Q_m(y)\,dz\,dy = W\int_0^L Q_m(y)\,dy. \tag{19}$$

Triode-region result

In the triode region, the channel potential varies approximately linearly,

$$V(y) \approx \frac{V_{DS}}{L}\,y,$$

and substituting into the expression for Q_T yields

$$\begin{aligned} Q_T &= -WC_{ox}\int_0^L \left(V_{GS} - V_T - \frac{V_{DS}}{L}y\right)dy \\ &= -WC_{ox}\left[(V_{GS} - V_T)\,L - \frac{1}{2}V_{DS}L\right]. \end{aligned} \tag{20}$$

Differentiating with respect to V_{GS} gives the intrinsic gate-to-source capacitance:

$$\boxed{C_{GS,i} = \frac{\partial Q_T}{\partial V_{GS}} = WLC_{ox}.} \tag{21}$$

Physical interpretation

This result shows that in the triode region, the intrinsic gate capacitance is equal to the full oxide capacitance over the channel area. The gate charge is shared between source and drain, but the total channel charge responds linearly to V_{GS}. More refined models later partition this charge into C_{GS} and C_{GD} components depending on operating region (triode vs. saturation).

This charge-based framework forms the foundation for accurate small-signal and high-frequency MOSFET models used in analog circuit design.

Current Mirror

A current mirror is a circuit that copies (or scales) a reference current into another branch while maintaining high output resistance. In MOS implementations, the mirroring action arises from enforcing identical gate–source voltages across matched transistors.

Physically, the mirror relies on the strong-inversion square-law relationship, so that two devices sharing the same V_{GS} will ideally conduct proportional currents if their (W/L) ratios are matched.

$$\boxed{I_D = \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS})} \tag{22}$$

A good current mirror should satisfy:

- I_{out} accurately tracks I_{REF} (good current transfer ratio).
- I_{out} is largely independent of V_{out} once in saturation (high output resistance).
- The minimum required V_{out} to sustain saturation is small (low compliance voltage).

Simple Current Mirror

The simplest MOS current mirror consists of two matched transistors. Transistor M_1 is diode-connected (gate tied to drain), and a reference current I_{REF} forces a gate–source voltage V_{GS1}. This same V_{GS1} is applied to M_2.

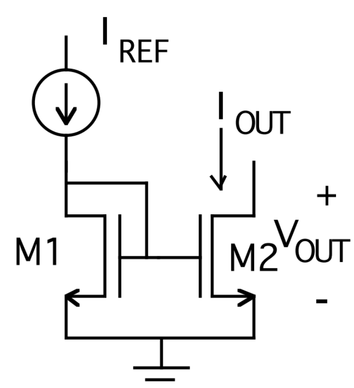

Figure 2: Circuit diagram of a simple current mirror.

Assuming matched devices ($\left(\frac{W}{L}\right)_1 = \left(\frac{W}{L}\right)_2$) and both operate in saturation. Since M_1 is diode-connected and $V_{DS1} = V_{GS1}$.

$$\begin{aligned} \text{For } M_1, \quad I_{\text{REF}} &= \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS1} - V_T)^2\,(1 + \lambda V_{GS1}) \\ \text{For } M_2, \quad I_{\text{out}} &= \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS1} - V_T)^2\,(1 + \lambda V_{\text{out}}) \end{aligned} \tag{23}$$

Therefore,

$$\boxed{\frac{I_{\text{out}}}{I_{\text{REF}}} = \frac{1 + \lambda V_{\text{out}}}{1 + \lambda V_{GS1}}} \tag{24}$$

If channel-length modulation is neglected ($\lambda = 0$), then

$$I_{\text{out}} = I_{\text{REF}}.$$

However, due to channel-length modulation:

- I_{out} depends weakly on V_{out}.
- The mirror exhibits finite output resistance.

Output Resistance In saturation, the small-signal output resistance r_{o2} of M_2 is $\frac{1}{\lambda I_D}$. Thus, the output resistance of the simple mirror is approximately

$$R_{\text{out}} \approx r_{o2} = \frac{1}{\lambda I_D}. \tag{25}$$

This limited output resistance causes the I_{out}–V_{out} curve to have a finite slope rather than being perfectly flat.

Compliance Voltage To maintain saturation in M_2:

$$V_{GS1} - V_T = V_{OV} \le V_{\text{out}} \tag{26}$$

Hence, the minimum output voltage required is approximately one overdrive voltage.

Cascode Current Mirror

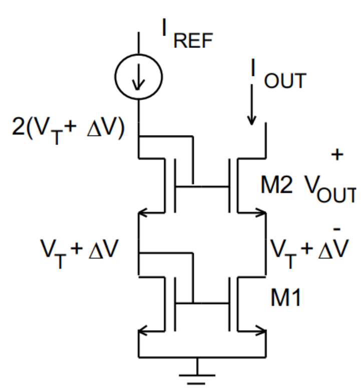

Figure 3: Circuit diagram of a cascode current mirror.

As shown in Figure 3, the cascode current mirror consists of a basic current mirror (M_1) stacked with a cascode device (M_2). The bias network establishes the gate voltages such that, at the reference current I_{REF},

$$\Delta V \quad V_{GS} - V_T \qquad \text{(evaluated at } I_D = I_{\text{REF}}). \tag{27}$$

The purpose of the cascode device is to reduce the variation of V_{DS1}, thereby increasing the small-signal output resistance. This comes at the cost of reduced voltage headroom.

Large-Signal Operation We examine the operation regions as a function of V_{OUT}.

1. M_1 and M_2 in Triode Region For sufficiently small V_{OUT}, both devices operate in triode. The boundary voltage V_1 is obtained by enforcing the triode condition for M_1:

$$V_{\text{OUT}} \le V_1 \; = \; \Delta V + \frac{\Delta V^2}{2\,(V_T + \Delta V)}. \tag{28}$$

In this region, the mirror no longer behaves as an ideal current source since both channel resistances strongly depend on V_{OUT}.

2. M_1 Saturated, M_2 in Triode For intermediate output voltage,

$$V_1 \le V_{\text{OUT}} \le V_T + 2\Delta V, \tag{29}$$

M_1 enters saturation while M_2 remains in triode. The current becomes less sensitive to V_{OUT}, but the cascode action is not yet fully established.

3. M_1 and M_2 in Saturation Full cascode operation occurs when

$$V_{\text{OUT}} \ge V_T + 2\Delta V. \tag{30}$$

In this regime, both devices satisfy

$$V_{DSi} \ge V_{GSi} - V_T, \tag{31}$$

and the output branch behaves as a high-resistance current source.

Drawback: Reduced Voltage Swing The minimum compliance voltage is approximately

$$V_{\text{OUT,min}} \approx V_T + 2\Delta V, \tag{32}$$

which can be prohibitively large in low-supply environments (e.g., $V_{DD} \sim 1$ V). Thus, while the cascode mirror improves output resistance, it reduces voltage headroom.

Small-Signal Analysis To quantify the improvement in output resistance, we consider the small-signal equivalent circuit.

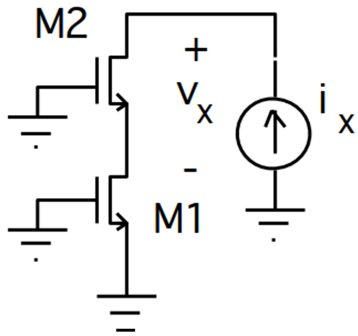

Figure 4: Small-signal equivalent circuit of the cascode mirror.

As shown in Figure 4, the output node sees:

- r_{o2} from M_2,
- r_{o1} from M_1,
- controlled sources $g_{m2}v_{gs2}$ and $g_{mb2}v_{bs2}$.

Injecting a test current i_x and measuring v_x, the output resistance is

$$\boxed{R_o = \frac{v_x}{i_x} = r_{o2}\left\{1 + \frac{r_{o1}}{r_{o2}} + (g_{m2} + g_{mb2})\, r_{o1}\right\}} \tag{33}$$

For $g_{m2}r_{o1} \gg 1$ and neglecting body effect,

$$R_o \approx g_{m2}\, r_{o2}\, r_{o1}. \tag{34}$$

Thus, compared to a simple mirror ($R_o \approx r_{o1}$), the output resistance is increased by approximately the intrinsic gain factor

$$a_v \approx g_{m2} r_{o2}. \tag{35}$$

Equivalent Interpretation Figure 5 illustrates the conceptual equivalent model: the lower transistor provides a high-resistance current source, while the upper transistor acts as a gain stage that amplifies variations at the intermediate node and suppresses output voltage fluctuations.

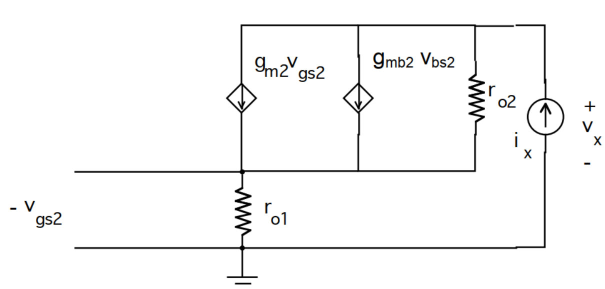

Figure 5: Conceptual equivalent model showing gain-boosted output resistance.

Physically, any increase in v_x produces a change in the intermediate node voltage that reduces v_{ds1}, counteracting the change in i_D. This negative feedback mechanism is responsible for the multiplication of output resistance.

Summary

- The cascode mirror increases output resistance by roughly $g_m r_o$.
- The improvement stems from local gain (intrinsic amplification).
- The price paid is reduced output voltage swing.