The g_m/I_D Design Methodology

BY MOHAMED WATFA

April 17, 2025

Abstract

As CMOS technologies have scaled down, their behavior increasingly deviates from the predictions made by the **square-law model**. As a result, there is a compelling need for an accurate yet tractable design **methodology** suitable for modern technologies. The g_m/I_D methodology is one approach that addresses this necessity.

1 Recap of MOSFET Square-Law Model Behavior

1.1 Regions of Operation

The region of operation of a MOSFET depends on two sets of voltages:

1. The **gate-to-source voltage** (V_{gs}) and the **threshold voltage** (V_t):
   - These voltages define the overdrive voltage $V_{ov} = V_{gs} - V_t$, which determines the inversion level (strong, moderate, weak) of the MOSFET.
2. The **drain-to-source voltage** (V_{ds}) and the **drain-to-source voltage at saturation** ($V_{ds,sat}$):
   - Given an inversion level, these voltages specify whether the MOSFET is biased in the triode region or in the saturation region.

1.2 Large Signal Currents

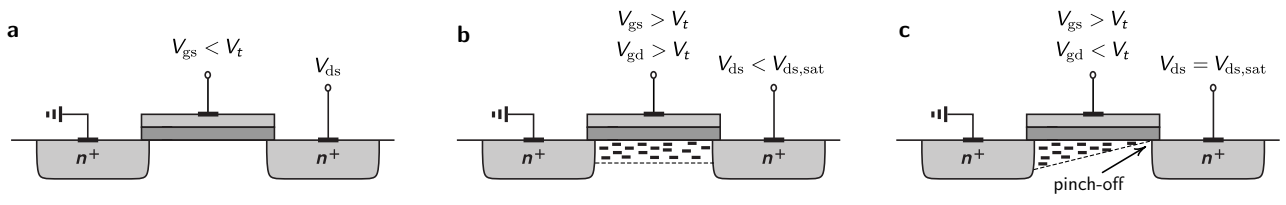

Figure 1. Regions of operation of an n-channel MOSFET: (a) cutoff region, (b) linear region, and (c) onset of saturation.

Consider an n-channel device:

1. **Cutoff Region**: When $V_{gs} < V_t$ (or $V_{ov} < 0$), no conductive channel exists between the source and the drain, and no current flows through the MOSFET.

$$I_{ds} = 0 \tag{1}$$

2. **Triode Region**: When $V_{gs} > V_t$ (or $V_{ov} > 0$) and $V_{ds} < V_{gs} - V_t$, a conductive channel extends uniformly between the source and the drain, and the MOSFET behaves as a non-linear voltage-controlled resistor.

$$I_{ds} = \mu_n C_{ox} \frac{W}{L} \left[\left(\underbrace{V_{gs} - V_t}_{V_{ov}} \right) V_{ds} - \frac{V_{ds}^2}{2} \right] \tag{2}$$

3. **Saturation Region**: When $V_{gs} > V_t$ (or $V_{ov} > 0$) and $V_{ds} \geq V_{gs} - V_t$, the channel pinches off near the drain, causing the current to become independent of V_{ds}. As a result, the MOSFET behaves as an ideal current source (since it is independent of the voltage across it).

$$I_{ds} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left(\underbrace{V_{gs} - V_t}_{V_{ov}} \right)^2 \tag{3}$$

1.3 Small-Signal Conductance/Resistance

Small-signal conductance/resistance indicates how the drain current I_{ds} changes for small variations in gate or drain voltage. This is given by a derivative of the form $\partial I_{ds} / \partial x$, where x is either V_{gs} or V_{ds}.

1.3.1 The Derivative With Respect to V_{ds}

When measuring the change in the drain current with respect to V_{ds}, the derivative, called the channel or output resistance r_{ds}, is given by:

$$r_{ds} = \left(\frac{\partial I_{ds}}{\partial V_{ds}} \right)^{-1} \tag{4}$$

Triode Region

$$I_{ds} = \mu_n C_{ox} \frac{W}{L} \left[(V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right],$$

$$r_{ds} = \frac{1}{\mu_n C_{ox} \frac{W}{L} [\, (V_{gs} - V_t) \; - \; V_{ds}]}. \tag{5}$$

Saturation Region

- In an ideal MOSFET, I_{ds} is completely independent of V_{ds} in the saturation region. As a result:

$$r_{ds} = \infty \tag{6}$$

- In a real MOSFET, I_{ds} continues to be dependent on V_{ds}, even in saturation.

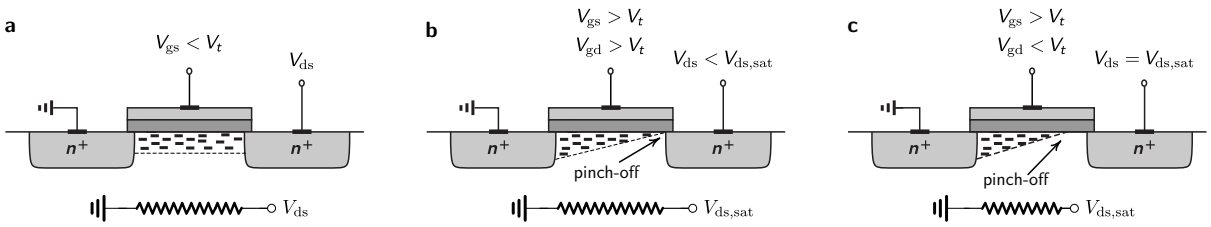

Figure 2. Channel length modulation.

- At the onset of saturation, when $V_{ds} = V_{ds,sat}$, the channel becomes pinched off near the drain, forming a triangular shape channel.

- As V_{ds} increases beyond $V_{ds,sat}$, the depletion region near the drain expands, causing the pinch-off point to move further away from the drain, reducing the effective channel length.

- As the voltage at pinch-off is by definition $V_{ds,sat}$, a shorter channel (due to a larger V_{ds}) implies a smaller channel resistance, resulting in a larger drain-source current. To account for this, the current in the saturation region is multiplied by a factor of $(1+\lambda V_{ds})$, where $0<\lambda<1$ is the channel length modulation parameter and is inversely proportional to the channel length.

$$I_{ds} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs}-V_t)^2(1+\lambda V_{ds}) \tag{7}$$

- Expanding:

$$I_{ds} = \underbrace{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs}-V_t)^2}_{f_1(V_{gs})} + \underbrace{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs}-V_t)^2\lambda V_{ds}}_{f_2(V_{gs},V_{ds})} \tag{8}$$

- Taking the derivative with respect to V_{ds}:

$$r_{ds} = \frac{1}{g_{ds}} = \frac{1}{f_1(V_{gs})\cdot\lambda} = \frac{1/\lambda}{f_1(V_{gs})} \tag{9}$$

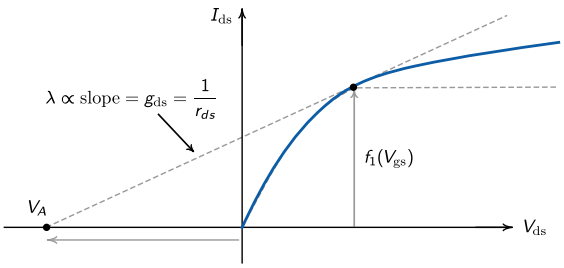

Figure 3. λ is proportional to the slope and inversely proportional to the Early voltage.

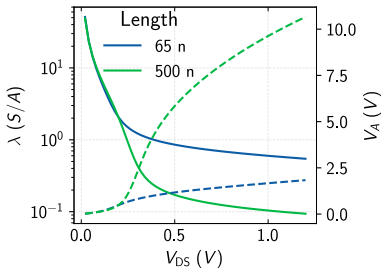

Figure 4. λ (V_A) gets smaller (larger) as the transistor is biased deeper into saturation.

- The slope intersects the x-axis at the point $-V_A$ called the Early voltage, which is approximately equal to $1/\lambda$. The current equation can be rewritten as:

$$I_{ds} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs}-V_t)^2\left(1+\frac{V_{ds}}{V_A}\right) \tag{10}$$

Significance of r_{ds}

From the above, we have the following:

$$r_{ds} \propto \frac{1}{\lambda} \propto V_A \tag{11}$$

A larger V_A means a smaller V_{ds}/V_A, implying a smaller variation in the drain current with respect to V_{ds}, or equivalently a large channel resistance. This is important in current mirrors and active loads, where a more constant current over voltage variations is desired.

Note that the absolute value of r_{ds} is not useful because it depends on the bias current. From $r_{ds} = V_A/I_{ds}$, for a given V_A, one device biased at a much smaller I_{ds} will have a much larger r_{ds} compared to another device biased at a much larger I_{ds}, even though both devices will have the same variation in I_{ds} with respect to V_{ds} (since they have the same V_{ds}/V_A).

The plots below are for a device biased in strong inversion.

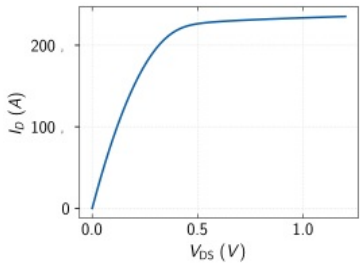

Figure 5. In strong inversion, $V_{ds,sat} \approx V_{ov}$.

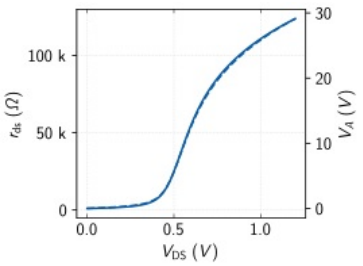

Figure 6. The variation of r_{ds} and V_A with respect to V_{ds}.

The plots below are for a device biased in weak inversion.

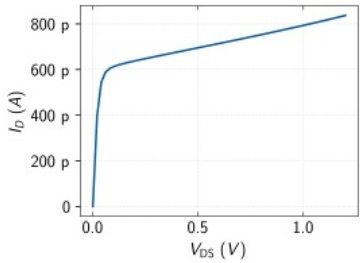

Figure 7. In weak inversion, $V_{ds,sat} \approx 3V_t \approx 80$ mV.

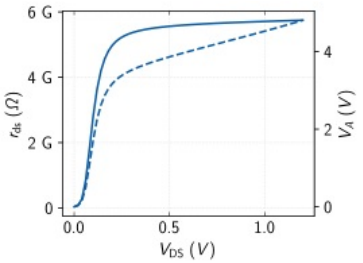

Figure 8. The variation of r_{ds} and V_A with respect to V_{ds}.

Even though the device in weak inversion has a much larger r_{ds}, its drain current shows a larger variation with V_{ds} when compared to the device in strong inversion. Hence, it is more useful to use V_A rather than r_{ds} as a design parameter (V_A is r_{ds} normalized by the current).

1.3.2 The Derivative With Respect to V_{gs}

When measuring the change in the drain current with respect to V_{gs}, the derivative, called the transconductance g_m, is given by:

$$g_m = \frac{\partial I_{ds}}{\partial V_{gs}} \tag{12}$$

Triode Region

$$I_{\text{ds}} = \mu_n C_{\text{ox}} \frac{W}{L}\left[(V_{\text{gs}} - V_t)V_{\text{ds}} - \frac{1}{2}V_{\text{ds}}^2\right],$$

$$g_m = \mu_n C_{\text{ox}} \frac{W}{L} V_{\text{ds}}. \tag{13}$$

Saturation Region

$$I_{\text{ds}} = \frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}(V_{\text{gs}} - V_t)^2(1 + \lambda V_{\text{ds}}),$$

$$g_m = \mu_n C_{\text{ox}} \frac{W}{L}(V_{\text{gs}} - V_t)(1 + \lambda V_{\text{ds}}). \tag{14}$$

Ignoring channel-length modulation, g_m can be rewritten in terms of I_{ds} as follows:

$$\begin{aligned} g_m &= \sqrt{2\mu_n C_{\text{ox}} \frac{W}{L} I_{\text{ds}}} \\ &= \frac{2I_{\text{ds}}}{V_{\text{ov}}} \end{aligned} \tag{15}$$

In the strong inversion:

- g_m is proportional to $\sqrt{I_{\text{ds}}}$ at a fixed W/L.
- g_m is proportional to $\sqrt{I_{\text{ds}}}$ at a fixed V_{ov}.
- g_m is proportional to $\sqrt{W/L}$ at a fixed I_{ds}.
- g_m is inversely proportional to V_{ov} at a fixed I_{ds}.

Significance of g_m

A larger g_m translates to a higher gain in amplifier circuits, because a small gate-voltage variation leads to a larger drain-current variation.

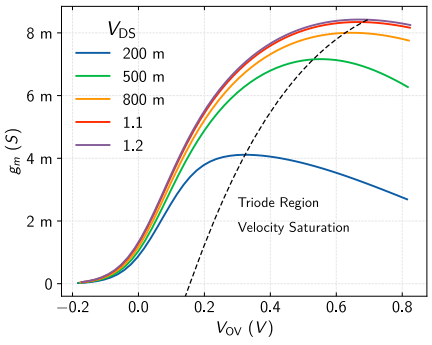

Figure 9. Variation of g_m with V_{ov} for for different V_{ds}.

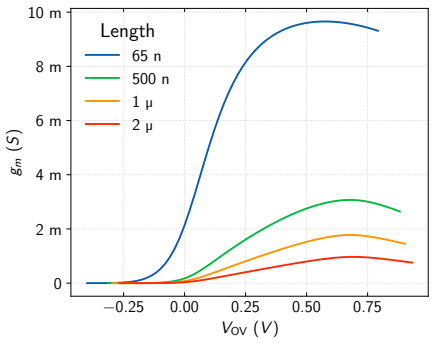

Figure 10. Variation of g_m with V_{ov} for different lengths.

- **Small V_{ov}**: When V_{ov} is near zero, the channel is just forming and the current is still small, so $g_m = \partial I_{\text{ds}}/\partial V_{\text{gs}}$ is also small.
- **Moderate V_{ov}**: As V_{ov} increases further into strong inversion, g_m grows strongly with $V_{\text{ov}}(1 + \lambda V_{\text{ds}})$.

- **Large V_{ov}**: At higher V_{ov}, two things can happen:
  - The device enters into the triode region and g_m stops increasing with V_{ov}.
  - The device becomes velocity saturated. In this regime, I_{ds} depends on V_{ov} to the first power. As a result, g_m becomes a constant.
- **Very Large V_{ov}**: Mobility degredation occurs at very large V_{ov}, causing the current, and therefore g_m, to fall.

2 Limitations of the Square-Law Model

In the square-law model, V_{gs} (or V_{ov}) serves as the primary controlling variable since it directly affects I_{ds}. However, this equation is only valid in certain cases.

Recall the equation of a mosfet in saturation according to the square-law model:

$$I_{ds} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}V_{ov}^2 \tag{16}$$

Computing the transconductance:

$$g_m = \frac{\partial I_{ds}}{\partial V_{gs}} = \mu_n C_{ox}\frac{W}{L}V_{ov} \tag{17}$$

Dividing g_m by I_{ds}:

$$\frac{g_m}{I_{ds}} = \frac{1}{V_{ov}} \tag{18}$$

A plot of the g_m/I_{ds} predicted by the square-law model vs. V_{ov} for different lengths is shown below.

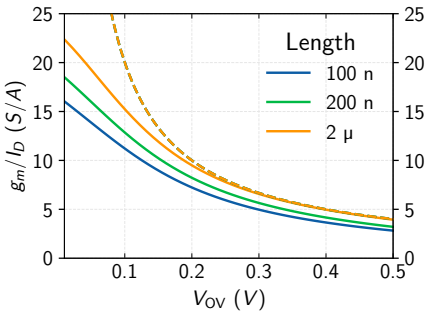

Figure 11. Comparing the g_m/I_{ds} predicted by the square-law model (dashed) with the actual g_m/I_d of the device (solid).

We observe two things:

- In **strong inversion** ($V_{ov} > 0.2$), the square-law model is only applicable for long channel devices. This is because the square-law model does not account for short-channel effects such as velocity saturation.
- In **moderate/weak inversion** ($V_{ov} < 0.2$), the square-law model breaks down altogether. This is because the square-law model assumes that the current is zero when $V_{ov} = 0$. However, in reality, current continues to flow by diffusion (like in a BJT) and the square-law model must be replaced with an exponential I–V relationship:

$$I_{ds} = I_{off}\, e^{\frac{V_{gs}}{n \cdot V_{th}}} \tag{19}$$

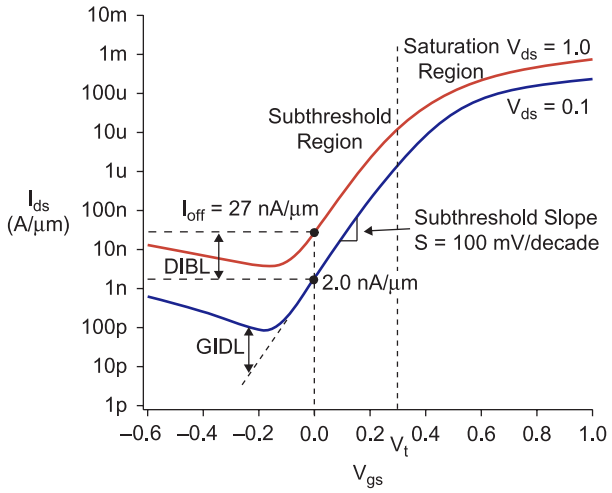

Figure 12. Plot of $\log_{10}(I_{ds})$ vs. V_{gs}. In the subthreshold regime the drain current falls exponentially but never reaches zero. For short-channel MOSFETs the depletion voltage V_{dep} also depends on the drain voltage V_{ds} through drain-induced barrier lowering (DIBL). Because DIBL raises V_{dep}, the entire curve shifts upward, increasing leakage at $V_{gs}=0$. When the gate near 0 V (or slightly negative) and the drain at a large positive bias, the large drain-to-gate voltage (V_{dg}) creates a strong electric field at the drain-channel junction. This field drives band-to-band tunneling, generating electron-hole pairs; electrons enter the drain, boosting the off-state current known as gate-induced drain leakage (GIDL).

3 Introduction to g_m/I_D Methodology

The g_m/I_D methodology is a methodology for designing analog circuits where the transistor's bias point is critical, such as in amplifiers and current mirrors. It differs from the V_{ov}-based methodology in the following ways:

1. Its controlling variables are the device length L and g_m/I_D (not V_{ov}).
2. It uses lookup tables (not equations) to provide accurate relationships between g_m, I_D, and other parameters.
3. It accounts for strong and moderate inversion.

3.1 Physical Meaning of g_m/I_D

g_m and I_D are two important parameters in the design of many analog circuits (e.g., amplifiers).

- g_m is directly connected to several important specifications such as the GBW and noise. The larger the value of g_m, the better.
- In analog ICs, we usually set the bias current (I_{ds}) rather than the bias voltage (V_{gs}).
- Thus, g_m/I_D tells us how much g_m we get out of a MOSFET when we invest a given I_D. Therefore, g_m/I_D serves as an indicator of how efficiently the transistor converts current into transconductance under the set bias conditions.

Note: A high $\mathrm{g_m/I_D}$ value does not mean that g_m itself is large. It simply means that a transistor biased at a high $\mathrm{g_m/I_D}$ has a larger g_m compared to a transistor biased at a lower $\mathrm{g_m/I_D}$ value given the same biasing current I_D.

3.2 Importance of g_m

3.2.1 Gain Banddwidth Product (GBW)

The GBW of a single-pole amplifier is the DC gain A_0 multiplied by the bandwidth $f_{3\mathrm{db}}$. If we assume the dominant pole is located at the output node, where the load capacitance C_L is connected, then

$$\mathrm{GBW} = A_0 \cdot f_{3\mathrm{db}} = [g_m \cdot R_{\mathrm{out}}] \cdot \left[\frac{1}{2\pi \cdot R_{\mathrm{out}} \cdot C_L}\right] = \frac{g_m}{2\pi \cdot C_L}. \tag{20}$$

The larger g_m is made, the larger is the GBW for the same load capacitance.

3.2.2 Noise

Consider the common-source amplifier shown below.

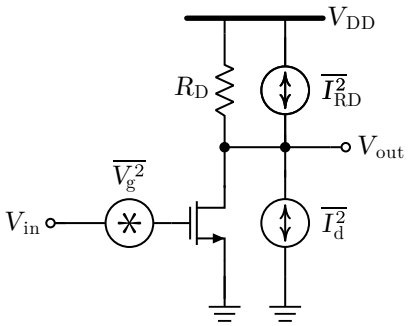

Figure 13. Noise sources in a common-source amplifier.

The input referred noise voltage is given by:

$$\overline{V^2_{n,\mathrm{in}}} = \frac{\overline{V^2_{n,\mathrm{out}}}}{A_V^2} = \frac{\overline{I^2_{n,\mathrm{out}}} \cdot R_D^2}{A_V^2} \tag{21}$$

where

$$\overline{I^2_{n,\mathrm{out}}} = [\overline{I^2_{R_D}}] + [\overline{I^2_d} + \overline{I^2_g}] \tag{22}$$

$$\overline{I^2_{n,\mathrm{out}}} = \left[\frac{4\,k\,T}{R_D}\right] + \left[4\,k\,T\,\gamma\,g_m + \frac{K}{W\,L\,C_{\mathrm{ox}}}\frac{1}{f} \cdot g_m^2\right] \tag{23}$$

and $A_V = -g_m \cdot R_D$.

Therefore:

$$\overline{V^2_{n,\mathrm{in}}} = 4\,k\,T \underbrace{\left(\frac{\gamma}{g_m} + \frac{1}{g_m^2\,R_D}\right)}_{\text{thermal noise}} + \underbrace{\left(\frac{K}{W\,L\,C_{\mathrm{ox}}}\frac{1}{f}\right)}_{\text{flicker noise}} \tag{24}$$

The larger g_m is made, the smaller is the input referred noise voltage.

3.3 g_m/I_D in BJTs

Recall that the collector current in BJTs is given by:

$$I_c = I_S \cdot e^{\left(\frac{V_{be}}{V_{th}}\right)} \tag{25}$$

where V_{th} is the thermal voltage.

From this, the transconductance is derived as:

$$g_m = \frac{\partial I_c}{\partial V_{be}} = \frac{I_c}{V_{th}} \tag{26}$$

Consequently:

$$\frac{g_m}{I_c} = \frac{1}{V_{th}} \approx \frac{1}{26\,\text{mV}} = 38.5\ \text{S/A}, \tag{27}$$

indicating that the transconductor efficiency in BJTs is independent of the bias point.

The constant g_m/I_D in BJTs arises from the exponential nature of I_c, which in turn dictates that g_m will be of the same nature. Any increase in I_c in a BJT results in the same proportional increase in g_m (scaled by $1/V_{th}$) regardless of the bias point. On the contrary, MOSFETs exhibit a constant g_m/I_D only in the subthreshold region, where they have exponential characteristics.

4 Trends

4.1 g_m/I_D and V_{OV}

In strong inversion, the square-law model predicts that $g_m = 2I_{ds}/V_{ov}$. Thus, $g_m/I_{ds} = 2/V_{ov}$ gives a rough idea of the relationship between g_m/I_D and V_{ov} for $V_{ov} > 0$. The actual relationship is more complex, as can be seen in the plots below.

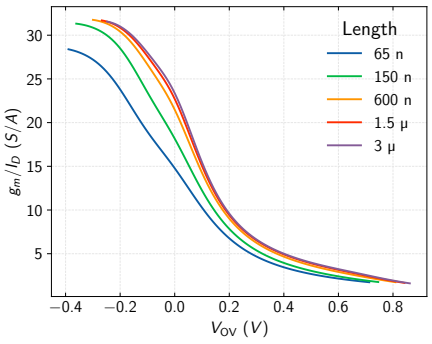

Figure 14. g_m/I_D vs. V_{ov}.

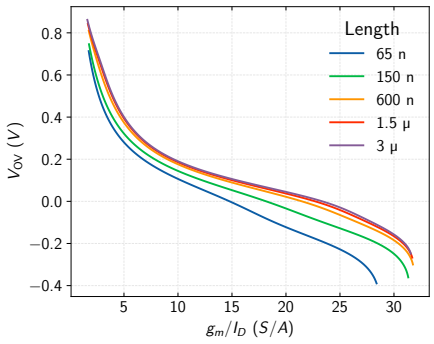

Figure 15. V_{ov} vs g_m/I_D.

g_m/I_D values typically range between 3 and 30 S/A (less than 38.5 S/A in BJTs), irrespective of the technology.

- Higher g_m/I_D values represent weak inversion.

- Midrange values, typically around $12-20\,\mathrm{S/A}$, represent moderate inversion.
- Lower values represent strong inversion.

4.2 g_m/I_D and Transistor Width

While high g_m/I_D values represent high transconductor efficiency, they correspond to low current densities (I_D/W).

Recall that the current in weak inversion (where g_m/I_D values are the highest) has exponential characteristics. This implies the need for exponentially increasing transistor widths to achieve a specific drain current. This comes at the cost of exponentially increasing area parasitic capacitances, thus lowering (or limiting) the speed of the device.

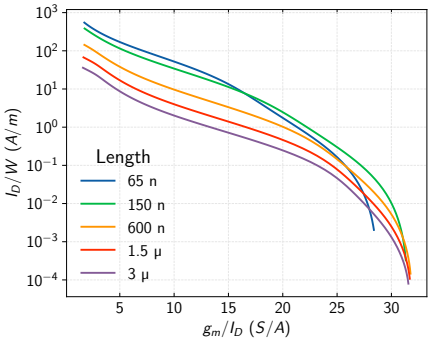

Figure 16. The current density decreases with increasing g_m/I_D.

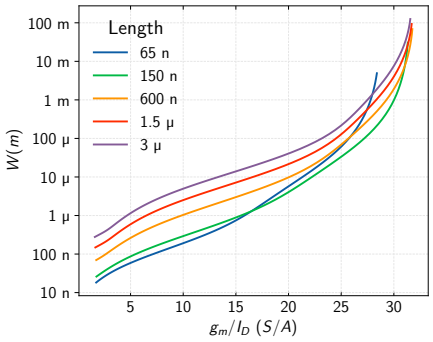

Figure 17. Biasing at high g_m/I_D requires exponentially increasing widths to achieve a specific drain current.

4.3 g_m/I_D and r_{ds}

Recall that the small-signal channel resistance is proportional to the Early voltage:

$$r_{ds} = \frac{1}{\lambda \cdot I_{ds}} \approx \frac{V_A}{I_{ds}} \tag{28}$$

- Early voltage reflects how the drain current changes with drain-source voltage.
- Higher g_m/I_D values correspond to small overdrive voltages. This means that the transistor is biased closer to the subthreshold region where the transistor's current is more sensitive to changes in the output voltage, implying lower small-signal resistance and, by extension, smaller Early voltage.

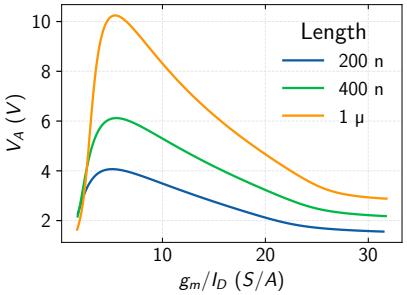

Figure 18. Early voltage decreases with g_m/I_D for a given V_{ds}

4.4 g_m/I_D and $V_{DS_{SAT}}$

$V_{DS_{SAT}}$ is the minimum drain-source voltage required for saturation.

- In the V_{ov} methodology, a transistor is assumed to be in saturation when $V_{ds} = V_{ov}$. However, we see from the plot below that this is only true for the case when $V_{ov} > 0.2$ (strong inversion).
- In the g_m/I_D methodology, the equivalent term is $V^\star = \frac{2}{g_m/I_D}$. It is always greater than $V_{DS_{SAT}}$ and therefore $V_{ds} \approx V^\star$ gurantees that the transistor is in saturation in all regions of operation.

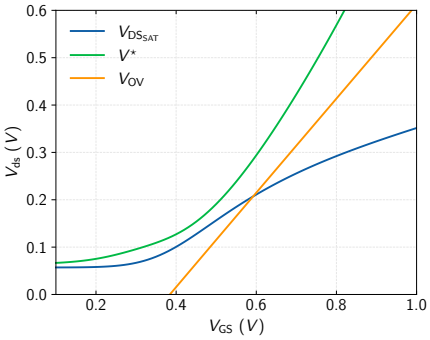

Figure 19. $V_{ds} \approx V^\star$ gurantees that the transistor is in saturation in all regions of opration.

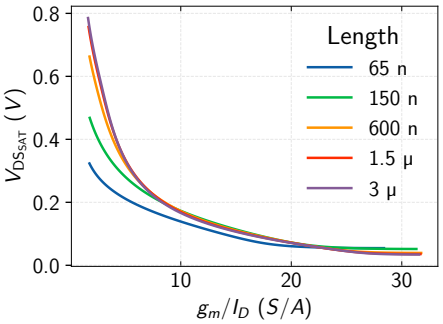

Figure 20. $V_{DS_{SAT}}$ decreases with g_m/I_D since higher g_m/I_D values correspond to smaller V_{ov}.

4.5 g_m/I_D and Intrinsic Gain

The intrinsic gain of a MOSFET is given by $g_m \cdot r_{ds}$, which can be rewritten as:

$$g_m \cdot r_{ds} = \frac{g_m}{I_{ds}} \cdot \frac{I_{ds}}{g_{ds}} = \left(\frac{g_m}{I_{ds}}\right) \cdot V_A \tag{29}$$

- Early voltage, and consquently the output resistance, generally decreases with increasing g_m/I_D.
- The large Early voltage at low g_m/I_D values causes the intrinsic gain to increase rapidly before it starts saturating at higher g_m/I_D values.

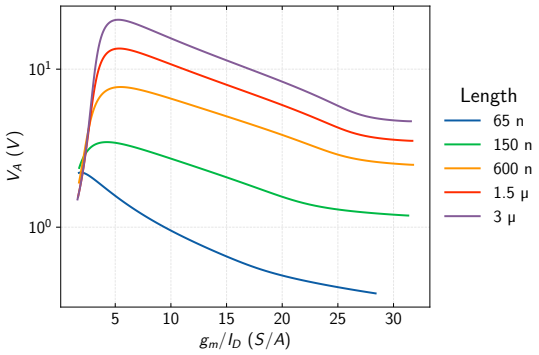

Figure 21. Early voltage vs. g_m/I_D.

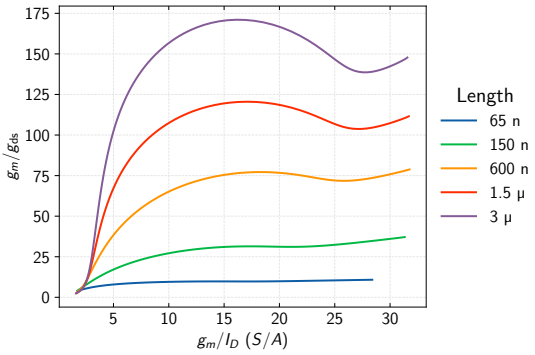

Figure 22. Intrinsic gain vs. g_m/I_D.

4.6 g_m/I_D and Transistor Speed

The transit frequency, f_T, is the frequency at which the current gain of the transistor becomes unity. It is a critical parameter that indicates the frequency limit for a MOSFET to act as an effective amplifier. It is determined by the gate length (L), gate capacitance (C_{gs}), carrier mobility (μ), and V_{OV}.

$$f_T = \frac{g_m}{2\pi C_{gs}} = \frac{2\mu}{2\pi \cdot C_{gs} \cdot 3L^2} V_{OV} \tag{30}$$

- The transit frequency is proportional to V_{OV}, whereas g_m/I_D is proportional to $\frac{1}{V_{OV}}$. Hence, f_T decreases with increasing g_m/I_D.
- The trade-off between g_m/I_D and f_T can be roughly translated to a tradeoff between gain and bandwidth.

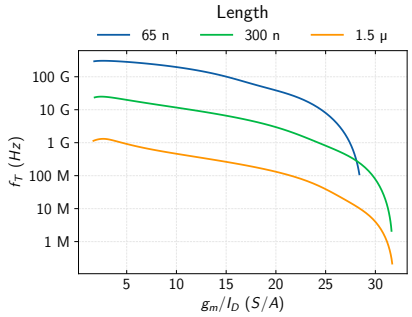

Figure 23. Transist frequency decreases with g_m/I_D.

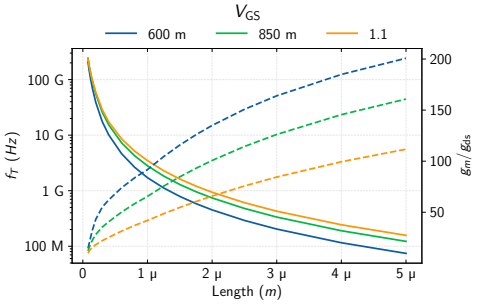

Figure 24. Transist frequency decreases with length, whereas the instrinsic gain increases with length.

5 Design Methodology

5.1 Guidelines

- **Gain**: Recall that the DC gain is given by $g_m \cdot r_{ds}$. Given a fixed current, use a large g_m/I_D for transistors whose g_m contribute to the gain to get a larger g_m.
- **Output Resistance**: Recall that the output resistance is given by $r_{ds} \approx V_A / I_D$. Since the Early voltage decreases with increasing g_m/I_D, use a small g_m/I_D for transistors that need a large output resistance.
- **Noise**: Recall that the input referred noise voltage for the common-source amplifier is given by:

$$\overline{V_{n,in}^2} = \underbrace{4kT\left(\frac{\gamma}{g_m} + \frac{1}{g_m^2 R_D}\right)}_{\text{thermal noise}} + \underbrace{\left(\frac{K}{WLC_{ox}}\frac{1}{f}\right)}_{\text{flicker noise}}$$

  - The g_m in this equation is due to the scaling of the output noise voltage by $1/A_V^2$. Since A_V is proportional to g_m, use a large g_m/I_D for transistors whose g_m contribute to the gain.
  - The thermal noise is also inversely propotional to R_D. If R_D is implemented via a MOSFET, use a small g_m/I_D for a large output resistance.

- The flicker noise can be reduced by increasing the device size WL. A large $\mathrm{g_m/I_D}$ implies a larger W and therefore low flicker noise. Further reduction in the flicker noise can be done by using longer devices.

- **Voltage Swing**: Recall that $V_{\mathrm{DS_{SAT}}}$ decreases with increasing $\mathrm{g_m/I_D}$. Thus, use a large $\mathrm{g_m/I_D}$ for transistors that contribute to the determination of the voltage swing. Less voltage will be required to keep these transistors in saturation, thereby maximizing the voltage swing.

- **Matching**: A large $\mathrm{g_m/I_D}$ implies a larger W. A large device area translates to less random mismatch.

5.2 Transistor Sizing

When g_m can be easily deduced from design specificiations, the sizing methodology is as follows:

- Determine g_m (from design specifications).
- Pick L:
  - short channel → high speed, small area.
  - long channel → high intrinsic gain, improved matching, ...
- Pick $\mathrm{g_m/I_D}$:
  - large $\mathrm{g_m/I_D}$ → low power, large signal swing (low $V_{\mathrm{ds,sat}}$).
  - small $\mathrm{g_m/I_D}$ → high speed, small area.
- Determine I_D (from g_m and $\mathrm{g_m/I_D}$).
- Determine W (from I_D and $\mathrm{I_D/W}$).