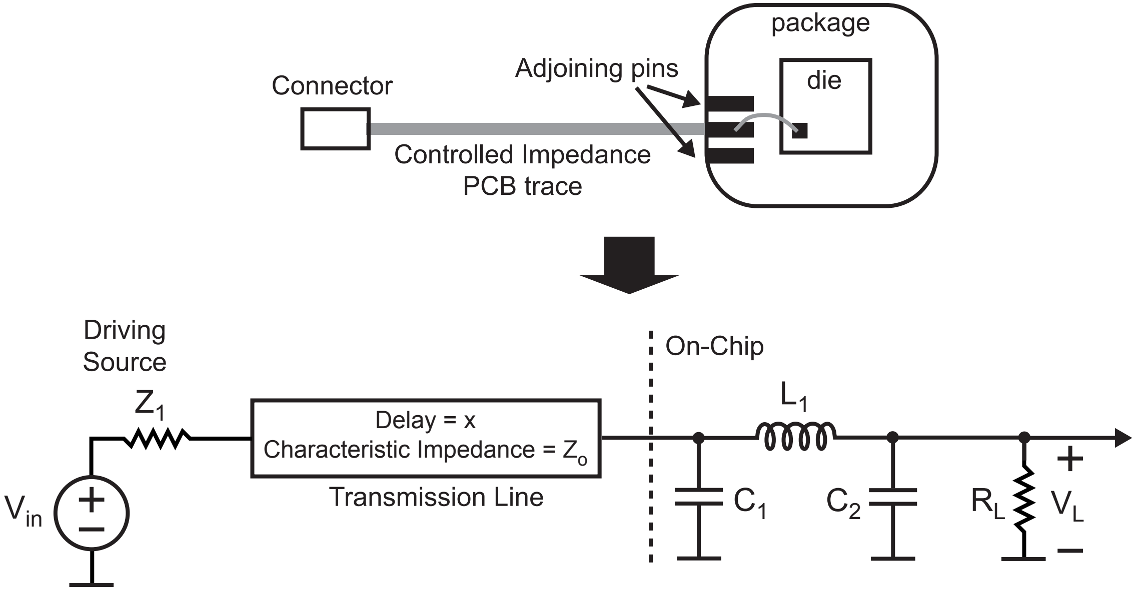
6.976
High Speed Communication Circuits and Systems
Lecture 5
High Speed, Broadband Amplifiers

Michael Perrott
Massachusetts Institute of Technology

Broadband Communication System

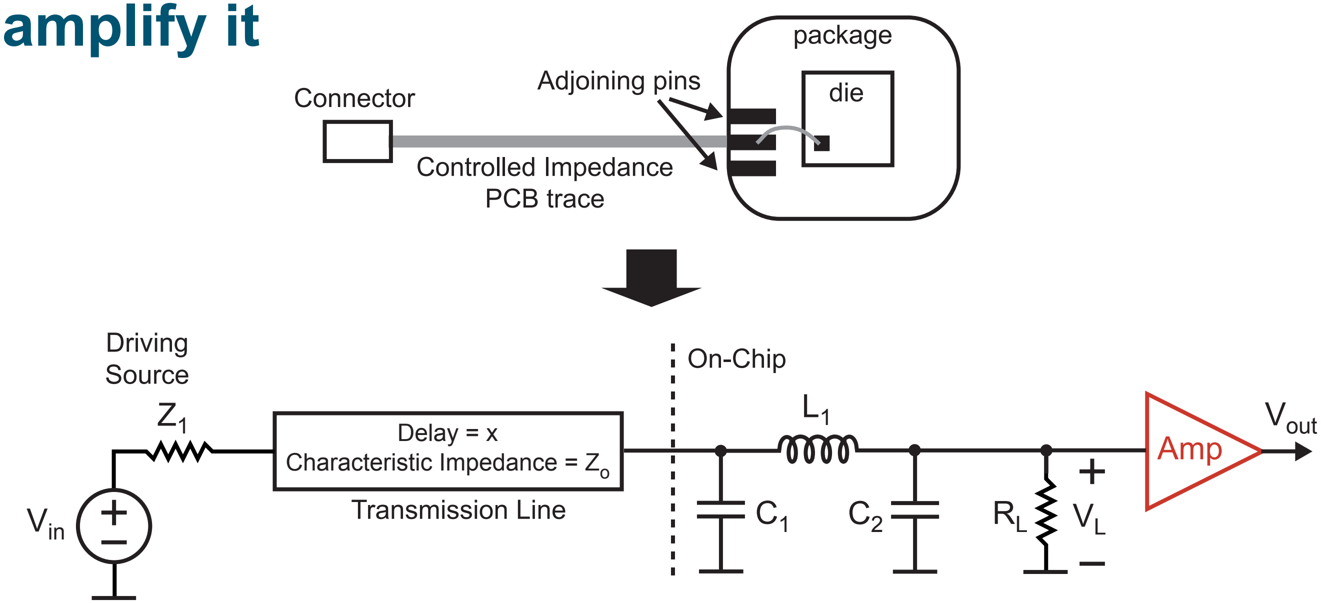
- **Example: high speed data link on a PC board**



  - **We’ve now studied how to analyze the transmission line effects and package parasitics**
  - **What’s next?**

High Speed, Broadband Amplifiers

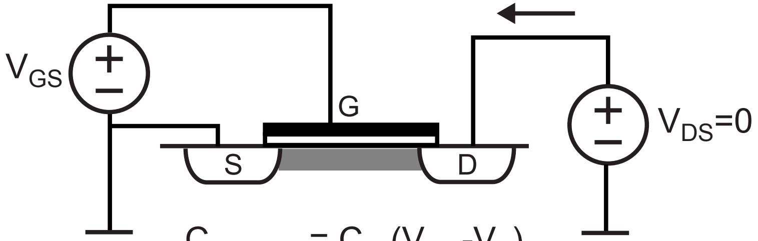
- The first thing that you typically do to the input signal is amplify it



- Function
  - Boosts signal levels to acceptable values
  - Provides reverse isolation
- Key performance parameters
  - Gain, bandwidth, noise, linearity

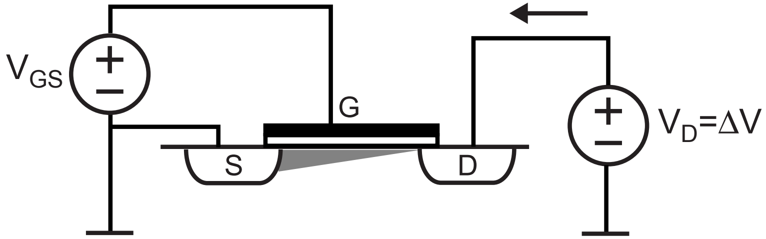
Basics of MOS Large Signal Behavior (Qualitative)

Triode

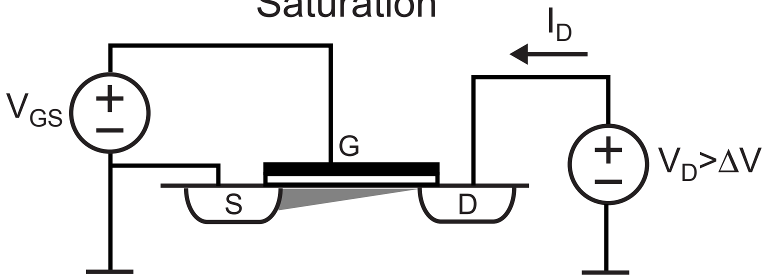


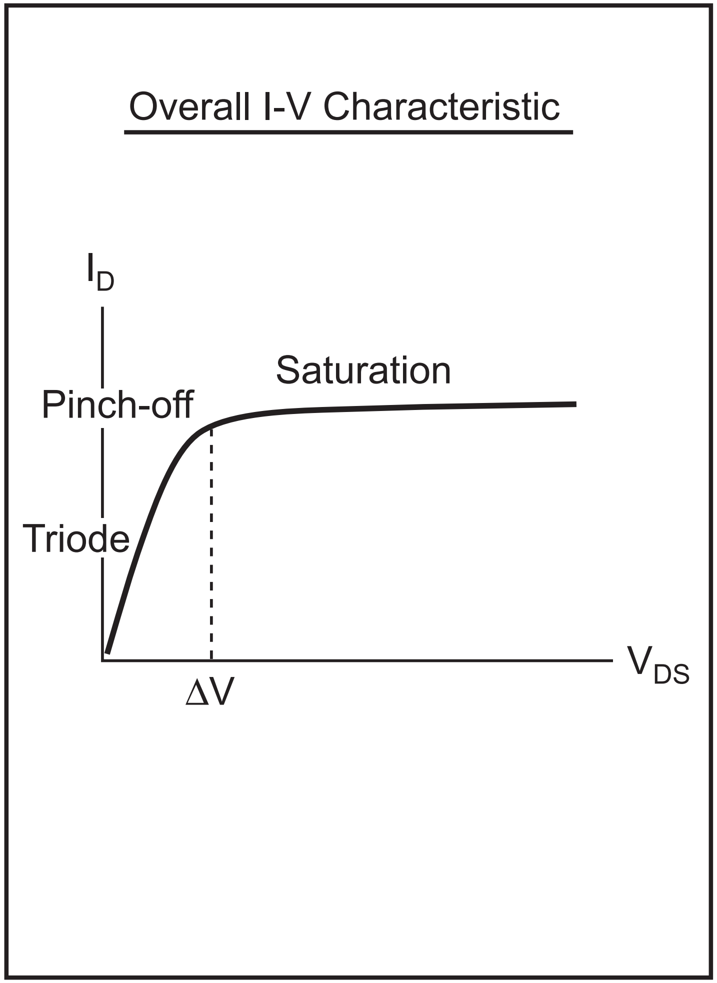
$C_{channel} = C_{ox}(V_{GS}-V_T)$

Pinch-off



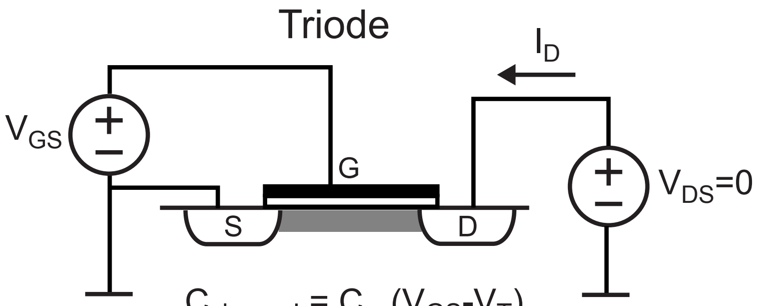
Saturation





Overall I-V Characteristic

Basics of MOS Large Signal Behavior (Quantitative)

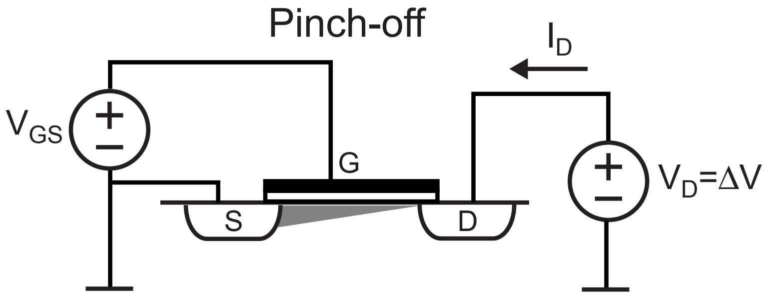


$C_{channel} = C_{ox}(V_{GS}-V_T)$

$$I_D = \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_T - V_{DS}/2)V_{DS}$$

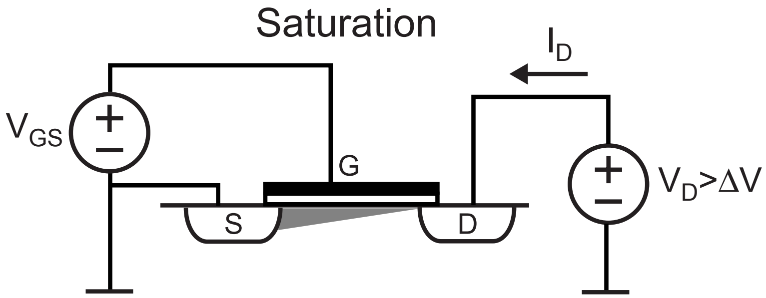
for $V_{DS} << V_{GS} - V_T$

$$I_D \approx \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_T)V_{DS}$$



$$\Delta V = V_{GS}-V_T$$

$$\Delta V = \sqrt{\frac{2I_DL}{\mu_n C_{ox}W}}$$

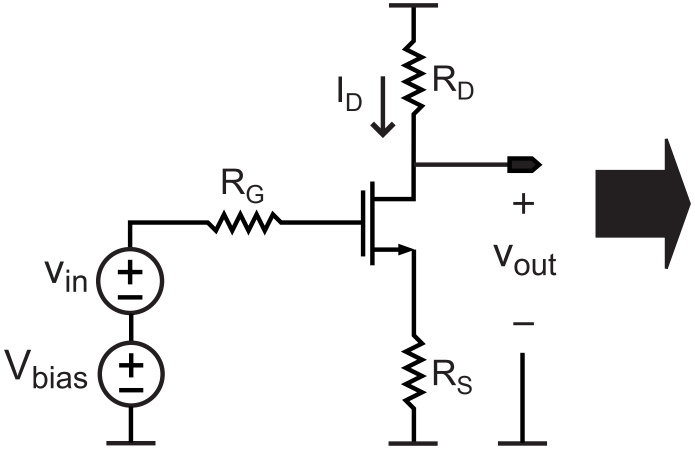


$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS}-V_T)^2(1+\lambda V_{DS})$$

(where λ corresponds to channel length modulation)

Analysis of Amplifier Behavior

- **Typically focus on small signal behavior**
  - **Work with a linearized model such as hybrid-π**
  - **Thevenin modeling techniques allow fast and efficient analysis**
- **To do small signal analysis:**

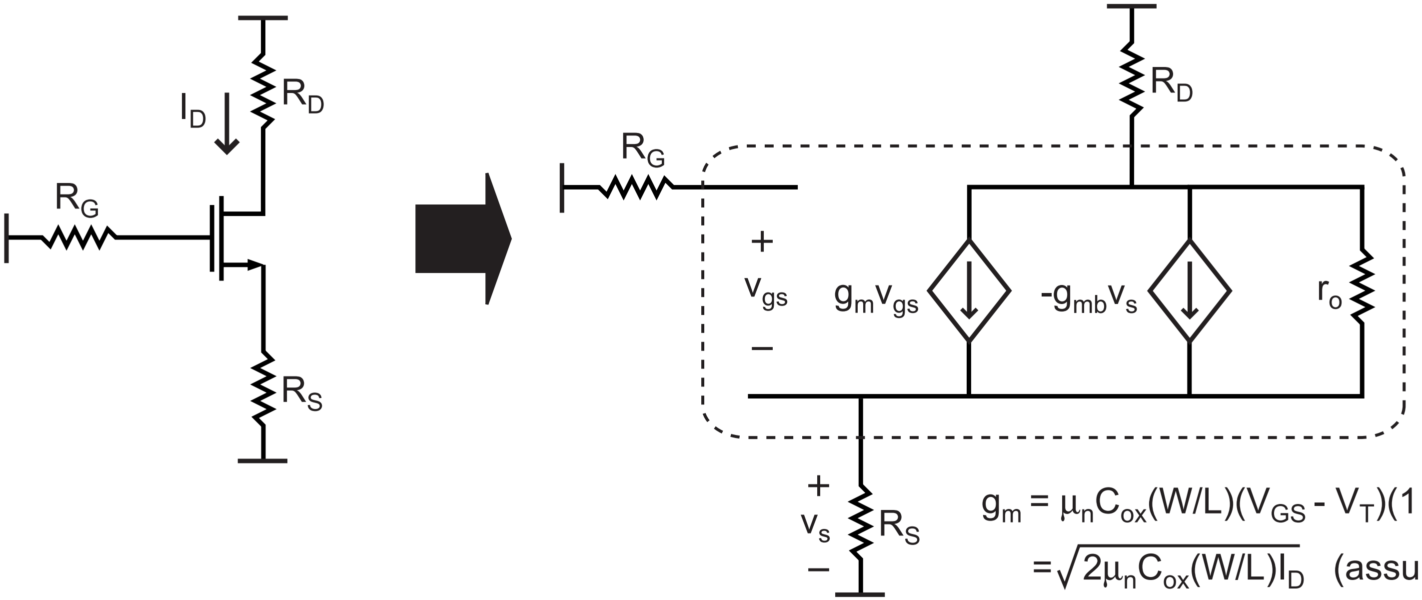


Small Signal Analysis Steps

1) Solve for bias current I_d
2) Calculate small signal parameters (such as g_m, r_o)
3) Solve for small signal response using transistor hybrid-π small signal model

MOS DC Small Signal Model

- **Assume transistor in saturation:**



$$g_m = \mu_n C_{ox}(W/L)(V_{GS} - V_T)(1 + \lambda V_{DS})$$

$$= \sqrt{2\mu_n C_{ox}(W/L) I_D} \quad \text{(assuming } \lambda V_{DS} << 1)$$

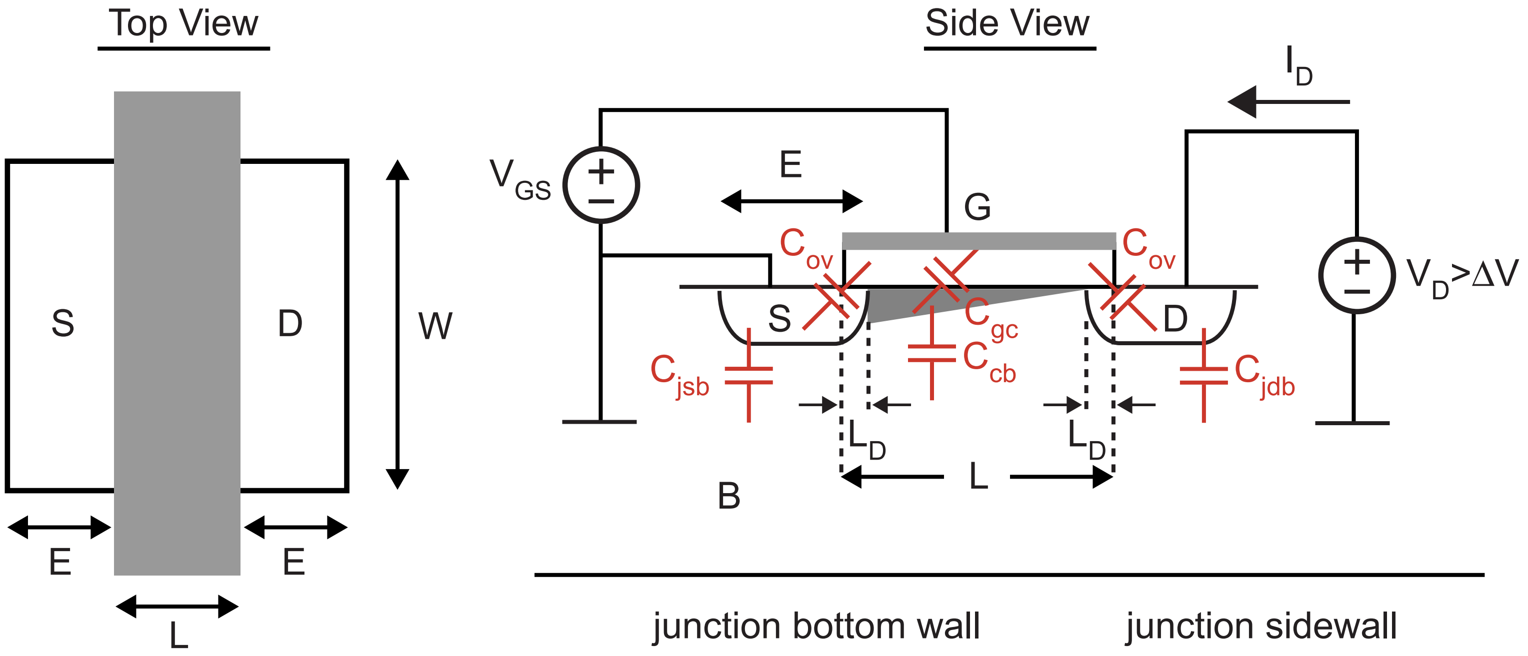
$$g_{mb} = \frac{\gamma g_m}{2\sqrt{2|\Phi_p| + V_{SB}}} \quad \text{where } \gamma = \frac{\sqrt{2q\varepsilon_s N_A}}{C_{ox}}$$

In practice: $g_{mb} = g_m/5$ to $g_m/3$

$$r_o = \frac{1}{\lambda I_D}$$

- **Thevenin modeling based on the above**

Capacitors For MOS Device In Saturation



junction bottom wall cap (per area)

junction sidewall cap (per length)

source to bulk cap: $C_{jsb} = \dfrac{C_j(0)}{\sqrt{1 + V_{SB}/|\Phi_B|}} WE + \dfrac{C_{jsw}(0)}{\sqrt{1 + V_{SB}/|\Phi_B|}} (W + 2E)$

(make 2W for "4 sided" perimeter in some cases)

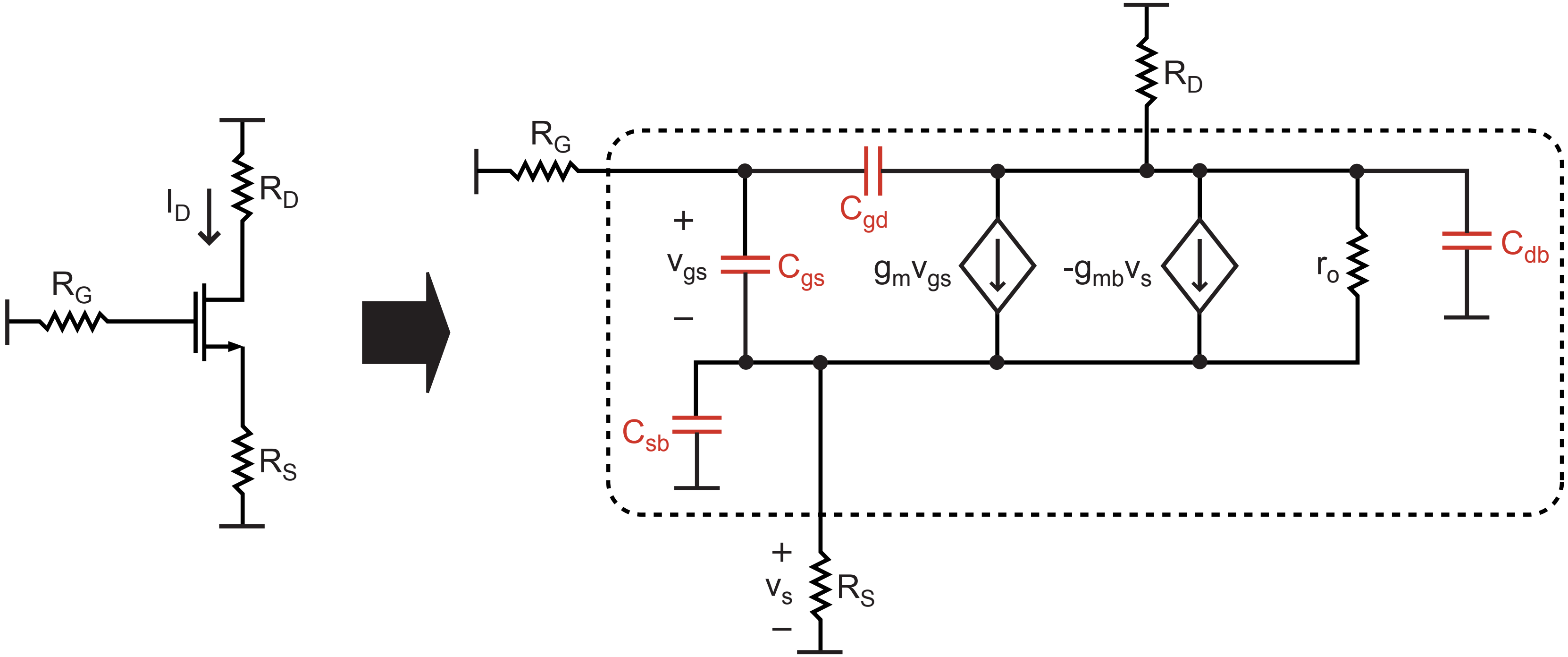
drain to bulk cap: $C_{jsd} = \dfrac{C_j(0)}{\sqrt{1 + V_{DB}/|\Phi_B|}} WE + \dfrac{C_{jsw}(0)}{\sqrt{1 + V_{DB}/|\Phi_B|}} (W + 2E)$

overlap cap: $C_{ov} = WL_DC_{ox} + WC_{fringe}$

gate to channel cap: $C_{gc} = \dfrac{2}{3} C_{ox}W(L\text{-}2L_D)$

channel to bulk cap: C_{cb} - ignore in this class

MOS AC Small Signal Model (Device in Saturation)



$$C_{gs} = C_{gc} + C_{ov} = \frac{2}{3} C_{ox} W(L - 2L_D) + C_{ov}$$

$$C_{gd} = C_{ov}$$

$C_{sb} = C_{jsb}$ (area + perimeter junction capacitance)

$C_{db} = C_{jdb}$ (area + perimeter junction capacitance)

Wiring Parasitics

- Capacitance
  - Gate: cap from poly to substrate and metal layers
  - Drain and source: cap from metal routing path to substrate and other metal layers
- Resistance
  - Gate: poly gate has resistance (reduced by silicide)
  - Drain and source: some resistance in diffusion region, and from routing long metal lines
- Inductance
  - Gate: poly gate has negligible inductance
  - Drain and source: becoming an issue for long wires

Extract these parasitics from circuit layout

Frequency Performance of a CMOS Device

- **Two figures of merit in common use**
  - **f_t : frequency for which current gain is unity**
  - **f_{max} : frequency for which power gain is unity**

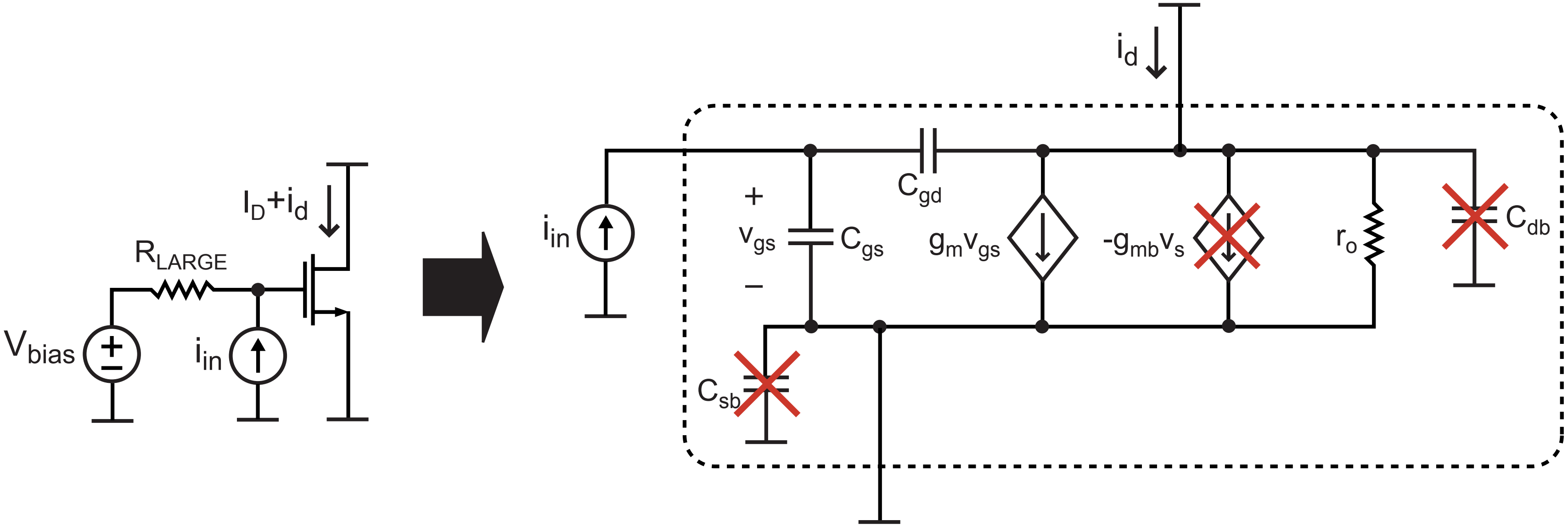
- **Common intuition about f_t**
  - **Gain, bandwidth product is conserved**

$$\Rightarrow \ \text{Gain} \cdot \text{Bandwidth} = f_t$$

  - **We will see that MOS devices have an f_t that shifts with bias**
    - This effect strongly impacts high speed amplifier topology selection

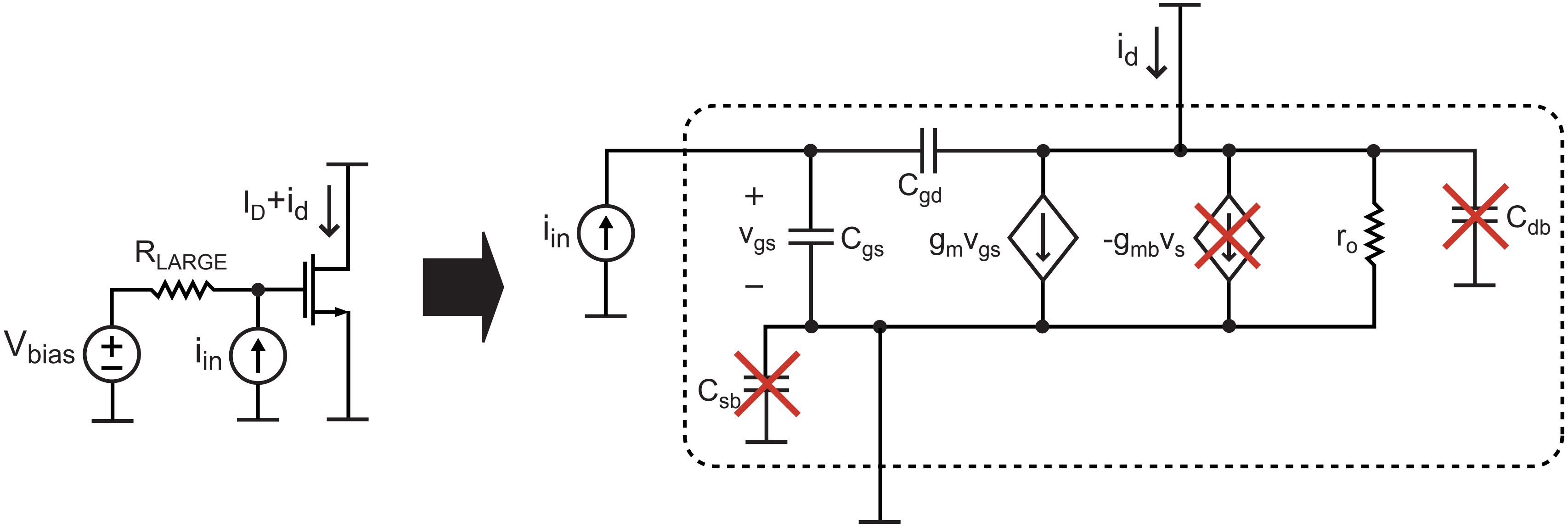
- **We will focus on f_t**
  - **Look at pages 70-72 of Tom Lee's book for discussion on f_{max}**

Derivation of f_t for MOS Device in Saturation



- **Assumption is that input is current, output of device is short circuited to a supply voltage**
  - **Note that voltage bias is required at gate**
    - The calculated value of f_t is a function of this bias voltage

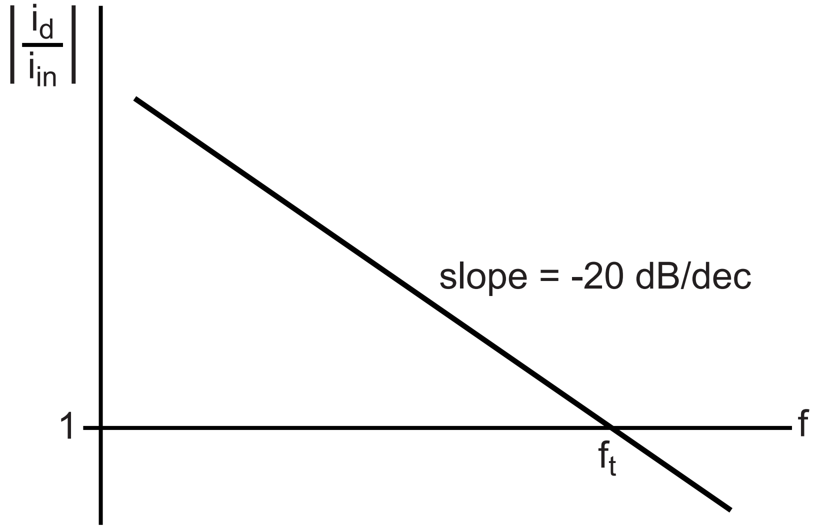
Derivation of f_t for MOS Device in Saturation



$$i_d = g_m v_{gs} = g_m \left(\frac{1}{s(C_{gs} + C_{gd})} \right) i_{in}$$

$$\Rightarrow \frac{i_d}{i_{in}} = \frac{g_m}{j2\pi f (C_{gs} + C_{gd})}$$

Derivation of f_t for MOS Device in Saturation



$$\Rightarrow \quad f_t = \frac{g_m}{2\pi(C_{gs} + C_{gd})}$$

$$i_d = g_m v_{gs} = g_m \left(\frac{1}{s(C_{gs} + C_{gd})}\right) i_{in}$$

$$\Rightarrow \quad \frac{i_d}{i_{in}} = \frac{g_m}{j2\pi f(C_{gs} + C_{gd})}$$

Why is f_t a Function of Voltage Bias?

$$f_t = \frac{g_m}{2\pi(C_{gs} + C_{gd})}$$

- f_t is a ratio of g_m to gate capacitance
  - g_m is a function of gate bias, while gate cap is not (so long as device remains biased)
- First order relationship between g_m and gate bias:

$$g_m \approx \mu_n C_{ox} \frac{W}{L}(V_{gs} - V_t)$$

  - The larger the gate bias, the higher the value for f_t
- Alternately, f_t is a function of current density

$$\frac{g_m}{C_{gs} + C_{gd}} \approx \frac{\sqrt{2\mu_n C_{ox}(W/L)I_d}}{(2/3)WLC_{ox} + W(C_{ov}/W)} \propto \sqrt{\frac{I_d}{W}}$$

  - So f_t maximized at max current density (and minimum L)

Speed of NMOS Versus PMOS Devices

$$f_t = \frac{g_m}{2\pi(C_{gs} + C_{gd})}$$

- **NMOS devices have much higher mobility than PMOS devices (in current, non-strained, bulk CMOS processes)**

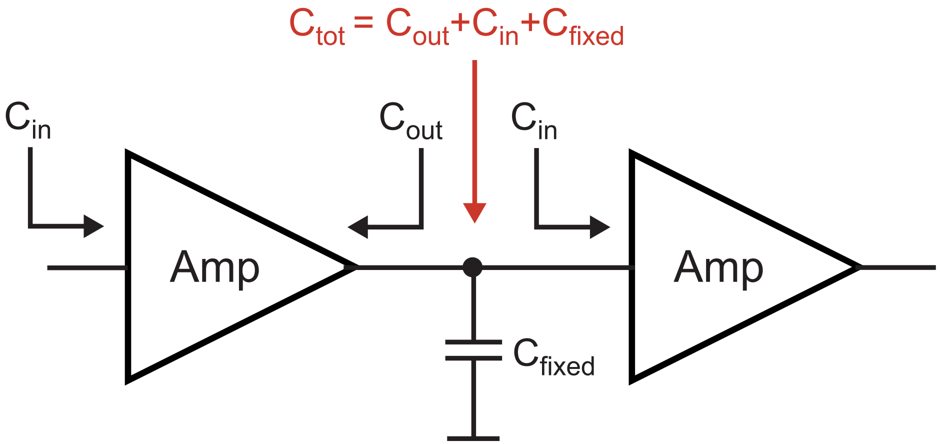
$$\mu_n \approx 2.5\mu_p \quad \text{for many processes}$$

$$\Rightarrow \; f_t \text{ of NMOS } \approx 2.5 \times \; f_t \text{ of PMOS}$$

  - **Intuition: NMOS devices provide approximately 2.5 x g_m for a given amount of capacitance and gate bias voltage**
  - **Also: NMOS devices provide approximately 2.5 x I_d for a given amount of capacitance and gate bias voltage**

Assumptions for High Speed Amplifier Analysis

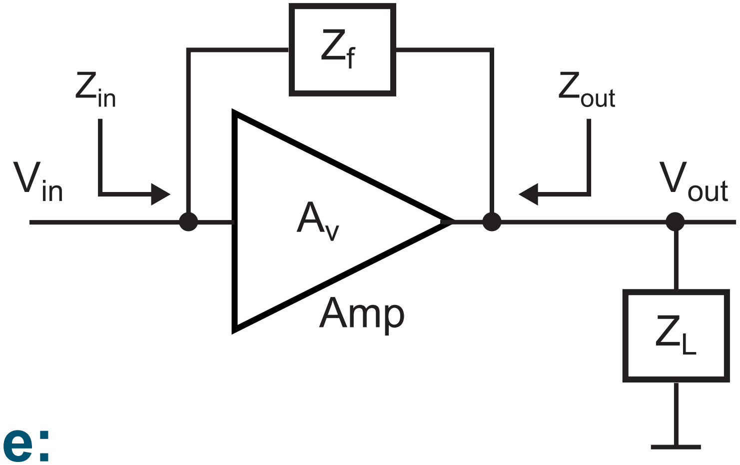
- **Assume that amplifier is loaded by an identical amplifier and by fixed wiring capacitance**



- **Intrinsic performance**
  - **Defined as the bandwidth achieved for a given gain when C_{fixed} is negligible**
  - **Amplifier approaches intrinsic performance as its device sizes (and current) are increased**
- **In practice, optimal sizing (and power) of amplifier is roughly where $C_{in}+C_{out} = C_{fixed}$**

The Miller Effect

- Concerns impedances that connect from input to output of an amplifier



- Input impedance:

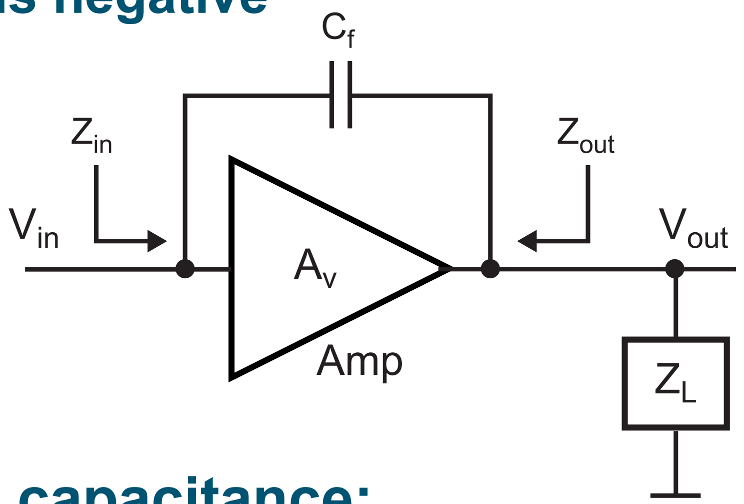
$$Z_{in} = \frac{V_{in}}{(V_{in} - V_{out})/Z_f} = \frac{Z_f}{1 - A_v}$$

- Output impedance:

$$Z_{out} = \frac{V_{out}}{(V_{out} - V_{in})/Z_f} = \frac{Z_f}{1 - 1/A_v} \approx Z_f \quad \text{for } |A_v| \gg 1$$

Example: The Impact of Capacitance in Feedback

- Consider C_{gd} in the MOS device as C_f
  - Assume gain is negative



- Impact on input capacitance:

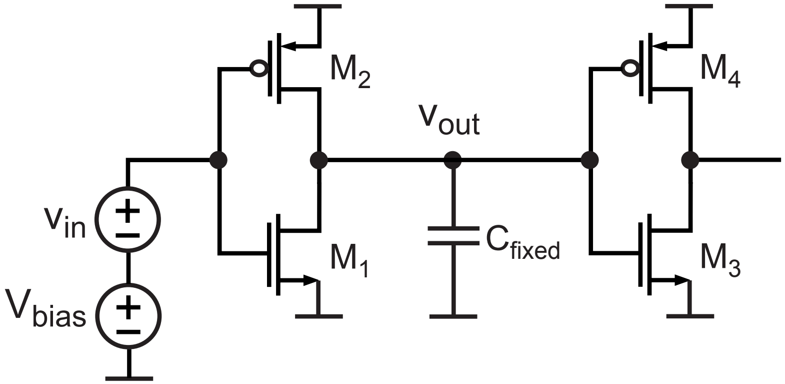
$$Z_{in} = \frac{1/(sC_f)}{1+|A_v|} = \frac{1}{sC_f(1+|A_v|)} \quad \Rightarrow \quad \text{looks like larger cap!}$$

- Output impedance:

$$Z_{out} = \frac{1/(sC_f)}{1+1/|A_v|} = \frac{1}{sC_f(1+1/|A_v|)} \quad \Rightarrow \quad \text{only slightly larger!}$$

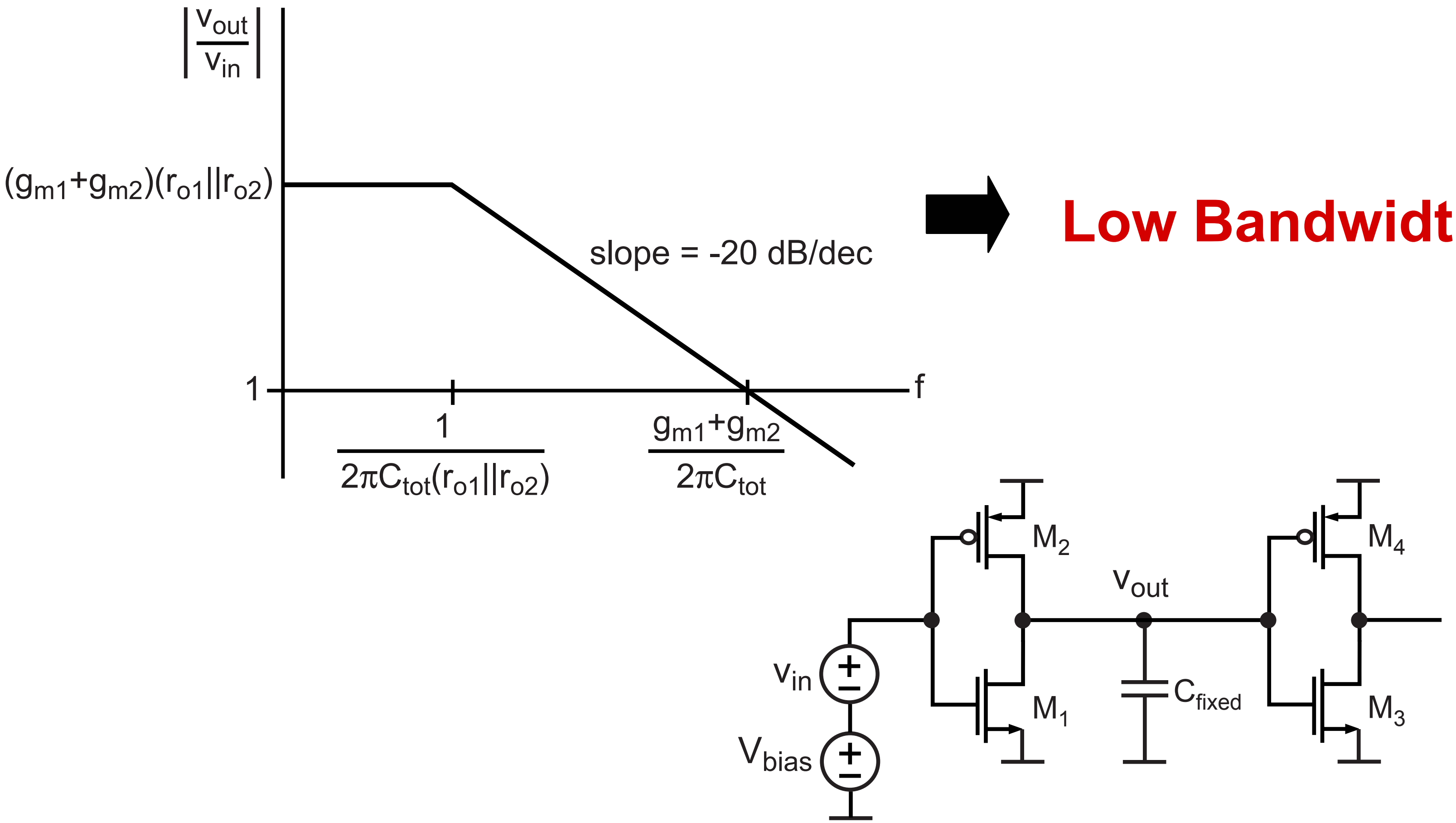
Amplifier Example – CMOS Inverter

- **Assume that we set V_{bias} such that the amplifier nominal output is such that NMOS and PMOS transistors are all in saturation**
  - **Note: this topology VERY sensitive to bias errors**



$$C_{tot} = \underset{\substack{\Uparrow\\(+C_{ov1}+C_{ov2})}}{C_{db1}+C_{db2}} + C_{gs3}+C_{gs4} + \underset{\substack{\Uparrow\\\text{Miller multiplication factor}}}{K}(C_{ov3}+C_{ov4}) + C_{fixed}$$

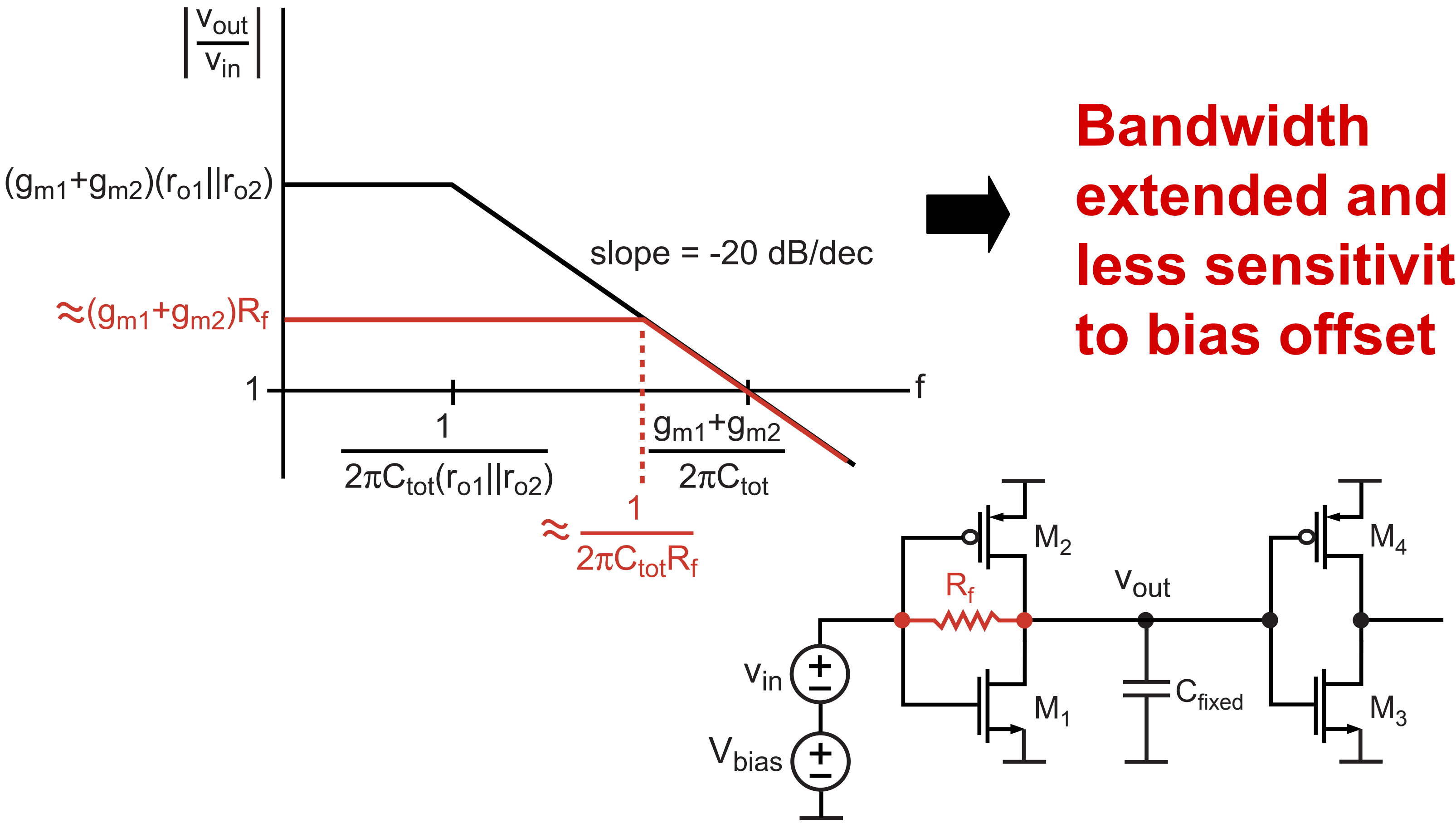
Transfer Function of CMOS Inverter



Low Bandwidth!

$$C_{tot} = \underset{\substack{\uparrow \\ (+C_{ov1}+C_{ov2})}}{C_{db1}+C_{db2}} + C_{gs3}+C_{gs4} + \underset{\substack{\uparrow \\ \text{Miller multiplication factor}}}{K}(C_{ov3}+C_{ov4}) + C_{fixed}$$

Add Resistive Feedback

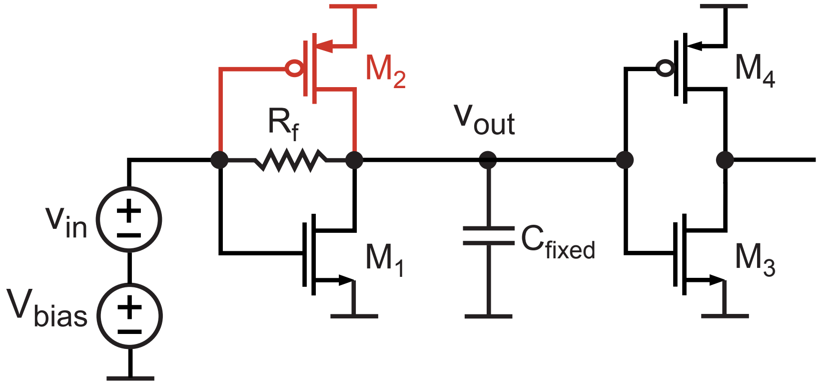


Bandwidth extended and less sensitivity to bias offset

$$C_{tot} = \underset{\substack{\Uparrow \\ (+C_{ov1}+C_{ov2})}}{C_{db1}+C_{db2}} + C_{gs3}+C_{gs4} + \underset{\substack{\Uparrow \\ \text{Miller multiplication factor}}}{K}(C_{ov3}+C_{ov4}) + C_{Rf}/2 + C_{fixed}$$

We Can Still Do Better

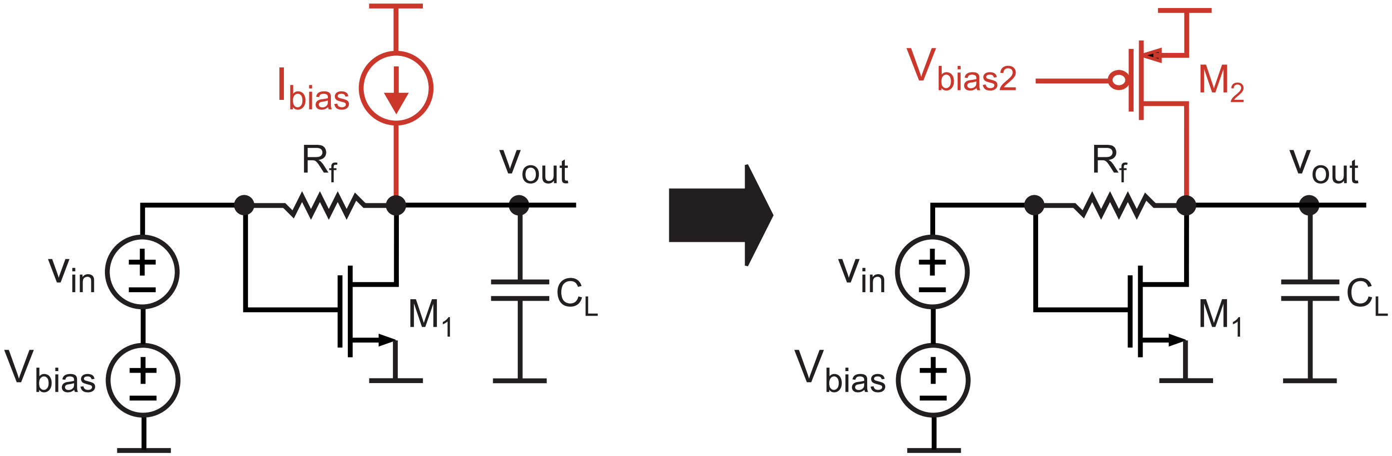
- We are fundamentally looking for high g_m to capacitance ratio to get the highest bandwidth
  - PMOS degrades this ratio
  - Gate bias voltage is constrained



$$C_{tot} = C_{db1}\underset{\Uparrow}{+}C_{db2} + C_{gs3}+C_{gs4} + \underset{\Uparrow}{K}(C_{ov3}+C_{ov4}) + C_{Rf}/2 + C_{fixed}$$

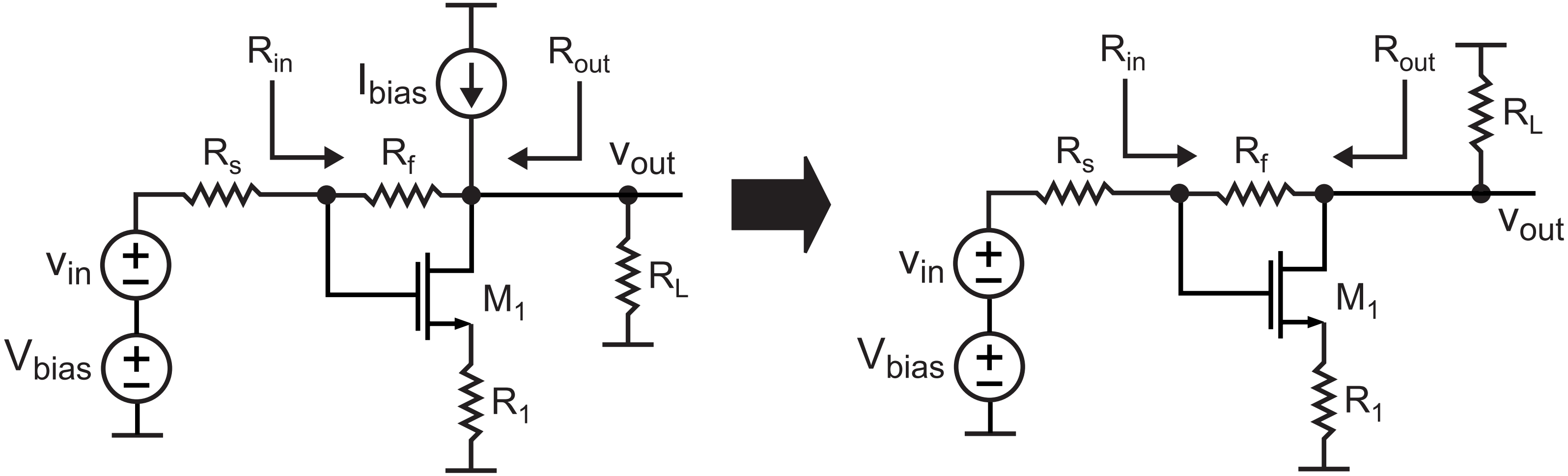
$(+C_{ov1}+C_{ov2})$ Miller multiplication factor

Take PMOS Out of the Signal Path



- **Advantages**
  - **PMOS gate no longer loads the signal**
  - **NMOS device can be biased at a higher voltage**
- **Issue**
  - **PMOS is not an efficient current provider (I_d/drain cap)**
    - Drain cap close in value to C_{gs}
  - **Signal path is loaded by cap of R_f and drain cap of PMOS**

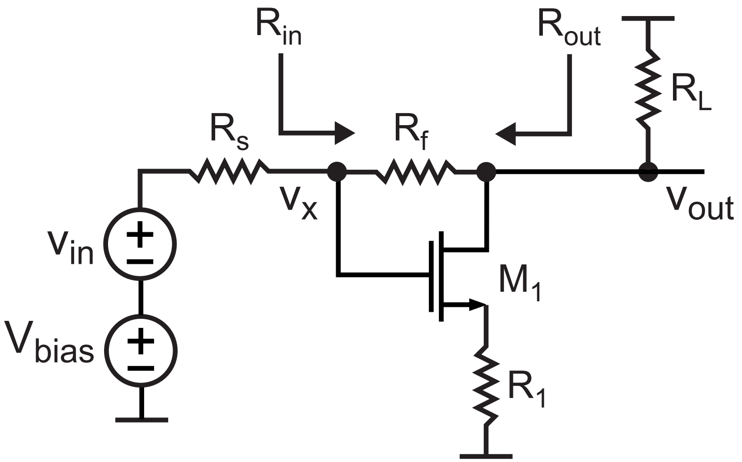
Shunt-Series Amplifier



- Use resistors to control the bias, gain, and input/output impedances
  - Improves accuracy over process and temp variations
- Issues
  - Degeneration of M_1 lowers slew rate for large signal applications (such as limit amps)
  - There are better high speed approaches – the advantage of this one is simply accuracy

Shunt-Series Amplifier – Analysis Snapshot

- **From Chapter 8 of Tom Lee's book (see pp 191-197):**
  - **Gain**



$$A_v = \frac{v_{out}}{v_{in}} = -\frac{R_L}{R_E}\left(\frac{R_f - R_E}{R_f + R_L}\right)$$

where: $R_E = 1/g_m + R_1$

Note: $\boxed{A_v \approx -\frac{R_L}{R_1}}$ for $R_f \gg R_L,\ R_f \gg R_E,\ R_1 \gg 1/g_m$

  - **Input resistance**

$$R_{in} = \frac{R_f}{1 - A_v} = \frac{R_E(R_f + R_L)}{R_E + R_L} \approx \frac{R_f}{1 + R_L/R_1}$$

for $R_f \gg R_L,$
$R_1 \gg 1/g_m$

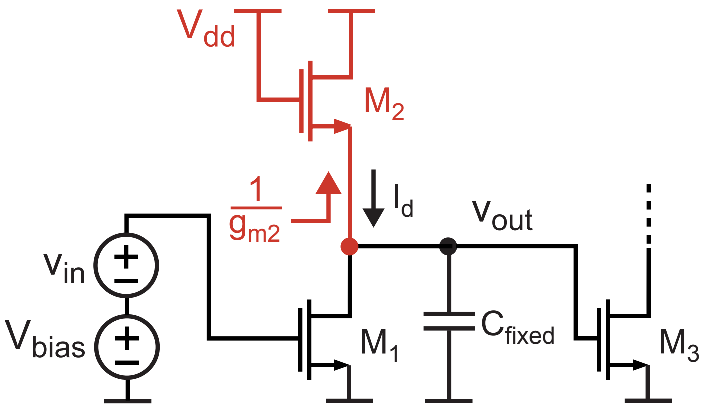
  - **Output resistance**

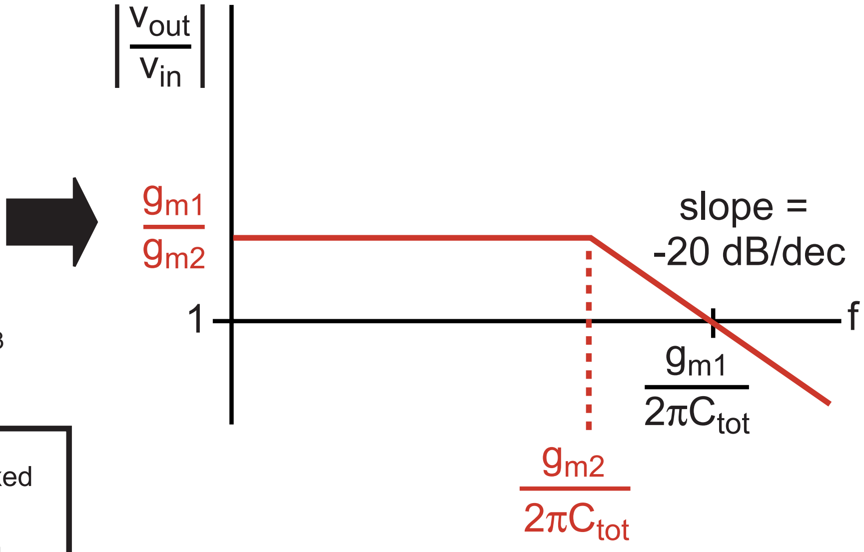
↔ **Same for $R_s = R_L$!**

$$R_{out} = \frac{R_E(R_f + R_s)}{R_E + R_s} \approx \frac{R_f}{1 + R_s/R_1}$$

for $R_f \gg R_s, R_1 \gg 1/g_m$

NMOS Load Amplifier





$C_{tot} = C_{db1} + C_{sb2} + C_{gs2} + C_{gs3} + KC_{ov3} + C_{fixed}$

(+C_{ov1}) Miller multiplication factor

- **Gain set by the relative sizing of M_1 and M_2**

$$M_1 :\ I_{d1} = (1/2)\mu_n C_{ox}(W_1/L_1)(V_{IN} - V_T)^2$$

$$M_2 :\ I_{d2} = (1/2)\mu_n C_{ox}(W_2/L_2)(V_{dd} - V_{out} - V_T)^2$$

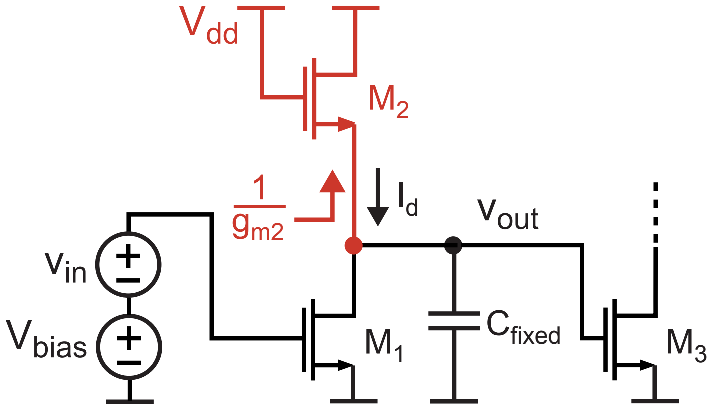
$$\Rightarrow I_{d1} = I_{d2}$$

$$\Rightarrow\ V_{out} = -AV_{IN} + V_{dd} + (A-1)V_T$$

$$(V_{IN} = V_{in} + V_{bias})$$

$$\boxed{\text{where}\ \ A = \sqrt{\frac{W_1/L_1}{W_2/L_2}}}$$

Design of NMOS Load Amplifier

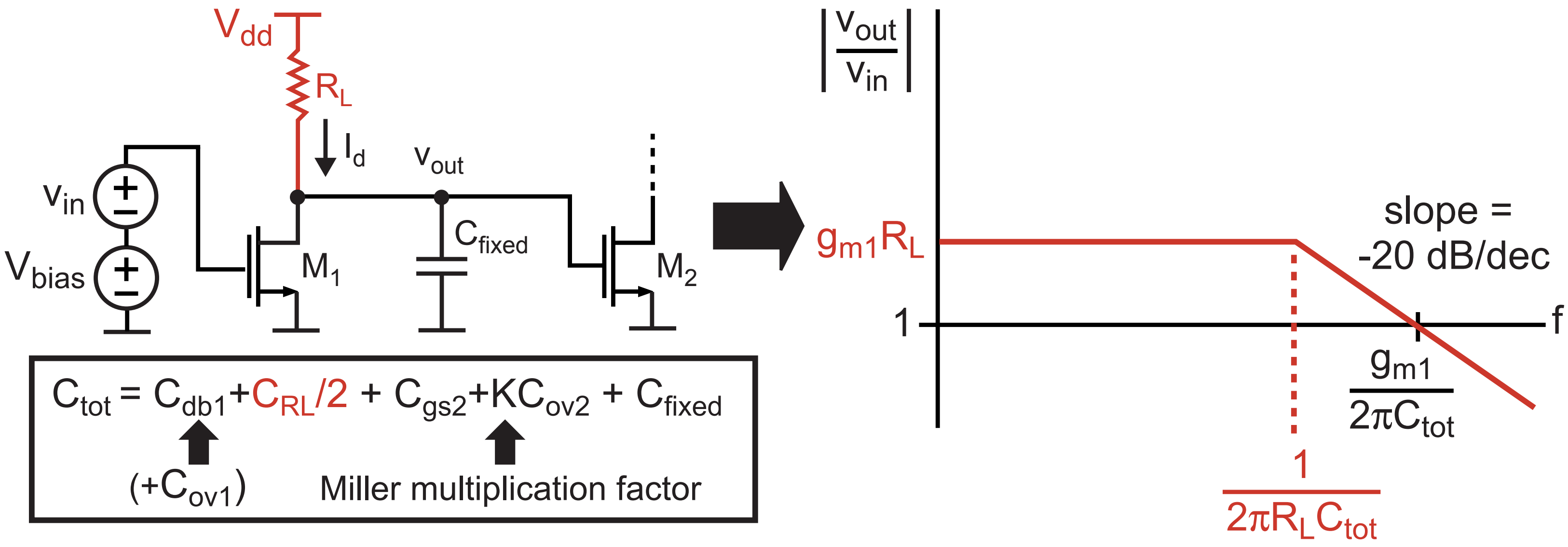


$$C_{tot} = \underset{\substack{\uparrow \\ (+C_{ov1})}}{C_{db1}} + C_{sb2} + C_{gs2} + C_{gs3} + \underset{\substack{\uparrow \\ \text{Miller multiplication factor}}}{K} C_{ov3} + C_{fixed}$$

$$A = \sqrt{\frac{W_1/L_1}{W_2/L_2}}$$

- **Size transistors for gain and speed**
  - **Choose minimum L for maximum speed**
  - **Choose ratio of W_1 to W_2 to achieve appropriate gain**
- **Problem: V_T of M_2 lowers the bias voltage of the next stage (thus lowering its achievable f_t)**
  - **Severely hampers performance when amplifier is cascaded**
  - **One person solved this issue by increasing V_{dd} of NMOS load (see Sackinger et. al., “A 3-GHz 32-dB CMOS Limiting Amplifier for SONET OC-48 receivers”, JSSC, Dec 2000)**

Resistor Loaded Amplifier (Unsilicided Poly)



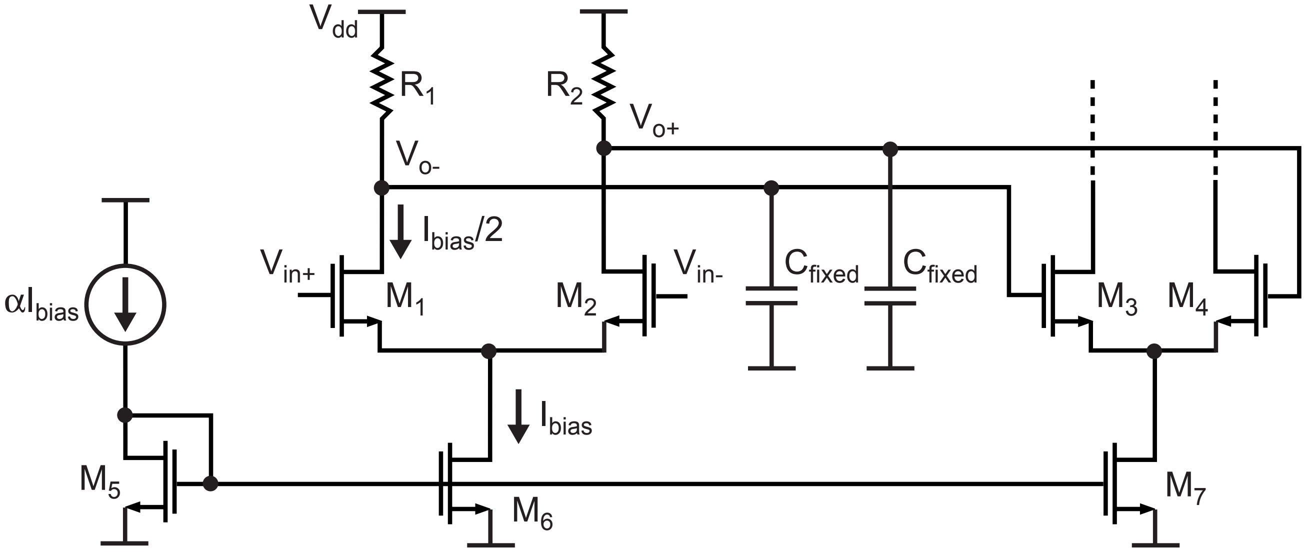
$C_{tot} = C_{db1} + C_{RL}/2 + C_{gs2} + KC_{ov2} + C_{fixed}$

$(+C_{ov1})$ Miller multiplication factor

- **This is the fastest non-enhanced amplifier I've found**
  - **Unsilicided poly is a pretty efficient current provider (i..e, has a good current to capacitance ratio)**
  - **Output swing can go all the way up to V_{dd}**
    - Allows following stage to achieve high f_t
  - **Linear settling behavior (in contrast to NMOS load)**

Implementation of Resistor Loaded Amplifier

- **Typically implement using differential pairs**



- **Benefits**
  - **Self-biased**
  - **Common-mode rejection**
- **Negative**
  - **More power than single-ended version**

The Issue of Velocity Saturation

- **We classically assume that MOS current is calculated as**

$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{gs} - V_T)^2$$

- **Which is really**

$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{gs} - V_T)V_{dsat,l}$$

  - **$V_{dsat,l}$ corresponds to the saturation voltage at a given length, which we often refer to as ΔV**

- **It may be shown that**

$$V_{dsat,l} \approx \frac{(V_{gs} - V_T)(LE_{sat})}{(V_{gs} - V_T) + (LE_{sat})} = (V_{gs}-V_T)||(LE_{sat})$$

  - **If V_{gs}-V_T approaches LE_{sat} in value, then the top equation is no longer valid**
    - We say that the device is in velocity saturation

Analytical Device Modeling in Velocity Saturation

- If L small (as in modern devices), than velocity saturation will impact us for even moderate values of V_{gs}-V_T

$$I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{gs} - V_T)[(V_{gs} - V_T)||(LE_{sat})]$$

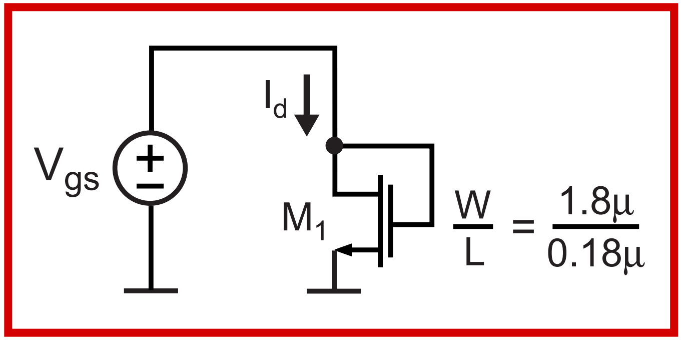
$$\boxed{\Rightarrow \ I_D \approx \frac{\mu_n C_{ox}}{2} W(V_{gs} - V_T)E_{sat}}$$

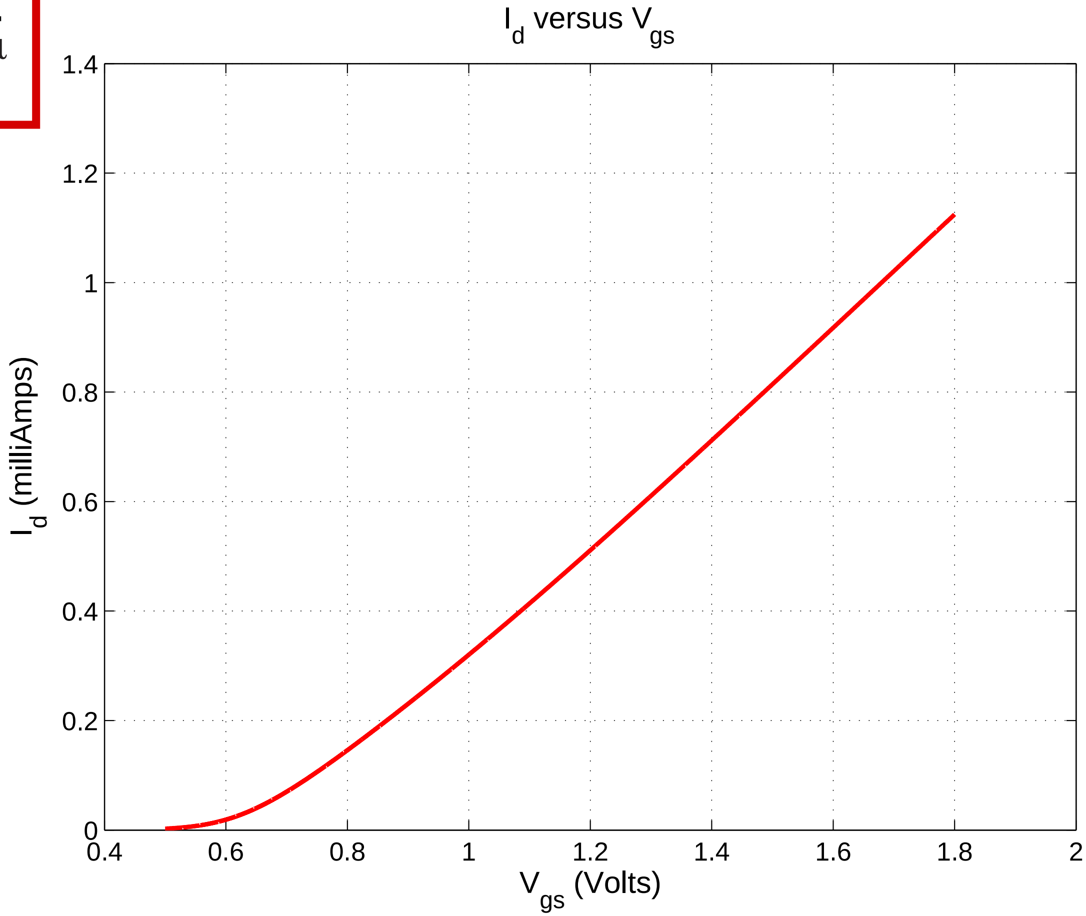
  - Current increases linearly with V_{gs}-V_T!
- Transconductance in velocity saturation:

$$g_m = \frac{dI_d}{dV_{gs}} \ \Rightarrow \boxed{g_m = \frac{\mu_n C_{ox}}{2} W E_{sat}}$$

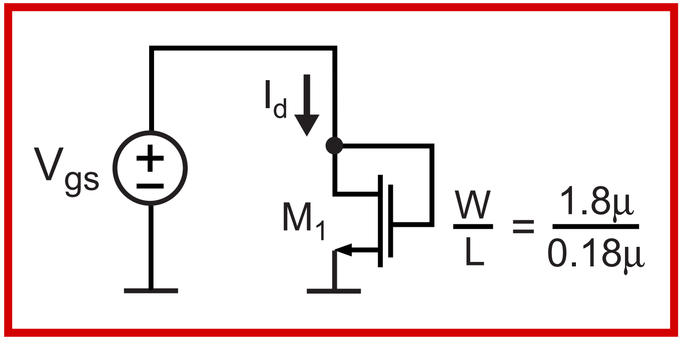
  - No longer a function of V_{gs}!

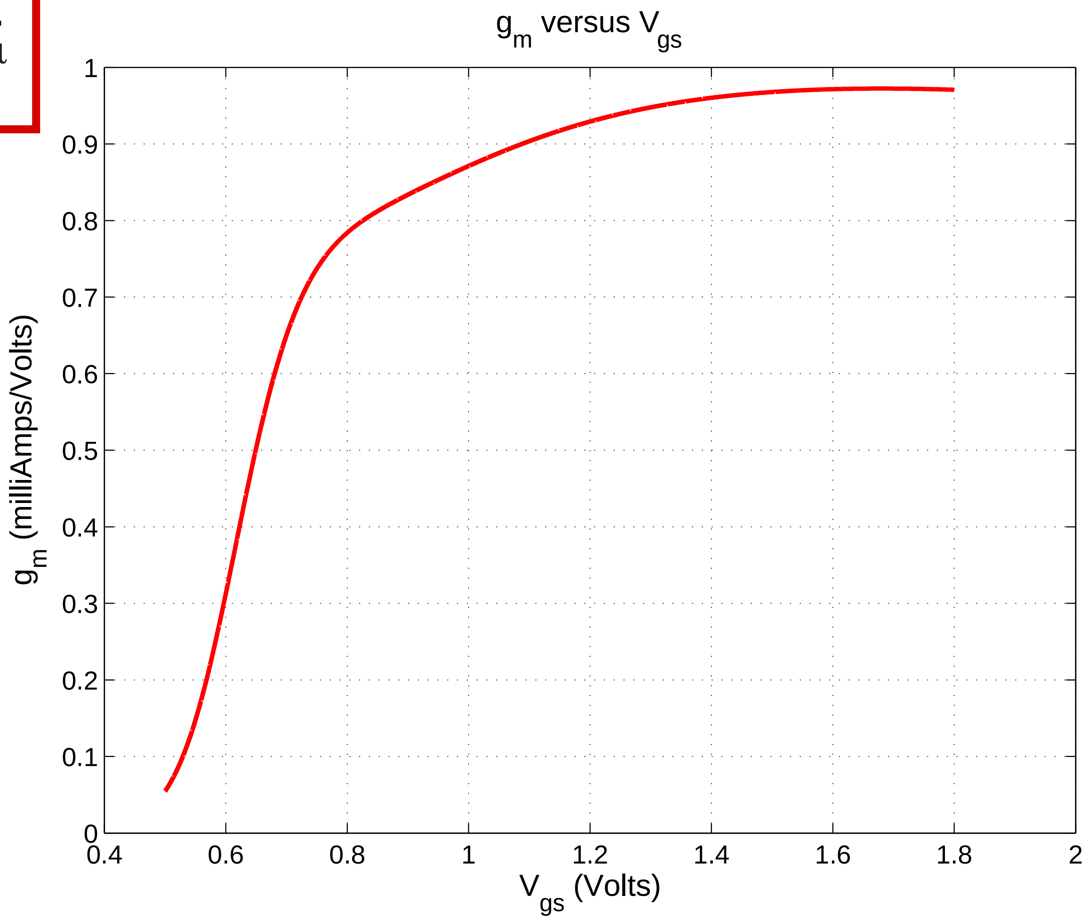
Example: Current Versus Voltage for 0.18μ Device



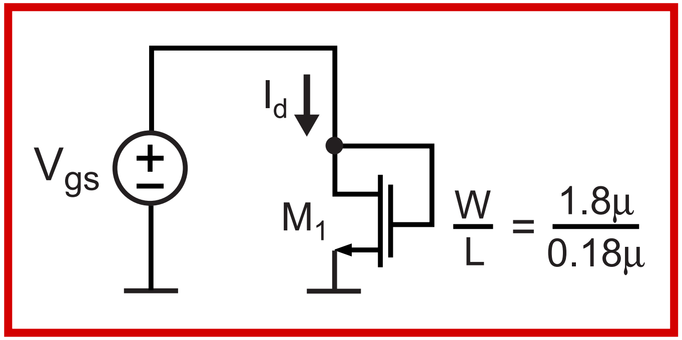


Example: G_m Versus Voltage for 0.18μ Device

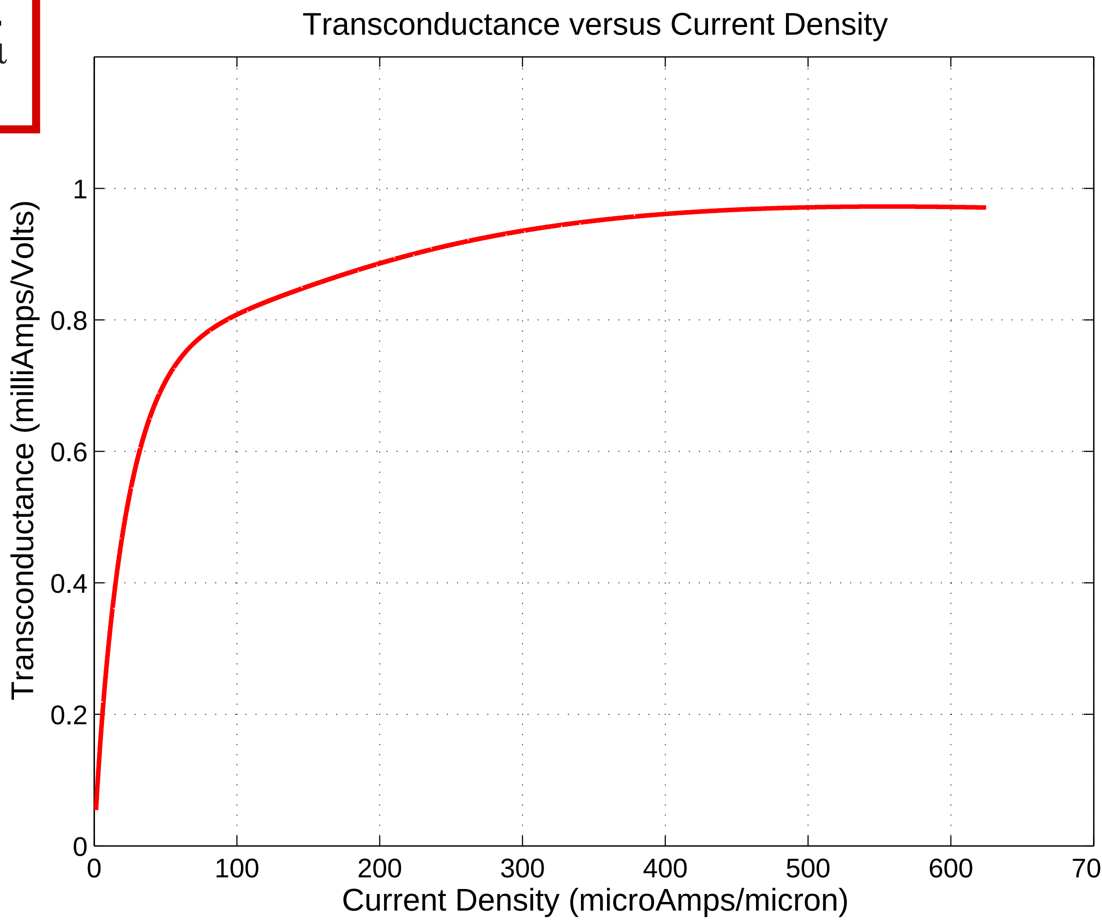




Example: G_m Versus Current Density for 0.18μ Device



$$\text{Note:}\quad I_{den} = \frac{I_d}{W} = \frac{I_d}{1.8\mu}$$



How Do We Design the Amplifier?

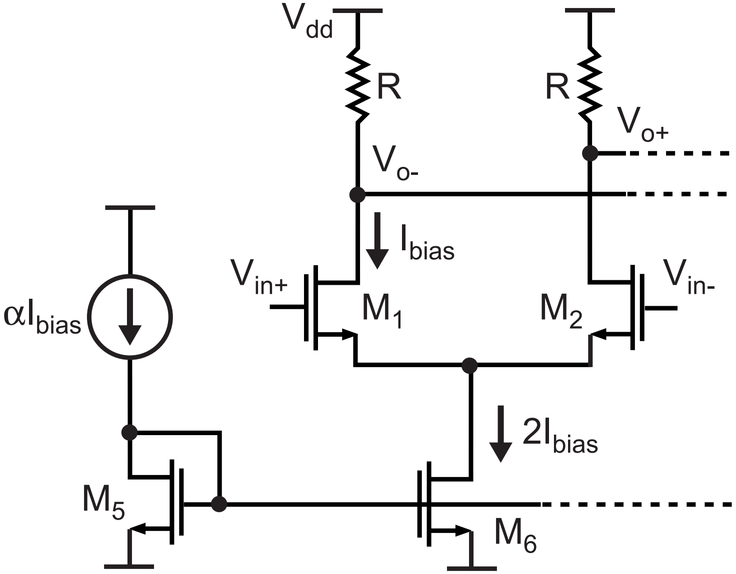
- **Highly inaccurate to assume square law behavior**
- **We will now introduce a numerical procedure based on the simulated g_m curve of a transistor**
  - **A look at g_m assuming square law device:**

$$g_m = \sqrt{2\mu C_{ox}\frac{W}{L}I_d} = W\sqrt{2\mu C_{ox}\frac{W}{L}\left(\frac{I_d}{W}\right)}$$

  - **Observe that if we keep the current density (I_d/W) constant, then g_m scales directly with W**
    - This turns out to be true outside the square-law regime as well
  - **We can therefore relate $g_m$of devices with different widths given that have the same current density**

$$\boxed{g_m(W, I_{den}) = \frac{W}{W_o}g_m(W_o, I_{den})}$$

A Numerical Design Procedure for Resistor Amp – Step 1



- **Two key equations**
  - **Set gain and swing (single-ended)**

$$(1)\quad g_m(W, I_{bias}/W)R = A$$

$$(2)\quad V_{sw} = 2I_{bias}R$$

- **Equate (1) and (2) through R**

$$\frac{A}{g_m(W, I_{bias}/W)} = \frac{V_{sw}}{2I_{bias}}$$

$$\Rightarrow \quad g_m(W, I_{bias}/W) = 2\frac{A}{V_{sw}}W\left(\frac{I_{bias}}{W}\right)$$

Can we relate this formula to a g_m curve taken from a device of width W_o?

A Numerical Design Procedure for Resistor Amp – Step 2

- **We now know:**

$$(1)\quad g_m(W, I_{bias}/W) = 2\frac{A}{V_{sw}}W\left(\frac{I_{bias}}{W}\right)$$

$$(2)\quad g_m(W, I_{den}) = \frac{W}{W_o}g_m(W_o, I_{den})$$

- **Substitute (2) into (1)**

$$\frac{W}{W_o}g_m(W_o, I_{bias}/W) = 2\frac{A}{V_{sw}}W\left(\frac{I_{bias}}{W}\right)$$

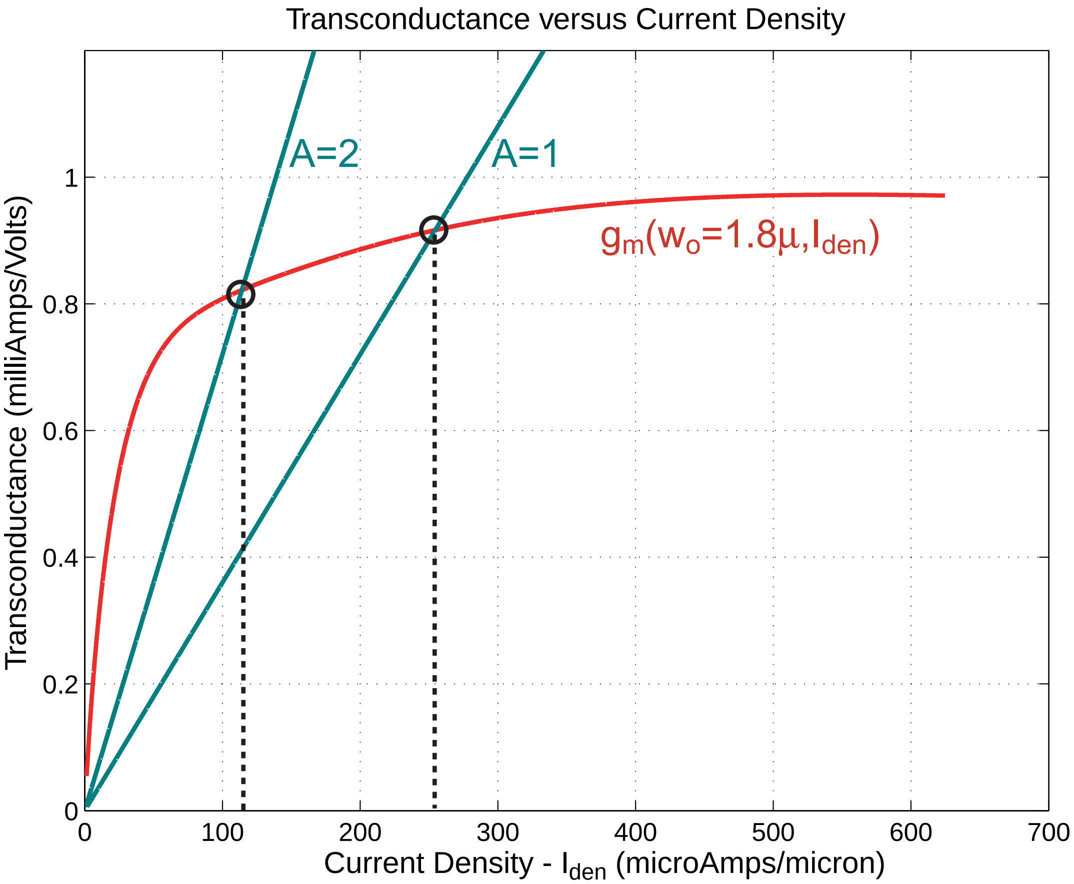
$$\boxed{\Rightarrow\quad g_m(W_o, I_{den}) = 2W_o\frac{A}{V_{sw}}I_{den}}$$

- **The above expression allows us to design the resistor loaded amp based on the g_m curve of a representative transistor of width W_o!**

Example: Design for Swing of 1 V, Gain of 1 and 2

$$\boxed{g_m(W_o, I_{den}) = 2W_o \frac{A}{V_{sw}} I_{den}}$$

- **Assume L=0.18μ, use previous g_m plot (W_o=1.8μ)**



- **For gain of 1, current density = 250 μA/μm**
- **For gain of 2, current density = 115 μA/μm**
- **Note that current density reduced as gain increases!**
  - **f_t effectively decreased**

Example (Continued)

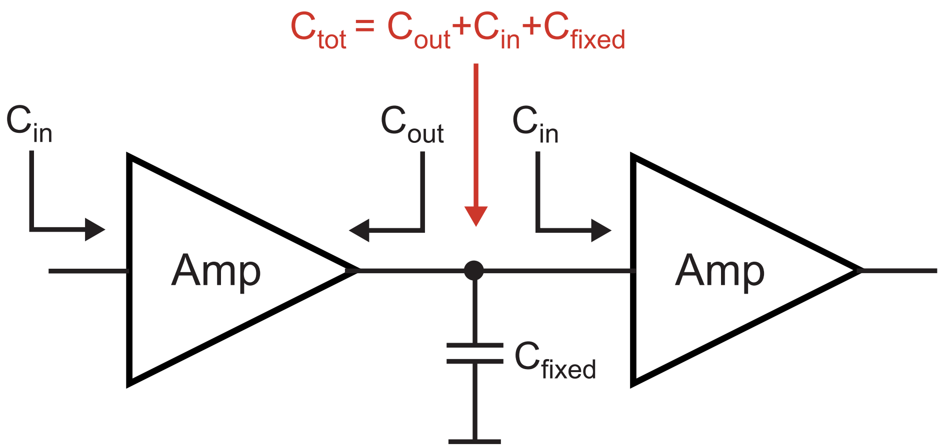
- Knowledge of the current density allows us to design the amplifier
  - Recall $V_{sw} = 2I_{bias}R$
  - Free parameters are W, I_{bias}, and R (L assumed to be fixed)
- Given I_{den} = 115 μA/μm (Swing = 1V, Gain = 2)
  - If we choose I_{bias} = 300 μA

$$I_{den} = \frac{I_{bias}}{W} \quad \Rightarrow \quad W = \frac{300}{115} = 2.6\mu m$$

$$V_{sw} = 2I_{bias}R \quad \Rightarrow \quad R = \frac{1}{2 \cdot 300 \times 10^{-6}} = 1.67k\Omega$$

- Note that we could instead choose W or R, and then calculate the other parameters

How Do We Choose I_{bias} For High Bandwidth?



- **As you increase I_{bias}, the size of transistors also increases to keep a constant current density**
  - **The size of C_{in} and C_{out} increases relative to C_{fixed}**
- **To achieve high bandwidth, want to size the devices (i.e., choose the value for I_{bias}), such that**
  - **C_{in}+C_{out} roughly equal to C_{fixed}**