وَمَا أُوتِيتُمْ مِنَ الْعِلْمِ إِلَّا قَلِيلًا

Analog IC Design

Lecture 13
gm/ID Design Methodology

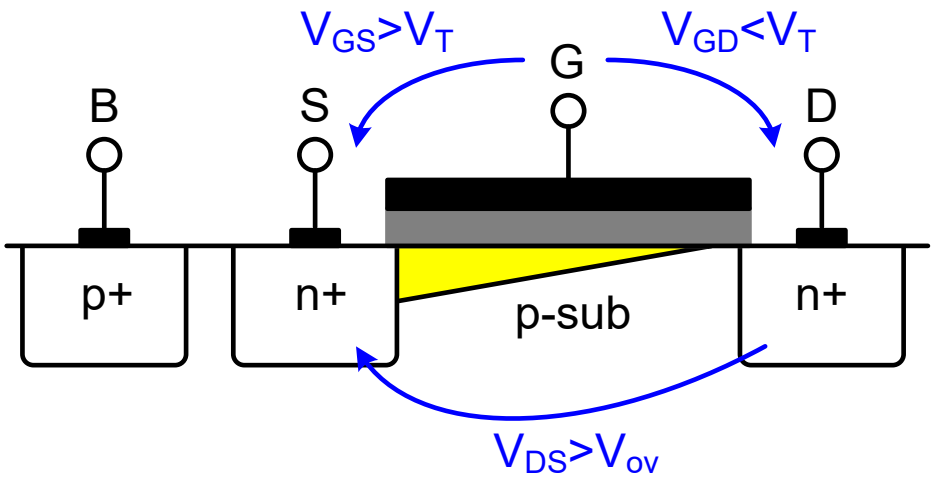
Dr. Hesham A. Omran

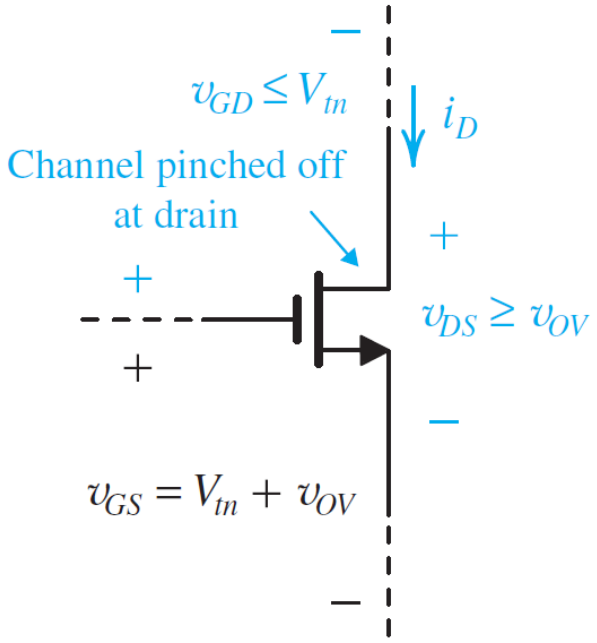
Integrated Circuits Lab (ICL)
Electronics and Communications Eng. Dept.
Faculty of Engineering
Ain Shams University

MOSFET in Saturation

- The channel is pinched off if the difference between the gate and drain voltages is not sufficient to create an inversion layer

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





[Sedra/Smith, 2015]

Regions of Operation Summary

OFF (Subthreshold)

$V_{GS} < V_T$

ON

$V_{GS} > V_T$

Triode

$V_{DS} < V_{ov}$

Or

$V_{GD} > V_T$

$$I_D = \mu C_{ox} \frac{W}{L}\left(V_{ov}V_{DS} - \frac{V_{DS}^2}{2}\right)$$

Pinch-Off (Saturation)

$V_{DS} \geq V_{ov}$

Or

$V_{GD} \leq V_T$

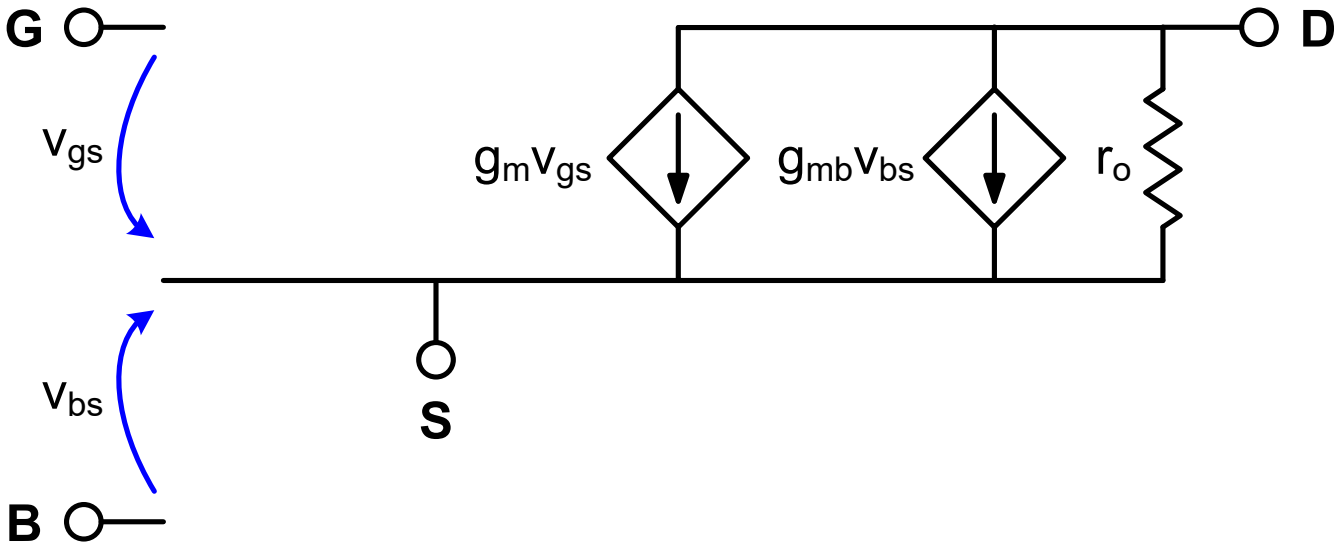
$$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}V_{ov}^2(1 + \lambda V_{DS})$$

Low-Frequency Small-Signal Model

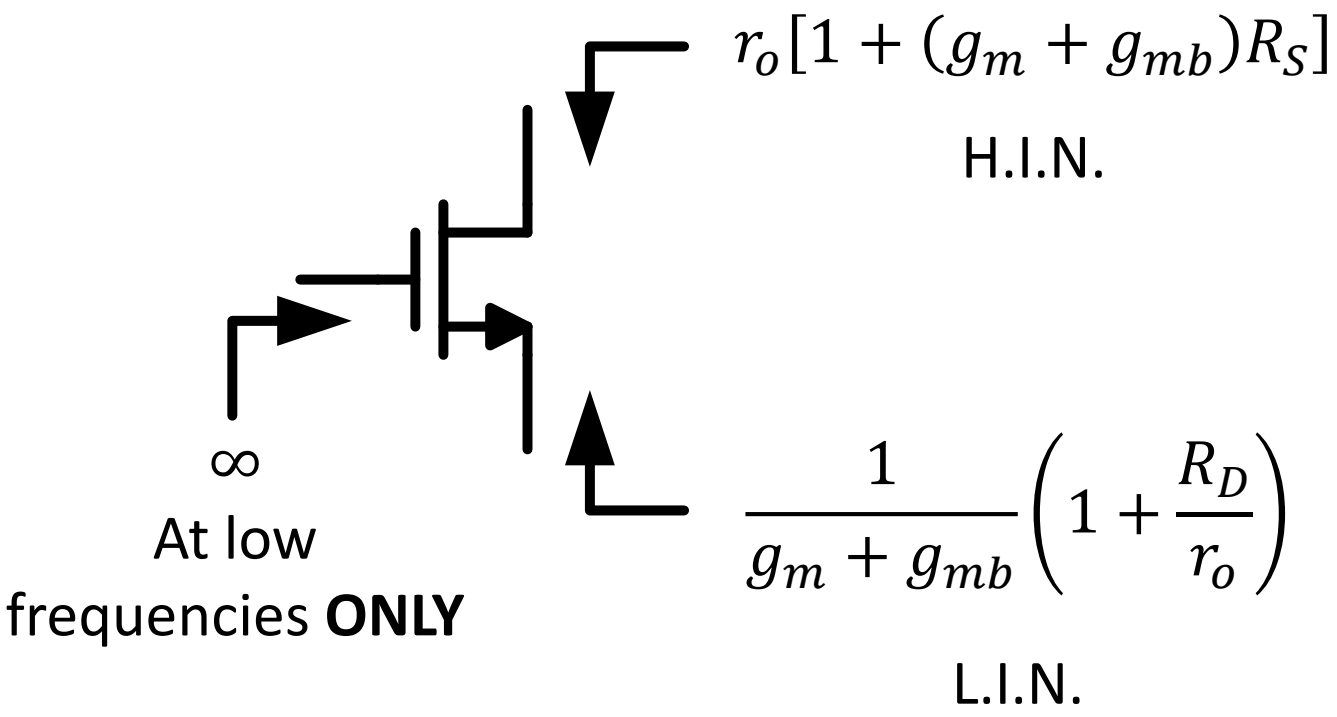
$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \mu C_{ox}\frac{W}{L}V_{ov} = \sqrt{\mu C_{ox}\frac{W}{L}\cdot 2I_D} = \frac{2I_D}{V_{ov}}$$

$$g_{mb} = \eta g_m, \qquad \eta \approx 0.1 - 0.25$$

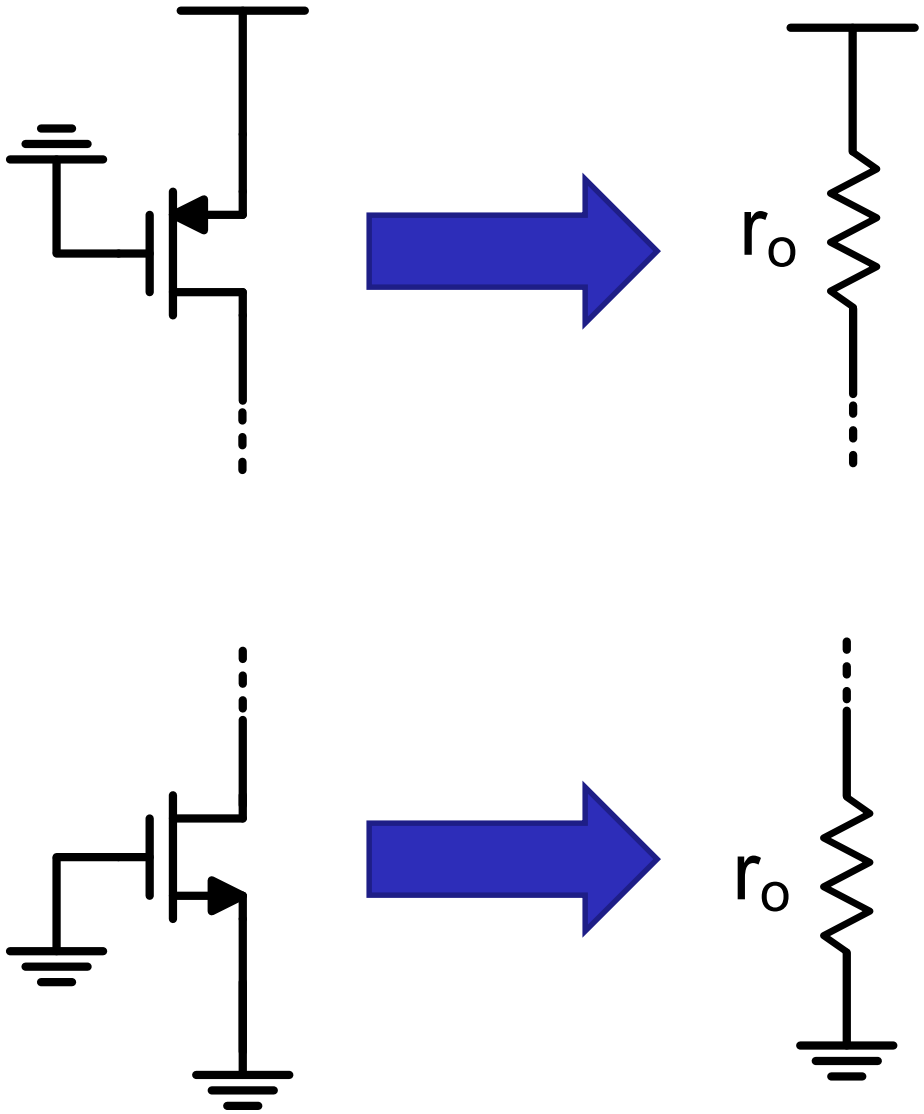
$$r_o = \frac{1}{\frac{\partial I_D}{\partial V_{DS}}} = \frac{1}{\lambda I_D}, \qquad \lambda \propto \frac{1}{L}$$



Rin/out Shortcuts Summary

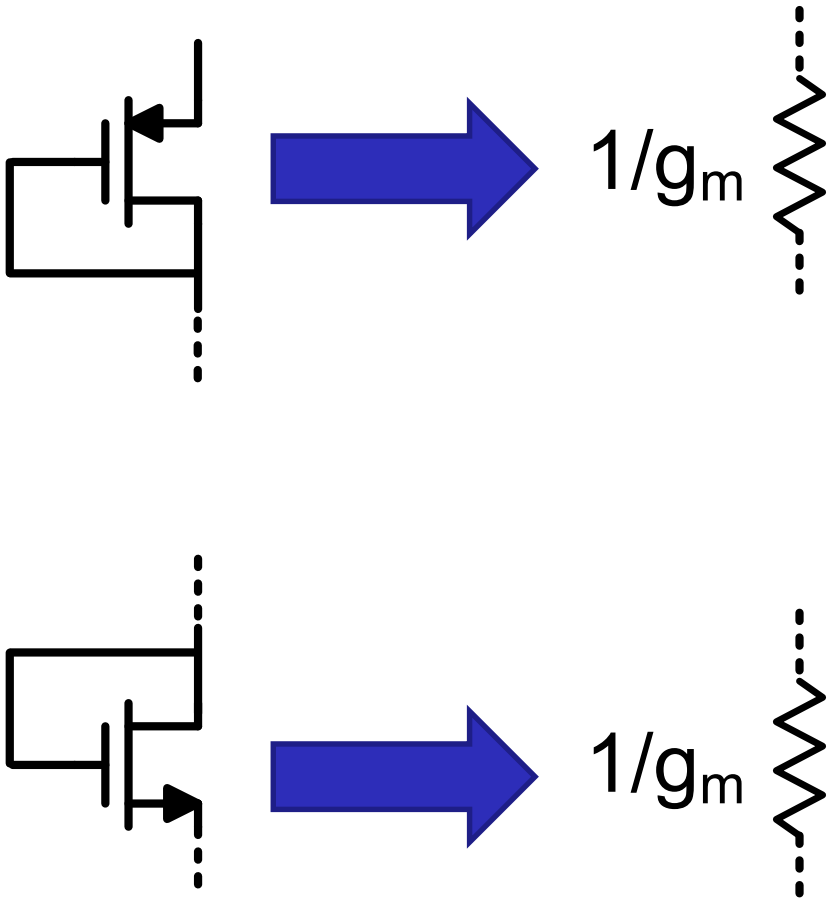


Active Load (Source OFF)



Diode Connected (Source Absorption)

- Always in saturation
- Bulk effect: $g_m \rightarrow g_m + g_{mb}$



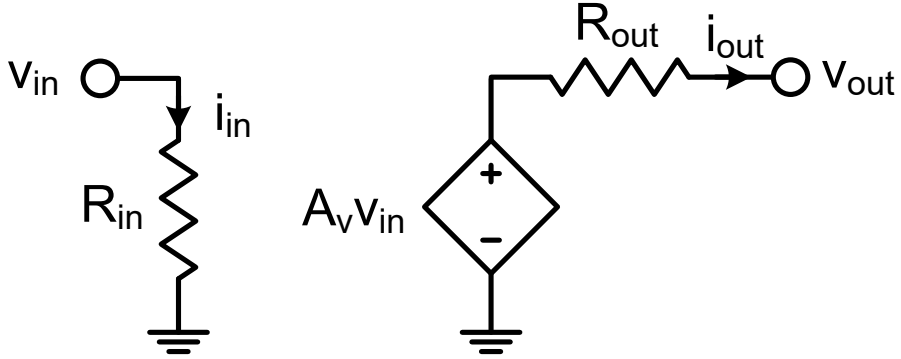
Why GmRout?

$$R_{out} = \frac{v_x}{i_x} \; @ \; v_{in} = 0$$

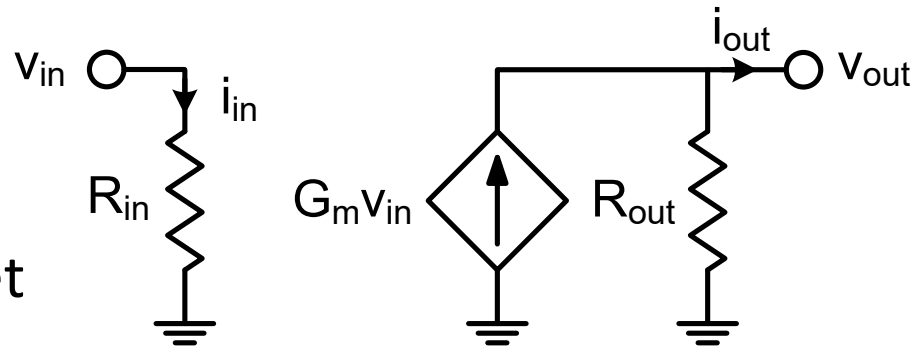
$$G_m = \frac{i_{out,sc}}{v_{in}}$$

$$A_v = G_m R_{out}$$

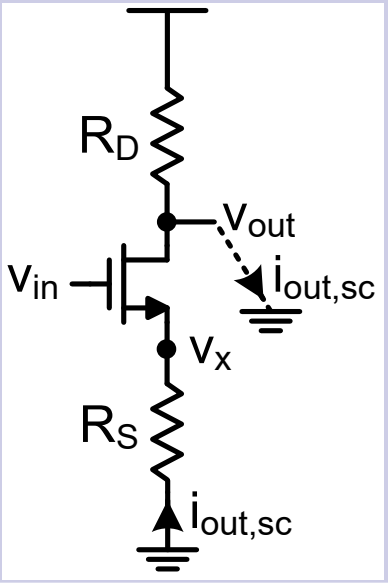
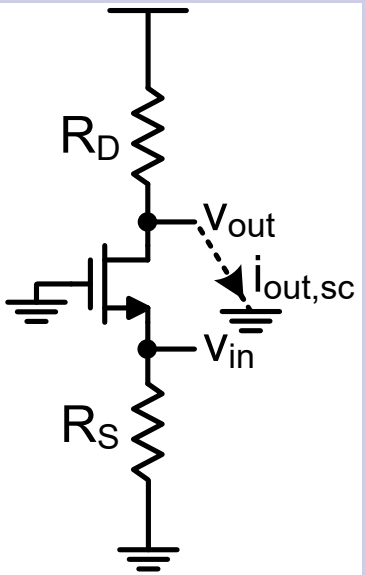
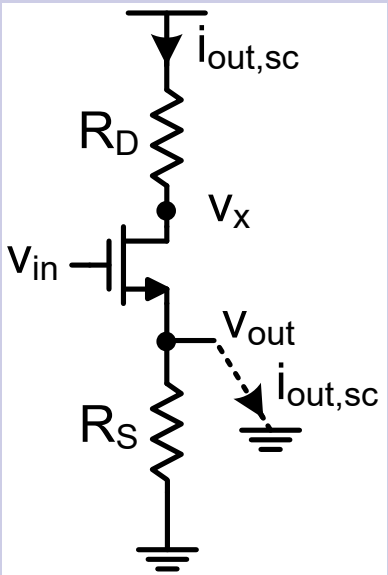
$$A_i = G_m R_{in}$$



- Divide and conquer
  - Rout simplified: vin=0
  - Gm simplified: vout=0
  - We already need Rin/out
  - We can quickly and easily get Rin/out from the shortcuts



Summary of Basic Topologies

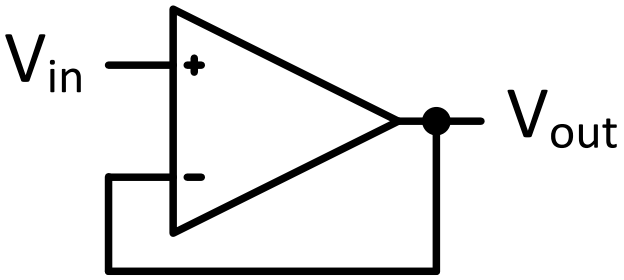
	CS	CG	CD (SF)
			
	Voltage & current amplifier	Current buffer	Voltage buffer
Rin	∞	$R_S // \frac{1}{g_m + g_{mb}}\left(1 + \frac{R_D}{r_o}\right)$	∞
Rout	$R_D // r_o[1 + (g_m + g_{mb})R_S]$	$R_D // r_o$	$R_S // \frac{1}{g_m + g_{mb}}\left(1 + \frac{R_D}{r_o}\right)$
Gm	$\frac{-g_m}{1 + (g_m + g_{mb})R_S}$	$g_m + g_{mb}$	$\frac{g_m}{1 + R_D/r_o}$

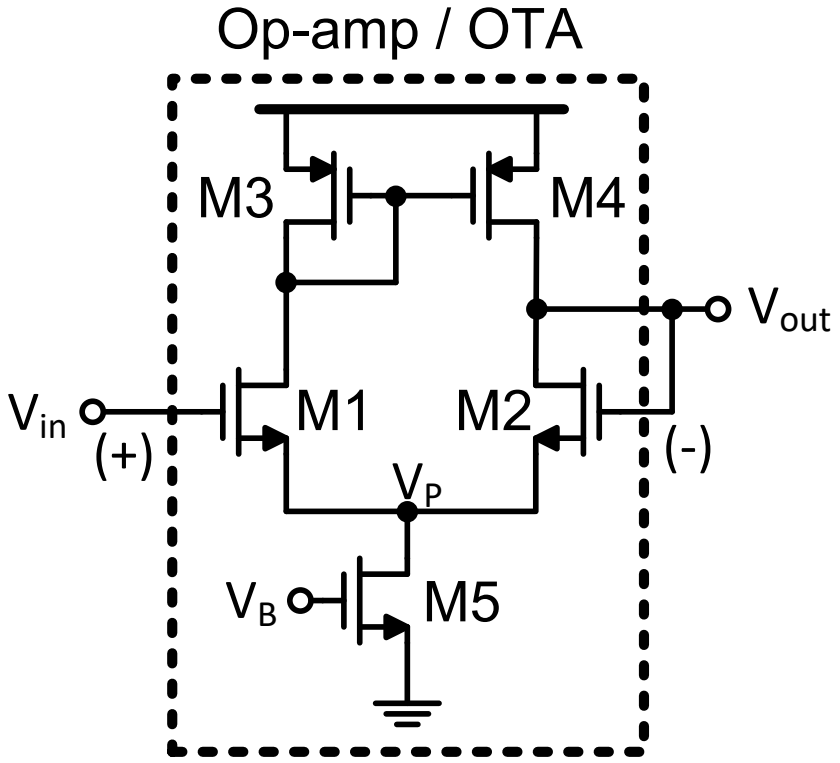
Differential Amplifier

	Pseudo Diff Amp	Diff Pair (w/ ideal CS)	Diff Pair (w/ R_{SS})
A_{vd}	$-g_m R_D$	$-g_m R_D$	$-g_m R_D$
A_{vCM}	$-g_m R_D$	0	$\frac{-g_m R_D}{1 + 2(g_m + g_{mb})R_{SS}}$
A_{vd}/A_{vCM}	1	∞	$2(g_m + g_{mb})R_{SS} \gg 1$

Op-Amp

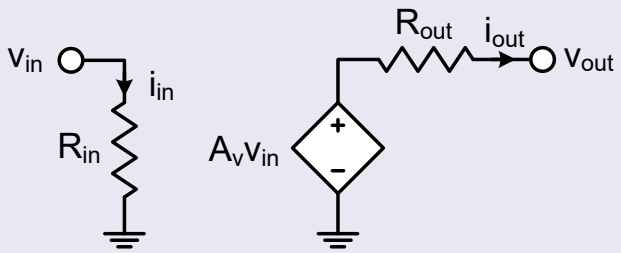
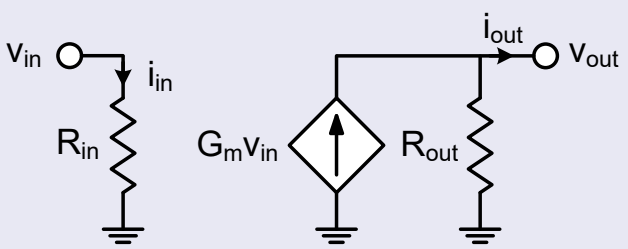
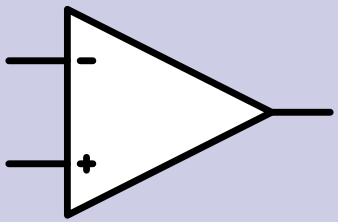
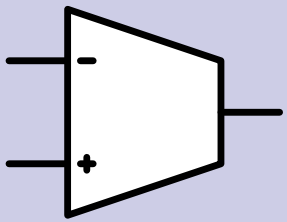
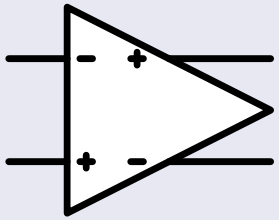
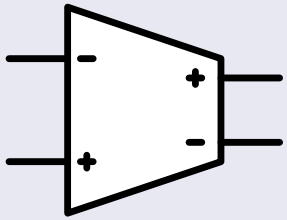
- An op-amp is simply a high gain differential amplifier
- The gain can be increased by using cascodes and multi-stage amplifiers





Op-Amp vs OTA

- An OTA is an op-amp without an output stage (buffer)
- Some designers just use op-amp name and symbol for both

	Op-amp	OTA
Rout	**LOW**	**HIGH**
Model		
Diff input, SE output		
Fully diff		

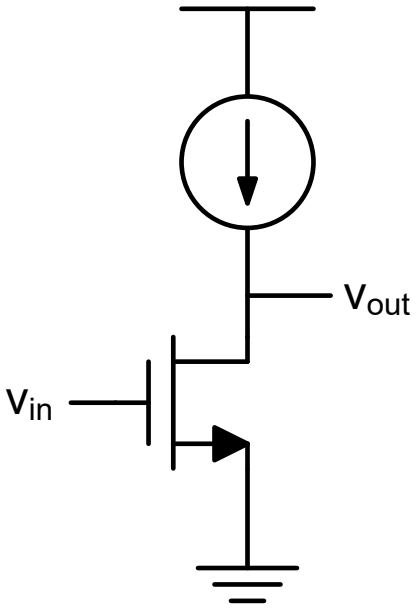
MOSFET Figures-of-Merit (FoM)

- Intrinsic gain (gm*ro)
- Intrinsic frequency (transit frequency) (gm/Cgg)
- Current efficiency (gm/ID)

Intrinsic Gain

$$v_{out} = -(g_m v_{in}) r_o$$

$$|A_v| = \left|\frac{v_{out}}{v_{in}}\right| = g_m r_o$$



- $g_m r_o$ is the max gain that can be obtained from a single transistor

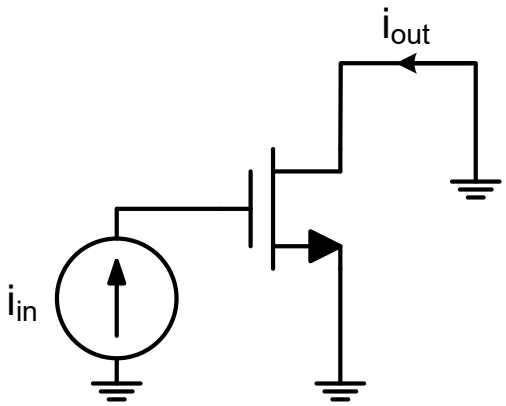
$$g_m r_o = \frac{2I_D}{V_{ov}} \cdot \frac{1}{\lambda I_D} = \frac{2}{\lambda V_{ov}}$$

- For higher gain
  - Lower V_{ov} (or equivalently lower I_D): weak inversion and subthreshold operation
  - Longer L (i.e., smaller λ)
  - Both come at the expense of speed

Intrinsic Frequency

- f_T is the frequency at which the MOSFET current gain drops to one (i.e., unity-gain frequency)

$$i_{out} = g_m v_{gs} = g_m \frac{i_{in}}{sC_{gg}}$$



@ $i_{out} = i_{in}$

$$f_T = \frac{g_m}{2\pi C_{gg}}$$

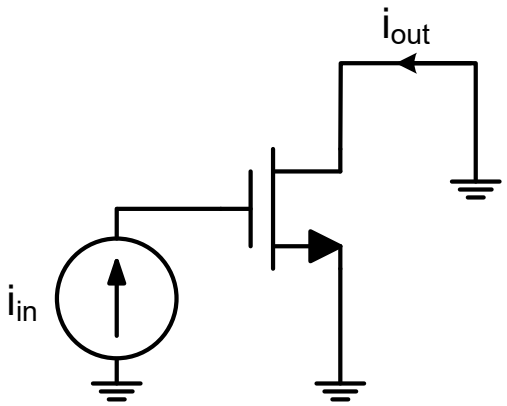
$$\approx \frac{1}{2\pi} \cdot \mu C_{ox} \frac{W}{L} V_{ov} \cdot \frac{1}{\frac{2}{3} WLC_{ox}}$$

$$\approx \frac{3\mu V_{ov}}{4\pi L^2}$$

Intrinsic Frequency (Speed)

- f_T is the frequency at which the MOSFET current gain drops to one (i.e., unity-gain frequency)

$$f_T \approx \frac{3\mu V_{ov}}{4\pi L^2}$$



- For higher speed
  - Higher V_{ov} (or equivalently higher I_D): strong inversion and higher power consumption
  - Shorter L (technology scaling helps!)
  - Just opposite to higher gain!
  - Analog design is all about trade-offs!
- After velocity sat, g_m and f_T saturate and become independent of V_{ov}

gm/ID

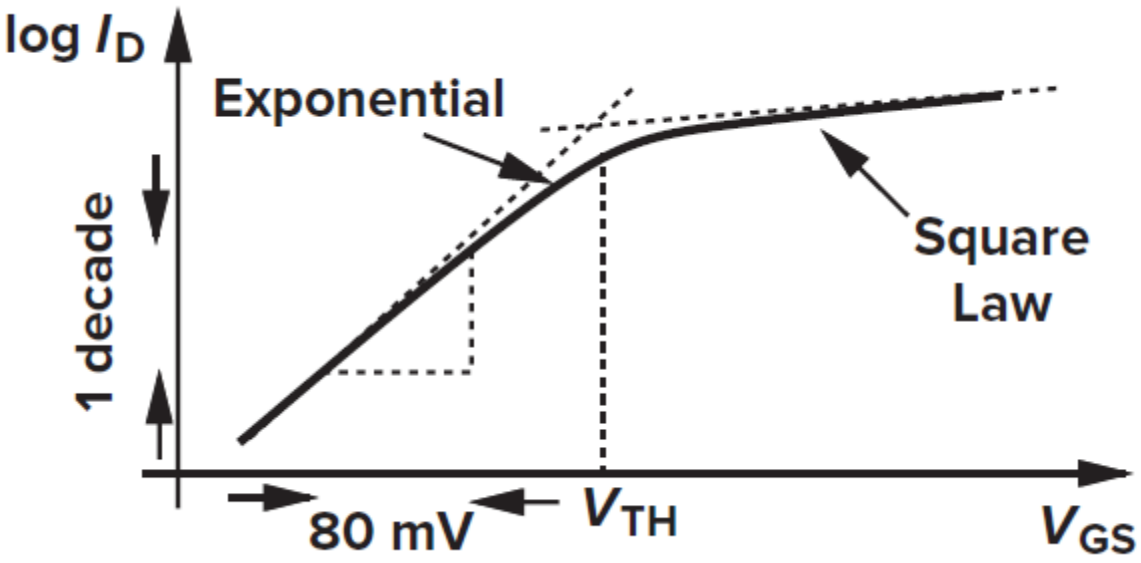
- gm/ID is the transconductance per unit current (a measure of energy efficiency)
  - How much transconductance (or GBW) can we get from each micro-amp of current

$$\frac{g_m}{I_D} = \frac{2}{V_{ov}}$$

- For higher efficiency
  - Lower V_{ov} (or equivalently lower I_D): weak inversion and subthreshold operation
  - Comes at the expense of speed

Subthreshold Operation

- Subthreshold slope (S) = 80mV/decade



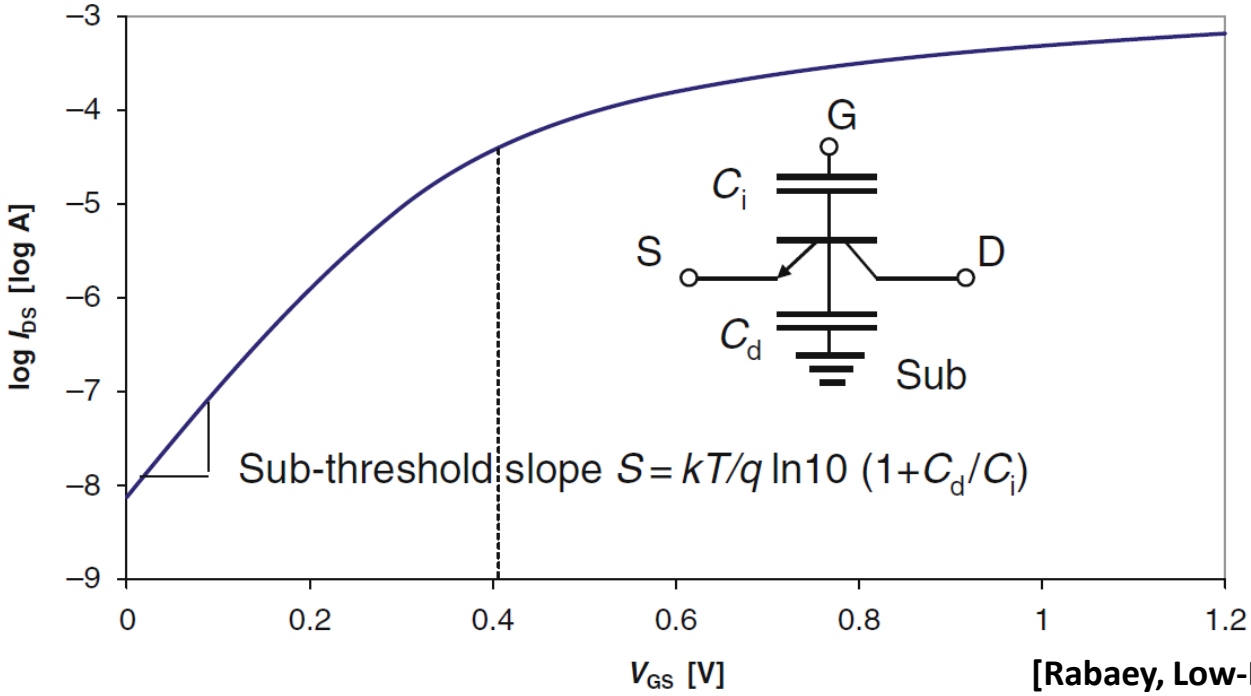
[Razavi, CMOS]

Subthreshold Operation

- MOSFET behaves as a BJT (npn for an NMOS) with its base coupled to the gate through capacitive divider

$$I_D = I_{Do}\frac{W}{L}e^{\frac{V_{GS}}{nV_T}}$$

$$n = \frac{C_i + C_d}{C_i} \approx 1.5$$



[Rabaey, Low-Power Essentials]

Subthreshold Operation

- MOSFET behaves as a BJT (npn for an NMOS) with its base coupled to the gate through capacitive divider

$$I_D = I_{Do}\frac{W}{L}e^{\frac{V_{GS}}{nV_T}}$$

$$V_T = \frac{kT}{q}$$

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{I_D}{nV_T} = \frac{2I_D}{V^*}$$

$$V^* = 2nV_T$$

- Onset of strong inversion

$$V^* = V_{ov} \rightarrow V_{ov} = 2nV_T \approx 80mV$$

- MOSFET in saturation if $V_{DS} > V^*$
- For strong inversion: $V^* = V_{ov}$

Subthreshold Intrinsic Gain

$$g_m r_o = \frac{2}{\lambda V^*} = \frac{1}{\lambda n V_T}$$

- Independent of V_{ov}
- No improvement as we go deeper in subthreshold

Subthreshold Intrinsic Speed

$$f_T = \frac{g_m}{2\pi C_{gg}}$$

$$\approx \frac{1}{2\pi} \cdot \frac{2I_D}{V^*} \cdot \frac{1}{C_{gg}}$$

$$\approx \frac{I_D}{2\pi n V_T C_{gg}}$$

- Continues to degrade as I_D decreases

Subthreshold Current Efficiency

$$\frac{g_m}{I_D} = \frac{2}{V^*} = \frac{1}{nV_T}$$

- Independent of V_{ov}
- No improvement as we go deeper in subthreshold

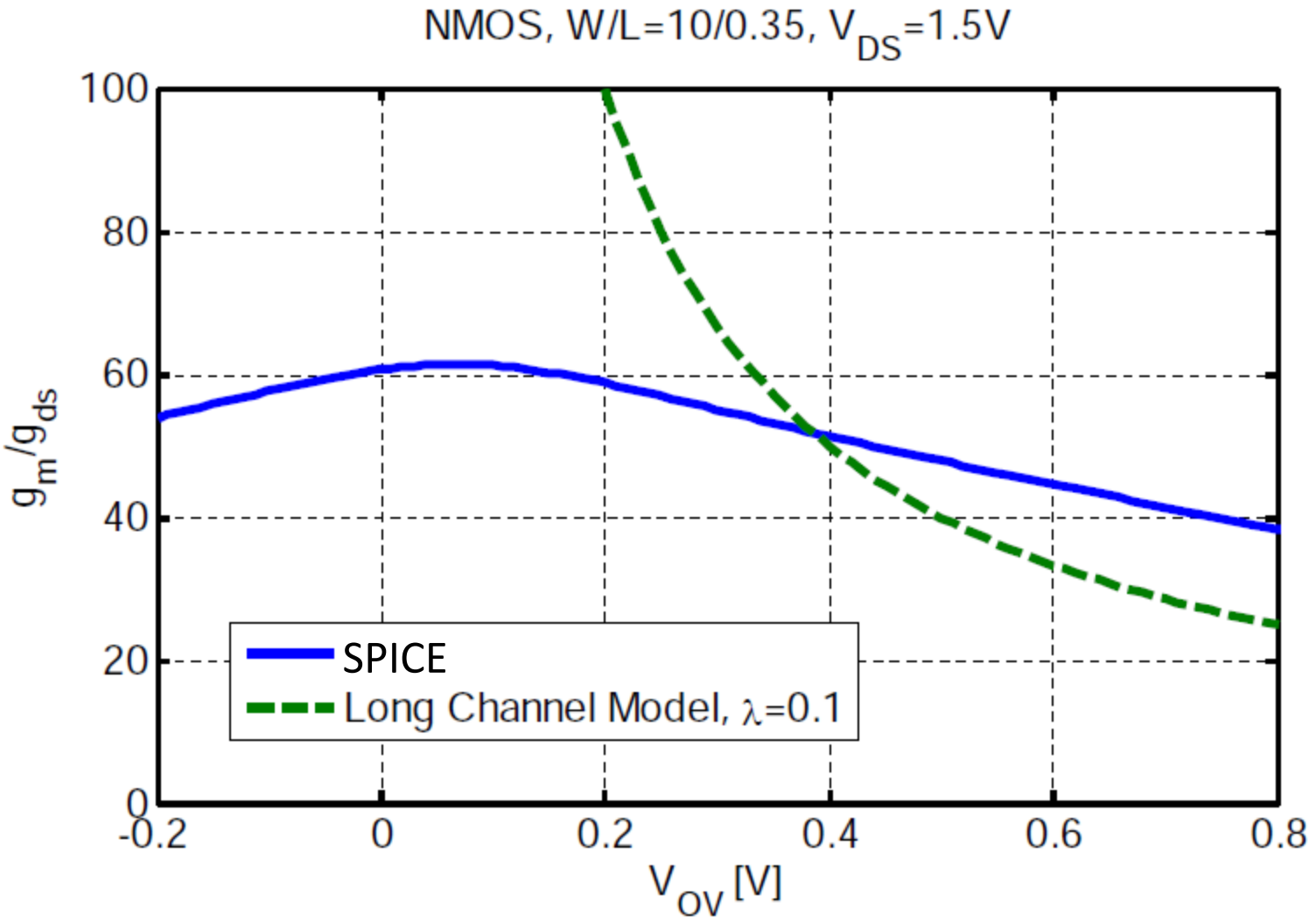
Subthreshold Operation

CAUTION

Exponential IV characteristics

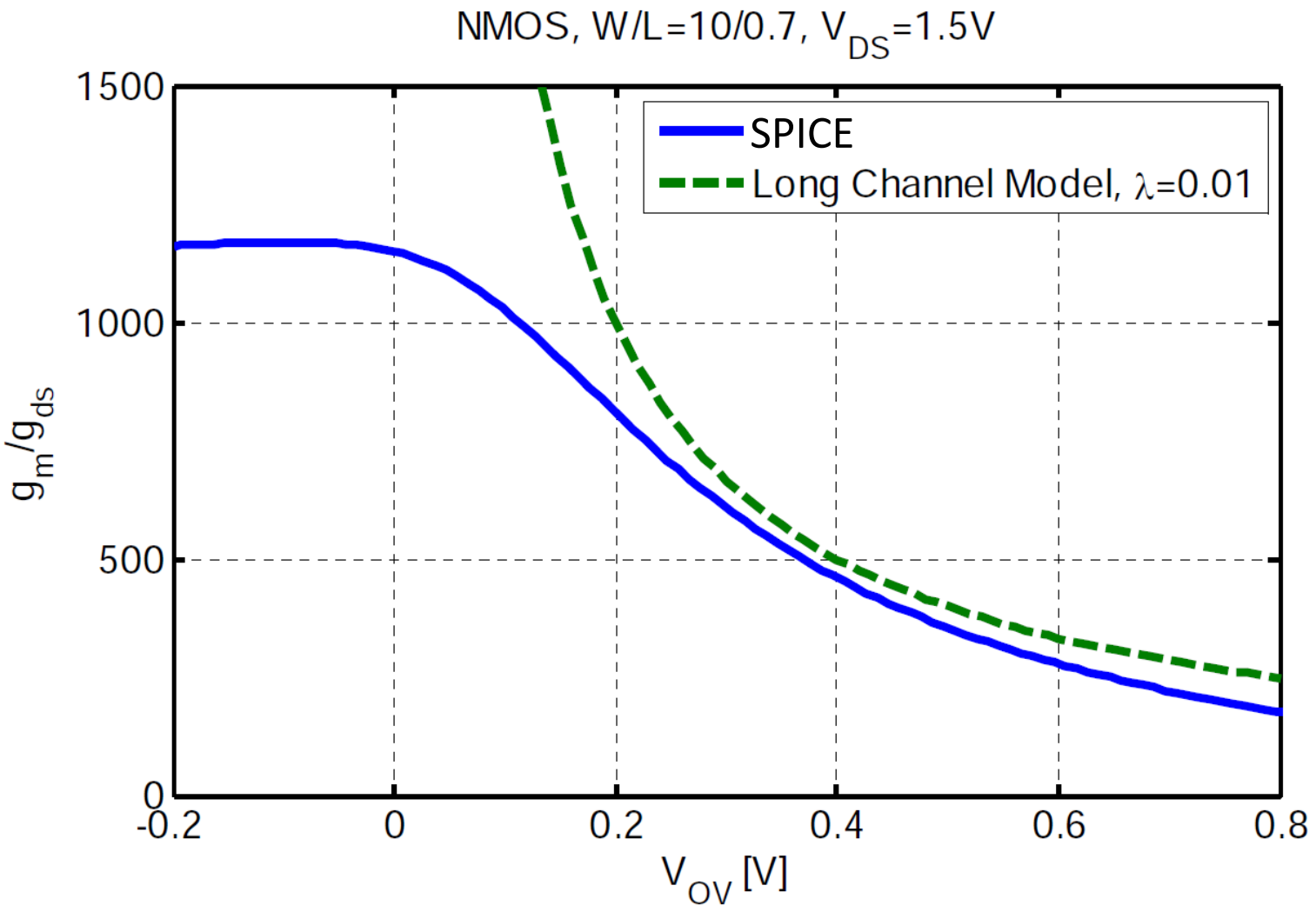
Very sensitive to PVT variations

SPICE vs Square Law



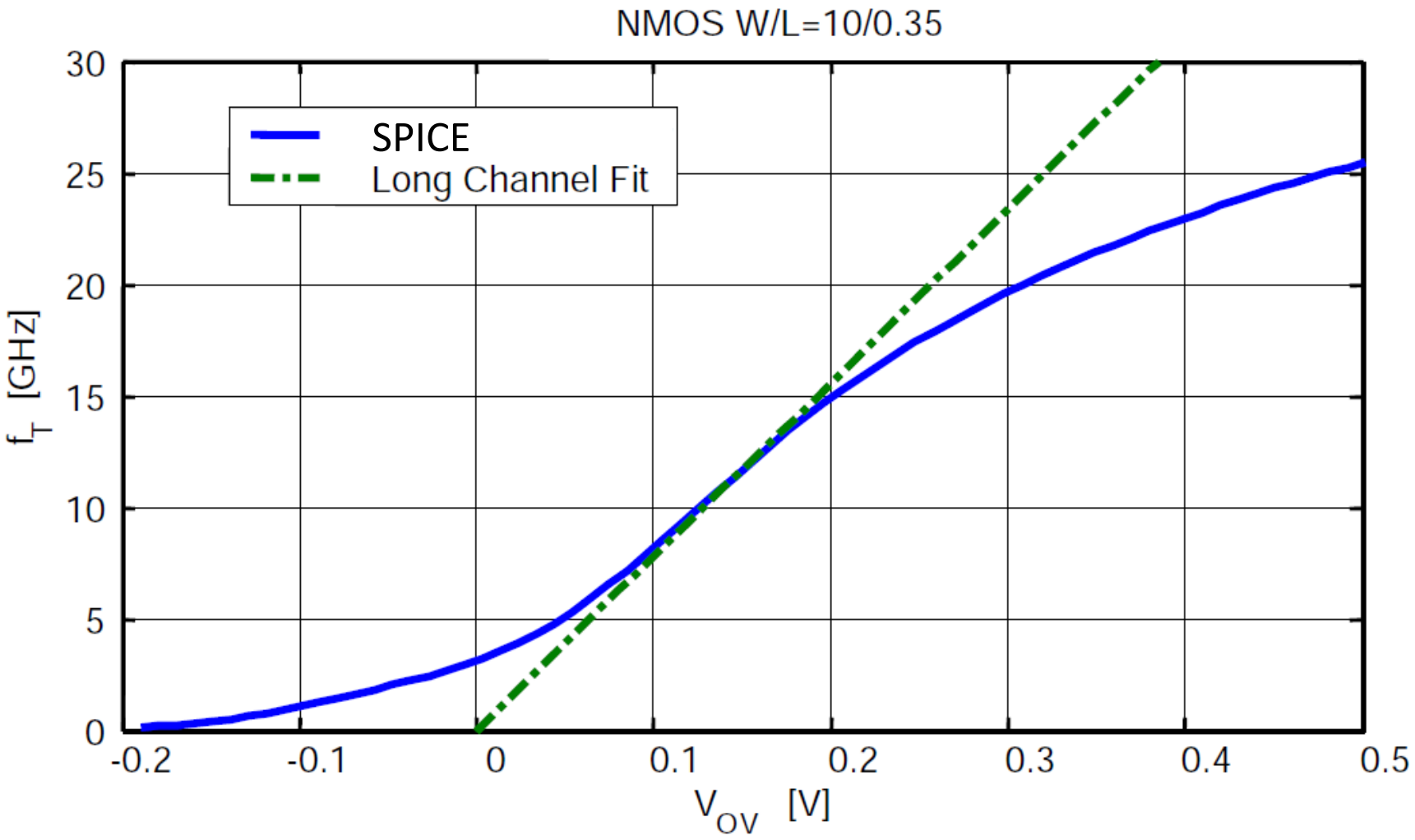
[Murmann, EE214B, Stanford]

SPICE vs Square Law



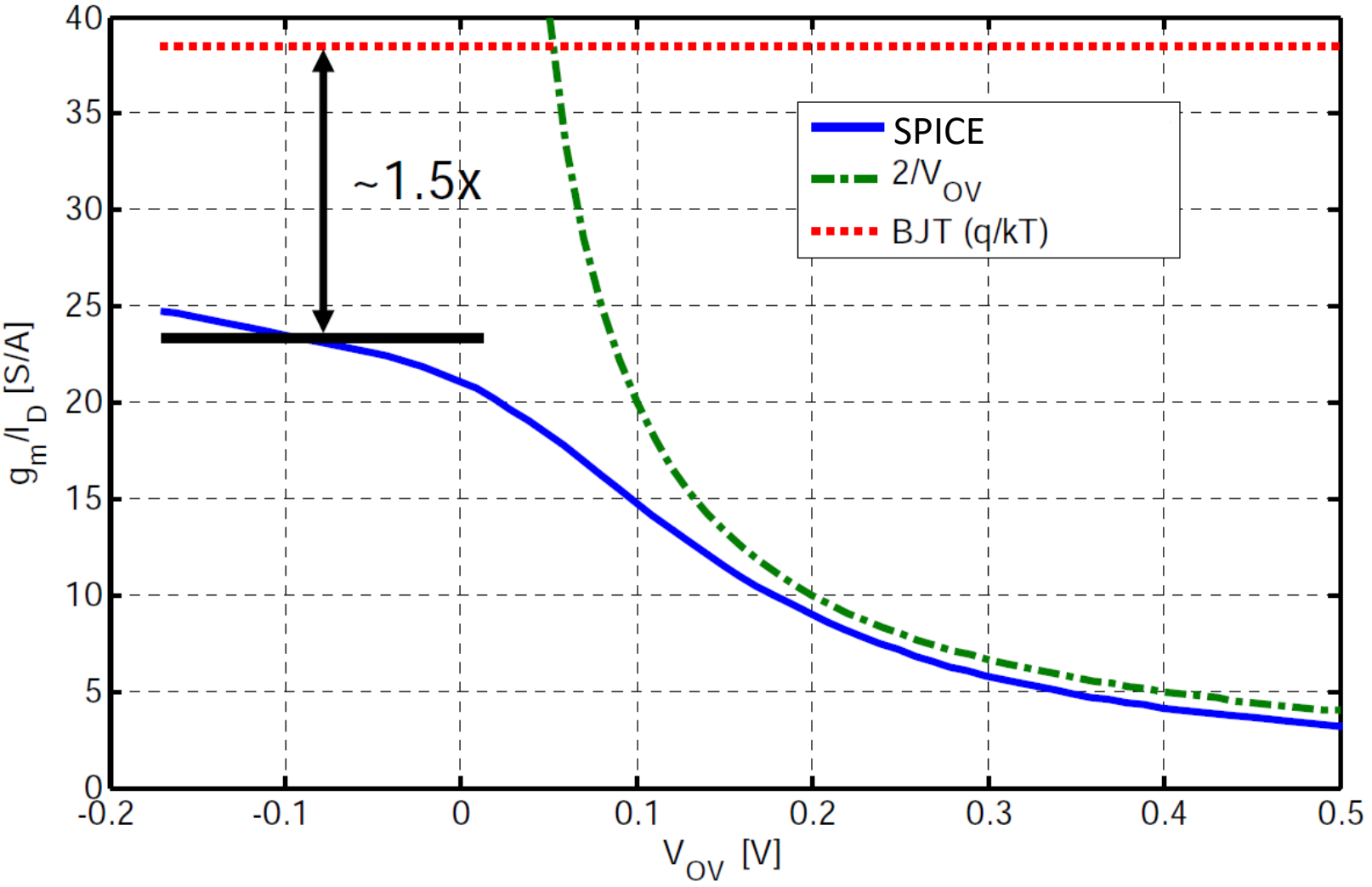
[Murmann, EE214B, Stanford]

SPICE vs Square Law



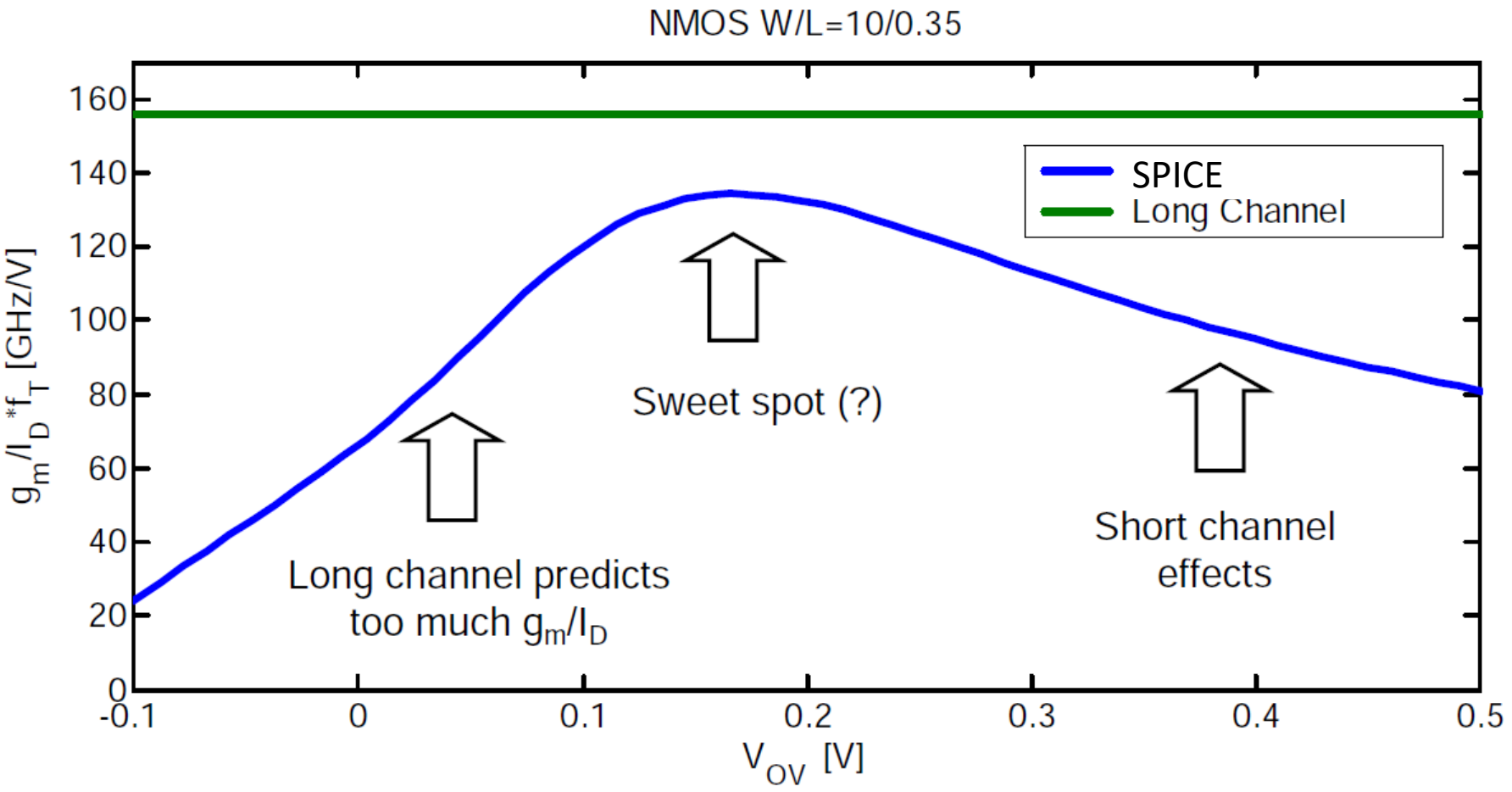
[Murmann, EE214B, Stanford]

SPICE vs Square Law



[Murmann, EE214B, Stanford]

SPICE vs Square Law



[Murmann, EE214B, Stanford]

Don’t be a SPICE Monkey!

- In absence of a clear methodology for hand analysis, many designers tend to converge toward a “SPICE monkey” design methodology
  - No hand calculations, iterate in SPICE until the circuit “somehow” meets the specifications
  - Typically results in sub-optimal designs, uninformed design decisions, etc.
- A SPICE monkey is someone who does not use hand analysis to figure out how to design a circuit, but rather plugs stuff into SPICE and uses whatever value works

gm/ID Design Methodology

- Traditionally, square-law was used in hand analysis to obtain initial design point
  - But short channel devices in recent technologies do not obey the square law
  - Square law is seldom used in nowadays designs
- The popular approach nowadays is using gm/Id (or equivalently V*) design methodology
  - V* is not equal to V_{ov} (also known as V_{eff}) except for a long channel devices at strong inversion
- Perform DC sweeps for both PMOS and NMOS to generate design charts vs gm/Id (or V*)
  - Use these charts to design your circuit to meet required specs

Design Tradeoffs

- There are always tradeoffs between gain, speed, and energy efficiency
- The design knobs that you use to control the tradeoffs are W and L
- Finding the best compromise for design tradeoffs given required specs is your job as a designer
- For practical designs, other factors like matching, signal-swing, and noise are also very important

OTA Design Example

- We are allowed to use ideal external 10uA DC current source

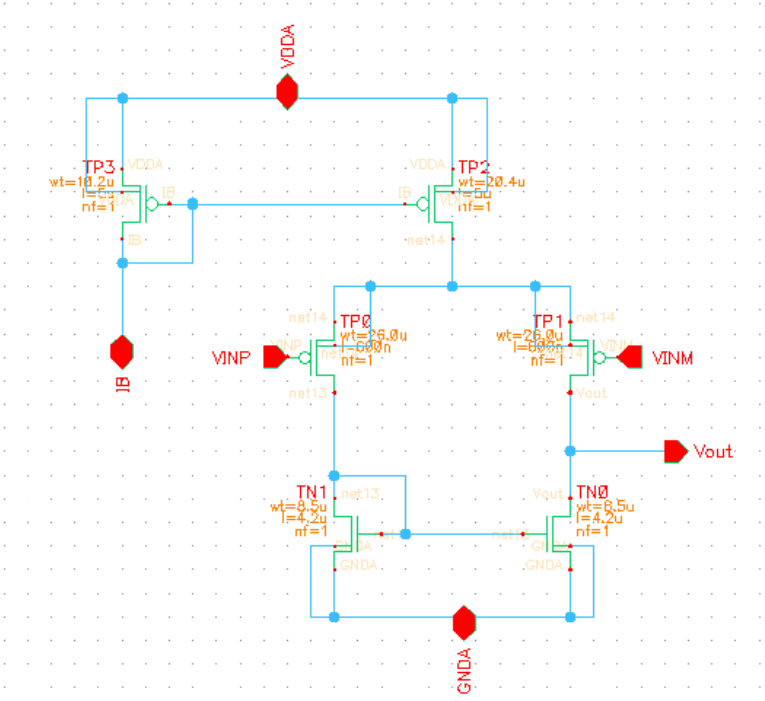
Technology	0.18um CMOS
Supply voltage	1.8V
Load	5pF
Cgg	<= 100fF
Open loop DC voltage gain	>= 40dB
Phase margin	>= 70^{o}
Total current consumption	<= 30uA
CM input range – low	<= 0.2V
CM input range – high	>= 1.1V
GBW	5MHz

Architecture Selection

- The required gain is not high (only 40dB=100) so it can be achieved by a simple single stage OTA
  - If the gain is high, we must use cascode or two stage OTA
- Since the required CMIR is close to the ground rail, we need to use PMOS input stage
- PMOS input stage has other advantages as well
  - PMOS usually has low flicker noise
  - PMOS input transistors can be placed in a separate well so they don't suffer from body effect

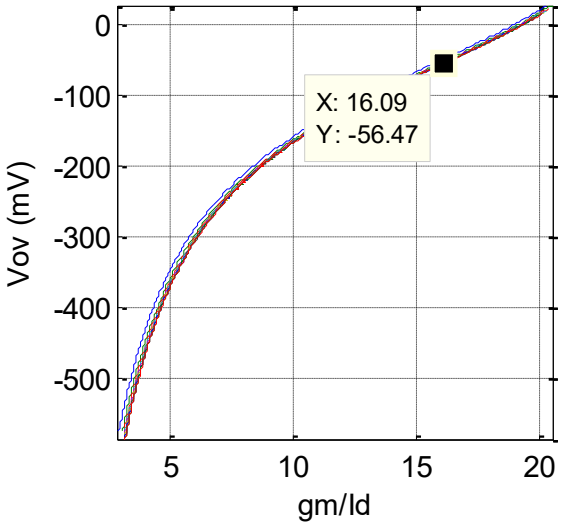
Architecture Selection

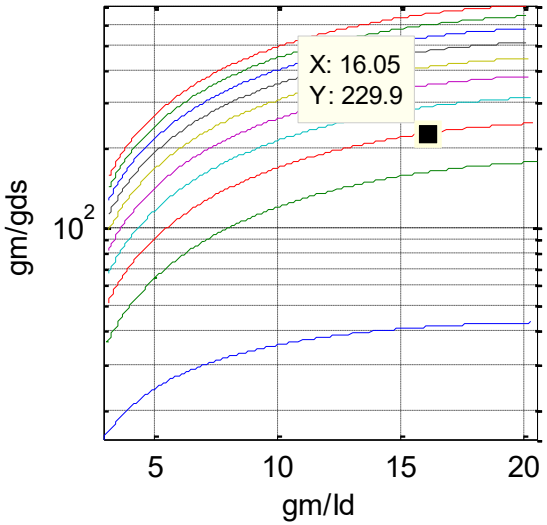
- Single stage is unconditionally sable, so we don't need to worry about PM
  - For two-stage you must use compensation network
- We use simple current mirror for biasing
  - The input current is 10uA, so all the remaining current (30-10=20uA) will go to our diff pair

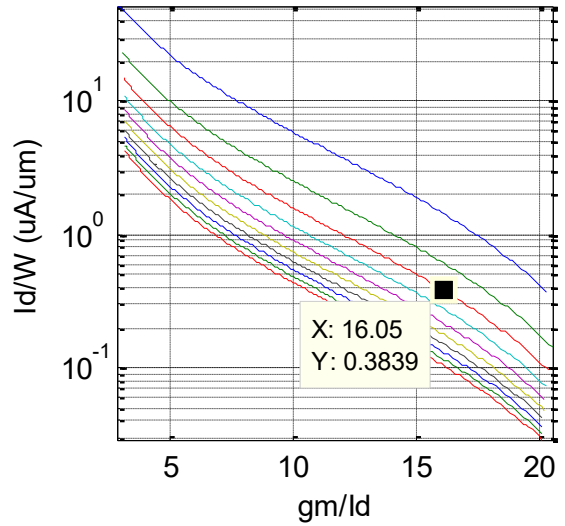


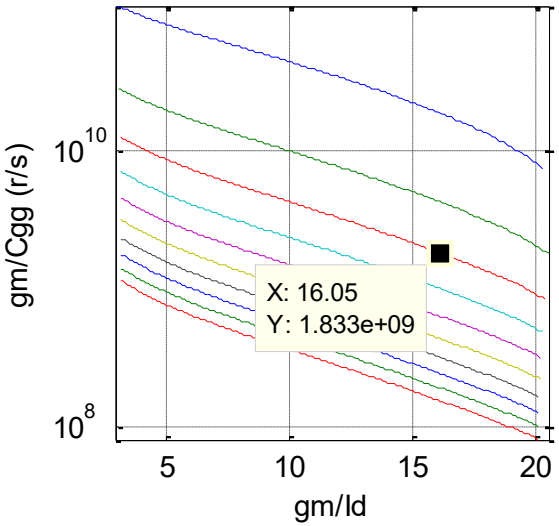
PMOS Design Charts

- L = 0.2um:0.2um:2um, VDS = 1V









PMOS Input Stage

- $GBW = \frac{g_m}{2\pi C_L}$
- $g_m = 2\pi \times 5p \times 5M \approx 160\mu S$
- $I_d = \frac{20\mu}{2} = 10\mu A$
- $V^* = \frac{2I_d}{g_m} = \frac{20\mu}{160\mu} = 125mV$
- $\frac{g_m}{I_d} = 16$
- From the design charts we find that

$$V_{ov} \approx 56mV$$

PMOS Input Stage

- Next, we need to find the channel length to get the required gain
- We assume PMOS and NMOS have same r_o

$$A_v = \frac{g_m r_o}{2} \rightarrow \frac{g_m}{g_{ds}} > 200$$

- From the design chart, we find that required length is **L=0.6um**
- We can also calculate $r_o \approx \frac{230}{g_m} \approx 1.44M\Omega$ (we will use this later when we design the NMOS load)
- Going again to the chart we find that

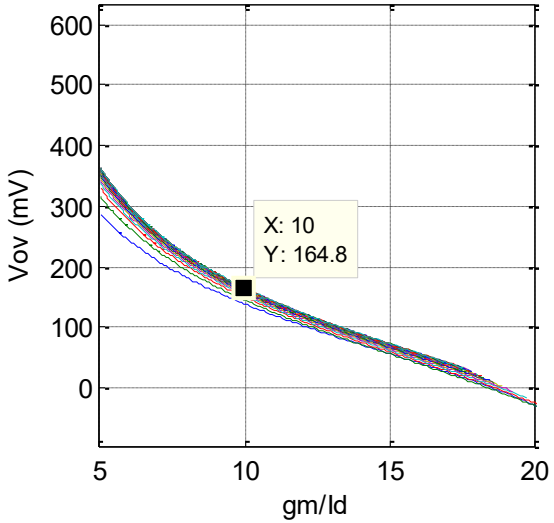
$$\frac{I_d}{W} = 0.384 \rightarrow W = \frac{10}{0.384} \approx 26\mu m$$

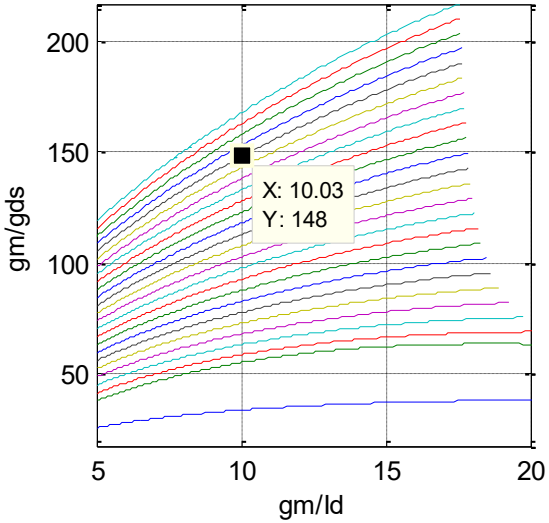
- Back to the chart to check the capacitance we find that

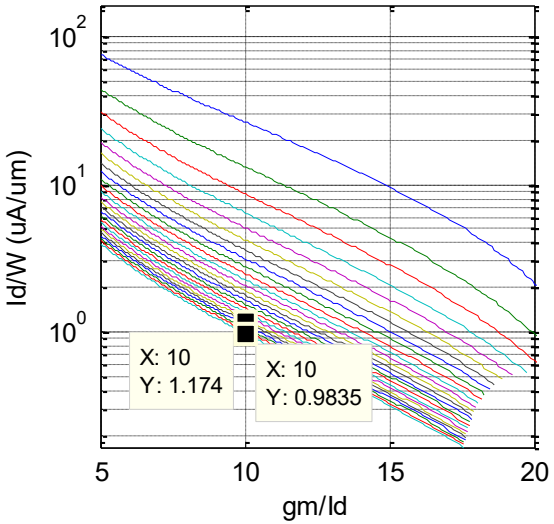
$$\frac{g_m}{C_{gg}} = 1.83e9 \rightarrow C_{gg} \approx 87fF$$

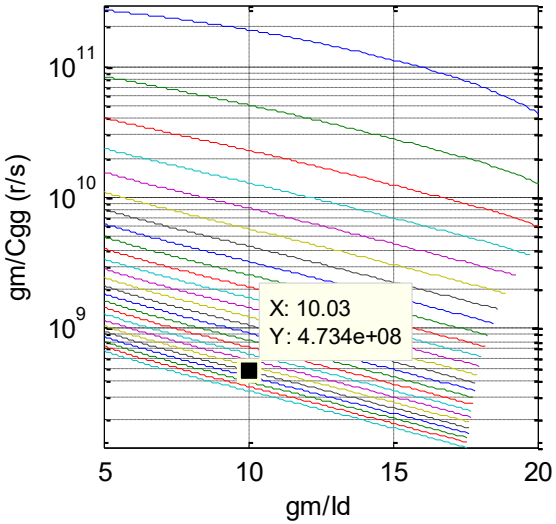
NMOS Design Charts

- L = 0.2um:0.2um:5um, VDS = 1V









NMOS Current Mirror Load

- The design of current mirror load is determined by noise, CMIR, and output swing specs
- Given our CMIR requirements (CMIR-low < 0.2V) and since there no noise spec, we can go for V*=200mV (gm/Id=10) since this is usually a good compromise

$$V^* = \frac{2I_d}{g_m} \rightarrow \frac{g_m}{I_d} = 10$$

$$g_m = 100\mu S$$

NMOS Current Mirror Load

- To select channel length we use gain spec to determine gm/gds

$$r_o = \frac{1}{g_{ds}} = 1.44M$$

$$\frac{g_m}{g_{ds}} = 100\mu \times 1.44M = 144$$

- To meet this requirement we select **L=4.2um** as shown in the NMOS design charts
  - We find this corresponds to $V_{ov} \approx 165mV$
- Going again to the chart we find that

$$\frac{I_d}{W} = 1.174 \rightarrow W = \frac{10}{1.174} \approx 8.5\mu m$$

Tail Current Source

- The design considerations here are as follows:
  - Long channel length means large output resistance: good CMRR @DC and good mirroring

$$A_{vCM} = \frac{V_{out}}{V_{iCM}} \approx -\frac{1}{2g_{m3,4}R_{SS}}$$

$$CMRR \approx g_{m1,2}(r_{o2}//r_{o4}) \cdot 2g_{m3,4}R_{SS}$$

  - But long L → large W → large parasitic cap: CMRR and power supply rejection ratio (PSRR) degrades at high frequencies
  - Large V* means better mirroring and better matching
  - But large V* means smaller swing

Tail Current Source

- Given no strict specs, we may go for V*=200mV (gm/Id=10) which also meets our swing requirements (CMIR-high > 1.1V), noting that V_T is ~0.4V (V_T is a function of W and L)
- Note that the tail current source has double the current

Notes

- The design charts (especially gm/gds) depend on V_{DS}
  - We neglected this dependence for simplicity
  - But practically you need to add some margin to your specs (e.g., 20%) to account for this
  - Or use multiple charts at multiple V_{DS}
- The total load capacitance is composed of external load capacitance and parasitic capacitance from the OTA itself
  - We did not consider parasitic cap in this tutorial for simplicity
  - But practically it should be considered by adding some margin to GBW or increasing effective load cap (or re-iterate)
- We did not consider process variations, mismatch, and noise

Thank you!