

# **EE214B**

## **Advanced Analog Integrated Circuit Design**

- Winter 2016 -

Boris Murmann  
Stanford University  
[murmann@stanford.edu](mailto:murmann@stanford.edu)

### **Table of Contents**

<b>Chapter 1</b>	<b>Introduction</b>
<b>Chapter 2</b>	<b>Bipolar Junction Transistors</b>
<b>Chapter 3</b>	<b>Elementary BJT Circuits</b>
<b>Chapter 4</b>	<b>MOS Transistor Modeling</b>
<b>Chapter 5</b>	<b><math>g_m/I_D</math>-Based Design</b>
<b>Chapter 6</b>	<b>Electronic Noise</b>
<b>Chapter 7</b>	<b>Introduction to Feedback</b>
<b>Chapter 8</b>	<b>Feedback TIA Design</b>
<b>Chapter 9</b>	<b>Root Locus and Frequency Compensation</b>
<b>Chapter 10</b>	<b>Fully Differential Amplifiers</b>
<b>Chapter 11</b>	<b>OTA Design Examples</b>
<b>Chapter 12</b>	<b>Wideband Amplifiers Using Local Feedback</b>
<b>Chapter 13</b>	<b>Mismatch</b>

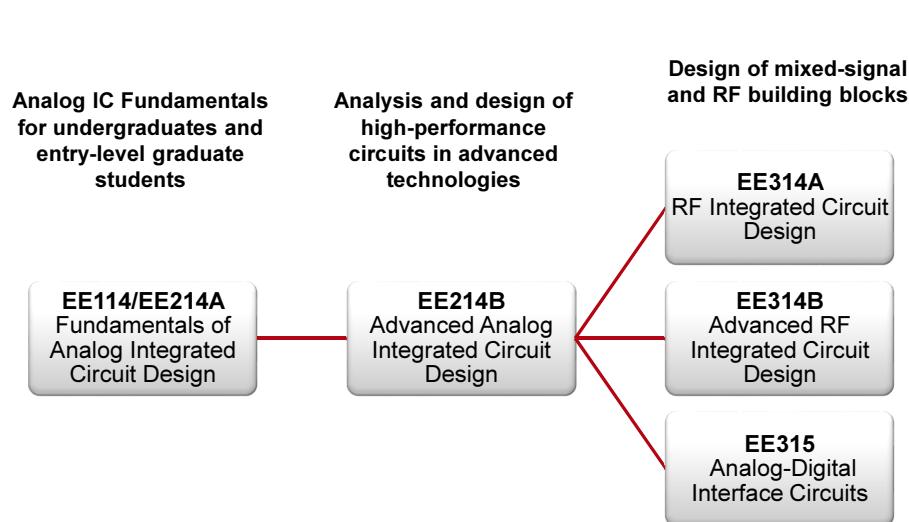
(This page is intentionally left blank)

# Chapter 1

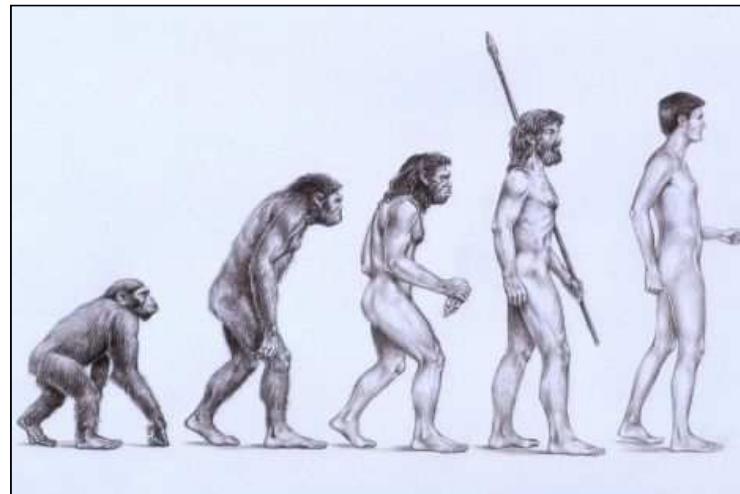
## Introduction

Boris Murmann  
Stanford University  
Winter 2015-16

### Analog Circuit Sequence



## The Evolution of a Circuit Designer



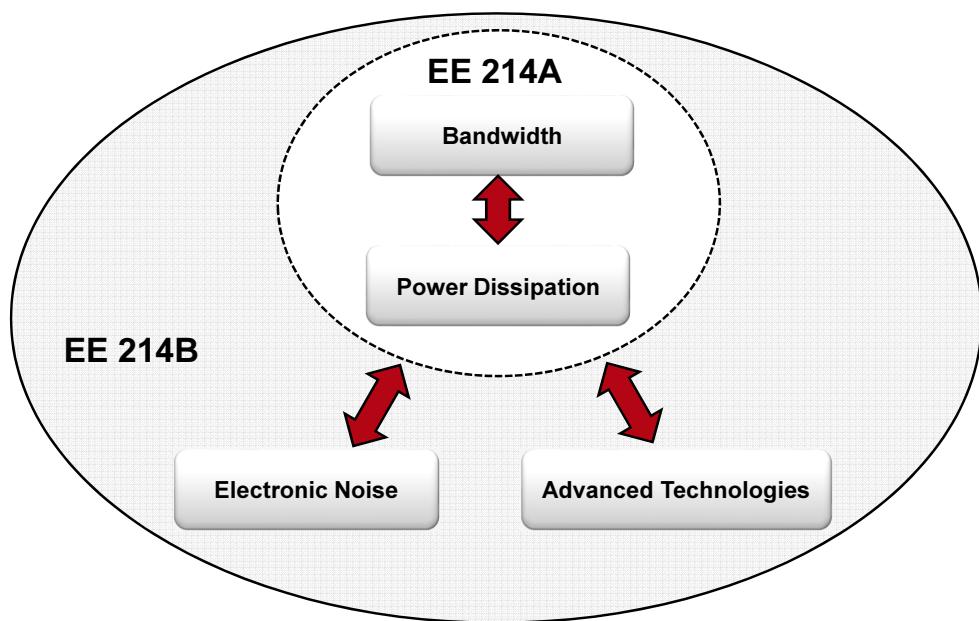
EE101A,B

EE114/  
EE214A

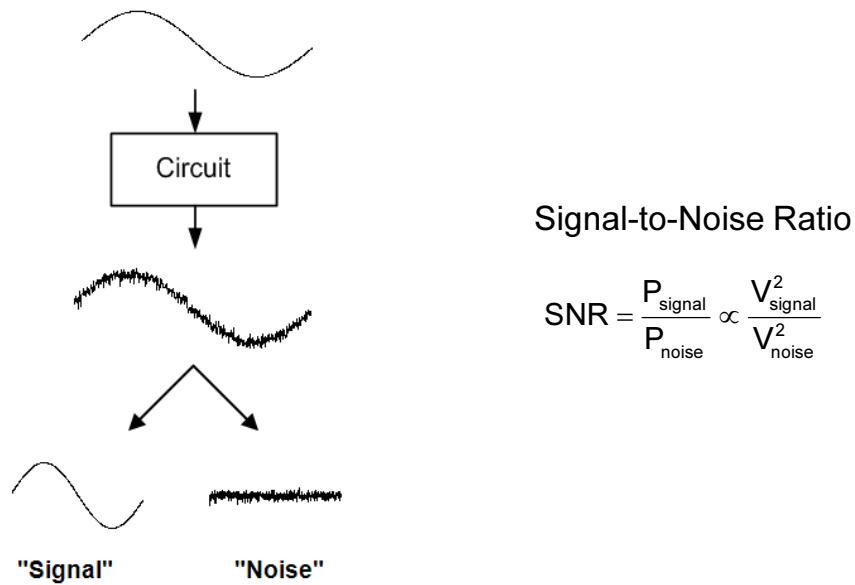
EE214B

EE314A,B  
EE315

## Key Elements of This Course



## Significance of Electronic Noise (1)



## Significance of Electronic Noise (2)

Example: Noisy image



<http://www.soe.ucsc.edu/~htakeda/kernelreg/kernelreg.htm>

## Significance of Electronic Noise (3)

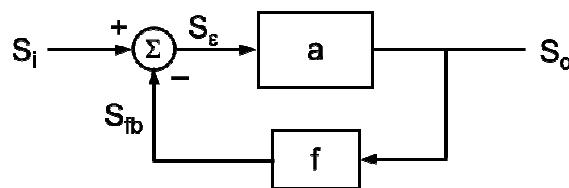
- The "fidelity" of electronic systems is often determined by their SNR
  - Examples
    - Audio systems
    - Imagers, cameras
    - Wireless and wireline transceivers
- Electronic noise directly trades with power dissipation and speed
  - In most circuits, low noise dictates large capacitors (and/or small R, large  $g_m$ ), which means high power dissipation
- Noise has become increasingly important in modern technologies with reduced supply voltages
  - $\text{SNR} \sim V_{\text{signal}}^2/V_{\text{noise}}^2 \sim (\alpha V_{\text{DD}})^2/V_{\text{noise}}^2$
- Designing a low-power, high-SNR circuit requires good understanding of electronic noise

## Noise Analysis in EE214B

- Main objective
  - Acquire the **basic** tools and intuition needed to analyze electronic noise in integrated circuits
  - Look at a few specific circuit examples to “get a feel” for situations where electronic noise may matter
- Leave application-specific examples for later
  - EE314A/B: Noise in LNAs, mixers, etc.
  - EE315: Noise in filters, sensor interfaces, samplers, A/D converters

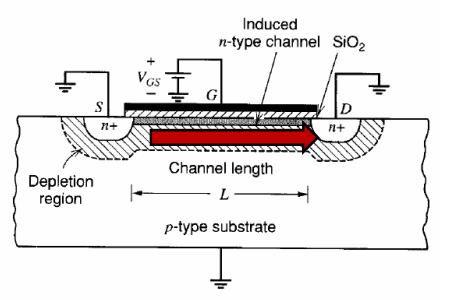
## Feedback Circuit Analysis in EE214B

- Builds on what you have already learned in EE114/214A
- Goals for this quarter
  - Reinforce your understanding of two-port and return ratio analysis
  - Learn about advanced frequency compensation techniques
  - Investigate the effect of feedback on noise and distortion



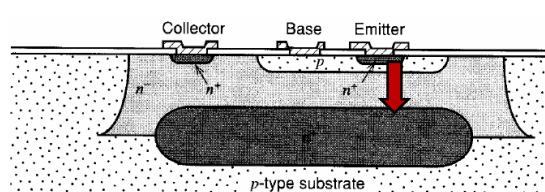
## Technology

MOSFET



EE214A

Bipolar Junction Transistor (BJT)

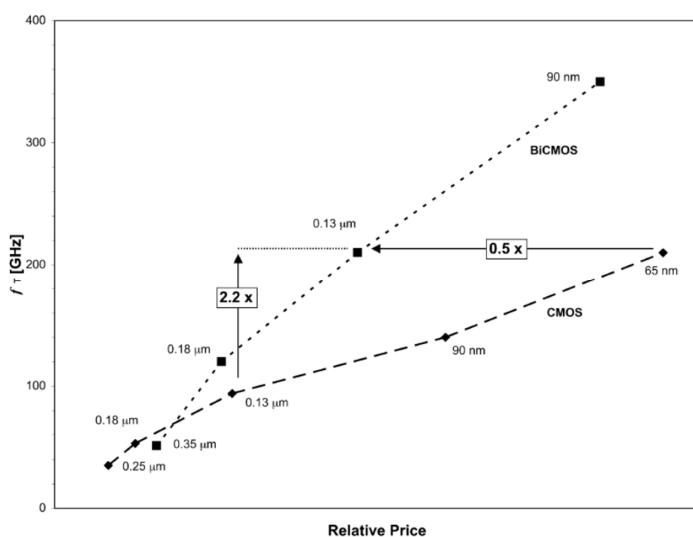


EE214B

## Bipolar vs. CMOS (1)

- Advantages of bipolar transistors
  - Higher  $f_T$  for a given feature size/lithography
  - Higher supply voltages
  - Higher intrinsic gain ( $g_m r_o$ )
  - Lower parametric variance
- Disadvantages of bipolar transistors
  - Lower integration density, larger features
  - Higher cost (due to higher fabrication process complexity)
  - Poor ohmic switch (as required for switched capacitor circuits)

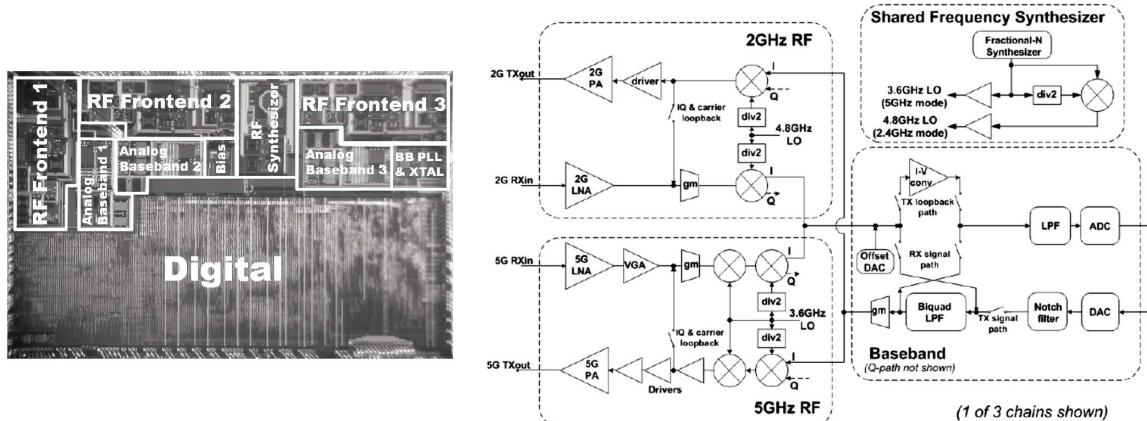
## Bipolar vs. CMOS (2)



A.J. Joseph, et al., "Status and Direction of Communication Technologies - SiGe BiCMOS and RFCMOS," Proceedings of the IEEE, vol. 93, no. 9, pp.1539-1558, September 2005.

- CMOS tends to require finer lithography to achieve same speed as BiCMOS process with advanced BJT

## Example that Leverages Densely Integrated CMOS: RF Transceiver System-on-a-Chip (SoC)



S. Abdollahi-Alibeik et al., "A 65nm dual-Band 3-Stream 802.11n MiMo WLAN SoC," ISSCC 2011

- In modern CMOS technology, millions of logic gates can be integrated together with moderate- to high-performance analog/RF blocks

B. Murmann

EE214B Winter 2015-16 – Chapter 1

13

## Example that Leverages High-Speed BJTs: Long Range Automotive Radar



	Freq.	BW	Modulation	Angle	Range	Resolution	Application
Short Range	24GHz	7GHz	Pulsed	70°	10m	< 10cm	Side-Crash Parking
Mid Range	24GHz	250MHz	FMCW	30°~60°	40m	~ 1m	Stop & Go
Long Range	77GHz	1GHz	FMCW	16°	150m	~ 1m	ACC

H.P. Forstner et al., A 77GHz 4-Channel Automotive Radar Transceiver in SiGe, RFIC 2008.

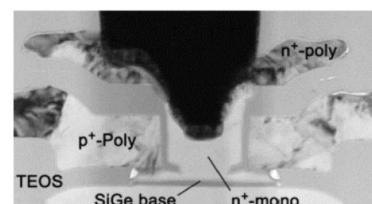
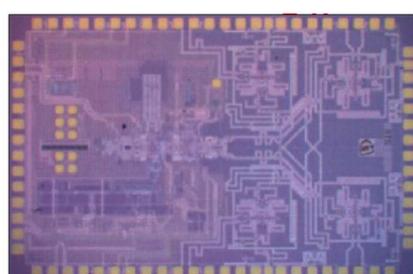
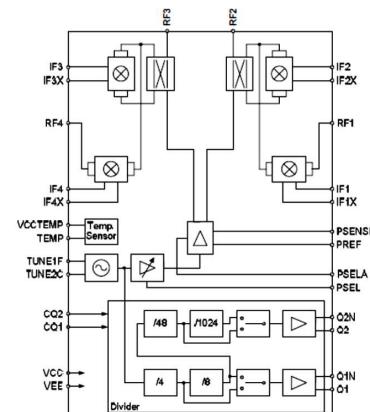


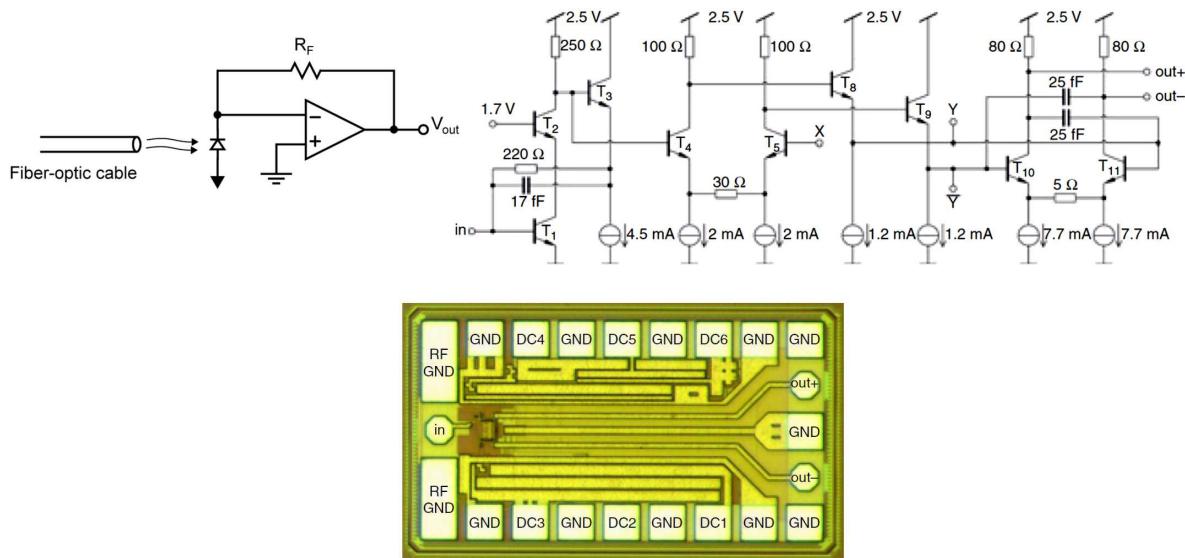
Fig. 1. TEM cross section of the emitter-base complex of a transistor with effective emitter width of 0.18 um

B. Murmann

EE214B Winter 2015-16 – Chapter 1

14

## Example that Leverages High-Speed BJTs: Transimpedance Amplifier for Optical Communications



C. Knochenhauer et al., “40 Gbit/s transimpedance amplifier with high linearity range in 0.13 μm SiGe BiCMOS,” Electronics Letters, May 12, 2012.

## Summary of Learning Goals

- Understand device behavior and models for transistors available in advanced integrated circuit technologies
  - SiGe BJT, short channel MOS
- Acquire the basic intuition and models for
  - Noise analysis
  - Feedback circuit analysis
  - Frequency compensation and related design techniques for broadband amplification
- Solidify the above topics in a hands-on project involving the design and optimization of a broadband amplifier circuit

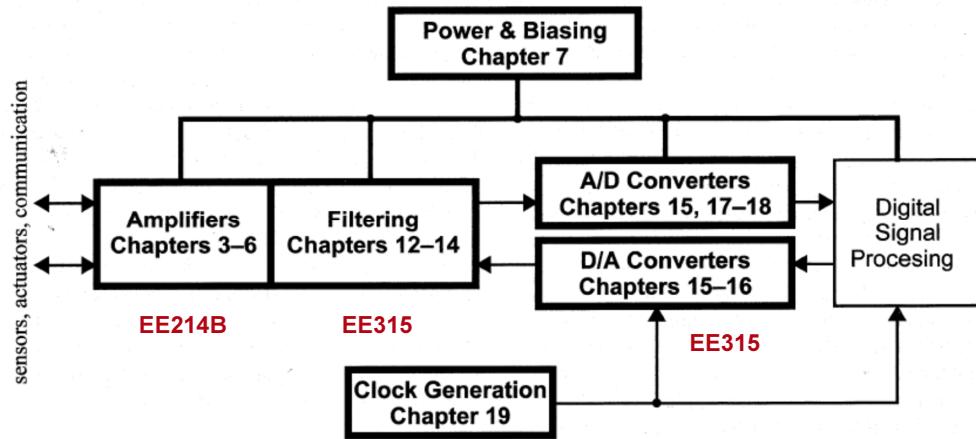
## Staff and Website

- Instructor
  - Boris Murmann
- Teaching assistant
  - Danny Bankman
- Administrative Assistant
  - Ann Guerra, Allen-207
- Lectures are available online
  - But please come to class to keep the discussion interactive!
- Web page
  - <http://coursework.stanford.edu/homepage/W16/W16-EE-214B-01.html>
- Online forum
  - <http://piazza.com/stanford/winter2016/ee214b>

## Text and Prerequisites

- Required textbook
  - Chan Carusone, Johns, Martin, *Analog Integrated Circuit Design*, 2<sup>nd</sup> Edition, Wiley, 2011
  - Errata: <http://analogicdesign.com/students/errata/>
- Reference texts
  - Gray, Hurst, Lewis and Meyer, *Analysis and Design of Analog Integrated Circuits*, 5<sup>th</sup> Edition, Wiley, 2008
  - B. Razavi, *Design of Integrated Circuits for Optical Communications*, 2<sup>nd</sup> Edition, McGraw-Hill, 2012
- Course prerequisite: EE114/214A or equivalent
  - Basic device physics and models
  - Frequency response, dominant pole approximation, ZVTC
  - Biasing, small-signal models
  - Common source, common gate, and common drain stages
  - Port impedance calculations
  - Feedback basics, two-port and return ratio analysis

## Textbook Content Overview



## Tools and Technology

- Simulation
  - HSpice + circuit netlists
  - Cscope and/or Matlab for post-processing
  - You can use your own tools/setups “at own risk”
- Getting started
  - Read “CAD basics” handout provided on the course website
- EE214B Technology
  - 0.18- $\mu$ m SiGe BiCMOS
  - BSIM3v3 models provided under /usr/class/ee214b/hspice
  - Models correspond to a typical technology as described in
    - Wada, et al., “A manufacturable 0.18- $\mu$ m SiGe BiCMOS technology for 40-Gb/s optical communication LSIs,” BCTM 2002

## Assignments

- Homework (20%)
  - Handed out on Wednesdays, due following Wednesdays in class
  - Lowest HW score will be dropped
  - Late submission penalty: 0.5 dB/hour
  - Policy for off-campus students
    - Email to SCPD before deadline stated on handout
- Midterm Exam (30%)
- Design Project (20%)
  - Design of an amplifier using HSpice (no layout)
  - Work in teams of two
  - OK to discuss your work with other teams, but no file exchange!
  - Late submission penalty: 0.5 dB/hour
- Final Exam (30%)

## Honor Code

- Please remember that you are bound by the honor code
  - We will trust you not to cheat
  - We will try not to tempt you
- But if you are found cheating it is very serious
  - There is a formal hearing
  - You can be thrown out of Stanford
- Save yourself a huge hassle and be honest
- For more info
  - <http://www.stanford.edu/dept/vpsa/judicialaffairs/guiding/pdf/honorcode.pdf>

## Course Outline

- Bipolar Junction Transistors
- Elementary BJT Circuits
- MOSFET Modeling and  $g_m/I_D$  Based Design
- Noise Analysis
- Feedback Circuits
  - Review of analysis tools
  - Frequency compensation techniques
  - Feedback TIA design
  - Fully differential amplifiers
  - OTA design

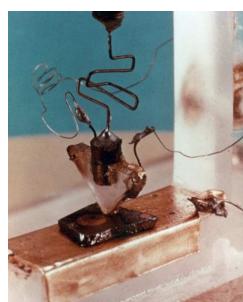
# Chapter 2

## Bipolar Junction Transistors

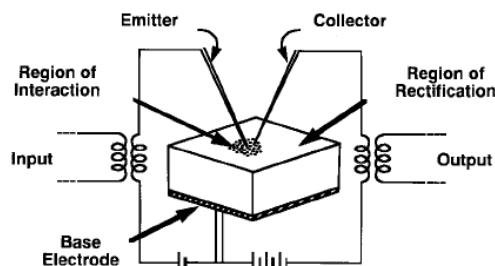
Boris Murmann  
Stanford University  
Winter 2015-16

Textbook Sections: 8.1–8.4, 8.6

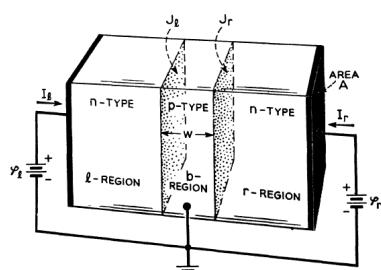
### History



Bardeen, Brattain, and Shockley, 1947

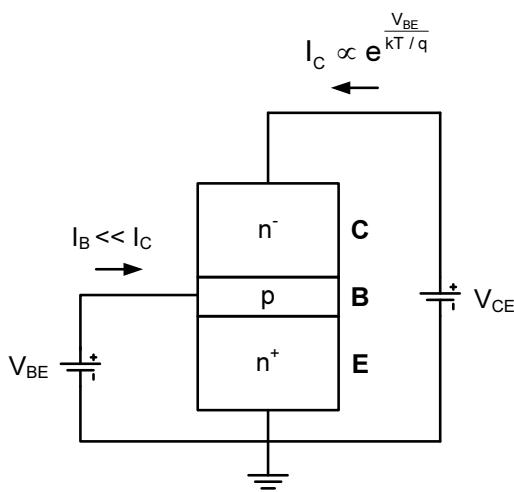


W. Brinkman, D. Hagan, and W. Troutman, "A history of the invention of the transistor and where it will lead us," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1858-1865, Dec. 1997.



W. Shockley, M. Sparks, and G. K. Teal, "P-N junction transistors," *Phys. Rev.* 83, pp. 151–162, Jul. 1951.

## Conceptual View of an NPN Bipolar Transistor (Active Mode)

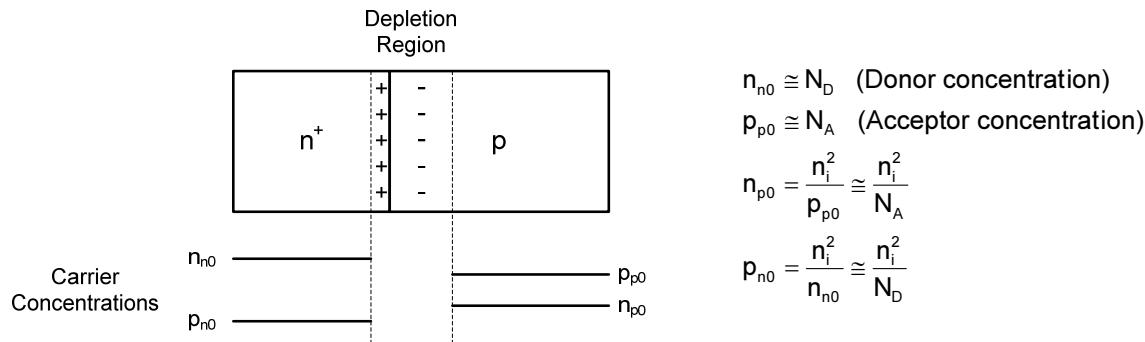


- Device acts as a voltage controlled current source
  - $V_{BE}$  controls  $I_C$
- The base-emitter junction is forward biased and the base-collector junction is reverse biased
- The device is built such that
  - The base region is very thin
  - The emitter doping is much higher than the base doping
  - The collector doping is much lower than the base doping

## Outline of Discussion

- In order to understand the operation principle of a BJT, we will look at
  - The properties of a forward biased  $p\text{n}^+$  junction
  - The properties of a reverse biased  $\text{p}\text{n}^-$  junction
  - And the idea of combining the two junctions such that they are joined by a very thin (p-type) base region
- The treatment in the following slides is meant to be short and qualitative
  - See any solid-state physics text for a more rigorous treatment (involving band diagrams, etc.)

## pn<sup>+</sup> Junction in Equilibrium (No Bias Applied)



$n_{n0} \approx N_D$  (Donor concentration)

$p_{p0} \approx N_A$  (Acceptor concentration)

$$n_{p0} = \frac{n_i^2}{p_{p0}} \approx \frac{n_i^2}{N_A}$$

$$p_{n0} = \frac{n_i^2}{n_{n0}} \approx \frac{n_i^2}{N_D}$$

$n_n$  Concentration of electrons on n side (majority carriers)

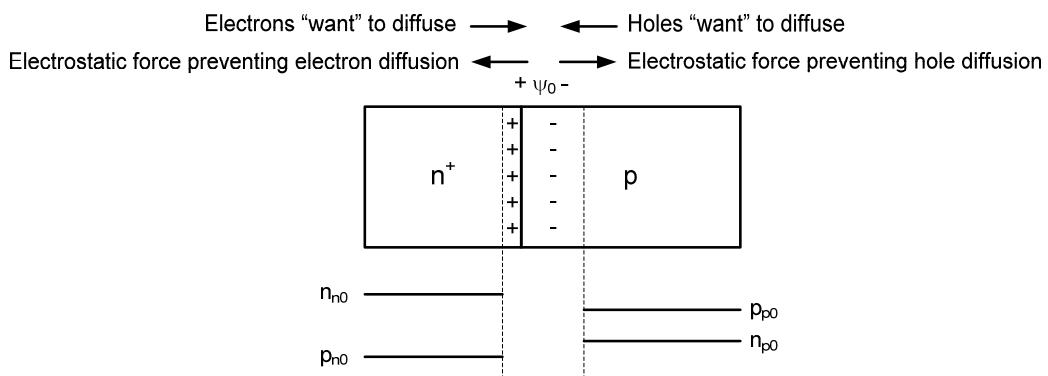
$p_n$  Concentration of holes on n side (minority carriers)

$n_p$  Concentration of electrons on p side (minority carriers)

$p_p$  Concentration of holes on p side (majority carriers)

The subscript "0" in the carrier concentrations denotes equilibrium (no bias applied)

## Built-in Potential

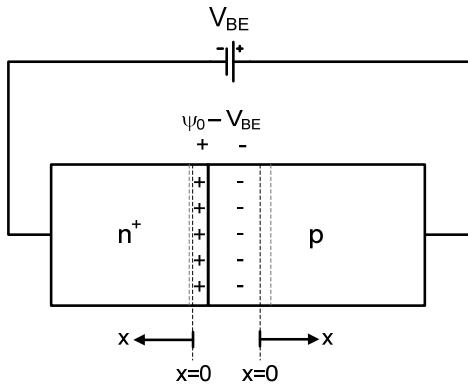


- The built in potential sets up an electric field that opposes the diffusion of mobile holes and electrons across the junction

$$(Drift) \quad q\mu_p pE = qD_p \frac{dp}{dx} \quad (Diffusion)$$

$$\Rightarrow \psi_0 = V_T \ln\left(\frac{p_{p0}}{p_{n0}}\right) = V_T \ln\left(\frac{n_{n0}}{n_{p0}}\right) \approx V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) \quad V_T = \frac{kT}{q}$$

## pn<sup>+</sup> Junction with Forward Bias (1)



- Depletion region narrows, diffusion processes are no longer balanced by electrostatic force
- At the edge of the depletion region ( $x=0$ ), the concentration of minority carriers [ $n_p(0)$ ] can be computed as follows

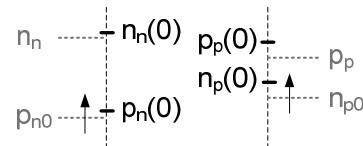
$$\psi_0 - V_{BE} = V_T \ln\left(\frac{n_n}{n_p(0)}\right) \approx V_T \ln\left(\frac{N_D}{n_p(0)}\right) \quad \therefore n_p(0) = \frac{N_D}{e^{\frac{\psi_0}{V_T}}} = n_{p0} e^{\frac{V_{BE}}{V_T}} \approx \frac{n_n^2}{N_A} e^{\frac{V_{BE}}{V_T}}$$

## pn<sup>+</sup> Junction with Forward Bias (2)

- The result on the previous slide shows that forward biasing increases the concentration of electrons at the “right” edge of the depletion region by a factor of  $\exp(V_{BE}/V_T)$
- The same holds for holes at the “left” edge of the depletion region

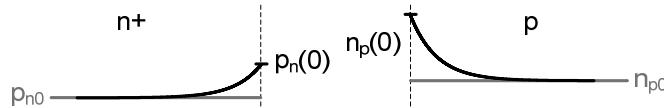
$$p_n(0) = p_{n0} \cdot e^{\frac{V_{BE}}{V_T}} \approx \frac{n_n^2}{N_D} \cdot e^{\frac{V_{BE}}{V_T}}$$

- Since  $N_D \gg N_A$ , it follows that  $p_n(0) \ll n_p(0)$ , i.e. the concentration of minority carriers is much larger at the lightly doped edge
- Since there must be charge neutrality in the regions outside the depletion region, the concentration of the majority carriers at the edge of the depletion region must also increase
  - However, this increase is negligible when  $n_p(0) \ll p_p \approx N_A$  (or  $p_n(0) \ll n_n \approx N_D$ )
  - These conditions are called “low-level injection”

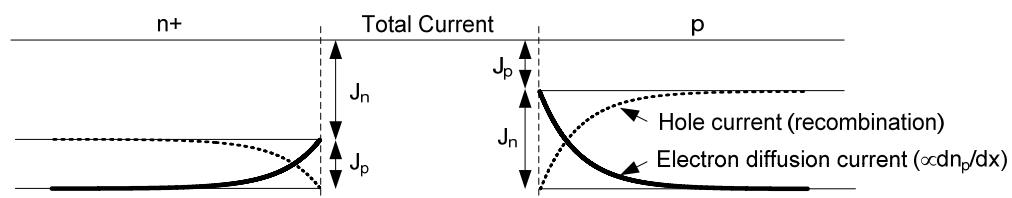


## What Happens with the Injected Minority Carriers?

- The carriers would “like” to diffuse further into the neutral regions, but quickly fall victim to recombination
- The number of minority carriers decays exponentially, and drops to 1/e of the at the so-called diffusion length ( $L_p$  or  $L_n$ , on the order of microns)



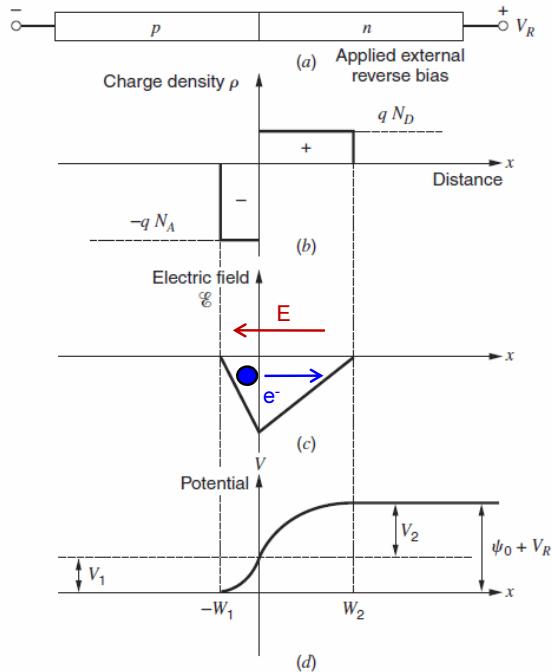
- In each region, there are now two types of currents
  - Diffusion of injected minority carriers due to non-zero  $d n_p / dx$  (or  $d p_n / dx$ )
  - Majority carrier currents for recombination



## Summary – Forward Biased pn+ junction

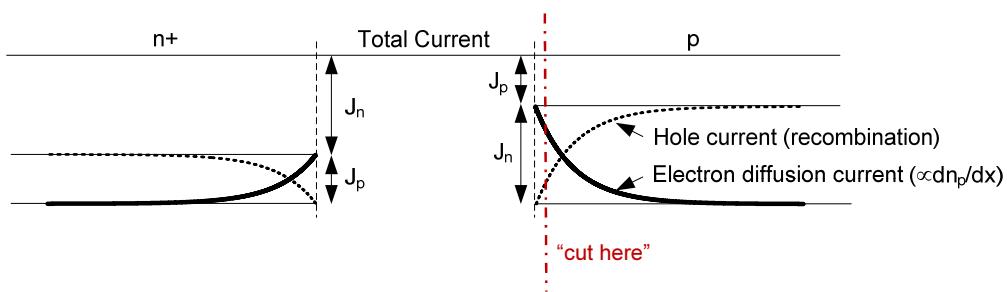
- Lots of electrons being injected into the p-region, not all that many holes get injected into the n<sup>+</sup> region
  - The heavier the n-side doping, the more pronounced this imbalance becomes
- The electrons injected in the p region cause a diffusion current that decays in the x-direction due to recombination
- The recombination necessitates a flow of holes to maintain charge neutrality; as the diffusion current decays, the hole current increases, yielding a constant current density along the device
- Near the edge of the depletion region, the electron diffusion current dominates over the hole current that supplies carriers for recombination
  - This is a very important aspect that we will come back to

## Reverse Biased pn<sup>-</sup> Junction



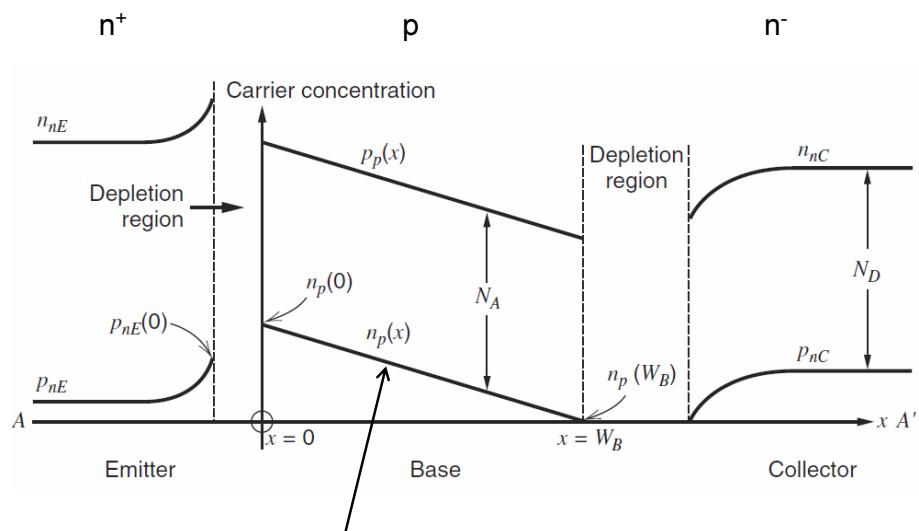
- Reverse bias increases the width of the depletion region and increases the electric field
- Depletion region extends mostly into n<sup>-</sup> side
- Any electron that would “somehow” make it into the depletion region will be swept through, into the n-region
  - Due to electric field

## Bipolar Junction Transistor – Main Idea



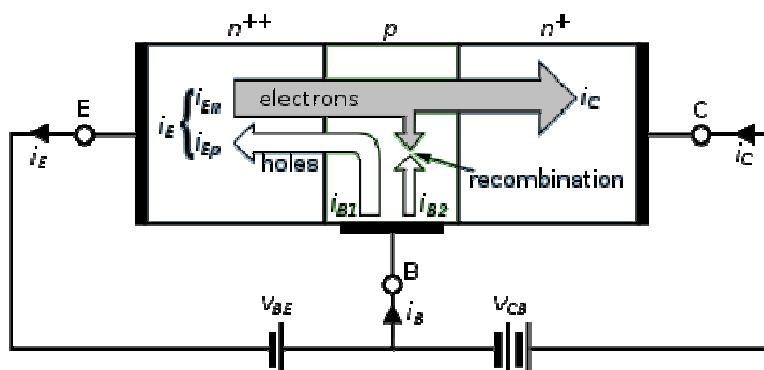
- Make the p-region of the  $\text{pn}^+$  junction very thin
- Attach an n<sup>-</sup> region that will “collect” and sweep across most of the electrons before there is a significant amount of recombination

## Complete Picture



Straight line because base is thin; negligible recombination  
("short base" electron profile)

## BJT Currents



[http://en.wikipedia.org/wiki/Bipolar\\_junction\\_transistor](http://en.wikipedia.org/wiki/Bipolar_junction_transistor)

- Primary current is due to electrons captured by the collector
- Two (undesired) base current components
  - Hole injection into emitter ( $\rightarrow 0$  for infinite emitter doping)
  - Recombination in the base ( $\rightarrow 0$  for base width approaching zero)

## First-Order Collector Current Expression

$$J_n = qD_n \frac{dn_p(x)}{dx} \approx -qD_n \frac{n_p(0)}{W_B}$$

Current density

$$I_C \approx qAD_n \frac{n_p(0)}{W_B}$$

A is the cross-sectional area  $W_B$   
is the base width

$$n_p(0) \approx \frac{n_i^2}{N_A} e^{\frac{V_{BE}}{V_T}}$$

Result from pn+ junction analysis

$$\therefore I_C \approx \frac{qAD_n n_i^2}{W_B N_A} e^{\frac{V_{BE}}{V_T}}$$

$$\boxed{\therefore I_C \approx I_S e^{\frac{V_{BE}}{V_T}}}$$

$$\boxed{I_S = \frac{qAD_n n_i^2}{W_B N_A}}$$

## Base Current

$$I_B = I_{B1} + I_{B2} \quad \text{where } I_{B1} = \text{Recombination in the base}$$

$$I_{B2} = \text{Injection into the emitter}$$

$I_{B1}$  follows from dividing the minority carrier charge in the base ( $Q_e$ ) by its "lifetime" ( $\tau_B$ )

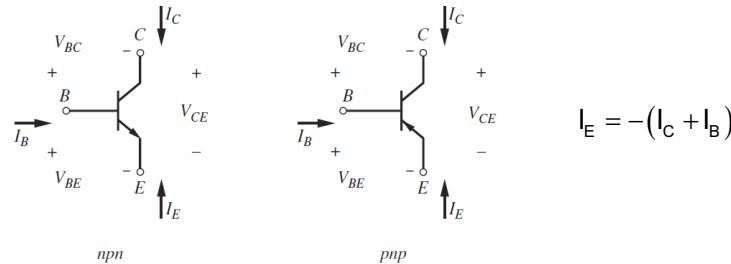
$$I_{B1} = \frac{Q_e}{\tau_b} = \frac{\frac{1}{2} n_p(0) W_B q A}{\tau_b} = \frac{1}{2} \frac{W_B q A n_i^2}{\tau_b N_A} e^{\frac{V_{BE}}{V_T}}$$

$I_{B2}$  depends on the gradient of minority carriers (holes) in the emitter. For a "long" emitter (all minority carriers recombine)

$$I_{B2} = -qAD_p \frac{dp_n(x)}{dx} \Big|_{x=0} = -qAD_p \left[ \frac{d}{dx} \left( \frac{n_i^2}{N_D} e^{\frac{V_{BE}}{V_T}} e^{-\frac{x}{L_p}} \right) \right]_{x=0} = \frac{qAD_p}{L_p} \frac{n_i^2}{N_D} e^{\frac{V_{BE}}{V_T}}$$

In modern narrow-base transistors  $I_{B2} \gg I_{B1}$ .

## Terminal Currents and Definition of $\alpha_F$ , $\beta_F$

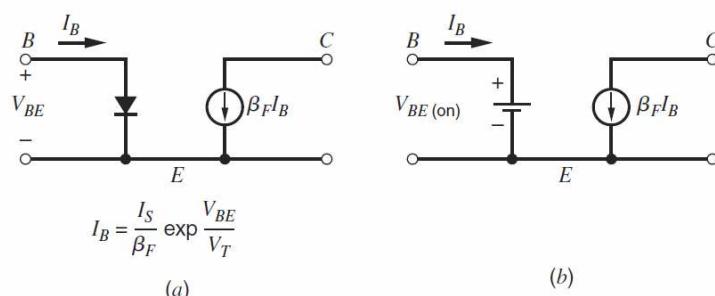


$$\beta_F = \frac{I_C}{I_B} \quad (\text{ideally infinite})$$

$$\alpha_F = \frac{I_C}{(-I_E)} = \frac{\beta_F}{1 + \beta_F} \quad (\text{ideally one})$$

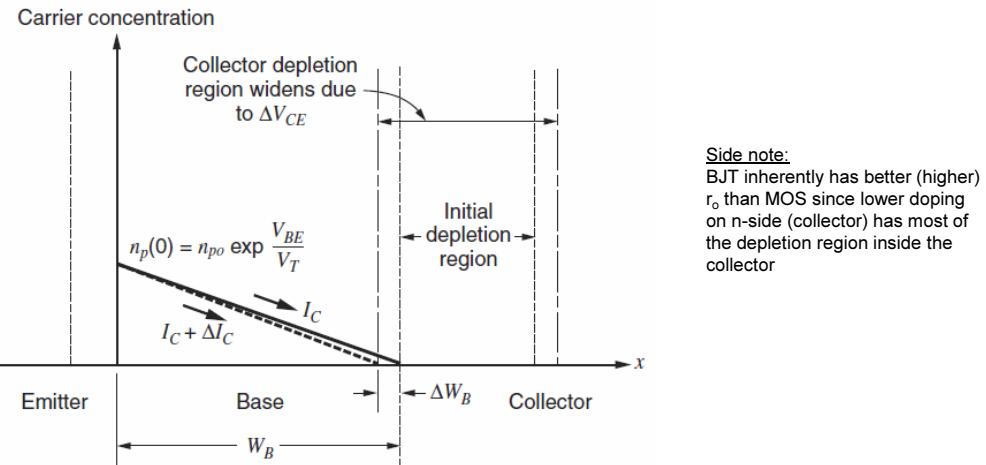
- The subscript “F” indicates that the device is assumed to operate in the forward active region (BE junction forward biased, BC reverse biased, as assumed so far)
  - More on other operating regions later...

## Basic Transistor Model



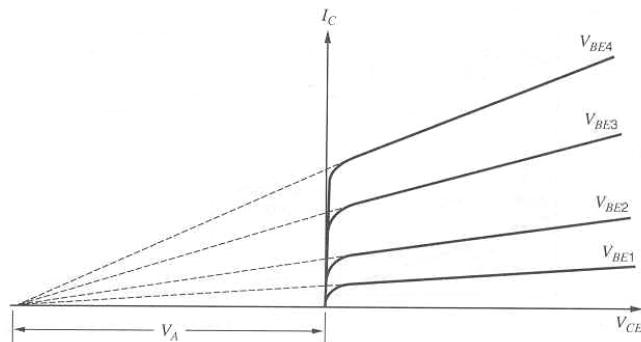
Simplified model; very useful for bias point calculations  
(assuming e.g.  $V_{BE(\text{on})} = 0.8V$ )

## Basewidth Modulation (1)



$$\frac{\partial I_C}{\partial V_{CE}} = \frac{\partial}{\partial V_{CE}} \left( \frac{qAD_n n_i^2}{W_B(V_{CE}) \cdot N_A} e^{\frac{V_{BE}}{V_T}} \right) = - \frac{I_C}{W_B} \frac{dW_B}{dV_{CE}}$$

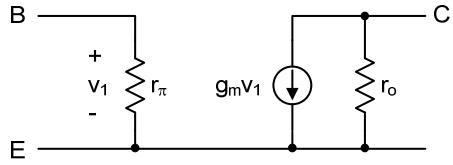
## Early Voltage ( $V_A$ )



$$V_A = \frac{I_C}{\frac{\partial I_C}{\partial V_{CE}}} = - \frac{W_B}{\frac{dW_B}{dV_{CE}}} = \text{const. (independent of } I_C \text{)}$$

$$I_C \approx I_S e^{\frac{V_{BE}}{V_T}} \left( 1 + \frac{V_{CE}}{V_A} \right)$$

## Small-Signal Model



$$g_m = \frac{dI_C}{dV_{BE}} = \frac{d}{dV_{BE}} \left[ I_S e^{\frac{V_{BE}}{V_T}} \left( 1 + \frac{V_{CE}}{V_A} \right) \right] = \frac{I_C}{V_T}$$

$$g_\pi = \frac{1}{r_\pi} = \frac{dI_B}{dV_{BE}} = \frac{d\left(\frac{I_C}{\beta_F}\right)}{dV_{BE}} = \frac{1}{\beta_F} \frac{I_C}{V_T} = \frac{g_m}{\beta_F} \quad (\text{assuming } \beta_F = \text{const.})$$

$$g_o = \frac{1}{r_o} = \frac{dI_C}{dV_{CE}} = \frac{d}{dV_{CE}} \left[ I_S e^{\frac{V_{BE}}{V_T}} \left( 1 + \frac{V_{CE}}{V_A} \right) \right] \approx \frac{I_C}{V_A}$$

## Intrinsic Gain

$$g_m r_o \approx \frac{I_C}{V_T} \cdot \frac{V_A}{I_C} = \frac{V_A}{V_T} \quad V_T \approx 26 \text{mV} \quad (\text{at room temperature})$$

- In the EE214B technology, the SiGe npn device has  $V_A = 90 \text{V}$ , thus

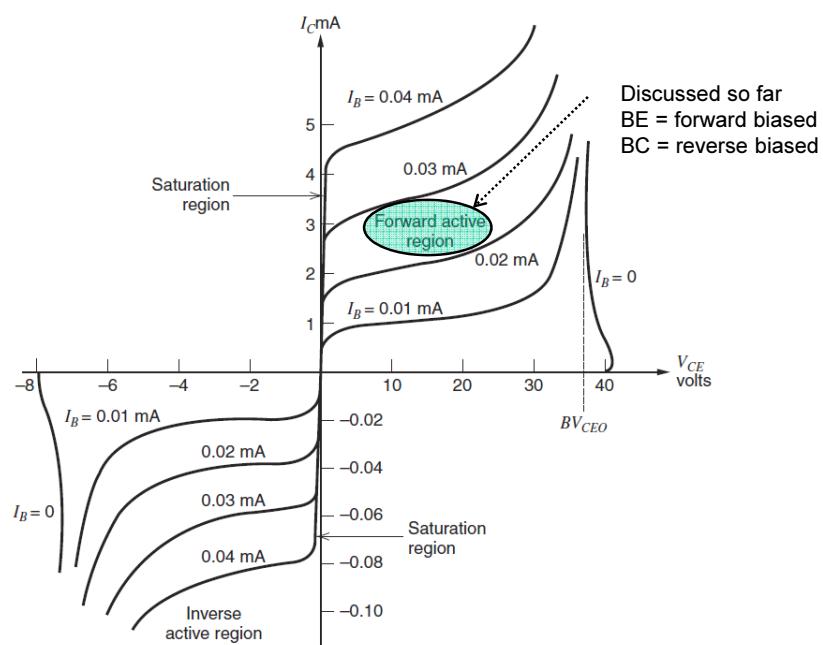
$$g_m r_o \approx \frac{90 \text{V}}{26 \text{mV}} = 3460$$

- Much larger than the intrinsic gain of typical MOSFET devices

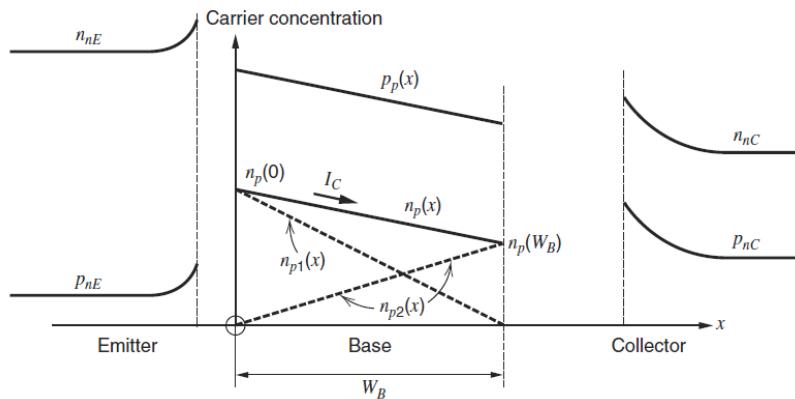
## Outline – Model Extensions and Technology

- Complete picture of BJT operating regions
- Dependence of  $\beta_F$  on operating conditions
- Device capacitances and resistances
- Technology
  - Junction isolated
  - Oxide isolated with polysilicon emitter
  - Heterojunction bipolar (SiGe base)
  - BiCMOS
  - Complementary bipolar

## BJT Operating Regions



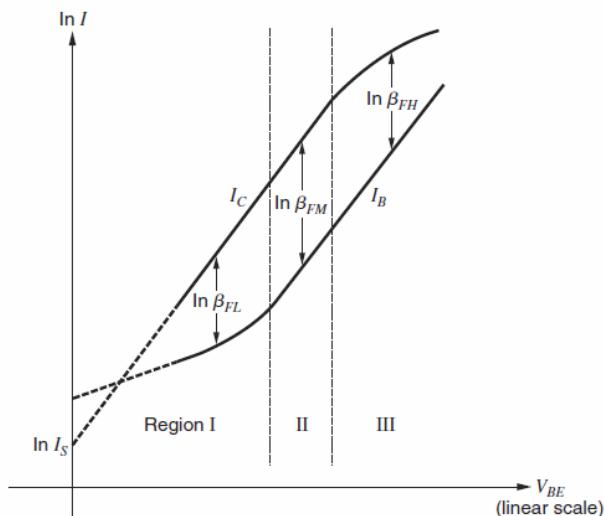
## Carrier Concentrations in Saturation



- Base-Collector junction is forward biased
- $n_p(W_B)$ , and therefore also  $I_C$ , strongly depend on  $V_{BC}$ ,  $V_{CE}$
- $V_{CE(sat)}$  is the voltage at which the device enters saturation,  $\sim 0.1 \dots 0.3V$ 
  - Related to the difference between the emitter and collector doping

## Gummel Plot

- A Gummel plot is a semi-log plot of  $I_C$  and  $I_B$  versus  $V_{BE}$  (linear scale)
- It reveals the regions for which high  $\beta_F$  is maintained (region II below)
- What happens in regions I and III?



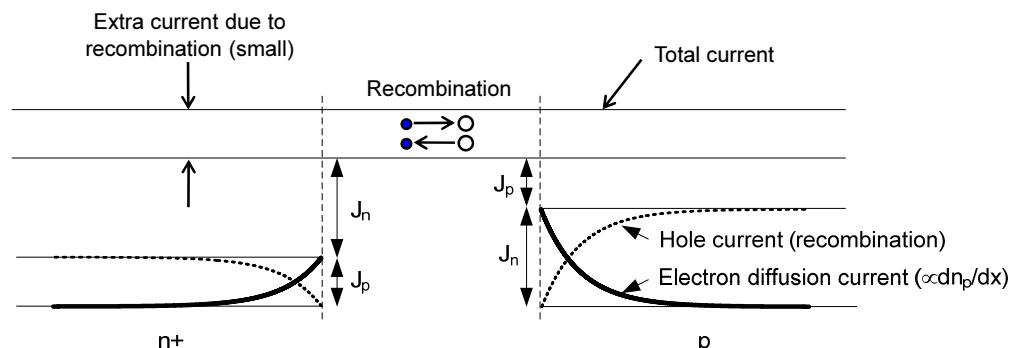
## $\beta_F$ Fall-Off

- Region III (high current density)
  - Injected electron charge in base region nears the level of doping (“high level injection”)
  - For this case, it can be shown that the injected carrier concentration rises with a smaller exponent (cut in half) and therefore

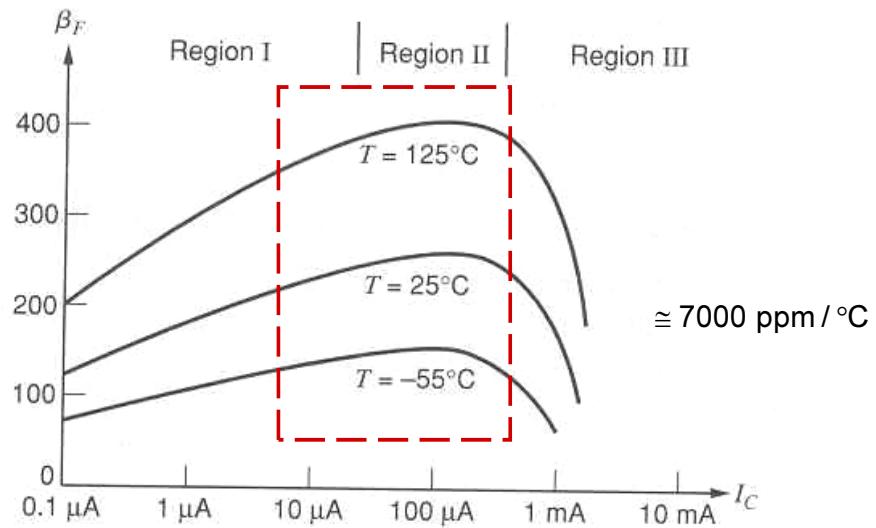
$$I_C = I_S e^{\frac{V_{BE}}{2V_T}}$$

- Region I (low current density)
  - There exists excess base current due to (unwanted) recombination in the depletion layer of the base-emitter junction
  - This current becomes significant at low current densities and sets a minimum for  $I_B$

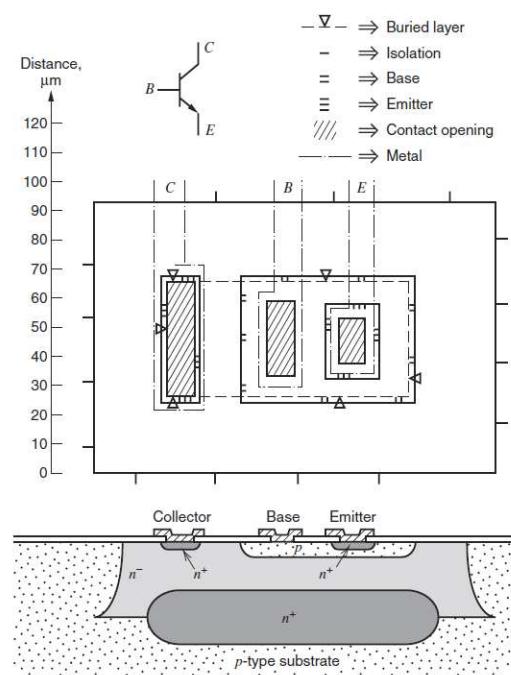
## Current Profile of a Forward Biased Diode Revisited



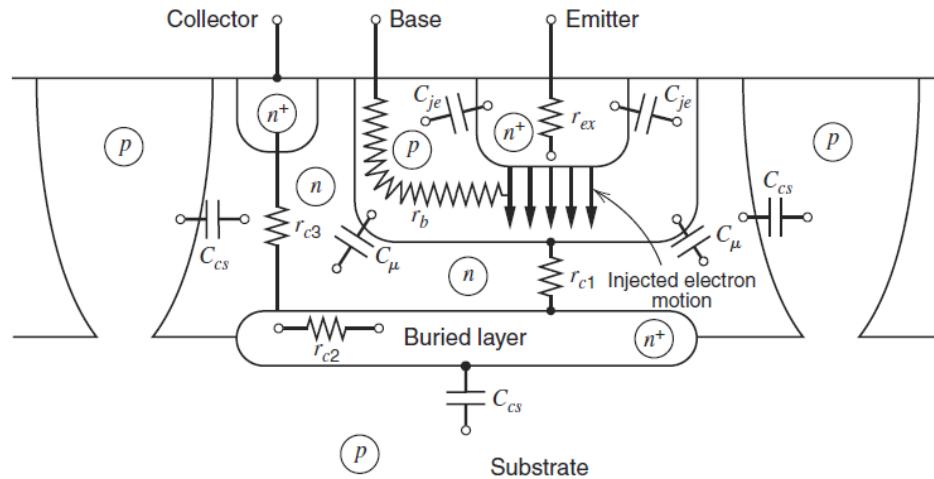
## $\beta_F$ vs. $I_C$ and Temperature



## Junction Isolated npn Transistor



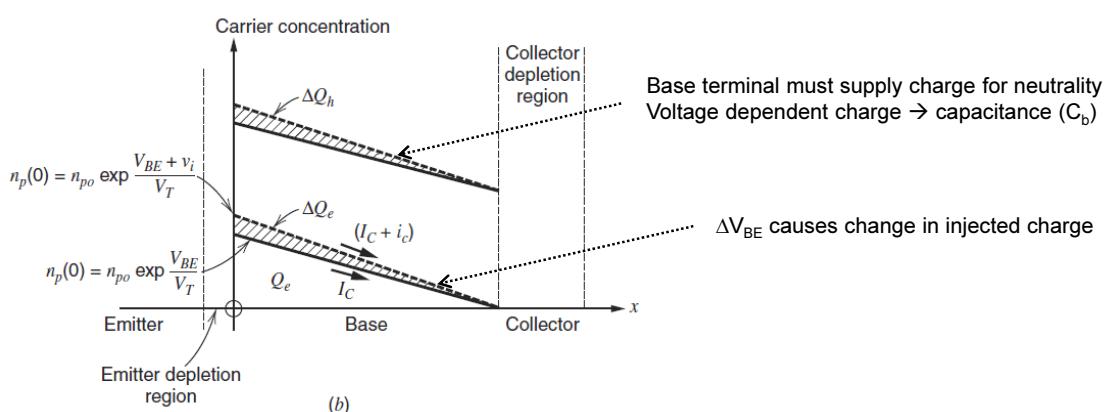
## Device Capacitances and Resistances



- Big mess!
- First focus on intrinsic elements

## Charge Storage

- In the intrinsic transistor, charge is stored in the junction capacitances,  $C_{je}$  and  $C_{jc} = C_\mu$ , and as minority carriers in the base and emitter
- Both minority carrier charge injected into the base and into the emitter, are proportional to  $\exp(V_{BE}/V_T)$ 
  - But the charge in the base is much larger, as discussed previously



## Base Charging Capacitance

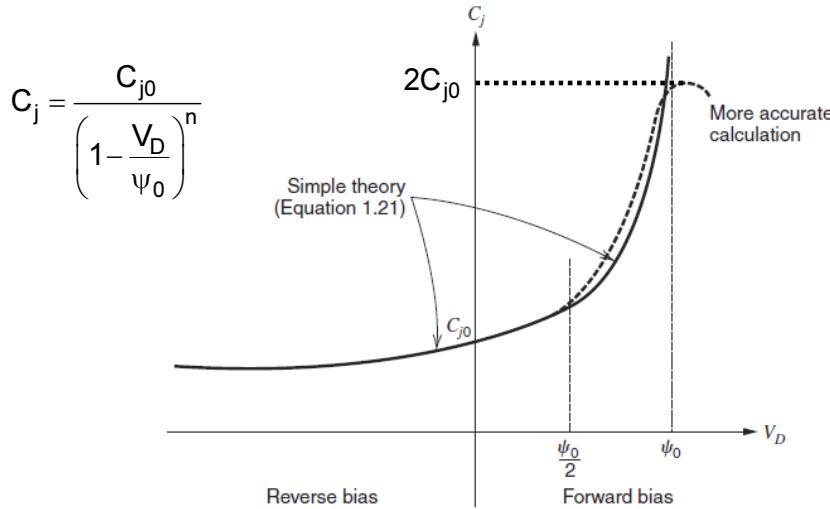
$$C_b = \frac{\partial Q_e}{\partial V_{BE}} = \frac{\partial Q_e}{\partial I_C} \frac{\partial I_C}{\partial V_{BE}} = \tau_F g_m$$

$$\tau_F = \frac{\partial Q_e}{\partial I_C} = \frac{\partial}{\partial I_C} \left( \frac{1}{2} n_p(0) W_B q A \right) \quad I_C = \frac{q A D_n n_p(0)}{W_B}$$

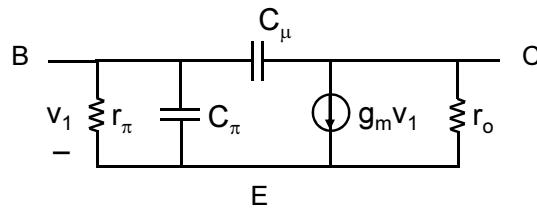
$$\tau_F = \frac{\partial}{\partial I_C} \left( \frac{1}{2} \frac{W_B^2}{D_n} I_C \right) = \frac{1}{2} \frac{W_B^2}{D_n}$$

- $\tau_F$  is called the base transit time (in forward direction)
- Typical values for high-speed transistors are on the order of 1...100ps

## Junction Capacitance



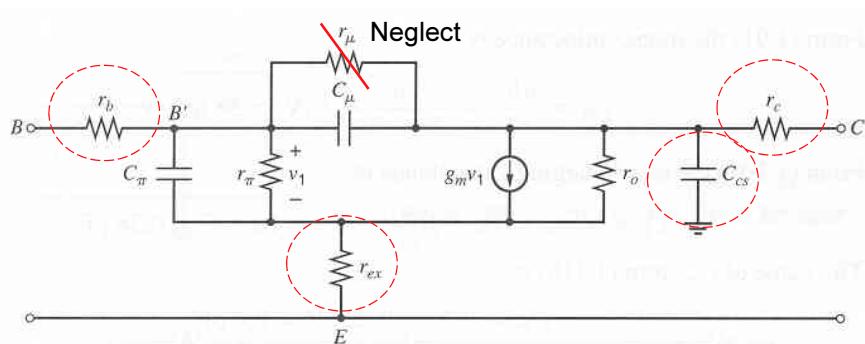
## Small-Signal Model with Intrinsic Capacitances



$$C_\pi = C_b + C_{je} = C_b + 2C_{je0}$$

$$C_\mu = C_{jc} = \frac{C_{jc0}}{\left(1 + \frac{V_{CB}}{\Psi_{0c}}\right)^n}$$

## Model with Additional Parasitics

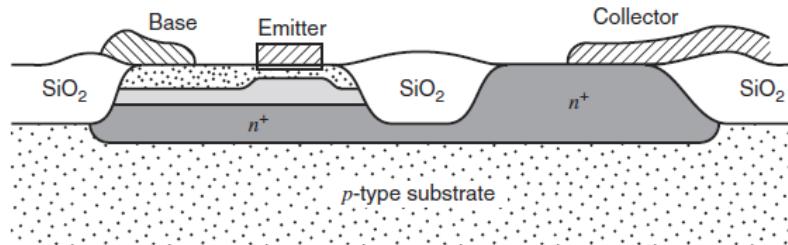


### Range of numbers

$r_e \sim 1-3\Omega$	}
$r_b \sim 50-500\Omega$	
$r_c \sim 20-500\Omega$	
$C_{cs} \sim 3-200fF$	

Values at high end of these ranges may have large impact on performance → Try to minimize through advanced processing & technology

## BJT in Advanced Technology



- Oxide isolated
- Self-aligned structure (base and emitter align automatically)
- Very thin base (~100nm or less) through ion implantation
- Reduced breakdown voltages compared to more traditional structures

## SiGe Heterojunction Bipolar Technology

- A heterojunction is a pn junction formed with different materials for the n and p regions
- Germanium is added to the base of a silicon bipolar transistor to create a heterojunction bipolar transistor (HBT)
  - Base formed by growing a thin epitaxial layer of SiGe
  - Results in a lower bandgap (and higher intrinsic carrier concentration) in the base than emitter
- In “band diagram speak” the bandgap mismatch increases the barrier to the injection of holes (in an npn transistor) from the base into the emitter
- One way to enumerate the benefits of a SiGe base is to look at the current gain expression

## HBT Current Gain

- Intrinsic carrier concentration in the SiGe base ( $n_{iB}$ ) is larger than intrinsic carrier concentration in the Si emitter ( $n_{iE}$ )

$$\beta_F = \frac{\frac{qAD_n n_{p0}}{W_B}}{\frac{1}{2} \frac{n_{p0} W_B qA}{\tau_b} + \frac{qAD_p n_{iE}^2}{L_p N_D}} \approx \frac{\frac{qAD_n n_{p0}}{W_B}}{\frac{qAD_p n_{iE}^2}{L_p N_D}} = \frac{\frac{qAD_n n_{iB}^2}{W_B N_A}}{\frac{qAD_p n_{iE}^2}{L_p N_D}} = \frac{D_n N_D L_p}{D_p N_A W_B} \frac{n_{iB}^2}{n_{iE}^2}$$

Added degree of freedom for HBT

- Base doping ( $N_A$ ) can be increased while maintaining same  $\beta_F$ 
  - Can reduce base width without affecting  $r_b$
  - Larger  $r_o$  due to decrease in base width modulation

## Device Parameter Comparison

Parameter	Vertical <i>npn</i> Transistor with $2 \mu\text{m}^2$ Emitter Area	SiGe npn HBT Transistor with $0.7 \mu\text{m}^2 = 0.22 \mu\text{m} \times 3.2 \mu\text{m}$ Emitter Area
$\beta_F$	120	300
$\beta_R$	2	2
$V_A$	35 V	90
$I_S$	$6 \times 10^{-18} \text{ A}$	$3.2 \times 10^{-17} \text{ A}$
$I_{CO}$	1 pA	1 pA
$BV_{CEO}$	8 V	2.0 V
$BV_{CBO}$	18 V	5.5 V
$BV_{EBO}$	6 V	3.3 V
$\tau_F$	10 ps	0.56 ps
$\tau_R$	5 ns	10 ps
$r_b$	$400 \Omega$	$25 \Omega$
$r_c$	$100 \Omega$	$60 \Omega$
$r_{ex}$	$40 \Omega$	$2.5 \Omega$
$C_{je0}$	5 fF	6.26 fF
$\psi_{0e}$	0.8 V	0.8 V
$n_e$	0.4	0.4
$C_{\mu 0}$	5 fF	3.42 fF
$\psi_{0c}$	0.6 V	0.6 V
$n_c$	0.33	0.33
$C_{cs0} (C_{bs0})$	20 fF	3.0 fF
$\psi_{0s}$	0.6 V	0.6 V
$n_s$	0.33	0.33

## BiCMOS Technology

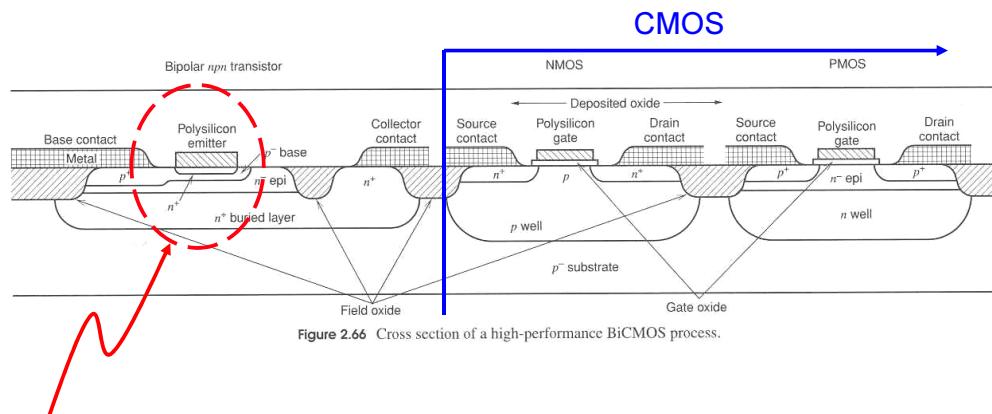


Figure 2.66 Cross section of a high-performance BiCMOS process.

Older BJTs used poly Si as “diffusion source” for emitter doping. Advanced (state-of-the-art) BJTs use epitaxial growth of both the SiGe base and Si emitter regions

## Advanced Complementary Bipolar Technology

### *BiCom3x* Overview

#### Technology Features:

- 200mm Wafers
- 0.35 um Features
- 5V Operation
- SOI Substrates
- Trench Isolation
- SiGe Bipolar (NPN & PNP)
- 115 A Gate Ox
- QLM (1.0 um Pitch)
- NiCrAl Thin Film Resistor
- Metal-Silicide Capacitor
- Cu Power Metal Option
- Laser Trimming
- 13K Gates/mm<sup>2</sup>
- Qualified 1Q06
- In Production

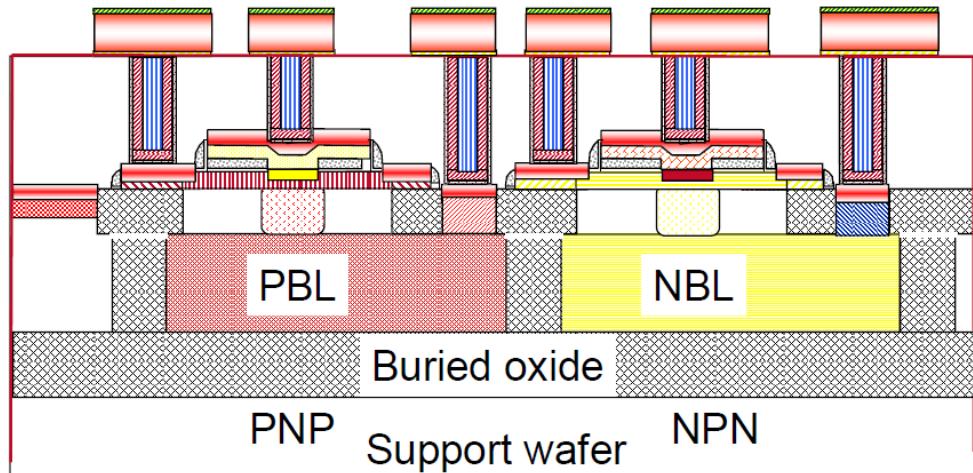
#### Component Set:

- CMOS: 5V & 3.3V
- Isolated CMOS
- 5V NPN
- 5V PNP
- Poly Resistors
- Well Resistors
- Thin Film Resistor
- TiN-Polyicide Capacitor
- Poly Fuse



[Texas Instruments]

## Cross Section



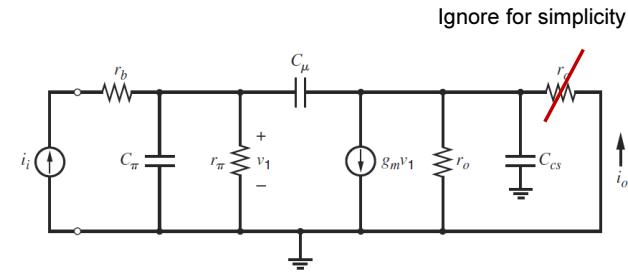
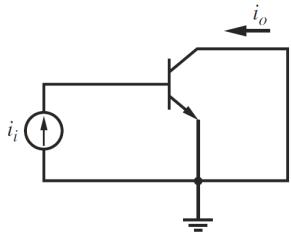
[Texas Instruments]

## Figures of Merit for BJTs

- Product of current gain and Early voltage,  $\beta \cdot V_A$
- Product of transit frequency and breakdown voltage,  $f_T \cdot BV_{CEO}$
- Maximum frequency of oscillation,  $f_{max}$ 
  - More in EE314A
- Transit (or transition) frequency,  $f_T$ 
  - Formally defined as the frequency for which the current gain of the device falls to unity
  - Important to keep in mind that the basic device model may fall apart altogether at this frequency
    - Lumped device models tend to be OK up to  $\sim f_T/5$
  - Therefore,  $f_T$  should be viewed as an extrapolated parameter, or simply as a proxy for device transconductance per capacitance

## Transit Frequency Calculation (1)

(AC circuit; DC biasing not shown)



$$v_1 = \frac{r_\pi}{1 + r_\pi s(C_\pi + C_\mu)} i_i$$

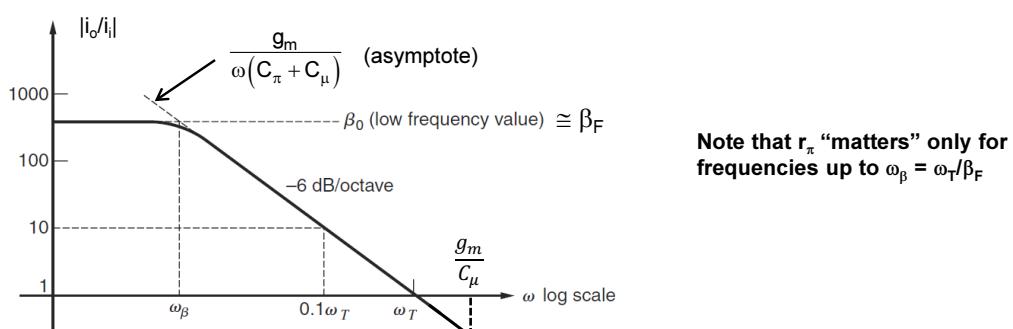
$$i_o = g_m v_1 + s C_\mu v_1$$

$$\frac{i_o}{i_i} = \frac{g_m r_\pi \left(1 - s \frac{C_\mu}{g_m}\right)}{1 + s r_\pi (C_\pi + C_\mu)}$$

**Low frequencies:**  $\frac{i_o}{i_i} \approx g_m r_\pi = \beta_F$  for  $\omega \ll \frac{1}{r_\pi (C_\pi + C_\mu)} = \omega_\beta$

**High frequencies:**  $\frac{i_o}{i_i} \approx \frac{g_m}{j\omega (C_\pi + C_\mu)}$  for  $\omega_\beta \ll \omega \ll \frac{g_m}{C_\mu}$

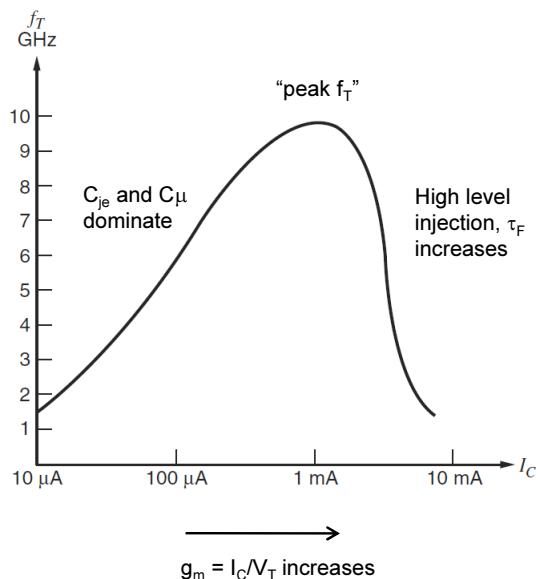
## Transit Frequency Calculation (2)



$$1 = \frac{g_m}{\omega_T (C_\pi + C_\mu)} \Rightarrow \omega_T = \frac{g_m}{C_\pi + C_\mu}$$

$$\tau_T = \frac{1}{\omega_T} = \frac{C_b}{g_m} + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} = \tau_F + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m}$$

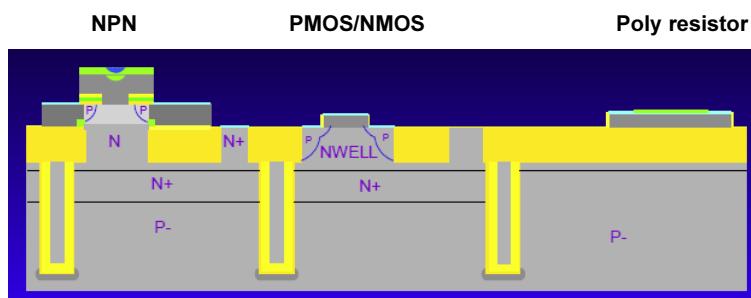
## $f_T$ versus $I_C$ plot



- The particular current value at which  $f_T$  is maximized depends on the parameters of a technology and the emitter area of the BJT

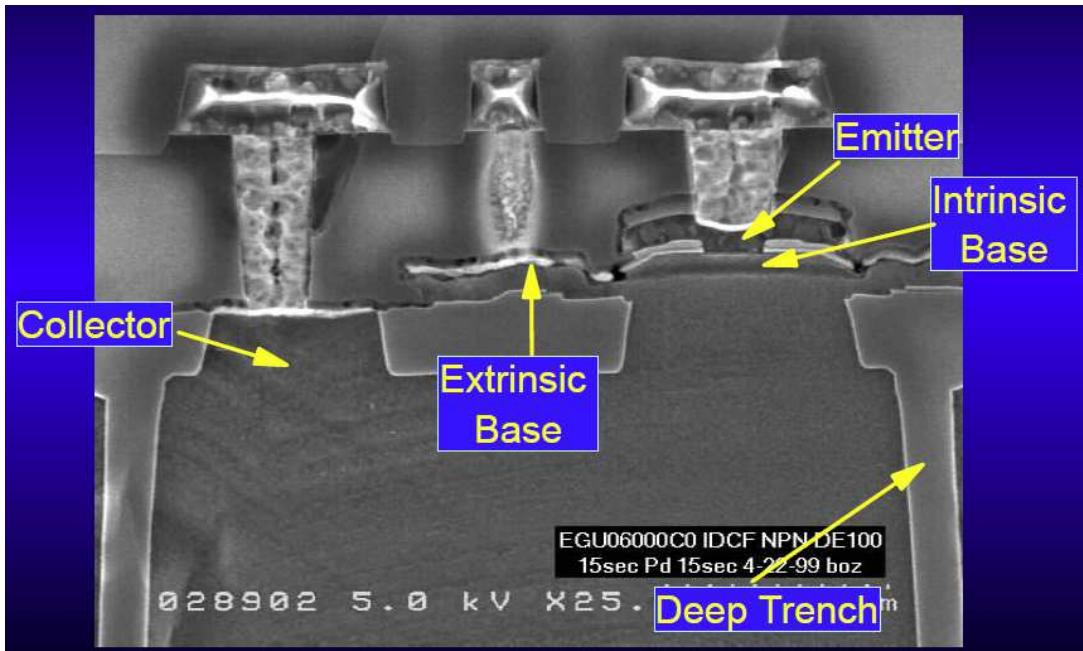
## EE214B Technology

- Assumed to be similar to a  $0.18\text{-}\mu\text{m}$  BiCMOS technology featuring a high-performance SiGe npn device
  - $V_{CC} = 2.5\text{V}$  (BJT),  $V_{DD} = 1.8\text{V}$  (MOS)
- See e.g.
  - Wada et al., BCTM 2002
  - Joseph et al., BCTM 2001
  - IBM 7HP documentation
    - [https://www-01.ibm.com/chips/techlib/techlib.nsf/products/BiCMOS\\_7HP](https://www-01.ibm.com/chips/techlib/techlib.nsf/products/BiCMOS_7HP)



<http://fuji.stanford.edu/events/spring01/slides/harameSlides.pdf>

## Cross Section of npn Device



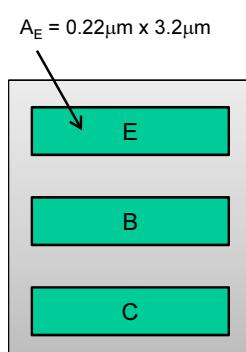
B. Murmann

EE214B Winter 2015-16 – Chapter 2

49

## EE214B npn Unit Device

- A technology typically comes with an optimized layout for a unit device of a certain size
- Great care is then taken to extract a Spice model for this particular layout using measured data
- Spice model (`/usr/class/ee214b/hspice/ee214_hspice.sp`)



```
.model npn214 npn
+ level=1 tref=25 is=.032f bf=300 br=2 vaf=90
+ cje=6.26f vje=.8 mje=.4 cjc=3.42f vjc=.6 mjc=.33
+ re=2.5 rb=25 rc=60 tf=563f tr=10p
+ xtf=200 itf=80m ikf=12m ikr=10.5m nkf=0.9
```

- Instantiation in a circuit netlist

```
*      C   B   E
q1    n1  n2  n3  npn214
```

## BJT Model Parameters

Table 5-3 BJT Model Parameters

Parameter	Description
DC	BF, BR, IBC, IBE, IS, ISS, NF, NR, NS, VAF, VAR
beta degradation	ISC, ISE, NC, NE, IKF, IKR
geometric	SUBS, BULK
resistor	RB, RBM, RE, RC, IRB
junction capacitor	CJC, CJE, CJS, FC, MJC, MJE, MJS, VJC, VJE, VJS, XCJC
parasitic capacitance	CBCP, CBEP, CCSP
transit time	ITF, PTF, TF, VT, VTF, XTF
noise	KF, AF

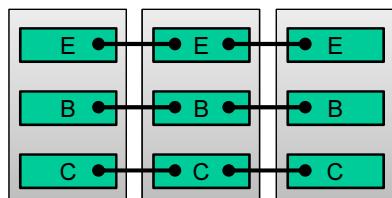
- For more info consult the HSpice documentation under

`/afs/ir.stanford.edu/class/ee/synopsys/B-2008.09-SP1/hspice/docs_help`

PDF files:

home.pdf hspice\_cmdref.pdf hspice\_integ.pdf hspice\_relnote.pdf hspice\_sa.pdf  
hspice\_devmod.pdf hspice\_mosmod.pdf hspice\_rf.pdf hspice\_si.pdf

## Adding Multiple Devices in Parallel



- For the unit device, there exists a practical upper bound for the collector current
  - Due to the onset of high level injection
- This means that the unit device can only deliver a certain maximum  $g_m$
- If more  $g_m$  is needed, "m" unit devices can be connected in parallel
- Instantiation in a circuit netlist ( $m=3$ )

```
*      C   B   E  
q1    n1  n2  n3  npn214  3
```

## npn Unit Device Characterization

```

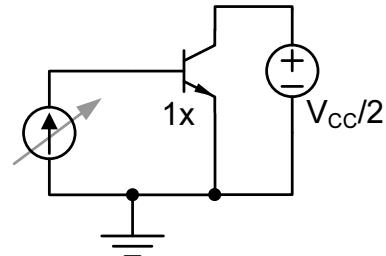
* ee214b npn device characterization

* C B E
q1 c b 0 npn214
Vc c 0 1.25
ib 0 b 1u

.op
.dc ib dec 10 10f 100u
.probe ib(q1) ic(q1) ie(q1)
.probe gm = par('gm(q1)')
.probe go = par('g0(q1)')
.probe cpi = par('cap_be(q1)')
.probe cmu = par('cap_ibc(q1)')
.probe beta = par('beta(q1)')

.options dccap post brief
.inc '/usr/class/ee214b/hspice/ee214_hspice.sp'
.end

```



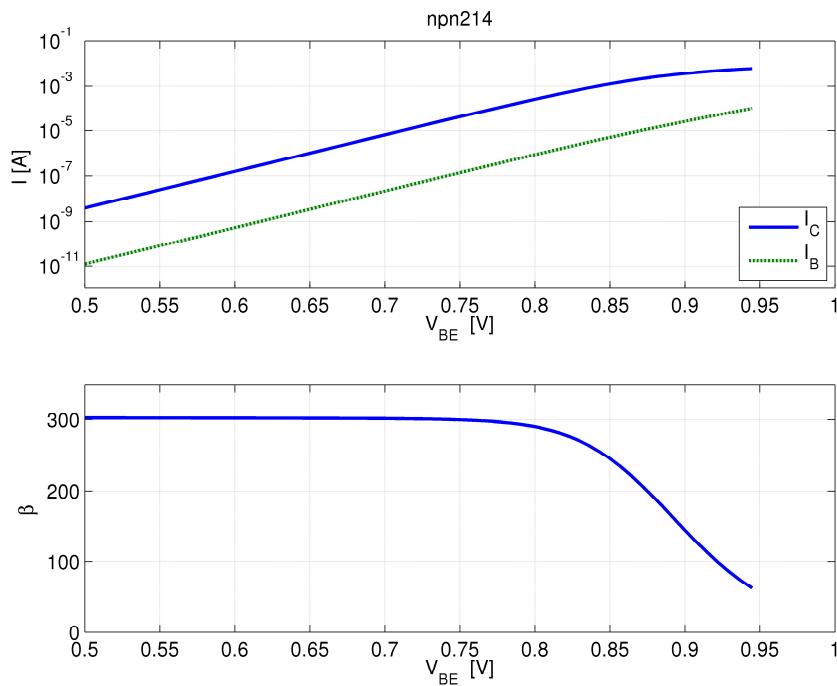
## DC Operating Point Output

Table 5-18 BJT DC Operating Point Output

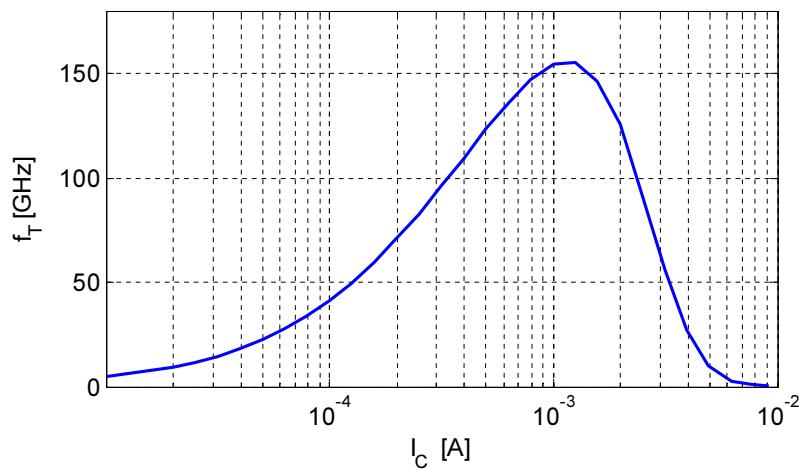
**** bipolar junction transistors	
element	0:q1
model	0:npn214
ib	999.9996n
ic	288.5105u
vbe	803.4402m
vce	1.2500
vbc	-446.5598m
vs	-1.2327
power	361.4415u
betad	288.5106
gm	10.2746m
rpi	26.8737k
rx	25.0000
ro	313.4350k
cpi	14.6086f
cmu	2.8621f
cbx	0.
ccs	0.
betaac	276.1163
ft	93.5999g

Quantities	Definitions
ib	base current
ic	collector current
is	substrate current
vbe	B-E voltage
vbc	B-C voltage
vcs	C-S voltage
vs	substrate voltage
power	power
betad(betadc)	beta for DC analysis
gm	transconductance
rpi	B-E input resistance
rmu(rmuV)	B-C input resistance
rx	base resistance
ro	collector resistance
cpi	internal B-E capacitance
cmu	internal B-C capacitance
cbx	external B-C capacitance
ccs	C-S capacitance
cbs	B-S capacitance
cxs	external substrate capacitance
betaac	beta for AC analysis
ft	unity gain bandwidth

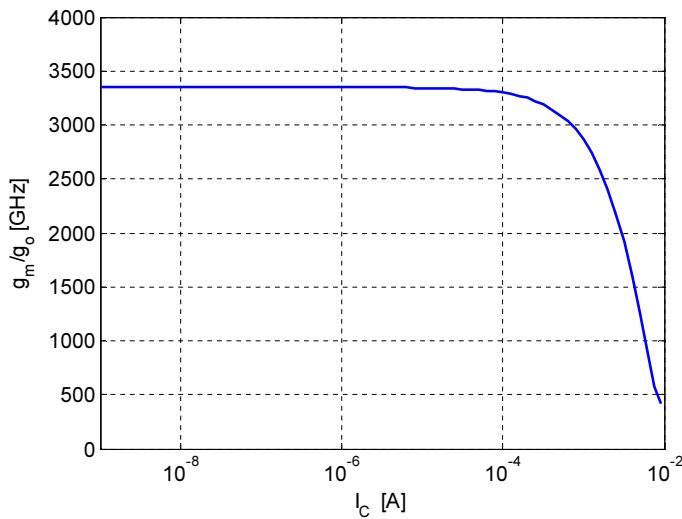
## Gummel Plot



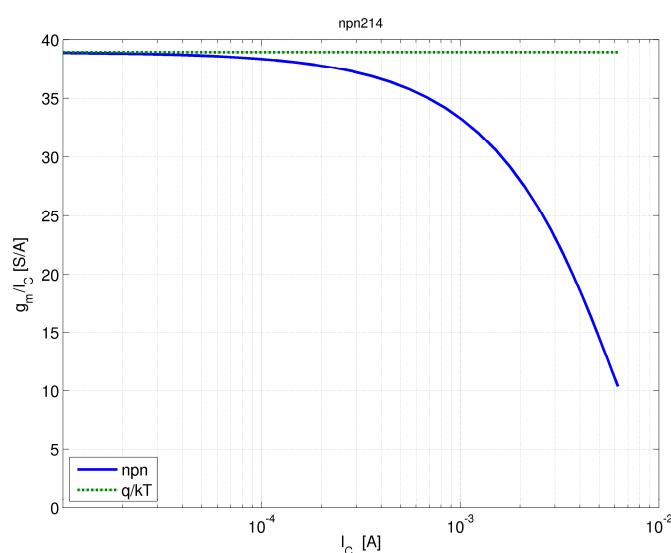
## Transit Frequency



## Intrinsic Gain

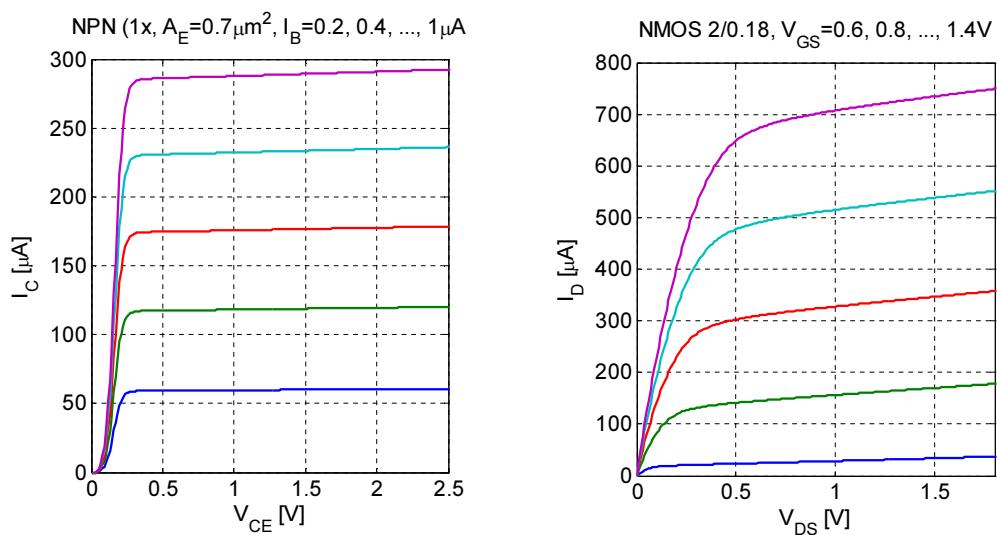


## $g_m/I_c$



- Important to realize that  $g_m$  will not be exactly equal to  $I_C/V_T$  at high currents

## I-V Curves

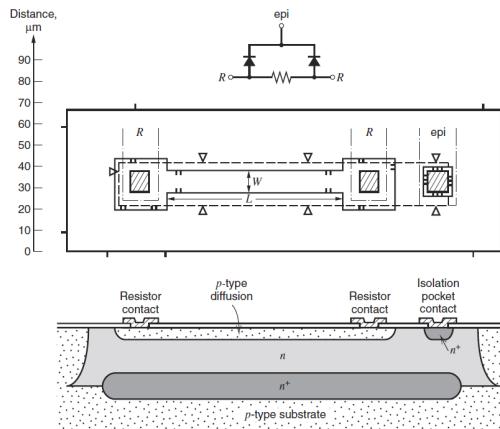


## Passive Components (1)

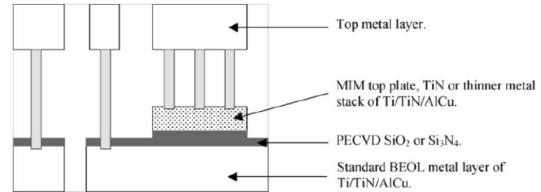
Resistors	$R_s$ (O/Sq)	TCR (ppm/C)
Subcollector	8.1	1430
N+ Diffusion	72	1910
P+ Diffusion	105	1430
P+ Polysilicon	270	50
P Polysilicon	1600	-1178
TaN	135	-750
Capacitors	$C_p$ (fF/ $\mu m^2$ )	VCR (+5/-5 ppm/V)
MIM	1	<45
MOS	2.6	5
Inductor	$L$ (nH)	Max Q at 5 GHz
Al - Spiral Inductor	$>=0.7$	21
Varactor	Tuning Range	Q @ 0.5 GHz
CB Junction	1.64:1	90
MOS Accumulation	3.1:1	300

## Passive Components (2)

**Diffusion Resistor**

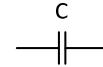


**MIM Capacitor**

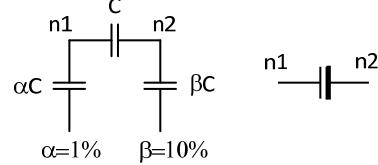


[Ng, Trans. Electron Dev. 7/2005]

**Ideal Capacitor**



**Typical MIM Capacitor**



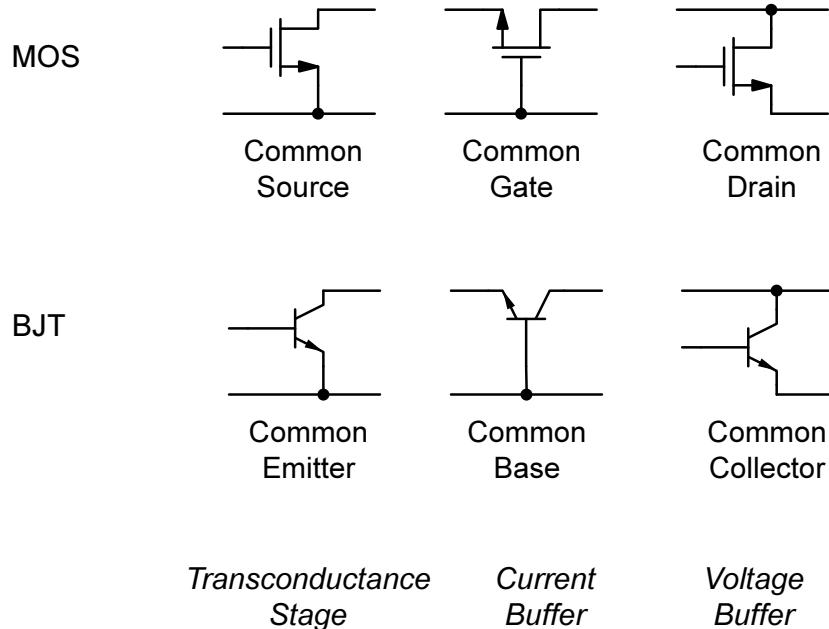
# Chapter 3

## Elementary BJT Circuits

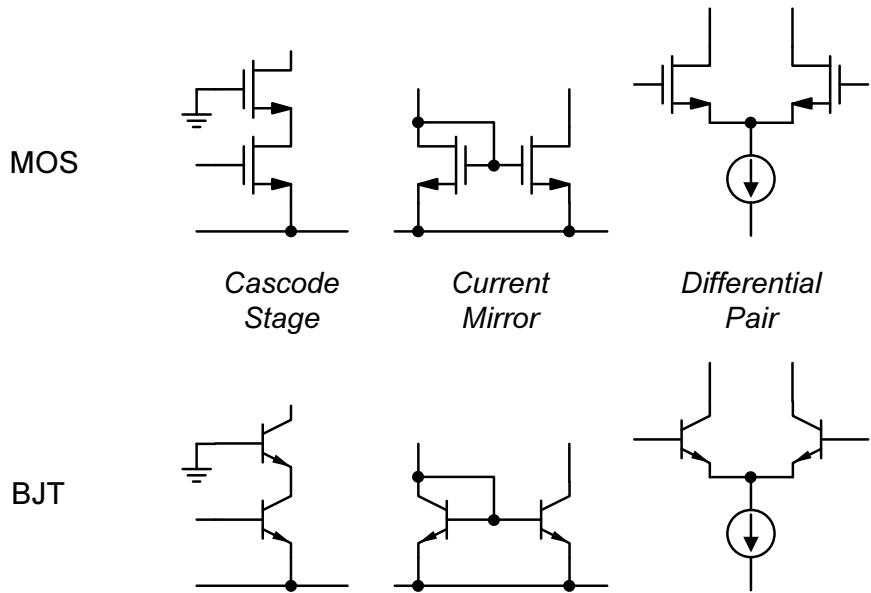
Boris Murmann  
Stanford University  
Winter 2015-16

Textbook Sections: 8.5, 7.3

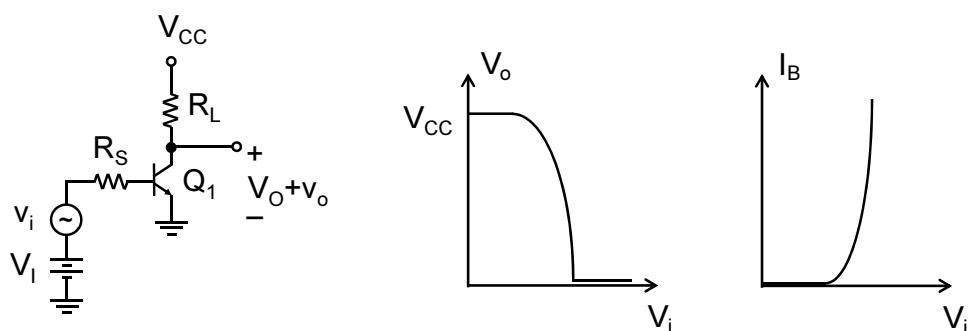
### Elementary Amplifier Configurations



## Widely Used Two-Transistor Circuits

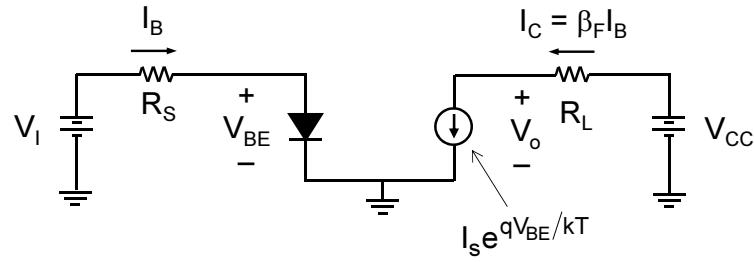


## Common-Emitter Stage



- DC input bias voltage ( $V_I$ ) biases  $Q_1$  in the forward active region
- Typically, want  $V_o \approx V_{CC}/2$
- Main differences to consider versus common-source stage (MOS)
  - Bias point sensitivity
  - Finite input resistance (due to  $r_\pi$ )
  - Base resistance ( $r_b$ ) sometimes significant

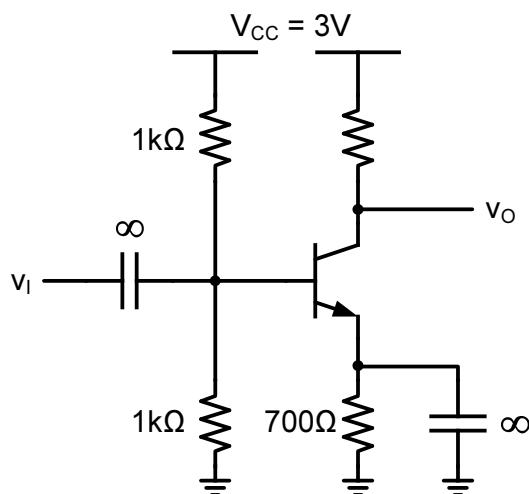
## Bias Point Sensitivity



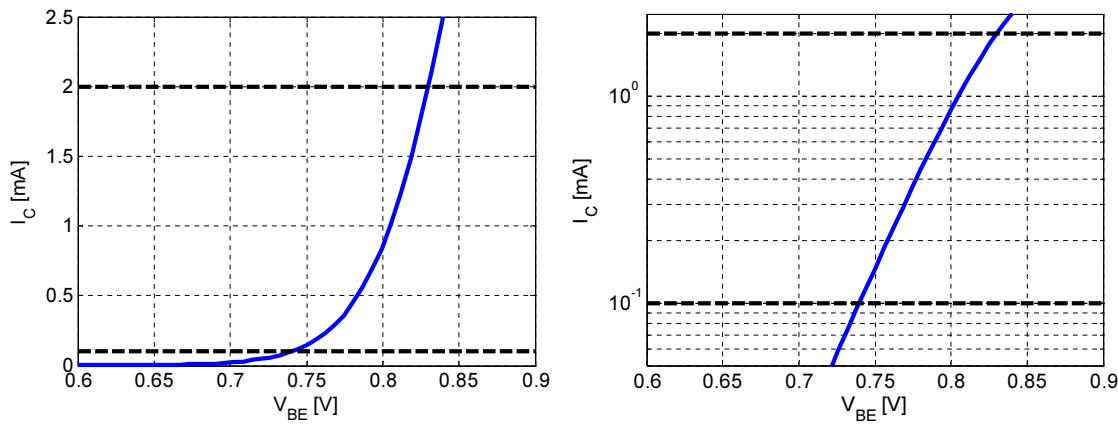
$$I_B \approx \frac{V_I - V_{BE(on)}}{R_S} \quad V_O = V_{CC} - I_C R_L = V_{CC} - \beta_F I_B R_L = V_{CC} - \frac{\beta_F}{R_S} (V_I - V_{BE(on)})$$

- The dependence on  $\beta_F$  makes “direct voltage biasing” impractical
- How to generate  $V_I$  so as to control  $V_o$ ?
- Practical configurations are usually based on feedback, replica biasing, ac coupling or differential pairs

## Simple Example



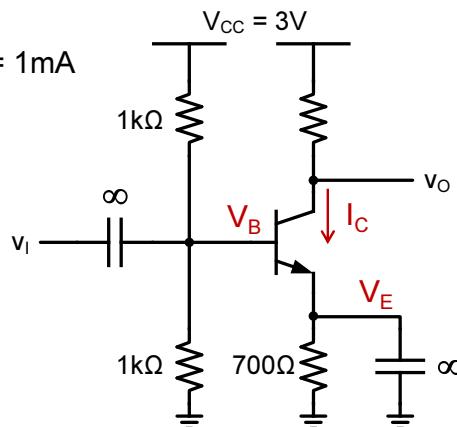
## $V_{BE}$ Approximation for First-Order Bias Calculations



- For the useful range of unit collector current between 0.1...2mA,  $V_{BE}$  changes by less than 100mV
- We can therefore approximate  $V_{BE} \approx \text{const.} \approx 0.8\text{V}$  for bias calculations

## Simple Example – Hand Analysis and Simulation

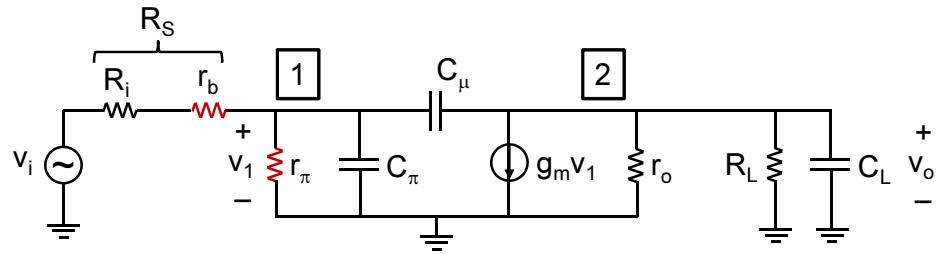
- Neglecting  $I_B$ , we have  $V_B = V_{CC}/2 = 1.5\text{V}$
- Assuming  $V_{BE} = 0.8\text{V}$ , we know that  $V_E = 0.7\text{V}$
- $I_C \approx I_E = 0.7\text{V}/700\Omega = 1\text{mA}$
- $g_m = I_C/V_T = 38\text{mS}$



HSpice .op Result:

```
element 0:q1
model 0:npn214
ib      3.8005u
ic      987.3900u
vbe    804.2664m
vce    1.9112
vbc    -1.1069
vs     -2.5458
power   1.8902m
betad  259.8058
gm     32.9729m
rpi    6.7601k
rx     25.0000
ro    92.2106k
cpi    29.7574f
cmu    2.4504f
betaac 222.9016
ft     162.9361g
```

## Small-Signal Equivalent Circuit for CE Stage



- For hand analysis, we will usually neglect  $r_c$  and  $r_e$  (from the BJT model)
- If significant,  $r_b$  can be included with  $R_i$ , i.e.  $R_s = R_i + r_b$
- Resulting low-frequency gain

$$A_{v0} = - \left( \frac{r_\pi}{R_s + r_\pi} \right) \cdot g_m R_{L\text{tot}} \quad R_{L\text{tot}} = r_o \parallel R_L$$

$\underbrace{\qquad\qquad\qquad}_{=1 \text{ for MOS}}$

## Frequency Response

- Using nodal analysis (neglecting  $r_\pi$ ), we find

$$A_v(s) = \frac{v_o(s)}{v_i(s)} = A_{v0} \frac{\left(1 - \frac{s}{z_1}\right)}{1 + b_1 s + b_2 s^2}$$

$$b_1 = R_s [C_\pi + C_\mu (1 + g_m R_{L\text{tot}})] + R_{L\text{tot}} (C_L + C_\mu)$$

$$b_2 = R_s R_{L\text{tot}} (C_\pi C_L + C_\pi C_\mu + C_L C_\mu)$$

$$z_1 = + \frac{g_m}{C_\mu}$$

- $z_1$  is a feedforward zero in the RHP that can typically be ignored
- If  $C_\pi \gg C_\mu$ , then

$$z_1 = + \frac{g_m}{C_\mu} \gg \frac{g_m}{C_\pi + C_\mu} = \omega_T$$

## Toolkit for Simplification (1)

- We typically want a much lower entropy result for design
- The following options are listed in order of simplicity (and accuracy)
- Bandwidth estimate using the Miller approximation
  - Approximate the gain across  $C_\mu$  as frequency independent for the frequency range of interest – must be checked for validity
  - A (slightly optimistic) estimate of the circuit's bandwidth is then:



$$\omega_{3dB} \approx \frac{1}{R_s [C_\pi + C_\mu (1 + g_m R_{Ltot})]}$$

## Toolkit for Simplification (2)

- Open-circuit time constant (OCT) analysis
  - The coefficient  $b_1$  can be found by summing all zero value time constants in the circuit and the corresponding (conservative) bandwidth estimate is
- If we need to know the location of the non-dominant pole, neither Miller nor OCT is useful
- However, if we know that one of the poles is indeed dominant, we can simplify using the so-called dominant pole approximation

$$\begin{aligned} \frac{1}{\left(1 - \frac{s}{p_1}\right)\left(1 - \frac{s}{p_2}\right)} &= \frac{1}{1 - \frac{s}{p_1} - \frac{s}{p_2} + \frac{s^2}{p_1 p_2}} \approx \frac{1}{1 - \frac{s}{p_1} + \frac{s^2}{p_1 p_2}} \\ \text{Given } \frac{1}{1 + b_1 s + b_2 s^2} &\Rightarrow p_1 \approx -\frac{1}{b_1}, \quad p_2 \approx -\frac{b_1}{b_2} \end{aligned}$$

## Dominant Pole Approximation for the CE Stage

- If a dominant pole condition exists, we can therefore write

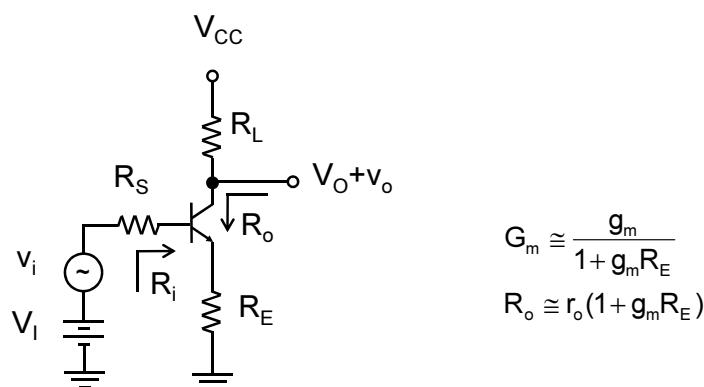
$$p_1 \approx -\frac{1}{b_1} = -\frac{1}{R_S [C_\pi + C_\mu (1 + g_m R_{Ltot})] + R_{Ltot} (C_L + C_\mu)}$$

$$p_2 \approx -\frac{b_1}{b_2} = -\frac{R_S [C_\pi + C_\mu (1 + g_m R_{Ltot})] + R_{Ltot} (C_L + C_\mu)}{R_S R_{Ltot} (C_\pi C_L + C_\pi C_\mu + C_L C_\mu)}$$

- If  $C_\mu \ll C_\pi, C_L$ , then

$$p_2 \approx -\frac{R_S [C_\pi + C_\mu g_m R_{Ltot}] + R_{Ltot} C_L}{R_S R_{Ltot} C_\pi C_L} = -\left( \frac{1}{R_L C_L} + \frac{g_m}{C_L} \cdot \frac{C_\mu}{C_\pi} + \frac{1}{R_S C_\pi} \right)$$

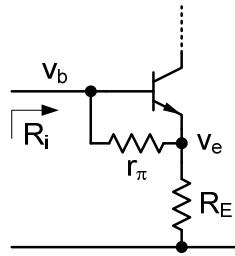
## Emitter Degeneration



- The degeneration resistor reduces the transconductance and increases the output resistance of the device (same as in the MOS CS stage)
- For the BJT version,  $R_E$  helps increase the input resistance

## Input Resistance Calculation

- Can always do a nodal analysis
- A more intuitive way to find  $R_i$  is via the Miller theorem

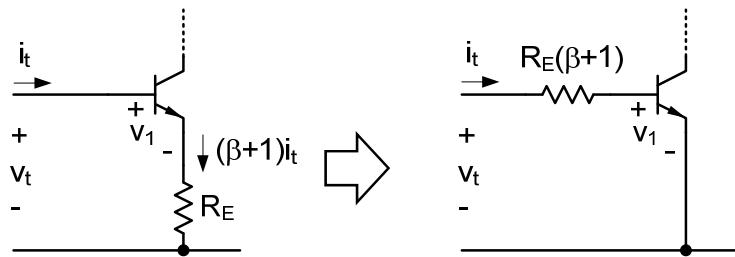


$$K = \frac{V_e}{V_b} \approx \frac{R_E}{\frac{1}{g_m} + R_E} = \frac{g_m R_E}{1 + g_m R_E}$$

$$R_i = \frac{r_\pi}{1-K} \approx \frac{r_\pi}{\left(1 - \frac{g_m R_E}{1 + g_m R_E}\right)} = r_\pi (1 + g_m R_E)$$

- The same “bootstrapping” effect applies to  $C_\pi$ , we see  $C_\pi/(1+g_m R_E)$  looking into the input
  - Assuming  $K = \text{constant}$  in the frequency range of interest

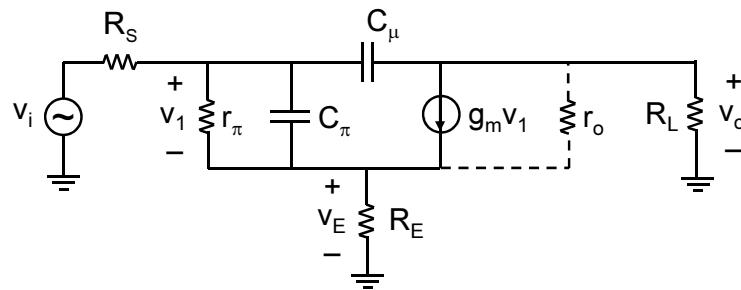
## Alternative “Trick” Calculation



$$R_i \approx r_\pi + R_E (1 + \beta) = r_\pi + R_E (1 + g_m r_\pi) \approx r_\pi (1 + g_m R_E)$$

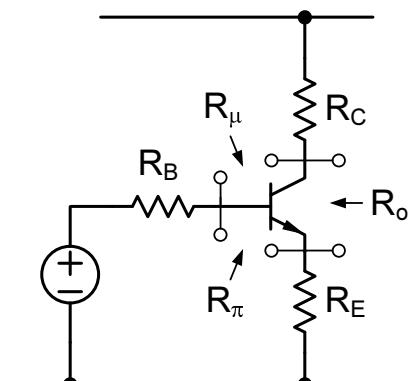
- Tricks of this kind are useful for reasoning about low frequency behavior
- A more detailed analysis is required when investigating frequency dependence

## Small-Signal Equivalent Circuit for Degenerated CE Stage



- Deriving the transfer function of this circuit requires solving a  $3 \times 3$  system of equations
- This is very messy, and so we will just look at an OCT-based bandwidth estimate to gain some basic insight

## Useful Expressions



$$R_o \cong r_o \parallel \frac{R_C + R_E}{1 + g_m R_E} \quad (\text{for } \beta \rightarrow \infty)$$

$$R_\pi \cong r_\pi \parallel \frac{R_B + R_E}{1 + g_m R_E}$$

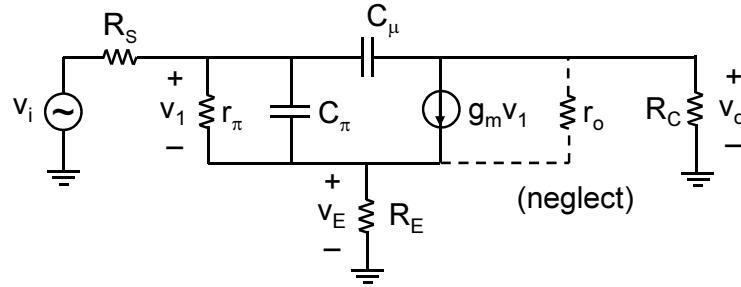
$$R_\mu = R_{left} + R_{right} + G_m R_{left} R_{right}$$

$$R_{left} \cong R_B \parallel r_\pi (1 + g_m R_E)$$

$$R_{right} \cong R_C$$

$$G_m = \frac{g_m}{1 + g_m R_E}$$

## Bandwidth Estimate for Degenerated CE Stage (1)



$$\begin{aligned} R_\mu &= R_s \parallel r_\pi (1 + g_m R_E) + R_C + G_m [R_s \parallel r_\pi (1 + g_m R_E)] R_C \\ &\cong R_s + R_C + G_m R_C R_s = R_s (1 + |A_{v0}|) + R_C \end{aligned}$$

$$R_\pi \cong r_\pi \parallel \frac{R_s + R_E}{1 + g_m R_E} \cong \frac{R_s + R_E}{1 + g_m R_E}$$

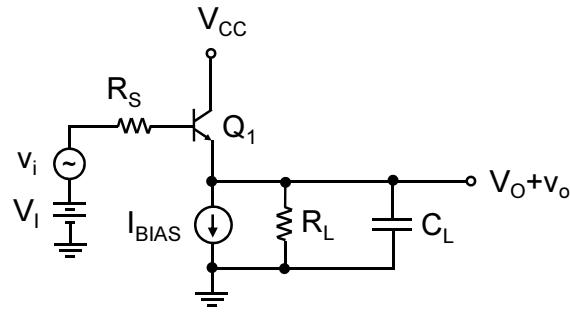
## Bandwidth Estimate for Degenerated CE Stage (2)

$$\tau = \left[ R_s (1 + |A_{v0}|) + R_C \right] C_\mu + \frac{1 + \frac{R_E}{R_s}}{1 + g_m R_E} R_s C_\pi \quad \omega_{-3dB} \cong \frac{1}{\tau}$$

- Compare to the case of  $R_E = 0$
- $\tau \cong [R_s (1 + |A_{v0}|) + R_C] C_\mu + R_s C_\pi$
- Adding  $R_E$  can help improve the bandwidth, provided that  $g_m > 1/R_s$ 
  - Note, however, that  $g_m$  (and hence the power dissipation must be increased) to maintain the same  $A_{v0}$
- Furthermore, it is interesting to consider a special case where  $g_m R_E \gg 1$  and the time constant due to  $C_\mu$  is negligible

$$\tau \cong \left( 1 + \frac{R_s}{R_E} \right) \frac{C_\mu}{g_m} \quad \omega_{-3dB} \cong \omega_T / \left( 1 + \frac{R_s}{R_E} \right) \quad \text{Near } \omega_T \text{ for small } R_s$$

## Common-Collector Stage (Emitter Follower)



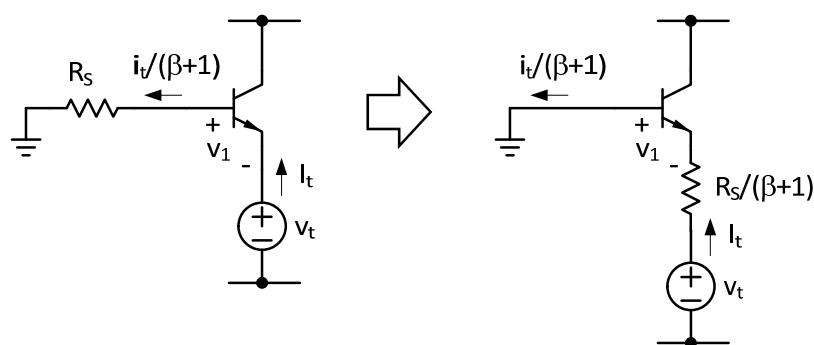
- Behavior is very similar to MOS common drain stage, except that
  - We do not need to worry about backgate effect
  - There is finite input resistance due to  $r_\pi$
  - The output resistance depends on  $R_S$  (in addition to  $1/g_m$ )

## Input and Output Resistance

- Input resistance (by inspection)

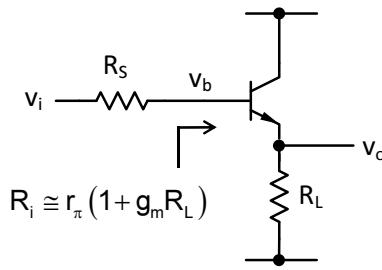
$$R_i \approx r_\pi (1 + g_m R_L)$$

- Output resistance (using push-through trick)



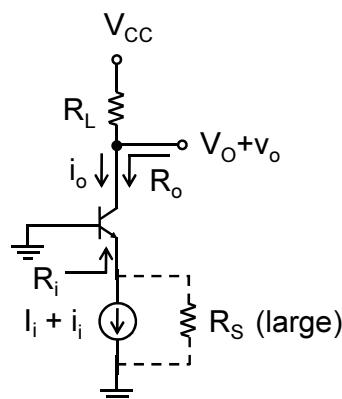
$$R_o \approx \frac{1}{g_m} + \frac{R_S}{\beta + 1} \approx \frac{1}{g_m} \left( 1 + \frac{R_S}{r_\pi} \right)$$

## Low Frequency Voltage Gain



$$\begin{aligned}
 A_{v0} &= \frac{v_o}{v_i} = \frac{v_b}{v_i} \frac{v_o}{v_b} \approx \frac{r_\pi(1+g_m R_L)}{r_\pi(1+g_m R_L) + R_s} \frac{g_m R_L}{1+g_m R_L} \\
 &\approx \frac{g_m R_L}{1+g_m R_L} \quad \text{for } r_\pi(1+g_m R_L) \gg R_s \\
 &\approx 1 \quad \text{for } g_m R_L \gg 1 \text{ and } r_\pi(1+g_m R_L) \gg R_s
 \end{aligned}$$

## Common-Base Stage



$$A_i = \frac{i_o}{i_i} \quad A_{io} = \frac{i_c}{i_E} = \frac{\beta}{\beta+1}$$

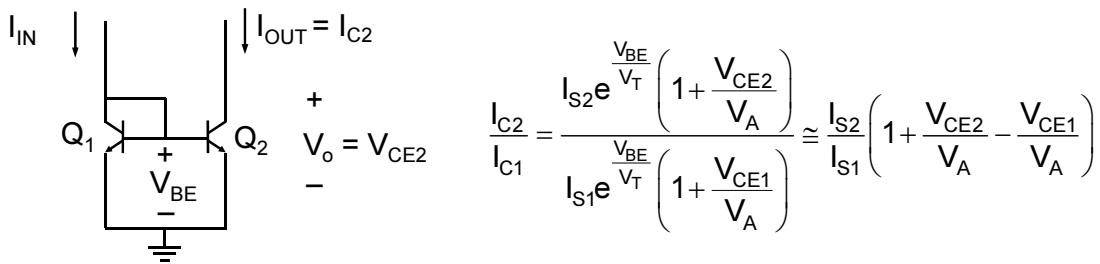
Neglecting  $r_b$ ,  $r_c$ ,  $r_e$  and  $r_\pi$ , we have

$$\begin{aligned}
 R_o &\approx r_o(1+g_m R_s) \\
 R_i &\approx \frac{1}{g_m} \left( 1 + \frac{R_L}{r_o} \right) \approx \frac{1}{g_m}
 \end{aligned}$$

Behavior is very similar to MOS common gate stage, except that

- We do not need to worry about backgate effect
- The DC current gain is not exactly unity, due to finite  $\beta$

## Basic BJT Current Mirror

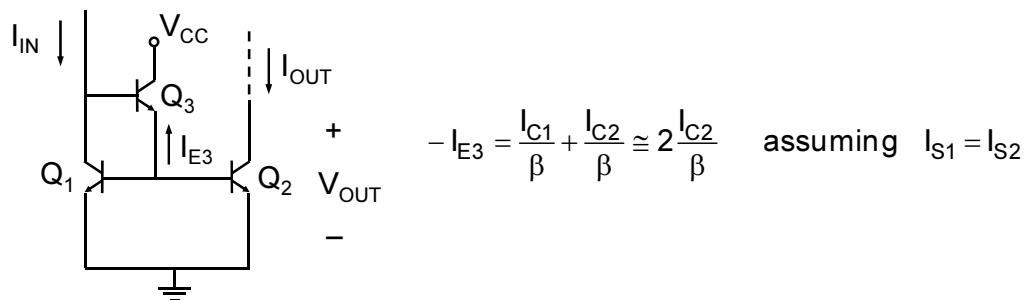


- Error due to base current

$$I_{IN} = I_{C1} + I_{B1} + I_{B2} = I_{C1} + \frac{I_{C1}}{\beta} + \frac{I_{C2}}{\beta} \approx I_{C2} \left(1 + \frac{2}{\beta}\right) \quad \text{for } I_{S1} = I_{S2}$$

$$\frac{I_{OUT}}{I_{IN}} \approx \frac{1}{\left(1 + \frac{2}{\beta}\right)} \approx 1 - \frac{2}{\beta}$$

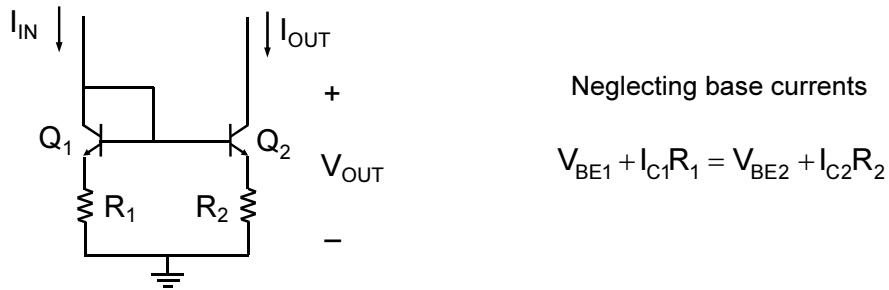
## BJT Current Mirror with “Beta Helper”



$$I_{B3} = -\frac{I_{E3}}{\beta + 1} \approx \frac{2I_{C2}}{\beta(\beta + 1)} \quad I_{IN} = I_{C1} + I_{B3} \approx I_{C1} + \frac{2I_{C2}}{\beta(\beta + 1)} \approx I_{C2} \left[1 + \frac{2}{\beta(\beta + 1)}\right]$$

$$\frac{I_{OUT}}{I_{IN}} \approx \frac{1}{1 + \left(\frac{2}{\beta^2 + \beta}\right)} \approx 1 - \frac{2}{\beta^2}$$

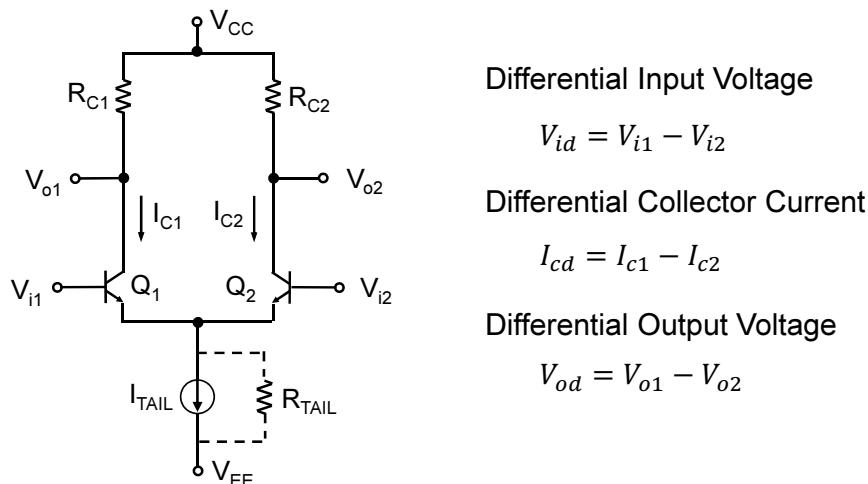
## BJT Current Mirror with Degeneration



$$I_{C2} = \frac{1}{R_2} \left\{ I_{C1}R_1 + V_T \ln \left[ \left( \frac{I_{C1}}{I_{C2}} \right) \left( \frac{I_{S2}}{I_{S1}} \right) \right] \right\} \cong I_{C1} \frac{R_1}{R_2} \quad \frac{I_{OUT}}{I_{IN}} \cong \frac{R_1}{R_2}$$

- Degeneration brings two benefits
  - Increased output resistance
  - Reduces sensitivity of mirror ratio to mismatches in  $I_S$
- However, the minimum  $V_{OUT}$  for which  $Q_2$  remains forward active is increased

## BJT Differential Pair



- The following large signal analysis neglects  $r_b$ ,  $r_c$ ,  $r_e$ , finite  $R_{EE}$  and assumes that the circuit is perfectly symmetric

## Large Signal Analysis

$$V_{i1} - V_{be1} + V_{be2} - V_{i2} = 0 \quad I_{C1} \approx I_{S1} e^{\frac{V_{be1}}{V_T}} \quad I_{C2} \approx I_{S2} e^{\frac{V_{be2}}{V_T}}$$

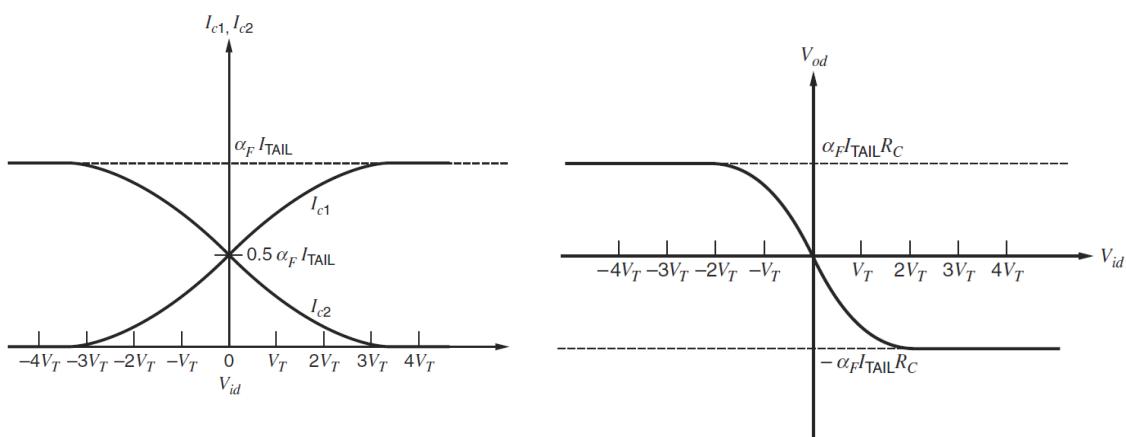
$$\Rightarrow \frac{I_{c1}}{I_{c2}} = e^{\frac{V_{be1}-V_{be2}}{V_T}} = e^{\frac{V_{i1}-V_{i2}}{V_T}} = e^{\frac{V_{id}}{V_T}}$$

$$I_{TAIL} = -(I_{e1} + I_{e2}) = \frac{1}{\alpha} (I_{c1} + I_{c2}) \quad \Rightarrow I_{c1} = \frac{\alpha I_{TAIL}}{1 + e^{-\frac{V_{id}}{V_T}}} \quad I_{c2} = \frac{\alpha I_{TAIL}}{1 + e^{+\frac{V_{id}}{V_T}}}$$

$$I_{cd} = I_{c1} - I_{c2} = \alpha_F I_{TAIL} \left[ \frac{1}{1 + e^{-\frac{V_{id}}{V_T}}} - \frac{1}{1 + e^{+\frac{V_{id}}{V_T}}} \right] = \alpha_F I_{TAIL} \tanh\left(\frac{V_{id}}{2V_T}\right)$$

$$V_{od} = I_{od} R_L = \alpha I_{TAIL} R_L \tanh\left(\frac{V_{id}}{2V_T}\right)$$

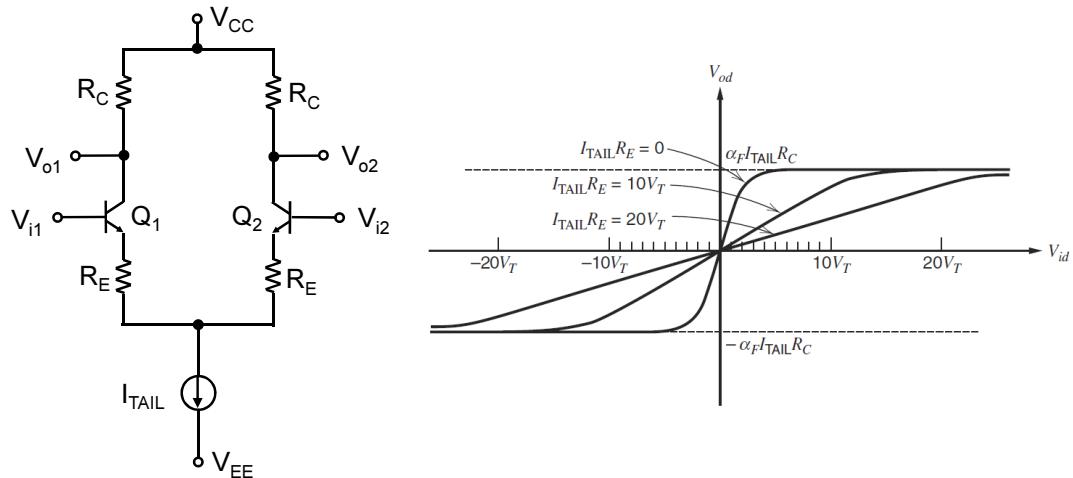
## Plot of Transfer Characteristics



- Linear region in  $V_{od}$  vs.  $V_{id}$  characteristic is narrow compared to MOS
  - Recall that full steering in a MOS pair occurs for  $V_{id} = \sqrt{2}V_{OV}$
- BJT differential pair is linear only for  $|V_{id}| < V_T \approx 26 \text{ mV}$

## Emitter Degeneration

- Can use emitter degeneration resistors to increase the range of input voltage over which the transfer characteristic of the pair is linear
- For large  $R_E$ , linear range is approximately equal to  $I_{TAIL}R_E$

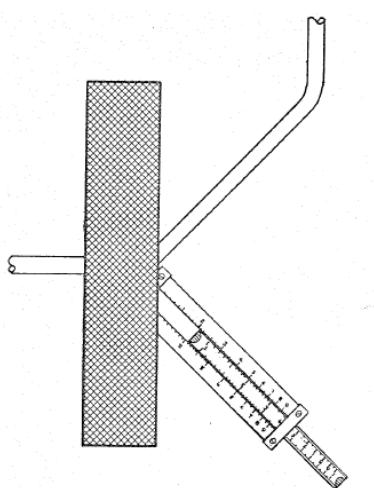


## Translinear Circuits

MORE ABOUT  
TRANSLINEAR  
CIRCUITS

THESE NOTES BEGAN 3rd SEPTEMBER  
AND COMPLETED AT GUNPOINT SOMETIME  
NEAR THE END OF 1978. Barrie Gilbert

TRANSLINEAR CIRCUITS FORM A DISTINCT CLASS OF BIPOLAR ANALOG CIRCUITS USEFUL IN PERFORMING A WIDE VARIETY OF LINEAR AND NONLINEAR SIGNAL MANIPULATION AND DESERVING OF BETTER FAMILIARITY IN THEIR OWN RIGHT. A SOMEWHAT RANDOM CROSS-SECTION OF REAL-LIFE EXAMPLES SHOULD DEMONSTRATE THEIR APPLICATIONS & UTILITY.



HAPPINESS IS....

A Logarithmic  $V_{BE}$

- “A translinear circuit is one having all inputs and outputs in the form of currents and whose primary function arises from the exploitation of the logarithmic behavior of forward-biased PN junctions arranged in pairs so as to result in fundamentally exact, temperature independent transformations in the amplitude domain”
- Most basic example:

$$V_{BE1} = V_{BE2}$$

$$V_T \ln\left(\frac{I_i}{I_{s1}}\right) = V_T \ln\left(\frac{I_o}{I_{s2}}\right)$$

$$\frac{I_i}{I_{s1}} = \frac{I_o}{I_{s2}}$$

$$I_o = \frac{I_{s2}}{I_{s1}} I_i$$

## Square-Root Circuit

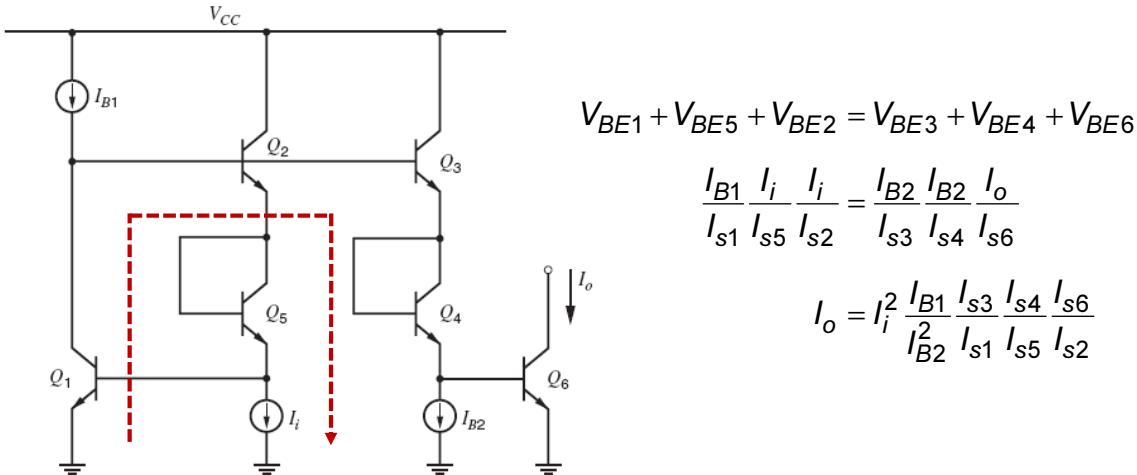
$$V_{BE1} + V_{BE2} = V_{BE3} + V_{BE4}$$

$$V_T \ln\left(\frac{I_B}{I_{s1}}\right) + V_T \ln\left(\frac{I_i}{I_{s2}}\right) = V_T \ln\left(\frac{I_o}{I_{s3}}\right) + V_T \ln\left(\frac{I_o}{I_{s4}}\right)$$

$$\frac{I_B}{I_{s1}} \frac{I_i}{I_{s2}} = \frac{I_o}{I_{s3}} \frac{I_o}{I_{s4}}$$

$$I_o = \sqrt{I_i} \sqrt{I_B} \sqrt{\frac{I_{s3}}{I_{s1}}} \sqrt{\frac{I_{s4}}{I_{s2}}}$$

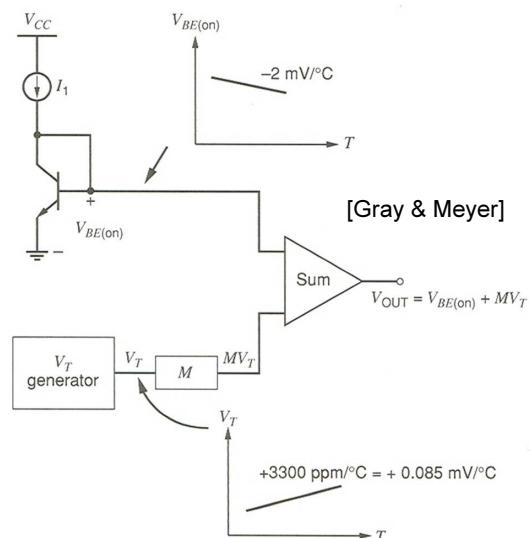
## Squaring Circuit



- Many more examples exist
- It is possible, but often not practical, to play similar tricks with MOSFETs
  - Key issues: offset voltages, deviation from idealized I-V law

## Bandgap Reference — Key Idea

- $V_{BE}$  of a BJT decreases with temperature
  - "CTAT" — complementary to absolute temperature
- $kT/q$  has a positive temperature coefficient
  - "PTAT" — proportional to absolute temperature
- Can combine PTAT + CTAT to a reference voltage with "near zero" temperature dependence
  - Useful in circuits that require a stable reference voltage
    - E.g. A/D converters



## A Closer Look at $V_{BE}$

$$V_{BE} = V_T \ln \left( \frac{I_C}{I_S} \right) \quad V_T = \frac{kT}{q} \quad I_S = \frac{qAD_n n_i^2}{W_B N_A}$$

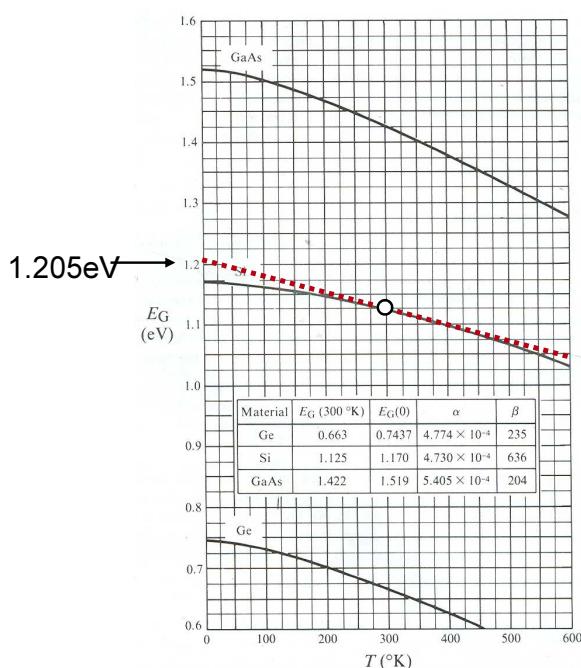
- Even though  $V_T$  increases with temperature,  $V_{BE}$  decreases because  $I_S$  increases rapidly with temperature (due to the  $n_i$  term)

$$I_S \cong I_0 e^{-\frac{V_{G0}}{V_T}}$$

$$V_{BE} \cong V_T \ln \left( \frac{I_C}{I_0} e^{-\frac{V_{G0}}{V_T}} \right) = V_{G0} - V_T \ln \left( \frac{I_C}{I_0} \right)$$

- $I_0$  is a device parameter, which is not completely independent of temperature (we'll ignore this)
- $V_{G0}$  is the bandgap voltage of silicon "extrapolated to 0°K"

## Extrapolated Bandgap



Pierret  
Advanced Semiconductor  
Fundamentals Page 85

$$V_{G0} = \frac{1.205eV}{q} = 1.205V$$

## Temperature Coefficient of $V_{BE}$

$$V_{BE} \cong V_{G0} - V_T \ln \left( \frac{I_C}{I_0} \right)$$

- Assuming that both  $I_0$  and  $I_C$  are constant over temperature:

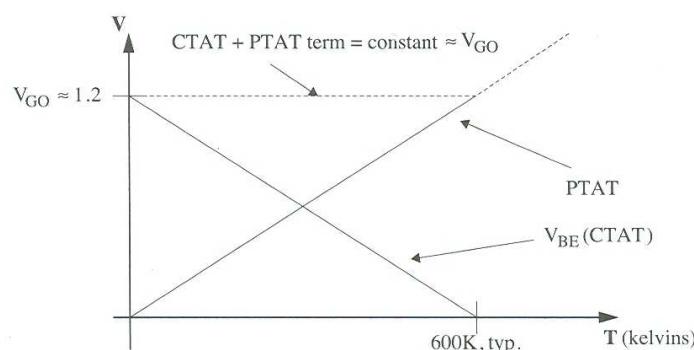
$$\frac{dV_{BE}}{dT} \cong \frac{V_T}{T} \ln \left( \frac{I_C}{I_0} \right) = \frac{V_{BE} - V_{G0}}{T}$$

- Example:

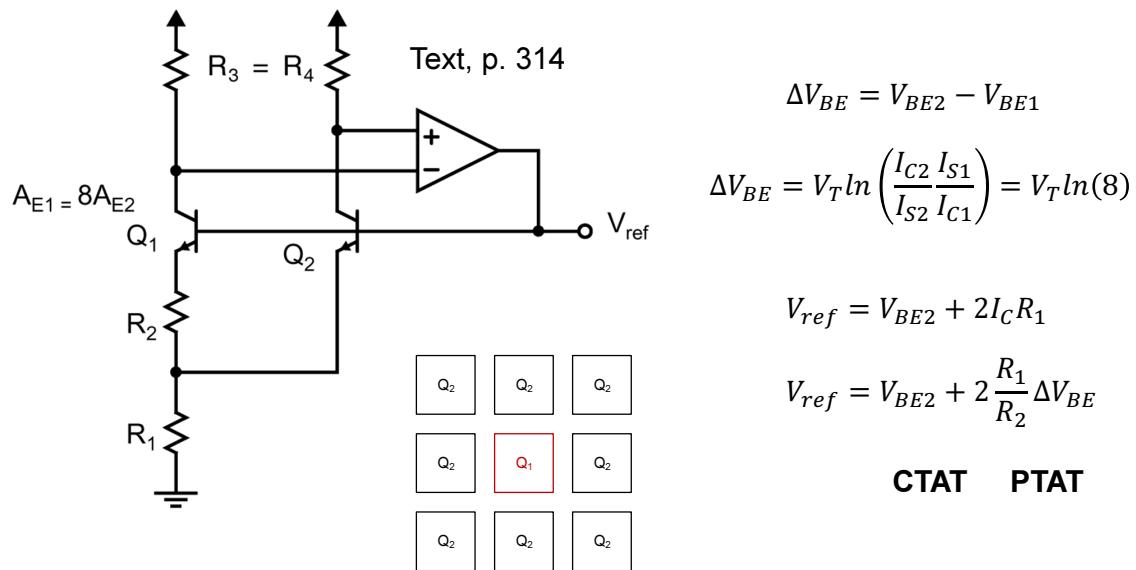
$$\frac{dV_{BE}}{dT} \cong \frac{0.6V - 1.205V}{300K} = -2.02 \frac{mV}{K}$$

## CTAT + PTAT

- We now want to build a circuit that adds a properly scaled version of  $V_T$  to  $V_{BE}$  so that a temperature independent voltage results
- No matter how we do this, the resulting voltage must be equal to  $V_{G0}$ 
  - This is by far the simplest criterion for finding the proper design parameters (no need to grind through derivatives)



## Classical Brokaw Bandgap



## Design Example

- From measurement data, we know that  $|V_{BE}|$  of a unit device is 600mV at room temperature and  $I_C = 50\mu A$

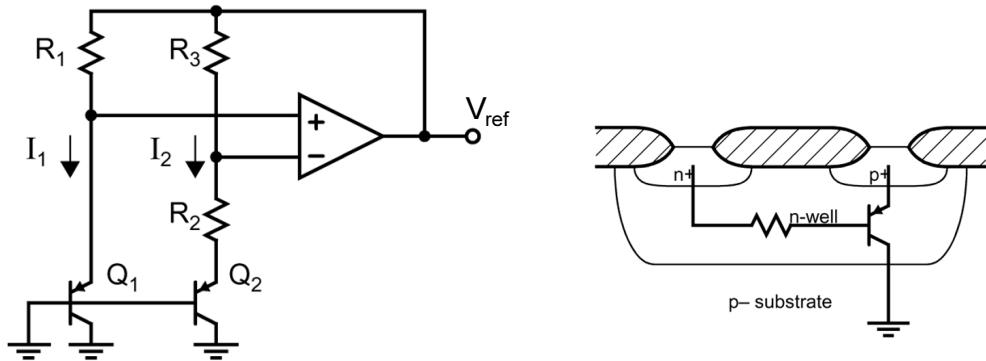
$$V_{ref} = V_{BE2} + 2 \frac{R_1}{R_2} \Delta V_{BE} = V_{G0}$$

$$\frac{R_1}{R_2} = \frac{1}{2} \frac{V_{G0} - V_{BE2}}{\Delta V_{BE}} = \frac{1}{2} \cdot \frac{1.205V - 0.6V}{26mV \cdot \ln(8)} = 5.6$$

$$R_2 = \frac{\Delta V_{BE}}{I_C} = \frac{26mV \cdot \ln(8)}{50\mu A} = 1.08k\Omega$$

$$R_1 = 6.06k\Omega$$

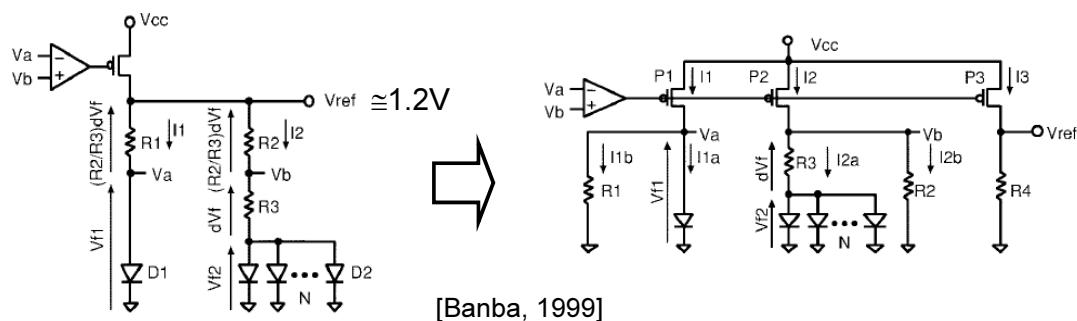
## Bandgap Using Substrate PNP Transistors



- Typical approach in a standard CMOS process where no dedicated BJTs are available
- Need to watch out for base resistance (keep currents small)

## Sub-1V Bandgap Reference

- Idea: Add currents proportional to V<sub>BE</sub> and V<sub>T</sub>, instead of adding voltages



$$V_{ref} = R_4 \left( \frac{V_{BE}}{R_2} + \frac{V_T \ln(N)}{R_3} \right)$$

## Nonidealities

- Temperature dependence of  $I_0$ , leading to “curvature”
  - Opamp offset voltages and their temperature coefficients
  - Resistor mismatch and temperature coefficient
  - Finite  $\beta$  and  $\beta$  mismatch
  - ...
- 
- Consequences
    - The voltage with minimum temperature coefficient typically lies several  $kT/q$  above  $V_{GO}$
    - The residual temperature dependence of mass-produced bandgaps is on the order of 20...50 ppm/K

## Selected References

- R. J. Widlar, "New developments in IC voltage regulators," IEEE J. Solid-State Circuits, pp. 2-7, Feb. 1971.
  - First report, LM309 5V regulator
- A. P. Brokaw, "A simple three-terminal IC bandgap reference," IEEE J. Solid-State Circuits, pp. 388-393, Dec. 1974.
  - A classic implementation
- C. Palmer and R. Dobkin, "A curvature corrected micropower voltage reference," IEEE Int. Solid-State Conference, pp. 58-59, Feb. 1981.
- G. Nicollini et al., "A CMOS bandgap reference for differential signal processing," IEEE J. Solid-State Circuits, pp. 41-50, Jan. 1991.
- T.L. Brooks et al., "A low-power differential CMOS bandgap reference," IEEE Int. Solid-State Conf., pp. 248-249, Feb. 1994.
- H. Banba et al. "A CMOS bandgap reference circuit with sub-1-V operation," IEEE J. Solid-State Circuits, pp. 670 - 674, May 1999.
- P. Malcovati et al., "Curvature-compensated BiCMOS bandgap with 1-V supply voltage," IEEE J. Solid-State Circuits, pp. 1076-1081, July 2001.

# **Chapter 4**

## **MOS Transistor Modeling**

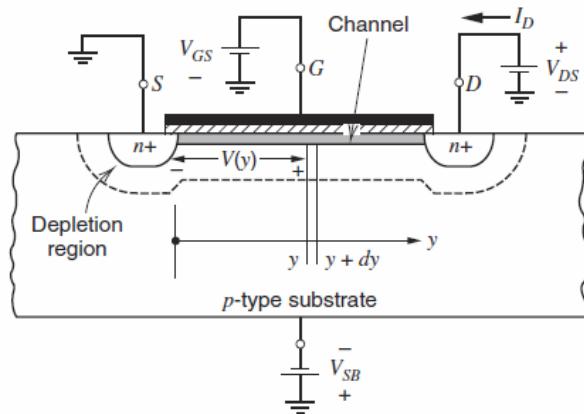
**Boris Murmann  
Stanford University  
Winter 2015-16**

**Textbook Sections: 1.4, 1.5**

### **Outline**

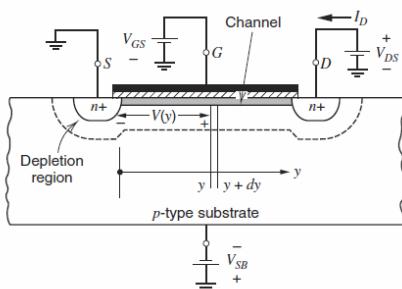
- Review of square-law model
- Inspection of EE214B MOSFETs ( $0.18\mu\text{m}$ )
- Analysis of relevant effects
  - Weak inversion
  - Short channel effects

## Basic MOSFET Operation (NMOS)



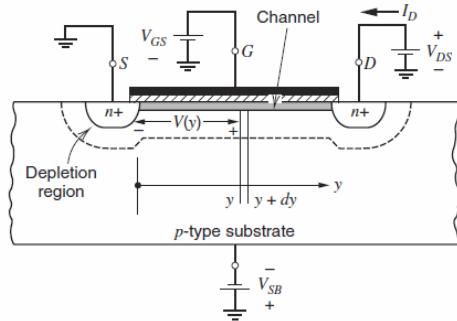
How to calculate the drain current ( $I_D$ ) current as a function of  $V_{GS}$ ,  $V_{DS}$ ?

## Simplifying Assumptions



- 1) Current is controlled by the mobile charge in the channel
- 2) Gradual channel approximation - the vertical field sets channel charge, so we can approximate the mobile charge through the voltage difference between the gate and the channel
- 3) The carrier velocity is proportional to the lateral field ( $v = \mu E$ ). This is equivalent to Ohm's law: velocity (current) is proportional to E-field (voltage)

## Derivation of Square-Law Characteristics



$$Q_n(y) = C_{ox} [V_{GS} - V(y) - V_t]$$

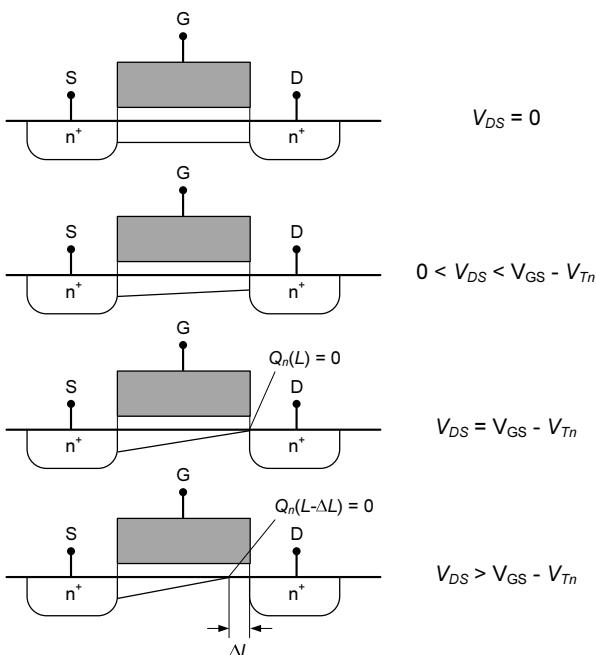
$$I_D = Q_n \cdot v \cdot W$$

$$v = \mu \cdot E$$

$$I_D = C_{ox} [V_{GS} - V(y) - V_t] \cdot \mu \cdot E \cdot W$$

$$I_D = \mu C_{ox} \frac{W}{L} \left[ (V_{GS} - V_t) - \frac{V_{DS}}{2} \right] \cdot V_{DS}$$

## Drain Current Saturation and Channel Length Modulation



$$Q_n(y) = C_{ox} [V_{GS} - V(y) - V_t]$$

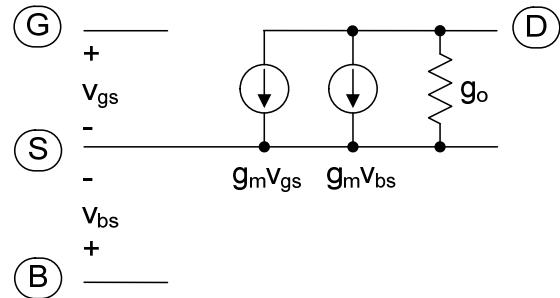
- Channel pinch-off occurs for  $V_{DS} \geq V_{GS} - V_t$  and the drain current saturates

- The saturation current has some dependence on  $V_{DS}$  due to channel length modulation

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

“Lambda Model”

## Small-Signal Model

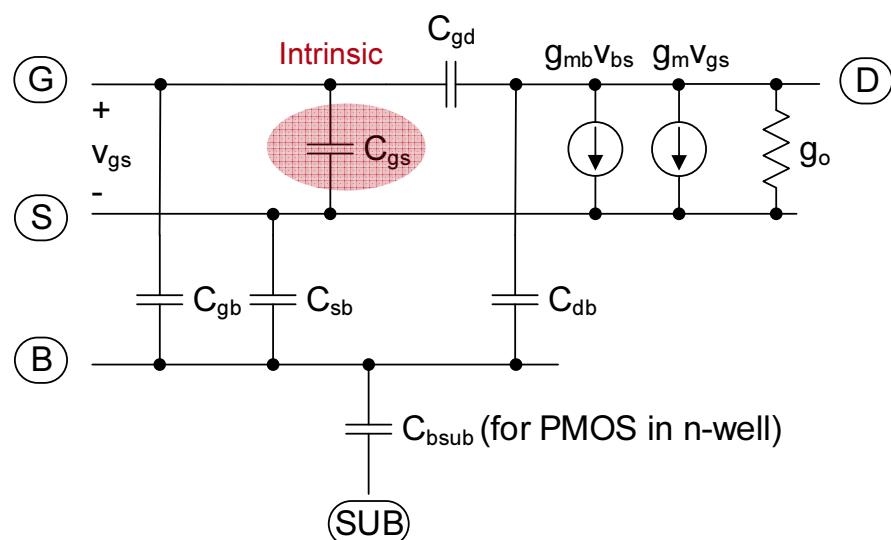


$$g_m = \frac{dI_D}{dV_{GS}} = \mu C_{ox} \frac{W}{L} V_{OV} (1 + \lambda V_{DS}) = \frac{2I_D}{V_{OV}} \quad V_{OV} = V_{GS} - V_t$$

$$g_o = \frac{dI_D}{dV_{DS}} = \frac{1}{2} \mu C_{ox} \frac{W}{L} V_{OV}^2 \cdot \lambda = \frac{\lambda I_D}{1 + \lambda V_{DS}} \cong \lambda I_D$$

$$g_{mb} = \frac{dI_D}{dV_{BS}} = \frac{g_m \gamma}{2\sqrt{2\phi_f + V_{SB}}}$$

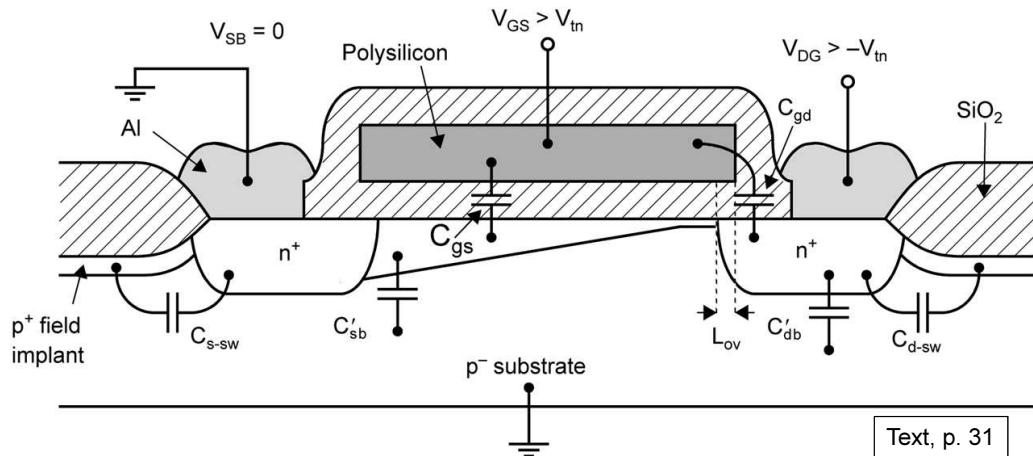
## Capacitances



$$C_{gg} = C_{gs} + C_{gb} + C_{gd}$$

$$C_{dd} = C_{db} + C_{gd}$$

## Capacitances

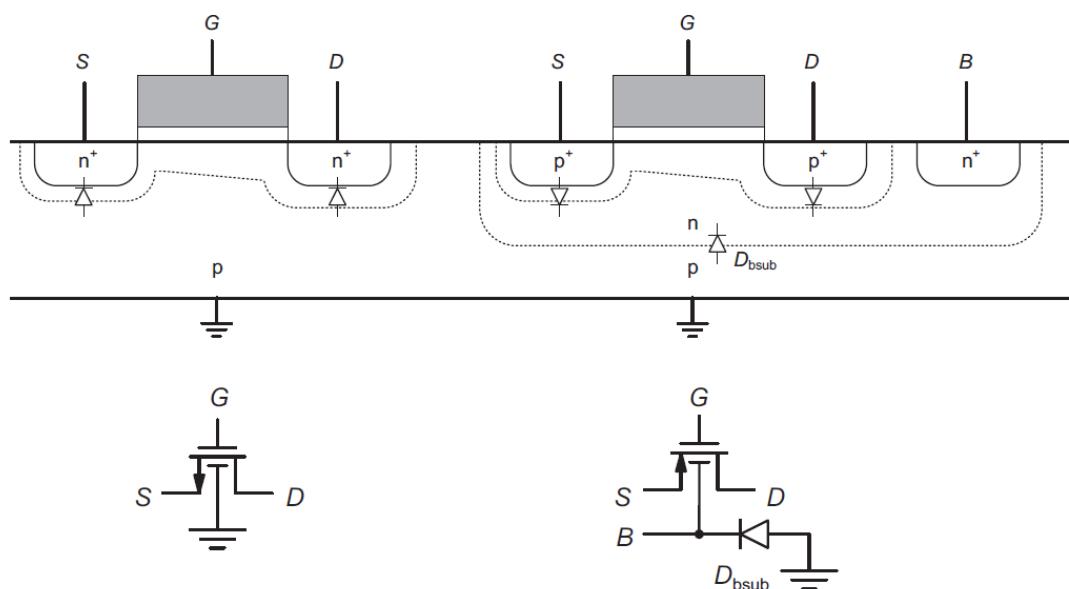


$$C_{db} = \frac{AD \cdot C_J}{\left(1 + \frac{V_{DB}}{PB}\right)^{MJ}} + \frac{PD \cdot C_{JSW}}{\left(1 + \frac{V_{DB}}{PB}\right)^{MJSW}}$$

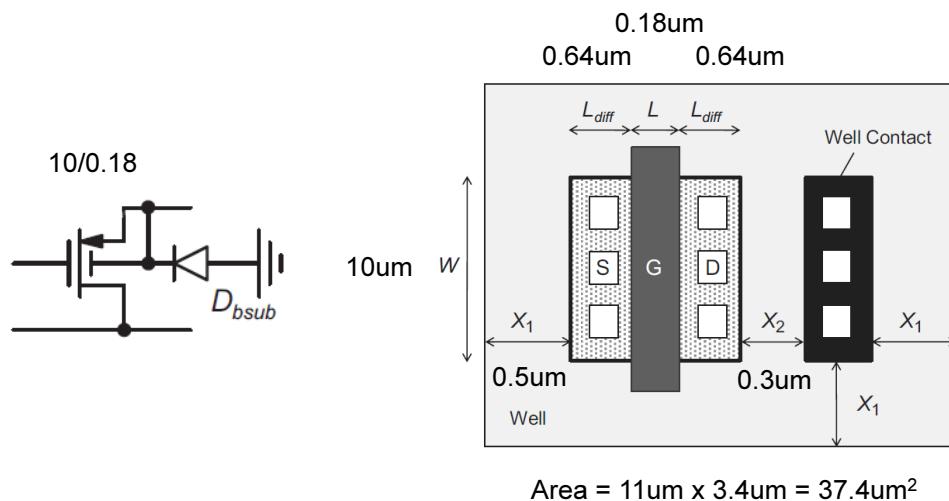
$$AD = WL_{diff}$$

$$PD = W + 2L_{diff}$$

## Well Capacitance (PMOS)



## Example



```
* HSpice Netlist
*
* d g s b
mp1 0 in out out pmos214 L=0.18um W=10um
d1 0 out      dwell    37.4p
```

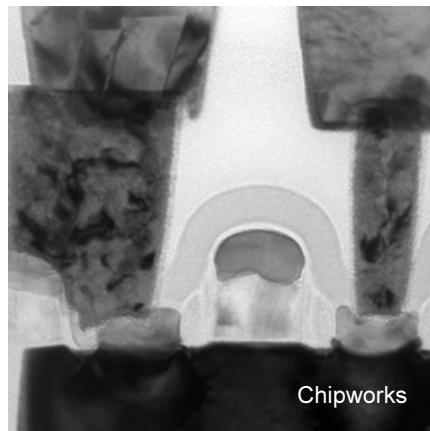
## Gate Capacitance Summary

	Subthreshold	Triode	Saturation
$C_{gs}$	$C_{ol}$	$\frac{1}{2} WLC_{ox} + C_{ol}$	$\frac{2}{3} WLC_{ox} + C_{ol}$
$C_{gd}$	$C_{ol}$	$\frac{1}{2} WLC_{ox} + C_{ol}$	$C_{ol}$
$C_{gb}$	$\left( \frac{1}{C_{js}} + \frac{1}{WLC_{ox}} \right)^{-1}$	0	0

$$C_{ol} = WC_{ol}'$$

“Overlap Capacitance”

## Aside: “Overlap” Capacitance of a 28nm MOSFET



- For the most recent generation of MOS transistors, the overlap and fringe capacitances from gate to drain/source are about as large as the intrinsic gate capacitance!

## Capacitance Parameters for EE214B Technology

PARAMETER	EE 214B Technology ( $0.18\mu\text{m}$ )	
	NMOS	PMOS
$C_{\text{ox}}$	$8.42 \text{ fF}/\mu\text{m}^2$	$8.42 \text{ fF}/\mu\text{m}^2$
$C'_{\text{ol}}$	$0.491 \text{ fF}/\mu\text{m}$	$0.657 \text{ fF}/\mu\text{m}$
$C_J$	$0.965 \text{ fF}/\mu\text{m}^2$	$1.19 \text{ fF}/\mu\text{m}^2$
$C_{\text{JSW}}$	$0.233 \text{ fF}/\mu\text{m}$	$0.192 \text{ fF}/\mu\text{m}$
$C_{\text{Jwell}}$	—	$0.2 \text{ fF}/\mu\text{m}^2$
PB	$0.8 \text{ V}$	$0.8 \text{ V}$
MJ	0.38	0.40
MJSW	0.13	0.33
LDIF	$0.64 \mu\text{m}$	$0.64 \mu\text{m}$

## What are $\mu C_{ox}$ ("KP") and $\lambda$ ("LAMBDA") for our Technology?

```
.MODEL nmos214 nmos
+ACM = 3          hdif = 0.32e-6      LEVEL = 49
+VERSION = 3.1    TNOM = 27           TOX = 4.1E-9
+XJ = 1E-7         NCH = 2.3549E17   VTH0 = 0.3618397
+K1 = 0.5916053   K2 = 3.225139E-3  K3 = 1E-3
+K3B = 2.3938862   W0 = 1E-7        NLX = 1.776268E-7
+DVTOW = 0         DVTIW = 0       DVT2W = 0
+DVT0 = 1.3127368 DVT1 = 0.3876801 DVT2 = 0.0238708
+U0 = 256.74093   UA = -1.585658E-9 UB = 2.528203E-18
+UC = 5.182125B-11 VSAT = 1.003268E5 AO = 1.981392
+AGS = 0.4347252   BO = 4.989266E-7 B1 = 5E-6
+KETA = -9.888408E-3 A1 = 6.164533E-4 A2 = 0.9388917
+RDSW = 128.705483 PRWG = 0.5     PRWB = -0.2
+NR = 1            WINT = 0       LINT = 1.617316E-8
+XL = 0            XW = -1E-8     DWG = -5.383413E-9
+IWB = 9.111767E-9 VOFF = -0.0854824 NFACCTOR = 2.2420572
+CIT = 0            CDSC = 2.4E-4   CDSCD = 0
+CDSCB = 0          ETAO = 2.981159E-3 ETAB = 9.289544E-6
+DSUB = 0.0159753   PC1M = 0.7245546 PDIBLC1 = 0.1568183
+PDIBLC2 = 2.543351E-3 PDIBLCB = -0.1 DRROUT = 0.7445011
+PSCE1 = 6E10       PSCE2 = 1.876443E-9 PVAG = 7.200284E-3
+DELTA = 0.01       RSH = 6.6      MOBMOD = 1
+FRT = 0            UTE = -1.5    KT1 = -0.11
+KT1L = 0           KT2 = 0.022   UA1 = 4.31E-9
+UB1 = -7.61E-18   UC1 = -5.6E-11 AT = 3.3E4
+WL = 0             WLN = 1      WW = 0
+WNW = 1            NWL = 0      LL = 0
+LLN = 1            LW = 0       LNN = 1
+LML = 0            CAPMOD = 2   XPART = 1
+CGDO = 4.91E-10   CGSO = 4.91E-10 CGBO = 1E-12
+CJ = 9.652028E-4  PB = 0.8     MJ = 0.3836899
+CJSW = 2.326465E-10 PBSW = 0.8   MJSW = 0.1253131
+CJSWG = 3.3E-10   PBSWG = 0.8  MJSWG = 0.1253131
+CF = 0             PVTHO = -7.714081E-4 PRDSW = -2.5827257
+FK2 = 9.619963E-4 WKETA = -1.060423E-4 LKETA = -5.373522E-3
+PU0 = 4.5760891   PUA = 1.4690208E-14 PUB = 1.783193E-23
+PVSAT = 1.19774E3 PETAO = 9.968409E-5 PKETA = -2.51194E-3
+nlev = 3           kf = 0.5e-25
```

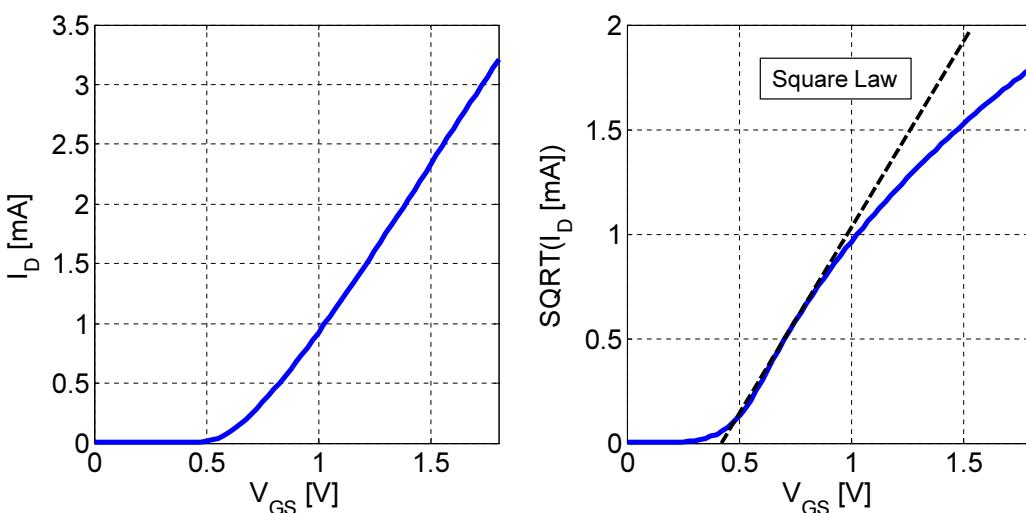
- The HSpice model for an NMOS device in our technology is shown to the left
- This is a 110-parameter BSIM3v3 model
  - More recent models may require even more parameters (e.g. PSP, BSIM6)
  - KP and LAMBDA are nowhere to be found
- It turns out that the I-V characteristics of a modern MOSFET cannot be accurately described by the square law

B. Murmann

EE214B Winter 2015-16 – Chapter 4

15

## Simulation (NMOS, 5/0.18μm, $V_{DS}=1.8V$ )



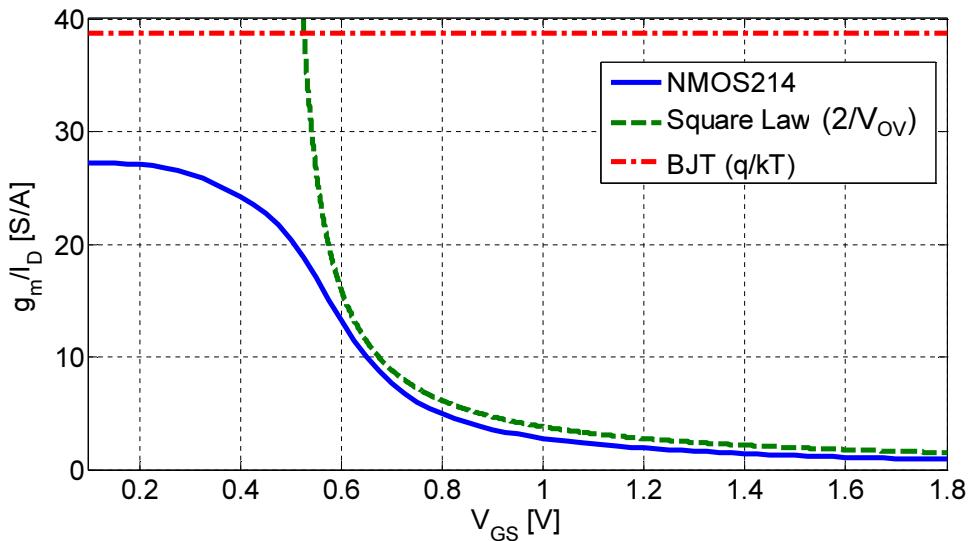
- Two observations
  - The transistor does not abruptly turn off at some  $V_t$
  - The current is not perfectly quadratic in  $(V_{GS} - V_t)$

B. Murmann

EE214B Winter 2015-16 – Chapter 4

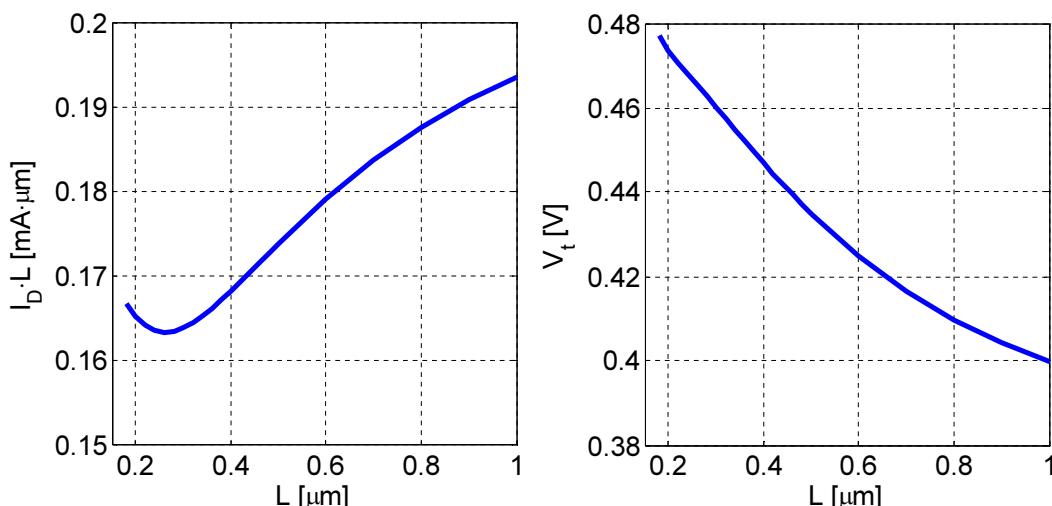
16

## Additional Issues (1)



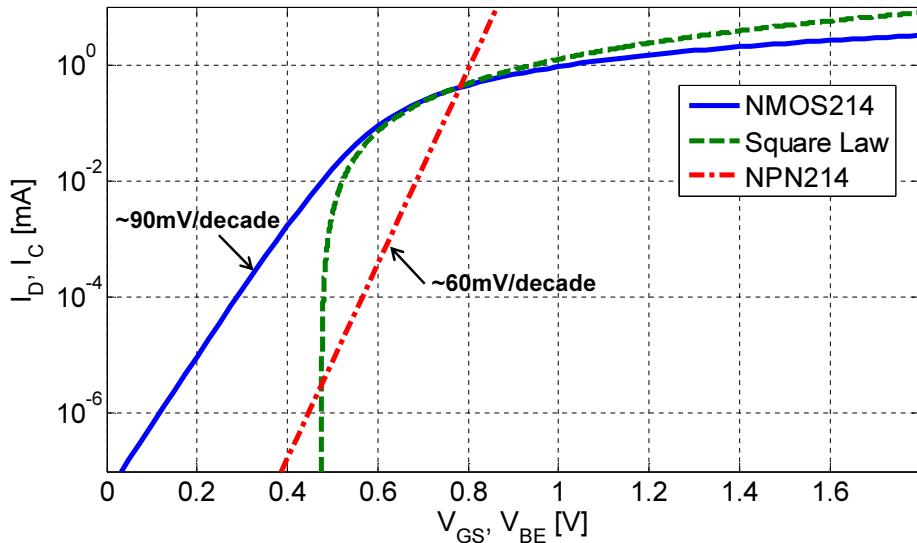
- The square law fails miserably at predicting  $g_m/I_D$  for low  $V_{GS}$

## Additional Issues (2)



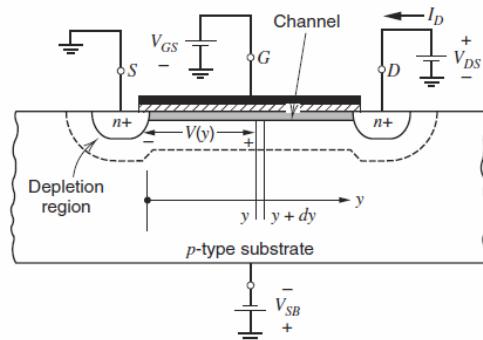
- The current does not scale perfectly with  $1/L$  ( $I_D \cdot L \neq \text{const.}$ )
- The threshold voltage of the device depends on the channel length

## Currents on a Log Scale



- What is  $V_t$ , anyway? The device does not turn off at all, but really approaches an exponential IV law for low  $V_{GS}$
- What determines the current at low  $V_{GS}$ ?

## Definition of $V_t$

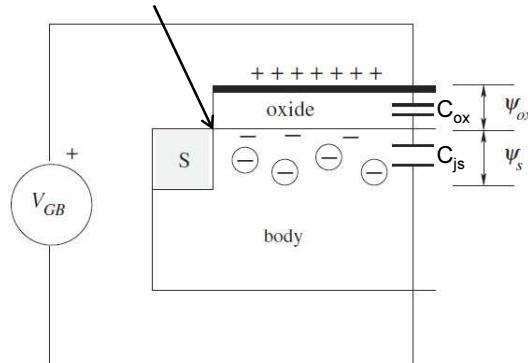


- $V_t$  is (roughly speaking) defined as the  $V_{GS}$  at which the number of electrons at the surface equals the number of doping atoms
- Seems arbitrary, but makes sense in terms of surface charge control
  - This is the point where the surface becomes inverted (no more holes to fill) and the relationship between mobile charge and gate voltage becomes linear,  $Q_n \propto C_{ox}(V_{GS} - V_t)$
  - Exactly what is assumed in the square law model

## Weak Inversion

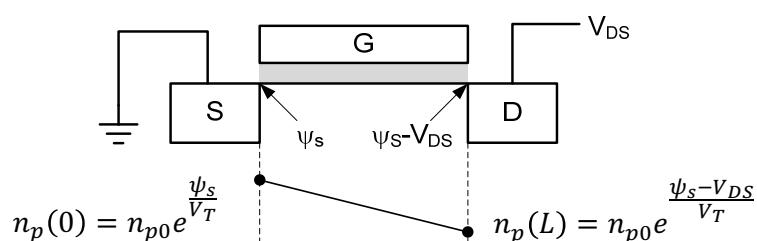
- Before inversion occurs, the electrostatic field from the gate forward-biases the source-side pn junction at the surface
- Physics governed by the “gated diode” model

Potential at this point is higher than body potential  $\rightarrow$  forward bias



D.L. Pulfrey, Understanding Modern Transistors and Diodes,  
Cambridge University Press, 2010.

## Resulting Diffusion Current



$$I_D = qAD_n \frac{n_p(0) - n_p(L)}{L}$$

$$I_D = \frac{1}{L} qAD_n n_p 0 e^{\frac{\psi_s}{V_T}} (1 - e^{-\frac{V_{DS}}{V_T}})$$

- The current grows exponentially with  $\psi_s$
- The current becomes independent of  $V_{DS}$  for  $V_{DS} > 3V_T$  ( $\sim 78mV$ )

## Capacitive Divider

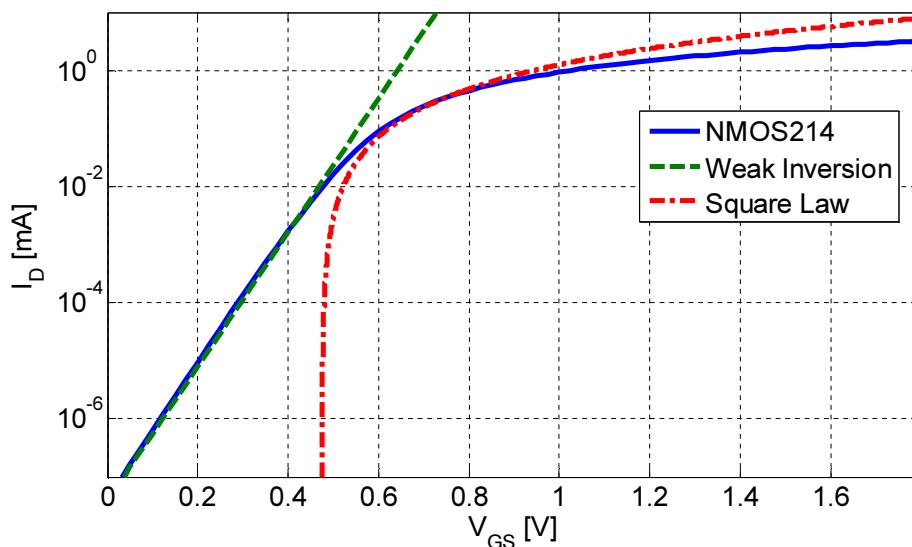
$$\frac{d\psi_s}{dV_{GS}} = \frac{C_{ox}}{C_{js} + C_{ox}} = \frac{1}{n}$$

- $n$  is called “subthreshold factor” or “nonideality factor”
- $n \approx 1.45$  for an NMOS device in the EE214B technology
- After including this relationship between  $\psi_s$  and  $V_{GS}$  and after a few additional manipulations, the final expression for the drain current becomes

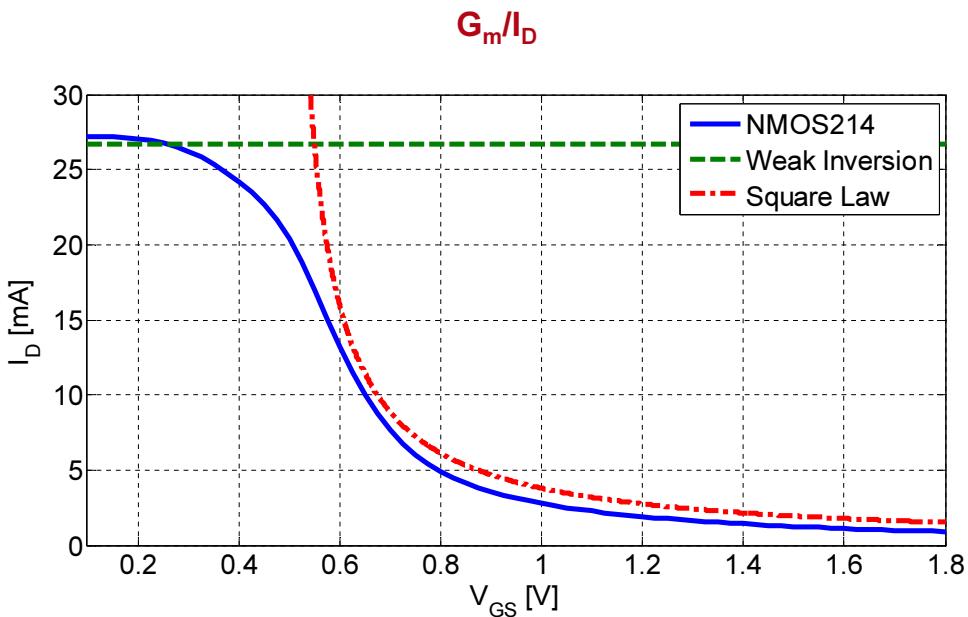
$$I_D = \frac{W}{L} I_{D0} e^{\frac{V_{GS}-V_t}{nV_T}} \left(1 - e^{-\frac{V_{DS}}{V_T}}\right)$$

where  $I_{D0}$  depends on technology ( $I_{D0} \approx 0.43\mu A$  for an NMOS device in EE214B technology)

## Combining the Weak Inversion Expression and Square Law



- Two remaining problems
  - The weak inversion expression and square law are disconnected
  - We still do not know what causes the discrepancies at high  $V_{GS}$



- We now have a better idea now about the maximum possible  $g_m/I_D$ , but we still do not know how to model the transition region between the two IV laws

## Moderate Inversion

- In the transition region between weak and strong inversion, the drain current consists of both drift and diffusion currents
- One can show that the ratio of drift/diffusion current in moderate inversion and beyond is approximately  $(V_{GS} - V_t)/(kT/q)$
- This means that the square law equation (which assumes 100% drift current) does not work unless the gate overdrive is several  $kT/q$ 
  - Recall that in EE214A, you used the square law model only for  $V_{GS} - V_t > 150\text{mV} \approx 6 kT/q$
- Is there a simple expression that works for all three regions (weak, moderate and strong inversion)?
- The so-called EKV model was developed with this intent
  - C. Enz and E. A. Vittoz, Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design, Wiley, 2006

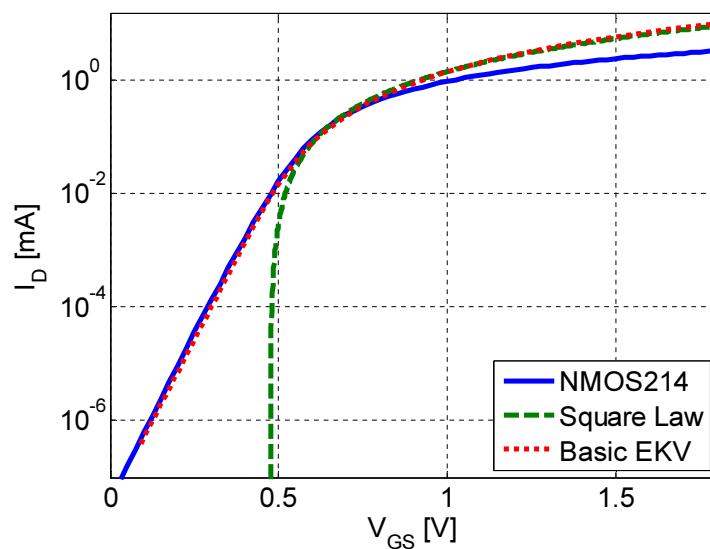
## Basic EKV Expression in Saturation

$$I_D = 2nV_T^2 \cdot \mu C_{ox} \frac{W}{L} \left( \underbrace{q_s^2}_{\text{Drift}} + \underbrace{\frac{q_s}{2}}_{\text{Diffusion}} \right)$$

$$V_{GS} - V_t = nV_T(2(q_s - 1) + \ln(q_s))$$

- This is a parametric equation set that cannot be solved directly
- Only three parameters:  $V_t$ ,  $n$ ,  $\mu C_{ox} W/L$
- $q_s$  is the normalized source charge density and typically runs from  $10^{-5}$  (weak inversion) to  $10\dots 100$  (strong inversion)
- It can be shown that the equation set approaches a square law expression for  $q_s \gg 1$  and the weak inversion expression for  $q_s \ll 1$

## Comparison



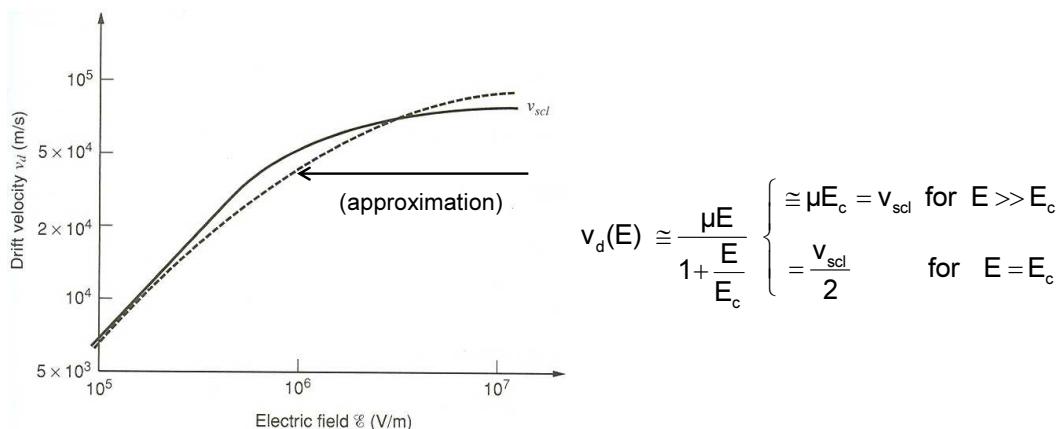
- The basic EKV model provides a proper transition region
- A more complex model is needed to iron out the remaining discrepancy at large  $V_{GS}$  (and to include the triode region, channel length modulation, etc.)

## Short Channel Effects

- The sub-square behavior at large  $V_{GS}$  is primarily due to a number of issues that fall under the category of “short channel effects”
  - Onset of velocity saturation due to high lateral field
  - Mobility degradation due to high vertical field
  - Strong  $V_{DS}$  dependence of drain current and output resistance
  - Threshold voltage depends on channel length and width
- Many more issues exist; we will once again only discuss the most relevant subset

## Velocity Saturation (1)

- In the derivation of the square law model, it is assumed that the carrier velocity is proportional to the lateral E-field,  $v = \mu E$
- Unfortunately, the speed of carriers in silicon is limited ( $v_{scl} \approx 10^5$  m/s)
  - At very high fields (high voltage drop across the conductive part of the channel), the carrier velocity saturates



## Velocity Saturation (2)

- It is important to distinguish the various regions in the above plot
  - Low field, the square law equation still hold
  - Moderate field, the square law becomes somewhat inaccurate
  - Very high field across the conducting channel – the velocity saturates completely and becomes essentially constant ( $v_{scl}$ )
- To get some feel for latter two cases, let's first estimate the E field using simple physics
- In saturation, for a transistor with  $V_{OV} = 200\text{mV}$ , the lateral field across the conducting part of the channel is

$$E = \frac{V_{OV}}{L} \quad \text{e.g.} \quad \frac{200\text{mV}}{0.18\mu\text{m}} = 1.11 \cdot 10^6 \frac{\text{V}}{\text{m}}$$

## Field Estimates

- In our  $0.18\mu\text{m}$  technology, we have for an NMOS device

$$E_c = \frac{v_{scl}}{\mu} \approx \frac{10^5 \frac{\text{m}}{\text{s}}}{150 \frac{\text{cm}^2}{\text{Vs}}} = 6.7 \cdot 10^6 \frac{\text{V}}{\text{m}}$$

Therefore

$$\frac{E}{E_c} = \frac{1.11 \cdot 10^6 \frac{\text{V}}{\text{m}}}{6.7 \cdot 10^6 \frac{\text{V}}{\text{m}}} \approx 0.16$$

- This means that for  $V_{OV}$  on the order of  $200\text{mV}$ , the carrier velocity is somewhat reduced, but the impairment is relatively small
- The situation changes when much larger  $V_{OV}$  are applied, as the case in digital circuits

## Short Channel $I_D$ Equation

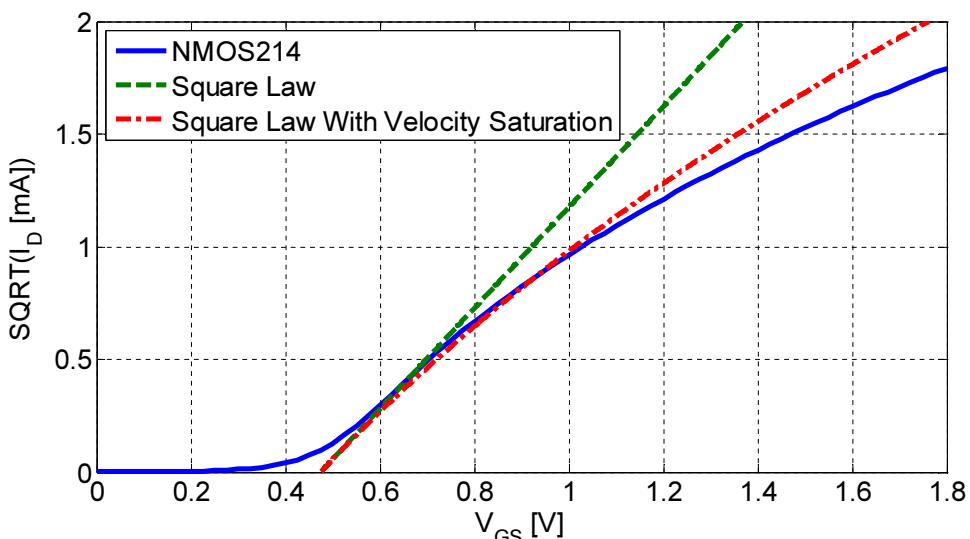
- A simple equation that captures the moderate deviation from the long channel drain current can be written as

$$I_D \approx \frac{1}{2} \mu C_{ox} \frac{W}{L} V_{ov}^2 \cdot \frac{1}{\left(1 + \frac{V_{ov}}{E_c L}\right)}$$

Minimum-length NMOS:  $E_c L = 6.7 \cdot 10^6 \frac{V}{m} \cdot 0.18 \mu m = 1.2V$

Minimum-length PMOS:  $E_c L = 16.75 \cdot 10^6 \frac{V}{m} \cdot 0.18 \mu m = 3V$

## Reality Check



- This is definitely a step in a good direction, but we still see residual discrepancies at high  $V_{GS}$  and the curvature is also not quite correct

## Additional Effect: Mobility Degradation due to Vertical Field

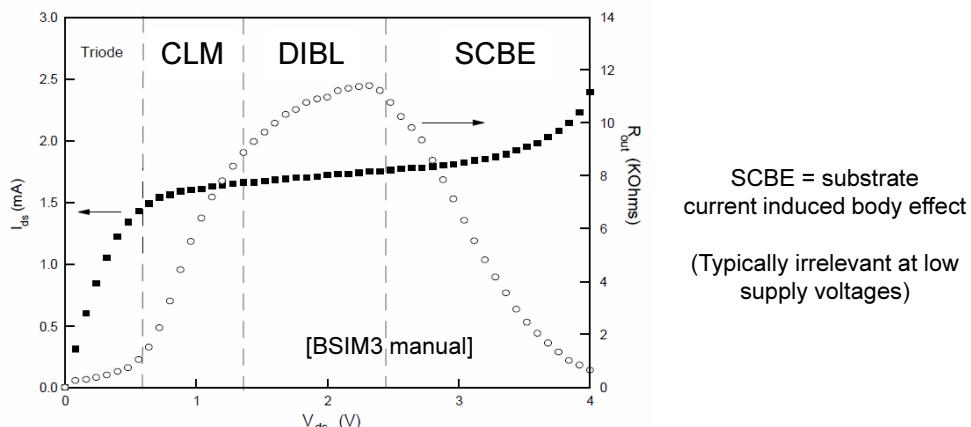
- In MOS technology, the oxide thickness has been continuously scaled down with feature size
  - ~6.5nm in 0.35μm, ~4nm in 0.18μm, ~1.8nm in 90nm CMOS
- As a result, the vertical electric field in the device has become quite large and tries to pull the carriers closer to the "dirty" silicon surface
  - Imperfections impede movement and thus mobility
- This net effect is a bias dependent mobility degradation that can be incorporated similar to velocity saturation
- The following is a possible equation that captures both velocity saturation and mobility degradation due to the vertical field

$$I_D \approx \frac{1}{2} \mu C_{ox} \frac{W}{L} V_{ov}^2 \cdot \frac{1}{(1 + [\theta V_{ov}]^m)^{\frac{1}{m}}} \quad (\text{see text, page 45})$$

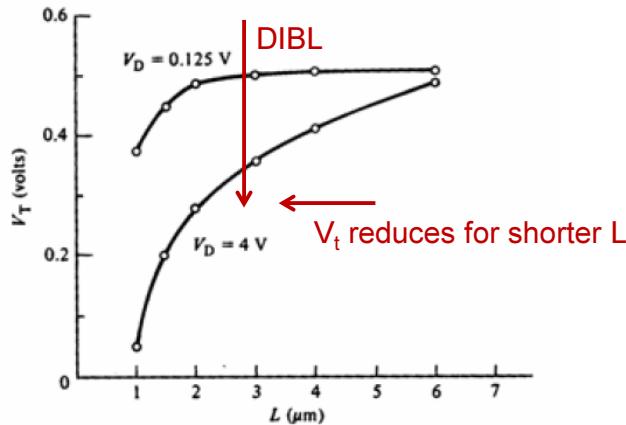
- The parameters  $m$  and  $\theta$  are "fudge factors" that allow us to fit this expression to measured data

## Drain Induced Barrier Lowering (DIBL)

- In the square law model, we attributed the  $I_D$ - $V_{DS}$  dependence (and thus finite intrinsic gain) primarily to channel length modulation
- In a short channel device, it turns out that the intrinsic gain is strongly affected by another effect called DIBL
- In essence, the drain can be viewed as an additional gate that modulates the inversion charge

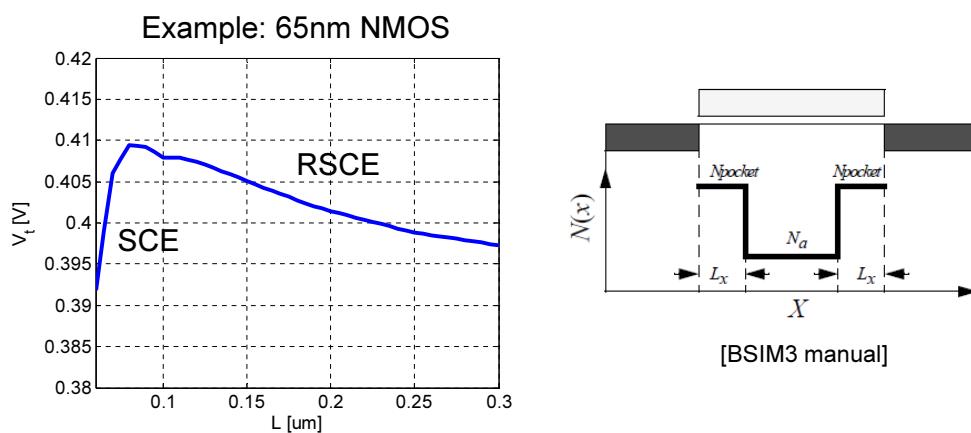


## “Short Channel Effect (SCE)”



## “Reverse Short Channel Effect (RSCE)”

- To reduce the width of the source/drain depletion regions, modern processes use pocket implants (“halos”)
- The average doping in the channel now increases for shorter L, which results in higher V<sub>t</sub> up to a point where the classic SCE effect takes over



# Chapter 5

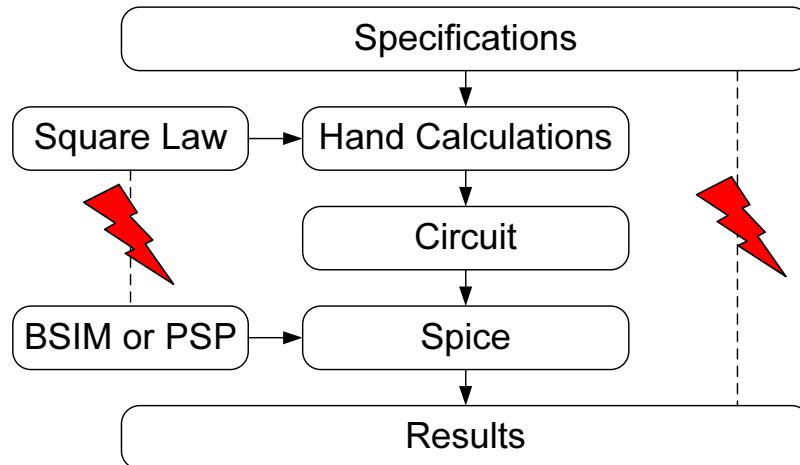
## $g_m/I_D$ -Based Design

Boris Murmann  
Stanford University  
Winter 2015-16

### Summary on MOSFET Modeling

- Modern MOSFETs are complicated!
- The IV-behavior in saturation can be roughly categorized according to the channel's inversion level: weak, moderate and strong inversion
- The current is due to diffusion in weak inversion and mostly due to drift in strong inversion; the transition is smooth and complicated
- The classic square law model is based on an ideal drift model, and applies only near the onset of strong inversion
  - And even then, the predictions are inaccurate unless short channel effects are taken into account
- The bottom line is that there is no modeling expression that is simple enough for hand analysis and sufficiently accurate to match real world device behavior

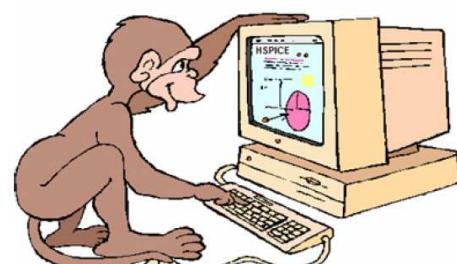
## The Problem



- Since there is a disconnect between actual transistor behavior and the simple square law model, any square-law driven design optimization will be far off from Spice results

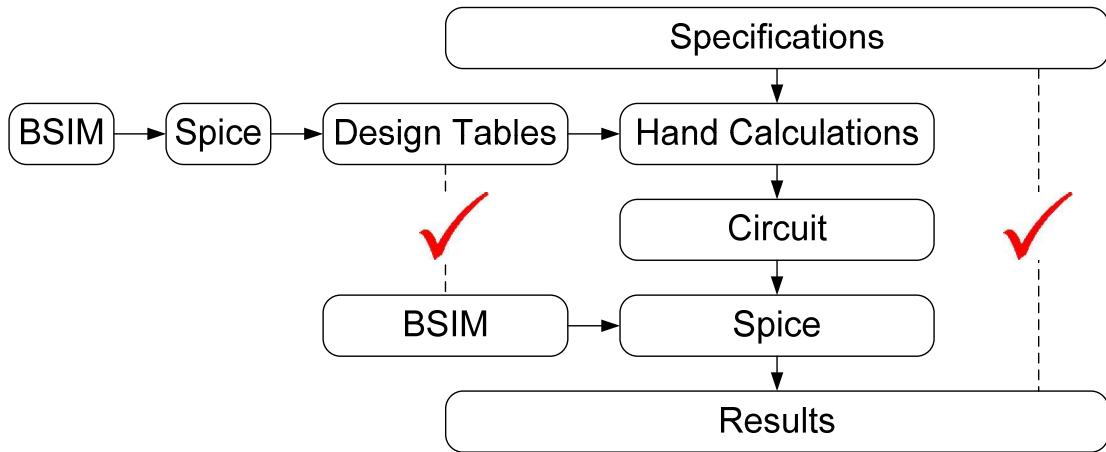
## Unfortunate Consequence

- In absence of a simple set of equations 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.
- Our goal
  - Maintain a systematic design methodology in absence of a set of compact MOSFET equations
- Strategy
  - Design using look-up tables or charts



[Courtesy Isaac Martinez]

## The Solution

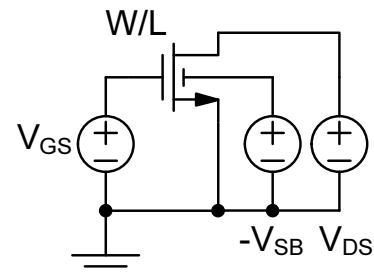


- Use pre-computed spice data in hand calculations

## Starting Point: Technology Characterization via DC Sweep

```
* /usr/class/ee214b/hspice/techsweep.sp
.inc ./models/ee214_hspice.sp
.inc techsweep_params.sp
vdsn vdn 0          dc 'ds'
vgsn vgn 0          dc 'gs'
vbsn vbn 0          dc '-sb'
vdsp vdp 0          dc '-ds'
vgsp vgp 0          dc '-gs'
vbsp vbp 0          dc 'sb'
mn   vdn vgn 0 vbn nmos214 L='length*1e-6' W=5u
mp   vdp vgp 0 vbp pmos214 L='length*1e-6' W=5u
.options dccap post brief accurate
.dc gs 0 1.8 25m ds 0 1.8 25m

.probe n_id    = par('i(mn)')
.probe n_vt    = par('vth(mn)')
.probe n_gm    = par('gmo(mn)')
.probe n_gmb   = par('gbso(mn)')
.probe n_gds   = par('gdso(mn)')
.probe n_cgg   = par('cgbo(mn)')
.probe n_cgs   = par('-cgbsb(mn)')
.probe n_cgd   = par('-cgdbo(mn)')
.probe n_cgb   = par('cgbbo(mn)')
.probe n_cdd   = par('cddbo(mn)')
.probe n_css   = par('-cgbsb(mn)-cbsbo(mn)')
```



## Matlab Wrapper

```
% /usr/class/ee214b/hspice/techsweep_hspice.m

% Path to hspice toolbox (http://www.cppsim.com/download\_hspice\_tools.html)
addpath('./HspiceToolbox')
% Load configuration
c = techsweep_config_bsim3_180_hspice;

for i = 1:length(c.LENGTH)
    for j = 1:length(c.VSB)
        % Write simulation parameters
        fid=fopen('techsweep_params.sp', 'w');
        fprintf(fid,'.param length = %d\n', c.LENGTH(i));
        fprintf(fid,'.param sb = %d\n', c.VSB(j));
        fclose(fid);

        % Run simulator
        [status,result] = system(c.simcmd);
        if(status)
            disp('Simulation did not run properly. Check techsweep.out.')
            return;
        end

        %Read and store results
        h = loadsig(c.outfile);
        for k = 1: length(c.outvars)
            nch.(c.outvars{k})(i,:,:,:) = evalsig(h, c.nvars{k});
            pch.(c.outvars{k})(i,:,:,:) = evalsig(h, c.pvars{k});
        end
    end
end
```

B. Murmann

EE214B Winter 2015-16 – Chapter 5

7

## Simulation Data in Matlab

```
% data stored in /usr/class/ee214b/matlab
>> load 180nch.mat;
>> nch

nch =
    ID: [4-D double]
    VT: [4-D double]
    GM: [4-D double]
    GMB: [4-D double]
    GDS: [4-D double]
    CGG: [4-D double]
    CGS: [4-D double]
    CGD: [4-D double]
    CGB: [4-D double]
    CDD: [4-D double]
    CSS: [4-D double]
INFO: 'Stanford EE214B models, 180nm CMOS, BSIM3'
    VGS: [73x1 double]
    VDS: [73x1 double]
    VSB: [11x1 double]
    L: [22x1 double]
    W: 5
    NFING: 1

>> size(nch.ID)
ans =
    22    73    73    11
```

Four-dimensional arrays

$$I_D(L, V_{GS}, V_{DS}, V_S)$$

$$V_t(L, V_{GS}, V_{DS}, V_S)$$

$$g_m(L, V_{GS}, V_{DS}, V_S)$$

...

B. Murmann

EE214B Winter 2015-16 – Chapter 5

8

## Lookup Function (For Convenience)

```
>> lookup(nch, 'ID', 'VGS', 0.5, 'VDS', 0.5, 'L', 0.25)

ans =
6.4737e-06

>> help lookup

The function "lookup" extracts a desired subset from the 4-dimensional
simulation data. The function interpolates when the requested points lie off
the simulation grid.

There are three usage modes:
(1) Simple lookup of parameters at some given (L, VGS, VDS, VSB)
(2) Lookup of arbitrary ratios of parameters, e.g. GM_ID, GM_CGG at given
    (L, VGS, VDS, VSB)
(3) Cross-lookup of one ratio against another, e.g. GM_CGG for some GM_ID

In usage modes (1) and (2) the input parameters (L, VGS, VDS, VSB) can be
listed in any order and default to the following values when not specified:

L = min(data.L); (minimum length used in simulation)
VGS = data.VGS; (VGS vector used during simulation)
VDS = max(data.VDS)/2; (VDD/2)
VSB = 0;
```

## Key Question

- How can we use all this data for **systematic** design?
- Many options exist
  - And you can invent your own, if you like
- Method taught in EE214B
  - Look at the transistor in terms of **width-independent** figures of merit that are intimately linked to design specification (rather than some physical modeling parameters that do not directly relate to circuit specs)
  - Think about the design tradeoffs in terms of the MOSFET's inversion level, using  $g_m/I_D$  as a proxy

## Figures of Merit for Design

### Square Law

- Transconductance efficiency
  - Want large  $g_m$ , for as little current as possible

$$\boxed{\frac{g_m}{I_D}}$$

$$= \frac{2}{V_{OV}}$$

- Transit frequency
  - Want large  $g_m$ , without large  $C_{gg}$

$$\boxed{\frac{g_m}{C_{gg}}}$$

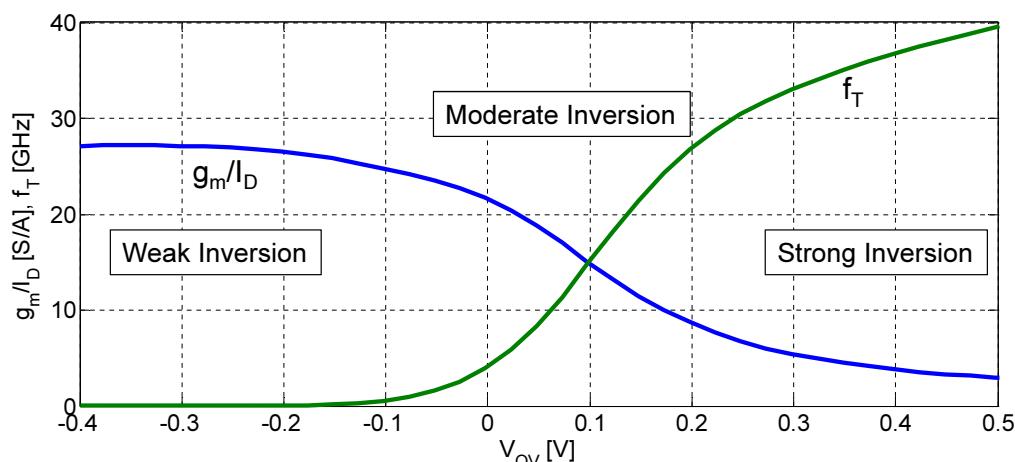
$$\approx \frac{3}{2} \frac{\mu V_{OV}}{L^2}$$

- Intrinsic gain
  - Want large  $g_m$ , but no  $g_o$

$$\boxed{\frac{g_m}{g_o}}$$

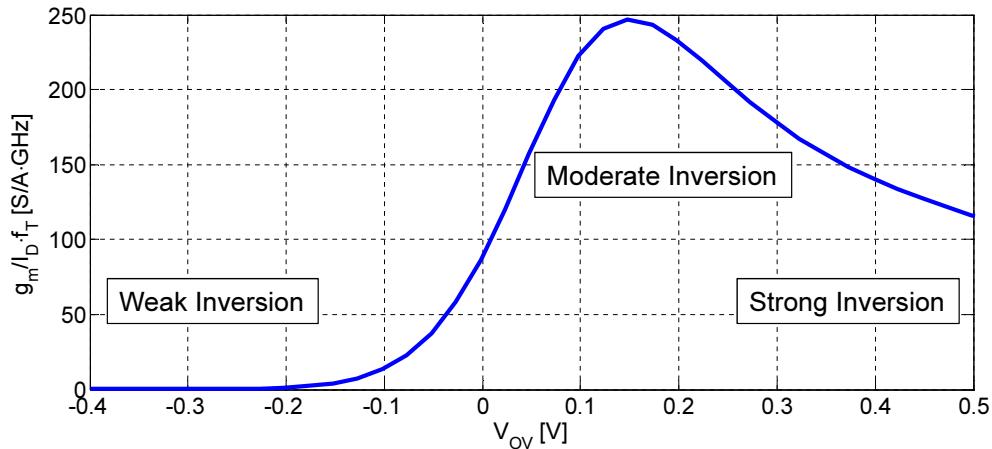
$$\approx \frac{2}{\lambda V_{OV}}$$

### Design Tradeoff: $g_m/I_D$ and $f_T$



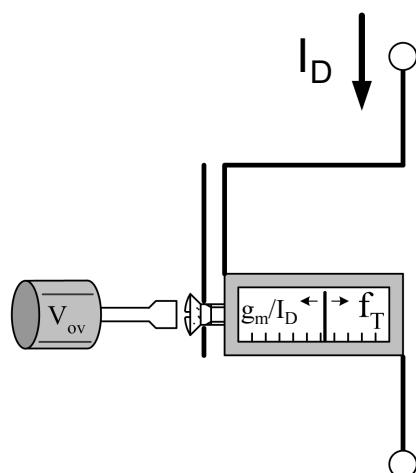
- Weak inversion: Large  $g_m/ID$  (>20 S/A), but small  $f_T$
- Strong inversion: Small  $g_m/ID$  (<10 S/A), but large  $f_T$

## Product of $g_m/I_D$ and $f_T$



- Interestingly, the product of  $g_m/I_D$  and  $f_T$  peaks in moderate inversion
- Operating the transistor in moderate inversion is optimal when we value speed and power efficiency equally
  - Not always the case

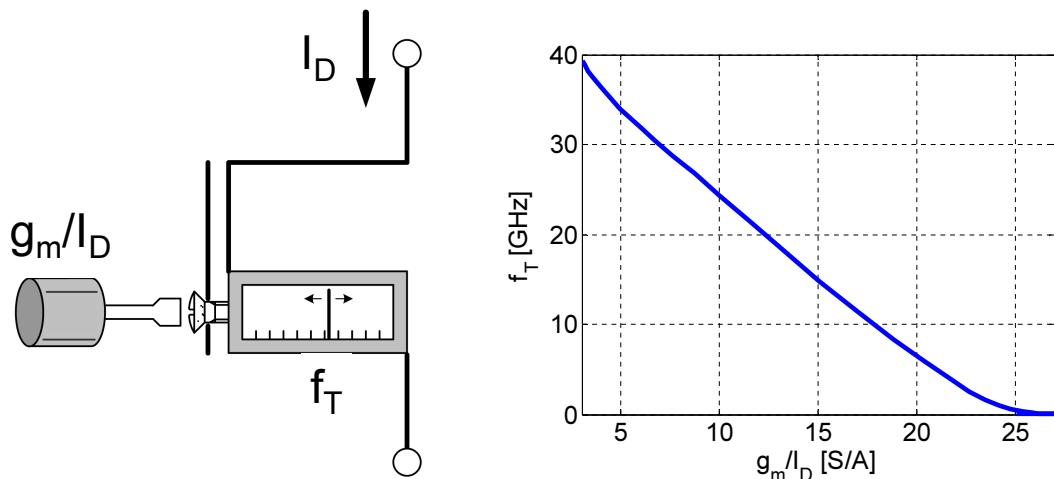
## Design in a Nutshell



- Choose the inversion level according to the proper tradeoff between speed ( $f_T$ ) and efficiency ( $g_m/I_D$ ) for the given circuit
- The inversion level is fully determined by the gate overdrive  $V_{ov}$ 
  - But,  $V_{ov}$  is not a very interesting parameter outside the square law framework; not much can be computed from  $V_{ov}$

## Eliminating $V_{ov}$

- The inversion level is also fully defined once we pick  $g_m/I_D$ , so there is no need to know  $V_{ov}$

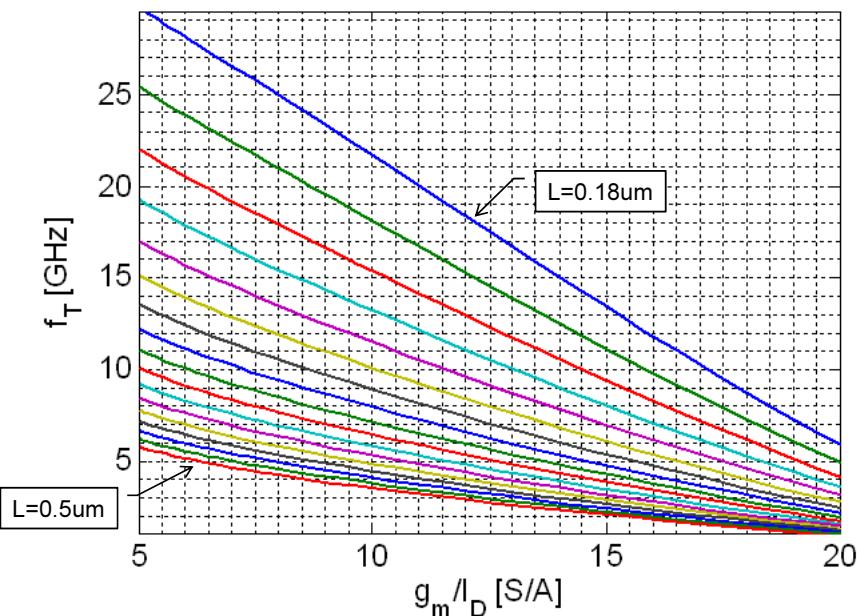


## $g_m/I_D$ -centric Technology Characterization

- Tabulate the following parameters for a reasonable range of  $g_m/I_D$  and channel lengths
  - Transit frequency ( $f_T$ )
  - Intrinsic gain ( $g_m/g_o$ )
- Also tabulate relative estimates of extrinsic capacitances
  - $C_{gd}/C_{gg}$  and  $C_{dd}/C_{gg}$
- Note that all of these parameters are (to first order) independent of device width
- In order to compute device widths, we need one more table that links  $g_m/I_D$  and current density  $I_D/W$

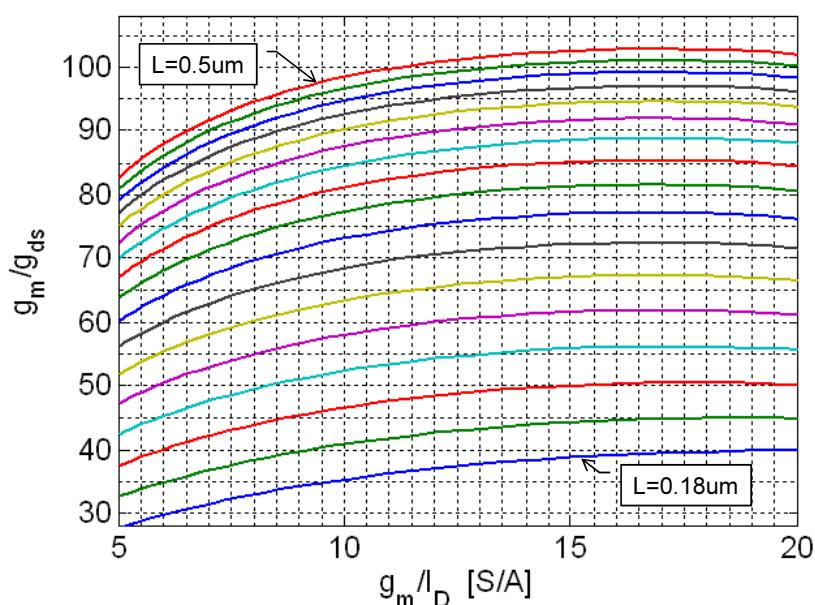
## Transit Frequency Chart

NMOS, 0.18...0.5um (step=20nm),  $V_{DS} = 0.9V$



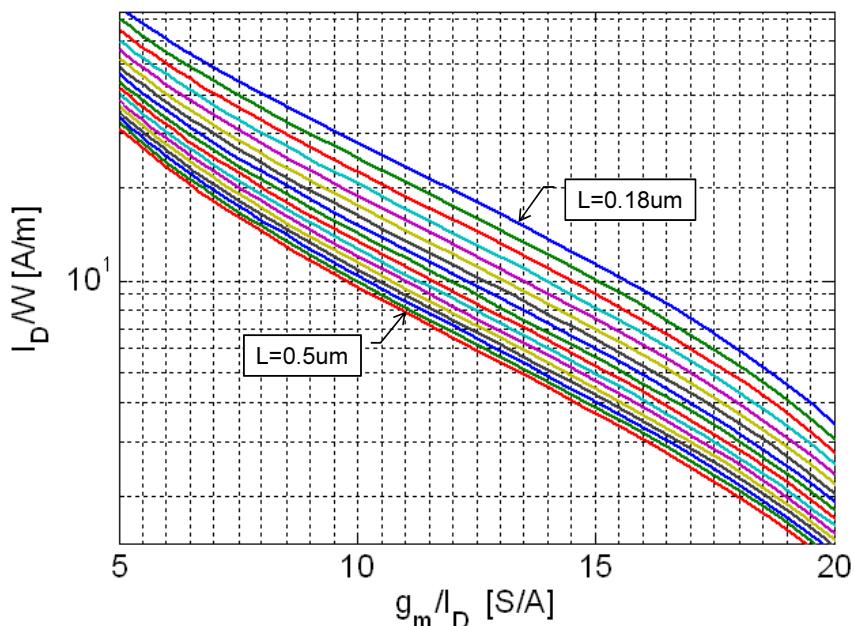
## Intrinsic Gain Chart

NMOS, 0.18...0.5um (step=20nm),  $V_{DS} = 0.9V$

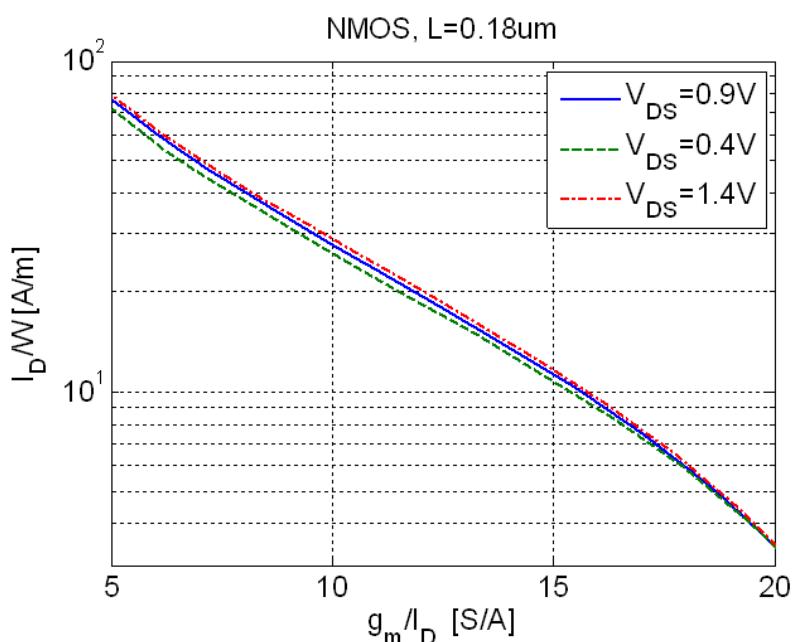


## Current Density Chart

NMOS, 0.18...0.5um (step=20nm),  $V_{DS} = 0.9V$



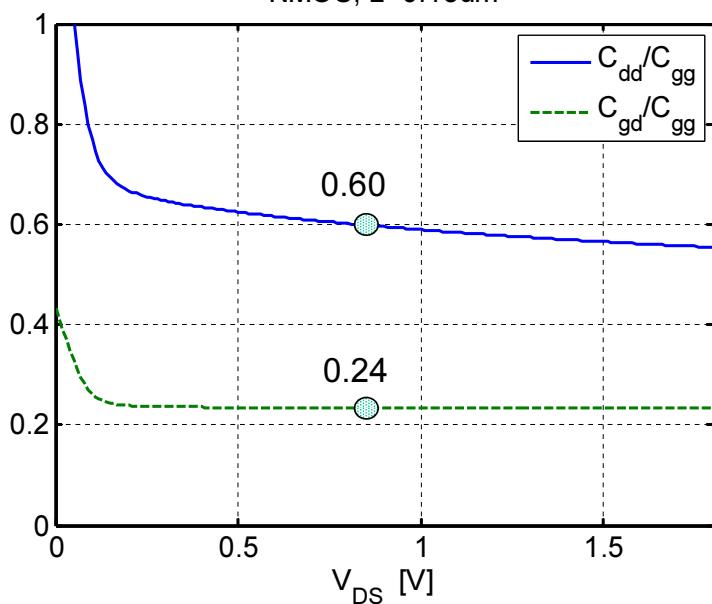
## $V_{DS}$ Dependence



- $V_{DS}$  dependence is relatively weak
- Typically OK to work with data generated for  $V_{DD}/2$

## Extrinsic Capacitances (1)

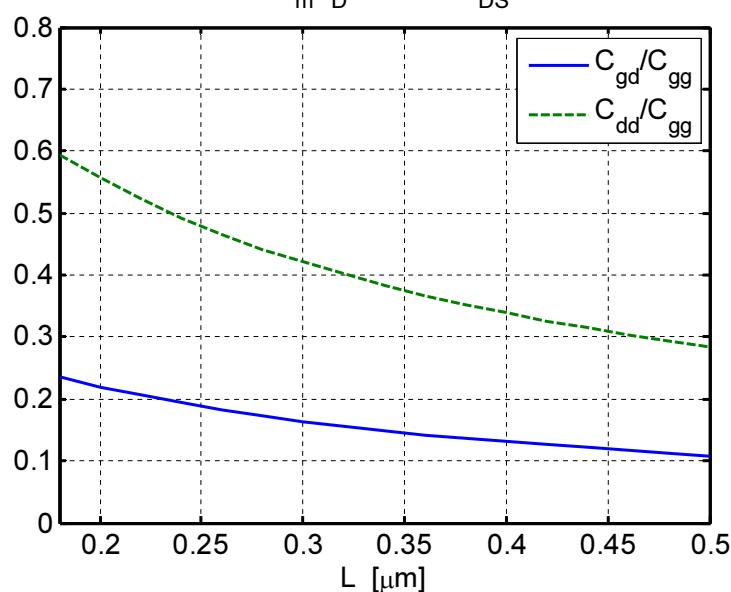
NMOS,  $L=0.18\mu\text{m}$



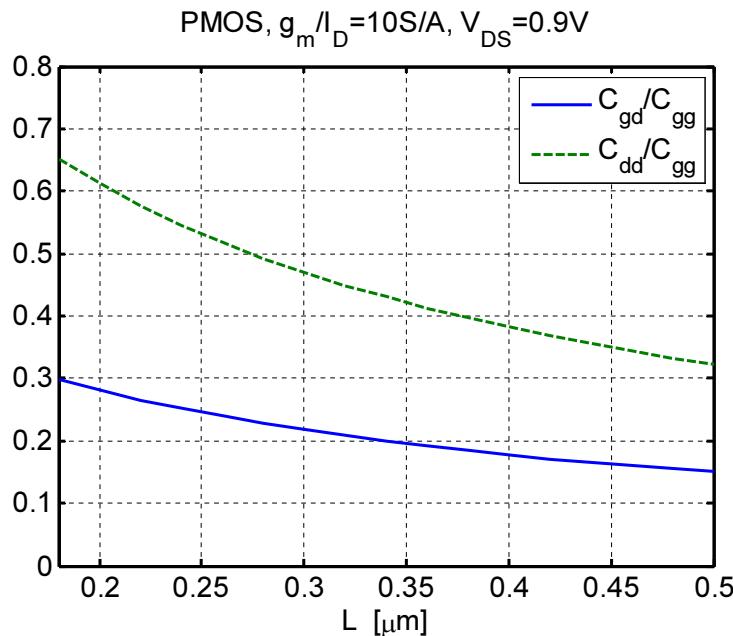
- Again, usually OK to work with estimates taken at  $V_{DD}/2$

## Extrinsic Capacitances (2)

NMOS,  $g_m/I_D = 10\text{S/A}$ ,  $V_{DS} = 0.9\text{V}$



### Extrinsic Capacitances (3)



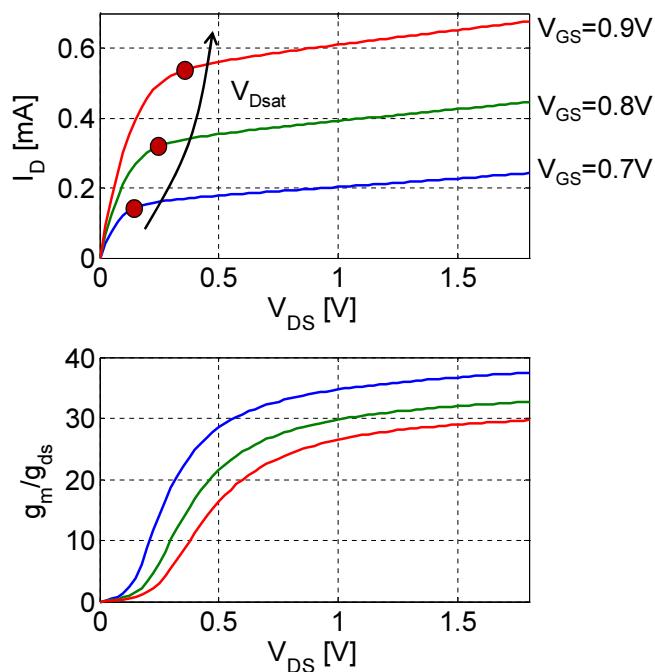
### Generic Design Flow

- 1) Determine  $g_m$  (from design objectives)
- 2) Pick  $L$ 
  - Short channel  $\rightarrow$  high  $f_T$  (high speed)
  - Long channel  $\rightarrow$  high intrinsic gain
- 3) Pick  $g_m/I_D$  (or  $f_T$ )
  - Large  $g_m/I_D \rightarrow$  low power, large signal swing (low  $V_{DSSat}$ )
  - Small  $g_m/I_D \rightarrow$  high  $f_T$  (high speed)
- 4) Determine  $I_D$  (from  $g_m$  and  $g_m/I_D$ )
- 5) Determine  $W$  (from  $I_D/W$ )

Many other possibilities exist (depending on circuit specifics, design constraints and objectives)

## How about $V_{Dsat}$ ?

- $V_{Dsat}$  tells us how much voltage we need across the transistor to operate in saturation
  - “High gain region”
- It is important to note that  $V_{Dsat}$  is not crisply defined in modern devices
  - Gradual increase of  $g_m/g_{ds}$  with  $V_{DS}$



## Relationship Between $V_{Dsat}$ and $g_m/I_D$

- It turns out that  $2/(g_m/I_D)$  is a reasonable first-order estimate for  $V_{Dsat}$

### Square Law

$$I_D = K(V_{GS} - V_t)^2$$

$$g_m = 2K(V_{GS} - V_t)$$

$$\frac{2}{(g_m / I_D)} = (V_{GS} - V_t) = V_{Dsat}$$

$\therefore$  Consistent with the classical first-order relationship

### Weak Inversion

$$I_D = I_{D0} e^{\frac{V_{GS}-V_t}{nV_T}} \left( 1 - e^{\frac{V_{DS}}{V_T}} \right)$$

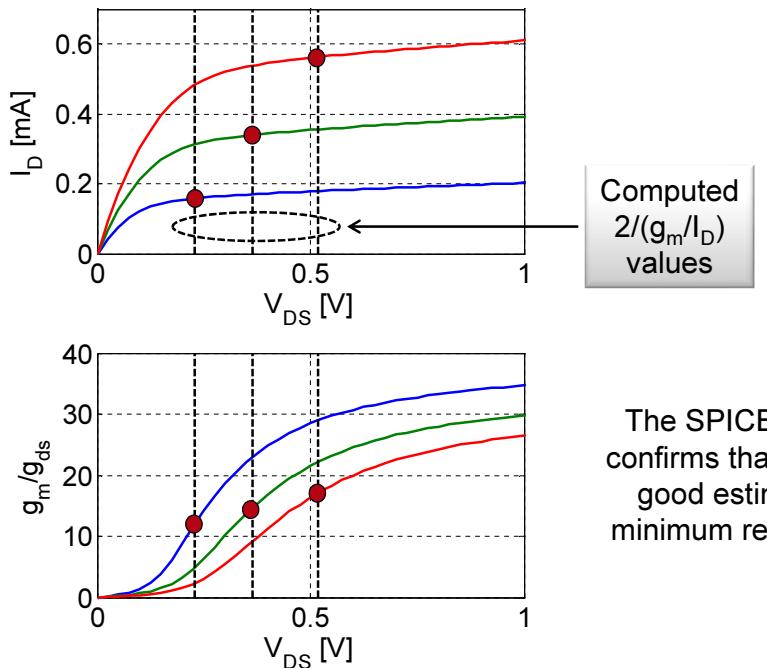
Need about  $3V_T$  for saturation

$$g_m = \frac{I_{D0}}{nV_T} e^{\frac{V_{GS}-V_t}{nV_T}} \left( 1 - e^{-\frac{V_{DS}}{V_T}} \right)$$

$$\frac{2}{(g_m / I_D)} = 2nV_T \cong 3V_T$$

$\therefore$  Corresponds well with the required minimum  $V_{DS}$

## Reality Check

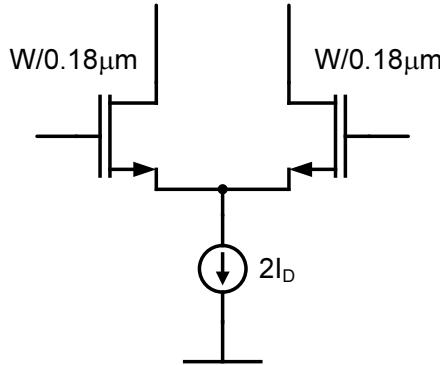


The SPICE model data confirms that  $2/(g_m/I_D)$  is a good estimate for the minimum reasonable  $V_{DS}$

## Examples

- Basic sizing example
- Sizing and bandwidth estimation in a simple gain stage
- Inversion level optimization in a simple gain stage

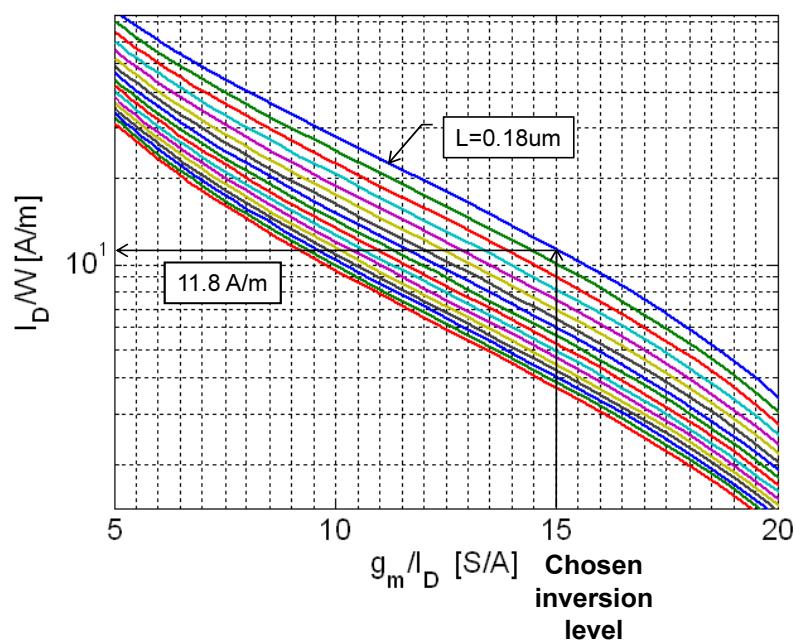
## Basic Sizing Example



- Determine the width of the differential pair transistors such that  $g_m=10\text{mS}$
- Consider various levels of inversion (weak, moderate, strong)
- For each case, compute the required  $I_D$ ,  $W$  and  $C_{gg}$  (total gate capacitance)

## Current Density Lookup

NMOS, 0.18...0.5um (step=20nm),  $V_{DS}=0.9\text{V}$

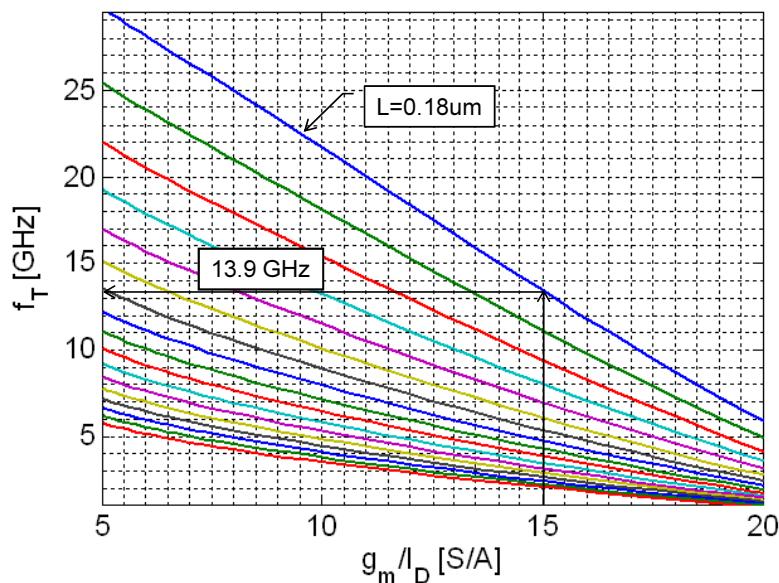


$$I_D = \frac{g_m}{g_m} = \frac{10\text{mS}}{15\frac{\text{S}}{\text{A}}} = 0.67\text{mA}$$

$$W = \frac{I_D}{I_D} = \frac{0.67\text{mA}}{11.8\frac{\text{A}}{\mu\text{m}}} = 56.6\mu\text{m}$$

## Determine $C_{gg}$ via $f_T$ Look-up

NMOS, 0.18...0.5um (step=20nm),  $V_{DS} = 0.9V$



$$C_{gg} = \frac{g_m}{\omega_T}$$

$$C_{gg} = \frac{10mS}{2\pi \cdot 13.9GHz} = 114fF$$

## Matlab Script

```
% EE214B, basic sizing example
clear all;
close all;

addpath('/usr/class/ee214b/matlab');
load 180nch.mat;

% Specification
gm = 10e-3;

% Chosen inversion levels and resulting drain current
gm_id = [25 15 5]';
ID = gm./gm_id

% Current density and width
JD = lookup(nch, 'ID_W', 'GM_ID', gm_id)
W = ID./JD

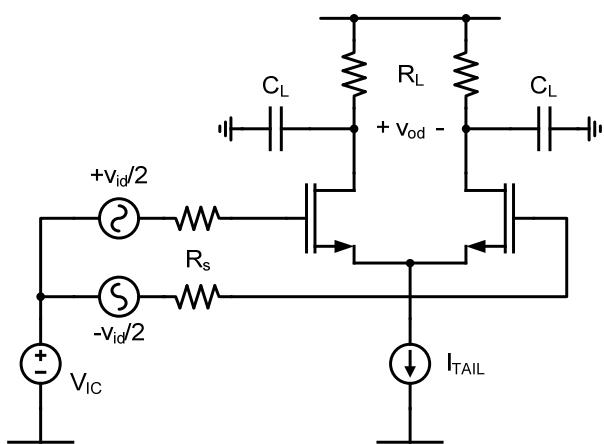
% Transit frequency and Cgg
wT = lookup(nch, 'GM_CGG', 'GM_ID', gm_id);
Cgg = gm./wT
```

## Result

$g_m/I_D$ [S/A]	25 (weak inversion)	15 (moderate inversion)	5 (strong inversion)
$g_m$ [mS]	10	10	10
$I_D$ [mA]	0.4	0.67	2
$I_D/W$ [A/m]	0.12	11.8	83.3
$W$ [ $\mu\text{m}$ ]	3243	56.7	24.0
$f_T$ [GHz]	0.37	13.9	32.3
$C_{gg}$ [fF]	4346	114	49

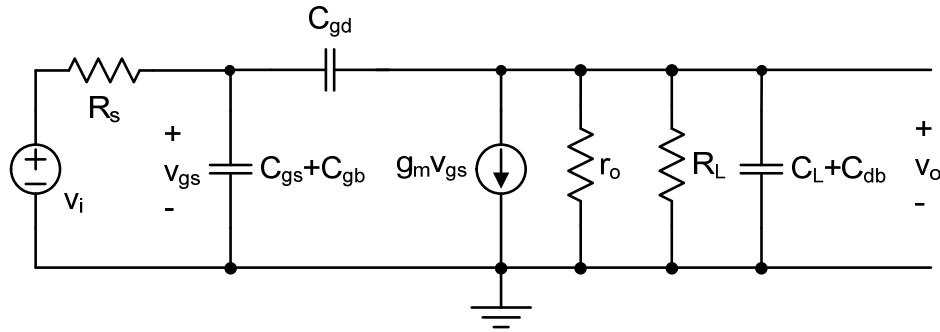
- Weak inversion
  - Small current, large device, large capacitance
- Strong inversion
  - Large current, small device, small capacitance
- Moderate inversion
  - A good compromise!

## Simple Gain Stage: Sizing and Bandwidth Estimation



- Given specs and objectives
  - Low frequency gain = -4
  - Parameters:  $R_L=1\text{k}$ ,  $C_L=50\text{fF}$ ,  $R_s=10\text{k}\Omega$ ,  $I_{TAIL}=600\mu\text{A}$ ,  $L=L_{min}=0.18\mu\text{m}$  (to maximize bandwidth)
  - Determine device width
  - Estimate dominant and non-dominant pole

## Small-Signal Half-Circuit Model



$$|A_{vo}| \approx g_m R_L = 4 \quad \Rightarrow \quad g_m = \frac{4}{1k\Omega} = 4mS \quad \frac{g_m}{I_D} = \frac{4mS}{300\mu A} = 13.3 \frac{S}{A}$$

Using table lookup:  $f_T = 16.9\text{GHz}$      $\frac{I_D}{W} = 16.1 \frac{A}{m}$

### Aside: Why can we Neglect $r_o$ ?

$$|A_{vo}| = g_m (R_L \parallel r_o)$$

$$= g_m \left( \frac{1}{R_L} + \frac{1}{r_o} \right)^{-1}$$

$$\frac{1}{|A_{vo}|} = \frac{1}{g_m R_L} + \frac{1}{g_m r_o}$$

$$\frac{1}{4} = \frac{1}{g_m R_L} + \frac{1}{g_m r_o}$$

- Even at  $L=L_{min}=0.18\mu m$ , we have  $g_m r_o > 30$
- $r_o$  is negligible in this design problem

## Zero and Pole Expressions

High frequency zero  
(negligible)       $\omega_z = \frac{g_m}{C_{gd}} \gg \omega_T$

Denominator coefficients       $b_1 = R_s [C_{gs} + C_{gd}(1 + |A_{v0}|)] + R_L(C_L + C_{gd})$   
 $b_2 = R_s R_L (C_{gs} C_L + C_{gs} C_{gd} + C_L C_{gd})$

Dominant pole       $\omega_{p1} \cong \frac{1}{b_1}$       ( $C_{db}$  can be added to  $C_L$  if significant)

Nondominant pole       $\omega_{p2} \cong \frac{b_1}{b_2}$

## Find Capacitances and Compute Poles

$$C_{gg} = \frac{1}{2\pi} \frac{4mS}{16.9GHz} = 37.8fF$$

$$C_{gd} = \frac{C_{gd}}{C_{gg}} C_{gg} = 0.24 \cdot 37.7fF = 9.0fF$$

$$C_{dd} = \frac{C_{dd}}{C_{gg}} C_{gg} = 0.60 \cdot 37.7fF = 22.6fF$$

$$C_{db} = C_{dd} - C_{gd} = 13.6fF$$

$$C_{gs} = C_{gg} - C_{gd} = 28.8fF$$

$f_{p1} \cong 200 \text{ MHz}$

$f_{p2} \cong 5.8 \text{ GHz}$

## Matlab Script

```

clear all; close all;
load 180nch.mat;

% Specs
Av0 = 4; RL = 1e3; CL = 50e-15; Rs = 10e3; ITAIL = 600e-6;

% Component calculations
gm = Av0/RL;
gm_id = gm/(ITAIL/2);
wT = lookup(nch, 'GM_CGG', 'GM_ID', gm_id);
cgd_cgg = lookup(nch, 'CGD_CGG', 'GM_ID', gm_id);
cdd_cgg = lookup(nch, 'CDD_CGG', 'GM_ID', gm_id);
cgg = gm/wT;
cgd = cgd_cgg*cgg;
cdd = cdd_cgg*cgg;
cdb = cdd - cgd;
cgs = cgg - cgd;

% pole calculations
b1 = Rs*(cgs + cgd*(1+Av0)) + RL*(CL+cgd);
b2 = Rs*RL*(cgs*CL + cgs*cgd + CL*cgd);
fp1 = 1/2/pi/b1
fp2 = 1/2/pi*b1/b2

% device sizing
id_w = lookup(nch, 'ID_W', 'GM_ID', gm_id);
w = ITAIL/2 / id_w

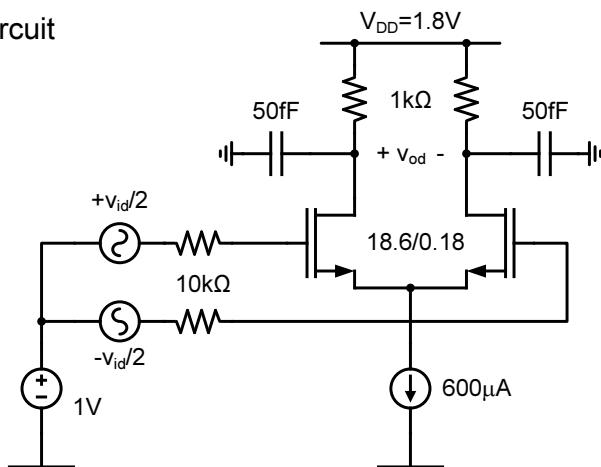
```

## Circuit For Spice Verification

Device width

$$W = \frac{I_D}{I_D} = \frac{300\mu A}{16.1A/m} = 18.6\mu m$$

Simulation circuit



## Circuit Netlist

```
* ee214 device models
.include /usr/class/ee214b/hspice/ee214_hspice.sp

vdd vdd 0 1.8
vic vic 0 1
vid vid 0 ac 1
x1 vid vic vip vim balun
x2 vod vom vop vom balun
rdum vod 0 1gig
it t 0 600u

m1 vop vgp t 0 nmos214 w=18.6u l=0.18u
m2 vom vgm t 0 nmos214 w=18.6u l=0.18u
rsp vip vgp 10k
rsm vim vgm 10k
rlp vop vdd 1k
rlm vom vdd 1k
clp vop 0 50f
clm vom 0 50f

.op
.ac dec 100 1e6 1000e9

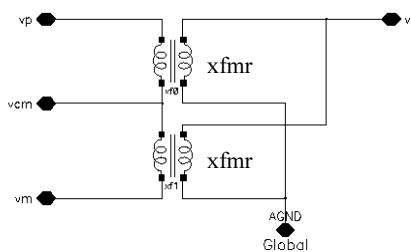
.pz v(vod) vid
.option post brief accurate
.end
```

B. Murmann

EE214B Winter 2015-16 – Chapter 5

41

## Aside: Ideal Balun



```
.subckt balun vdm vcm vp vm
e1 vp vcm transformer vdm 0 2
e2 vcm vm transformer vdm 0 2
.ends balun
```

- Useful for separating CM and DM signal components
- Bi-directional, preserves port impedance
- Uses ideal, inductorless transformers that work down to DC
- Not available in all simulators

B. Murmann

EE214B Winter 2015-16 – Chapter 5

42

## Simulated DC Operating Point

element	0:m1	0:m2
model	0:nmos214	0:nmos214
region	Saturati	Saturati
id	300.0000u	300.0000u
vgs	682.4474m	682.4474m
vds	1.1824	1.1824
vbs	-317.5526m	-317.5526m
vth	564.5037m	564.5037m
vdsat	109.0968m	109.0968m
vod	117.9437m	117.9437m
beta	37.2597m	37.2597m
gam eff	583.8490m	583.8490m
gm	4.0718m	4.0718m
gds	100.9678u	100.9678u
gmb	887.2111u	887.2111u
cdtot	20.8290f	20.8290f
cgtot	37.4805f	37.4805f
cstot	42.2382f	42.2382f
cbtot	31.5173f	31.5173f
cgs	26.7862f	26.7862f
cgd	8.9672f	8.9672f

Good agreement!

Design values

$$g_m = 4 \text{ mS}$$

$$C_{dd} = 22.6 \text{ fF}$$

$$C_{gg} = 37.8 \text{ fF}$$

$$C_{gd} = 9.0 \text{ fF}$$

## HSpice .OP Capacitance Output Variables

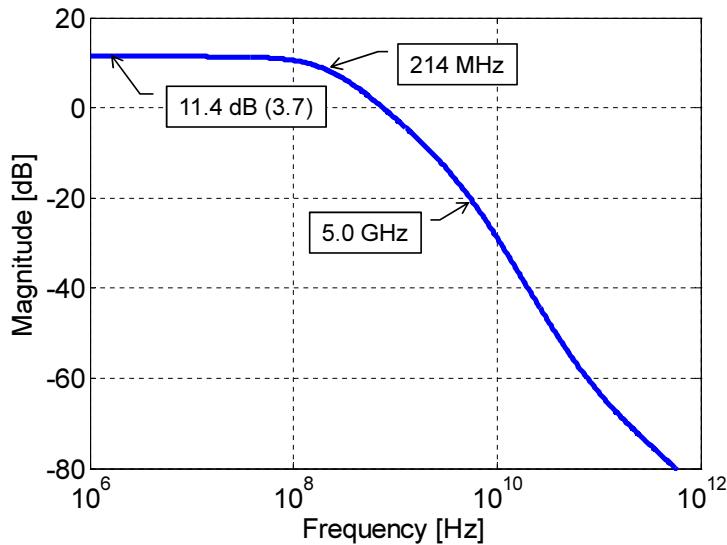
HSpice (.OP)

cdtot	20.8290f
cgtot	37.4805f
cstot	42.2382f
cbtot	31.5173f
cgs	26.7862f
cgd	8.9672f

Corresponding Small Signal  
Model Elements

$$\begin{aligned} cdtot &\equiv C_{gd} + C_{db} \\ cgtot &\equiv C_{gs} + C_{gd} + C_{gb} \\ cstot &\equiv C_{gs} + C_{sb} \\ cbtot &\equiv C_{gb} + C_{sb} + C_{db} \\ cgs &\equiv C_{gs} \\ cgd &\equiv C_{gd} \end{aligned}$$

## Simulated AC Response



- Calculated values:  $|A_{v0}|=12$  dB (4.0),  $f_{p1} = 200$  MHz,  $f_{p2}= 5.8$  GHz

## Plotting HSpice Results in Matlab

```
clear all;
close all;
addpath('/usr/class/ee214b/matlab/hspice_toolbox');

h = loadsig('gm_id_example1.ac0');
lssig(h)

f = evalsig(h,'HERTZ');
vod = evalsig(h,'vod');
magdb = 20*log10(abs(vod));
av0 = abs(vod(1))
f3dB = interp1(magdb, f, magdb(1)-3, 'spline')

figure(1);
semilogx(f, magdb, 'linewidth', 3);
xlabel('Frequency [Hz]');
ylabel('Magnitude [dB]');
axis([1e6 1e12 -80 20]);
grid;
```

## Using .pz Analysis

### Netlist statement

```
.pz v(vod) vid
```

### Output

```
*****
input = 0:vid          output = v(vod)

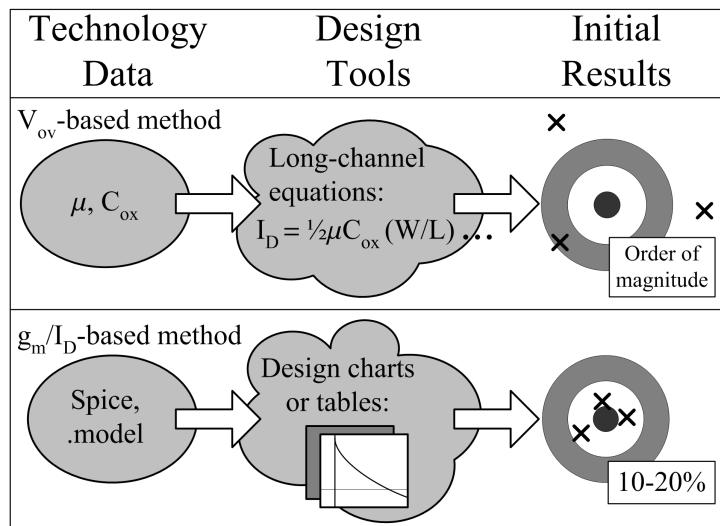
      poles (rad/sec)           poles ( hertz)
real      imag            real      imag
-1.35289g    0.        -215.319x    0.
-31.6307g    0.        -5.03418g    0.

      zeros (rad/sec)           zeros ( hertz)
real      imag            real      imag
445.734g    0.        70.9407g    0
```

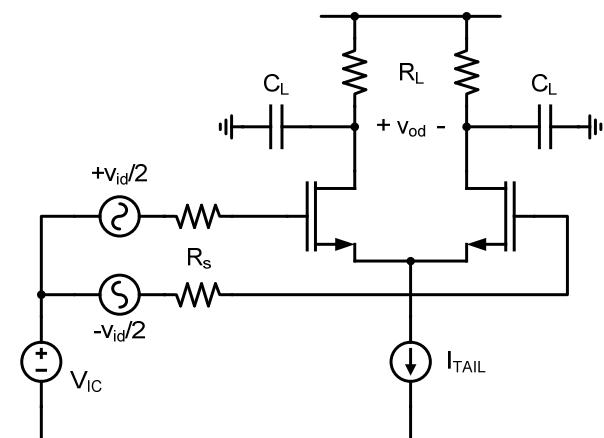
## Observations

- The design is essentially right on target!
  - Typical discrepancies are no more than 10-20%, due to  $V_{DS}$  dependencies, finite output resistance, etc.
- We accomplished this by using pre-computed spice data in the design process
- Even if discrepancies are more significant, there's always the possibility to track down the root causes
  - Hand calculations are based on parameters that also exist in Spice, e.g.  $g_m/I_D$ ,  $f_T$ , etc.
  - Different from square law calculations using  $\mu C_{ox}$ ,  $V_{OV}$ , etc.
    - Based on artificial parameters that do not exist or have no significance in the spice model

## Comparison



## Simple Gain Stage: Inversion Level Optimization



- Same parameters as before, but this time  $I_{TAIL}$  is not given
- Want to inspect the design space and look for potential “sweet spots”

## $g_m/I_D$ sweep

```

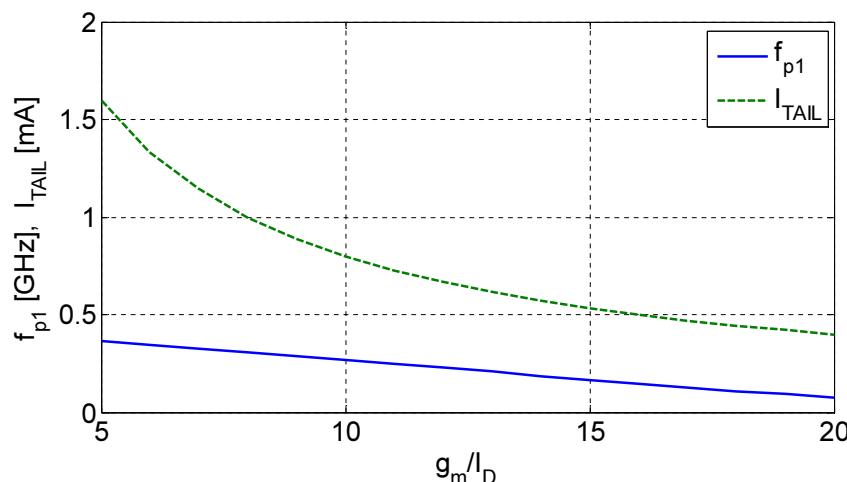
clear all; close all;
load 180nch.mat;

% Specs
Av0 = 4; RL = 1e3; CL = 50e-15; Rs = 10e3;

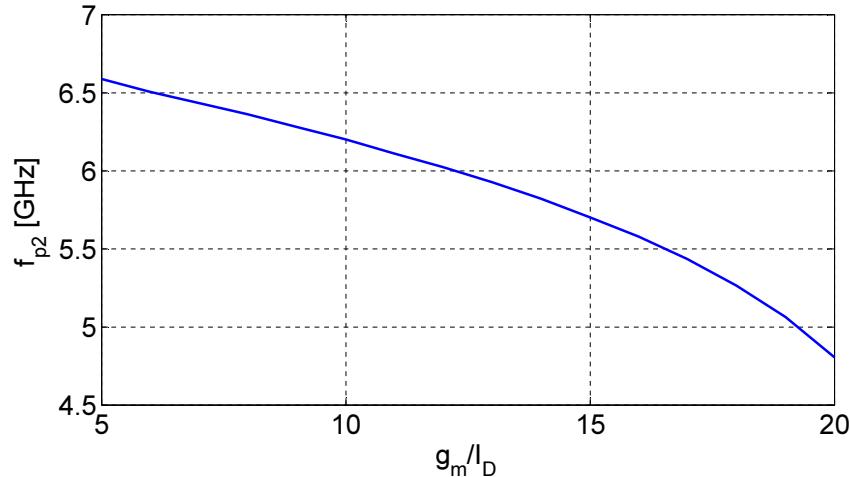
gm = Av0/RL;
gm_id = 5:20;
itail = 2*gm./gm_id;
for i = 1:length(gm_id);
    wT = lookup(nch, 'GM_CGG', 'GM_ID', gm_id(i));
    cgd_cgg = lookup(nch, 'CGD_CGG', 'GM_ID', gm_id(i));
    cdd_cgg = lookup(nch, 'CDD_CGG', 'GM_ID', gm_id(i));
    cg = gm/wT;
    cgd = cgd_cgg*cgg;
    cdd = cdd_cgg*cgg;
    cdb = cdd - cgd;
    cgs = cg - cgd;
    % pole calculations
    b1 = Rs*(cgs + cgd*(1+Av0)) + RL*(CL+cgd);
    b2 = Rs*RL*(cgs*CL + cgs*cgd + CL*cgd);
    fp1(i) = 1/2/pi/b1;
    fp2(i) = 1/2/pi*b1/b2;
end

```

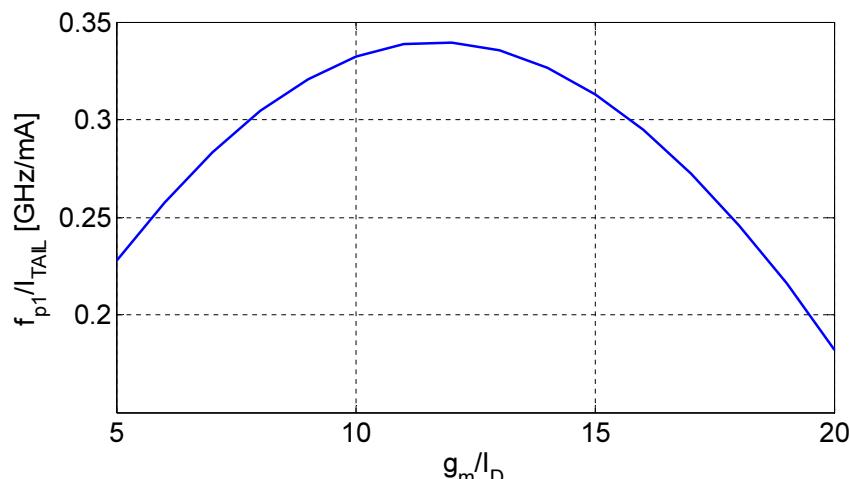
## Dominant Pole and Current as a Function of $g_m/I_D$



## Nondominant Pole as a Function of $g_m/I_D$



## Bandwidth/Current Ratio as a Function of $g_m/I_D$



- Key point

- Using a lookup table based design flow allows you to visualize the design space, and shapes/enhances your understanding of the circuit tradeoffs → very hard to do using a Spice monkey approach

## References

- F. Silveira et al. "A  $g_m/I_D$  based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA," IEEE J. Solid-State Circuits, Sep. 1996, pp. 1314-1319.
- D. Foty, M. Bucher, D. Binkley, "Re-interpreting the MOS transistor via the inversion coefficient and the continuum of  $g_{ms}/I_d$ ," Proc. Int. Conf. on Electronics, Circuits and Systems, pp. 1179-1182, Sep. 2002.
- B. E. Boser, "Analog Circuit Design with Submicron Transistors," IEEE SSCS Meeting, Santa Clara Valley, May 19, 2005, <http://www.ewh.ieee.org/r6/scv/ssc/May1905.htm>
- P. Jespers, The  $g_m/I_D$  Methodology, a sizing tool for low-voltage analog CMOS Circuits, Springer, 2010.
- T. Konishi, K. Inazu, J.G. Lee, M. Natsu, S. Masui, and B. Murmann, "Optimization of High-Speed and Low-Power Operational Transconductance Amplifier Using  $g_m/I_D$  Lookup Table Methodology," IEICE Trans. Electronics, Vol. E94-C, No.3, Mar. 2011.
- P.G.A. Jespers and B. Murmann, "Calculation of MOSFET Distortion Using the Transconductance-to-Current Ratio ( $g_m/I_D$ )," in Proc. IEEE Int. Symp. Circuits Syst., May 2015, pp. 529-532.

# **Chapter 6**

## **Electronic Noise**

**Boris Murmann  
Stanford University  
Winter 2015-16**

**Textbook Sections: 9.1, 9.2, 9.3, 9.4**

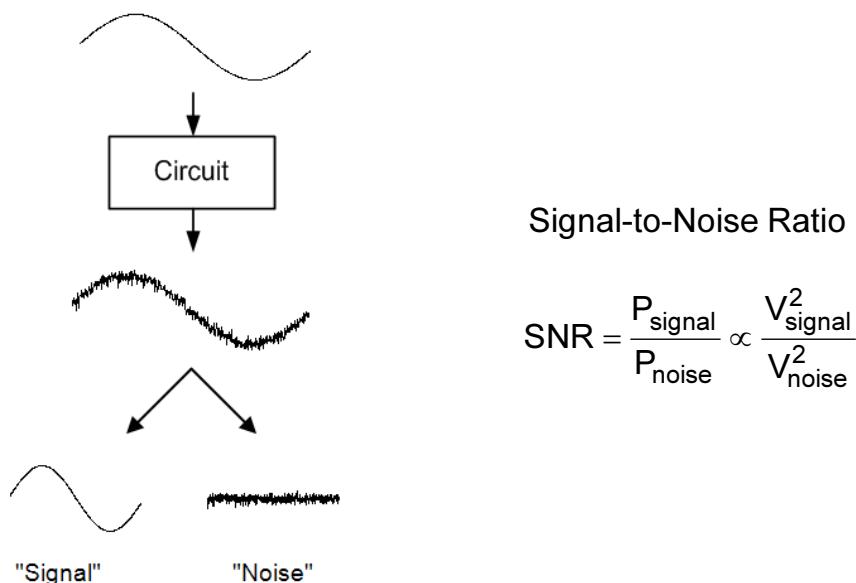
### **Outline**

- Noise in components
  - Resistors
  - MOSFETs
  - BJTs
- Noise analysis in circuits
  - Output- and input-referred noise with ideal voltage drive
  - Total integrated noise
  - Output and input referred noise with finite source resistance
  - Equivalent voltage and current generators for arbitrary source resistance

## Types of Noise

- "Man made noise" or interference noise
  - Signal coupling
  - Substrate coupling
  - Finite power supply rejection
  - Solutions
    - Fully differential circuits
    - Layout techniques
- "Electronic noise" or "device noise" (focus of this discussion)
  - Fundamental
    - E.g. "thermal noise" caused by random motion of carriers
  - Technology related
    - "Flicker noise" caused by material defects and "roughness"

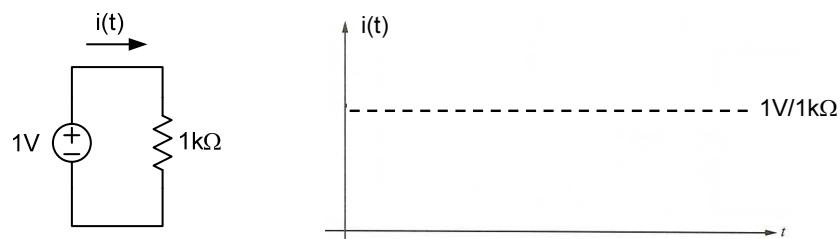
## Significance of Electronic Noise (1)



## Significance of Electronic Noise (2)

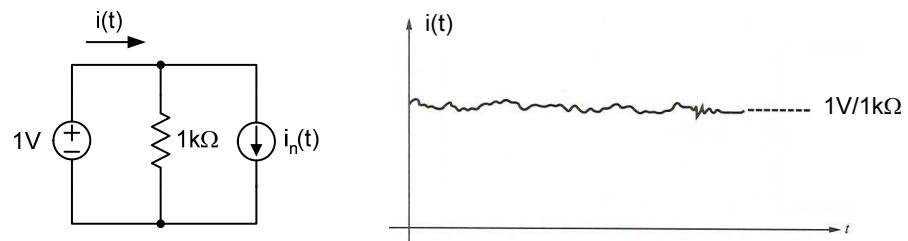
- The "fidelity" of electronic systems is often determined by their SNR
  - Examples
    - Audio systems
    - Imagers, cameras
    - Wireless and wireline transceivers
  - Electronic noise directly trades with power dissipation and speed
  - Noise has become increasingly important in modern technologies with reduced supply voltages
    - $\text{SNR} \sim V_{\text{signal}}^2/V_{\text{noise}}^2 \sim (\alpha V_{\text{DD}})^2/V_{\text{noise}}^2$
- Topics
  - How to model noise of circuit components
  - How to calculate/simulate the noise performance of a complete circuit
    - In which circuits and applications does thermal noise matter?

## Ideal Resistor



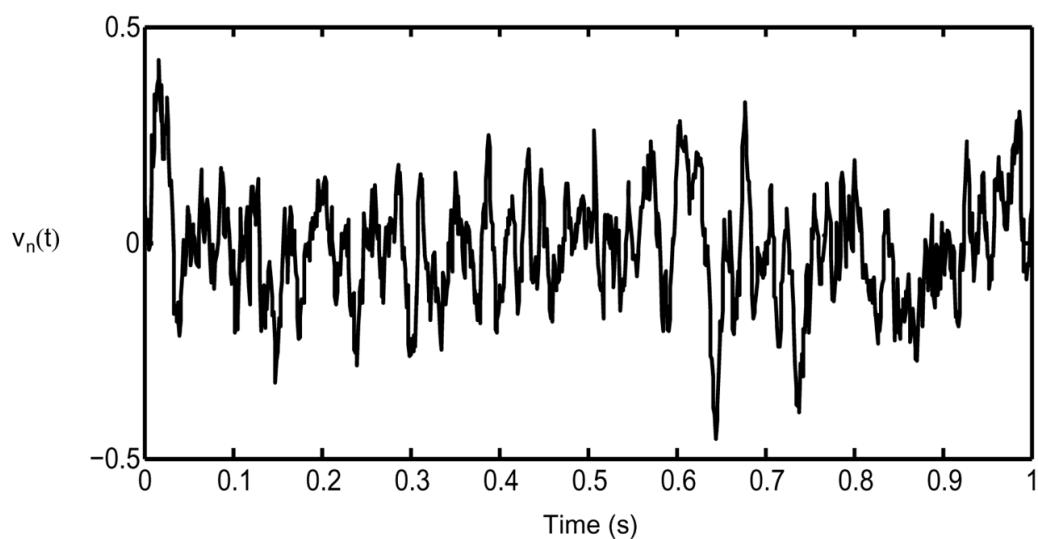
- Constant current, independent of time
- Non-physical
  - In a physical resistor, carriers "randomly" collide with lattice atoms, giving rise to small current variations over time

## Physical Resistor



- "Thermal Noise" or "Johnson Noise"
  - J.B. Johnson, "Thermal Agitation of Electricity in Conductors," Phys. Rev., pp. 97-109, July 1928
- Can model random current component using a noise current source  $i_n(t)$

## Example Voltage Noise Waveform



Text, p. 364

## Properties of Thermal Noise

- Present in any conductor
- Independent of DC current flow
- Instantaneous noise value is unpredictable since it is a result of a large number of random, superimposed collisions with relaxation time constant  $\tau_0 \approx 0.17\text{ps}$ 
  - Consequences:
    - Gaussian amplitude distribution
    - Knowing  $i_n(t)$  does not help predict  $i_n(t+\Delta t)$ , unless  $\Delta t$  is on the order of 0.17ps (cannot sample signals this fast)
    - The power generated by thermal noise is spread up to very high frequencies ( $1/\tau_0 \approx 6,000\text{Grad/s}$ )
- The only predictable property of thermal noise is its average power!

## Average Power

- For a deterministic current signal with period T, the average power is

$$P_{av} = \frac{1}{T} \int_{-T/2}^{T/2} i^2(t) \cdot R \cdot dt$$

- This definition can be extended to random signals
- Assuming a real, stationary and ergodic random process, we can write

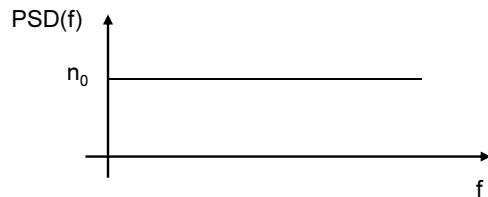
$$P_n = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} i_n^2(t) \cdot R \cdot dt$$

- For notational convenience, we typically drop R in the above expression and work with "mean square" currents (or voltages)

$$\bar{i_n^2} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} i_n^2(t) \cdot dt$$

## Thermal Noise Spectrum

- The so-called power spectral density (PSD) shows how the power is distributed across frequency
- In the case of thermal noise, the power is spread uniformly up to very high frequencies (about 10% drop at 2,000GHz)



- The total average noise power  $P_n$  in a particular frequency band is found by integrating the PSD

$$P_n = \int_{f_1}^{f_2} \text{PSD}(f) \cdot df$$

## Thermal Noise Power

- Nyquist showed that the noise PSD of a resistor is

$$\text{PSD}(f) = n_0 = 4 \cdot kT$$

- $k$  is the Boltzmann constant and  $T$  is the absolute temperature
- $4kT = 1.66 \cdot 10^{-20}$  Joules at room temperature
- The total average noise power of a resistor in a certain frequency band is therefore

$$P_n = \int_{f_1}^{f_2} 4kT \cdot df = 4kT \cdot (f_2 - f_1) = 4kT \cdot \Delta f$$

## Equivalent Noise Generators

- We can model the noise using either an equivalent voltage or current generator

$$\overline{v_n^2} = P_n \cdot R = 4kT \cdot R \cdot \Delta f$$

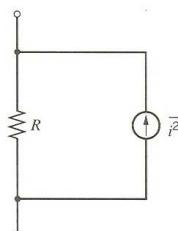
$$\overline{i_n^2} = \frac{P_n}{R} = 4kT \cdot \frac{1}{R} \cdot \Delta f$$

For  $R = 1\text{k}\Omega$ :

$$\frac{\overline{v_n^2}}{\Delta f} = 16 \cdot 10^{-18} \frac{V^2}{\text{Hz}}$$

$$\sqrt{\frac{\overline{v_n^2}}{\Delta f}} = 4\text{nV} / \sqrt{\text{Hz}}$$

$$\Delta f = 1\text{MHz} \Rightarrow \sqrt{\overline{v_n^2}} = 4\mu\text{Vrms}$$



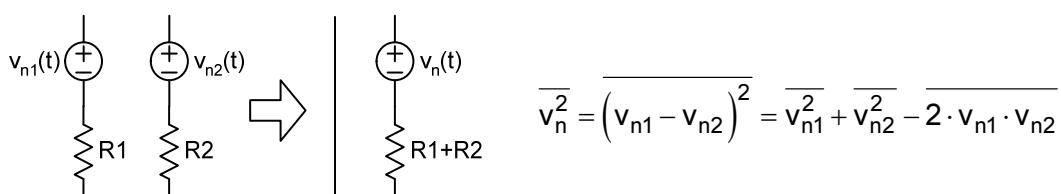
For  $R = 1\text{k}\Omega$ :

$$\frac{\overline{i_n^2}}{\Delta f} = 16 \cdot 10^{-24} \frac{A^2}{\text{Hz}}$$

$$\sqrt{\frac{\overline{i_n^2}}{\Delta f}} = 4\text{pA} / \sqrt{\text{Hz}}$$

$$\Delta f = 1\text{MHz} \Rightarrow \sqrt{\overline{i_n^2}} = 4\text{nArms}$$

## Two Resistors in Series

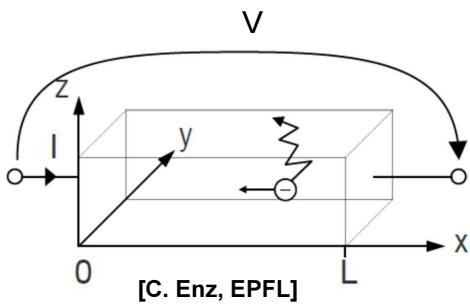


- Since  $v_{n1}(t)$  and  $v_{n2}(t)$  are statistically independent, we have

$$\overline{v_n^2} = \overline{v_{n1}^2} + \overline{v_{n2}^2} = 4 \cdot kT \cdot (R_1 + R_2) \cdot \Delta f$$

- Always remember to add independent noise sources using mean squared quantities
  - Never add RMS values!

## Derivation of Resistor Noise PSD (1)



PSD of  $V$  is equal to Fourier transform of autocorrelation function

Model for autocorrelation  
 $\tau_0$  is the relaxation time (mean time between collisions)

Volume of  $N = n \cdot A \cdot L$  electrons  
 $A$  = cross section  
 $L$  = length  
 $n$  = electron density  
 $v_i$  = velocity of individual  $e^-$  in  $x$ -direction

$$V = R \cdot I = R \cdot q \cdot \frac{\sum_{i=1}^N v_i}{L}$$

$$S_V = \mathcal{F}(R_V) = N \left( \frac{Rq}{L} \right)^2 \mathcal{F}(R_{v_i})$$

$$R_{v_i} = R_{v_i}(0) e^{-|\tau|/\tau_0}$$

$$S_{v_i} = R_{v_i}(0) \cdot \frac{2\tau_0}{1 + (2\pi f \tau_0)^2} \cong R_{v_i}(0) \cdot 2\tau_0$$

## Derivation of Resistor Noise PSD (2)

Equipartition theorem: every energy storage element has an average noise of  $kT/2$

Intermediate result for the PSD of  $V$

Resistance expression

Carrier mobility

Resulting two-sided PSD  
 (positive and negative frequencies)

Single sided PSD

$$\frac{1}{2} m \overline{v_i^2} = \frac{kT}{2} \quad \overline{v_i^2} = R_{v_i}(0) = \frac{kT}{m}$$

$$S_V = N \left( \frac{Rq}{L} \right)^2 \frac{kT}{m} \cdot 2\tau_0$$

$$\left. \begin{aligned} R &= \frac{L}{\sigma A} = \frac{L}{qunA} = \frac{L^2}{q\mu N} \\ \mu &= \frac{q\tau_0}{m} \end{aligned} \right\} N = \frac{L^2 m}{q^2 R \tau_0}$$

$$S_V = 2kTR$$

$S_V = 4kTR$

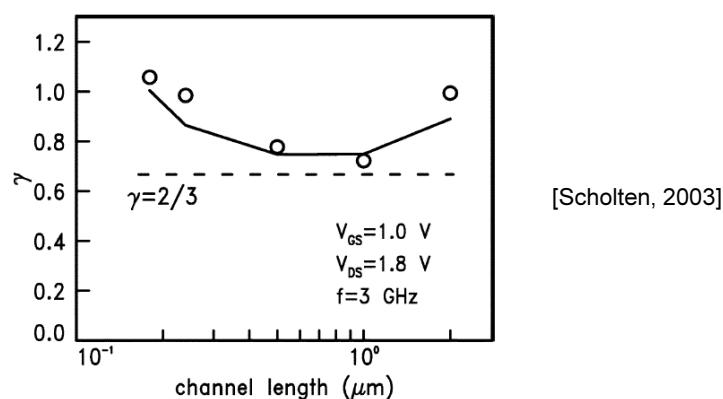
## MOSFET Thermal Noise (1)

- The noise of a MOSFET operating in the triode region is approximately equal to that of a resistor
- In the saturation region, the thermal noise of a MOSFET can be modeled using a drain current source with spectral density

$$\overline{i_d^2} = 4kT \cdot \gamma \cdot g_m \cdot \Delta f$$

- For an idealized long channel MOSFET, it can be shown that  $\gamma=2/3$
- For the past 10-15 years, researchers have been debating the value of  $\gamma$  in short channels
- Preliminary (wrong) results had suggested that in short channels  $\gamma$  can be as high as 5 due to “hot carrier” effects

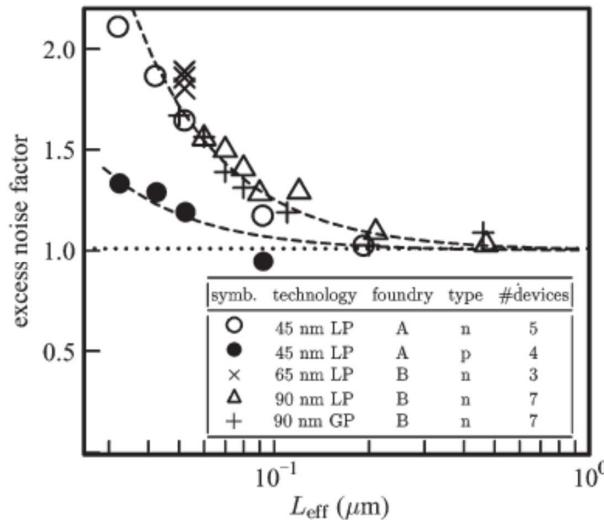
## MOSFET Thermal Noise (2)



- At moderate gate bias in strong inversion, the  $\gamma$  values for short-channel MOSFETs are slightly larger than 2/3
  - A. J. Scholten et al., "Noise modeling for RF CMOS circuit simulation," IEEE Trans. Electron Devices, pp. 618-632, Mar. 2003.
  - R. P. Jindal, "Compact Noise Models for MOSFETs," IEEE Trans. Electron Devices, pp. 2051-2061, Sep. 2006.

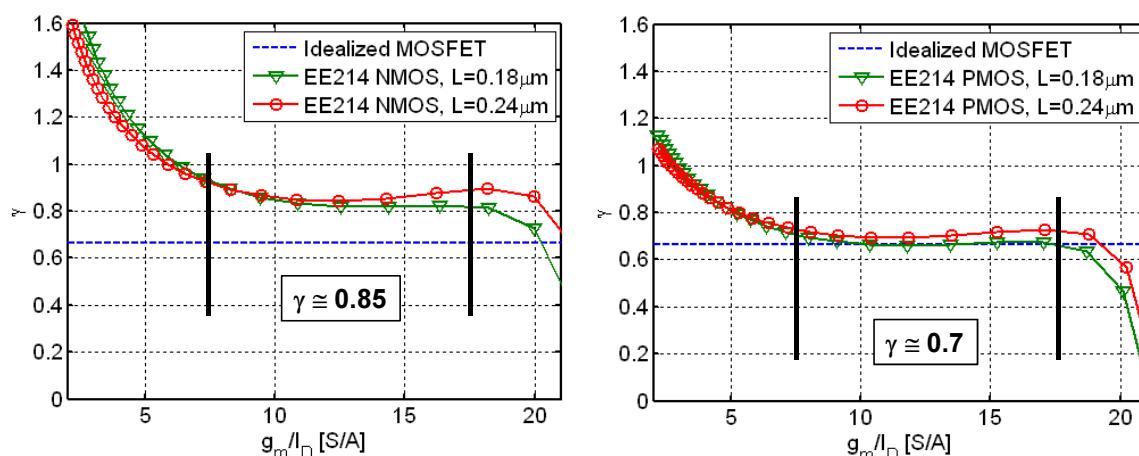
### MOSFET Thermal Noise (3)

Measured  $\gamma$   
relative to 2/3



G.D.J. Smit, A.J. Scholten, R.M.T. Pijper, R. van Langevelde, L.F. Tiemeijer, and D.B.M. Klaassen, "Experimental Demonstration and Modeling of Excess RF Noise in Sub-100-nm CMOS Technologies," IEEE Electron Device Letters, vol.31, no.8, pp. 884-886, Aug. 2010.

### Thermal Noise in EE214B MOSFET Devices



- Parameter  $\gamma$  depends on biasing conditions, but is roughly constant within a reasonable range of  $g_m/I_D$  used for analog design
- The EE214B HSpice models are inaccurate in the weak inversion region

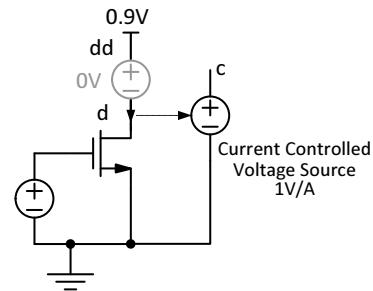
## Spice Simulation (1)

```
* EE214B MOS device noise simulation

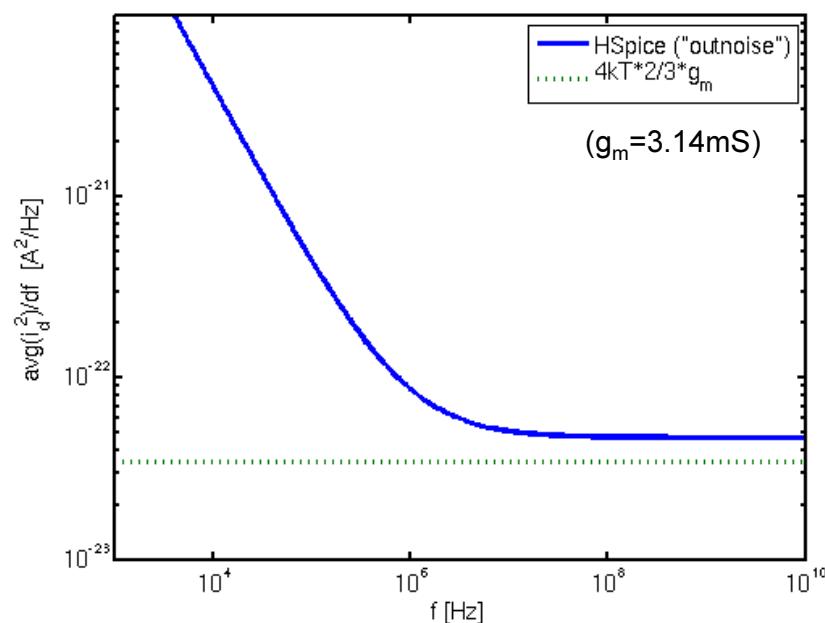
vd      dd  0  0.9
vm      dd  d  0
vg      g  0  dc 0.7 ac 1
mn1    d  g  0  0  nmos214  L=0.18u W=10u
h1      c  0  ccvv  vm  1

.op
.ac dec 100 10k 1gig
.noise v(c) vg

.options post brief
.inc '/usr/class/ee214b/hspice/ee214_hspice.sp'
.end
```



## Spice Simulation (2)



## 1/f Noise

- Also called "flicker noise" or "pink noise"
  - Caused by traps near Si/SiO<sub>2</sub> interface that randomly capture and release carriers, and also by mobility fluctuations
  - Occurs in virtually any device, but is most pronounced in MOSFETs
- Several (empirical) expressions exist to model flicker noise
  - The following expression is used in the EE214B HSpice models

$$\overline{i_{1/f}^2} = \frac{K_f}{C_{ox}} \frac{g_m^2}{W \cdot L} \frac{\Delta f}{f}$$

- For other models, see HSpice manual or
  - D. Xie et al., "SPICE Models for Flicker Noise in n-MOSFETs from Subthreshold to Strong Inversion," IEEE Trans. CAD, Nov. 2000
- K<sub>f</sub> is strongly dependent on technology; numbers for EE214B:
  - K<sub>f,NMOS</sub> = 0.5 · 10<sup>-25</sup> V<sup>2</sup>F
  - K<sub>f,PMOS</sub> = 0.25 · 10<sup>-25</sup> V<sup>2</sup>F

## 1/f Noise Corner Frequency

- By definition, the frequency at which the flicker noise density equals the thermal noise density

$$\frac{K_f}{C_{ox}} \frac{g_m^2}{W \cdot L} \frac{\Delta f}{f_{co}} = 4kT\gamma \cdot g_m \cdot \Delta f$$

$$\Rightarrow f_{co} = \frac{K_f}{4kT\gamma} \frac{1}{C_{ox}} \frac{g_m}{W \cdot L} = \frac{K_f}{4kT\gamma} \frac{1}{C_{ox}} \frac{1}{L} \left( \frac{g_m}{I_D} \right) \left( \frac{I_D}{W} \right)$$

- For a given g<sub>m</sub>/I<sub>D</sub> the only way to achieve lower f<sub>co</sub> is to use longer channel devices
  - In the above expression, both 1/L and I<sub>D</sub>/W are reduced for increasing L
- Example
  - EE214B NMOS, L = 0.18 μm, g<sub>m</sub>/I<sub>D</sub> = 12 S/A, ⇒ I<sub>D</sub>/W = 20 A/m  
⇒ f<sub>co</sub> = 560 kHz
- In newer technologies, f<sub>co</sub> can be on the order of 10 MHz

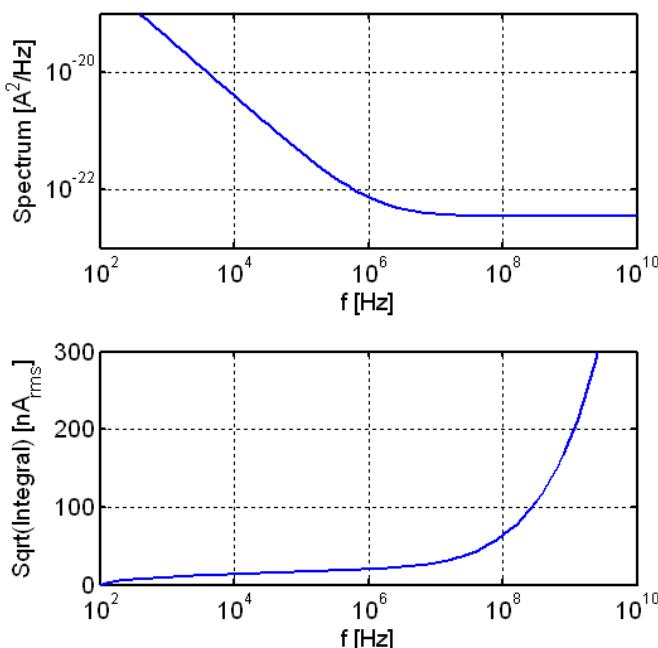
## 1/f Noise Contribution (1)

- Just as with white noise, the total 1/f noise contribution is found by integrating its power spectral density

$$\begin{aligned}\overline{i_{1/f, \text{tot}}^2} &= \int_{f_1}^{f_2} \frac{K_f}{C_{\text{ox}}} \frac{g_m^2}{W \cdot L} \frac{\Delta f}{f} \\ &= \frac{K_f}{C_{\text{ox}}} \frac{g_m^2}{W \cdot L} \ln\left(\frac{f_2}{f_1}\right) = \frac{K_f}{C_{\text{ox}}} \frac{g_m^2}{W \cdot L} 2.3 \log\left(\frac{f_2}{f_1}\right)\end{aligned}$$

- The integrated flicker noise depends on the number of frequency decades
  - The frequency range from 1Hz ... 10Hz contains the same amount of flicker noise as 1GHz ... 10GHz
  - Note that this is very different from thermal noise
- So, does flicker noise matter?
  - Let's look at the total noise integral (flicker and thermal noise)

## 1/f Noise Contribution (2)



- In the example shown on the left, the noise spectrum is integrated from 100Hz to 10GHz
- The contribution of the flicker noise is relatively small, even though its PSD dominates at low frequencies
- For circuits with very large bandwidth (beyond the 1/f corner), flicker noise is insignificant

## Analysis

$$PSD_{1/f} = \frac{K}{f} \quad PSD_{1/f}(f_{co}) = \frac{K}{f_{co}} = PSD_{th} \quad \Rightarrow K = PSD_{th} f_{co}$$

Integrated 1/f noise:  $N_{1/f} = \int_{f_1}^{f_2} \frac{K}{f} df = PSD_{th} f_{co} \ln\left(\frac{f_2}{f_1}\right) = PSD_{th} f_{co} \cdot 2.3 \log\left(\frac{f_2}{f_1}\right)$

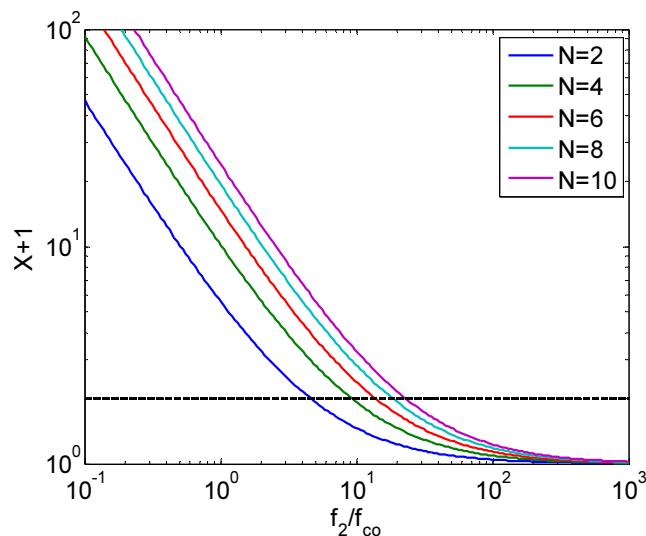
Integrated thermal noise:  $N_{th} = \int_{f_1}^{f_2} PSD_{th} df = PSD_{th}(f_2 - f_1) \cong PSD_{th} f_2$

Sum of thermal and 1/f:  $N_{th+1/f} \cong PSD_{th} f_2 \underbrace{\left(1 + 2.3 \frac{f_{co}}{f_2} \log\left(\frac{f_2}{f_1}\right)\right)}$

Relative flicker noise contribution  
 → Becomes small for large  $f_2/f_{co}$

## Plot

$$\left(1 + 2.3 \frac{f_{co}}{f_2} \log\left(\frac{f_2}{f_1}\right)\right) = \left(1 + 2.3 \frac{f_{co}}{f_2} N\right) = (1 + X) \quad N = \text{number of frequency decades}$$



Example:

For N=6, the thermal and 1/f components are equal (X=1) for  $f_2/f_{co} = 13.8$ .

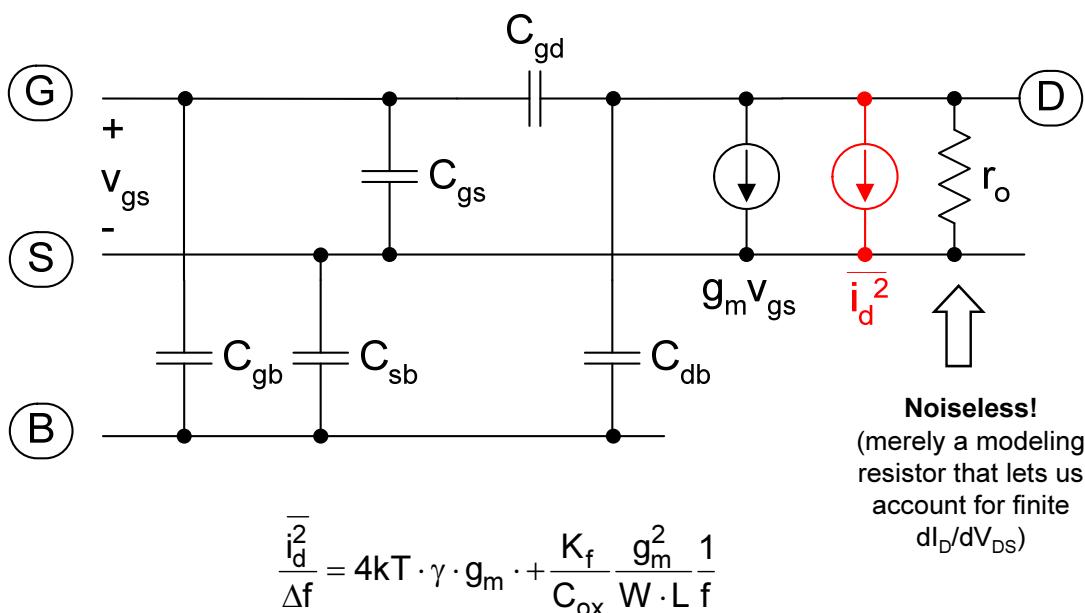
Beyond this frequency, the 1/f component decays rapidly.

Rule of thumb: Regardless of N, the 1/f contribution is small for  $f_2/f_{co} > 100$ .

## Lower Integration Limit

- Does the flicker noise PSD go to infinity for  $f \rightarrow 0$ ?
  - See e.g. E. Milotti, "1/f noise: a pedagogical review," available at <http://arxiv.org/abs/physics/0204033>
- Detailed analysis that considers the 1/f noise as a non-stationary process shows that the 1/f PSD holds up only to frequencies down to  $1/T_{ob}$ , where  $T_{ob}$  is the observation time
- Example
  - Say we are sensing a signal for a very long time (down to a very low frequency), e.g. 1 year  $\cong 32$  Msec, 1/year  $\cong 0.03$   $\mu\text{Hz}$
  - Number of frequency decades in 1/year to 100Hz  $\cong 10$
  - For the example on the previous slide, this means that the integration band changes from 8 to  $8+10=18$  decades
  - $\sqrt{18/8} = 1.5 \rightarrow$  Only 50% more flicker noise!

## MOS Model with Noise Generator

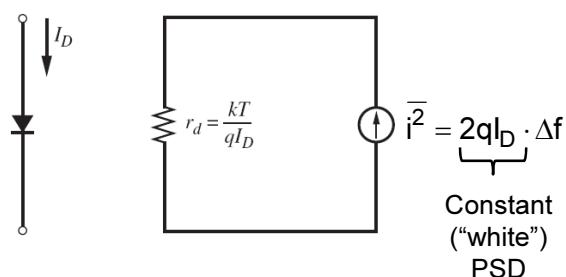


## Other MOSFET Noise Sources

- Gate noise
  - "Shot noise" from gate leakage current
  - Noise due to finite resistance of the gate material
  - Channel-induced gate noise (coupling via  $C_{gs}$ )
    - Relevant only at very high frequencies
    - See EE314A
- Bulk noise
- Source barrier noise in very short channels
  - Shot noise from carriers injected across source barrier
  - R. Navid, C. Jungemann, T. H. Lee and R. W. Dutton, "High-frequency noise in nanoscale metal oxide semiconductor field effect transistors," Journal of Applied Physics, vol. 101(12), pp. 101-108, June 15, 2007.
  - J. Jeon, J. Lee, J. Kim, C. H. Park, H. Lee, H. Oh, H.-K. Kang, B.-G. Park, and H. Shin, "The first observation of shot noise characteristics in 10-nm scale MOSFETs," in Proc. Symp. VLSI Technol., 2009, pp. 48-49.

## Shot Noise in a PN Junction

- Shot noise is generally associated with the flow of a DC current
- In a forward biased diode, shot noise occurs due to randomness in the carrier transitions across the PN junction (energy barrier)
- The power spectral density of this noise is white up to very high frequencies
- The noise can be included in the small-signal model as shown below

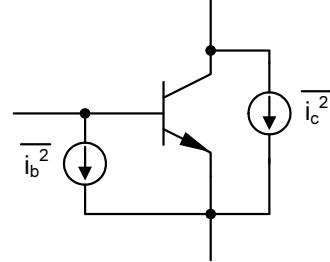


## Shot Noise in a Bipolar Transistor

- In a bipolar transistor, the flow of DC current into the base and collector causes shot noise
- The noise can be modeled via equivalent current generators

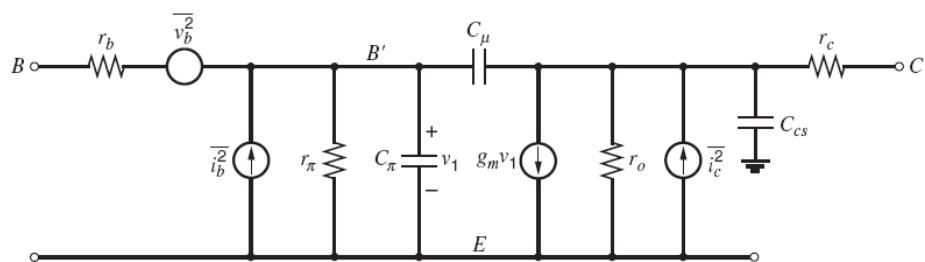
$$\overline{i_c^2} = 2qI_C\Delta f = 2kTg_m\Delta f$$

$$\overline{i_b^2} = 2qI_B\Delta f = 2kT \frac{g_m}{\beta} \Delta f = 2kT \frac{1}{r_\pi} \Delta f$$



- The base and collector noise currents are statistically independent as they arise from separate physical mechanisms
  - This will be important in the context of circuit noise calculations

## BJT Small Signal Model with Noise Generators



Thermal noise due to physical  $r_b$ :  $\overline{v_b^2} = 4k T r_b \Delta f$

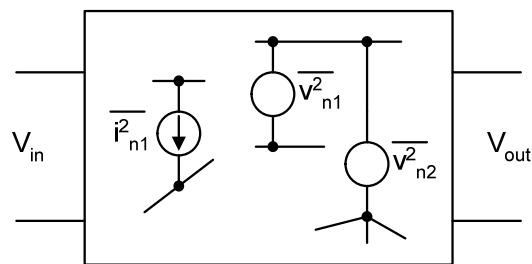
Collector shot noise:

$$\overline{i_c^2} = 2qI_C \Delta f \quad \text{Typically negligible}$$

Base noise components:

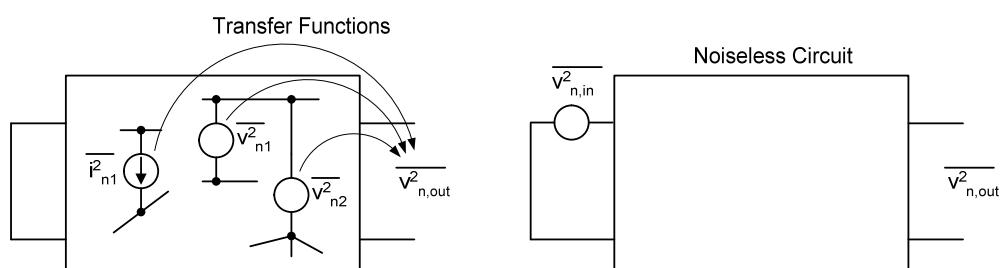
$$\overline{i_b^2} = \underbrace{2qI_B \Delta f}_{\text{Shot noise}} + K_1 \underbrace{\frac{I_B^a}{f} \Delta f}_{\text{Flicker noise}} + K_2 \underbrace{\frac{I_B^c}{1 + \left(\frac{f}{f_c}\right)^2} \Delta f}_{\text{Burst noise}}$$

## Noise in Circuits (1)



- Most circuits have more than one relevant noise source
- In order to quantify the net effect of all noise sources, we must refer the noise sources to a single "interesting" port of the circuit
  - Usually the output or input
- In the following discussion, we will first consider only circuits with a perfect voltage drive, i.e. no source resistance  $R_s$ 
  - Inclusion of finite  $R_s$  will be discussed later

## Noise in Circuits (2)



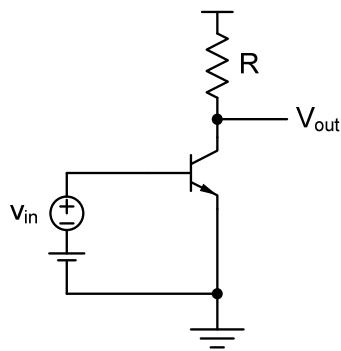
### Output referred noise

- Refer noise to output via individual noise transfer functions
- Physical concept, exactly what one would measure in the lab

### Input referred noise

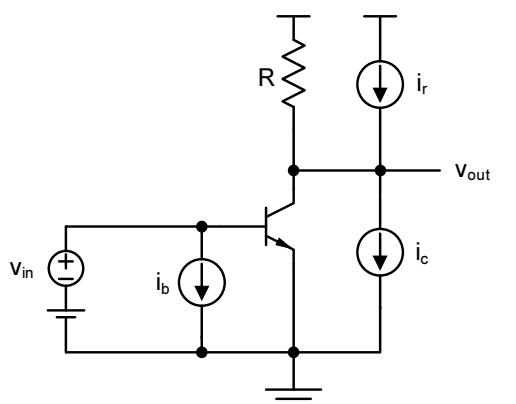
- Represent total noise via a fictitious input source that captures all circuit-internal noise sources
- Useful for direct comparison with input signal

## Circuit Example



- For simplicity, let's neglect
  - Source impedance
  - All capacitances
  - Burst and flicker noise
  - $r_o$ ,  $r_b$ , and  $r_\mu$

## Circuit With Noise Sources

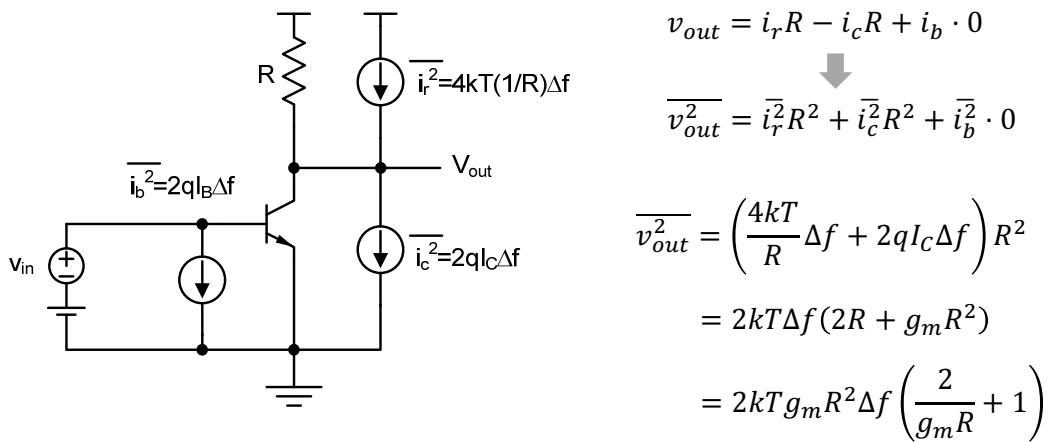


Output voltage due to the noise sources (superposition):

$$v_{out} = i_r R - i_c R + i_b \cdot 0$$

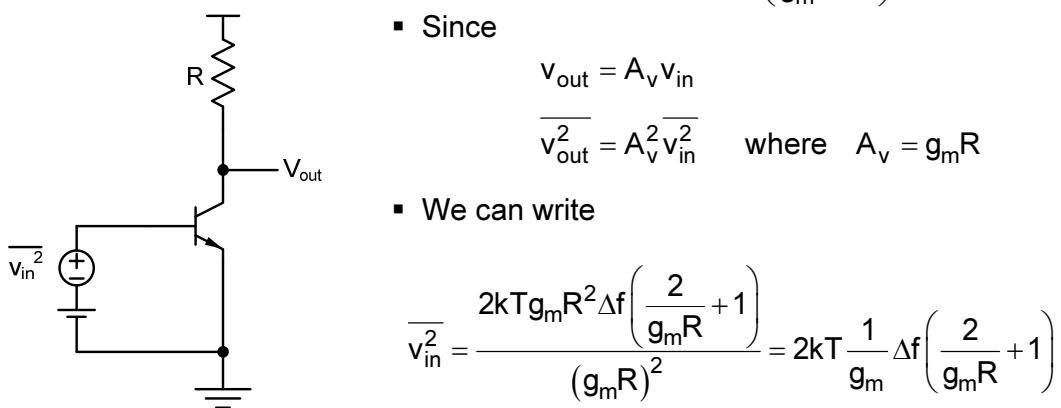
- To get started with noise analysis, simply treat the noise sources like any other deterministic current injected into the circuit
- Shown polarities of  $i_b$ ,  $i_r$  and  $i_c$  are arbitrary
- The noise due to  $i_b$  is irrelevant (absorbed by the ideal input source)

## Output Noise PSD



- Key observation: For large voltage gain ( $g_m R$ ), the collector shot noise ( $i_c$ ) dominates

## Input Referred Noise PSD



- Larger  $g_m$  translates into lower input referred voltage noise

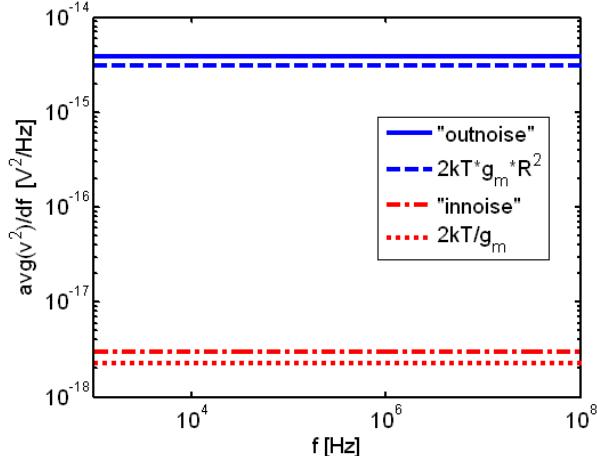
## Spice Simulation

```
*** EE214B BJT noise example
*** biasing
ib vcc vb      100u
q1 vb  vb  0  npn214
c1 vb  0      1

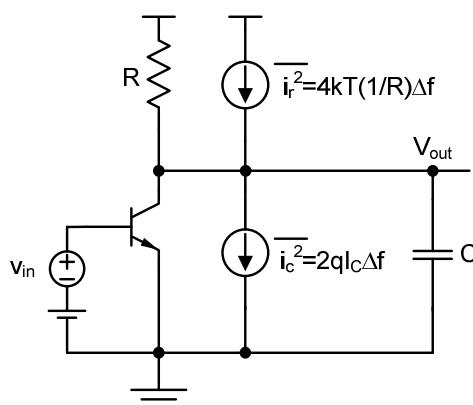
*** main circuit
v1 vcc 0      2.5
vi vi  vb     ac 1
rl vcc vo     10k
q2 vo  vi  0  npn214

.op
.ac dec 100 100 10e9
.noise v(vo) vi
.options post brief
.inc 'ee214_hspice.sp'
.end
```

$$g_m = 3.67 \text{ mS}, R = 10 \text{ k}\Omega, A_v = 36.7$$



## Output Referred Noise PSD with Load Capacitance



$$\begin{aligned} \overline{V_{out}^2} &= \left( 4kT \frac{1}{R} \Delta f + 2qI_C \Delta f \right) \cdot \left| R \parallel \frac{1}{j\omega C} \right|^2 \\ &= \left( 4kT \frac{1}{R} \Delta f + 2qI_C \Delta f \right) \cdot R^2 \left| \frac{1}{1 + j\omega RC} \right|^2 \\ &= 2kT g_m \Delta f \cdot R^2 \left( \frac{2}{g_m R} + 1 \right) \cdot \left| \frac{1}{1 + j\omega RC} \right|^2 \end{aligned}$$

- Same calculation as before, except that now the noise current drops into the parallel combination of R and C
- Output PSD is shaped by squared magnitude of first order response

## Input Referred Noise PSD with Load Capacitance

- Same calculation as before, except that the voltage gain is now frequency dependent

$$\overline{v_{in}^2}(\omega) = \frac{\overline{v_{out}^2}(\omega)}{|A_v(j\omega)|^2} \quad \text{where} \quad |A_v(j\omega)|^2 = A_v(0) \left| \frac{1}{1+j\omega RC} \right|^2$$

$$= \frac{2kTg_m \Delta f \cdot R^2 \left( \frac{2}{g_m R} + 1 \right) \cdot \left| \frac{1}{1+j\omega RC} \right|^2}{A_v^2(0) \left| \frac{1}{1+j\omega RC} \right|^2}$$

- Input referred noise is frequency independent, because the output noise and gain have the same frequency roll-off

$$\therefore \overline{v_{in}^2} = 2kT \frac{1}{g_m} \Delta f \left( \frac{2}{g_m R} + 1 \right)$$

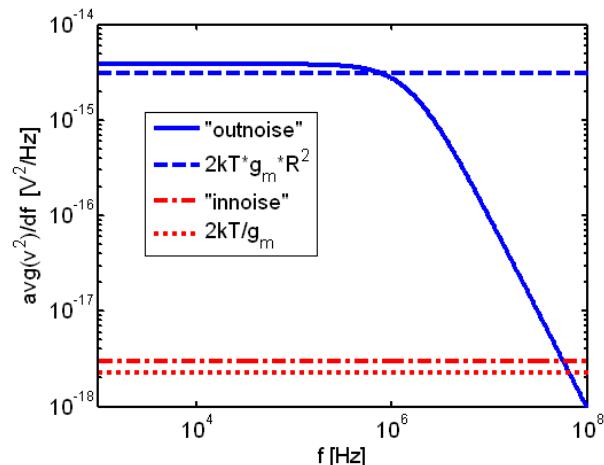
## Spice Simulation

```
*** EE214B BJT noise example
*** biasing
ib vcc vb      100u
q1 vb  vb  0  npn214
c1 vb  0       1

*** main circuit
v1 vcc 0      2.5
vi vi  vb     ac 1
rl vcc vo     10k
C1 vo  0      10p
q2 vo  vi  0  npn214

.op
.ac dec 100 100 10e9
.noise v(vo) vi
.options post brief
.inc 'ee214_hspice.sp'
.end
```

$g_m=3.67\text{mS}$ ,  $R=10\text{k}$ ,  $A_v=36.7$ ,  $C=10\text{pF}$



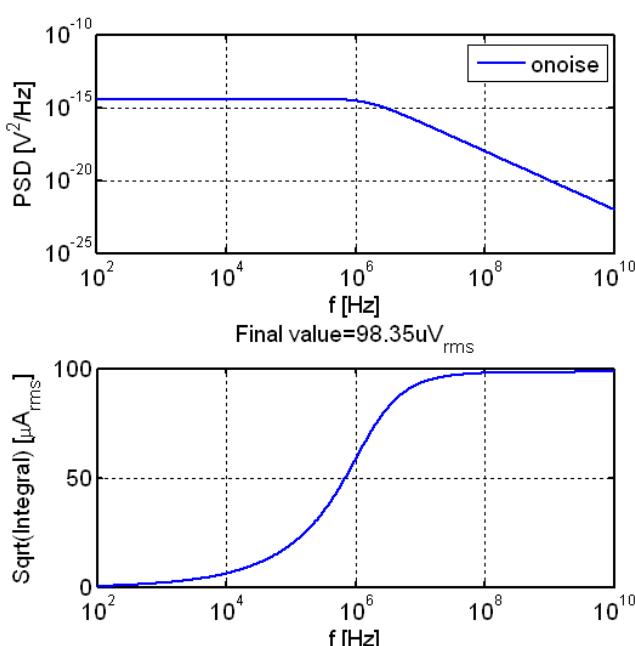
## Signal-to-Noise Ratio

- Assuming a sinusoidal signal, we can compute the SNR at the output of the circuit using

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}} = \frac{\frac{1}{2} \hat{V}_{\text{out}}^2}{\int_{f_1}^{f_2} \frac{V_{\text{out}}^2}{\Delta f} \cdot df}$$

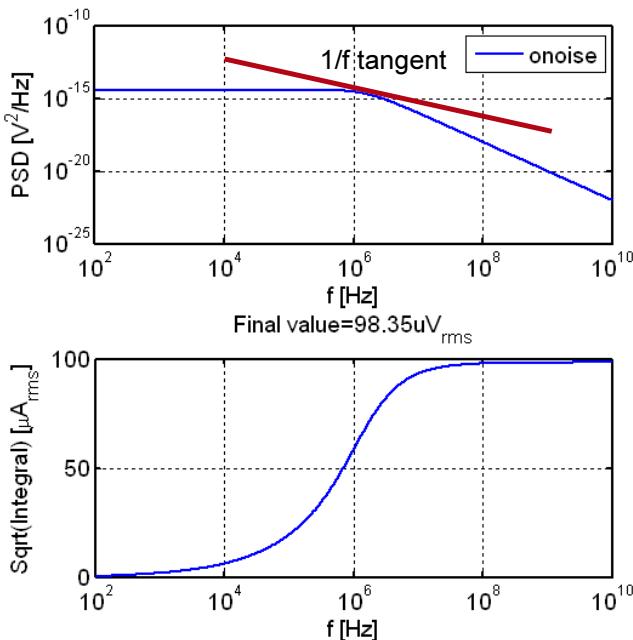
- Over which bandwidth should we integrate the noise?
- Two interesting cases
  - The output is measured or observed by a system with finite bandwidth (e.g. human ear, or another circuit with finite bandwidth)
    - Use frequency range of that system as integration limits
    - Applies on a case by case basis
  - Total integrated noise
    - Integrate noise from zero to “infinite” frequency

## A Closer Look at The Circuit’s Noise Integral



- The noise integral converges for upper integration limits that lie beyond the circuit’s pole frequency
- The total integrated noise (from “0” to “infinity”) is a reasonable metric to use
  - For convenience in comparing circuits without making bandwidth assumptions
  - In a circuit where the output is observed without any significant band limiting
    - E.g. in a sampling circuit
    - See EE315

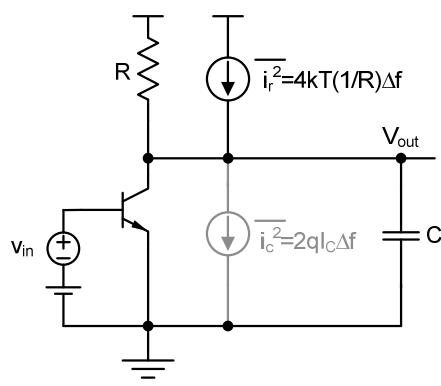
## Aside: A Cool Trick



- To find out at which frequency most of the noise rolls in, drop a  $1/f$  tangent in the log-log PSD plot (from the top)
- The frequency range where the tangent touches the PSD is the strongest contributor to the integrated noise
- See text (p. 377) for more info

## Total Integrated Noise Calculation

- Let us first consider only the noise from the resistor

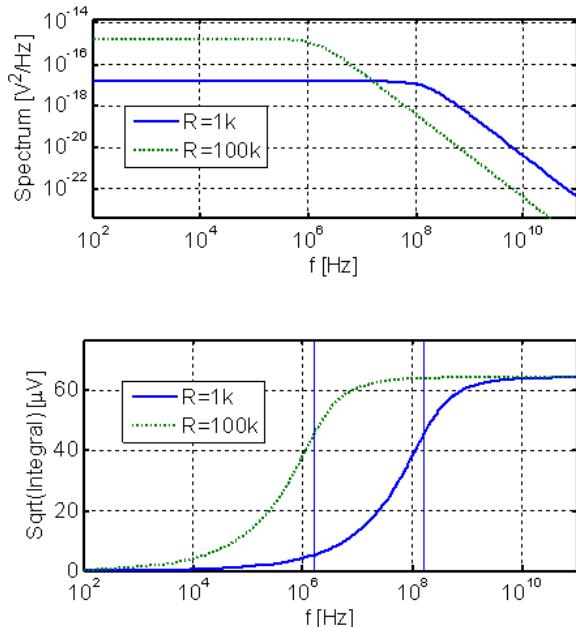


$$\begin{aligned}
 \overline{v_{out,tot}^2} &= \int_0^\infty 4kTR \cdot \left| \frac{1}{1+j2\pi f \cdot RC} \right|^2 df \\
 &= 4kTR \int_0^\infty \frac{df}{1+(2\pi f RC)^2}; \quad \int \frac{du}{1+u^2} = \tan^{-1} u \\
 &= 4kTR \cdot \frac{1}{4RC} \\
 &= \frac{kT}{C}
 \end{aligned}$$

- Interesting result
  - The total integrated noise at the output depends only on  $C$  (even though  $R$  is generating the noise)

## Effect of Varying R

- Increasing R increases the noise power spectral density, but also decreases the bandwidth
  - R drops out in the end result
- For C=1pF (example to the right), the total integrated noise is approximately  $64\mu V_{rms}$



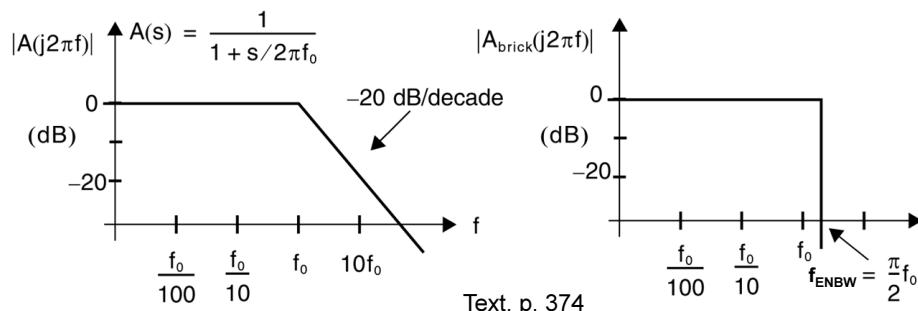
## Equivalent Noise Bandwidth

$$\overline{v_{out,tot}^2} = \int_0^\infty 4kTR \cdot \left| \frac{1}{1 + j2\pi f / \omega_0} \right|^2 df$$

$$= 4kTR \cdot \frac{\omega_0}{4} = 4kTR \cdot f_{ENBW}$$

$$f_{ENBW} = \frac{\omega_0}{4} = \frac{\pi}{2} f_0$$

- The equivalent noise bandwidth ( $f_{ENBW}$ ) is defined as the bandwidth of a brick-wall filter that results in the same total noise power as the filter in question
- For a first order filter, the equivalent noise bandwidth is  $\pi/2$  times its 3-dB corner frequency



Text, p. 374

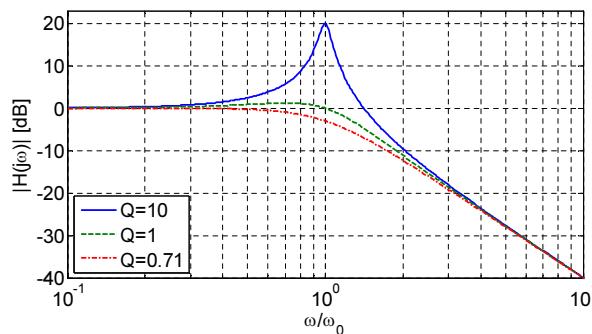
## Aside: Equivalent Noise Bandwidth for a Second Order Filter

$$\int_0^\infty 4kTR \cdot \left| \frac{1}{1 + s \frac{Q}{\omega_0} + \frac{s^2}{\omega_0^2}} \right|^2 df$$

$$= 4kTR \cdot \frac{\omega_0 Q}{4} = 4kTR \cdot f_{ENBW}$$

$$f_{ENBW} = \frac{\omega_0 Q}{4}$$

- For a second order filter, the equivalent noise bandwidth is proportional to Q
- Q controls the frequency domain peaking of the transfer function
- Large Q → large noise
  - This is also clear from the noise tangent argument shown earlier



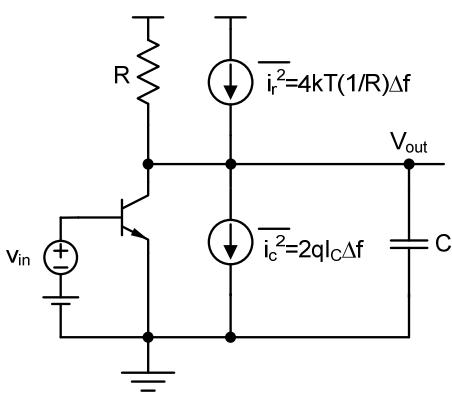
## Alternative Derivation of kT/C Result

- The equipartition theorem says that each degree of freedom (or energy state) of a system in thermal equilibrium holds an average energy of  $kT/2$
- In our circuit, the quadratic degree of freedom is the energy stored on the capacitor

$$\overline{\frac{1}{2}Cv_{out}^2} = \frac{1}{2}kT$$

$$\overline{v_{out}^2} = \frac{kT}{C}$$

## Total Integrated Noise Calculation for the Complete Circuit



$$\overline{V_{out,tot}^2} = \int_0^\infty (4kTR + 2kTg_mR^2) \cdot \left| \frac{1}{1+j2\pi f \cdot RC} \right|^2 df$$

Was  $4kTR$  in previous analysis

$$\begin{aligned}\overline{V_{out,tot}^2} &= \frac{kT}{C} \cdot \frac{4kTR + 2kTg_mR^2}{4kTR} \\ &= \frac{kT}{C} \left( 1 + \frac{1}{2} g_m R \right)\end{aligned}$$

- Taking the BJT's collector shot noise into account, the total integrated noise becomes a multiple of  $kT/C$

## Example SNR Calculation

- Assumptions
  - Output carries a sinusoid with 1V peak amplitude
  - We observe the output without significant band limiting and thus use the total integrated noise in the SNR expression

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{\frac{1}{2} \hat{V}_{out}^2}{\frac{kT}{C} \left( 1 + \frac{1}{2} g_m R \right)} = \frac{0.5V^2}{\frac{kT}{10pF} \left( 1 + \frac{1}{2} 3.67mS \cdot 10k\Omega \right)} = \frac{0.5V^2}{(763\mu V)^2} = 8.59 \cdot 10^6$$

$$SNR[\text{dB}] = 10 \log(8.59 \cdot 10^6) = 69.3 \text{ dB}$$

- Typical system requirements
  - Audio:  $SNR \geq 100 \text{ dB}$
  - Video:  $SNR \geq 60 \text{ dB}$
  - Gigabit Ethernet Transceiver:  $SNR \geq 35 \text{ dB}$

## Noise/Power Tradeoff

- Assuming that we're already using the maximum available signal swing, improving the SNR by 6dB means
  - Increase C by 4x
  - Decrease R by 4x to maintain bandwidth
  - Increase  $g_m$  by 4x to preserve gain
  - Increase collector current by 4x
- Bottom line
  - Improving the SNR in a noise limited circuit by 6dB ("1bit") **QUADRUPLES power dissipation !**

## MDS and DR

- Minimum detectable signal (MDS)
  - Quantifies the signal level in a circuit that yields SNR=1, i.e. noise power = signal power
- Dynamic range (DR) is defined as

$$DR = \frac{P_{\text{signal,max}}}{MDS}$$

- If the noise level in the circuit is independent of the signal level (which is often, but not always the case), it follows that the DR is equal to the "peak SNR," i.e. the SNR with the maximum signal applied

## Does Thermal Noise Always Matter?

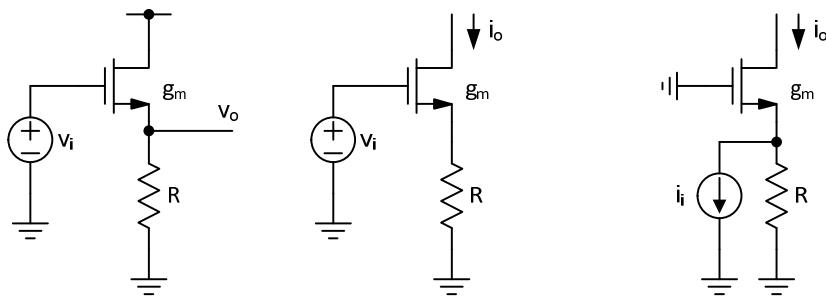
- Let's look at the SNR of an RC circuit with a 1-V sinusoid applied, considering the total integrated noise ( $kT/C$ )

SNR [dB]	C [pF]	
20	0.00000083	
40	0.000083	
60	0.0083	
80	0.83	
100	83	
120	8300	
140	830000	

Hard to make such small capacitors...  
 Designer will be concerned about thermal noise; component sizes often set by SNR  
 A difficult battle with thermal noise ...

- Rules of thumb
  - Up to SNR  $\sim 30\text{-}40\text{dB}$ , integrated circuits are usually not limited by thermal noise
  - Achieving SNR  $>100\text{dB}$  is extremely difficult
    - Must usually rely on external components, or reduce bandwidth and remove noise by a succeeding filter
    - See e.g. oversampling ADCs in EE315

## Noise of Other Elementary Stages



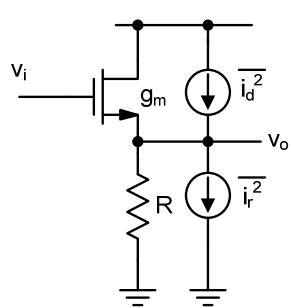
$$A_v = \frac{v_o}{v_i} = \frac{g_m R}{1 + g_m R}$$

$$G_m = \frac{i_o}{v_i} = \frac{g_m}{1 + g_m R}$$

$$A_i = \frac{i_o}{i_i} = \frac{g_m R}{1 + g_m R}$$

- Neglecting finite  $r_o$ , backgate effect and flicker noise for simplicity

## Source Follower



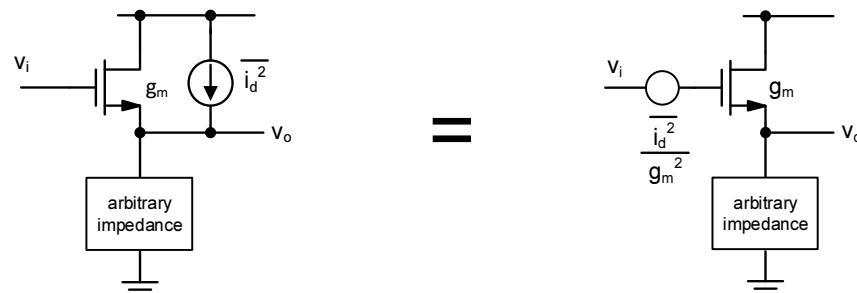
$$\overline{v_o^2} = \left( \overline{i_d^2} + \overline{i_r^2} \right) \cdot \left( \frac{1}{g_m + \frac{1}{R}} \right)^2 = \frac{\overline{i_d^2} + \overline{i_r^2}}{g_m^2} \cdot A_v^2$$

Often negligible

$$\overline{v_i^2} = \frac{\overline{i_d^2}}{g_m^2} + \frac{\overline{i_r^2}}{g_m^2} = 4kT \frac{1}{g_m} \Delta f \left( \gamma + \frac{1}{g_m R} \right)$$

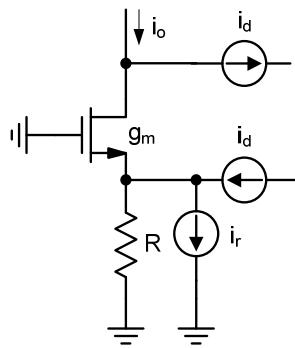
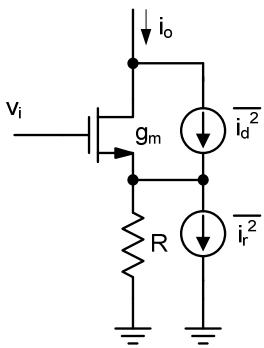
- The noise in a resistively loaded source follower is typically dominated by the contribution from the transistor
- The input referred noise voltage can be approximated by the drain current noise reflected through the device's transconductance

## Source Follower – Useful Generalization



- Why does this work?
- One way to think about this
  - First assume “arbitrary impedance” is removed → Equivalence is obvious in this case
  - Connecting arbitrary impedance back affects both circuits in the same way, hence equivalence is maintained

## Degenerated Common Source Stage



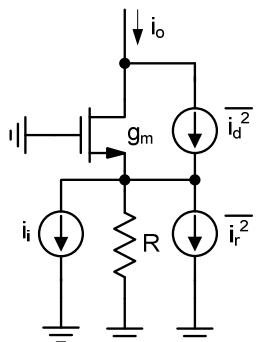
$$\begin{aligned}
 i_o &= i_d + (i_r - i_d) \frac{g_m}{g_m + \frac{1}{R}} \\
 &= i_r \frac{g_m R}{1 + g_m R} + i_d \frac{1}{1 + g_m R} \\
 &= i_r G_m R + i_d \frac{G_m}{g_m}
 \end{aligned}$$

$$\overline{i_o^2} = \left( \frac{\overline{i_d^2}}{g_m^2} + \overline{i_r^2} R^2 \right) \cdot G_m^2$$

$$\overline{v_i^2} = \frac{\overline{i_d^2}}{g_m^2} + \overline{i_r^2} R^2 = 4kT\gamma \frac{1}{g_m} \Delta f + 4kTR\Delta f$$

- The input referred voltage noise consists of drain current noise, reflected through  $g_m$ , plus the resistor's voltage noise

## Common Gate Stage



$$\begin{aligned}
 i_o &= i_r \frac{g_m R}{1 + g_m R} + i_d \frac{1}{1 + g_m R} \\
 &= i_r A_i + i_d \frac{A_i}{g_m R}
 \end{aligned}$$

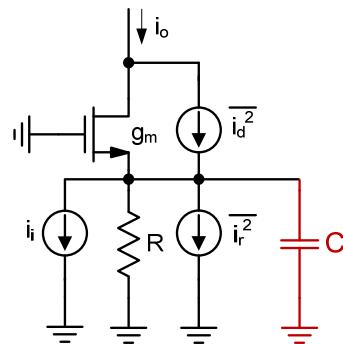
$$\overline{i_o^2} = \left( \frac{\overline{i_d^2}}{g_m^2 R^2} + \overline{i_r^2} \right) \cdot A_i^2$$

$$\overline{i_r^2} = \frac{\overline{i_d^2}}{g_m^2 R^2} + \overline{i_r^2} = 4kT \frac{1}{R} \Delta f \left( 1 + \frac{\gamma}{g_m R} \right)$$

Often negligible

- The input referred current noise from the transistor is often negligible (at low frequencies)
- The noise tends to be dominated by the devices providing the source and drain bias currents (resistors or current sources)

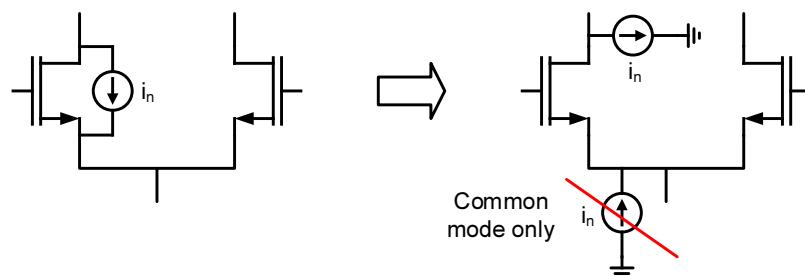
## Common Gate Stage at High Frequencies



$$\begin{aligned}\overline{i_i^2} &= 4kT \frac{1}{R} \Delta f + 4kT\gamma g_m \Delta f \frac{\left| \frac{1}{R} + j\omega C \right|^2}{g_m^2} \\ &\approx 4kT \frac{1}{R} \Delta f + 4kT\gamma g_m \Delta f \left( \frac{\omega C}{g_m} \right)^2\end{aligned}$$

- The input referred current noise from the transistor can be significant at high frequencies (near the cutoff frequency of the current transfer)

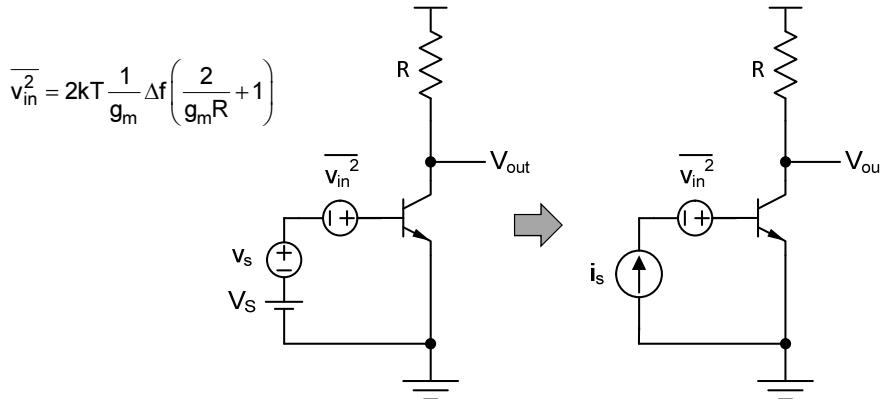
## Noise in a Differential Pair



- Bottom line: The noise of a differential pair can be analyzed using a common-source half circuit

## CE Example Revisited: More on Input-Referred Noise

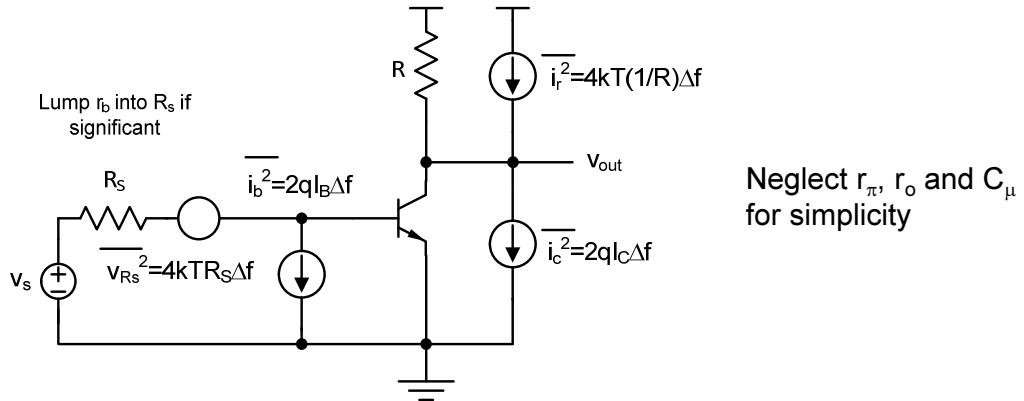
- In our previous analysis, we showed that the circuit's noise can be lumped into a single input-referred voltage noise generator
- It is important to remember that this result was based on the assumption of ideal voltage drive
- To see why this matters, consider driving the circuit with a current source → The circuit would appear noiseless at the output!



## How to Properly Deal with Input Referred Noise

- Case 1: Applies to 99% of what we do in this class
  - Compute the input source referred noise for the given (fixed) source type (voltage or current) and take the given (fixed) source resistance into account
  - The recommended procedure is the same as before
    - Refer all noise sources to the output
    - Divide by the transfer function from the input source to the output
- Case 2: What if we don't know the details of the driving network?
  - This case occurs only in very specific cases
  - Examples
    - You are trying to design/sell a general purpose OpAmp
    - You are trying to benchmark a transistor outside a specific circuit
  - This case requires the use of both input-referred current and voltage generators

## Example with Known and Finite $R_s$ (Case 1)



$$H(s) = \frac{v_{out}}{v_s} = -\frac{g_m R}{1 + s R_s C_\pi}$$

$$\overline{\frac{v_{out}^2}{\Delta f}} = (4kT R_s + 2qI_B R_s^2) |H(s)|^2 + \left( 4kT \frac{1}{R} + 2qI_C \right) \cdot R^2$$

## Input Source Referred Noise PSD

$$\overline{\frac{v_{out}^2}{\Delta f}} = (4kT R_s + 2qI_B R_s^2) |H(s)|^2 + \left( 4kT \frac{1}{R} + 2qI_C \right) \cdot R^2$$

$$\overline{\frac{v_s^2}{\Delta f}} = \frac{\overline{\frac{v_{out}^2}{\Delta f}}}{|H(s)|^2} = 4kT R_s + 2qI_B R_s^2 + \frac{\left( 4kT \frac{1}{R} + 2qI_C \right) \cdot R^2}{|H(s)|^2}$$

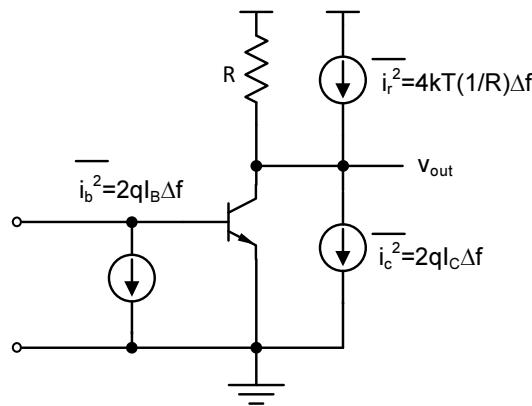
For low frequencies ( $|H(s)| \approx g_m R$ ) we obtain

$$\overline{\frac{v_s^2}{\Delta f}} = 4kT R_s + 2qI_B R_s^2 + \frac{\left( 4kT \frac{1}{R} + 2qI_C \right) \cdot R^2}{(g_m R)^2}$$

$$\overline{\frac{v_s^2}{\Delta f}} = 4kT \left( R_s + \frac{g_m R_s^2}{2\beta} + \frac{R}{(g_m R)^2} + \frac{1}{2g_m} \right)$$

Note that this equation has a minimum for a specific value of  $g_m$  (and thus  $I_C$ )

## Example with Unknown $R_s$ (Case 2)



- How to communicate the input-referred noise performance to another engineer without assuming anything about  $R_s$ ?

## Example: General Purpose OpAmp



**Ultraprecision  
Operational Amplifier**

**OP177**

### FEATURES

Ultralow offset voltage  
 $T_A = 25^\circ C$ ,  $25 \mu V$  maximum  
 Outstanding offset voltage drift  $0.1 \mu V/\text{ }^\circ C$  maximum  
 Excellent open-loop gain and gain linearity  
 $12 \text{ V}/\mu \text{V}$  typical  
 CMRR:  $130 \text{ dB}$  minimum  
 PSRR:  $115 \text{ dB}$  minimum  
 Low supply current  $2.0 \text{ mA}$  maximum  
 Fits industry-standard precision op amp sockets

### PIN CONFIGURATION

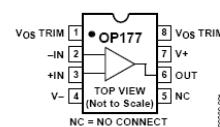
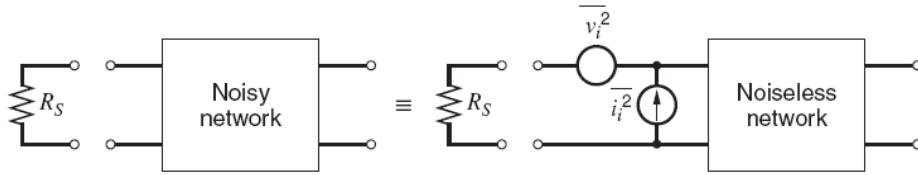


Figure 1.8-Lead PDIP (P-Suffix),  
 8-Lead SOIC (S-Suffix)

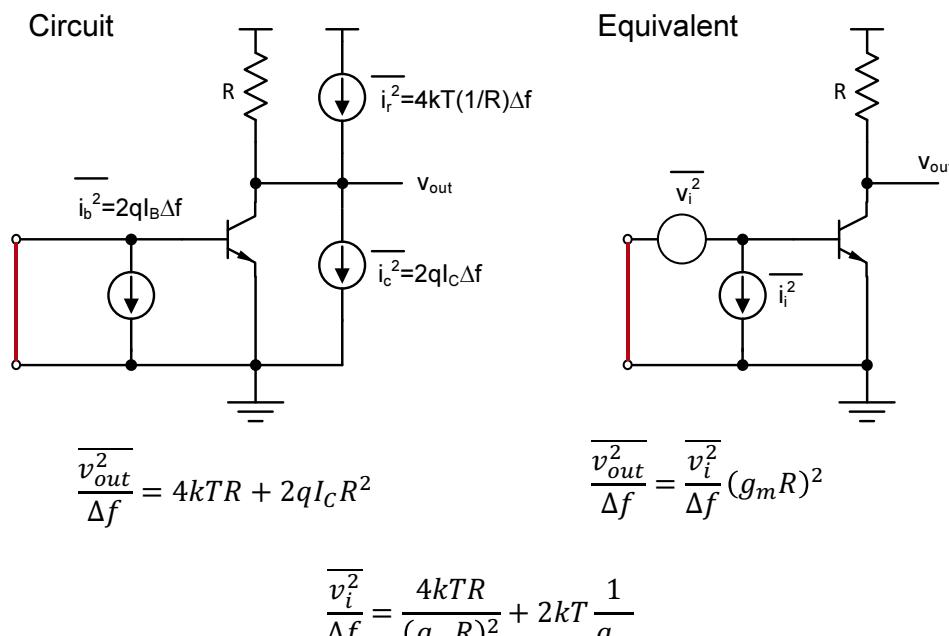
Parameter	Symbol	Conditions	OP177F			OP177G			Unit
INPUT OFFSET VOLTAGE	$V_{os}$		Min	Typ	Max	Min	Typ	Max	$\mu V$
LONG-TERM INPUT OFFSET <sup>1</sup>									
Voltage Stability	$\Delta V_{os}/\text{time}$			0.3			0.4		$\mu V/\text{mo}$
INPUT OFFSET CURRENT	$I_{os}$			0.3	1.5		0.3	2.8	nA
INPUT BIAS CURRENT	$I_B$		-0.2	+1.2	+2	-0.2	+1.2	+2.8	nA
INPUT NOISE VOLTAGE	$e_n$	$f_0 = 1 \text{ Hz to } 100 \text{ Hz}^2$		118	150		118	150	nVrms
INPUT NOISE CURRENT	$i_n$	$f_0 = 1 \text{ Hz to } 100 \text{ Hz}^2$		3	8		3	8	pA rms

## Using Equivalent Voltage and Current Noise Generators



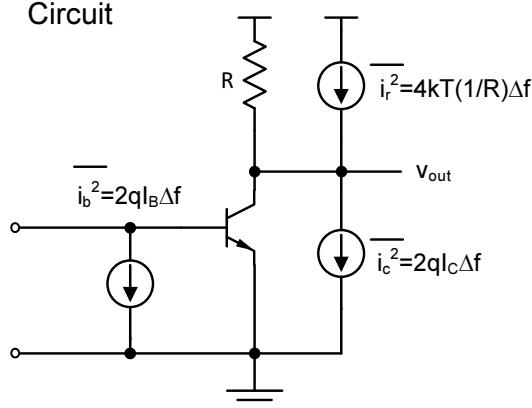
- Short-circuit both inputs and equate output noise
  - This yields  $\overline{v_i^2}$
- Open-circuit both inputs and equate output noise
  - This yields  $\overline{i_i^2}$
- This representation is valid for “any” source impedance
- Must consider correlation between equivalent voltage and current generator, but often times only one of the two generators matters in the target application
  - If both generators matter (and they are correlated), it is usually best to avoid working with input-referred noise representations!

## Step 1

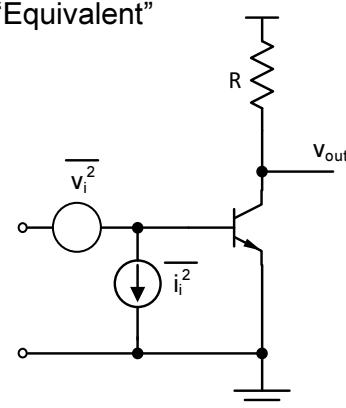


## Step 2

Circuit



“Equivalent”



$$\frac{\overline{v_{out}^2}}{\Delta f} = 4kTR + 2qI_C R^2 + 2qI_B \beta^2 R^2$$

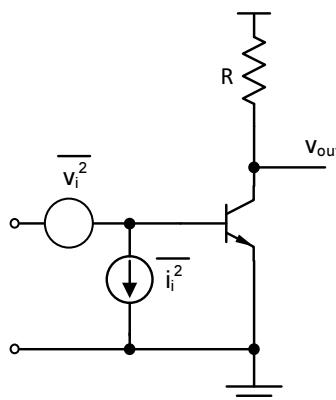
$$\frac{\overline{v_{out}^2}}{\Delta f} = \frac{\overline{i_i^2}}{\Delta f} \beta^2 R^2$$

$$\frac{\overline{i_i^2}}{\Delta f} = \frac{4kT}{\beta^2} \frac{1}{R} + \frac{2qI_C}{\beta^2} + 2qI_B$$

## Resulting Model

$$\frac{\overline{v_i^2}}{\Delta f} = \frac{4kTR}{(g_m R)^2} + 2kT \frac{1}{g_m}$$

$$\frac{\overline{i_i^2}}{\Delta f} = \frac{4kT}{\beta^2} \frac{1}{R} + \frac{2qI_C}{\beta^2} + 2qI_B$$



- Homework problem
  - Drive this model with a voltage source and given  $R_s$
  - Compute input source referred noise
  - Explain why the result is not the same as on slide 67!

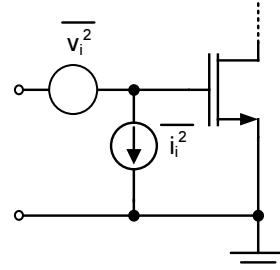
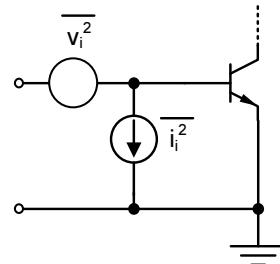
## Comparison: BJT vs. MOS

$$\frac{\overline{v_i^2}}{\Delta f} = 2kT \frac{1}{g_m} + 4kTr_b$$

$$\frac{\overline{i_i^2}}{\Delta f} = \frac{2qI_C}{|\beta(j\omega)|^2} + 2qI_B$$

$$\frac{\overline{v_i^2}}{\Delta f} = 4kT\gamma \frac{1}{g_m} + \frac{K_f}{WLC_{ox}f} \frac{1}{f}$$

$$\frac{\overline{i_i^2}}{\Delta f} = \left(\frac{\omega}{\omega_T}\right)^2 \left(4kT\gamma g_m + \frac{K_f}{WLC_{ox}} g_m^2\right)$$



## Comparison: BJT vs. MOS

- For low source impedance → voltage noise will dominate
  - BJT is usually superior
    - Need less  $gm$  for approximately same noise
    - $g_m/I$  is higher, making it easier to achieve low noise at a given current budget
  
- For high source impedance → current noise will dominate
  - MOS is usually superior
    - Since there is no “base current”
    - Unless the gate oxide becomes very thin, as in a 45-nm CMOS process that does not use high-k gate dielectrics

## Additional Topics in Noise Analysis

- Later in this course: Effect of noise in feedback circuits
- Covered in EE314A
  - RF-centric metrics
    - Noise figure
    - Receiver sensitivity
  - Phase noise in oscillators
- Covered in EE315
  - Noise in filters and switched capacitor circuits
- Other
  - Cyclostationary noise
    - Noise in circuits that are driven by a periodic waveform that modulates the power spectral densities
    - E.g. mixers

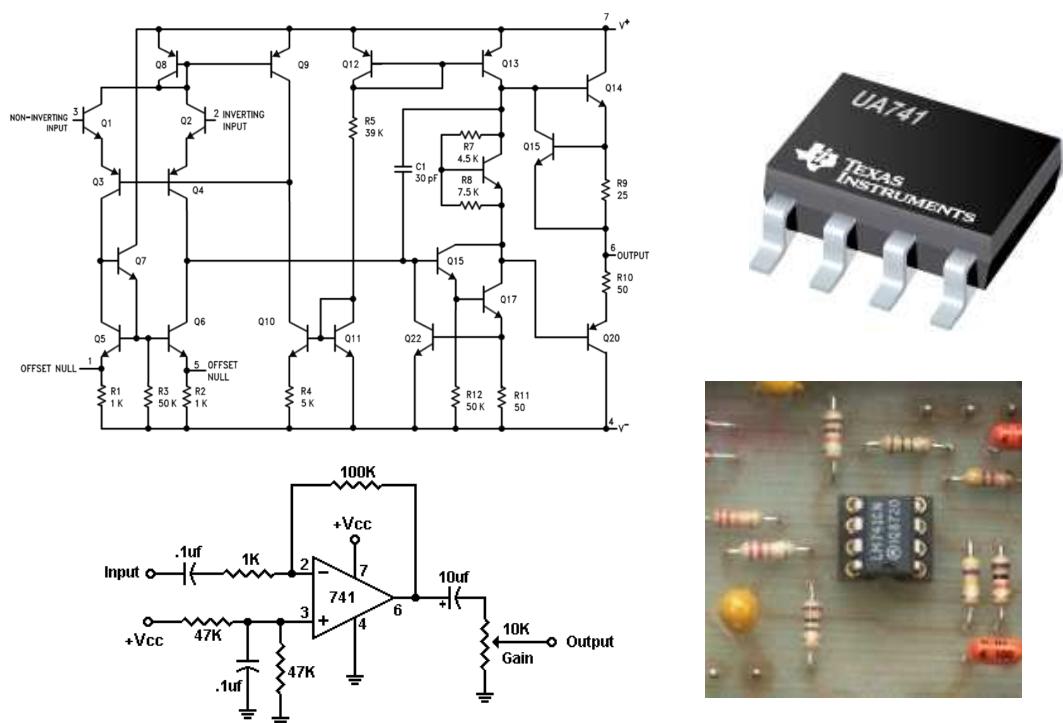
# Chapter 7

## Introduction to Feedback

Boris Murmann  
Stanford University  
Winter 2015-16

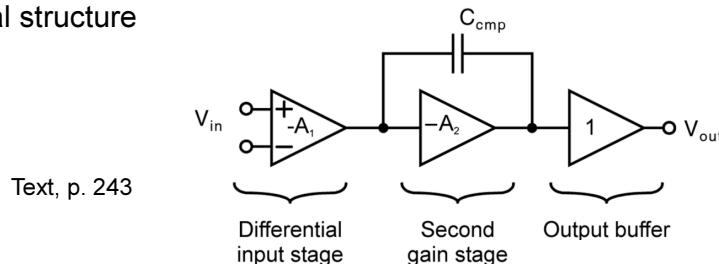
Textbook Sections: 5.1, 5.2, 5.3, 5.4

### Discrete Feedback Circuits Using General Purpose OpAmps

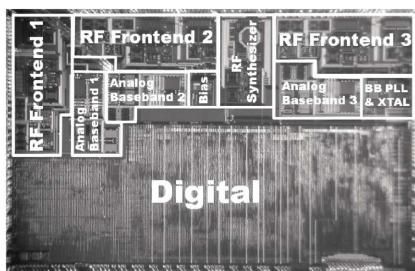


## Properties of General Purpose OpAmps

- Low integration density, up to a few OpAmps per package
  - Typically no more than a few hundred transistors
- Often fabricated in old or exotic fabrication processes
  - Purely bipolar process, JFETs, etc.
- Common denominator design (rather than optimized for one specific application)
  - Input stage often designed for wide common mode range
  - Output stage often designed for heavy loads
  - Internally compensated for unity gain stability
- Typical structure

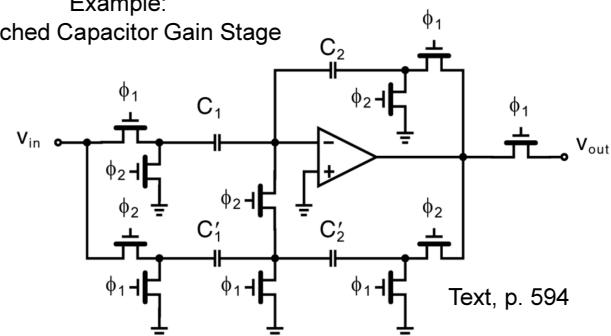


## “Mainstream” Integrated Feedback Amplifiers in CMOS SoCs



S. Abdollahi-Alibeik et al., “A 65nm dual-Band 3-Stream 802.11n MiMo WLAN SoC,” ISSCC 2011

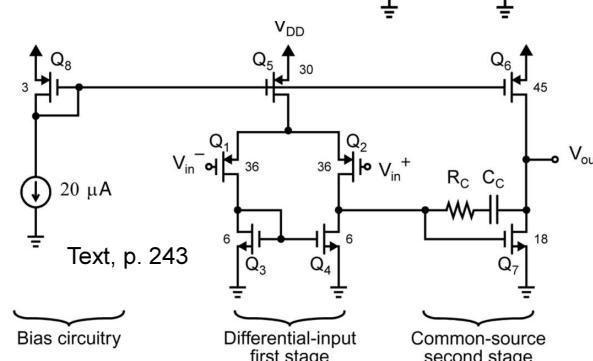
Example:  
Switched Capacitor Gain Stage



Text, p. 594

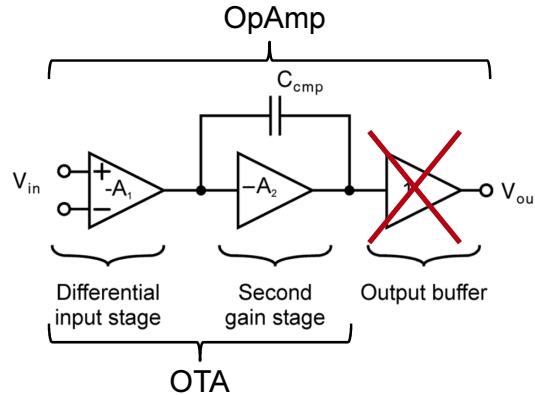
Example:  
Two-Stage Operational Transconductance Amplifier (OTA)

Text, p. 243



## Properties of Integrated Feedback Amplifiers

- Optimized for one specific task
  - Input stage may not need to handle wide common mode input range
  - Output may need to drive only moderate capacitive loads
    - Output buffer usually not needed
  - Stability must be ensured only for the given feedback network
    - The amplifier is not necessarily unity gain stable
  - Noise level is adjusted to meet specific target (no overdesign)

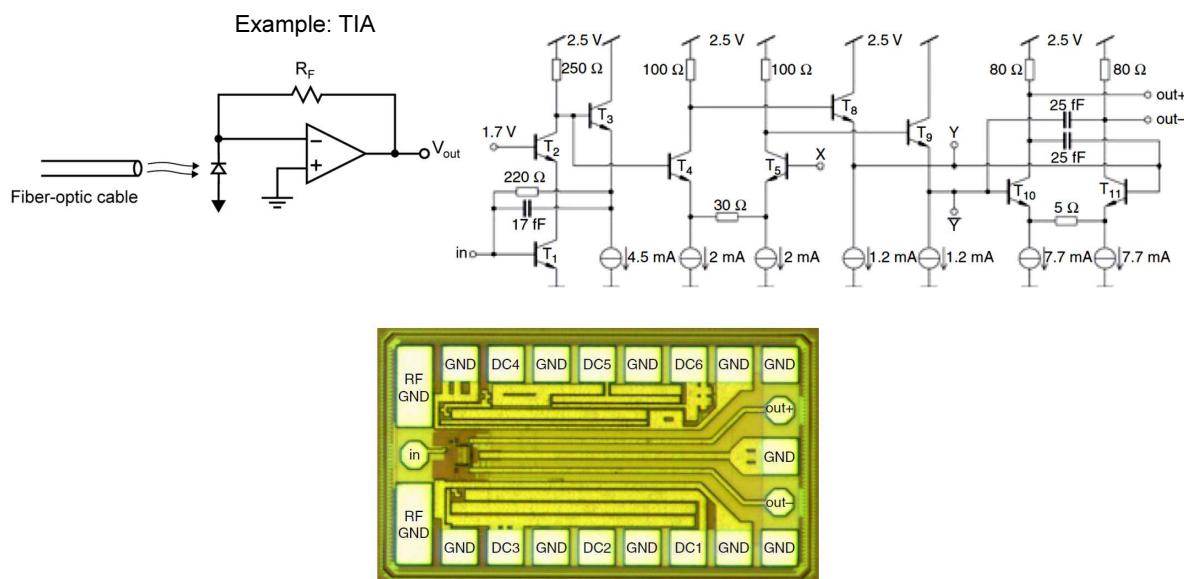


B. Murmann

EE214B Winter 2015-16 – Chapter 7

5

## Special Purpose Feedback Circuits



C. Knochenhauer et al., "40 Gbit/s transimpedance amplifier with high linearity range in 0.13  $\mu$ m SiGe BiCMOS, Electronics Letters, May 12, 2012.

B. Murmann

EE214B Winter 2015-16 – Chapter 7

6

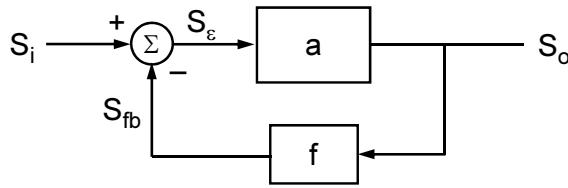
## Special Purpose Feedback Amplifiers

- Wide variety of applications
  - Voltage regulators
  - Transimpedance amplifiers for optical communications
  - Transimpedance amplifiers for instrumentation of physics experiments
  - RF power amplifiers (e.g. Cartesian feedback)
  - Audio power amplifiers
  - ...
- Sometimes integrated within a larger SoC, sometimes not
  - Depending on whether the goals are compatible with fine line CMOS (or BiCMOS, if available)

## Feedback Circuits in EE214B

- We will use the transimpedance amplifier example as a driver to advance/solidify your understanding of feedback circuits
  - Frequency compensation and strategic pole placement
  - Noise analysis in feedback circuits
  - Two-port feedback circuit analysis
- After developing the key tools using the TIA example, we will look into other examples and further generalizations to broaden your understanding

## Review: Idealized Negative Feedback System



### Assumptions for an ideal feedback system

1. No loading effects
2. Unilateral transmission in both the forward amplifier and feedback network

$$\left. \begin{array}{l} S_o = a \cdot S_e \\ S_{fb} = f \cdot S_o \\ S_e = S_i - S_{fb} \end{array} \right\} \Rightarrow S_o = (S_i - S_{fb}) = a(S_i - f \cdot S_o)$$

$$\left. \begin{array}{l} \textbf{Closed-Loop Gain: } A = \frac{S_o}{S_i} = \frac{a}{1+af} \\ \textbf{Loop Gain: } T = af = \frac{S_{fb}}{S_e} \end{array} \right\} \Rightarrow A = \frac{1}{f} \frac{T}{1+T}$$

If  $T \gg 1$ , then

$$A \approx \frac{1}{f}$$

(Also, note that if  $T \ll 1$ , then  $A \approx a$ )

For large loop gain, the feedback acts to minimize the error signal ( $S_e$ ), thus forcing  $S_{fb}$  to track  $S_i$

$$S_e = S_i - f \cdot S_o = S_i - f \cdot \left( \frac{a}{1+af} \right) S_i = \left( 1 - \frac{af}{1+af} \right) \cdot S_i$$

$$\therefore \frac{S_e}{S_i} = 1 - \frac{T}{1+T} = \frac{1}{1+T} \quad \text{and} \quad \frac{S_{fb}}{S_i} = a \cdot f \left( \frac{S_e}{S_i} \right) = \frac{T}{1+T}$$

## Forward Gain Desensitization

The feedback network is typically a precision passive network with an insensitive, well-defined transfer function  $f$ . The forward amplifier gain is generally large, but not well controlled.

Feedback acts to reduce not only the gain, but also the fractional gain error by the factor  $1+T$

$$\begin{aligned}\frac{dA}{da} &= \frac{d}{da} \left( \frac{a}{1+af} \right) = \frac{1}{1+af} + a \frac{d}{da} \left( \frac{1}{1+af} \right) \\ &= \frac{(1+af) - af}{(1+af)^2} = \frac{1}{(1+af)^2} = \frac{1}{(1+T)^2}\end{aligned}$$

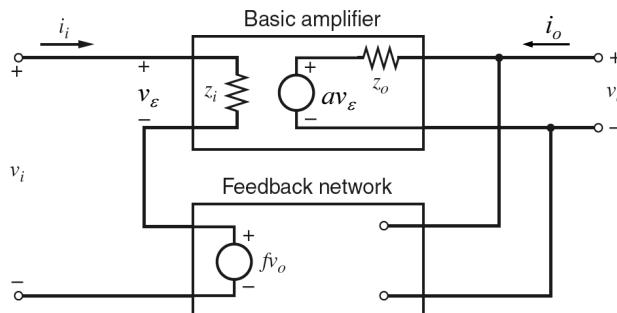
For a change  $\delta a$  in  $a$

$$\begin{aligned}\delta a &= \frac{dA}{da} \delta a = \frac{\delta a}{(1+T)^2} \\ \therefore \frac{\delta A}{A} &= \frac{\delta a}{(1+T)^2} \left( \frac{1+T}{a} \right) = \left( \frac{1}{1+T} \right) \frac{\delta a}{a}\end{aligned}$$

## Closed-Loop Impedances

To illustrate the influence of feedback on the input and output impedances of an amplifier, consider the following voltage-in, voltage-out example

- Include finite input and output impedances in a simple model for the forward amplifier
- Assume that the feedback network has ideal input and output impedances so as not to load the forward amplifier



### Input Impedance

$$Z_i = \left. \frac{V_i}{I_i} \right|_{I_o=0}$$

With  $I_o = 0$

$$\begin{aligned} V_o &= aV_\varepsilon \\ V_i &= V_\varepsilon + fV_o = (1+af)V_\varepsilon \\ &= (1+T)V_\varepsilon \\ I_i &= \frac{V_\varepsilon}{Z_i} = \frac{1}{Z_i} \left( \frac{1}{1+T} \right) V_i \end{aligned}$$

$$Z_i = \frac{V_i}{I_i} = Z_i (1+T)$$

### Output Impedance

$$Z_o = \left. \frac{V_o}{I_o} \right|_{V_i=0}$$

With  $V_i = 0$

$$\begin{aligned} V_\varepsilon + fV_o &= V_i = 0 \\ I_o &= \frac{V_o - aV_\varepsilon}{Z_o} = \frac{1}{Z_o} (1+af)V_o \\ &= \frac{1}{Z_o} (1+T)V_o \end{aligned}$$

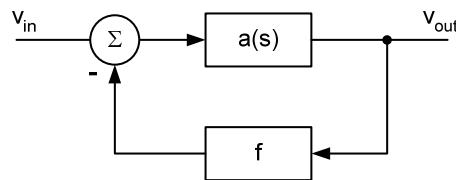
$$Z_o = \frac{V_o}{I_o} = \frac{Z_o}{1+T}$$

### Blackman's Impedance Formula

$$Z_{\text{port}} = Z_{\text{port}} (\text{gain set to zero}) \cdot \frac{1+T(\text{port shorted})}{1+T(\text{port open})}$$

- A very convenient tool for finding the port impedances of feedback circuits by inspection
- It is common that one of the two loop gains (shorted or open) is zero
  - Thus, depending on the configuration, feedback tends to decrease or increase the port impedance by the loop gain

## Negative Feedback and Bandwidth

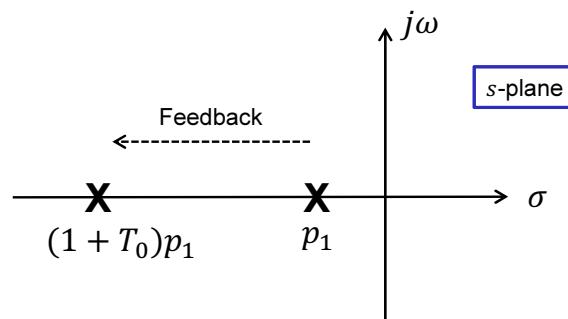


Example:  $a(s) = \frac{a_0}{1 - \frac{s}{p_1}}$

Closed-loop transfer function

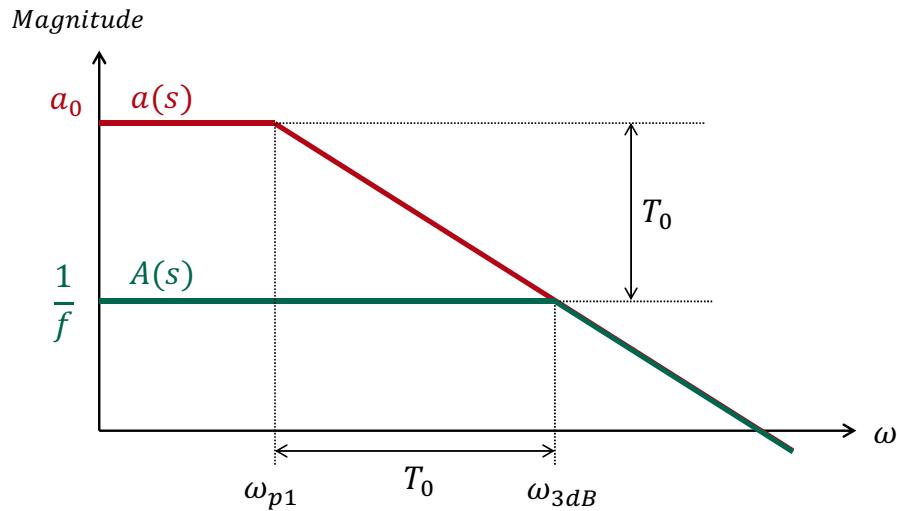
$$A(s) = \frac{a(s)}{1 + a(s)f} = \frac{a_0}{1 + a_0 f} \cdot \frac{1}{1 - \frac{s}{p_1(1 + a_0 f)}} = \underbrace{\frac{1}{f(1 + T_0)} \cdot \frac{1}{1 - \frac{s}{p_1(1 + T_0)}}}_{A_0}$$

- Bandwidth increases by  $(1 + T_0)$ !
  - But gain is reduced by the same factor
- Product of gain and bandwidth remains constant

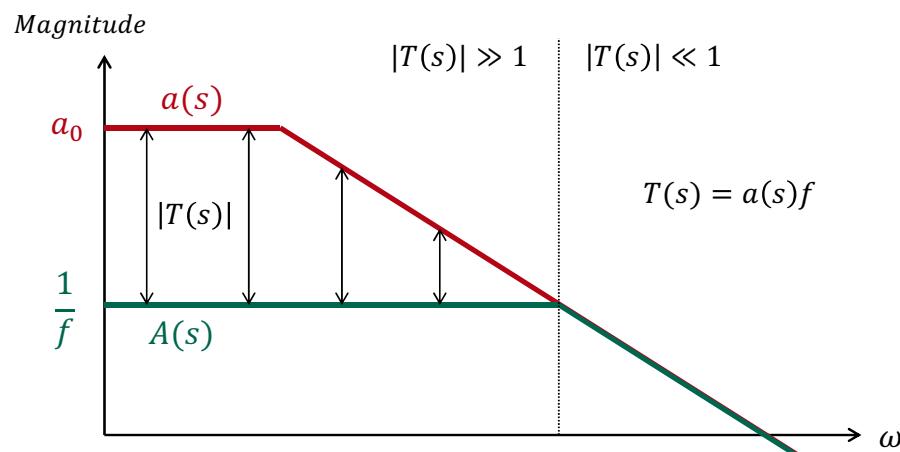


- Applying feedback has increased the bandwidth by  $(1 + T_0)$
- But, we have sacrificed gain in the process, since  $A_0 = a_0 / (1 + T_0)$
- Essentially, we have traded gain for bandwidth

## Bode Plot (Assuming $T_0 \gg 1$ )

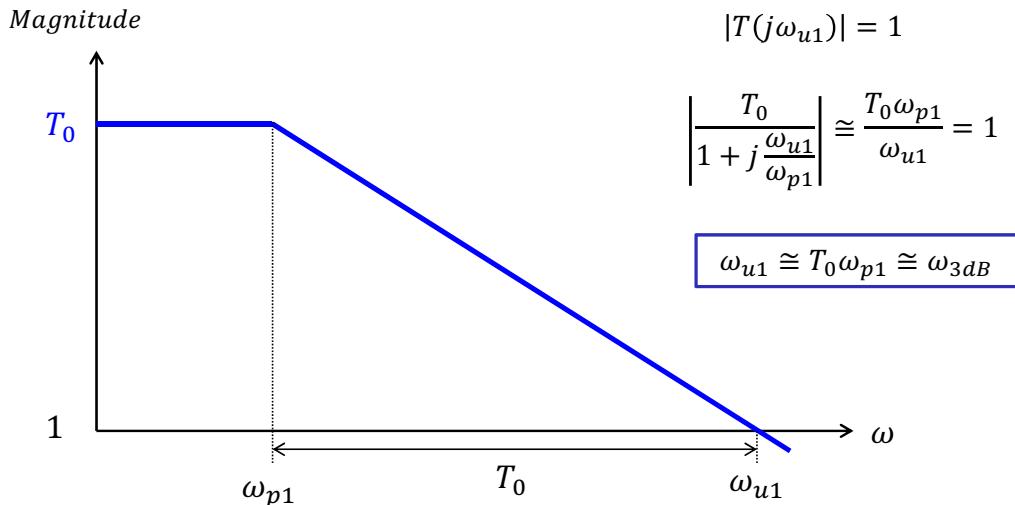


$$A(s) = \frac{1}{f} \frac{T(s)}{1 + T(s)} = \frac{a(s)}{1 + T(s)} \cong \begin{cases} 1/f \text{ for } |T(s)| \gg 1 \\ a(s) \text{ for } |T(s)| \ll 1 \end{cases}$$



- The loop forces  $A$  to be close to  $1/f$  until it runs out of loop gain
  - This is where the closed loop pole appears

## Bode Plot of Loop Gain



- The frequency  $\omega_{u1}$  is called the unity gain frequency of the loop or loop gain-bandwidth product; it is approximately equal to the closed-loop bandwidth

## Instability

At a frequency where the phase shift around the loop of a feedback amplifier reaches  $\pm 180^\circ$  the feedback becomes positive. In that case, if the loop gain is greater than unity, the circuit is unstable.

For a “single-pole” forward path amplifier stability is assured because the maximum phase shift is  $90^\circ$ . However, if  $a(s)$ , or in general  $T(s)$ , has multiple poles, the amount of loop gain that can be used is constrained.

The stability of a feedback amplifier can be assessed from:

- Nyquist diagram (polar plot of loop gain with frequency as a parameter)
- Bode plot (plot of loop gain and phase as functions of frequency)
- Locus of poles (root locus) of  $A(s)$  in the s-plane

## Bode Stability Criterion

A feedback system is unstable when  $|T(j\omega)| > 1$  at the frequency where  $\text{Phase}[T(j\omega)] = -180^\circ$

Phase margin

- Defined at the frequency where  $|T(j\omega)| = 1 \rightarrow \omega_0$

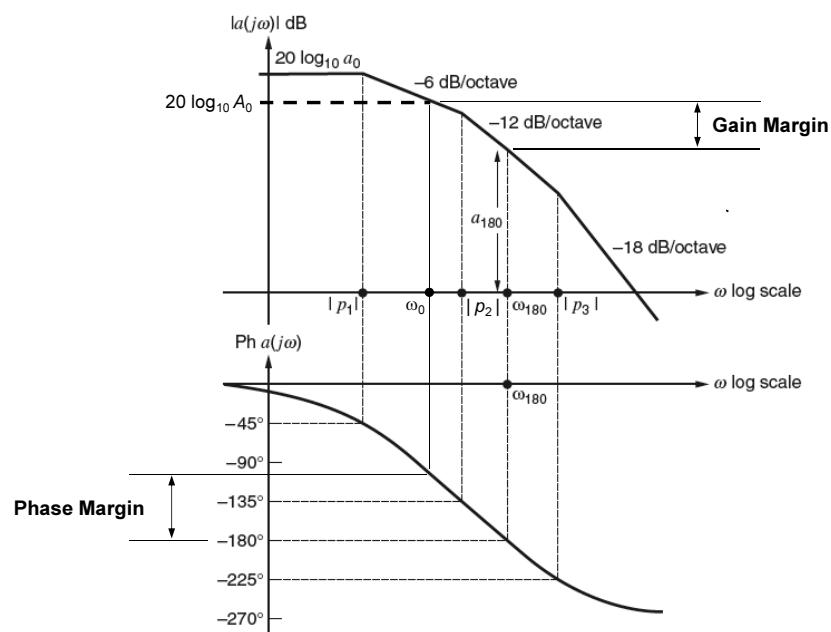
$$\boxed{\text{PM} = 180^\circ + \text{Phase}[T(j\omega_0)]}$$

Gain margin

- Defined at the frequency where  $\text{Phase}[T(j\omega)] = -180^\circ \rightarrow \omega_{180}$

$$\boxed{\text{GM} = \frac{1}{|T(j\omega_{180})|}}$$

## Example



## Practical Phase Margins

- Practical circuits typically use phase margins greater  $45^\circ$ 
  - For continuous time amplifiers, a common target is  $\sim 60^\circ$ 
    - More later
  - For switched capacitor circuits, a phase margin of  $\sim 75^\circ$  is desirable
    - More later
- In order to see the need for phase margin  $>45^\circ$ , investigate the closed-loop behavior at  $\omega = \omega_u$

$$|T(j\omega_u)| = |a(j\omega_u) \cdot f| = 1 \quad \Rightarrow \quad |a(j\omega_u)| = \frac{1}{f} \quad (\text{assuming } f \text{ is real})$$

$$\begin{aligned} A(j\omega_u) &= \frac{a(j\omega_u)}{1 + a(j\omega_u)/|a(j\omega_u)|} \\ &= \frac{a(j\omega_u)}{1 + e^{j\phi[a(j\omega_u)]}} = \frac{a(j\omega_u)}{1 + e^{j(\text{PM}-180^\circ)}} \end{aligned}$$

$$\underline{\text{PM} = 45^\circ} \quad A(j\omega_u) = \frac{a(j\omega_u)}{1 + e^{-j135^\circ}} = \frac{a(j\omega_u)}{1 - 0.7 - 0.7j}$$

$$\therefore |A(j\omega_u)| = \frac{|a(j\omega_u)|}{0.76} = \frac{1.3}{f} \approx 1.3A_0 \quad \text{"Peaking"}$$

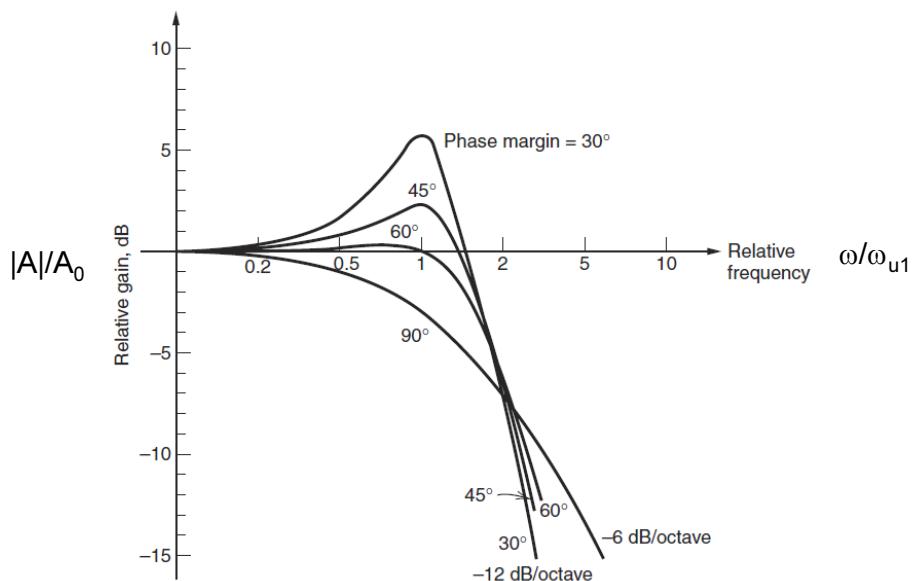
$$\underline{\text{PM} = 60^\circ} \quad A(j\omega_u) = \frac{a(j\omega_u)}{1 + e^{-j120^\circ}} = \frac{a(j\omega_u)}{1 - 0.5 - 0.87j}$$

$$\therefore |A(j\omega_u)| = |a(j\omega_u)| = \frac{1}{f} \approx A_0 \quad \text{"Flat"}$$

$$\underline{\text{PM} = 90^\circ} \quad A(j\omega_u) = \frac{a(j\omega_u)}{1 + e^{-j90^\circ}} = \frac{a(j\omega_u)}{1 - j}$$

$$\therefore |A(j\omega_u)| = \frac{|a(j\omega_u)|}{1.4} = \frac{0.7}{f} \approx 0.7A_0 \quad \text{"Drooping"}$$

## Typical Closed-Loop Response for Various Phase Margins

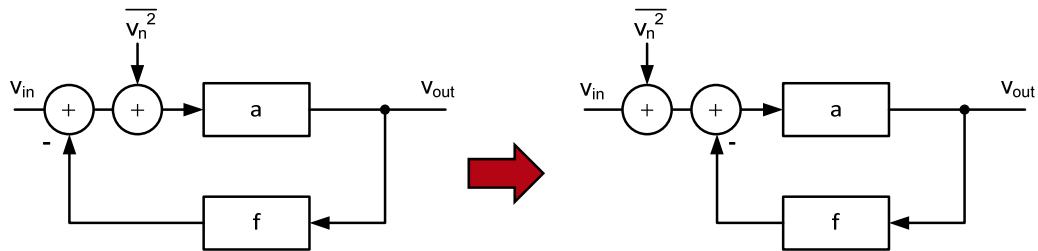


(These curves are based on a two-pole loop response)

## Frequency Compensation

- Frequency compensation refers to the means by which the frequency response of the loop gain is altered to ensure stability and adequate phase margin
- Frequency compensation can be categorized into three groups
  - Internal compensation = alter the frequency response of  $a(s)$
  - External compensation = alter the frequency response of  $f(s)$
  - Or alter both!
- Consequently, there exist many different options for frequency compensation, and it is often not immediately obvious which one is ideal for the design problem at hand
  - We will explore various options “just in time” as we work our way through practically relevant feedback amplifier circuits
  - Then generalize in Chapter 9

## Feedback and Noise



- To first order, feedback has no impact on noise performance
  - The amplifier's input referred noise can be simply pushed outside the loop
- To second order, we will see that feedback will in most cases slightly increase the noise
  - For example, there may be additional noise contributed by resistors in the feedback network
- Again, we will explore these subtleties using concrete examples

## Summary – Effect of Negative Feedback

Bandwidth	$\uparrow (1+T)$ Approximately equal to unity gain frequency of loop gain
Gain	$\downarrow (1+T)$ Defined primarily by the feedback network $\sim 1/f$
Port Impedances	$\uparrow$ or $\downarrow (1+T)$ (typically)
Noise	Moderate increase

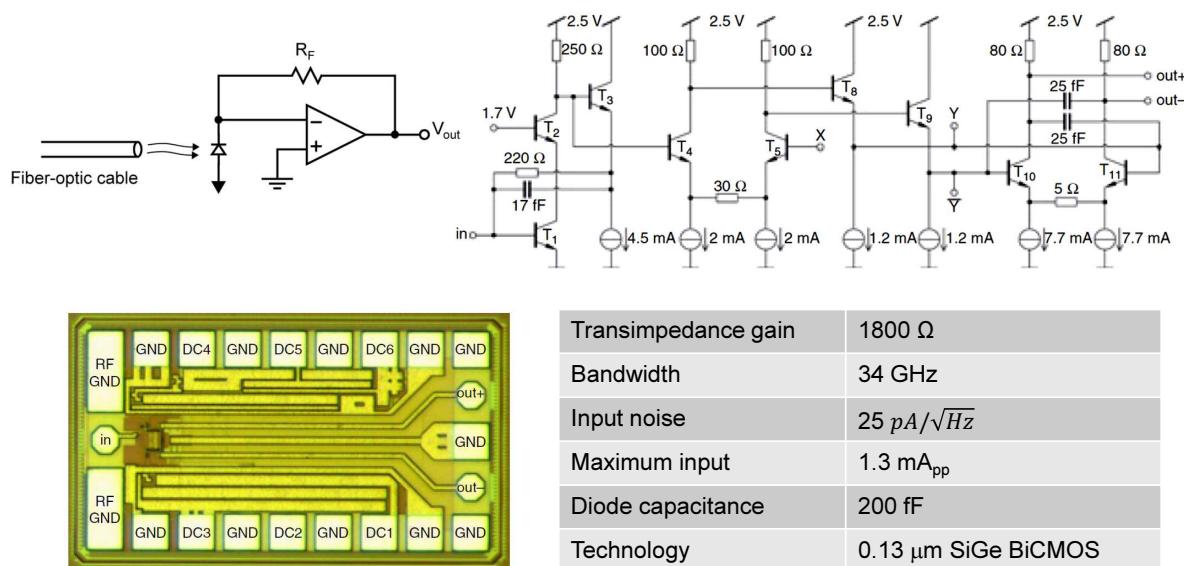
# Chapter 8

## Feedback TIA Design

Boris Murmann  
Stanford University  
Winter 2015-16

Textbook Sections: 5.4, 9.4.4

### Motivating Example: TIA for High-Speed Optical Networks



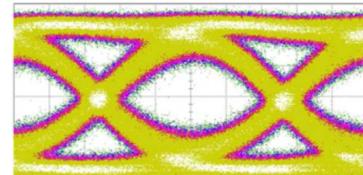
C. Knochenhauer et al., "40 Gbit/s transimpedance amplifier with high linearity range in 0.13  $\mu$ m SiGe BiCMOS," Electronics Letters, May 12, 2012.

## Overview

- We will use the high-speed TIA example as a driver to establish a number of useful design and analysis techniques
  - “Project-based learning”
- The idea is to start from the beginning and understand all (or most) steps that a designer would go through to arrive at the circuit shown on the previous slide

## Where does the Noise Specification come from?

- Let's consider a very simple example
- Assuming two-level signaling (“NRZ”), we need  $\text{SNR} > 16.9\text{dB}$  for a bit error rate (BER) of  $10^{-12}$ 
  - This follows from the “Q-function” (see any course on communication systems)
- In the given design, the rms noise is approximately



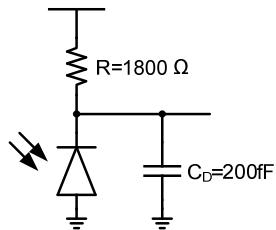
$$\sqrt{N} = \sqrt{\frac{\pi}{2} \cdot BW \cdot PSD} = \sqrt{\frac{\pi}{2} \cdot 34\text{GHz}} \cdot \frac{25\text{pA}}{\sqrt{\text{Hz}}} = 5.8\mu\text{A}_{rms}$$

- The smallest signal we can receive while maintaining the given BER is

$$\sqrt{S} = \sqrt{N} \cdot 10^{\frac{16.9}{20}} = 40.6\mu\text{A}_{rms}$$

- This number then defines the minimum optical input power, maximum length of the fiber, etc., typically in compliance with a certain standard

## Why use a TIA and not just a Resistor?

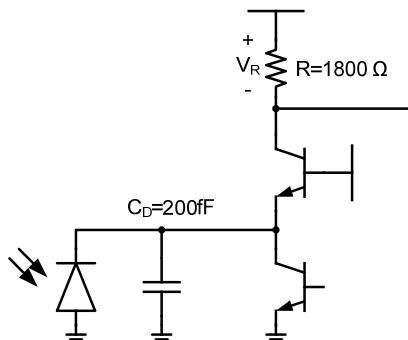


$$\frac{\bar{i^2}}{\Delta f} = 4kT \frac{1}{R} = \left( \frac{3pA}{\sqrt{Hz}} \right)^2$$

$$f_{3dB} = \frac{1}{2\pi R C_D} = 442MHz$$

- Assuming that transimpedance gain is fixed, both noise and  $f_{3dB}$  are fixed and beyond our control
  - The noise happens to be much better than needed
  - The bandwidth happens to be much worse than needed
- This circuit lacks degrees of freedom that let us adjust gain, bandwidth and noise independently (within some reasonable range)

## How about using a CB (or CG) Configuration?



$$f_{3dB} = \frac{g_m}{2\pi(C_D + C_\pi + C)} \approx \frac{g_m}{2\pi C_D}$$

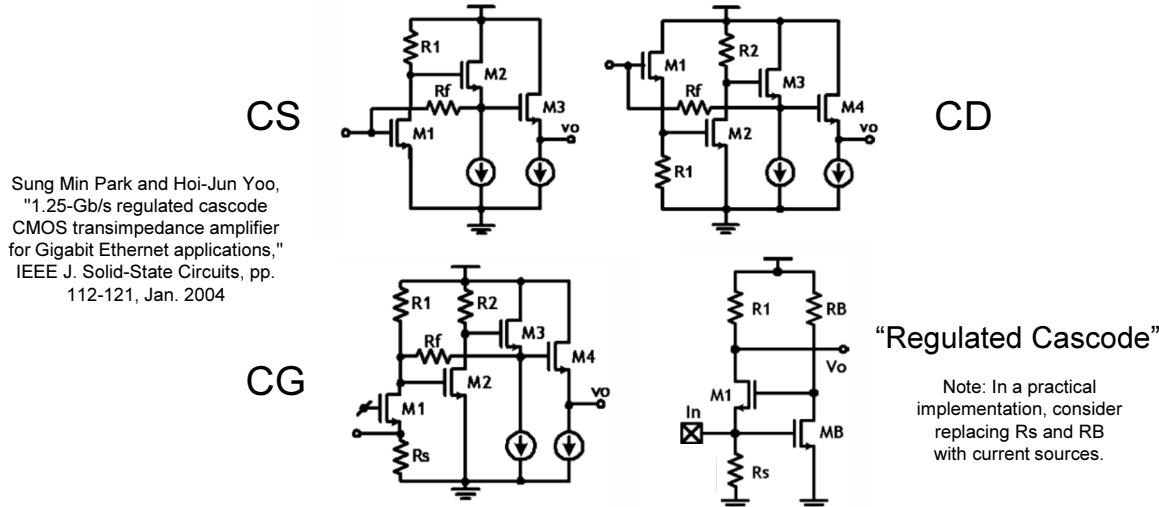
To get the bandwidth we want, chose  $g_m \approx 43mS$  ( $1/g_m = 23\Omega$ ),  $I_C = 1.1mA$

$$\frac{\bar{i^2}}{\Delta f} = 4kT \frac{1}{R} + 2qI_C = \left( \frac{19pA}{\sqrt{Hz}} \right)^2$$

- Good news
  - We are now meeting the gain, bandwidth and noise specification
- Bad news
  - The circuit has very small signal headroom:  $V_R = 1.1mA \cdot 1800\Omega \approx 2V$
  - The circuit cannot handle the maximum signal of  $1.3mA_{pp}$
  - Any capacitance we connect to the output will kill the bandwidth
  - Whatever we can do to try and fix these issues will drive us outside the noise specification...

## Commonly used Feedback TIAs

- Feedback TIAs incorporate additional degrees of freedom that mitigate these issues
- Over the years, the architectures shown below have evolved as commonly used solutions (in both MOS and Bipolar technologies)



B. Murmann

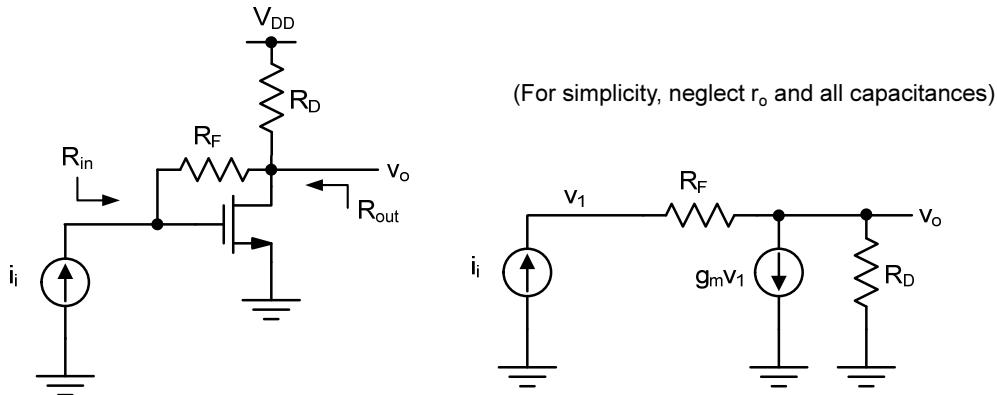
EE214B Winter 2015-16 – Chapter 8

7

## Optimality

- None of these (or other) options are known to be “globally” optimum
- Which one of these circuits performs best depends on the given specifications and the target technology
  - Welcome to analog design!
- In the following discussion, we will explore the CS architecture in more detail and develop know-how that is generally useful for feedback amplifier design

## Starting Point: 1-Transistor CS TIA



- Attractive property
  - Both  $R_{in}$  and  $R_{out}$  are low, e.g.  $\sim 1/g_m$ , for  $R_D \gg R_F$  (prove this)

## Analysis Methods

- If all we wanted is an expression for the closed-loop gain, we can abandon the “af” feedback model and simply analyze the circuit from first principles (KCL, KVL)
  - However, this yields no information about loop gain (which we need for stability analysis)
- There are two ways to map the circuit into a feedback block diagram
- Return ratio analysis
  - “Asymptotic” method, does not attempt to break the circuit into pieces
  - Find loop gain by injecting a test signal at the controlled source
  - Find ideal closed-loop gain by setting the controlled source to infinity
  - Proposed by Hendrik Bode
- Two-port analysis
  - “Massage” the circuit until we can squeeze it into the standard “af” block diagram
  - Proposed by Harold Black

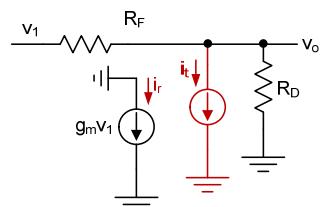
## Nodal Analysis Result for the Closed-Loop Gain

$$\left. \begin{aligned} 0 &= \frac{v_1 - v_o}{R_F} - i_i \\ 0 &= \frac{v_o - v_1}{R_F} + g_m v_1 + \frac{v_o}{R_D} \end{aligned} \right\} \quad A = \frac{v_o}{i_i} = -R_F \frac{1 - \frac{1}{g_m R_F}}{1 + \frac{1}{g_m R_D}}$$

- For large  $g_m R_D$  and  $g_m R_F$ , the transresistance approaches  $-R_F$
- Whichever analysis method we use (return ratio or two-port), the obtained closed loop gain expression should match this result
  - The above expression is therefore a useful reference point

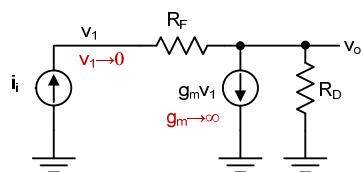
## Return Ratio Analysis

- Step 1: Find the loop gain



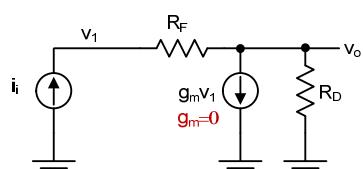
$$T = -\frac{i_r}{i_t} = -\frac{g_m v_1}{i_t} = -\frac{g_m (-i_t R_D)}{i_t} = g_m R_D$$

- Step 2: Find the ideal closed-loop gain



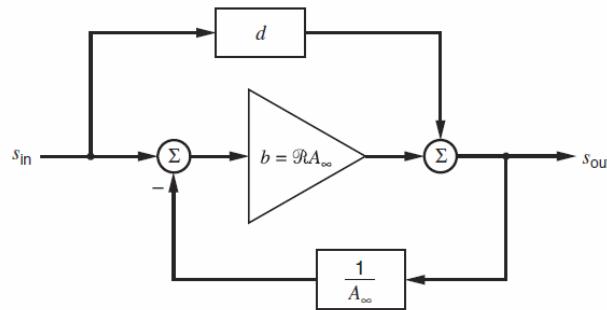
$$v_o = -i_i R_F \quad A_\infty = \left. \frac{v_o}{i_i} \right|_{g_m \rightarrow \infty} = -R_F$$

- Step 3: Find the feedforward term



$$v_o = i_i r_o \quad d = \left. \frac{v_o}{i_i} \right|_{g_m=0} = R_D$$

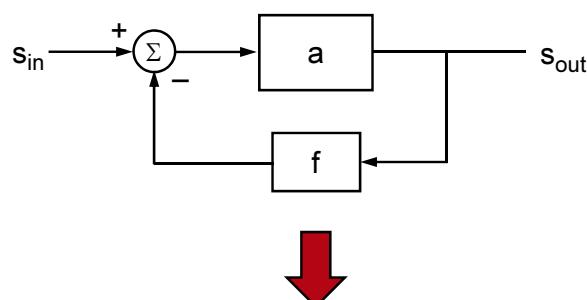
## Computing the Closed-Loop Gain using Return Ratio Parameters



$$A = A_\infty \frac{T}{1+T} + \frac{d}{1+T} = -R_F \frac{g_m R_D}{1+g_m R_D} + \frac{R_D}{1+g_m R_D} = -R_F \frac{1 - \frac{1}{g_m R_F}}{1 + \frac{1}{g_m R_D}}$$

- The result for the closed loop gain (A) matches the nodal analysis expression perfectly

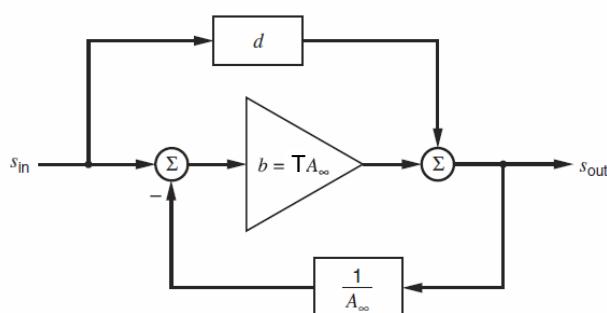
## Similarity



- If  $d=0$  (no feedforward), the return ratio model is identical to the classical “af” model

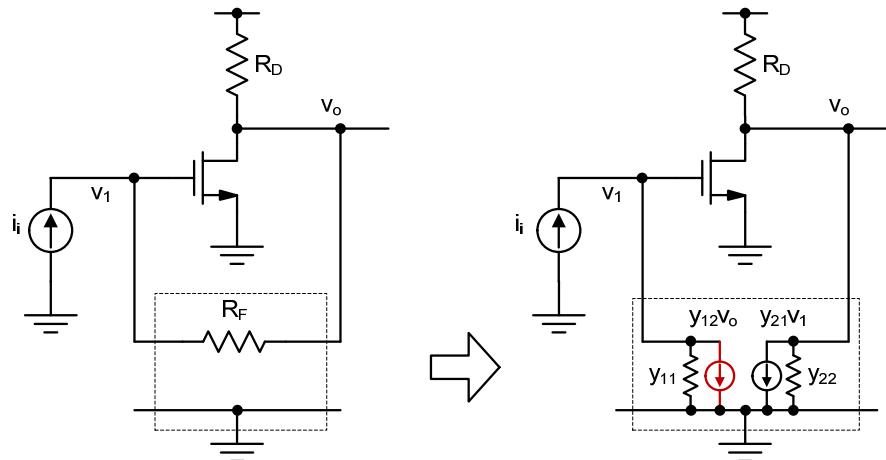
$$f \rightarrow \frac{1}{A_\infty}$$

$$a \rightarrow T A_\infty = \frac{T}{f}$$



- The return ratio model can therefore be viewed as an extension that can capture feedforward
  - There exist circuits where feedforward is relevant (see e.g. EE315A), but in the majority of cases it is not

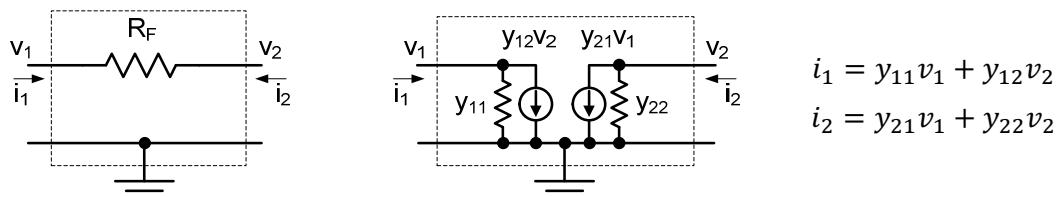
## Two-Port Modeling Approach (1)



- The two-port modeling approach insists on the “af” two-box model and looks to identify an idealized “f” block such that
  - The feedback block does not load (or alter) the forward gain
  - The feedback signal sums perfectly with the input (as in the “af” model)
- Key to achieving the latter is the controlled source  $y_{12}$ , which subtracts from the input; the other elements in the block are needed for equivalence

## Two-Port Modeling Approach (2)

- We want the two networks below to be equivalent

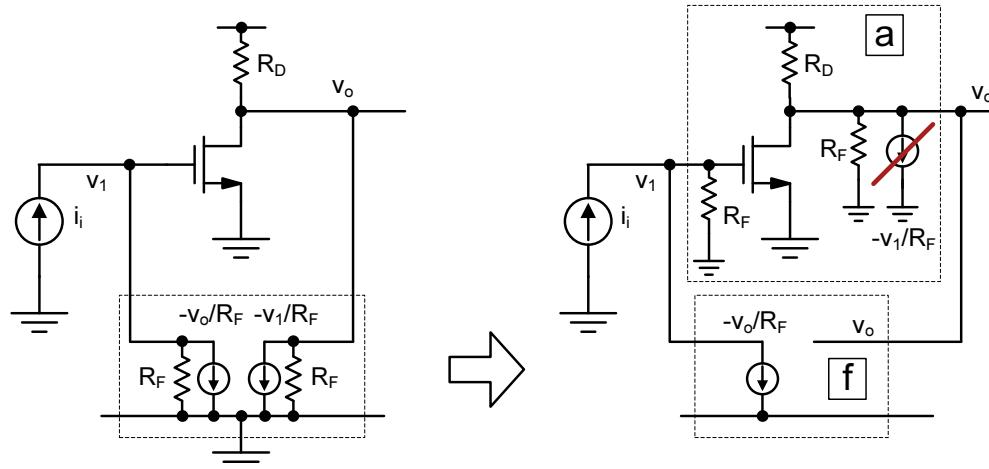


- This requires

$$y_{11} = \left. \frac{i_1}{v_1} \right|_{v_2=0} = \frac{1}{R_F} \quad y_{12} = \left. \frac{i_1}{v_2} \right|_{v_1=0} = -\frac{1}{R_F}$$

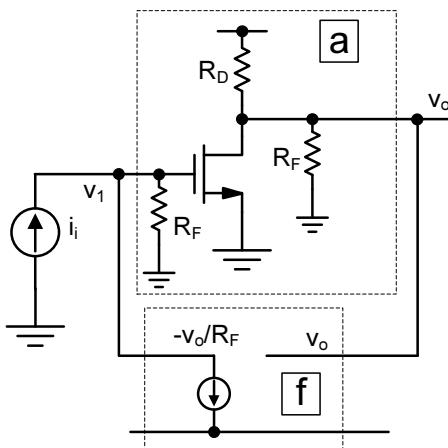
$$y_{21} = \left. \frac{i_2}{v_1} \right|_{v_2=0} = -\frac{1}{R_F} \quad y_{22} = \left. \frac{i_2}{v_2} \right|_{v_1=0} = \frac{1}{R_F}$$

## Two-Port Modeling Approach (3)



- Final steps
  - Absorb all elements other than the ideal feedback source into the “a” block
  - Neglect feedforward

## Final Version of the Two-Port Model



- Assuming  $R_F \gg R_D$

$$a = -R_F g_m \frac{R_F R_D}{R_F + R_D} \cong -g_m R_D R_F$$

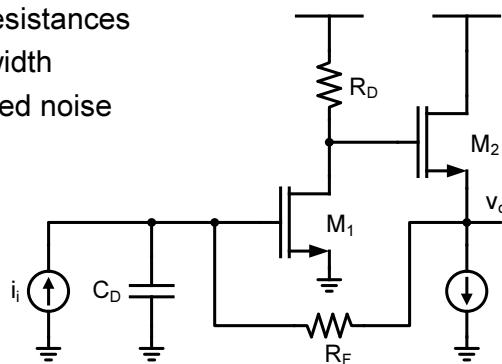
$$f = -1/R_F$$

$$T = af \cong g_m R_D$$

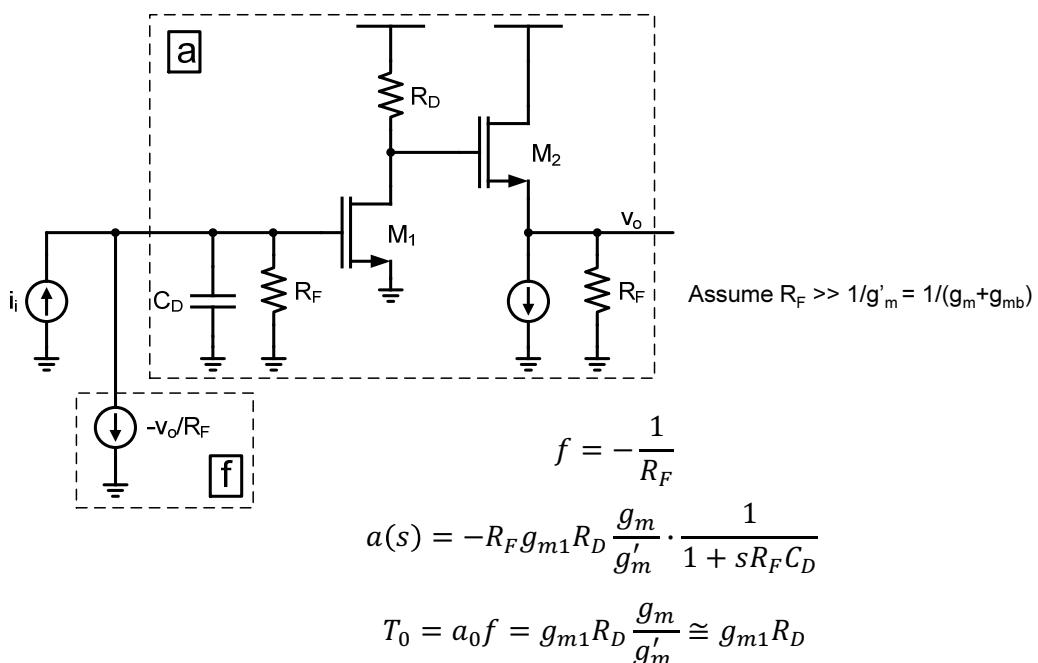
- The loop gain is consistent with the prediction from the return ratio analysis
  - But feedforward is not modeled

## Continuing our Design

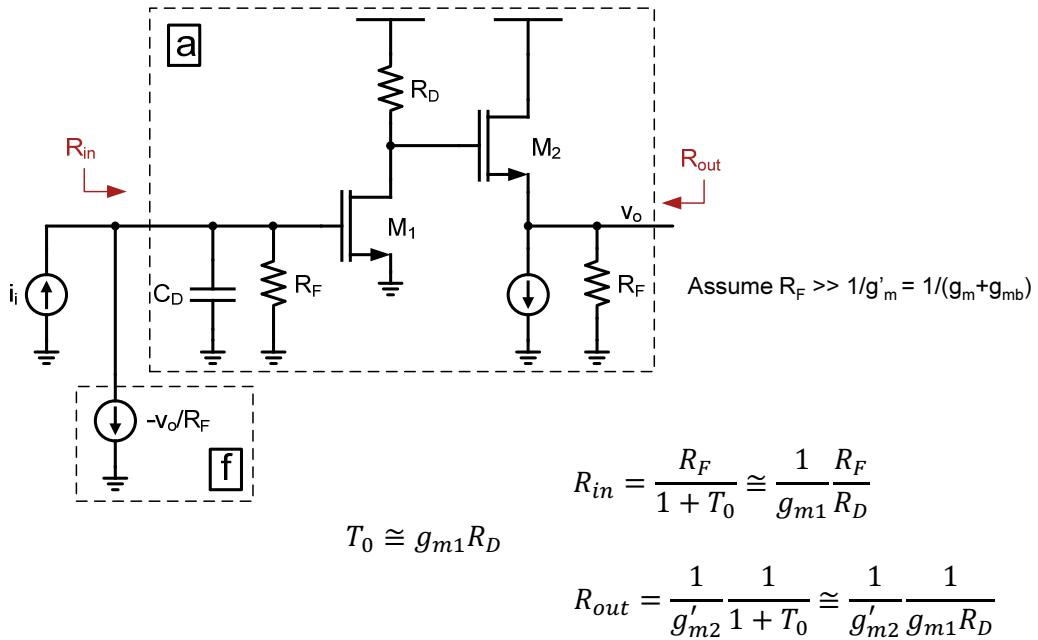
- Consider now a slightly more sophisticated CS TIA with a source-follower output stage
  - For now, ignore all capacitances, except  $C_D$
- Use this circuit as an example to walk though a complete set of calculations
  - Find “ $a$ ” and “ $f$ ”
  - Compute input and output resistances
  - Compute closed-loop bandwidth
  - Compute input source referred noise



## Two-Port Model



## Input and Output Resistance



## Closed-Loop Bandwidth

$$a(s) = -R_F g_{m1} R_D \frac{g_m}{g'_m} \cdot \frac{1}{1 + s R_F C_D}$$

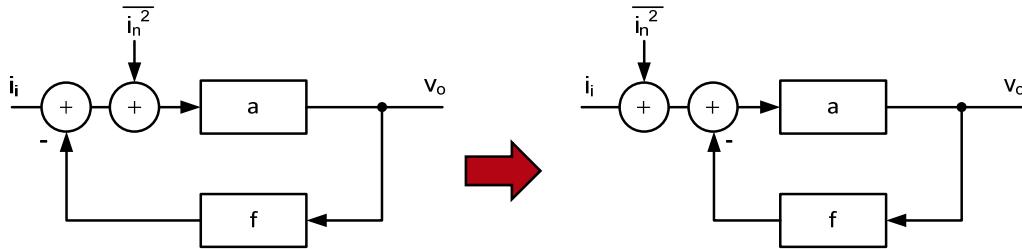
$$T_0 = a_0 f = g_{m1} R_D \frac{g_m}{g'_m} \cong g_{m1} R_D$$

$$\omega_{3dB,CL} = (1 + T_0) \omega_{3dB,OL}$$

$$\omega_{3dB,CL} \cong g_{m1} R_D \frac{1}{R_F C_D} = \frac{1}{R_{in} C_D}$$

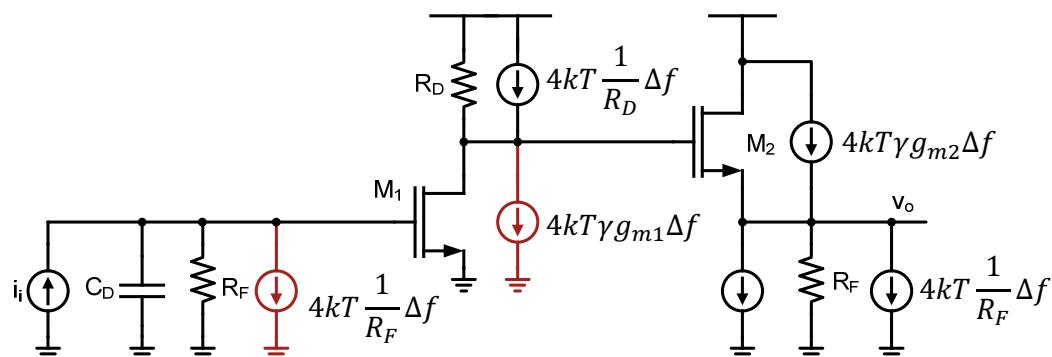
- The closed-loop bandwidth increases (relative to the bandwidth of  $a(s)$ ) consistent with the reduction in  $R_{in}$

## Noise Analysis



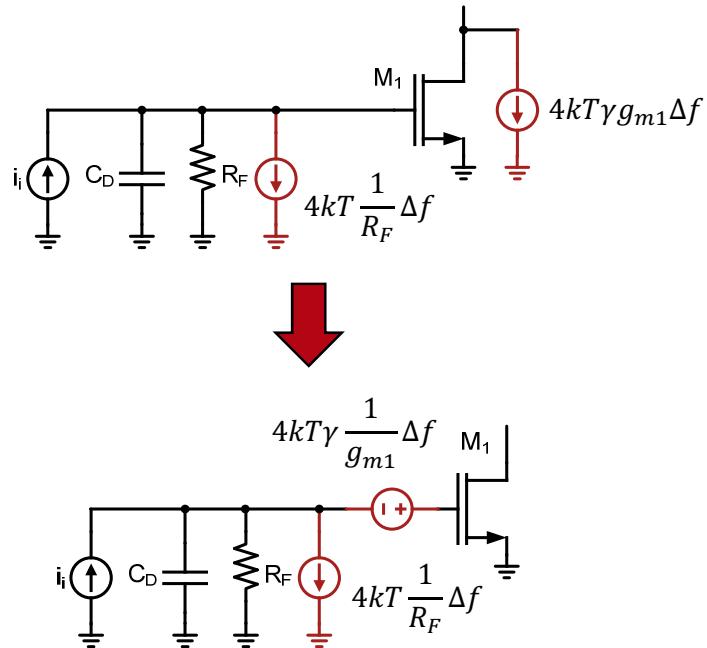
- All we need to compute is the input referred current noise of “a”, including the effect of  $C_D$ ,  $R_F$ , etc.
- The input referred current noise of the amplifier directly refers back to the input current source
- Note that “f” is noiseless in the employed two-port model
  - The noise from  $R_F$  is captured in “a”

## Noise Analysis of “a”

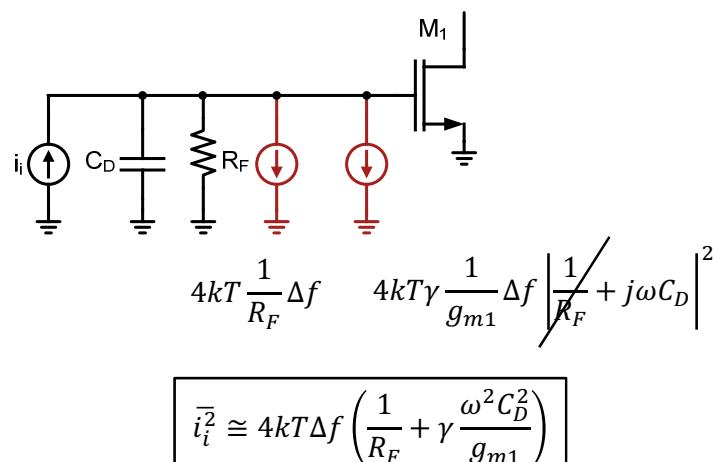


- All noise sources, except the noise of  $M_1$  and  $R_F$  at the input are negligible in practice
- We already saw in our CE noise example that the  $R_D$  noise is negligible as long as  $g_m R_D$  is large
  - All other noise sources are irrelevant by the same argument

## Noise Analysis of “a”

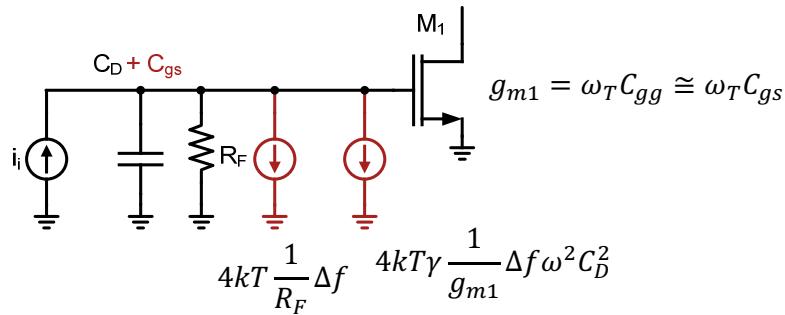


## Noise Analysis of “a”



- At low frequencies, the noise is usually dominated by  $R_F$
- At high frequencies, extra noise from  $M_1$  rolls in
  - This noise can be minimized by increasing  $g_{m1} \rightarrow$  larger device and/or more current and capacitance added to the input node

## Inclusion of $C_{gs}$



$$\bar{i_i^2} \cong 4kT\Delta f \left( \frac{1}{R_F} + \gamma \frac{\omega^2 (C_D + C_{gs})^2}{g_{m1}} \right) \cong 4kT\Delta f \left( \frac{1}{R_F} + \gamma \frac{\omega^2 (C_D + C_{gs})^2}{\omega_T C_{gs}} \right)$$

- This expression is minimized for  $C_{gs} = C_D$ , in which case we have

$$\bar{i_i^2} \cong 4kT\Delta f \left( \frac{1}{R_F} + 4\gamma \frac{\omega^2 C_D}{\omega_T} \right)$$

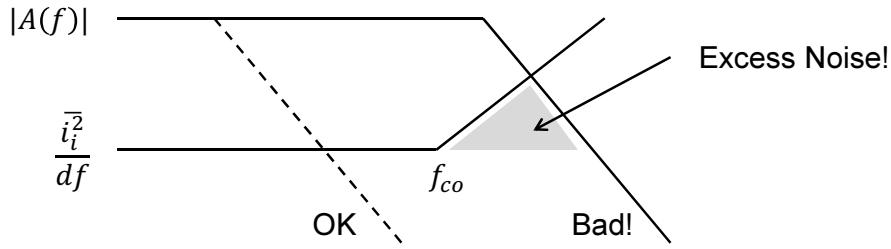
- We can now solve for the frequency where the noise of  $M_1$  takes over

$$\frac{1}{R_F} = 4\gamma \frac{\omega^2 C_D}{\omega_T}$$

$$\omega_{co} = \frac{1}{2} \sqrt{\frac{\omega_T}{\gamma R_F C_D}}$$

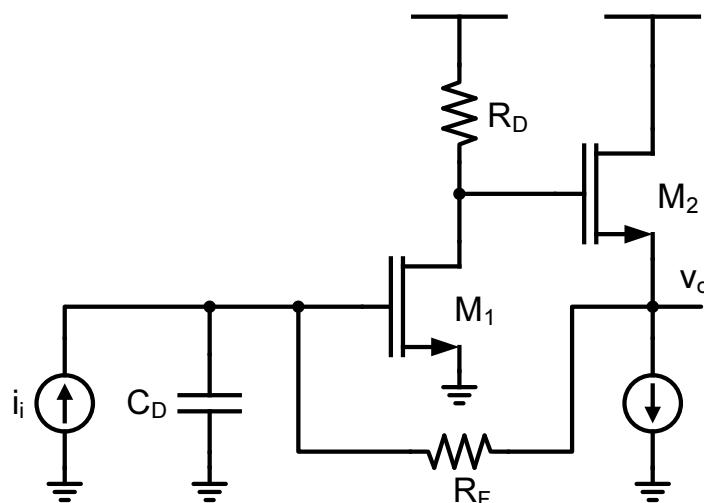
- In a practical design, this frequency must be somewhat above the circuit's bandwidth (or else we will ruin the noise performance)
- For a given  $C_D$  and process  $\omega_T$ , this limits how much transimpedance gain ( $R_F$ ) or bandwidth we can extract from the amplifier
- Example:  
 $f_T = 150 \text{ GHz}$ ,  $C_D = 200 \text{ fF}$ ,  $R_F = 220 \Omega$ ,  $\gamma = 1$   
 $\Rightarrow f_{co} = 46 \text{ GHz}$

## Scenarios

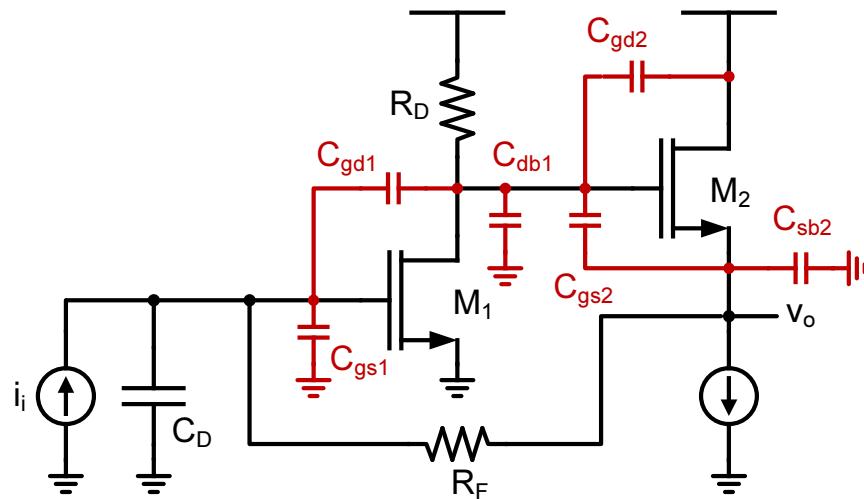


- If the input referred noise peaks noise peaks before the closed-loop transfer function rolls off, the output RMS noise will be completely swamped by the shaded region
  - Recall the 1/f tangent argument

## Continuing our Design

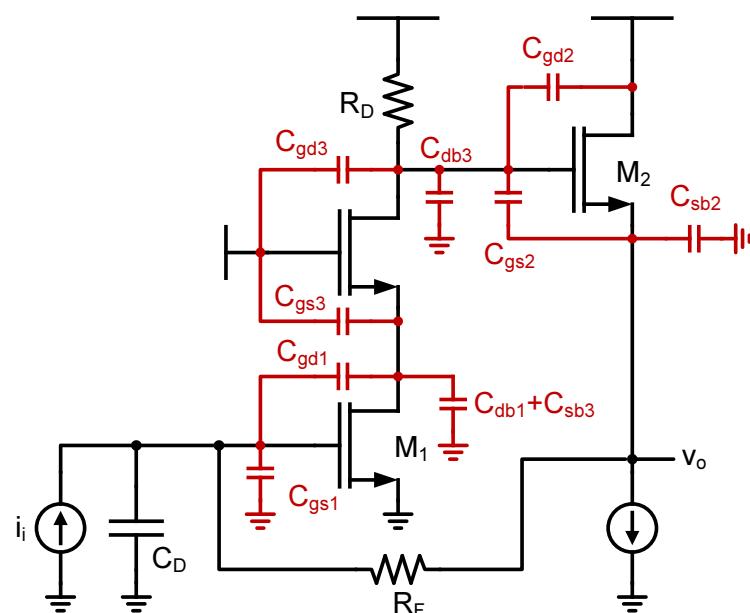


## Relevant Device Capacitances

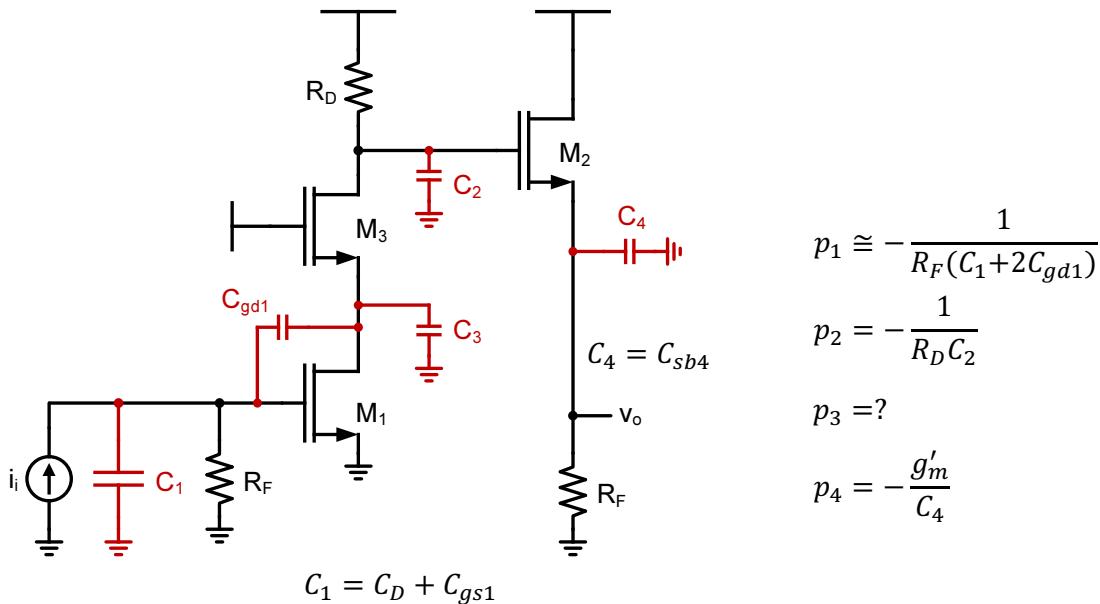


- Problem: Miller multiplication of  $C_{gd1}$

## Circuit with Cascode Input Stage



## Model of Forward Amplifier a(s)



## Pole at Drain of M1

- To compute the pole associated with the drain node, it is best to go back to the CE analysis result from chapter 3, slide 13:

$$p_2 \cong -\left( \frac{1}{R_L C_L} + \frac{g_m}{C_L} \cdot \frac{C_\mu}{C_\pi} + \frac{1}{R_S C_\pi} \right)$$

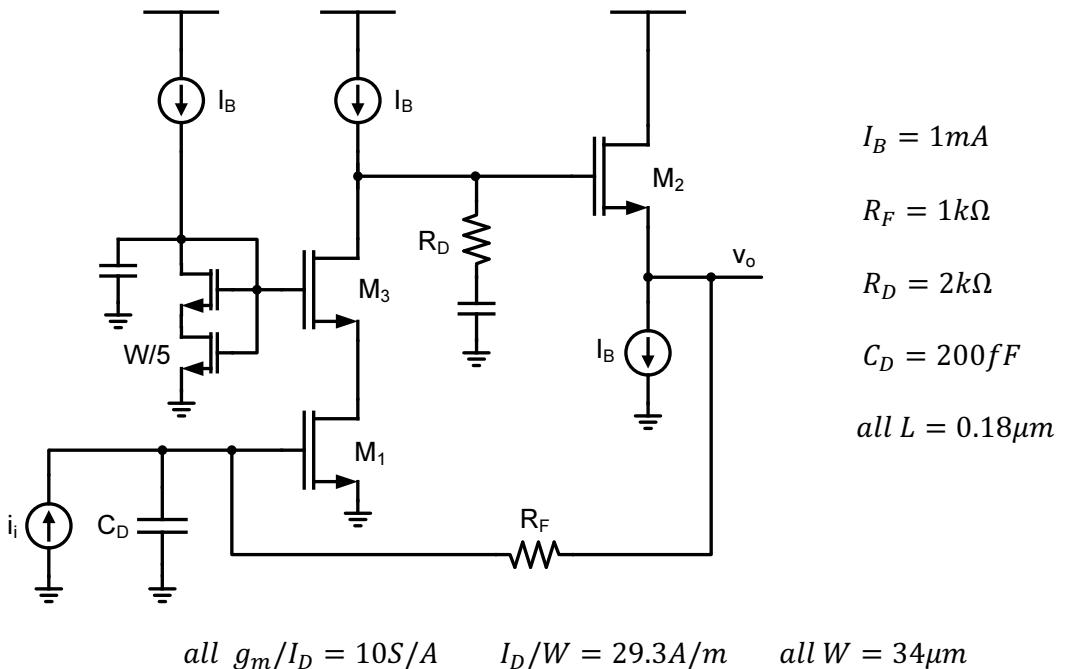
- Substitutions:

$$p_2 \rightarrow p_3 \quad R_L \rightarrow \frac{1}{g'_m} \quad C_L \rightarrow C_3 \quad C_\mu \rightarrow C_{gd1} \quad C_\pi \rightarrow C_1 \quad R_S \rightarrow R_F$$

$$p_3 \cong -\left( \frac{g'_m}{C_3} + \frac{g_m}{C_3} \cdot \frac{C_{gd1}}{C_1} + \frac{1}{R_F C_1} \right)$$

$$p_3 \cong -\frac{g'_m}{C_3}$$

## Prototype Circuit



## Operating Point

element 0:m1	0:m2	0:m3
model 0:nmos214	0:nmos214	0:nmos214
region Saturati	Saturati	Saturati
id 996.5217u	1.0000m	996.5217u
vgs 689.4266m	812.8655m	726.8479m
vds 290.6305m	1.1106	1.2117
vbs 0	-689.42m	-290.63m
vth 491.3958m	649.1707m	557.9840m
vdsat 147.8615m	143.2316m	138.8442m
vod 198.0308m	163.6948m	168.8639m
gm 9.3602m	10.1505m	10.1110m
gds 487.0970u	281.1118u	268.4636u
gmb 2.2263m	2.0191m	2.2033m
cdtot 43.6454f	37.1183f	37.8625f
cgtot 69.8437f	68.8173f	69.2204f
cstot 81.7366f	74.9554f	78.1306f
cbtot 66.9125f	52.9684f	57.4258f
cgs 49.7990f	49.9630f	49.8352f
cgd 16.5154f	16.3953f	16.3916f

## Pole Calculations

$$C_{gs} := 49\text{fF} \quad C_{gd} := 16\text{fF} \quad C_{db} := 20\text{fF} \quad C_{sb} := 25\text{fF} \quad C_D := 200\text{fF}$$

$$g_m := 10\text{mS} \quad g_{mb} := 2.2\text{mS} \quad R_f := 1\text{k}\Omega \quad R_D := 2\text{k}\Omega$$

$$A_{v2} := \frac{g_m}{g_m + g_{mb} + \frac{1}{R_f}} = 0.758 \quad T_0 := g_m \cdot R_D \cdot A_{v2} = 15.152$$

$$C_{1x} := C_D + C_{gs} + 2 \cdot C_{gd} = 281\text{fF}$$

$$f_{p1} := \frac{1}{2 \cdot \pi} \cdot \frac{1}{R_f \cdot C_{1x}} = 566.388\text{MHz}$$

$$C_2 := 2C_{gd} + (1 - A_{v2}) \cdot C_{gs} + C_{db} = 63.879\text{fF}$$

$$f_{p2} := \frac{1}{2 \cdot \pi} \cdot \frac{1}{R_D \cdot C_2} = 1.246\text{GHz}$$

$$C_3 := C_{gs} + C_{sb} + C_{db} = 94\text{fF}$$

$$f_{p3} := \frac{1}{2 \cdot \pi} \cdot \frac{g_m + g_{mb}}{C_3} = 20.656\text{GHz}$$

$$C_4 := C_{sb} = 25\text{fF}$$

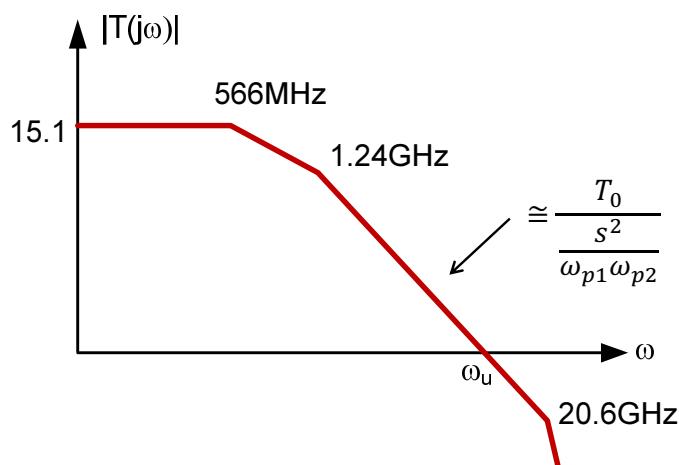
$$f_{p4} := \frac{1}{2 \cdot \pi} \cdot \frac{g_m + g_{mb}}{C_4} = 77.668\text{GHz}$$

Relevant

Somewhat relevant

Irrelevant

## Issue: Phase Margin



$$T(s) = \frac{T_0}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)}$$

$$\omega_u \cong \sqrt{T_0 \omega_{p1} \omega_{p2}}$$

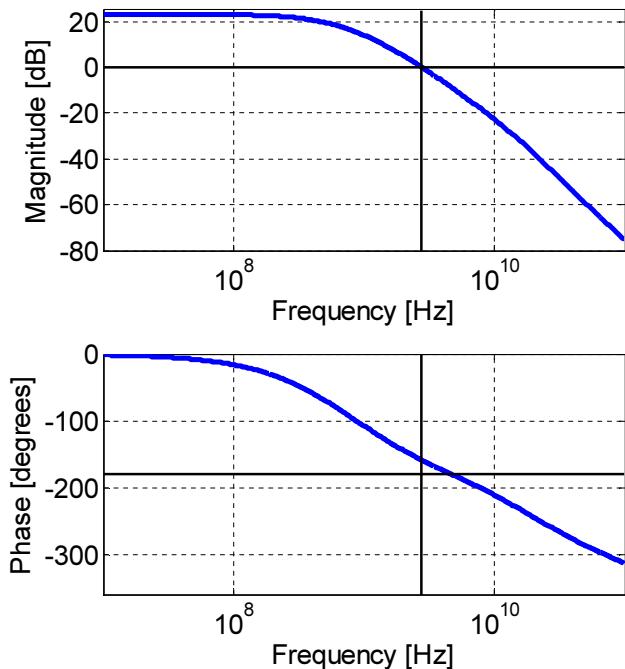
$$f_u \cong 3.2\text{GHz}$$

Side note:  $f_u$  cannot be increased by making  $R_D$  in stage 1 larger. Prove this.

$$PM = 180^\circ - \text{atan}\left(\frac{f_u}{f_{p1}}\right) - \text{atan}\left(\frac{f_u}{f_{p2}}\right) - \text{atan}\left(\frac{f_u}{f_{p3}}\right)$$

$$PM = 180^\circ - 80^\circ - 69^\circ - 12^\circ = 19^\circ$$

## Loop Gain Simulation



```
*** LSTB analysis ***
gain_margin(dB) =
8.711791

phase_margin(deg) =
20.96923

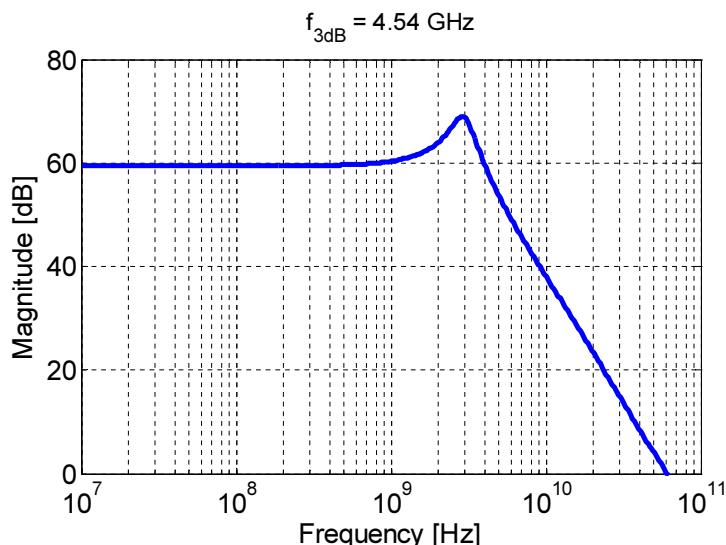
phase_margin_freq(Hz) =
2.7774433E+09

gain_margin_freq(Hz) =
4.6721000E+09

loop_gain_at_min_freq(dB) =
23.16820
```

See appendix for  
setup of LSTB  
analysis in HSpice

## Closed-Loop Response



- Lots of peaking, as expected
- Detrimental for RMS output noise,  $Q \sim 3 \rightarrow 3x$  noise penalty!

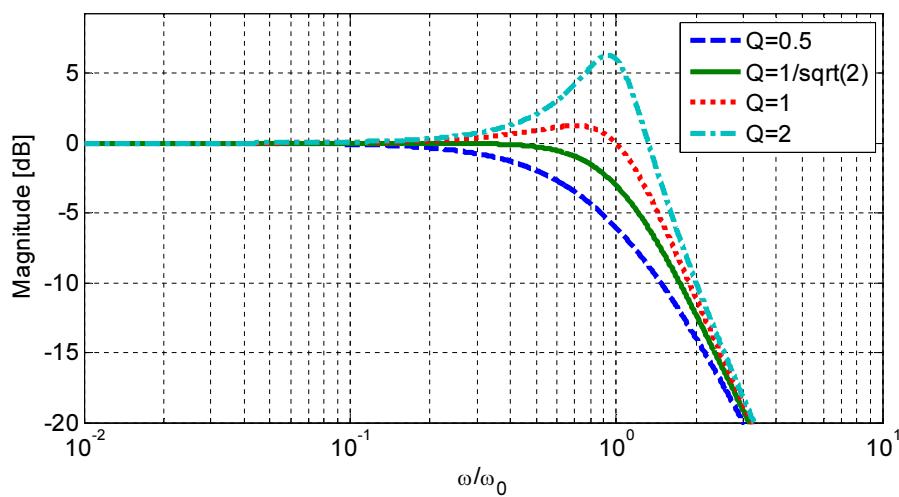
## Derivation of Closed-Loop Transfer Function

$$T(s) = \frac{T_0}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)}$$

$$A(s) = \frac{1}{f} \frac{T(s)}{1 + T(s)} = \frac{1}{f} \frac{T_0}{1 + T_0} \cdot \frac{1}{1 + \frac{\omega_{p1} + \omega_{p2}}{(1 + T_0)\omega_{p1}\omega_{p2}}s + \frac{1}{(1 + T_0)\omega_{p1}\omega_{p2}}s^2}$$

$$A(s) = \frac{A_0}{1 + \frac{s}{\omega_0 Q} + \frac{s^2}{\omega_0^2}}$$

$$\omega_0 = \sqrt{(1 + T_0)\omega_{p1}\omega_{p2}} \cong \omega_u \quad Q = \frac{\sqrt{(1 + T_0)\omega_{p1}\omega_{p2}}}{\omega_{p1} + \omega_{p2}}$$



$$H(s) = \frac{1}{1 + \frac{s}{\omega_0 Q} + \frac{s^2}{\omega_0^2}} \quad |H(s = j\omega_0)| = Q$$

## Closed-Loop Poles

- The closed-loop poles are the roots of the denominator polynomial

$$1 + \frac{s}{\omega_0 Q} + \frac{s^2}{\omega_0^2} = 0$$

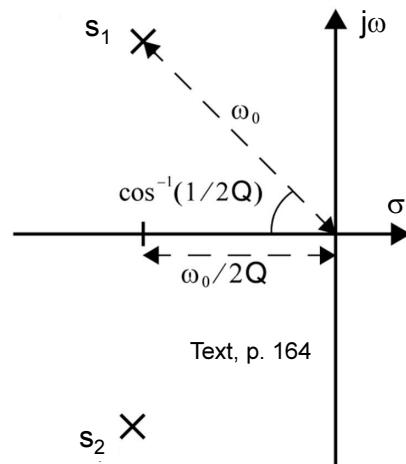
Location in the s-plane  
for  $Q > 0.5$

for  $Q > 0.5$        $s_{1,2} = -\frac{\omega_0}{2Q} \left( 1 \pm j\sqrt{4Q^2 - 1} \right)$

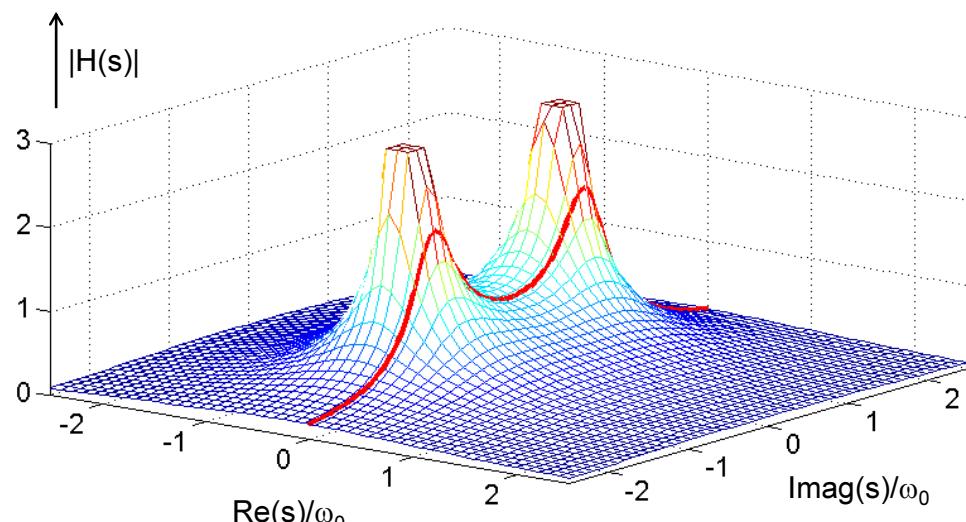
→ Complex Conjugate Poles

for  $Q \leq 0.5$        $s_{1,2} = -\frac{\omega_0}{2Q} \left( 1 \pm \sqrt{1 - 4Q^2} \right)$

→ Real Poles

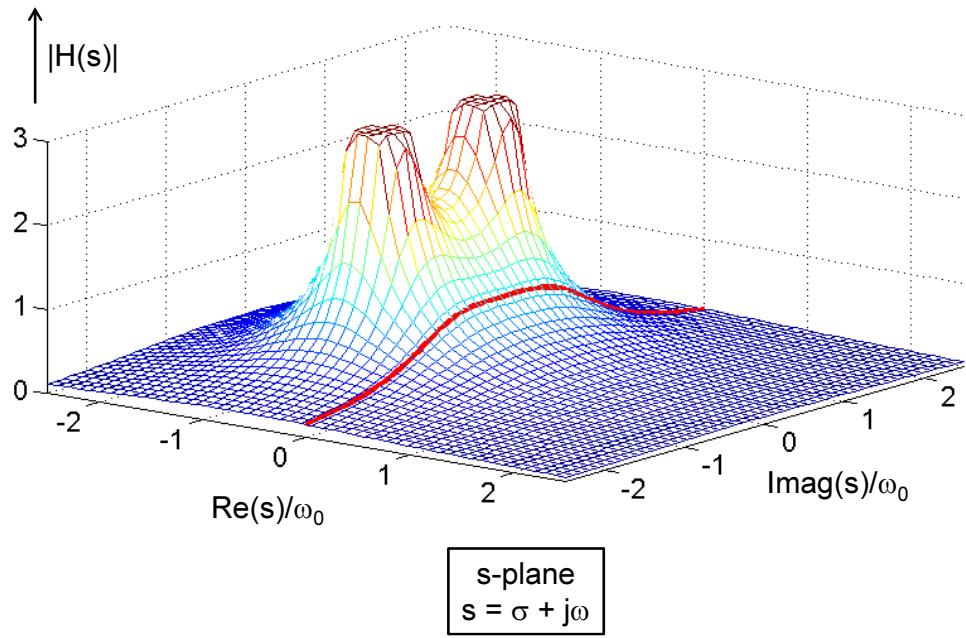


**$Q = 2$**

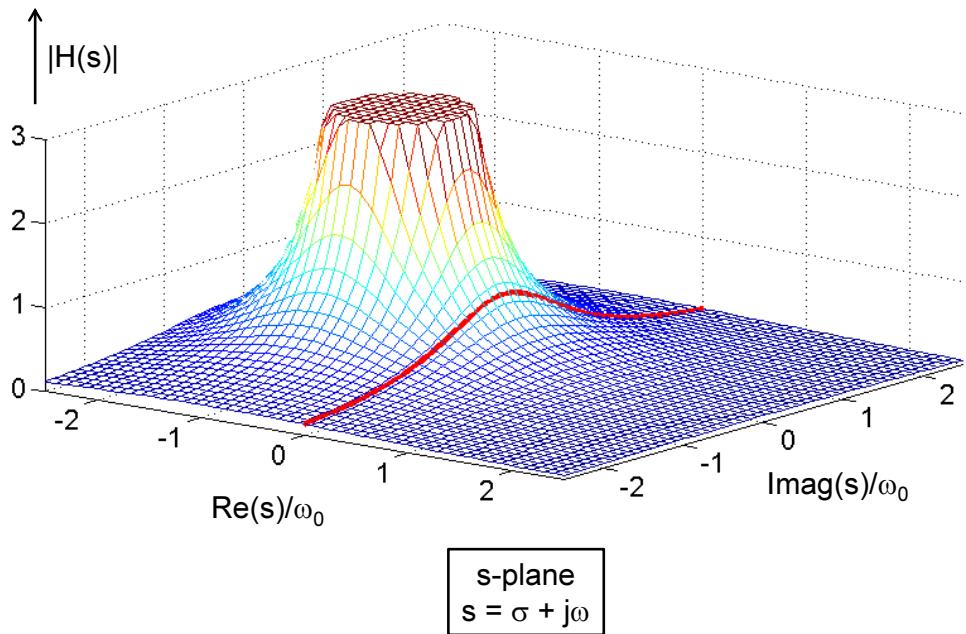


s-plane  
 $s = \sigma + j\omega$

$$Q = 1/\sqrt{2}$$



$$Q = 0.5$$

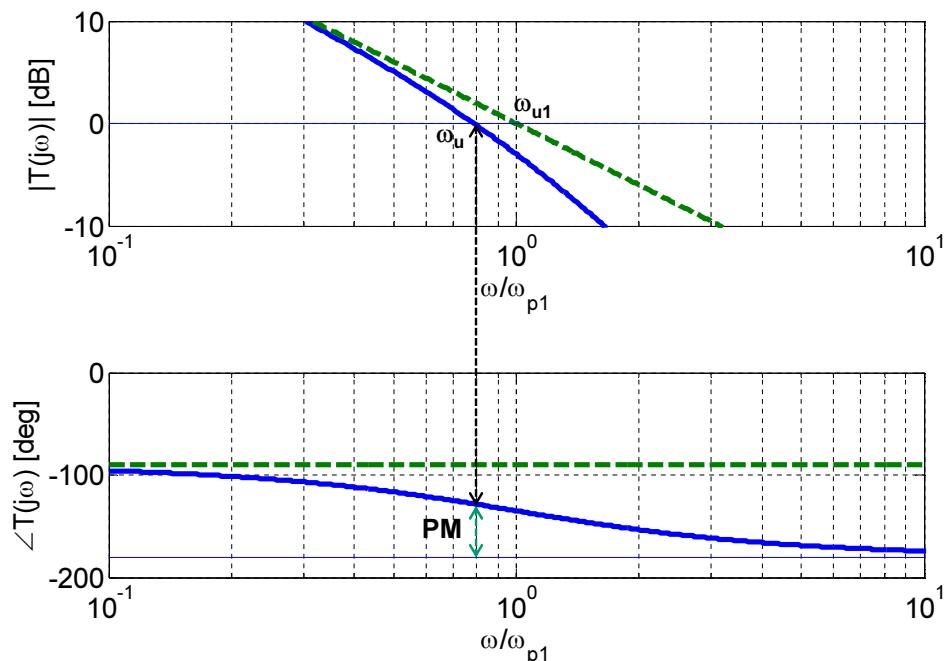


## Relationship Between Phase Margin and Q

$\omega_{p2}/\omega_{u1}$	$Q$	$\omega_u/\omega_{u1}$	PM (°)	$\omega_{3dB}/\omega_{u1}$
1	1	0.786	51.8	1.27
2	0.707	0.910	65.5	1.41
4	0.5	0.972	76.3	1.28
10	0.316	0.995	84.3	1.11
$\infty$	—	1	90	1

$\omega_u \rightarrow$  exact unity gain frequency  
 $\omega_{u1} \rightarrow$  approximate  $\omega_u$  based on linear extrapolation  
 $\omega_{3dB} \rightarrow$  closed loop – 3dB frequency

## Detailed View of Second Order Crossover



## Relationship Between Parameters

- We can find the relationship between  $\omega_u$  and  $\omega_{u1}$  using

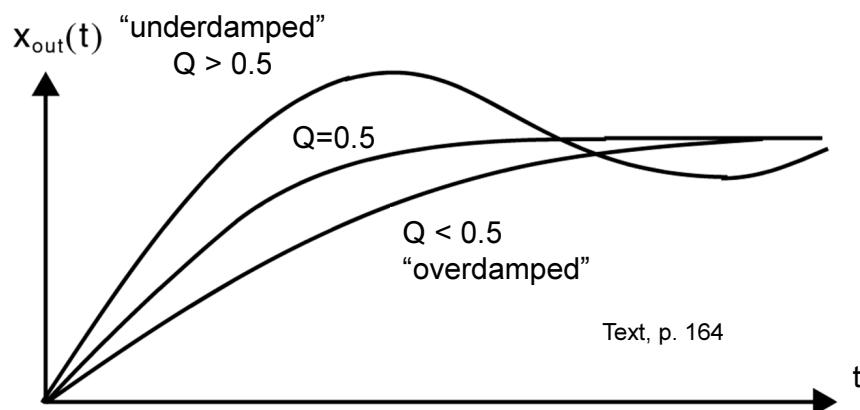
$$\left| \frac{T_0}{\left(1 + j\frac{\omega_u}{\omega_{p1}}\right) \left(1 + j\frac{\omega_u}{\omega_{p2}}\right)} \right| \approx \left| \frac{\omega_{u1}}{j\omega_u \left(1 + j\frac{\omega_u}{\omega_{p2}}\right)} \right| = 1$$

$$\Rightarrow \left(\frac{\omega_{u1}}{\omega_u}\right)^2 = \frac{1}{2} + \sqrt{\frac{1}{4} + \left(\frac{\omega_{u1}}{\omega_{p2}}\right)^2}$$

- Once we know  $\omega_u$ , we can easily compute the exact phase margin

$$PM = 180^\circ - 90^\circ - \arctan\left(\frac{\omega_u}{\omega_{p2}}\right)$$

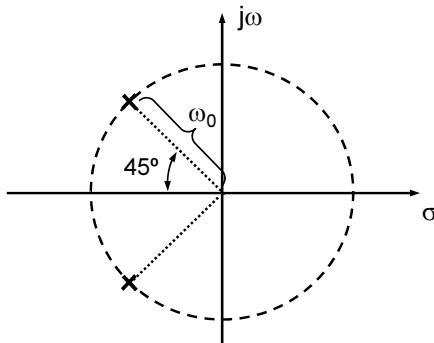
## Aside: Step Response



- Ringing for  $Q > 0.5$
- The case of  $Q = 0.5$  is called maximally damped response (fastest settling without any overshoot)

## What Should We Design For?

- A typical goal in wideband amplifier design is to achieve the maximum possible bandwidth without any peaking in the frequency response
- For a second-order system the corresponding solution is  $Q = 1/\sqrt{2}$ 
  - Closed-loop poles have equal real and imaginary parts
  - The closed loop bandwidth is equal to  $\omega_0$
  - The phase margin is approximately 65 degrees
- This response is called Maximally Flat Magnitude (MFM) or Butterworth response

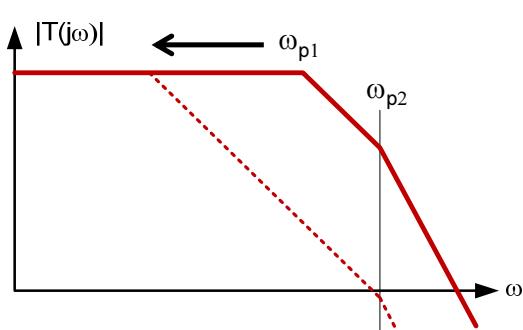


## How to Achieve a MFM Response?

- Our current design is quite far away from a MFM response

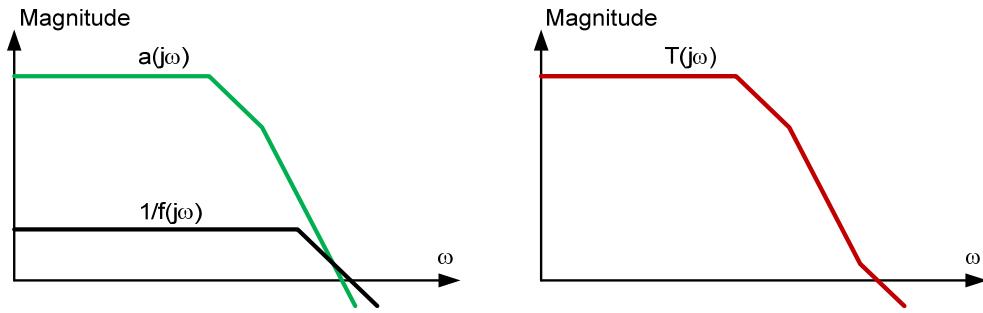
$$Q = \frac{\sqrt{(1 + T_0)\omega_{p1}\omega_{p2}}}{\omega_{p1} + \omega_{p2}} = \frac{\sqrt{(1 + 15.2)566MHz \cdot 1.24GHz}}{566MHz + 1.24GHz} = 1.9$$

- Getting the phase margin we need for an MFM response would require that one of the poles occurs beyond the unity gain frequency of the loop



- One (non-preferred) solution that achieves this is narrowbanding
  - Reduce  $\omega_{p1}$  (e.g. by increasing  $C_1$ ) until unity crossover occurs before  $\omega_{p2}$
- Key issue: substantial loss in bandwidth, as seen by the reduction in  $\omega_u$

## Better Idea: Feedback (or “Phantom”) Zero Compensation



- Leave the poles of  $a(s)$  untouched and introduce a left half plane zero in  $f(s)$ , close to the unity gain frequency of the loop
- Top first order, the phase margin will be 45 degrees (instead of 0 degrees) when the zero is placed at the unity gain frequency of the loop
- Two questions
  - How can we realize the zero?
  - How exactly should we position it to get a maximally flat response?

## Introducing a Zero in the Feedback Network

The figure shows two circuit diagrams. The left diagram illustrates a single-pole low-pass filter with input voltage  $v_1$  and output voltage  $v_2$ . The right diagram illustrates a two-pole low-pass filter with four admittance components ( $y_{11}$ ,  $y_{12}$ ,  $y_{21}$ ,  $y_{22}$ ) and input voltages  $v_1$  and  $v_2$ .

$$i_1 = y_{11}v_1 + y_{12}v_2$$

$$i_2 = y_{21}v_1 + y_{22}v_2$$

$$y_{11} = \left. \frac{i_1}{v_1} \right|_{v_2=0} = \frac{1}{R_F} + sC_F$$

$$y_{12} = \left. \frac{i_1}{v_2} \right|_{v_1=0} = -\left( \frac{1}{R_F} + sC_F \right)$$

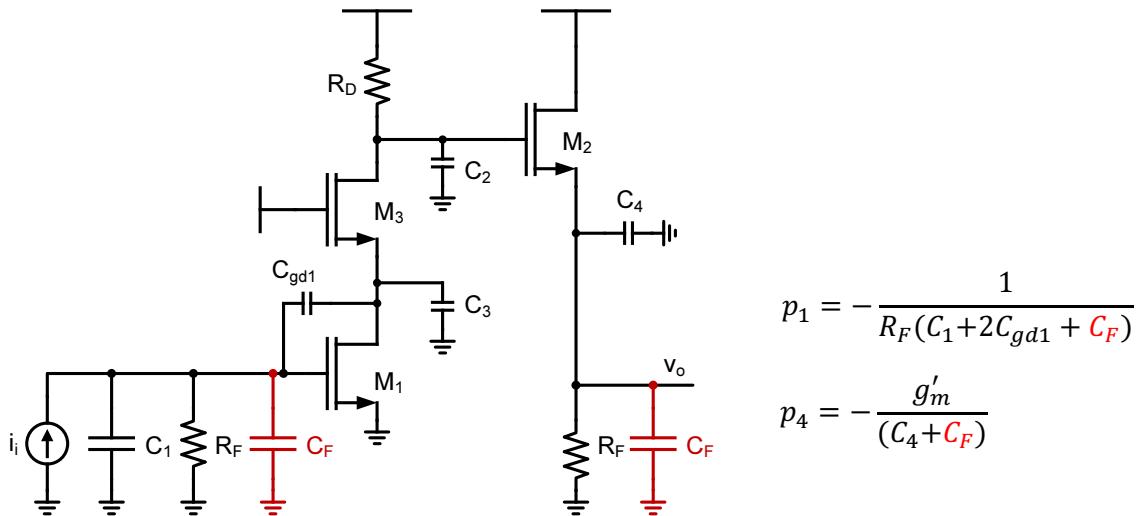
$$y_{21} = \left. \frac{i_2}{v_1} \right|_{v_2=0} = -\left( \frac{1}{R_F} + sC_F \right)$$

$$y_{22} = \left. \frac{i_2}{v_2} \right|_{v_1=0} = \frac{1}{R_F} + sC_F$$

$$f(s) = y_{12} = -\frac{1}{R_F} (1 + sR_F C_F)$$

$$z = -\frac{1}{R_F C_F}$$

## Modified Model of Forward Amplifier a(s)



- $C_F$  moves two of the poles to lower frequencies.
- However, we will see that  $C_F$  is typically very small, and thus the shift is usually not all that significant

## Derivation of the new Closed-Loop Transfer Function

$$T(s) = \frac{T_0 \left(1 + \frac{s}{\omega_z}\right)}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)} \quad f(s) = f_0 \left(1 + \frac{s}{\omega_z}\right)$$

$$A(s) = \frac{1}{f(s)} \frac{T(s)}{1 + T(s)} = \frac{1}{f_0} \frac{T_0}{1 + T_0} \cdot \frac{1}{1 + \frac{s}{\omega_0 Q} + \frac{s^2}{\omega_0^2}}$$

$$\omega_0 = \sqrt{(1 + T_0)\omega_{p1}\omega_{p2}} \cong \omega_u \quad \text{Unchanged}$$

$$\frac{1}{Q} = \frac{\omega_{p1} + \omega_{p2}}{\omega_0} + \frac{\omega_0}{\omega_z} \quad \text{New degree of freedom}$$

## Positioning the Zero

$$\frac{1}{Q} = \frac{\omega_{p1} + \omega_{p2}}{\omega_0} + \frac{\omega_0}{\omega_z} = \sqrt{2}$$

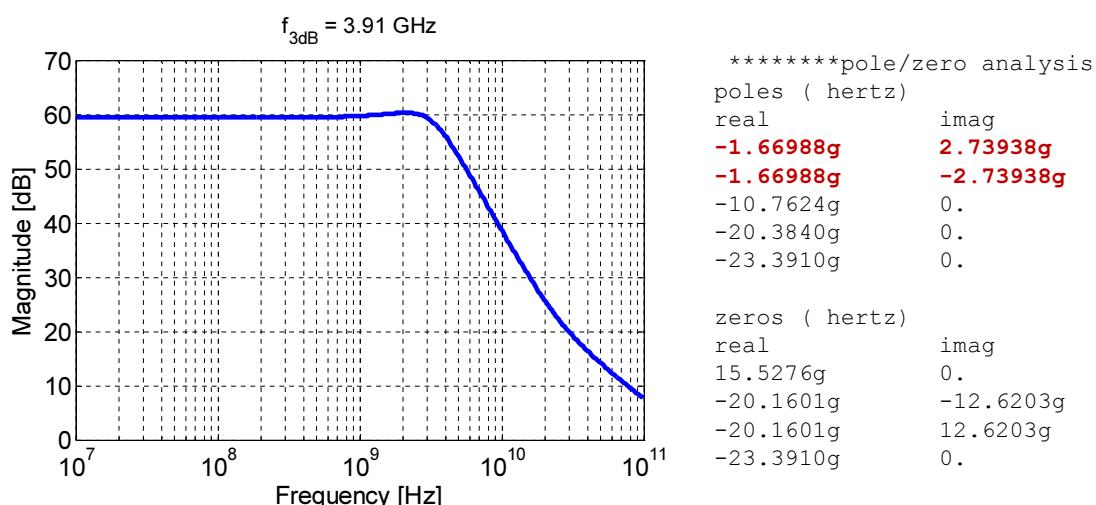
$$\omega_z = \frac{\omega_0}{\sqrt{2} - \frac{\omega_{p1} + \omega_{p2}}{\omega_0}} \cong \frac{\omega_0}{\sqrt{2}}$$

- For our design:

$$f_z = \frac{3.27\text{GHz}}{\sqrt{2} - \frac{566\text{MHz} + 1.24\text{GHz}}{3.27\text{GHz}}} = 3.8\text{GHz}$$

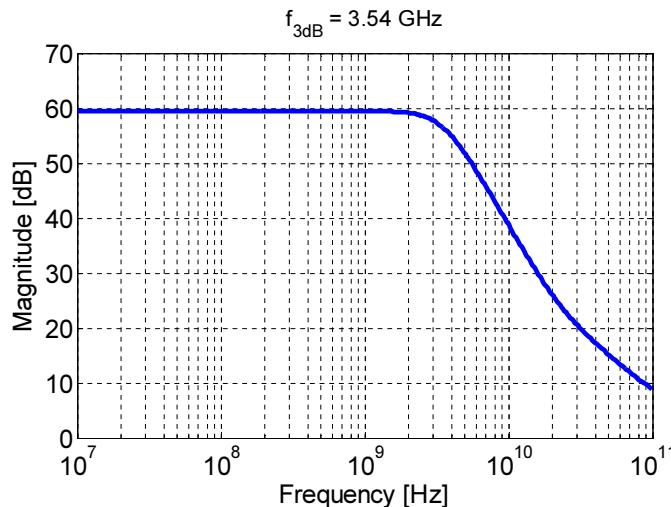
$$C_F = \frac{1}{2\pi f_z R_F} = 42\text{fF}$$

## Closed-Loop Response with $C_F = 42\text{fF}$



- We still see some minor peaking
- This is due to the fact that we have neglected higher frequency poles and the extra loading in the forward amplifier due to  $C_F$

## Closed-Loop Response with $C_F$ Tweaked to 55fF

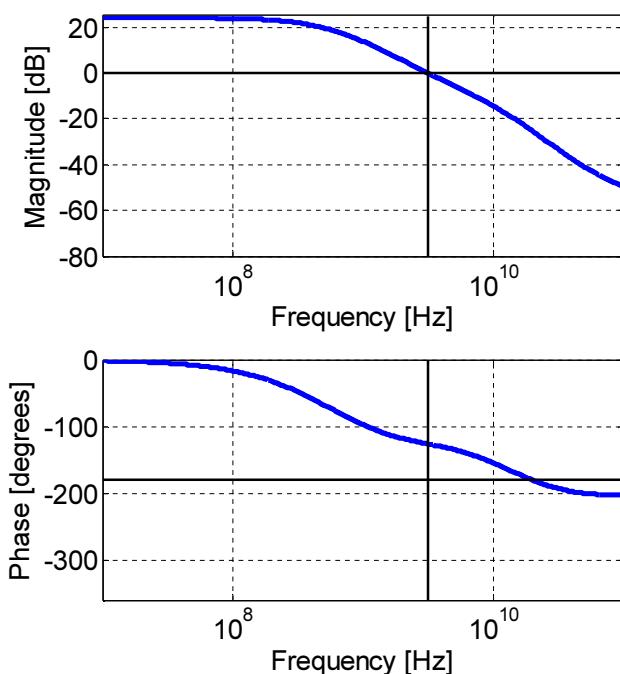


```
*****pole/zero analysis
poles ( hertz)
real           imag
-2.23916g    -2.62969g
-2.23916g    2.62969g
-8.09147g    0.
-20.3512g   0.
-23.3910g   0.

zeros (rad/sec)
zeros ( hertz)
real           imag
14.0106g    0.
-18.6215g   -11.4893g
-18.6215g   11.4893g
-23.3910g   0.
```

- The poles are still not exactly angled at 45 degrees, but this is OK
  - A high frequency pole (not analyzed) helps reduce any residual peaking

## Loop Gain Simulation



```
*** LSTB analysis ***
gain_margin(dB) =
24.02825

phase_margin(deg) =
55.81567

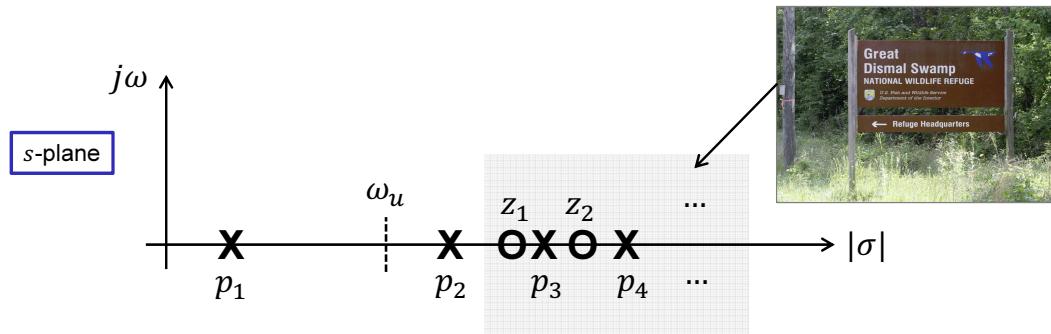
phase_margin_freq(Hz) =
3.1666851E+09

gain_margin_freq(Hz) =
1.9178802E+10

loop_gain_at_min_freq(dB) =
23.16766
```

Phase margin somewhat smaller than expected (again, due to high frequency poles that we did not consider)

## Impact of High-Frequency Poles/Zeros



- The majority of practical circuits have a dominant pole and a non-dominant pole past the loop's unity gain frequency
- In addition, there are usually several poles (LHP) and zeros (LHP and RHP) beyond the second pole
- Despite these extra poles/zeros, we can usually still approximate the system as second order
  - The extra poles/zeros mainly affect the phase margin and don't have much bearing on  $ω_u$

$$T(s) = \frac{T_0}{\left(1 - \frac{s}{p_1}\right)\left(1 - \frac{s}{p_2}\right)} \frac{\left(1 - \frac{s}{z_1}\right)\left(1 - \frac{s}{z_2}\right) \dots \left(1 - \frac{s}{z_m}\right)}{\left(1 - \frac{s}{p_3}\right) \dots \left(1 - \frac{s}{p_n}\right)}$$

$$T(jω) \cong \frac{T_0}{\left(1 + j \frac{\omega}{ω_{p1}}\right)\left(1 + j \frac{\omega}{ω_{p2eq}}\right)}$$

- The equivalent non-dominant pole frequency is given by the frequency at which the overall phase shift of the loop (with all poles and zeros included) is  $-135^\circ$
- This approximation is particularly convenient for interpreting circuit simulation results
- Some rules of thumb
  - A LHP pole or RHP zero 10x past  $ω_u$  steals about  $5.5^\circ$  of PM
  - A LHP pole or RHP zero 5x past  $ω_u$  steals about  $11^\circ$  of PM

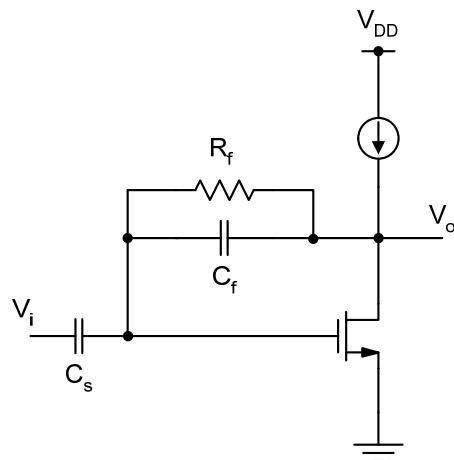
## APPENDIX

### Loop Gain Simulation

### References

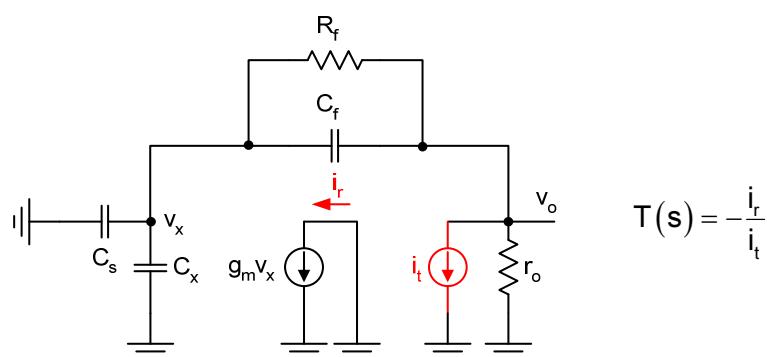
- H.W. Bode, Network Analysis and Feedback Amplifier Design, Van Nostrand, New York, 1945.
- R.D. Middlebrook, "Measurement of Loop Gain in Feedback Systems," Int. J. Electronics, Vol. 38, No.4, pp. 485-512, 1975.
- S. Rosenstark, "Loop Gain Measurement in Feedback Amplifiers," Int. J. Electronics, Vol. 57, No.3., pp. 415-421, 1984.
- P.J. Hurst, "Exact Simulation of Feedback Circuit Parameters," Trans. on Circuits and Systems, pp.1382-1389, Nov. 1991.
- P.J. Hurst, S.H. Lewis, "Simulation of Return Ratio in Fully Differential Feedback Circuits," Proc. CICC 1994, pp.29-32.
- M. Tian, V. Visvanathan, J. Hantgan, K. Kundert, "Striving for small-signal stability," IEEE Circuits and Devices Magazine, pp. 31-41, January 2001.
- F. Wiedmann, "Loop Gain Simulation,"  
<https://sites.google.com/site/frankwiedmann/loopgain>

## Circuit Example



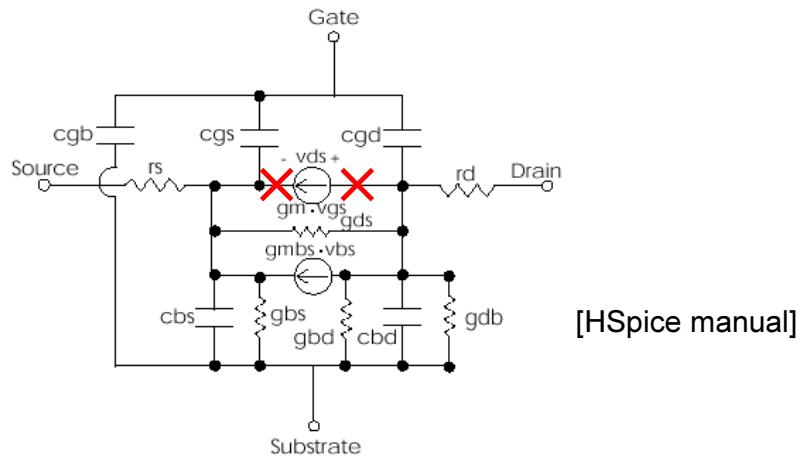
- What is the loop gain in this circuit?

## Return Ratio Analysis



- Hand analysis, for example using the return ratio method, is straightforward
- How can we simulate  $T(j\omega)$  in Spice?
  - Using "real" transistor models

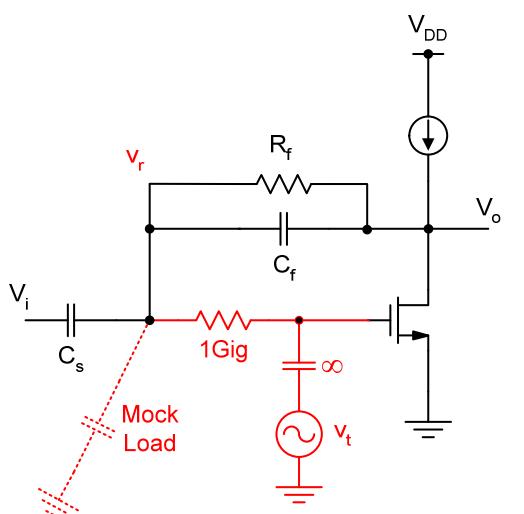
## Spice MOSFET AC Simulation Model



Nodes of the controlled source are not accessible!

- Cannot break loop at  $g_m$  generator

## Popular (but Non-Preferred) Simulation Approach

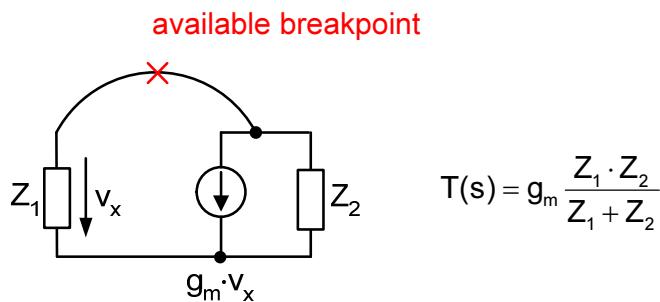


$$T(j\omega) \approx -\frac{V_r}{V_t}$$

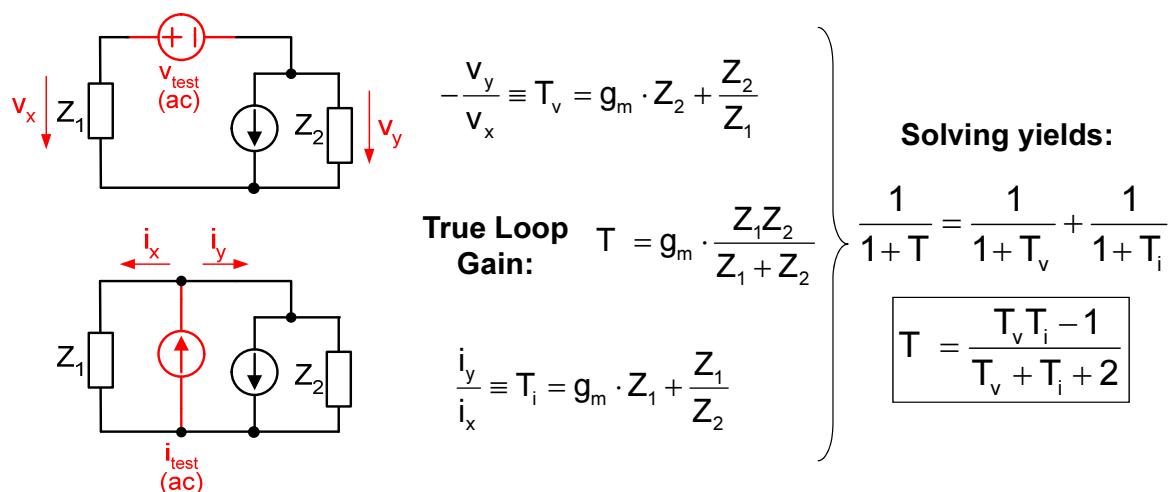
- Inaccurate
- Hard to estimate mock load
- May get different results for different breakpoints
- Ideally, we'd like to avoid all of the above issues
- Solution: Middlebrook method

## Problem Generalization

- Middlebrook argued that any single loop feedback circuit can be partitioned as shown below
- Hence, there is always some "nonideal" breakpoint between impedances
  - How can we use this breakpoint to find the loop gain?



## Double Injection Trick



- No “DC“ break in the loop, all loading effects included!
- Measure  $T_v$  and  $T_i$  separately, then calculate actual  $T$

## Implementation in Circuit Simulators

- The Middlebrook method (or a variant thereof) is implemented in most modern circuit simulators such as HSpice and Spectre
  - LSTB analysis in HSpice
  - STB analysis in Spectre
- For details, refer to HSpice manual
  - /usr/class/ee/synopsys/hspice/F-2011.09 SP2/hspice/docs\_help>/hspice\_aasa.pdf
- Important note
  - In class and in all major textbooks, the loop phase is defined as zero for negative feedback
  - Unfortunately, LSTB and STB analyses report the phase 180 degrees shifted
    - Easy to fix; don't get confused by this...

## LSTB Syntax

### Loop Stability Analysis Usage

```
.LSTB mode=[single|diff|comm]
+ vsource=[vlstb|vlstbp,vlstbn]
```

### Examples

Single-mode loop analysis on loop indicated by vx voltage source:

```
.LSTB mode=single vsource=vx
```

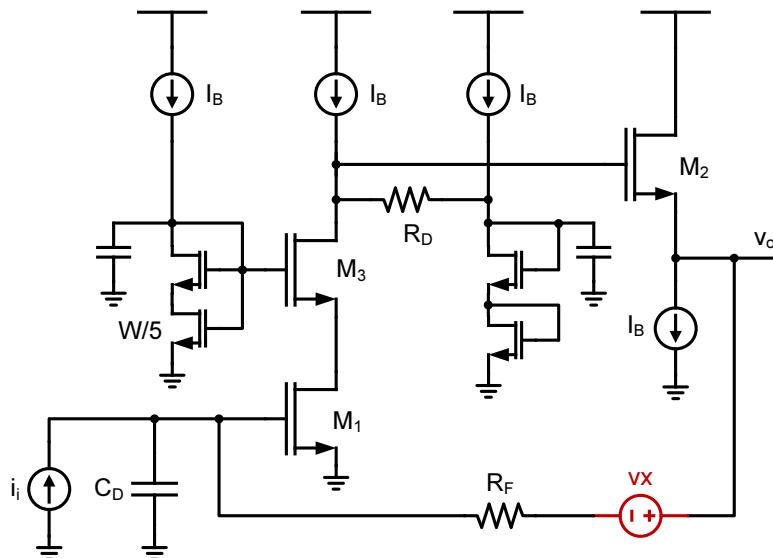
Differential-mode loop analysis on loops indicated by vp and vn voltage sources:

```
.LSTB mode=diff vsource=vp,vn
```

Common-mode loop analysis on loops indicated by vp and vn voltage sources:

```
.LSTB mode=comm vsource=vp,vn
```

## LSTB Simulation Setup for TIA Example



```
.ac dec 100 10e6 100e9
.lstb mode=single vsource=vx
.probe ac lstb(db) lstb(p)
```

## How About Multiple Feedback Loops?

- Any practical feedback circuit has multiple feedback loops
  - Fully differential circuits have CM/DM loops (see EE315A)
  - Local device feedback through  $C_{gd}$ ,  $R_{source}$
  - ...
- Solutions
  - Decompose fully differential circuit into CM/DM loops
  - If a local feedback loop can be modeled as a combination of a stable controlled source and passive impedances, the multi-loop circuit reduces to a single loop [Hurst 94]
  - If there is a common breakpoint that breaks all feedback loops simultaneously, stability can be checked by finding the return ratio at the single breakpoint [Hurst 94]

## Example

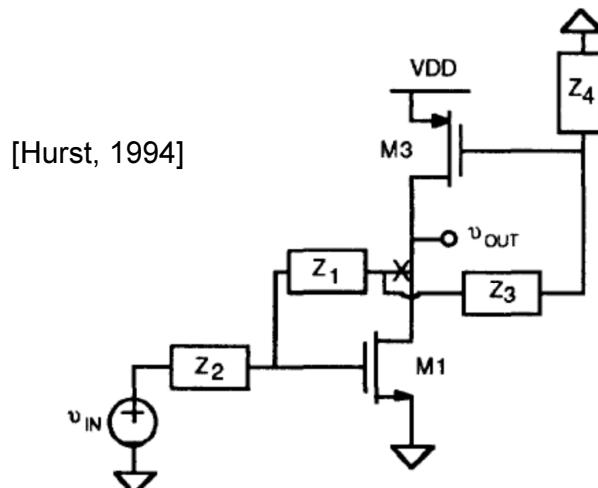


Fig. 6. An example of a circuit with multiple loops and one break point that opens all feedback loops.

## Last Resort: General Nyquist Criterion

[Bode 45]:

"If a circuit is stable when all its tubes have their nominal gains, the total number of clockwise and counterclockwise encirclements of the critical point must be equal to each other in the series of Nyquist diagrams for the individual tubes obtained by beginning with all tubes dead and restoring the tubes successively in any order to their nominal gains"

[You may want to take a controls class if you are interested in this...]



## Another Useful Quote

[Bode 45]:

“... thus the circuit may sing when the tubes begin to lose their gain because of age, and it may also sing, instead of behaving as it should, when the gain increases from zero as power is supplied to the circuit...”



**Always run one or more transient analyses for a  
"true" stability check!**

# Chapter 9

## Root Locus and

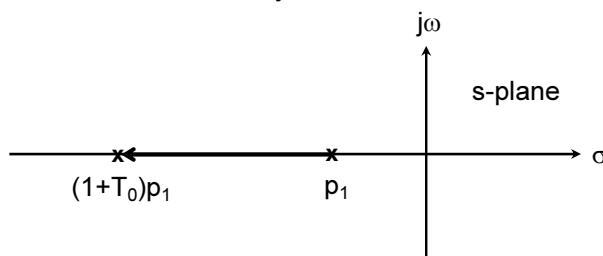
## Frequency Compensation

Boris Murmann  
Stanford University  
Winter 2015-16

Reference: Gray, Hurst, Lewis & Meyer, Chapter 9  
Textbook Section: 6.2

### Root Locus

- As we have seen from our analysis of first- and second-order feedback systems, applying feedback around an amplifier moves the open-loop poles to a new location
- Example: 1<sup>st</sup> order feedback system



- A so-called “root locus” plot shows the movement of the poles in the s-plane as we vary the low-frequency loop gain  $T_0$
- Root locus plots are most commonly used in control theory, but can provide valuable intuition for the design of amplifiers and other electronic circuits

## Second-Order System

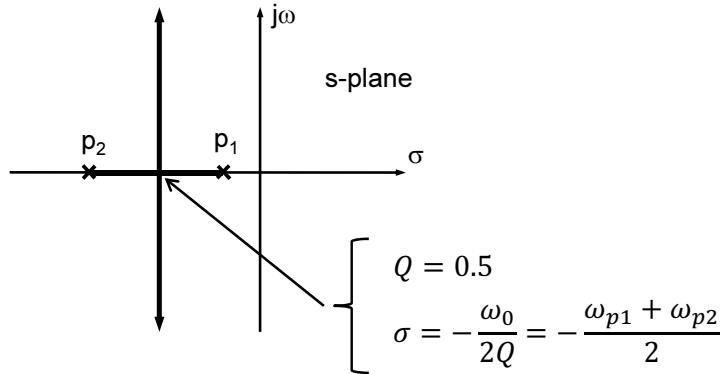
$$H(s) = \frac{1}{1 + \frac{s}{\omega_0 Q} + \frac{s^2}{\omega_0^2}}$$

*for  $Q > 0.5$*        $s_{1,2} = -\frac{\omega_0}{2Q} (1 \pm j\sqrt{4Q^2 - 1})$

$$\omega_0 = \sqrt{(1 + T_0)\omega_{p1}\omega_{p2}}$$

*for  $Q \leq 0.5$*        $s_{1,2} = -\frac{\omega_0}{2Q} (1 \pm \sqrt{1 - 4Q^2})$

$$Q = \frac{\sqrt{(1 + T_0)\omega_{p1}\omega_{p2}}}{\omega_{p1} + \omega_{p2}}$$



## Third Order System

- Consider a feedback network consisting of a forward amplifier with three identical poles, and a feedback network with a constant transfer function  $f$

$$a(s) = \frac{a_0}{\left(1 - \frac{s}{p_1}\right)^3}$$

$$A(s) = \frac{a(s)}{1 + a(s)f} = \frac{\frac{a_0}{\left(1 - \frac{s}{p_1}\right)^3}}{1 + a \frac{a_0}{\left(1 - \frac{s}{p_1}\right)^3} f} = \frac{a_0}{\left(1 - \frac{s}{p_1}\right)^3 + T_0}$$

$$T_0 = a_0 f$$

- The poles of  $A(s)$  are therefore the solution to

$$\left(1 - \frac{s}{p_1}\right)^3 + T_0 = 0$$

$$\left(1 - \frac{s}{p_1}\right)^3 = -T_0$$

$$\left(1 - \frac{s}{p_1}\right) = \sqrt[3]{-T_0} = -\sqrt[3]{T_0} \quad \text{or} \quad \left(1 - \frac{s}{p_1}\right) = \sqrt[3]{T_0} e^{j60^\circ} \quad \text{or} \quad \left(1 - \frac{s}{p_1}\right) = \sqrt[3]{T_0} e^{-j60^\circ}$$

$$s_1 = p_1 \left(1 + \sqrt[3]{T_0}\right)$$

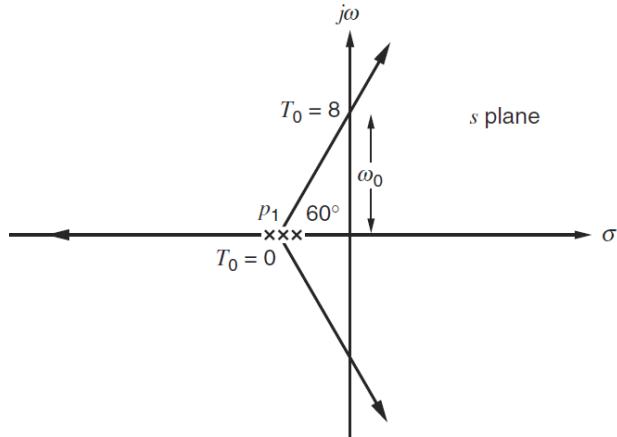
$$s_2 = p_1 \left(1 - \sqrt[3]{T_0} e^{j60^\circ}\right)$$

$$s_3 = p_1 \left(1 - \sqrt[3]{T_0} e^{-j60^\circ}\right)$$

$$0 = 1 - \operatorname{Re} \left( \sqrt[3]{T_0} e^{j60^\circ} \right)$$

$$0 = 1 - \sqrt[3]{T_0} \cos(60^\circ)$$

$$\Rightarrow T_0 = 8$$



- Conclusion: An amplifier with three identical poles is unstable unless we limit the low-frequency loop gain to less than eight

## Generalization

The poles of  $A(s)$  are the roots of  $1 + T(s) = 0$ .

In general,

$$a(s) = a_0 \frac{1 + a_1 s + a_2 s^2 + \dots}{1 + b_1 s + b_2 s^2 + \dots} = a_0 \frac{N_a(s)}{D_a(s)}$$

and

$$f(s) = f_0 \frac{1 + c_1 s + c_2 s^2 + \dots}{1 + d_1 s + d_2 s^2 + \dots} = f_0 \frac{N_f(s)}{D_f(s)}$$

Thus,

$$A(s) = \frac{a_0 N_a(s) D_f(s)}{D_a(s) D_f(s) + T_0 N_a(s) N_f(s)}$$

where  $T_0 = a_0 f_0$ .

The **zeros** of  $A(s)$  are the **zeros** of  $a(s)$  and the **poles** of  $f(s)$ .

The **poles** of  $A(s)$  are the roots of

$$D_a(s)D_f(s) + T_0 N_a(s)N_f(s) = 0$$

As  $T_0$  increases from 0 to  $\infty$ , the poles of  $A(s)$  move in the s-plane from the poles of  $a(s)$  to the zeros of  $f(s)$ .

$$T_0=0 \quad A(s) = \frac{a_0 N_a(s) D_f(s)}{D_a(s) D_f(s) + T_0 N_a(s) N_f(s)}$$

$$T_0 \rightarrow \infty \quad A(s) = \frac{a_0 N_a(s) D_f(s)}{D_a(s) D_f(s) + T_0 N_a(s) N_f(s)}$$

We can establish general root-locus construction rules using

$$1 + T(s) = 1 + T_0 \cdot \frac{N_a(s) \cdot N_f(s)}{D_a(s) \cdot D_f(s)} = 0 \quad T_0 \cdot \frac{N_a(s) \cdot N_f(s)}{D_a(s) \cdot D_f(s)} = -1$$

$$T_0 \cdot \frac{(1-s/z_{a1})(1-s/z_{a2})\cdots(1-s/z_{f1})(1-s/z_{f2})\cdots}{(1-s/p_{a1})(1-s/p_{a2})\cdots(1-s/p_{f1})(1-s/p_{f2})\cdots} = -1$$

where

$z_{a1}, z_{a2}, \dots$  = zeros of  $a(s)$

$z_{f1}, z_{f2}, \dots$  = zeros of  $f(s)$

$p_{a1}, p_{a2}, \dots$  = poles of  $a(s)$

$p_{f1}, p_{f2}, \dots$  = poles of  $f(s)$

The above equation can be rewritten as

$$T_0 \cdot \left[ \frac{(-p_{a1})(-p_{a2}) \cdots (-p_{f1})(-p_{f2}) \cdots}{(-z_{a1})(-z_{a2}) \cdots (-z_{f1})(-z_{f2}) \cdots} \right] \\ \times \left[ \frac{(s - z_{a1})(s - z_{a2}) \cdots (s - z_{f1})(s - z_{f2}) \cdots}{(s - p_{a1})(s - p_{a2}) \cdots (s - p_{f1})(s - p_{f2}) \cdots} \right] = -1$$

Expect all poles of  $T(s)$  to be in the left half plane (LHP).

If all zeros of  $T(s)$  are in the LHP, or if there are an **even** number of zeros in the RHP, then the first bracketed term in the above equation is positive, in which case

$$T_0 \cdot \left[ \frac{|p_{a1}| |p_{a2}| \cdots |p_{f1}| |p_{f2}| \cdots}{|z_{a1}| |z_{a2}| \cdots |z_{f1}| |z_{f2}| \cdots} \right] \\ \times \left[ \frac{(s - z_{a1})(s - z_{a2}) \cdots (s - z_{f1})(s - z_{f2}) \cdots}{(s - p_{a1})(s - p_{a2}) \cdots (s - p_{f1})(s - p_{f2}) \cdots} \right] = -1$$

Values of  $s$  satisfying the above equation are the poles of  $A(s)$ . These values simultaneously fulfill both a **phase condition** and a **magnitude condition**, and these conditions define the points of the root locus.

### Phase Condition

$$\left[ \angle(s - z_{a1}) + \angle(s - z_{a2}) + \cdots + \angle(s - z_{f1}) + \angle(s - z_{f2}) + \cdots \right] \\ - \left[ \angle(s - p_{a1}) + \angle(s - p_{a2}) + \cdots + \angle(s - p_{f1}) + \angle(s - p_{f2}) + \cdots \right] \\ = (2n - 1)\pi$$

### Magnitude Condition

$$T_0 \cdot \left[ \frac{|p_{a1}| |p_{a2}| \cdots |p_{f1}| |p_{f2}| \cdots}{|z_{a1}| |z_{a2}| \cdots |z_{f1}| |z_{f2}| \cdots} \right] \cdot \left[ \frac{|s - z_{a1}| |s - z_{a2}| \cdots |s - z_{f1}| |s - z_{f2}| \cdots}{|s - p_{a1}| |s - p_{a2}| \cdots |s - p_{f1}| |s - p_{f2}| \cdots} \right] = +1$$

For the case where there are an **odd** number of zeros in the RHP, the **magnitude condition** remains the same as above, but the **phase condition** is changed. Specifically,

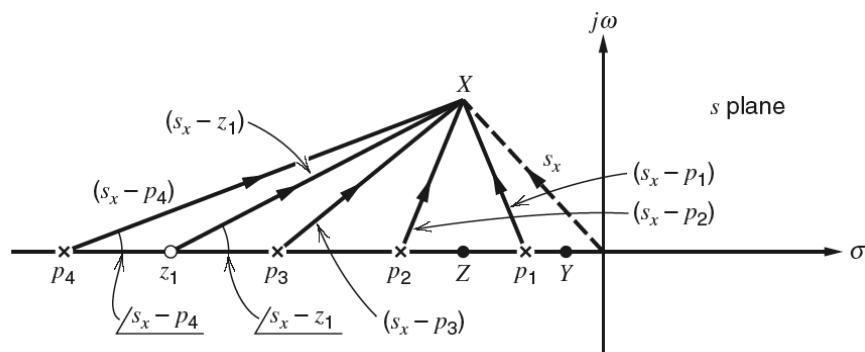
$$T_0 \cdot \left[ \frac{|p_{a1}| |p_{a2}| \cdots |p_{f1}| |p_{f2}| \cdots}{|z_{a1}| |z_{a2}| \cdots |z_{f1}| |z_{f2}| \cdots} \right] \cdot \left[ \frac{|s - z_{a1}| |s - z_{a2}| \cdots |s - z_{f1}| |s - z_{f2}| \cdots}{|s - p_{a1}| |s - p_{a2}| \cdots |s - p_{f1}| |s - p_{f2}| \cdots} \right] = +1$$

and

$$\begin{aligned} & [\angle(s - z_{a1}) + \angle(s - z_{a2}) + \cdots + \angle(s - z_{f1}) + \angle(s - z_{f2}) + \cdots] \\ & - [\angle(s - p_{a1}) + \angle(s - p_{a2}) + \cdots + \angle(s - p_{f1}) + \angle(s - p_{f2}) + \cdots] \\ & = 2n\pi \end{aligned}$$

The rules for constructing the root locus are based on the **phase condition**. The **magnitude condition** determines where, for a given  $T_0$ , the poles of  $A(s)$  actually lie on along the locus.

To determine if a point  $X$  in the  $s$ -plane lies on the root locus, draw vectors from the poles and zeros of  $T(s)$  to the point  $X$ . The angles of these vectors are then used to check the **phase condition**.



## Root-Locus Construction Rules

**Rule 1:** Branches of the root locus start at the poles of  $T(s)$ , where  $T_0 = 0$ , and terminate on the zeros of  $T(s)$ , where  $T_0 = \infty$ . If  $T(s)$  has more poles than zeros, some branches terminate at infinity.

**Rule 2:** If  $T(s)$  has all its zeros in the LHP or if  $T(s)$  has an even number of RHP zeros, the locus is situated along the real axis wherever there is an odd number of poles and zeros of  $T(s)$  to the right. If  $T(s)$  has an odd number of RHP zeros, the locus is situated along the real axis wherever there is an even number of poles and zeros of  $T(s)$  to the right.

**Rule 3:** All segments of the locus on the real axis between pairs of poles, or pairs of zeros, must branch out from the real axis.

**Rule 4:** The locus is symmetric with respect to the real axis because complex roots occur only in conjugate pairs.

**Rule 5:** Branches leaving the real axis do so at right angles to it.

**Rule 6:** Branches break away from the real axis at points where the vector sum of reciprocals of distances to the poles of  $T(s)$  equals the vector sum of reciprocals of distances to the zeros of  $T(s)$ .

**Rule 7:** Branches that terminate at infinity do so asymptotically to straight lines with angles to the real axis of

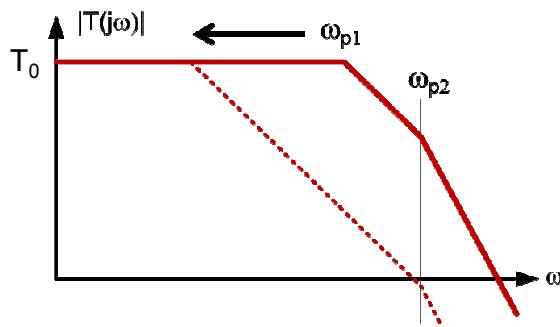
$$(2n - 1)\pi/(N_p - N_z),$$

where  $N_p$  = # of poles of  $T(s)$  and  $N_z$  = # of zeros of  $T(s)$ .

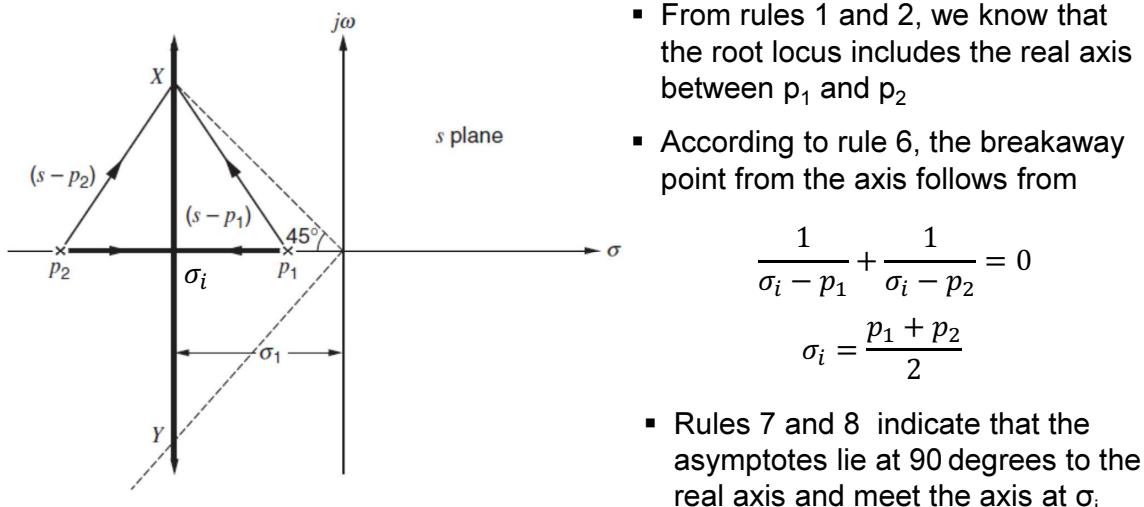
**Rule 8:** The asymptotes of the branches terminating at infinity all intersect the real axis at a single point given by

$$\sigma_a = \frac{\sum [\text{poles of } T(s)] - \sum [\text{zeros of } T(s)]}{N_p - N_z}$$

## Example: Root Locus for Narrowbanding Compensation



- Problem statement
  - For the given low-frequency loop gain  $T_0$  and high-frequency pole  $\omega_{p2}$ , find the proper value of  $\omega_{p1}$  that gives a maximally flat magnitude response for the closed-loop amplifier



- To find the point where the poles lie at an angle of 45 degrees, we invoke the magnitude condition

$$T_0 \frac{|p_1||p_2|}{|s - p_1||s - p_2|} = 1$$

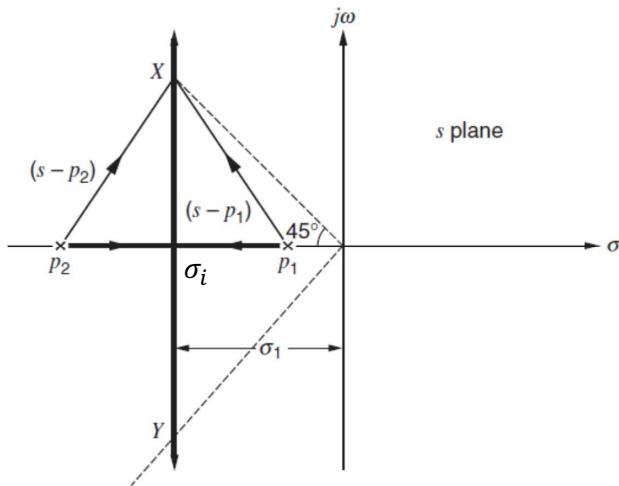
- If we approximate  $|p_1| \ll |p_2|$ , then

$$\sigma_i \cong \frac{p_2}{2}$$

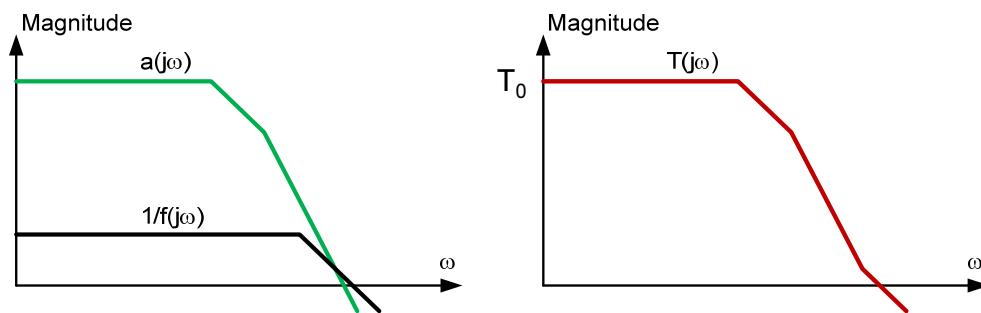
$$|s - p_1| = |s - p_2| \cong \sqrt{2} \frac{|p_2|}{2}$$

$$T_0 \frac{|p_1||p_2|}{\frac{|p_2|^2}{2}} = 1 \quad |p_1| = \frac{1}{2} \frac{|p_2|}{T_0}$$

$$\boxed{\omega_{p1} = \frac{1}{2} \frac{\omega_{p2}}{T_0}}$$

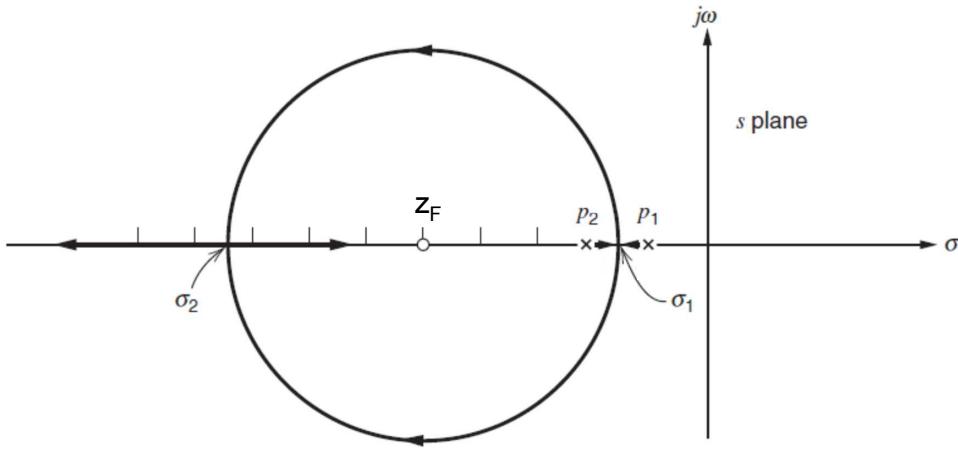


## Example: Root Locus for Feedback Zero Compensation

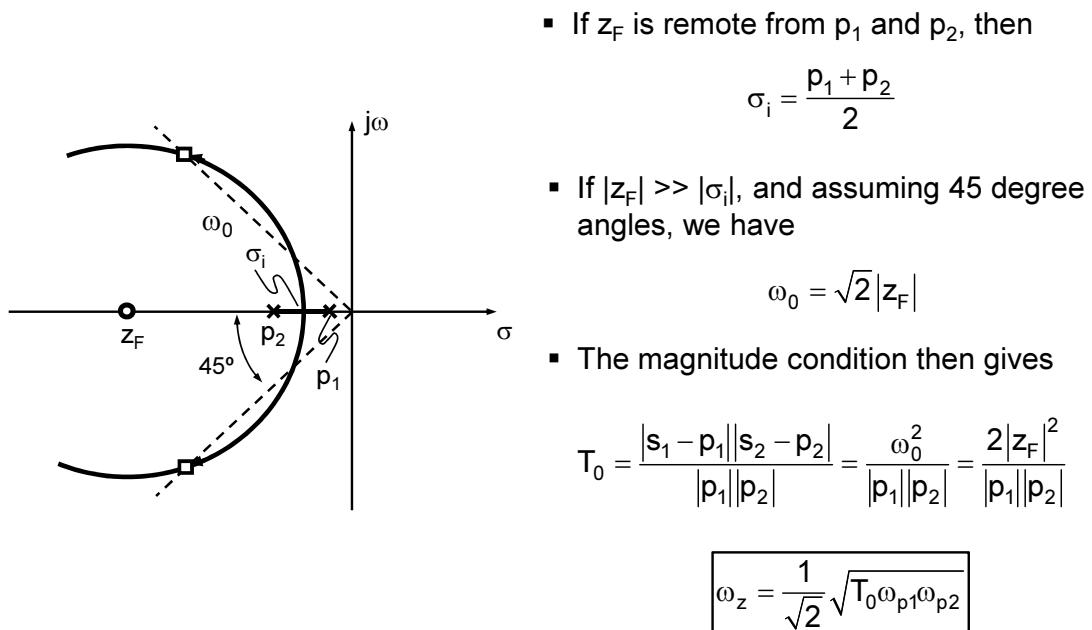


- Problem statement

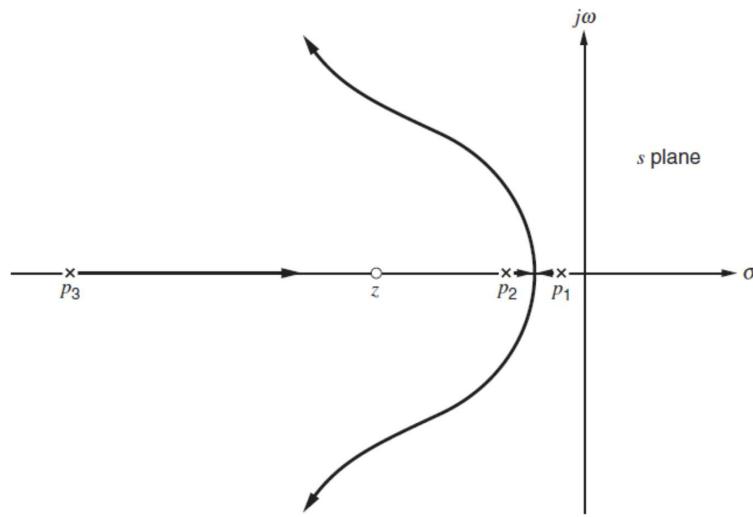
- For the given low-frequency loop gain and poles of  $a(s)$ , find the proper position of the feedback zero that gives a maximally flat response



- From rules 1 and 2, we know that the root locus lies between  $p_1$  and  $p_2$  and also to the left of  $z_F$
- Applying rule 6 gives two breakpoints, one between  $p_1$  and  $p_2$ , and one where the poles return to the real axis after circling around  $z$
- In general, the locus tends to bend toward zeros as if attracted and tends to bend away from poles as if repelled



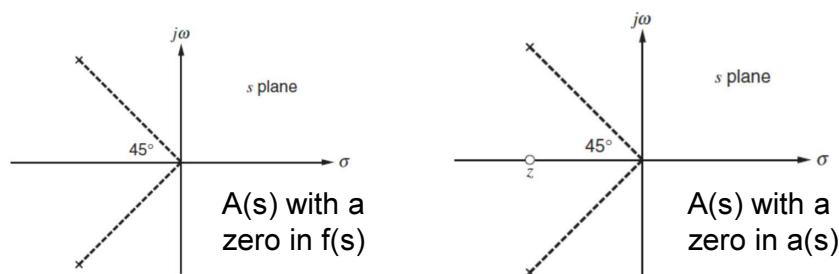
## Root Locus With a Third Pole



- The above-computed results can be affected by a high-frequency third pole, mandating a certain amount of “tweaking”

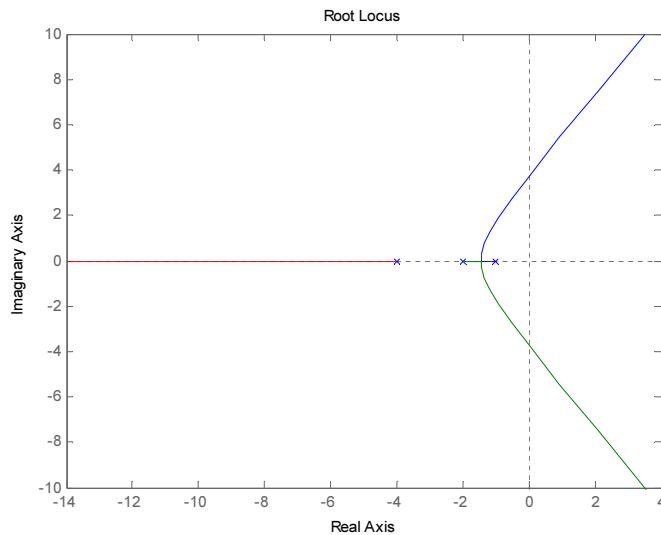
## “Phantom Zero”

- Feedback zero compensation is also called “phantom zero” compensation since the feedback zero affects the root locus, but does not appear as a zero of the overall amplifier
- The zero in the root locus plot is contributed by  $f(s)$  and is not a zero of the overall feedback amplifier
  - Recall that the zeros of the overall feedback amplifier are the zeros of basic amplifier  $a(s)$  and the poles of feedback network  $f(s)$
- If the zero in  $T(s)$  was contributed by  $a(s)$ , it would also appear as a zero in the overall amplifier



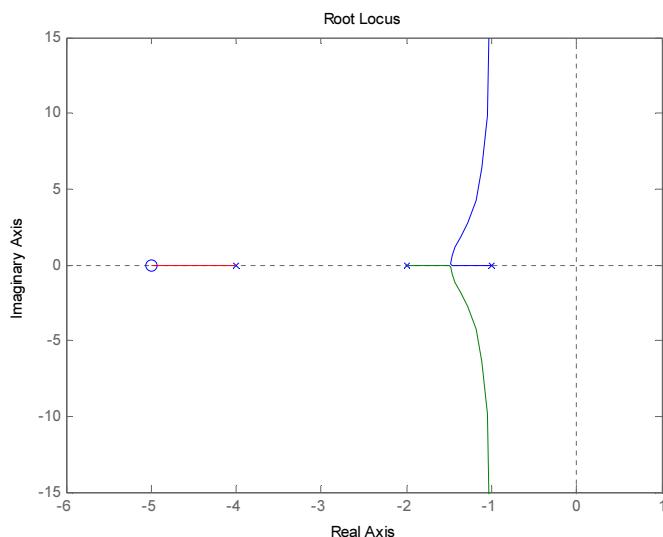
## Root Locus Example in Matlab

```
s = tf('s');
p1=-1; p2=-2; p3=-4;
T = 1 / [ (1-s/p1) * (1-s/p2) * (1-s/p3) ]
rlocus(T)
```



## Adding a Zero

```
s = tf('s');
z=-5; p1=-1; p2=-2; p3=-4;
T = (1-s/z) / [ (1-s/p1) * (1-s/p2) * (1-s/p3) ]
rlocus(T)
```



This example visualizes how introducing a zero in  $T(s)$  can be used to stabilize a feedback amplifier with three poles.

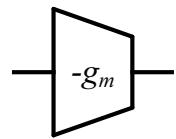
## Minor Issue with Root Locus Plots for Circuit Design

- In a root locus plot, it is implicitly assumed that we can change  $T_0$  without affecting the poles of the amplifier
- This is almost never true for practical circuits
  - Example: Changing the DC gain of a common source stage by increasing its drain resistance also changes its output pole; the gain-bandwidth product remains constant
- In principle, we could fix this issue by re-parameterizing the root locus plot using circuit parameters
  - Example: Plot the root locus as a function of drain resistance (rather than  $T_0$ )
- However, this is not needed, since all we want from the root locus is
  - Gain qualitative insight on the general movement of poles, eye-ball ways to stabilize the amplifier, etc.
  - Gain quantitative insight for one single point in the diagram, e.g. find the conditions for an MFM response
    - This works irrespective of the above-raised issue

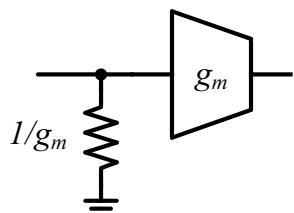
## Overview: Frequency Compensation Techniques

- So far, we have seen two methods for frequency compensation
  - Narrowbanding
  - Feedback zero compensation
- We will now broaden the picture by inspecting a few additional techniques that find their use in practice
  - Miller compensation
  - Ahuja compensation
  - Nested Miller compensation
  - Feedforward compensation
- Many more techniques exist

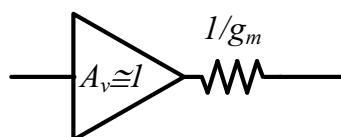
## Building Blocks



Transconductor  
(e.g. CS or CE stage, differential pair)

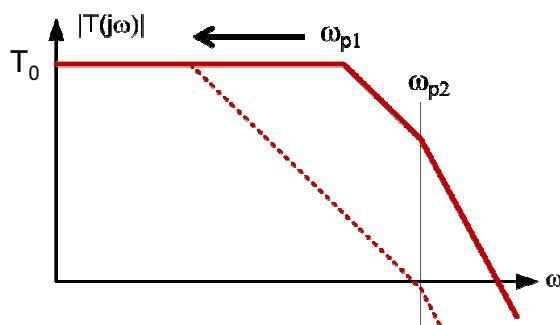


Current Buffer  
(CG or CB stage)



Voltage Buffer  
(CD or CC stage)

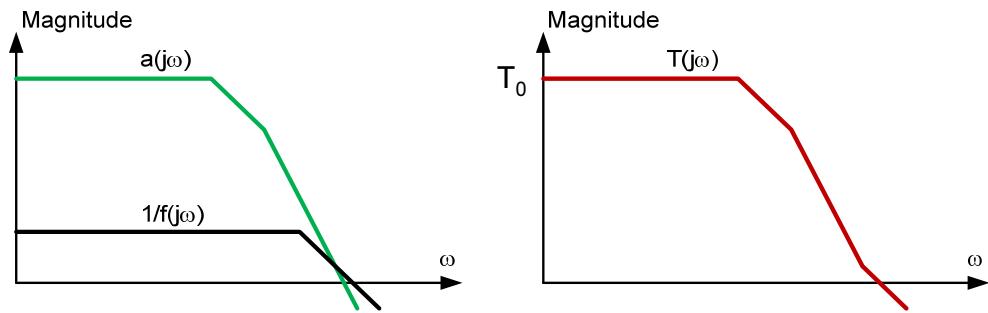
## Re-cap: Narrowbanding Compensation



- Idea

- Make one of the loop poles dominant, leave other poles unchanged

## Re-cap: Feedback Zero Compensation



- Leave amplifier poles unchanged and introduce a zero in the feedback network

## Benchmarking

- In order to compare the merit of various compensation techniques, it makes sense to inspect the loop's unity gain frequency before and after the compensation is applied
- Rationale: The maximum bandwidth we can possibly expect from a feedback amplifier is the unity gain frequency of the loop
  - Regardless of the order of the feedback system, the closed-loop response departs from  $1/f$  as  $|T(s)|$  crosses unity

$$A(s) = \frac{a(s)}{1 + a(s)f(s)} \begin{cases} \cong \frac{1}{f(s)} & \text{for } |T(s)| \gg 1 \\ \cong a(s) & \text{for } |T(s)| \ll 1 \end{cases}$$

## Loop Gain Pole Product (LP Product)

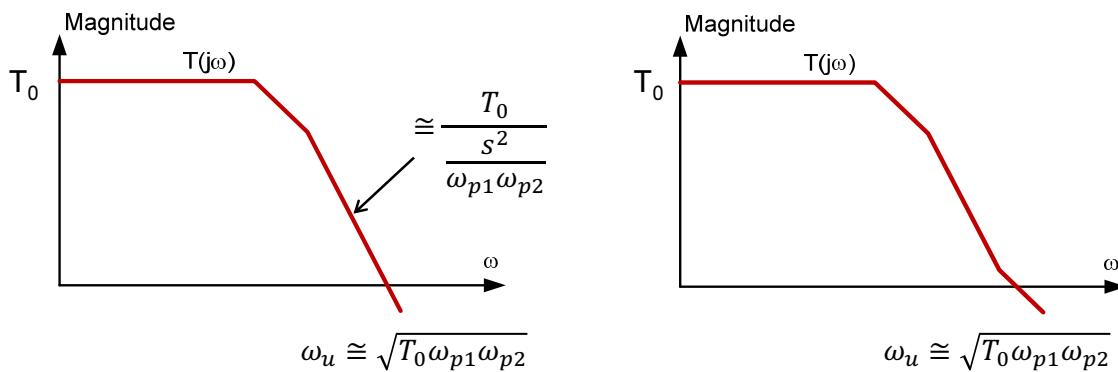
- Consider a loop transfer function with n “dominant” poles that occur before the unity crossover

$$T(s) = \frac{T_0}{(1 - \frac{s}{p_1})(1 - \frac{s}{p_2}) \dots (1 - \frac{s}{p_n})} = \frac{T_0}{1 + \dots + \frac{s^n}{p_1 p_2 \dots p_n}}$$

$$\omega_u \cong \sqrt[n]{T_0 |p_1| |p_2| \dots |p_n|} = \sqrt[n]{LP}$$

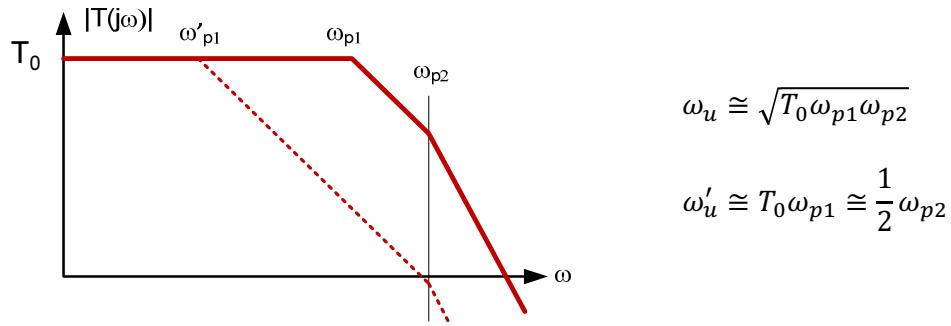
- The product of the low frequency loop gain and all dominant poles is called the loop gain pole product (LP product)
  - Nordholt, Design of High-Performance Negative-Feedback Amplifiers, 1983
- Note that in a first-order system, the LP product is simply the gain-bandwidth product

## Efficiency of Feedback Zero Compensation



- To first order, since we do not change the poles, the LP product and  $\omega_u$  are (approximately) unchanged
  - Feedback zero compensation is therefore bandwidth efficient, since we do not need to sacrifice bandwidth to stabilize the circuit
- To second order, the LP product will change slightly due to loading from the capacitance added in the feedback network

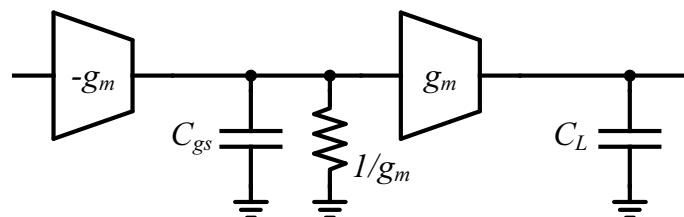
## Efficiency of Narrowbanding



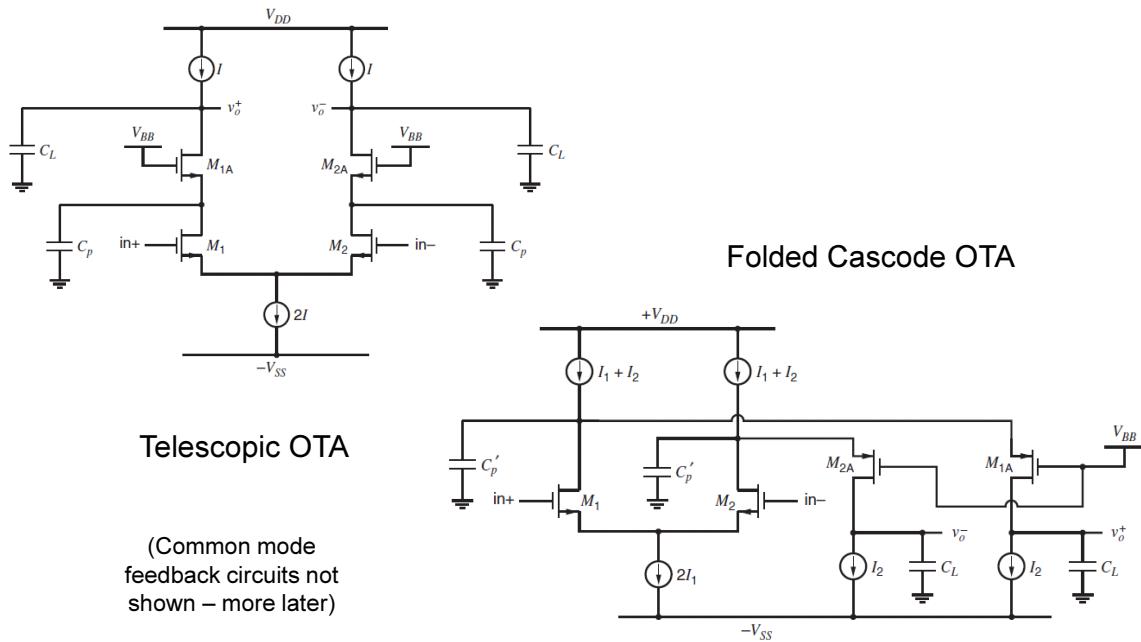
- The crossover and bandwidth is limited to some fraction of  $\omega_{p2}$
- At first glance, this makes narrowbanding appear to be inefficient
- However, provided that  $\omega_{p2}$  is close to the transit frequency of the process technology, this compensation approach is acceptable and hard to surpass in terms of absolute achievable closed-loop speed

## Example: Single Gain Stage with Current Buffer

- Consider the amplifier  $a(s)$  shown below
  - $C_{gs}$  introduces a non-dominant pole at high frequencies  $\omega_{p2} = \omega_T$
  - $C_L$  is adjusted until the circuit achieves the desired phase margin
- This type of narrowbanding is called “load compensation,” since stability is ensured by properly sizing the load capacitance  $C_L$

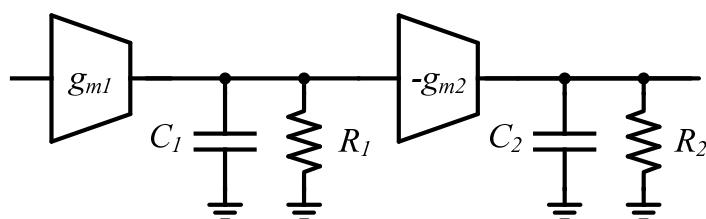


## Example Realizations

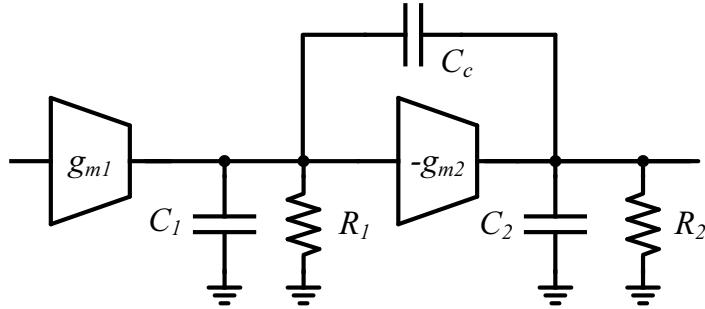


## Two-Stage OTA

- The two-stage amplifier shown below has two comparable poles
  - Assuming that there is no additional significant pole from the feedback network
- If the feedback network is resistive, we may be able to compensate the amplifier by introducing a feedback zero
- What can we do if the feedback is capacitive?
- Narrowbanding is not a good idea, since both poles are at low frequencies



## Miller Compensation



- Purposely connect a capacitor across the second transconductor
- Two interesting things happen
  - Low frequency input capacitance of second stage becomes large – moves the first pole to a lower frequency
  - Qualitatively speaking, at high frequencies,  $C_c$  turns the second stage into a “diode connected device” – low impedance, i.e. large  $\omega_{p2}$

- From the CS/CE analysis of chapter 3, using the dominant pole approximation, we find (after proper variable substitution)

$$p_1 \approx -\frac{1}{R_1[C_1 + C_c(1 + g_{m2}R_2)] + R_2(C_2 + C_c)}$$

$$p_2 \approx -\frac{R_1[C_1 + C_c(1 + g_{m2}R_2)] + R_2(C_2 + C_c)}{R_1R_2(C_1C_2 + C_1C_c + C_2C_c)}$$

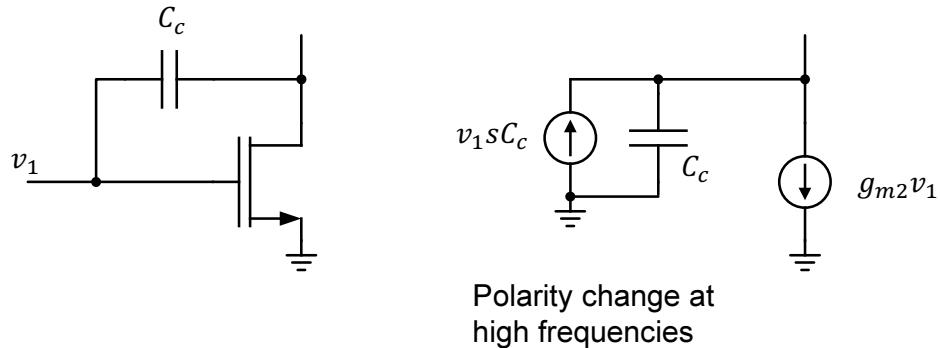
$z = +\frac{g_{m2}}{C_c}$   
RHP zero

- We can approximate further as shown below

$$\omega_{p1} \approx \frac{1}{g_{m2}R_2R_1C_c} \quad a_0\omega_{p1} \approx \frac{g_{m1}R_1g_{m2}R_2}{g_{m2}R_2R_1C_c} = \boxed{\frac{g_{m1}}{C_c}} \quad \text{GBW set by } g_{m1} \text{ and } C_c$$

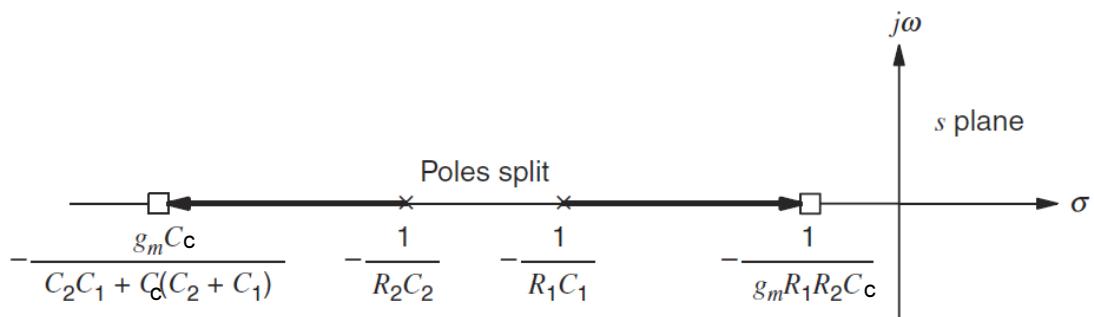
$$\omega_{p2} \approx \frac{g_{m2}C_c}{C_1C_2 + C_c(C_1 + C_2)} \quad \frac{1}{\omega_{p2}} \approx \left( \frac{C_1}{g_{m2}} + \frac{C_2}{g_{m2}} \right) \left( 1 + \frac{\frac{C_1C_2}{C_1 + C_2}}{C_c} \right)$$

## Where does the RHP zero Come From?



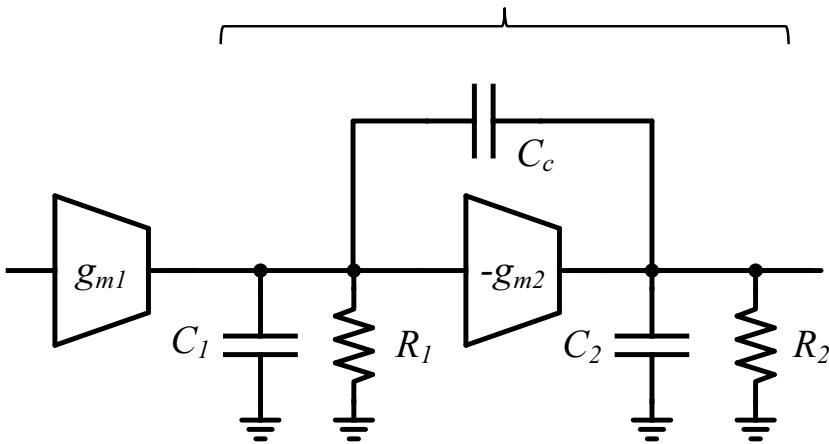
## “Pole Splitting”

- Increasing  $C_c$  reduces  $\omega_{p1}$ , and increases  $\omega_{p2}$ 
  - A very nice “knob” for adjusting the phase margin of the circuit



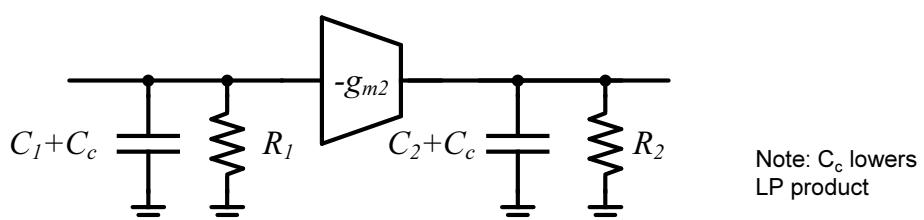
## Feedback Perspective on Miller Compensation

Can analyze this as a feedback circuit



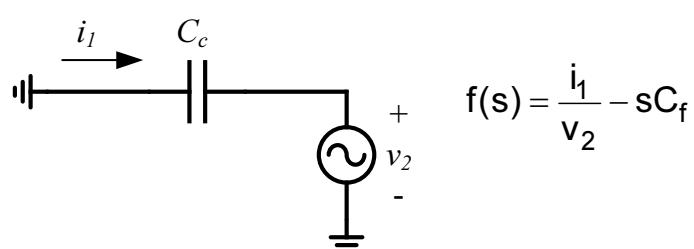
## Shunt-Shunt Feedback Analysis

Forward amplifier model  $a(s)$  has two low frequency poles



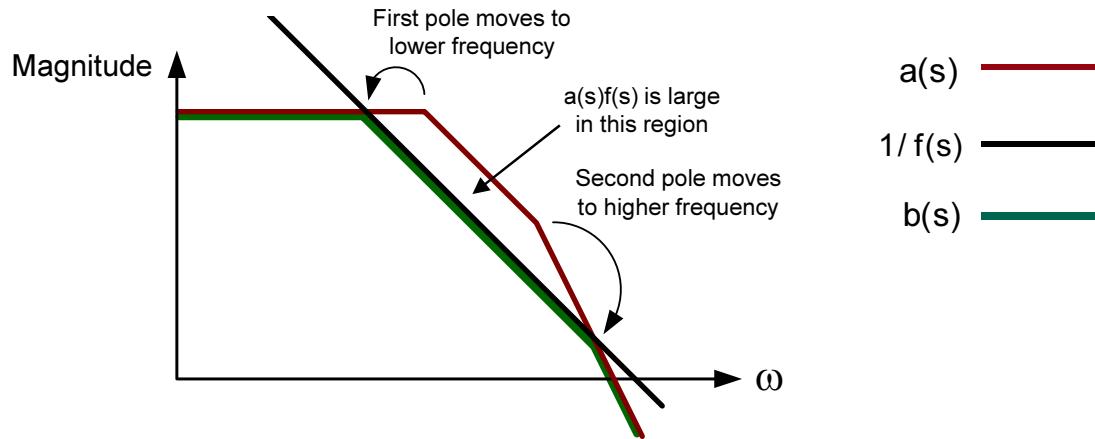
Note:  $C_c$  lowers LP product

Feedback network



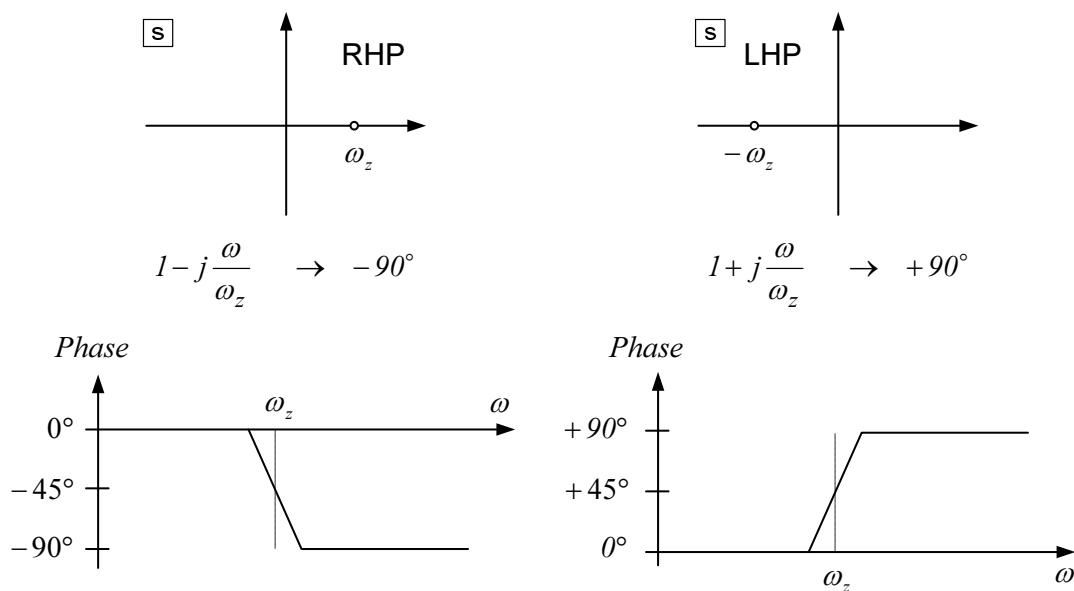
## Graphical View of Pole Splitting

$$b(s) = \frac{a(s)}{1 + a(s)f(s)} \approx \begin{cases} a(s) & \text{when } |a(s)f(s)| \ll 1 \\ \frac{1}{f(s)} & \text{when } |a(s)f(s)| \gg 1 \end{cases}$$

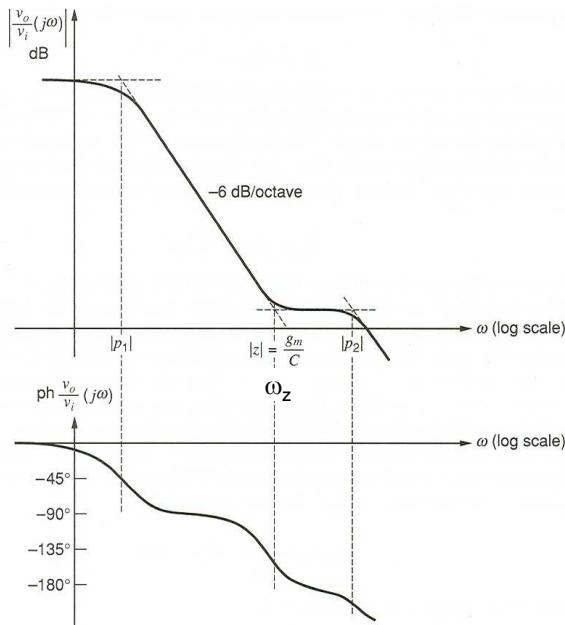


## RHP Zero

- Unfortunately, the right half plane zero due to  $C_c$  can destroy the PM



## Issue with RHP Zero

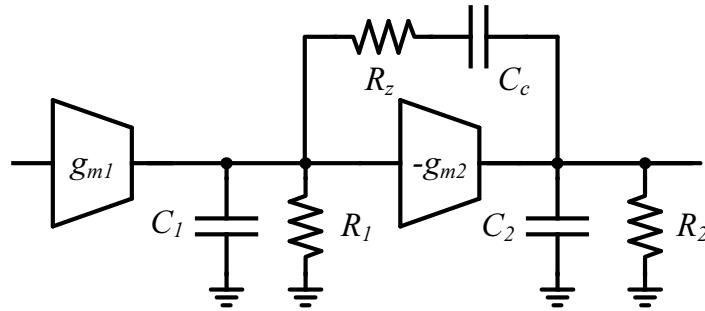


- RHP zero can destroy the phase margin if it occurs before or near the crossover frequency

## Mitigating the Impact of RHP Zero

- Somehow create “unilateral” feedback through  $C_c$ 
  - Source follower from output to drive  $C_c$ 
    - Additional power & swing reduction issues
  - Introduce a nulling resistor
    - Most popular approach
  - Ahuja compensation (also called cascode compensation)
    - Ahuja, IEEE JSSC, 12/1983
    - Ribner, IEEE JSSC, 12/1984

## Nulling Resistor

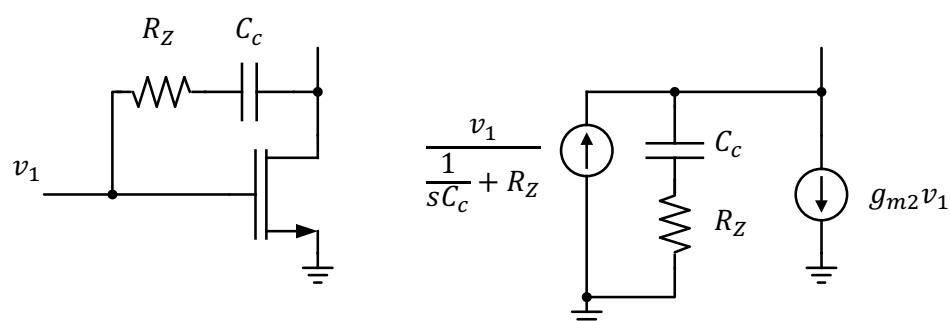


The new transfer function becomes

$$a(s) \approx a_{v0} \cdot \frac{1 - sC_c \left( \frac{1}{g_{m2}} - R_z \right)}{\left( 1 - \frac{s}{p_1} \right) \cdot \left( 1 - \frac{s}{p_2} \right) \cdot \left( 1 - \frac{s}{p_3} \right)}$$

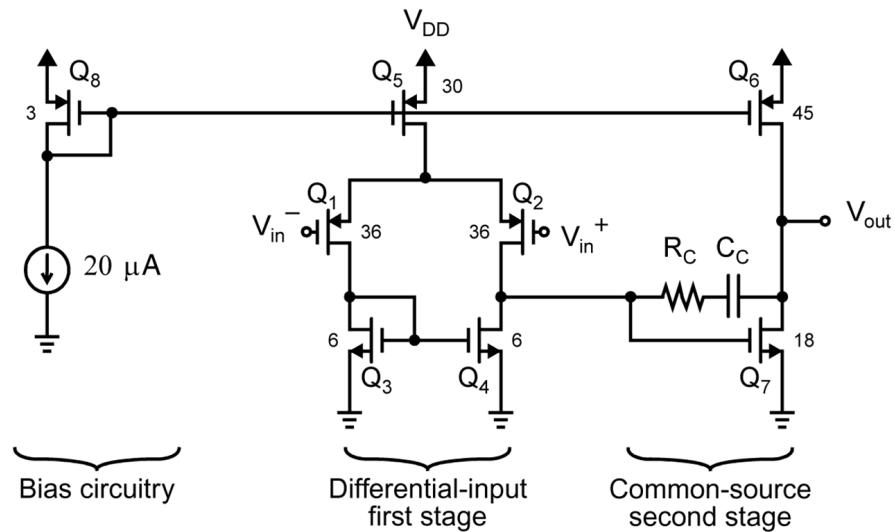
- $p_1$  and  $p_2$  unchanged, new pole  $p_3$ , and a “knob” to tune the zero

## Nulling Resistor

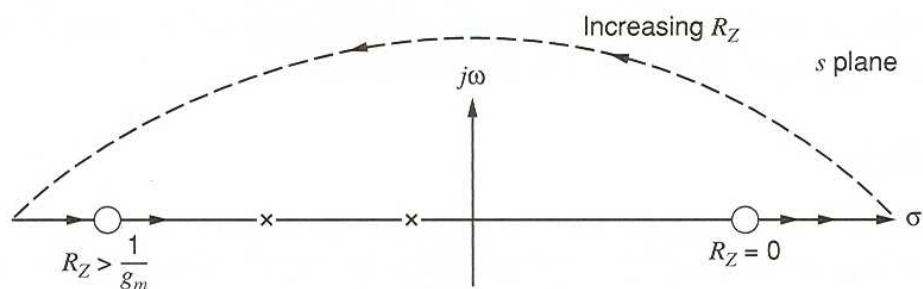


Polarity cannot  
change if  $R_z \geq 1/g_{m2}$

## Implementation Example



Text, page 243



- $R_z = 1/g_{m2}$  pushes the zero to  $+\infty$
- $R_z \approx (1+C_2/C_c)/g_{m2}$  places the zero such that it cancels  $p_2$ !

## Third Pole

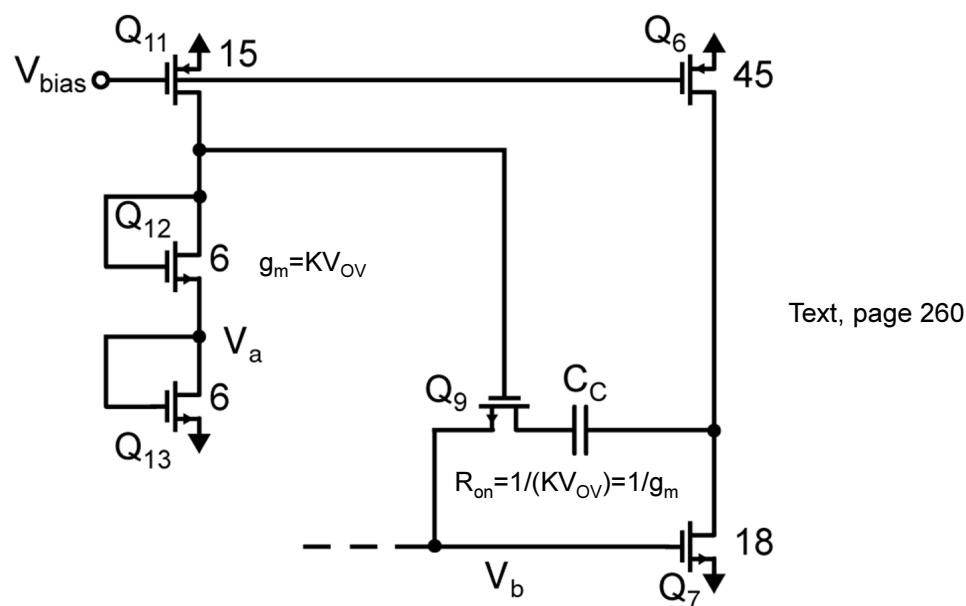
$$\omega_{p3} \cong \frac{1}{R_z C_1}$$

- For  $R_z = 1/g_{m2}$   $\omega_{p3} \cong \frac{g_{m2}}{C_1} \cong \omega_T$

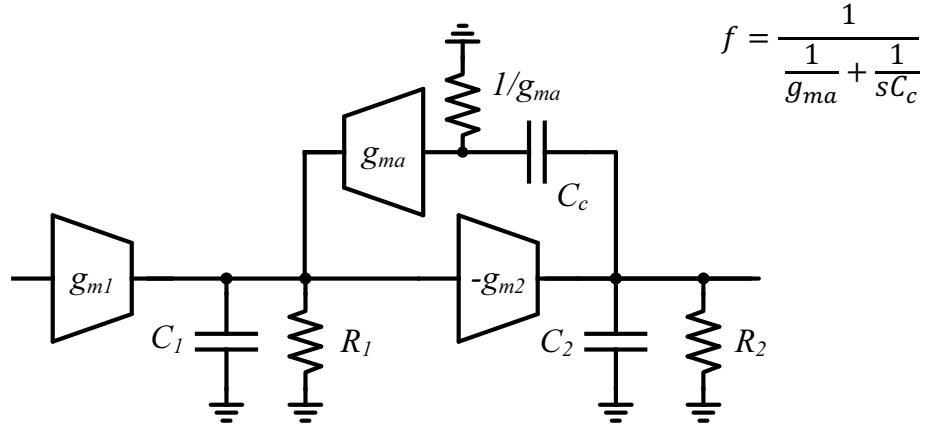
- For  $R_z = (1+C_2/C_c)/g_{m2}$   $\omega_{p3} \cong \frac{g_{m2}}{\left(1 + \frac{C_2}{C_c}\right)C_1} \cong \frac{\omega_T}{\left(1 + \frac{C_2}{C_c}\right)}$

- Thus, as we try to cancel the second pole, the third pole moves to a lower frequency, and may move to a frequency that is comparable to the original  $\omega_{p2}$  before cancellation
- My recommendation: Simply push the zero to infinity
- The textbook's recommendation: Spice-monkey the zero into the LHP and try to squeeze out some phase margin

## Process Insensitive Implementation of $R_z$

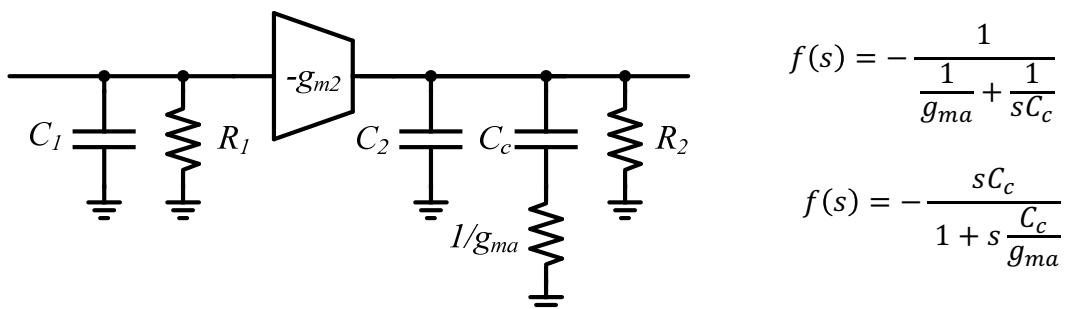


## Ahuja Compensation



- Idea: insert a current buffer (instead of  $R_z$ ) to obtain unilateral feedback
- This removes the feedforward zero and leads to a smaller LP product degradation than in plain Miller compensation
  - We can see this from the two-port model of the inner loop

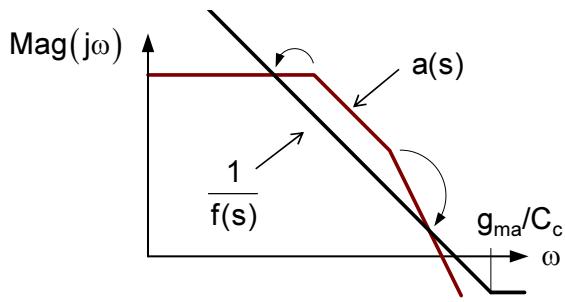
## Two-Port Model of Inner Loop



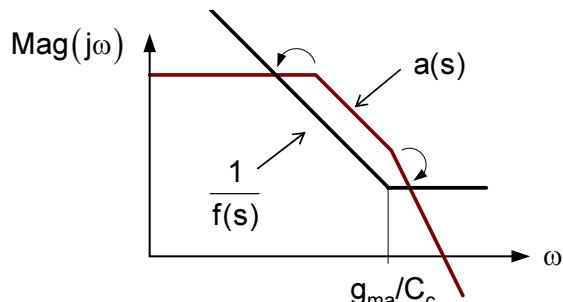
- Good news:  $C_c$  does not contribute extra loading at the input
  - Less LP product degradation than Miller compensation
- Bad news:  $f(s)$  of the inner loop contains a pole
  - Bad for stability of the inner loop

## Possible Scenarios

Pole in  $f(s)$  occurs after crossover of inner loop:



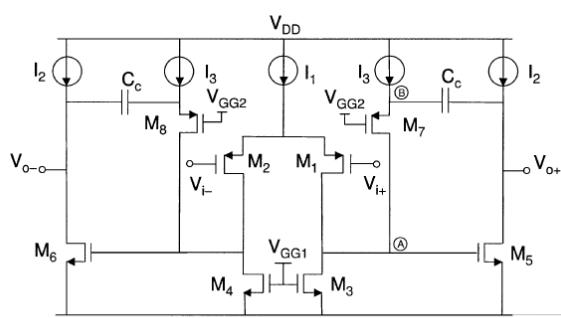
Pole in  $f(s)$  occurs before crossover of inner loop:



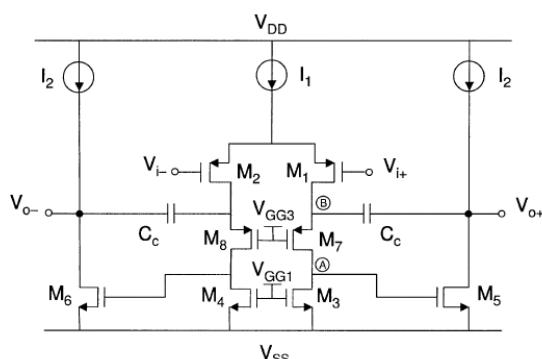
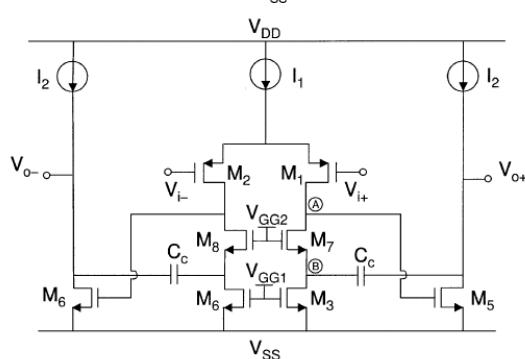
Inner loop has no phase margin!  
Expect peaking, complex poles in the amplifier  
even before outer feedback is applied

- Ahuja compensation is somewhat harder to design, but can bring significant benefits when used properly

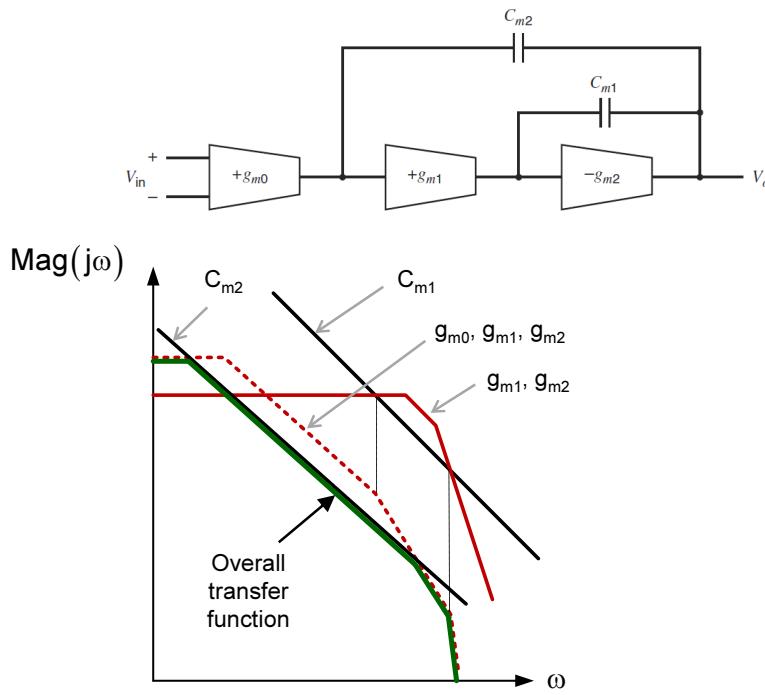
## Implementation Options



P.J. Hurst et al. "Miller compensation using current buffers in fully differential CMOS two-stage operational amplifiers," *IEEE Transactions on Circuits and Systems I*, vol.51, no.2, pp. 275-285, Feb. 2004.



## Nested Miller Compensation

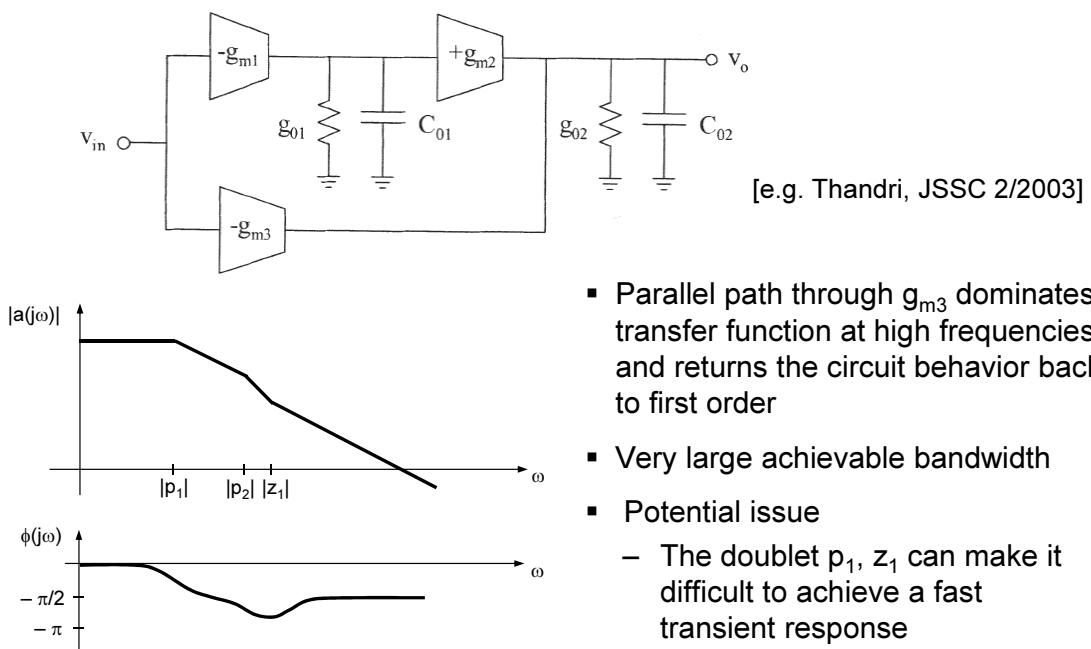


B. Murmann

EE214B Winter 2015-16 – Chapter 9

57

## Feedforward Compensation



B. Murmann

EE214B Winter 2015-16 – Chapter 9

58

## Issues with Pole-Zero Doublet (1)

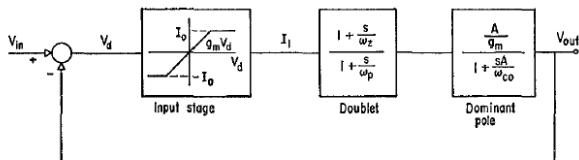
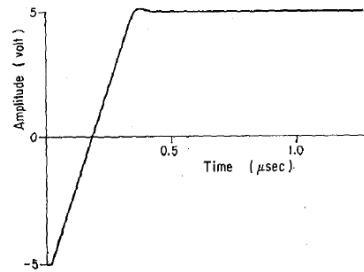


Fig. 1. Operational amplifier model for calculation of transient response.



$$V_{\text{out}}(t) \simeq V(1 - k_1 \exp[-\omega_{co}t] + k_2 \exp[-(t/\tau_2)]), \quad \text{for } t > T, \quad (1)$$

where

$$k_2 \simeq \frac{\omega_z - \omega_p}{\omega_{co}} \quad (2)$$

$$\tau_2 \simeq \frac{1}{\omega_z} \quad (3)$$

$T_s$  slewing period

$\omega_z$  doublet zero frequency

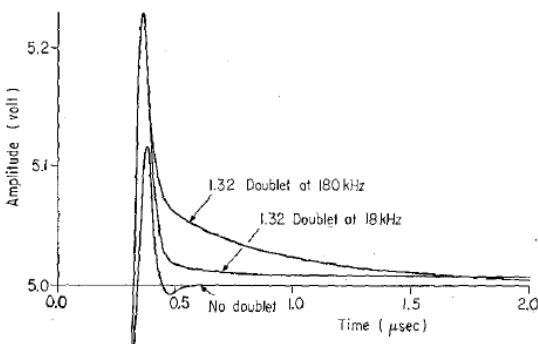
$\omega_p$  doublet pole frequency

$\omega_{co}$   $A \times (\text{amplifier dominant pole}) \simeq$  unity-gain bandwidth

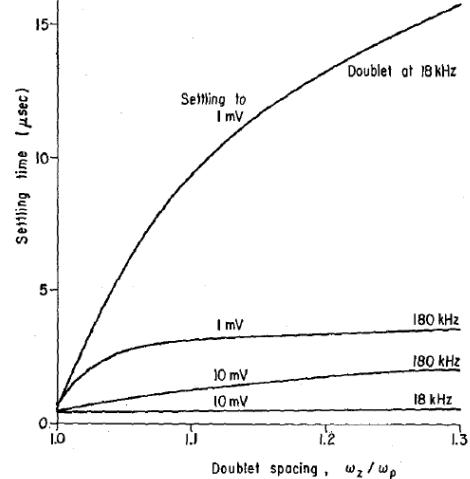
$A$  open-loop low frequency gain.

B.Y.T. Kamath, R.G. Meyer and P.R. Gray, "Relationship between frequency response and settling time of operational amplifiers," IEEE JSSC, Vol. 9, No. 6, pp.347–352, Dec. 1974.

## Issues with Pole-Zero Doublet (2)

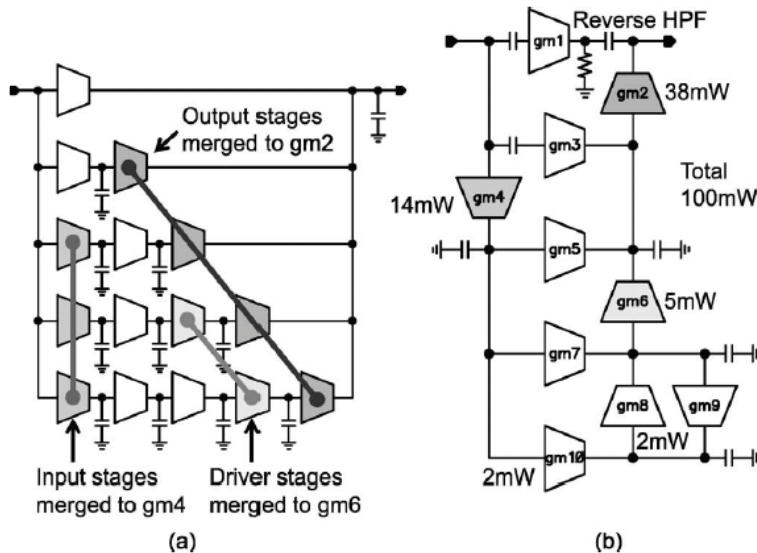


For fast and accurate settling, need either small pole-zero spacing or large  $w_z$



B.Y.T. Kamath, R.G. Meyer and P.R. Gray, "Relationship between frequency response and settling time of operational amplifiers," IEEE JSSC, Vol. 9, No. 6, pp.347–352, Dec. 1974.

## Recent Example of a Feedforward OTA



[Shibata et al., JSSC 12/2012]

## Summary

Technique	Internal	External	Lead	Lag	Comments
Narrowbanding	X	X		X	Practical mostly in single-stage amplifiers
Feedback zero			X	X	Preferred method whenever applicable; highest bandwidth efficiency
Miller	X			X	Easy to design and robust when zero is moved to infinity
Ahuja	X			X	Hard to design, but potentially large gains in achievable bandwidth
Feedforward	X		X		Do not use in circuits that demand fast and precise transient settling

# Chapter 10

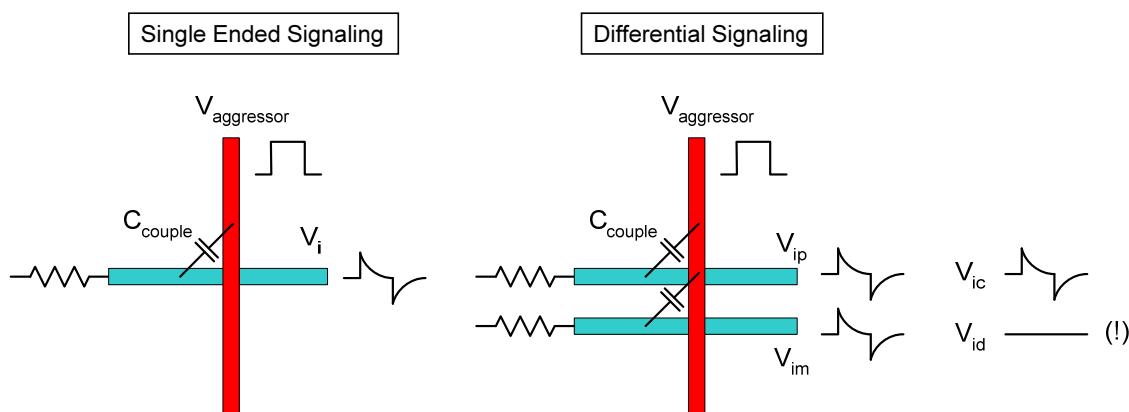
## Fully Differential Amplifiers

Boris Murmann  
Stanford University  
Winter 2015-16

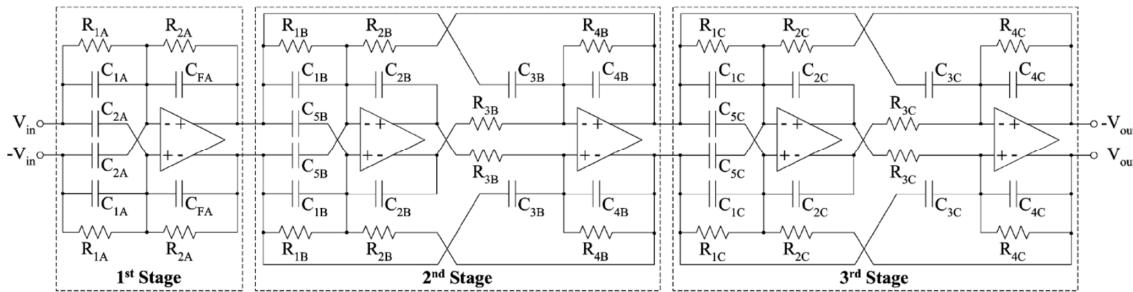
Textbook Sections: 6.7, 6.8

### Motivation

- Differential signaling is used in a variety of integrated circuit building blocks to reject several forms of coupling “noise”
  - Power supply noise, substrate noise, wire coupling noise, etc.

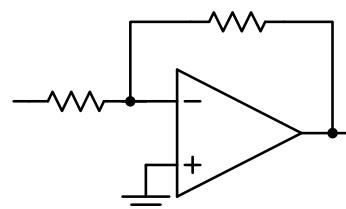
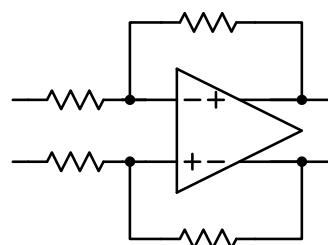


## Example: Lowpass Filter



A. Vasilopoulos, G. Vitzilaios, G. Theodoratos, and Y. Papananos, "A Low-Power Wideband Reconfigurable Integrated Active-RC Filter With 73 dB SFDR," IEEE J. Solid-State Circuits, pp. 1997-2008, Sept. 2006

## Amplifiers with Differential Outputs vs. Single Ended Outputs



- Symmetrical
  - Immune to coupling and power supply noise
  - Easy to analyze
- Can invert signal via wire crossing
- Requires common mode feedback (CMFB)
- Lower complexity
  - Important for PCB design
- Can build non-inverting unity gain buffer without any feedback components

## Basic Example and Review of Voltage Decomposition

Common-Mode Voltages

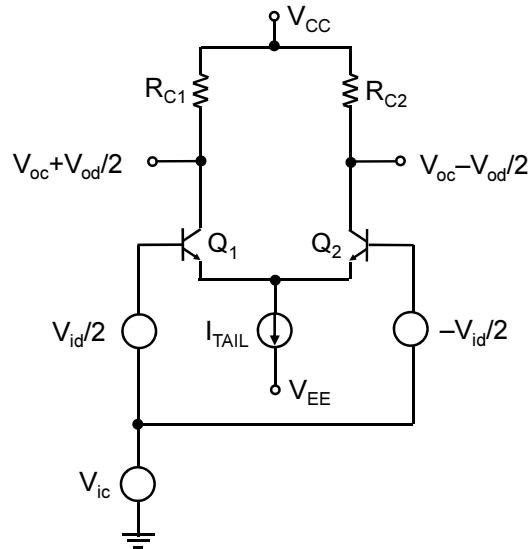
$$V_{ic} = \frac{1}{2}(V_{i1} + V_{i2})$$

$$V_{oc} = \frac{1}{2}(V_{o1} + V_{o2})$$

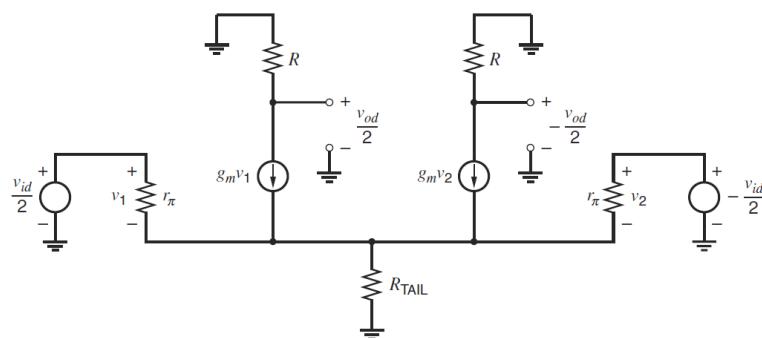
Inputs  $V_{i1}$  and  $V_{i2}$  are decomposed into a combination of differential- and common-mode voltage sources

$$V_{i1} = V_{ic} + \frac{1}{2}V_{id}$$

$$V_{i2} = V_{ic} - \frac{1}{2}V_{id}$$



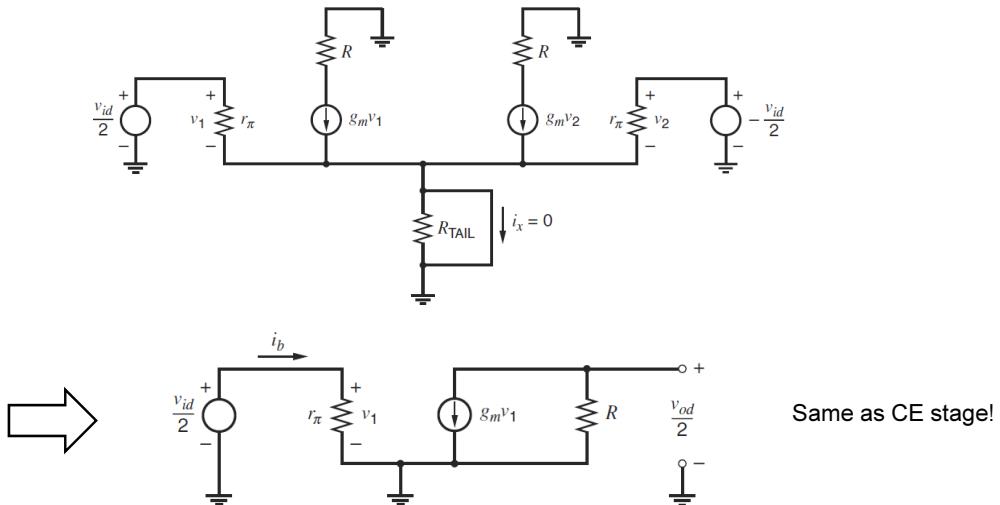
### Small Signal Model for $v_{ic} = 0$



- Define differential mode gain as

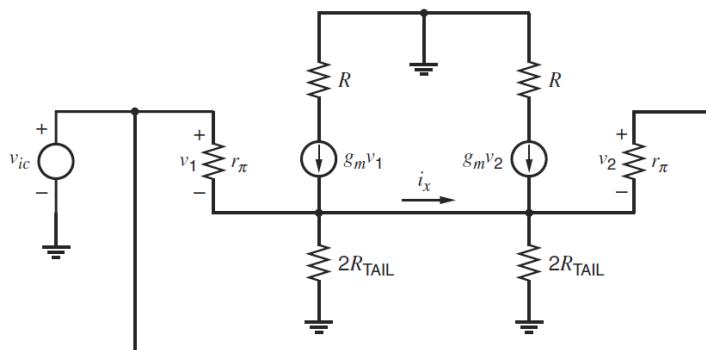
$$A_{dm} = \left. \frac{V_{od}}{V_{id}} \right|_{V_{ic}=0}$$

## Differential Mode Half Circuit



$$\frac{v_{od}}{2} = -g_m R \frac{v_{id}}{2} \quad A_{dm} = \frac{v_{od}}{v_{id}} = -g_m R$$

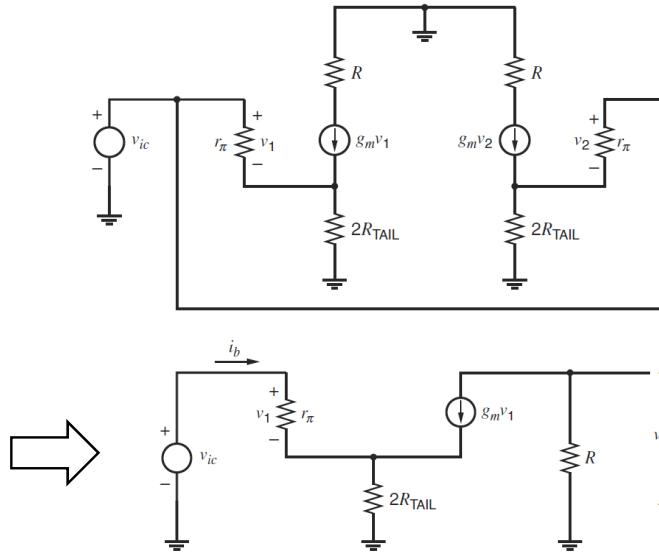
## Small Signal Model for $v_{id} = 0$



- Define common mode gain as

$$A_{cm} = \left. \frac{V_{oc}}{V_{ic}} \right|_{v_{id}=0}$$

## Common Mode Half Circuit



$$v_{oc} = -G_m R v_{ic} \quad A_{cm} = \frac{v_{oc}}{v_{ic}} = -G_m R = -\frac{g_m R}{1 + 2g_m R_{TAIL}}$$

## Interaction of Common Mode and Differential Mode

$$A_{cdm} = \left. \frac{V_{od}}{V_{ic}} \right|_{V_{id}=0} \quad \text{and} \quad A_{dcm} = \left. \frac{V_{oc}}{V_{id}} \right|_{V_{ic}=0}$$

- In a perfectly balanced (symmetric) circuit,  $A_{cdm} = A_{dcm} = 0$
- In practice,  $A_{cdm}$  and  $A_{dcm}$  are not zero because of component mismatch
- $A_{cdm}$  is important because it indicates the extent to which a common-mode input will corrupt the differential output (which contains the actual signal information)
- The calculation of  $A_{cdm}$  is often algebraically complex (see Hurst, Lewis, Gray & Meyer, chapter 12), and best determined via simulation (with proper mismatch/imbalance included)

## Common-Mode Rejection Ratio (CMRR)

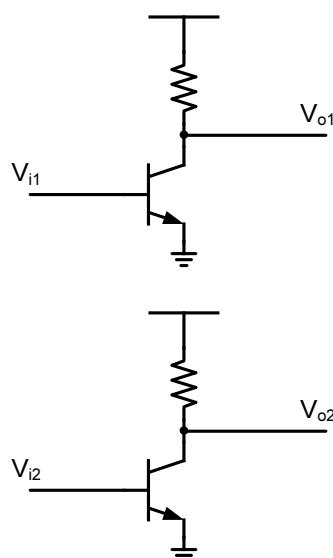
For fully differential amplifiers, the CMRR is defined as

$$\text{CMRR} = \left| \frac{A_{dm}}{A_{cdm}} \right|$$

This is different from the “textbook” definition for amplifiers with differential inputs and single-ended outputs

$$\text{CMRR}_{SE} = \left| \frac{A_{dm}}{A_{cm}} \right|$$

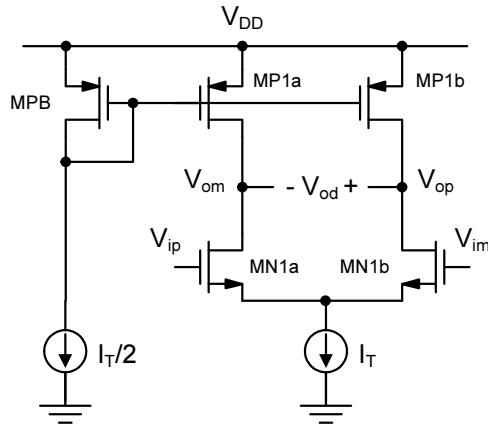
## Pseudo-Differential Amplifier



$$A_{cm} = A_{dm} = -g_m R$$

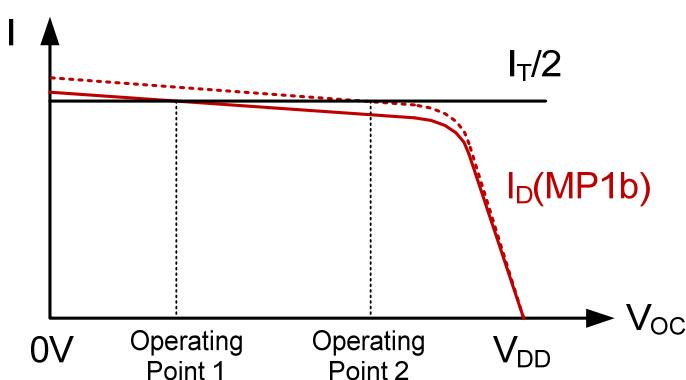
- Difficult to cascade such stages
- Common mode gets amplified (instead of being rejected)

## Differential Amplifier with Active Load



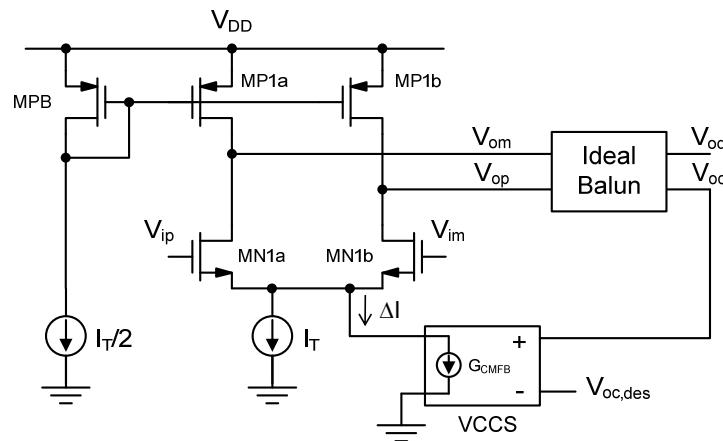
- Suppose that at the operating point  $V_{ip}=V_{im}$ , i.e.  $V_{id}=0$
- What is the output common mode voltage  $V_{oc} = (V_{om}+V_{op})/2$  ?

## Operating Point Sensitivity



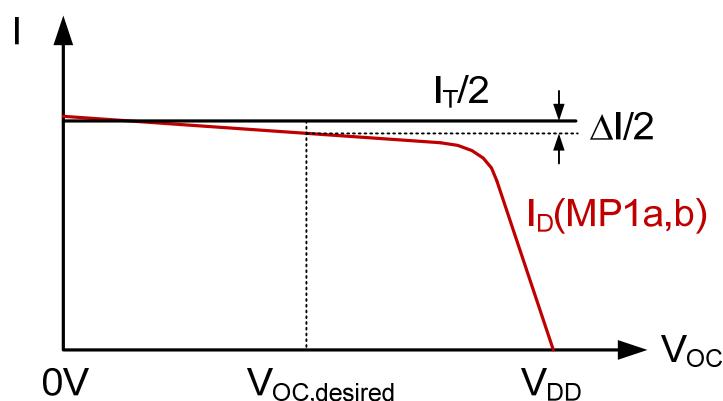
- The operating point is very sensitive to small changes in the device characteristics
- Solution: Common mode feedback (CMFB)

## Idealized CMFB Implementation



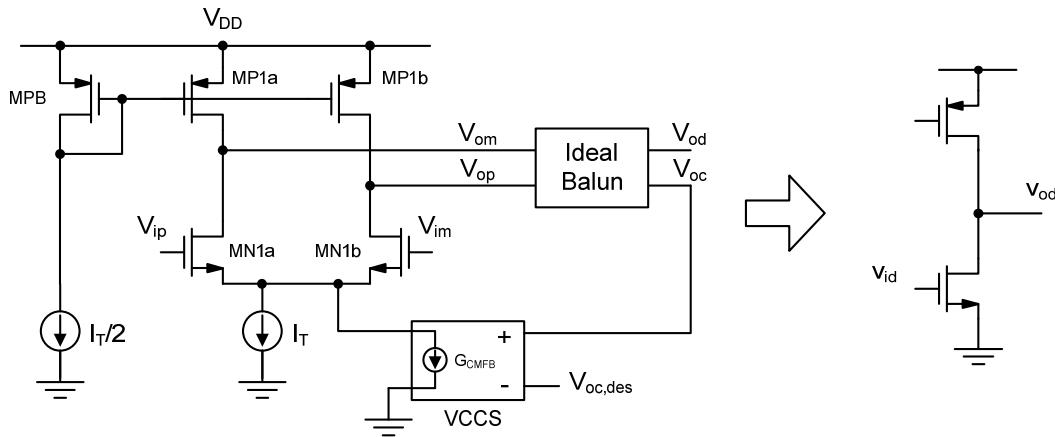
$$V_{oc} = V_{oc,des} + \frac{\Delta I}{G_{CMFB}} = V_{oc,des} \quad \text{for} \quad G_{CMFB} \rightarrow \infty$$

## CMFB Operation



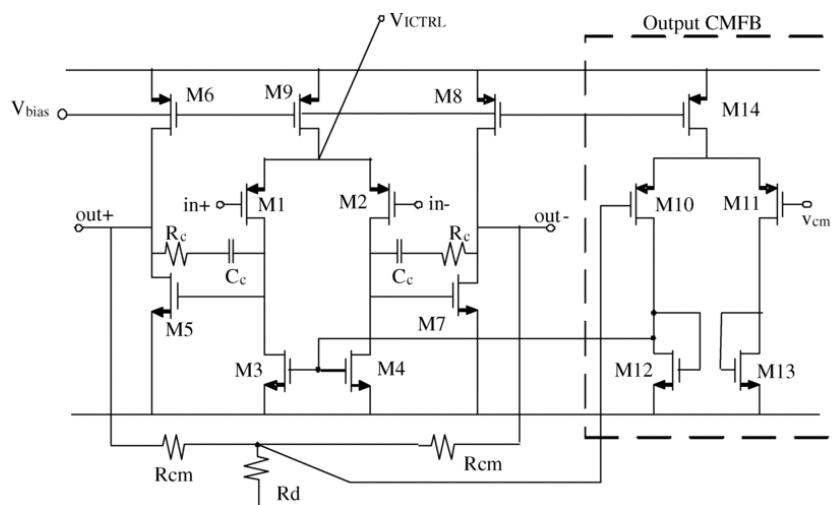
- The common mode feedback loop computes  $\Delta I$  such that  $V_{oc}$  is very close to the desired voltage

## Differential Mode Small-Signal Half Circuit



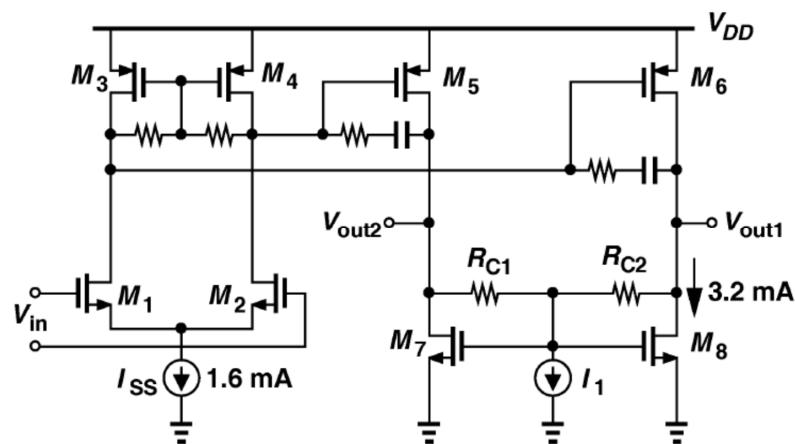
- With the circuit biases at the proper operating point, we can analyze its small-signal behavior using a differential mode half circuit model
- Note that (to first order) the CMFB loop does not influence the behavior of the differential mode signals

## CMFB Implementation Example (1)



M. De Matteis, S. D'Amico, and A. Baschirotto, "A 0.55 V 60 dB-DR Fourth-Order Analog Baseband Filter," IEEE J. Solid-State Circuits, pp. 2525-2534, Sept. 2009.

## CMFB Implementation Example (2)



A. Verma and B. Razavi, "A 10-Bit 500-MS/s 55-mW CMOS ADC," IEEE J. Solid-State Circuits, pp. 3039-3050, Nov. 2009.

# **Chapter 11**

## **OTA Design Examples**

**Boris Murmann**  
**Stanford University**  
**Winter 2015-16**

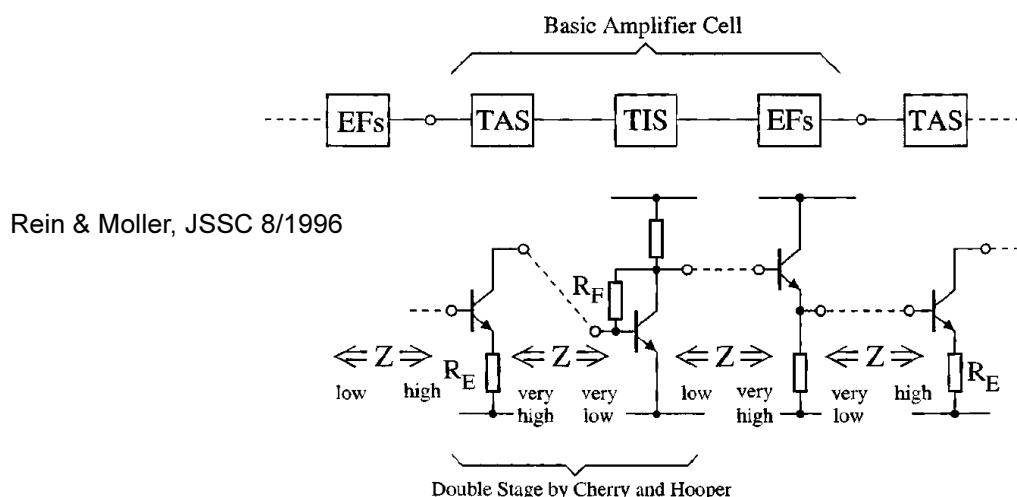
**See Supplementary Handout**

# Chapter 12

## Wideband Amplifiers Using Local Feedback

Boris Murmann  
Stanford University  
Winter 2015-16

### Wideband Amplifier Design with Local Feedback



In a cascade of amplifier stages, the interaction (loading) between adjacent stages can be minimized by alternating local series and shunt feedback circuits.

## Advantages & Disadvantages of Local Feedback Cascades

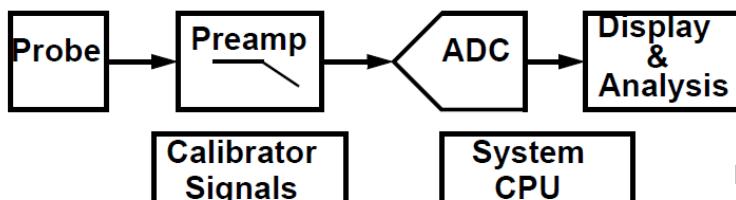
### Advantages

- No instability
- Lower gain sensitivity to component and device parameter variations than amplifiers without local feedback

### Disadvantages

- Higher gain sensitivity (less loop gain) than amplifiers with multi-stage feedback
- Bandwidth obtainable is typically slightly less than that possible in amplifiers with multi-stage feedback

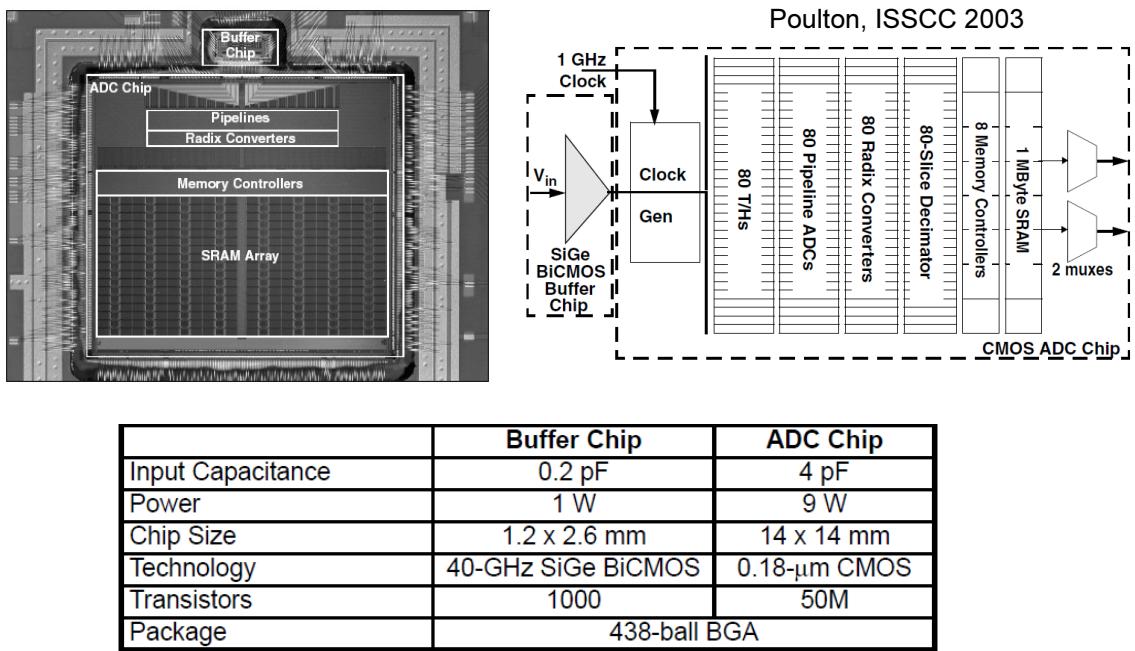
## Application Example (1)



Poulton, ISSCC 2003

Simplified oscilloscope block diagram

## Application Example (2)

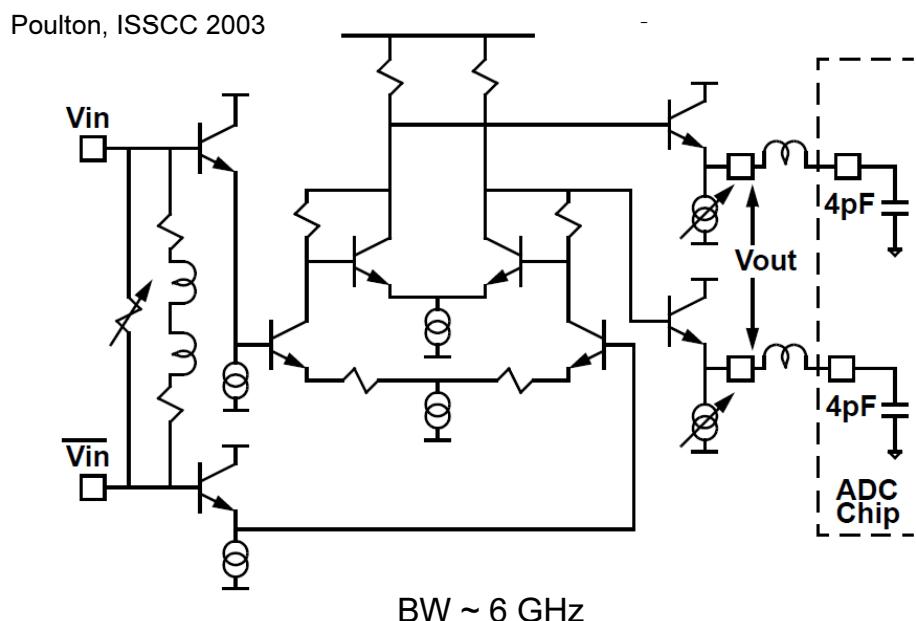


B. Murmann

EE214B Winter 2015-16 – Chapter 12

5

## Application Example (3)

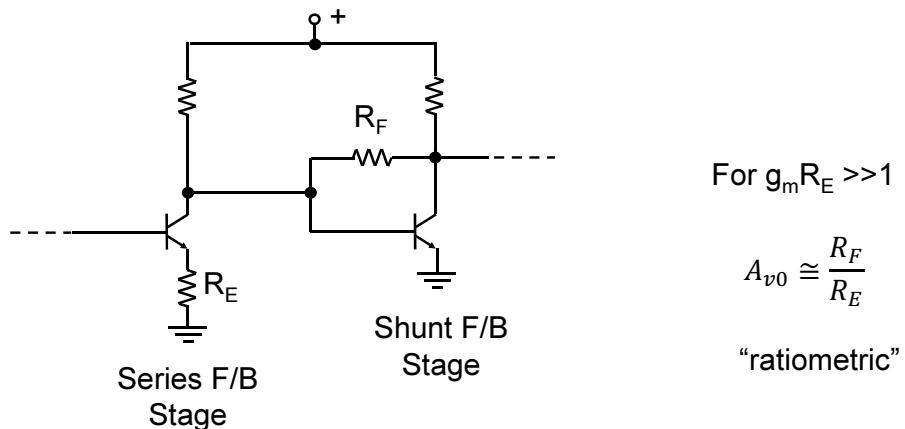


B. Murmann

EE214B Winter 2015-16 – Chapter 12

6

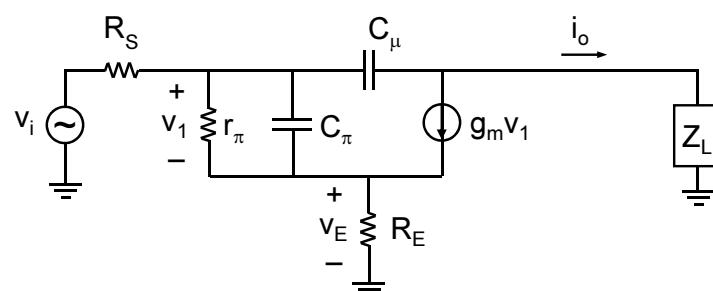
## Cherry-Hooper Amplifier



E. Cherry and D. Hooper, Proc. IEE, Feb. 1963

### Series Feedback Stage

Small-signal equivalent circuit:



Include transistor  $r_b$  in  $R_S$

Neglect  $r_\mu$ ,  $r_o$ ,  $r_e$ , and  $r_c$

Assume that  $Z_L \approx 0$ ; then there is no Miller effect and  $C_\mu$  is just a (small) capacitance to ground which often has negligible impact.

Define

$$Y_\pi = \frac{1}{r_\pi} + sC_\pi$$

Then

$$\frac{v_i - (v_1 + v_E)}{R_S} = v_1 Y_\pi$$

$$v_1 Y_\pi + g_m v_1 = \frac{v_E}{R_E}$$

Substituting for  $v_E$  in the first of these two equations

$$v_i - v_1 [1 + (g_m + Y_\pi) R_E] = v_1 Y_\pi R_S$$

$$\therefore \frac{v_1}{v_i} = \frac{1}{1 + (g_m + Y_\pi) R_E + Y_\pi R_S}$$

Since  $i_o = -g_m v_1$

$$\begin{aligned} \frac{i_o}{v_i} &= -\frac{g_m}{1 + g_m R_E + (R_S + R_E) Y_\pi} \\ &= -\left(\frac{g_m}{1 + g_m R_E}\right) \left[ \frac{1}{1 + \left(\frac{R_S + R_E}{1 + g_m R_E}\right) \left( \frac{1}{r_\pi} + sC_\pi \right)} \right] \\ &= -\left(\frac{g_m}{1 + g_m R_E}\right) \left[ \frac{1}{1 + \left(\frac{R_S + R_E}{r_\pi}\right) \left( \frac{1}{1 + g_m R_E} \right) + sC_\pi \left( \frac{R_S + R_E}{1 + g_m R_E} \right)} \right] \end{aligned}$$

Usually,

$$\left( \frac{R_S + R_E}{r_\pi} \right) \left( \frac{1}{1 + g_m R_E} \right) = \frac{g_m (R_S + R_E)}{\beta_0} \left( \frac{1}{1 + g_m R_E} \right) \ll 1$$

Then

$$\frac{i_o}{v_i} \cong -g_{\text{meq}} \left( \frac{1}{1 - s/p_1} \right)$$

where

$$g_{\text{meq}} = \frac{g_m}{1 + g_m R_E}$$

$$p_1 = - \left( \frac{1 + g_m R_E}{C_\pi} \right) \left( \frac{1}{R_E + R_S} \right)$$

If  $g_m R_E \gg 1$

$$g_{\text{meq}} \cong \frac{1}{R_E}$$

and

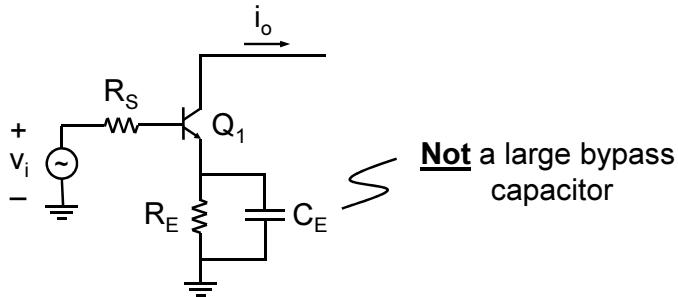
$$p_1 \cong - \frac{g_m}{C_\pi} \left( \frac{R_E}{R_E + R_S} \right) = -\omega_T \left( \frac{C_\pi + C_\mu}{C_\pi} \right) \left( \frac{R_E}{R_E + R_S} \right) \cong -\omega_T \left( \frac{R_E}{R_E + R_S} \right)$$

Note that this result corresponds to the dominant time constant found via ZVTC analysis in chapter 3.

Thus, for  $R_E \gg R_S$ ,  $p_1 \rightarrow -\omega_T$ . In some cases, the bandwidth may then be actually limited by  $C_\mu$  (which was neglected in this analysis).

## Emitter Peaking

The bandwidth of the series feedback stage can be increased by introducing an “emitter peaking” capacitor in shunt with  $R_E$ .



To analyze this circuit, substitute  $Z_E$  for  $R_E$  in the preceding analysis, where

$$Z_E = \frac{1}{Y_E} = \frac{R_E}{1 + sR_E C_E}$$

Then, defining  $\tau_E = R_E C_E$  and assuming  $\tau_T \approx \frac{C_\pi}{g_m}$

$$\begin{aligned} \frac{i_o}{v_i} &= -\frac{g_m}{1 + g_m Z_E + (R_S + Z_E) \left( \frac{1}{r_\pi} + s C_\pi \right)} \\ &= -\frac{g_m}{1 + \frac{R_S}{r_\pi} + s C_\pi R_S + \left( \frac{R_E}{1 + s \tau_E} \right) \left( g_m + \frac{1}{r_\pi} + s C_\pi \right)} \\ &\approx -\frac{g_m}{1 + \frac{R_S}{r_\pi} + s C_\pi R_S + \left( \frac{g_m R_E}{1 + s \tau_E} \right) \left( 1 + s \frac{C_\pi}{g_m} \right)} \end{aligned}$$

$$\text{since } g_m \gg \frac{1}{r_\pi}$$

$$\therefore \frac{i_o}{v_i} \cong -\frac{g_m}{1 + \frac{R_s}{r_\pi} + sC_\pi R_s + g_m R_E \left( \frac{1+s\tau_T}{1+s\tau_E} \right)}$$

If  $C_E$  is chosen so that

$$\tau_E = \tau_T$$

and it is true that

$$g_m R_E \gg \frac{R_s}{r_\pi} = \frac{g_m R_s}{\beta_0}$$

Then

$$\frac{i_o}{v_i} \cong -\left( \frac{g_m}{1 + g_m R_E} \right) \left[ \frac{1}{1 + sC_\pi \left( \frac{R_s}{1 + g_m R_E} \right)} \right]$$

The result is a single-pole response with

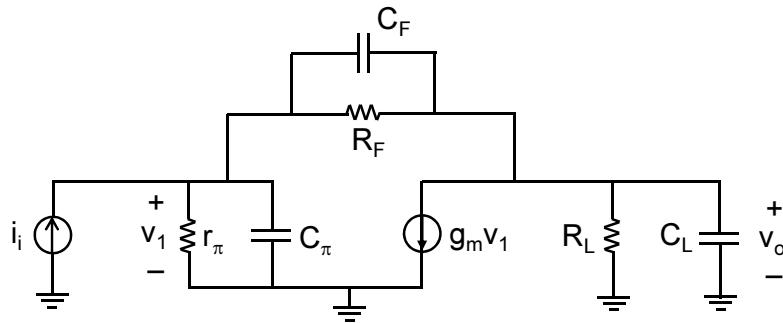
$$p_1 = -\frac{1 + g_m R_E}{R_s C_\pi} \cong -\omega_T \left( \frac{R_E}{R_s} \right)$$

In this case, if  $R_E > R_s$ , then  $|p_1| > \omega_T$  (!)

The bottom line is that this stage has no significant bandwidth limitations; the frequency response of the overall Cherry-Hooper amplifier will mostly depend on the second stage.

## Shunt Feedback Stage

Small-signal equivalent circuit



Include transistor  $C_\mu$  in  $C_F$

Neglect  $R_S$  or include it in  $r_\pi$

Neglect  $r_b$ ,  $r_c$ ,  $r_e$ , and  $r_\mu$

Include  $r_o$  and  $C_{cs}$  in  $R_L$  and  $C_L$

## Analysis (1)

We can analyze this circuit from first principles using KCL and then simplify to get to a low-entropy result. However, it is much faster to look at the circuit using a two-port feedback approach, and use our understanding from chapter 8.

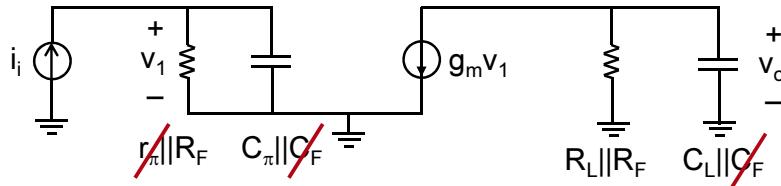
$$a(s) = \frac{a_0}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)} \quad f(s) = f_0 \left(1 + \frac{s}{\omega_z}\right)$$

$$A(s) = \frac{A_0}{1 + \frac{s}{\omega_0 Q} + \frac{s^2}{\omega_0^2}} \quad \omega_0 = \sqrt{(1 + T_0)\omega_{p1}\omega_{p2}} \cong \omega_u$$

$$\text{For a maximally flat response (Q=0.707), set: } \omega_z = \frac{1}{R_F C_F} \cong \frac{\omega_u}{\sqrt{2}}$$

## Analysis (2)

Model of  $a(s)$ :



$$a_0 = \frac{v_o}{i_i} = -R_F g_m (R_L || R_F) \quad \omega_{p1} = \frac{1}{R_F C_\pi} \quad \omega_{p2} = \frac{1}{(R_L || R_F) C_L}$$

$$\omega_u \cong \sqrt{T_0 \omega_{p1} \omega_{p2}} = \sqrt{\frac{g_m}{C_L} \cdot \frac{1}{R_F C_\pi}} \cong \sqrt{\omega_T \cdot \frac{1}{R_F C_L}}$$

## Analysis (3)

Now, assuming that  $C_F$  is properly set to yield a MFM response, the bandwidth of the overall Cherry-Hooper amplifier is

$$\omega_{3dB} = \omega_0 \cong \omega_u \cong \sqrt{\omega_T \cdot \frac{1}{R_F C_L}}$$

Further insight can be gained from this result by considering the input capacitance of the overall Cherry-Hooper amplifier

$$C_{in} \cong \frac{C_\pi}{g_m R_E} = \frac{1}{\omega_T R_E} = \frac{C_L}{k} \quad k = \text{"fanout"}$$

$$\omega_{3dB} \cong \sqrt{\omega_T \cdot \frac{1}{R_F \left( \frac{k}{\omega_T R_E} \right)}} = \frac{\omega_T}{\sqrt{k A_{v0}}}$$

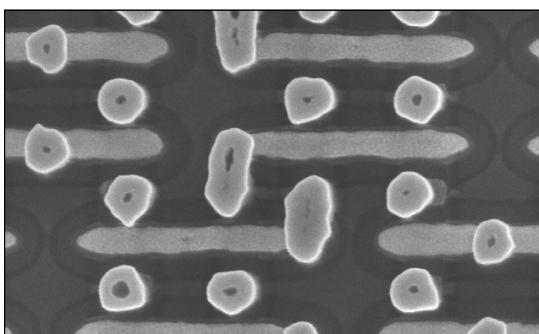
# Chapter 13

## Mismatch

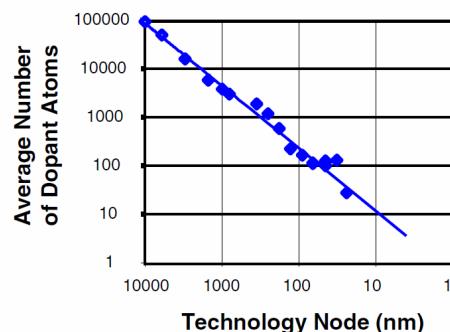
Boris Murmann  
Stanford University  
Winter 2015-16

Textbook Sections: 2.3

### Motivation



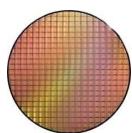
Courtesy A. Bowling  
Texas Instruments



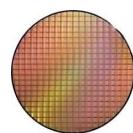
K. Kuhn et al., "Managing Process Variation in Intel's 45nm CMOS Technology," <http://www.intel.com/technology/itj/2008/v12i2/3-managing/1-abstract.htm>

- As transistors become smaller and more intricate, it becomes increasingly important to understand device mismatch
- Important distinctions
  - Global process variations
  - Device-to-device mismatch: Random & systematic errors

## Global Process Variations



Wafer made yesterday  
All NMOS are “slow”  
All PMOS are “nominal”  
All R are nominal  
All C are “fast”

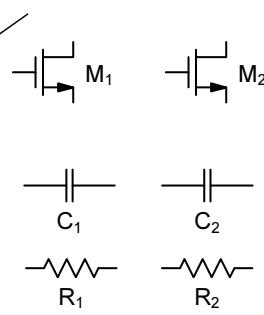
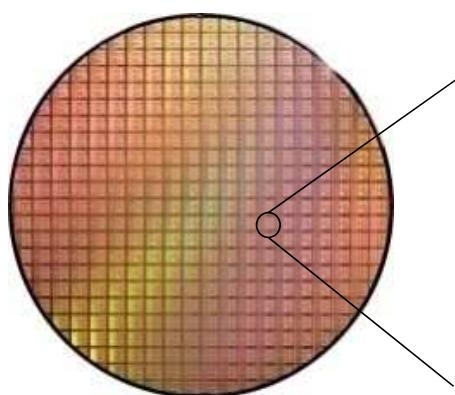


Wafer made today  
All NMOS are “fast”  
All PMOS are “fast”  
All R are nominal  
All C are “slow”

Parameter	“Slow”	“Nominal”	“Fast”
$V_t$	0.4V	0.3V	0.2V
$mC_{ox}$ (NMOS)	240 mA/V <sup>2</sup>	300 mA/V <sup>2</sup>	360 mA/V <sup>2</sup>
$mC_{ox}$ (PMOS)	80 mA/V <sup>2</sup>	100 mA/V <sup>2</sup>	120 mA/V <sup>2</sup>
$R_{poly}$	60Ω/◻	50Ω/◻	40Ω/◻
$R_{nwell}$	1.4 kΩ/◻	1 kΩ/◻	0.6 kΩ/◻
$C_{MIM}$	1.15 fF/mm <sup>2</sup>	1 fF/mm <sup>2</sup>	0.85 fF/mm <sup>2</sup>

## Device-to-Device Mismatch

- Device parameters not only vary from lot-to-lot or wafer-to-wafer, but there are also differences between closely spaced, nominally identical devices on the same chip
  - These differences are called mismatch



$$V_{t1} - V_{t2} = \Delta V_t$$

$$\left( \mu C_{ox} \frac{W}{L} \right)_1 - \left( \mu C_{ox} \frac{W}{L} \right)_2 = \Delta \beta$$

$$C_1 - C_2 = \Delta C$$

$$R_1 - R_2 = \Delta R$$

## Statistical Model

- Experiments over the past decades have shown that device-to-device mismatch ( $\Delta V_t$ ,  $\Delta C$ , ...) for properly laid out devices (no systematic errors) is “random” and well-described by a Gaussian distribution
  - With zero mean and a standard deviation that depends on the process and the size of the device
- The standard deviation of the parameter mismatch between two closely spaced devices is modeled using the following expression

$$\sigma_{\Delta P} = \frac{A_p}{\sqrt{WL}}$$

Sometimes referred to as  
“Pelgrom’s rule”  
 $A_p$ : “Pelgrom coefficient”

where  $W \cdot L$  represents the area of the device, and  $P$  is the device parameter under consideration

[M. Pelgrom et al., “Matching Properties of MOS Transistors,” JSSC, Oct. 1989]

## Parameters for a Typical 0.18- $\mu\text{m}$ Process

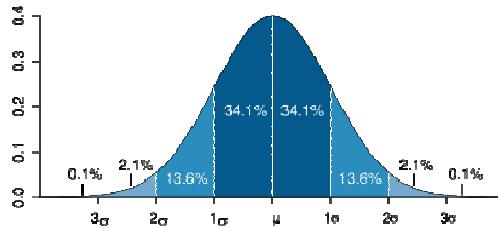
Parameter	Value
$A_{vt}$ (MOSFET)	5 mV- $\mu\text{m}$
$A_{\Delta\beta/\beta}$ (MOSFET)	1 %- $\mu\text{m}$
$A_{\Delta I_s/I_s}$ (Bipolar transistor)	2 %- $\mu\text{m}$
$A_{\Delta\beta/\beta}$ (Bipolar transistor)	4 %- $\mu\text{m}$
$A_{\Delta C/C}$ (MIM capacitor)	1 %- $\mu\text{m}$
$A_{\Delta R/R}$ (Poly resistor)	3 %- $\mu\text{m}$

## Example

Example: MIM-capacitor with  $W=10\mu m$ ,  $L=10\mu m$  ( $\sim 100-200$  fF)

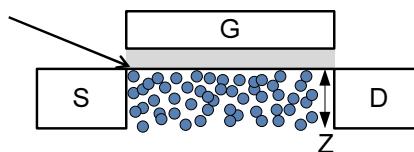
$$\sigma_{\Delta C/C} = \frac{1\%}{\sqrt{100}} = 0.1\% \quad 3\sigma_{\Delta C/C} = 0.3\%$$

[http://en.wikipedia.org/wiki/Image:Standard\\_deviations\\_diagram.svg](http://en.wikipedia.org/wiki/Image:Standard_deviations_diagram.svg)



## $V_t$ Mismatch due to Doping Fluctuations (First Order Analysis)

Average number of dopants  
 $N = N_A WLZ$



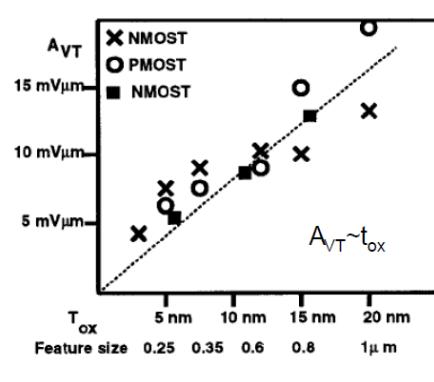
- To first order, the number of doping atoms in the depletion region follow a Poisson distribution with mean  $\lambda$  and standard deviation  $\sqrt{\lambda}$ .
  - Example: 1000 dopants on average  $\rightarrow$  standard deviation of  $\sim 30$

$$V_t = V_{FB} + \frac{qN}{WLC_{ox}}$$

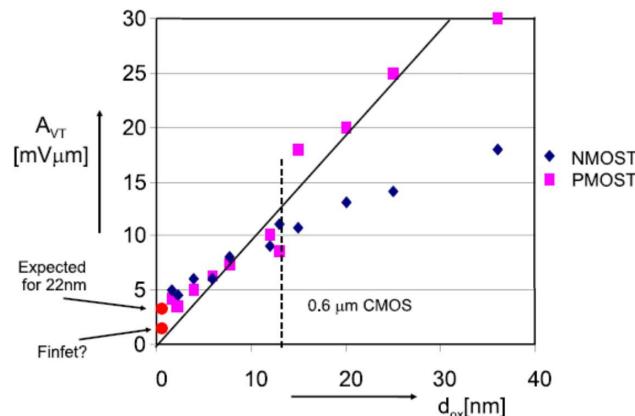
$$stdev(V_t) = \frac{q\sqrt{N}}{WLC_{ox}} = \frac{\frac{t_{ox}}{\varepsilon_{ox}} q \sqrt{N_A Z}}{\sqrt{WL}}$$

$$stdev(\Delta V_t) = \sqrt{2} \frac{\frac{t_{ox}}{\varepsilon_{ox}} q \sqrt{N_A Z}}{\sqrt{WL}} = \frac{A_{vt}}{\sqrt{WL}}$$

## A<sub>Vt</sub> Data Confirming (Approximate) t<sub>ox</sub> Proportionality



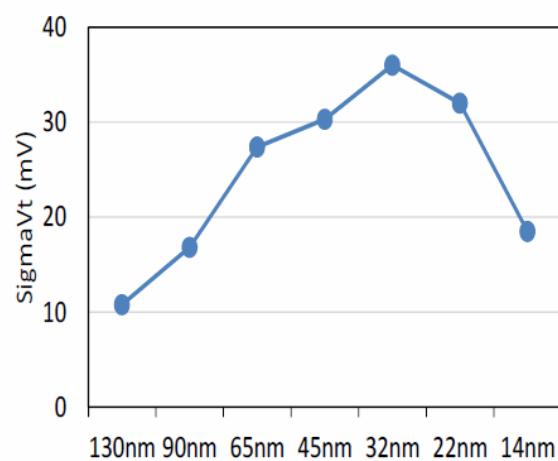
[M. Pelgrom, IEDM 1998]



[Pelgrom, "A Designer's View on Mismatch," in Analog Circuit Design, Springer, 2013]

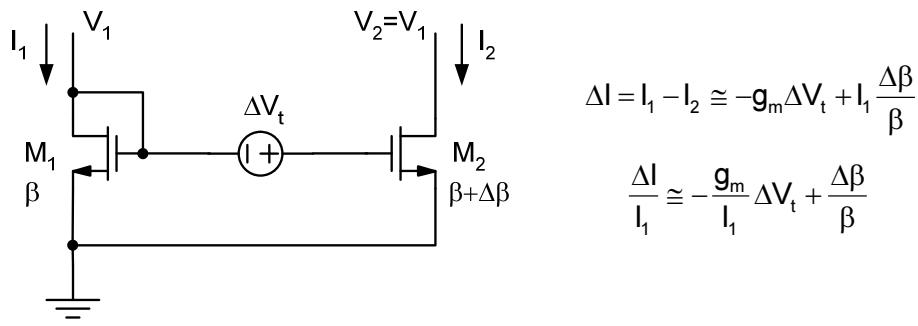
- For constant device area (WL),  $V_t$  mismatch improves as technology is scaled
- Unfortunately,  $A_{Vt}$  reduces not as fast as minimum device area (WL), and hence  $V_t$  mismatch for minimum-size devices worsens

## V<sub>t</sub> Mismatch on Minimum-Size Devices



Natarajan, IEDM 2014

## Mismatch in a MOS Current Mirror



Example:  $W=20\mu\text{m}$ ,  $L=0.2\mu\text{m}$ ,  $g_m/I_D=10\text{S/A}$ ,  $A_{vt}=5\text{mV}\mu\text{m}$ ,  $A_\beta=1\%\mu\text{m}$

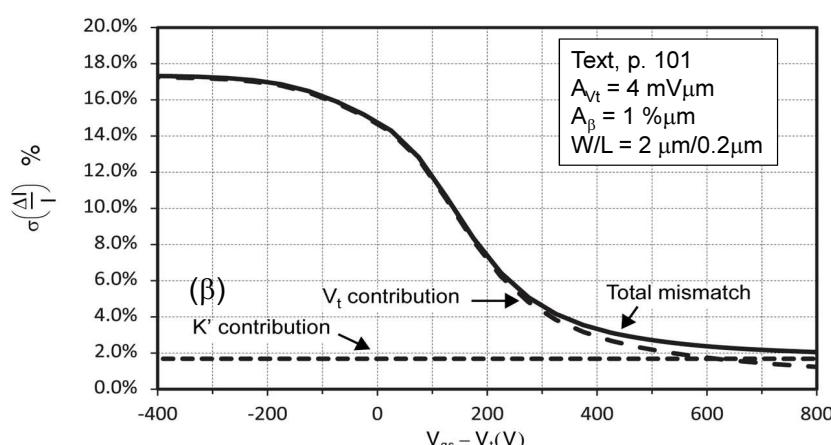
$$\sigma_{\frac{\Delta I}{I_1}} = \sqrt{\left(10 \frac{\text{S}}{\text{A}} \cdot 2.5\text{mV}\right)^2 + (0.5\%)^2} = \sqrt{(2.5\%)^2 + (0.5\%)^2} = 2.54\%$$

- Threshold mismatch usually dominates, unless  $g_m/I_D$  is impractically low

## Design Scenarios (1)

- If the device area is fixed, choose small  $g_m/I_D$  (large  $V_{OV}$ ) for improved matching

$$var\left(\frac{\Delta I_D}{I_D}\right) \cong \left(\frac{g_m}{I_D}\right)^2 \frac{A_{vt}^2}{WL}$$

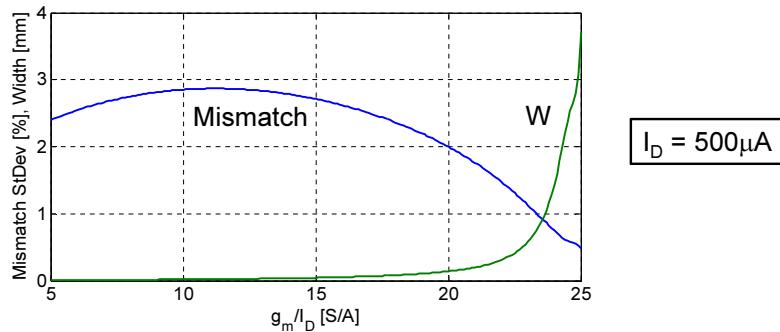


## Design Scenarios (2)

- More interesting case: Mirror input current is fixed

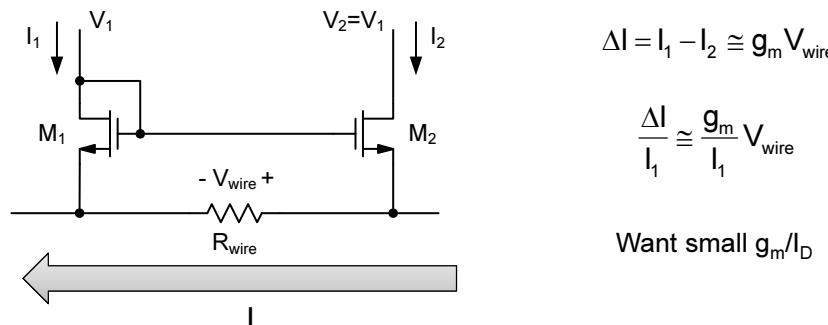
$$\text{var}\left(\frac{\Delta I_D}{I_D}\right) \cong \left(\frac{g_m}{I_D}\right)^2 \frac{A_{VT}^2}{WL} = \left(\frac{2}{V_{OV}}\right)^2 \frac{A_{VT}^2}{WL} \propto \left(\sqrt{\frac{W}{L}}\right)^2 \frac{1}{WL} = \frac{1}{L^2} \quad V_{OV} = \frac{2I_D}{\mu C_{ox} \frac{W}{L}}$$

- The primary “knob” to improve matching is the channel length
- For a real MOSFET, the W terms do not cancel perfectly and the best matching is achieved when the transistor operates in weak inversion

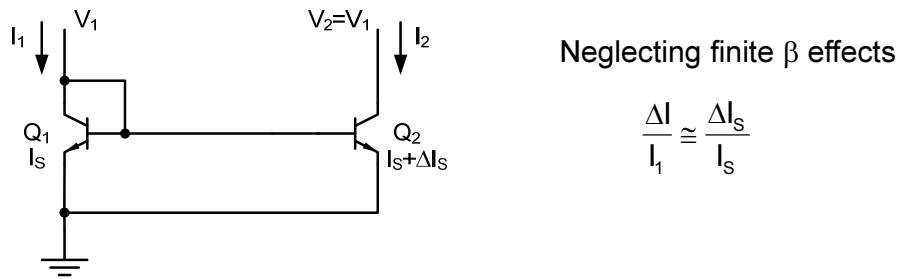


## Design Scenarios (3)

- Obvious issues with pushing the mirror into weak inversion
  - The devices get huge and their parasitics (e.g. junction caps) may significantly load the signal path
  - With large  $g_m/I_D$ , the thermal noise contribution from the mirror may become unacceptably high
  - Usually not worth it...
- Additional issue in practice: sensitivity to systematic errors, e.g. IR drop



## Mismatch in a BJT Current Mirror



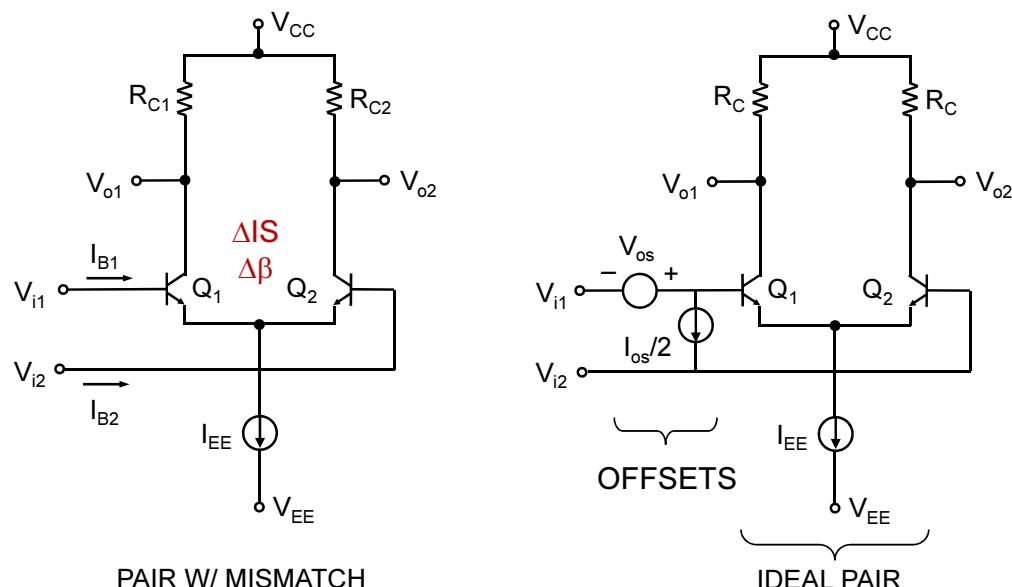
Example:  $A_E = 0.7 \mu\text{m}^2$  (EE214B unit BJT),  $A_{\Delta I_s/I_s} = 2\%/\mu\text{m}$

$$\sigma_{\frac{\Delta I}{I_1}} = \frac{2\%}{\sqrt{0.7}} = 2.4\%$$

- Error magnitude is similar to that of 20/0.2 MOSFET
- Can use multiple unit devices or resistive degeneration to achieve a smaller mirror error

## Input-Referred Offset due to Mismatch

- In a perfectly symmetric differential pair,  $V_{id} = 0$  yields  $V_{od} = 0$
- Imbalances can be modeled as input referred offsets



## Analysis (1)

$$\begin{aligned}
 V_{os} - V_{BE1} + V_{BE2} &= 0 \\
 \therefore V_{os} &= V_T \ln \left( \frac{I_{C1}}{I_{S1}} \right) - V_T \ln \left( \frac{I_{C2}}{I_{S2}} \right) \\
 &= V_T \ln \left[ \left( \frac{I_{C1}}{I_{C2}} \right) \left( \frac{I_{S2}}{I_{S1}} \right) \right], \quad \text{where } V_T = \frac{kT}{q}
 \end{aligned}$$

If  $V_{od} = 0$ , then

$$\begin{aligned}
 I_{C1}R_{C1} &= I_{C2}R_{C2} \\
 \therefore \frac{I_{C1}}{I_{C2}} &= \frac{R_{C2}}{R_{C1}}
 \end{aligned}$$

Thus

$$V_{os} = V_T \ln \left[ \left( \frac{R_{C2}}{R_{C1}} \right) \left( \frac{I_{S2}}{I_{S1}} \right) \right]$$

## Analysis (2)

- For small mismatches  $\Delta R_C \ll R_C$  and  $\Delta I_S \ll I_S$ , it follows that

$$V_{os} \approx V_T \left( -\frac{\Delta R_C}{R_C} - \frac{\Delta I_S}{I_S} \right) = \left( \frac{g_m}{I_D} \right)^{-1} \left( -\frac{\Delta R_C}{R_C} - \frac{\Delta I_S}{I_S} \right)$$

- And similarly

$$I_{os} \approx -\frac{I_C}{\beta} \left( \frac{\Delta R_C}{R_C} + \frac{\Delta \beta}{\beta} \right)$$

- Offset voltage drift

$$\frac{dV_{os}}{dT} = \frac{d}{dT} \left[ \frac{kT}{q} \left( -\frac{\Delta R_C}{R_C} - \frac{\Delta I_S}{I_S} \right) \right] = \frac{V_{os}}{T}$$

- Example

- $V_{os}$  was determined to be 2 mV through a measurement at 300°K
- Means that the offset voltage will drift by  $2 \text{ mV}/300^\circ\text{K} = 6.6 \mu\text{V}/^\circ\text{K}$

## Comparison of $V_{OS}$ for MOS and BJT Differential Pairs

$$V_{OS,BJT} \approx \left( \frac{g_m}{I_D} \right)^{-1} \left( -\frac{\Delta R}{R} - \frac{\Delta I_S}{I_S} \right)$$

$$V_{OS,MOS} \approx \Delta V_t + \left( \frac{g_m}{I_D} \right)^{-1} \left( -\frac{\Delta R}{R} - \frac{\Delta \beta}{\beta} \right) \quad \sigma_{\Delta V_t} = \frac{A_{V_t}}{\sqrt{WL}}$$

Extra      Worse  
 term      than BJT

→ Want large area

- Example (using previous numbers and neglecting resistor mismatch)

$$\text{std}(V_{OS,BJT}) \approx \text{std} \left[ \left( \frac{g_m}{I_D} \right)^{-1} \left( \frac{\Delta I_S}{I_S} \right) \right] = 26 \text{mV} \cdot 2.4\% = 0.62 \text{mV}$$

$$\text{std}(V_{OS,MOS}) \approx \text{std} \left[ \Delta V_t + \left( \frac{g_m}{I_D} \right)^{-1} \left( \frac{\Delta \beta}{\beta} \right) \right] \approx \sqrt{(2.5 \text{mV})^2 + (100 \text{mV} \cdot 1\%)^2} = 2.69 \text{mV}$$

- MOS offset is typically 5-10 times worse than BJT
- Need large MOS devices for low offset
  - Or use offset cancellation tricks – see EE315

## Mismatch “Gotchas” in Modern CMOS Technology

- In CMOS technologies at/below 130nm, a variety of matching “pitfalls” have been discovered
  - Well proximity effect
  - Shallow-Trench Isolation (STI) edge effects
  - Metal coverage effects
  - Matching degradation due to pocket implants
  - Increased ratio of mismatch/process variation
  - ...

## Well Proximity Effect

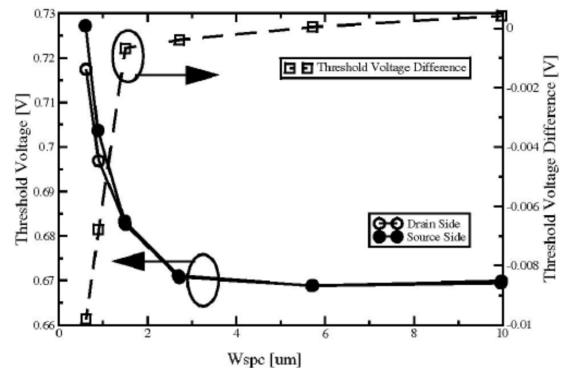
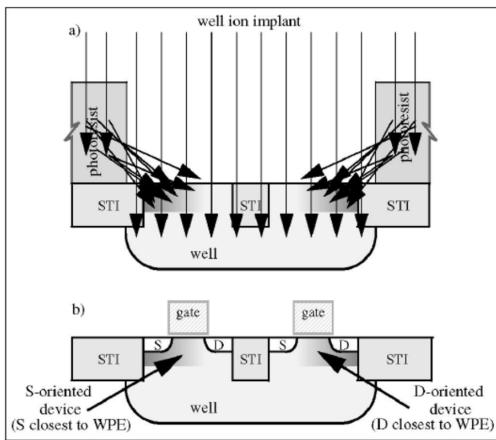
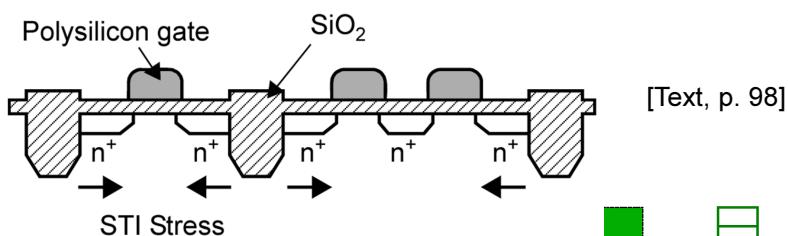


Fig. 3:  $V_t$  versus well-edge distance for 3.3V nMOS device on a 0.13 $\mu m$  technology.

P. G. Drennan et al., "Implications of Proximity Effects for Analog Design," Proc. CICC, pp. 169-176, Sep. 2006.

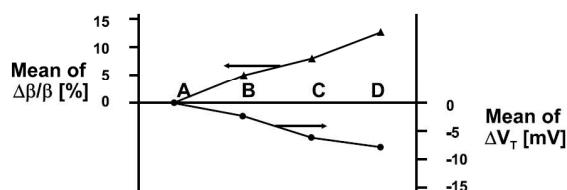
- Transistors near the well edge may see a threshold shift of up to 50mV

## STI Stress Effects



[Pelgrom, "A Designer's View on Mismatch," in Analog Circuit Design, Springer, 2013]

- Good matching requires proper matching of STI stress
  - Match edge distances, avoid placing critical fingers at STI edge (use dummies), etc.



**Figure 1 :** An experiment shows the influence of the STI edge on the drain current and threshold voltage, [8]. Top: a 65-nm 2/0.5  $\mu m$  NMOS reference transistor is designed with the STI edge of source and drain at 2.0  $\mu m$ . A second device has a similar STI distance (A) or at 0.525 (B), 0.35 (C), and 0.16  $\mu m$  (D).

## Metal Coverage Stress Effects

- Covering transistors with metal may reduce mobility
  - Change in annealing
  - Change in stress pattern
- Must match metal coverage and keep wiring away from matching sensitive structures (tens of microns)

[H.P. Tuinhout, M.J.M. Pelgrom, R. Penning de Vries, M. Vertregt, "Effects of Metal Coverage on MOSFET Matching," IEDM Digest, pp. 735-739, 1996]

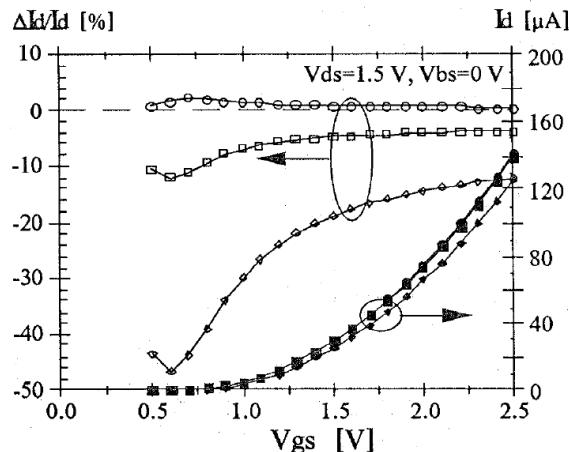


Figure 3. Mismatch characteristic, combined with Drain current

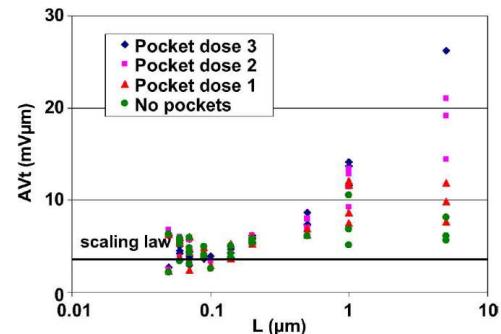
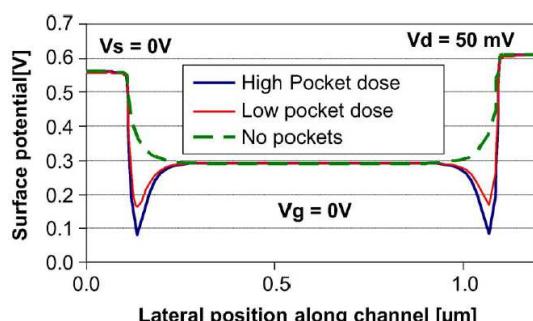
of LEFT transistor of the n-channel 10/10 pairs.

Circles : Both transistors not covered with metal (ref. module)

Squares : Left transistor covered with METAL-2

Diamonds : Left Transistor covered with METAL-1.

## Matching Degradation due to Pocket Implants

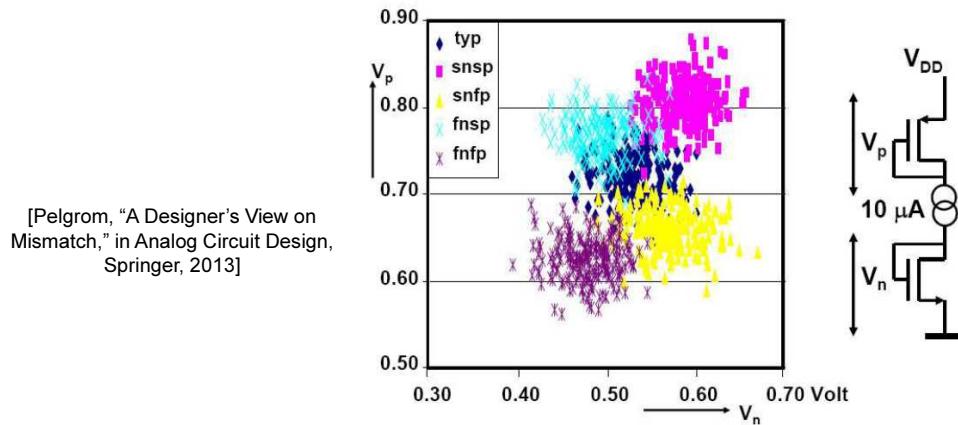


[C.M. Mezzomo, et al., "Characterization and Modeling of Transistor Variability in Advanced CMOS Technologies," IEEE Transactions on Electron Devices, vol. 58, no. 8, pp. 2235-2248, Aug. 2011]

- Doping concentration and electrostatic control is not uniform in a device with pocket implants
- Increasing L does not improve matching, since the drain current is mostly controlled by the pocket regions → Apparent increase in  $A_{vt}$
- Need a better mismatch model, e.g.

$$\sigma_{\Delta V_t} = \frac{A}{\sqrt{WL}} + \frac{B}{\sqrt{W}}$$

## Increased Ratio of Mismatch/Process Variation



**Figure 7 :** Simulation of 200 0.2/0.1 P- and NMOS transistors in their 90-nm process corners. The notation “snfp” refers to slow NMOS and fast PMOS transistors. The variation due to mismatch is of equal importance as the process variation.

- For minimum size transistors, device-to-device mismatch has become comparable in magnitude to global process variations
  - Why still bother with corner models (slow, nom, fast)?