EE214
Advanced Analog Integrated Circuit Design

- Winter 2011 -

Boris Murmann
Stanford University
murmann@stanford.edu

Table of Contents

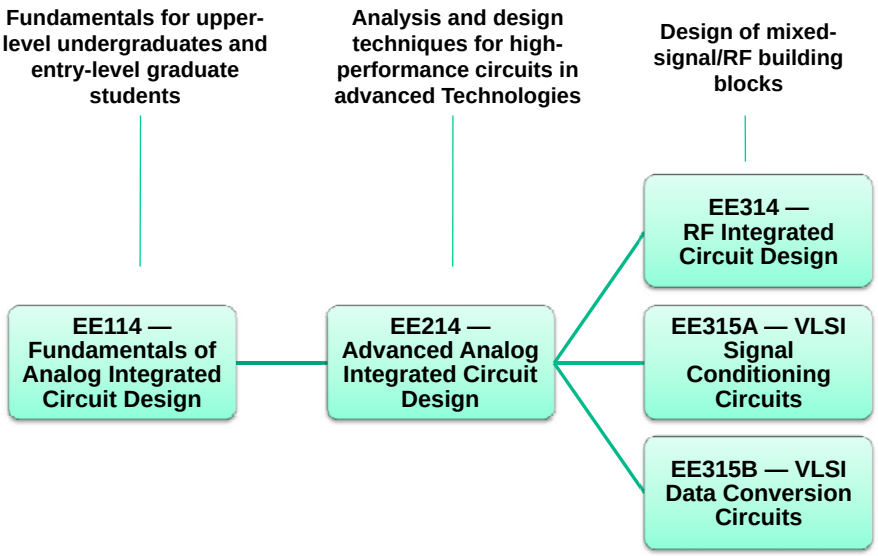
Chapter 1

Introduction

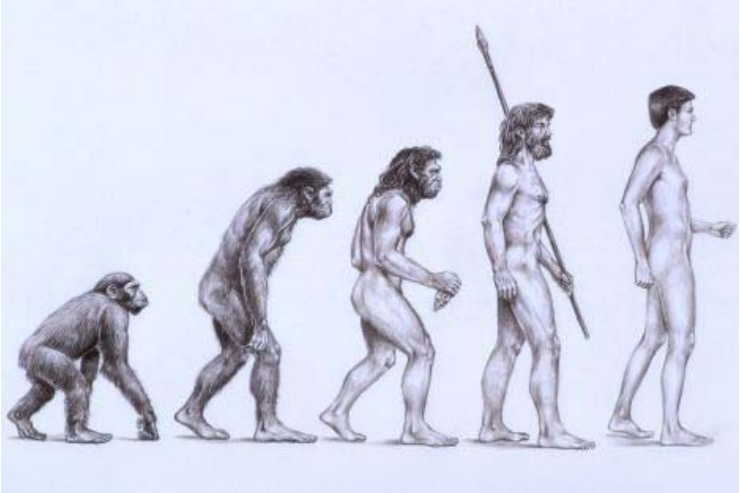
B. Murmann

Stanford University

Analog Circuit Sequence

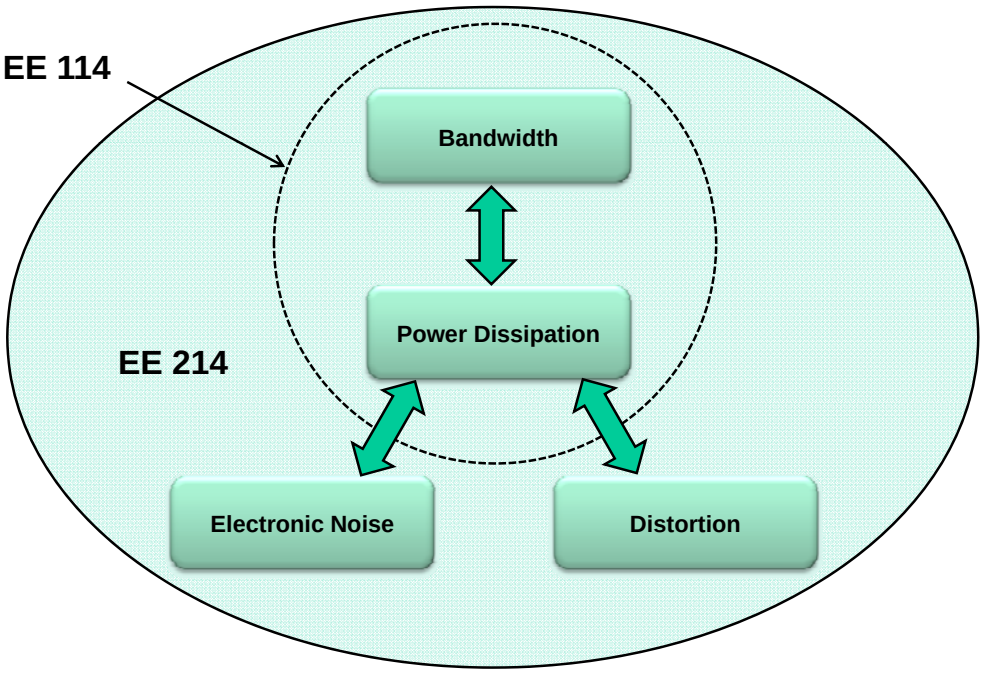


The Evolution of a Circuit Designer…

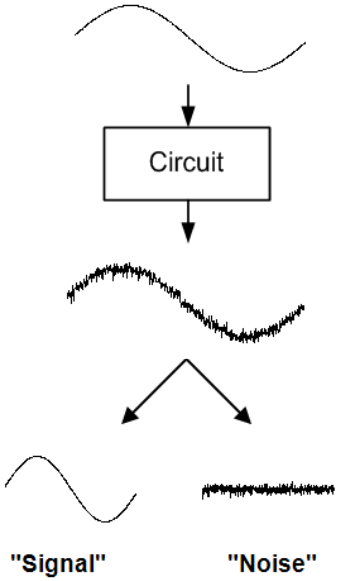


EE101A,B **EE114** **EE214** **EE314**
EE315A,B

Analog Design



Significance of Electronic Noise (1)



Signal-to-Noise Ratio

$$SNR = \frac{P_{signal}}{P_{noise}} \propto \frac{V^2_{signal}}{V^2_{noise}}$$

Significance of Electronic Noise (2)

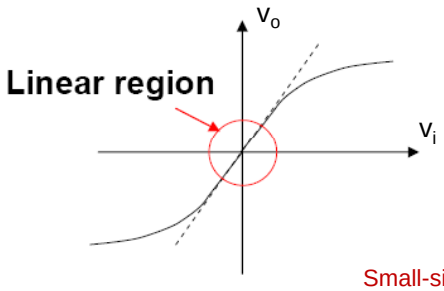
Example: Noisy image

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
  - $SNR \sim V_{signal}^2/V_{noise}^2 \sim (\alpha V_{DD})^2/V_{noise}^2$
- Designing a low-power, high-SNR circuit requires good understanding of electronic noise

Distortion



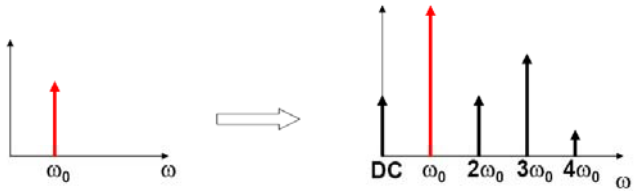
Small-signal approximation

$$v_o = f(v_{in}) \approx c_0 + c_1 v_i + c_2 v_i^2 + c_3 v_i^3 + \ldots$$

- All electronic circuits exhibit some level of nonlinear behavior
  - The resulting waveform distortion is not captured in linearized small-signal models
- The distortion analysis tools covered in EE214 will allow us to quantify the impact of nonlinearities on sinusoidal waveforms

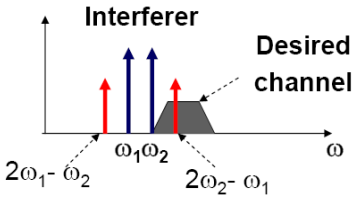
Significance of Distortion

- For a single tone input, the nonlinear terms in a circuit's transfer function primarily result in signal harmonics



- For a two-tone input, the nonlinear terms in a circuit's transfer function result in so-called "intermodulation products"

Example: Two interferer tones create an intermodulation product that corrupts the signal in a desired (radio-) channel



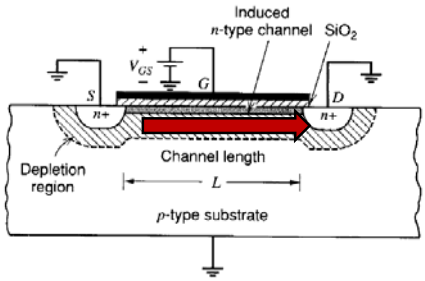
Noise and Distortion Analysis in EE214

- Main objective
  - Acquire the **basic** tools and intuition needed to analyze noise and distortion in electronic circuits
  - Look at a few specific circuit examples to "get a feel" for situations where noise and/or distortion may matter

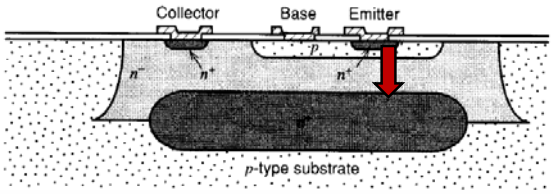
- Leave application-specific examples for later
  - EE314: Noise and distortion in LNAs, mixers and power amplifiers
  - EE315A: Noise and distortion in filters and sensor interfaces
  - EE315B: Noise and distortion in samplers, A/D & D/A converters

Technology

MOSFET



Bipolar Junction Transistor (BJT)



EE114

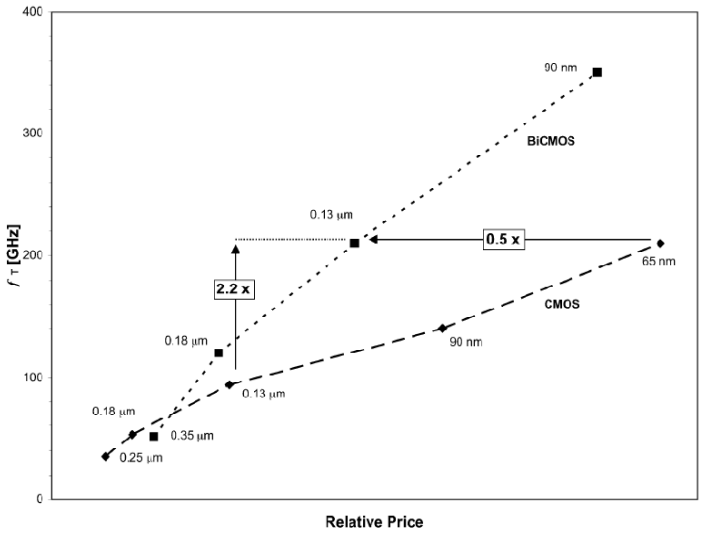
EE214

Bipolar vs. CMOS (1)

- Advantages of bipolar transistors
  - Lower parametric variance
  - Higher supply voltages
  - Higher intrinsic gain ($g_m r_o$)
  - Higher f_T for a given feature size/lithography

- Disadvantages of bipolar transistors
  - Lower integration density, larger features
  - Higher cost

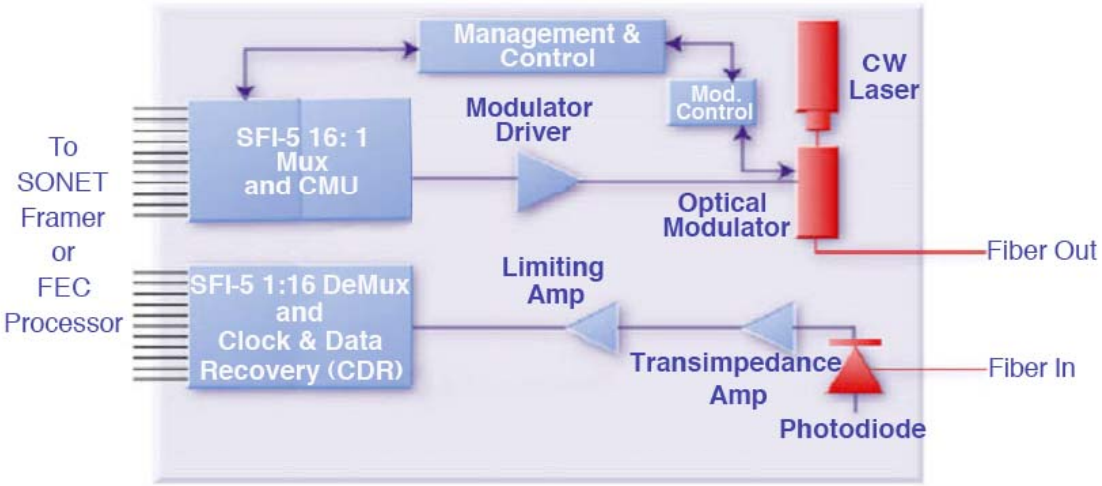
Bipolar vs. CMOS (2)



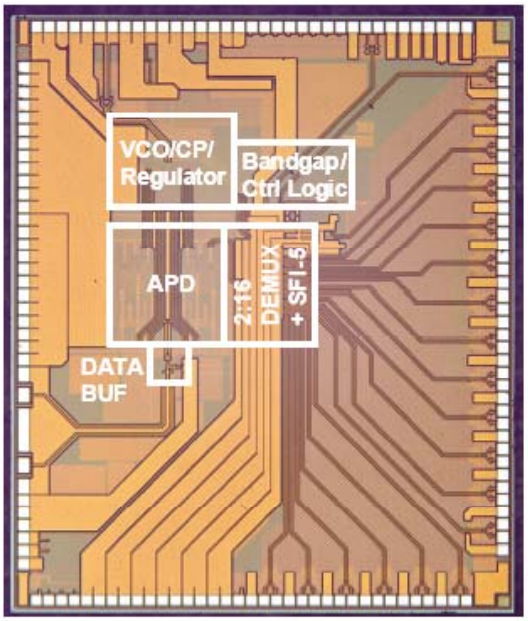
A.J. Joseph, et al., "Status and Direction of Communication Technologies - SiGe BiCMOS and RFCMOS," *Proceedings of the IEEE*, vol.93, no.9, pp.1539-1558, September 2005.

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

Example that Leverages High-Speed BJTs: 40Gb/s Integrated Optical Transponder



Die Photo of 40Gb/s CDR Circuit *



- 120-GHz f_T / 100GHz f_{max} 0.18µm SiGe BiCMOS
- 144 pins
- 3.5mm x 4.2mm
- +1.8V and –5.2V supplies
- 7.5W power dissipation

* A. Ong, et al., ISSCC 2003

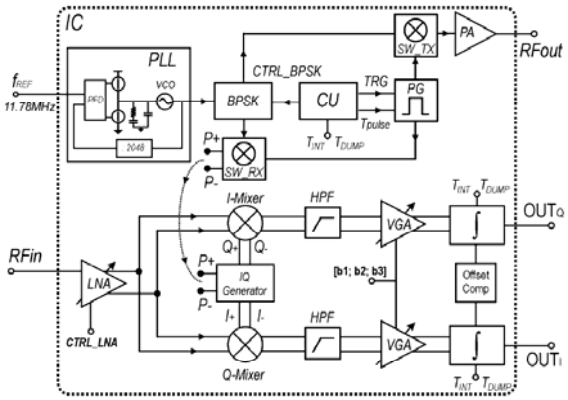
Radar Sensor

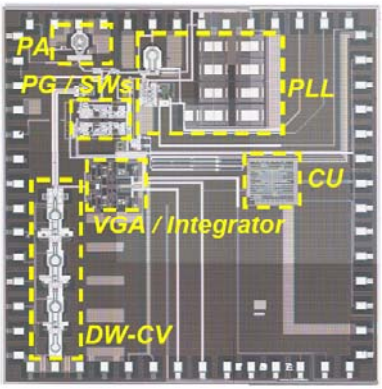
1. Collision warning and mitigation
2. Pre-crash sensing
3. Blind spot detection
4. Parking aid
5. Adaptive Cruise Control (ACC)

BiCMOS process

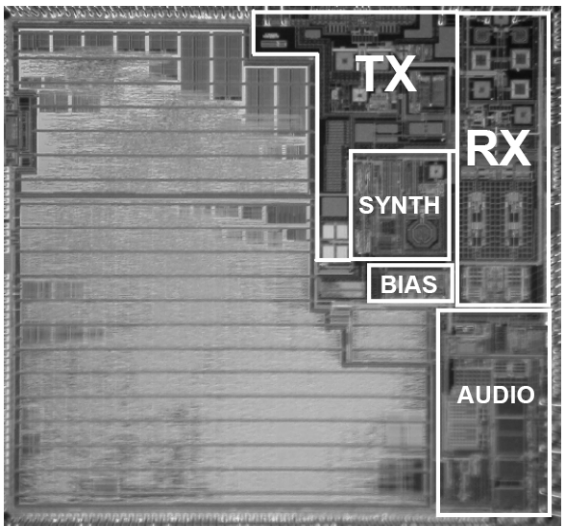
- 0.13-μm CMOS core (1.2/2.5V)
- 1.7V BV_{CEO} SiGe HBT
- 166/175GHz f_t/f_{max}
- 6 metal layers (1 thick)
- Capacitors MIM (2 fF/μm^2)
- Spiral PGS inductors

Ragonese et al., ISSCC 2009





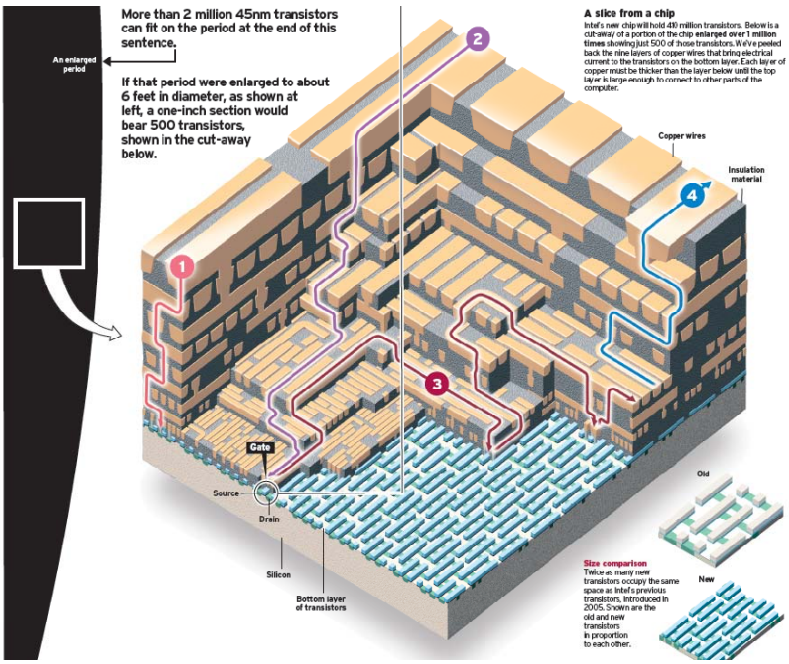
Example that Leverages Densely Integrated CMOS: RF Transceiver System-on-a-Chip (SoC)



- In modern CMOS technology, millions of logic gates can be integrated on a chip
  - Together with moderate- to high performance analog blocks

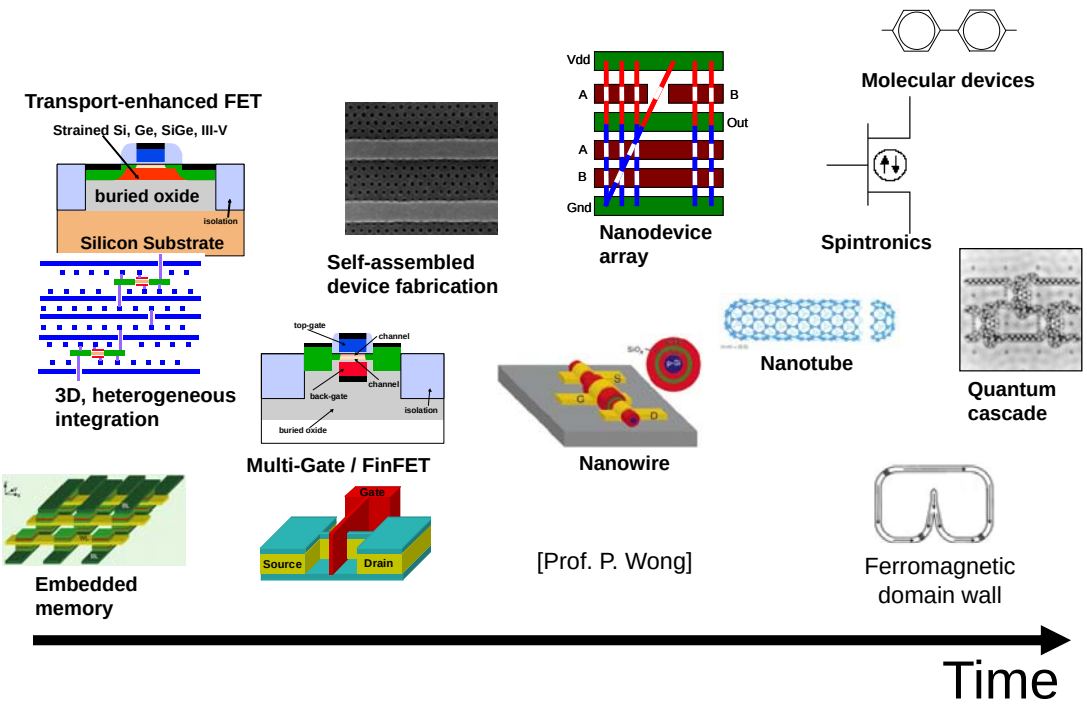
Mehta et al., "A 1.9GHz Single-Chip CMOS PHS Cellphone," ISSCC 2006.

45nm CMOS (Intel)



Steve Cowden
THE OREGONIAN
July 2007

Research in Device Technology



Thoughts on Device Technology

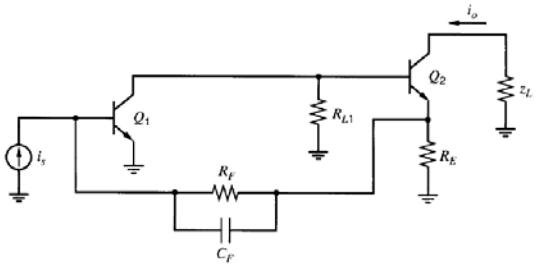
- In the future, innovative circuit designers must embrace "whichever" technology is most suitable (in terms of performance, cost, reliability, etc. for their specific problem
  - Regardless of the respective I-V law and associated nonidealities

- In EE214, we will use bipolar **and** MOS technology to illustrate the similarities and differences between two advanced technologies
- The device parameters and simulation models of the "EE214 technology" correspond to a modern 0.18-µm SiGe BiCMOS technology
  - See e.g. S. Wada, et al., "A manufacturable 0.18-um SiGe BiCMOS technology for 40-Gb/s optical communication LSIs," in *Proc. 2002 Bipolar/BiCMOS Circuits and Technology Meeting, pp. 84–87.*

Analysis of Feedback Circuits

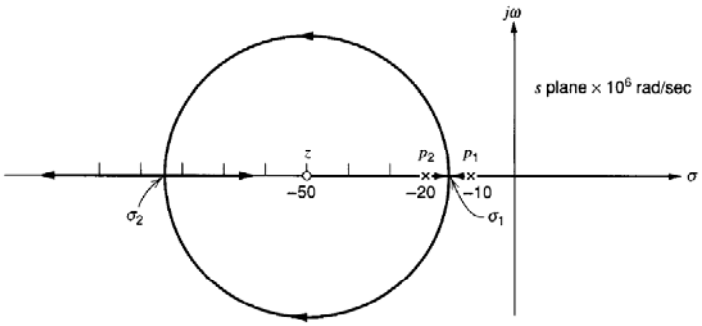
- Feedback circuits can be studied in several ways
  - Return ratio analysis (EE114)
  - Two-port analysis (EE214)
- Both methods have their own merits and demerits, and a good circuit designer should understand both approaches
- Two-port analysis nicely captures a number of practical scenarios in which the forward amplifier ("a") and feedback network ("f") can be intuitively identified and separated (while maintaining loading effects)
  - Shunt-shunt, shunt-series, series-shunt, series-series configurations

Example:
Shunt-series feedback circuit



Root Locus Techniques

- Provides intuitive guidance on "where the poles move" when the loop gain is varied
- Valuable for stability analysis and frequency compensation



Course Outline

- BJT & short channel MOS device models
- Review of elementary amplifier stages (BJT focus)
- Two-port feedback circuit analysis
- Root locus
- Wideband amplifiers
- Noise analysis
- Distortion analysis
- OpAmps and output stages

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
  - Distortion analysis
  - Noise analysis
  - Two-port feedback circuit analysis
  - Root locus techniques and their application to broadband amplifiers
- Solidify the above topics in a hands-on project involving the design and optimization of a broadband amplifier circuit

Staff and Website

- Instructors
  - Boris Murmann, Drew Hall
- Teaching assistants
  - Kamal Aggarwal, Pedram Lajevardi
- Administrative support
  - Ann Guerra, CIS 207
- Lectures are televised
  - But please come to class to keep the discussion interactive!
- Web page: http://ccnet.stanford.edu/ee214
  - Check regularly, especially bulletin board
  - Register for online access to grades and solutions

Text and Prerequisites

- EE214 Course reader
  - Hardcopies available at Stanford Bookstore (~1/3)
- Required textbook
  - Gray, Hurst, Lewis and Meyer, *Analysis and Design of Analog Integrated Circuits*, 5th ed., Wiley
- Reference text
  - B. Razavi, *Design of Integrated Circuits for Optical Communications*, McGraw-Hill, 2002
- Course prerequisite: EE114 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

Assignments

- Homework (20%)
  - Handed out on Wed, due following Wed in class
  - Lowest HW score will be dropped
  - Policy for off-campus students
    - Fax or 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!
- Final Exam (30%)

Honor Code

- Please remember 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

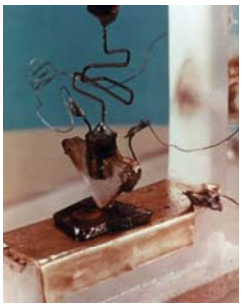
Chapter 2
Bipolar Junction Transistors

B. Murmann

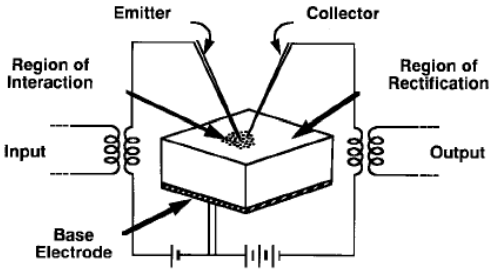
Stanford University

Reading Material: Sections 1.1, 1.2, 1.3, 1.4, 2.5, 2.6, 2.7, 2.11, 2.12

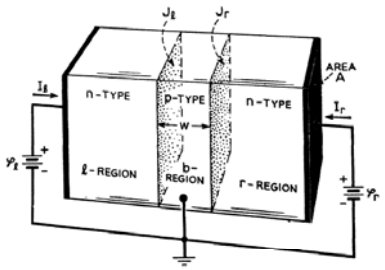
History



Bardeen, Brattain, and Shockley, 1947

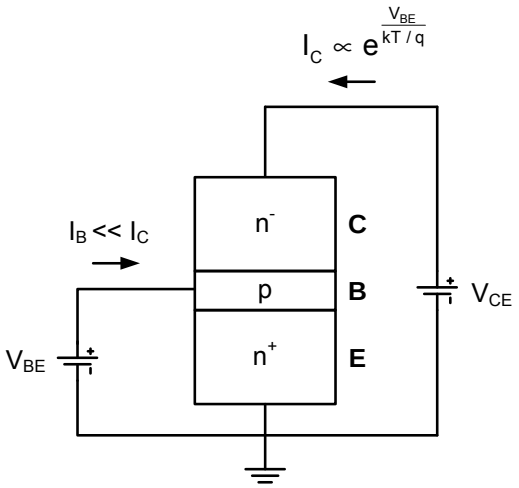


W. Brinkman, D. Haggan, and W. Troutman, "A history of the invention of the transistor and where it will lead us," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1858-1865, Dec. 1997.



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)

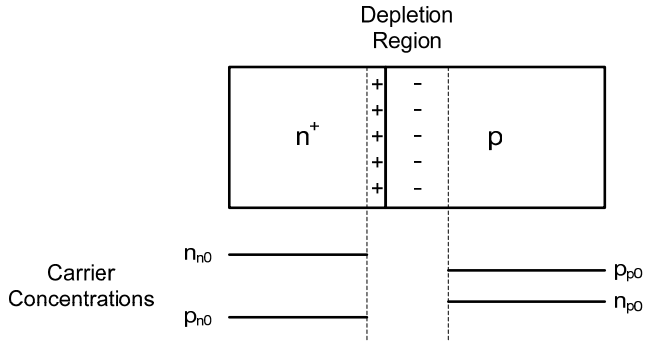


- Device acts as a voltage controlled current source
  - V_{BE} controls I_C
- The base-emitter junction is forward biased and the base-collector junction is reverse biased
- The device is built such that
  - The base region is very thin
  - The emitter doping is much higher than the base doping
  - 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 pn^+ junction
  - The properties of a reverse biased pn^- 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⁺ Junction in Equilibrium (No Bias Applied)



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

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

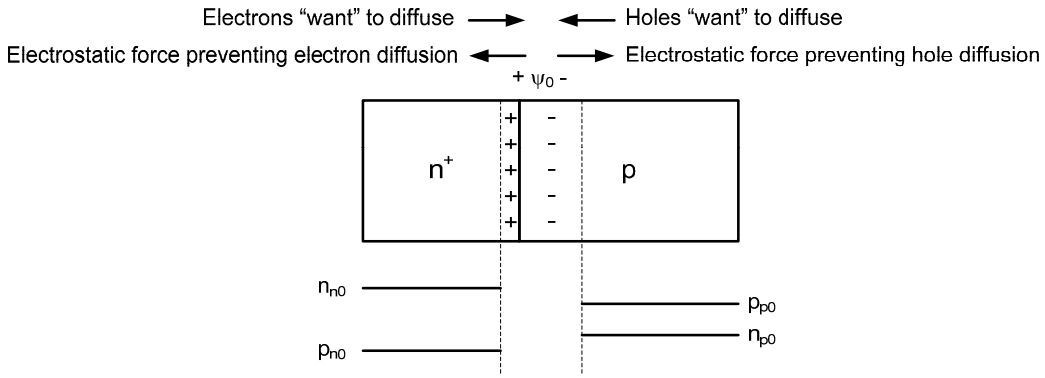
$$n_{p0} = \frac{n_i^2}{p_{p0}} \cong \frac{n_i^2}{N_A}$$

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

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

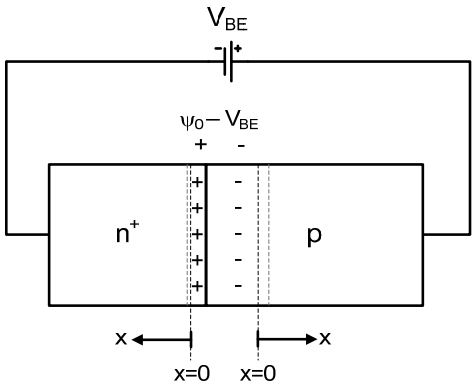


- The built in potential sets up an electric field that opposes the diffusion of mobile holes and electrons across the junction

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

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

pn⁺ 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

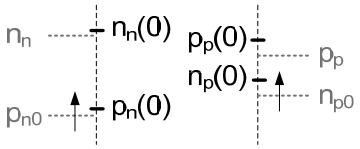
$$\psi_0 - V_{BE} = V_T \ln\left(\frac{n_n}{n_p(0)}\right) \cong V_T \ln\left(\frac{N_D}{n_p(0)}\right) \qquad \therefore n_p(0) = \frac{N_D}{e^{\frac{\psi_0}{V_T}}} \cdot e^{\frac{V_{BE}}{V_T}} = n_{p0} e^{\frac{V_{BE}}{V_T}} \cong \frac{n_i^2}{N_A} e^{\frac{V_{BE}}{V_T}}$$

pn+ 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

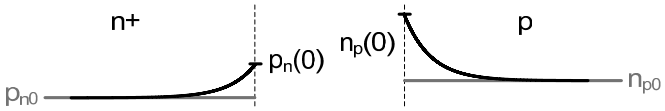
$$p_n(0) = p_{n0} \cdot e^{\frac{V_{BE}}{V_T}} \cong \frac{n_i^2}{N_D} \cdot e^{\frac{V_{BE}}{V_T}}$$

- Since $N_D >> N_A$, it follows that $p_n(0) << n_p(0)$, 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 $n_p(0) << p_p \cong N_A$ (or $p_n(0) << n_n \cong N_D$)
  - These conditions are called "low-level injection"

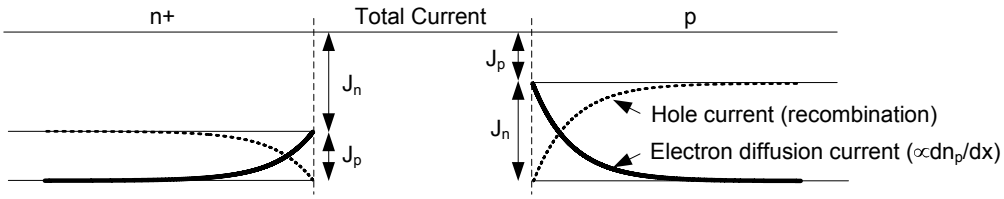


What Happens with the Injected Minority Carriers?

- 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)



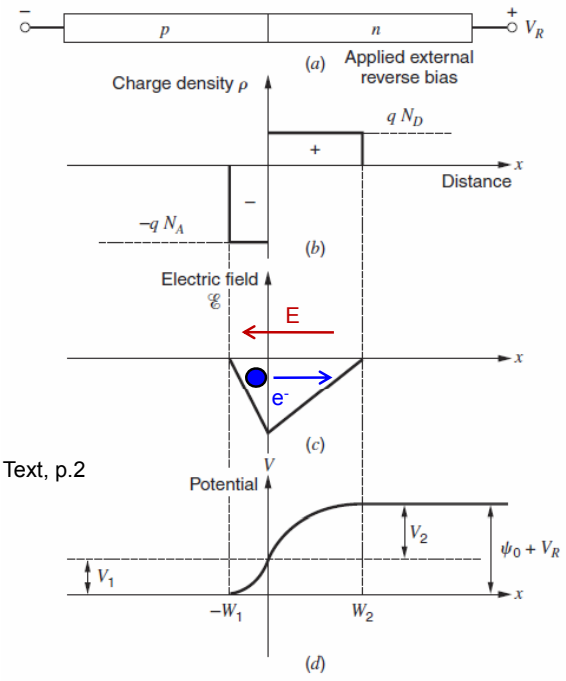
- In each region, there are now two types of currents
  - Diffusion of injected minority carriers due to non-zero dn_p/dx (or dp_n/dx)
  - Majority carrier currents for recombination



Summary – Forward Biased pn+ junction

- Lots of electrons being injected into the p-region, not all that many holes get injected into the n^+ region
  - The heavier 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

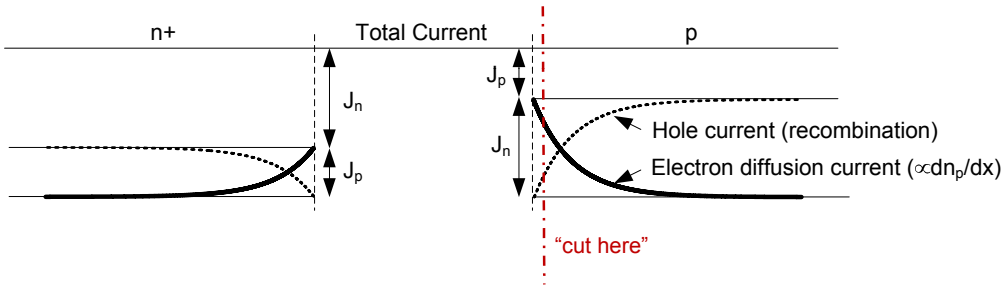
Reverse Biased pn⁻ Junction



Text, p.2

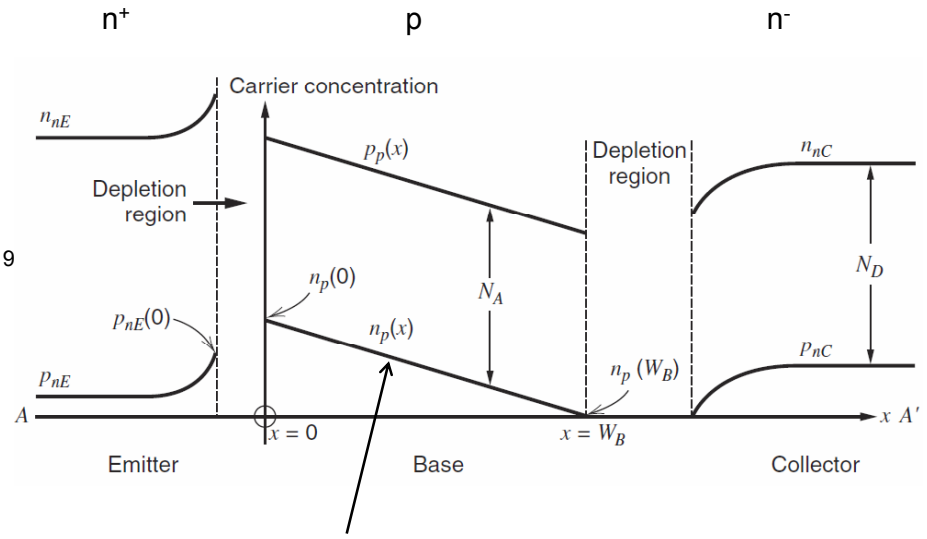
- Reverse bias increases the width of the depletion region and increases the electric field
- Depletion region extends mostly into n⁻ side
- Any electron that would "somehow" make it into the depletion region will be swept through, into the n-region
  - Due to electric field

Bipolar Junction Transistor – Main Idea



- Make the p-region of the pn⁺ junction very thin
- Attach an n⁻ region that will "collect" and sweep across most of the electrons before there is a significant amount of recombination

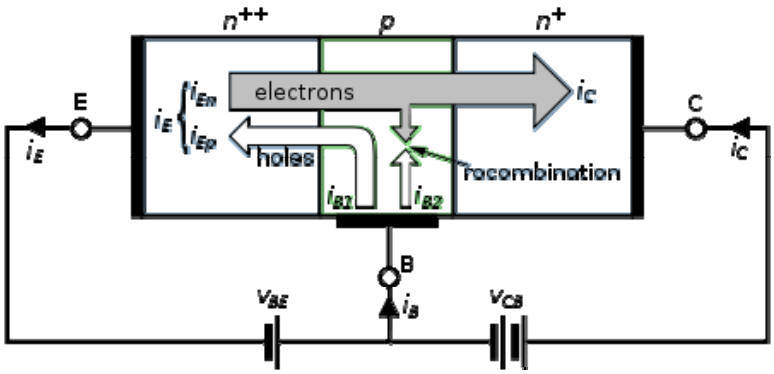
Complete Picture



Text, p. 9

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

BJT Currents



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 (→ 0 for infinite emitter doping)
  - Recombination in the base (→ 0 for base width approaching zero)

First-Order Collector Current Expression

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

$$I_C \cong qAD_n \frac{n_p(0)}{W_B}$$ A is the cross-sectional area
W_B is the base width

$$n_p(0) \simeq \frac{n_i^2}{N_A} e^{\frac{V_{BE}}{V_T}}$$ Result from slide 7

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

$$\boxed{\therefore I_C \cong I_S e^{\frac{V_{BE}}{V_T}}} \qquad \boxed{I_S = \frac{qAD_n n_i^2}{W_B N_A}}$$

Base Current

$I_B = I_{B1} + I_{B2}$ where I_{B1} = Recombination in the base
I_{B2} = Injection into the emitter

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

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

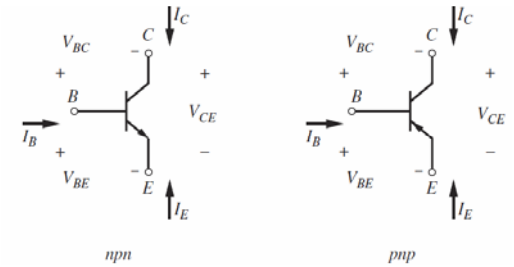
I_{B2} depends on the gradient of minority carriers (holes) in the emitter. For a "long" emitter (all minority carriers recombine)

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

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

Terminal Currents and Definition of α_F, β_F

Text, p. 9



$$I_E = -(I_C + I_B)$$

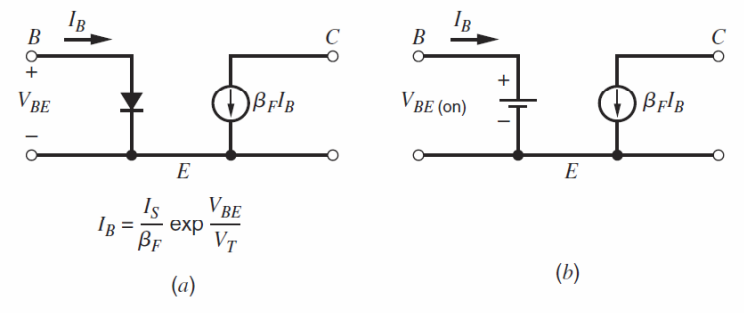
$$\beta_F \triangleq \frac{I_C}{I_B} \quad \text{(ideally infinite)}$$

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

- 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

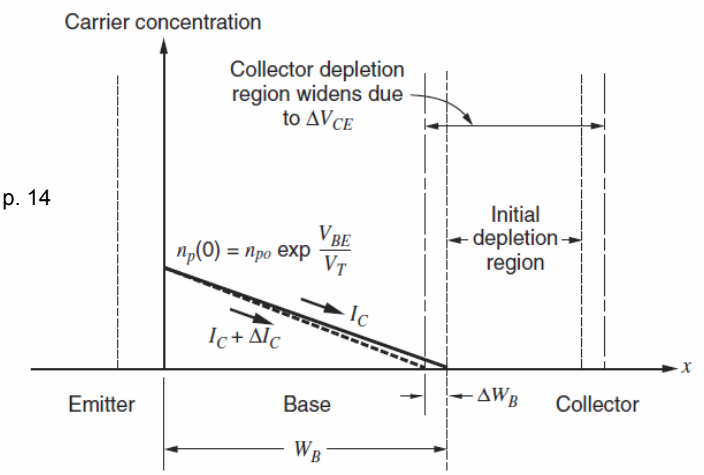
Text, p. 13



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

Basewidth Modulation (1)

Text, p. 14



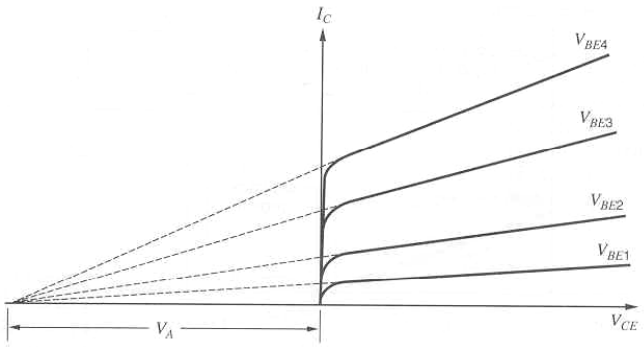
Side note:
BJT inherently has better (higher) r_o than MOS since lower doping on n-side (collector) has most of the depletion region inside the collector

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

(See eq. (1.18) for dW_B/dV_{CE} term)

Early Voltage (V_A)

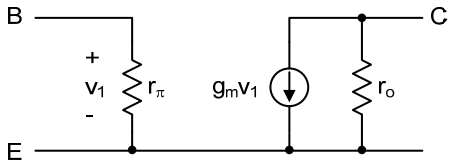
Text, p. 15



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

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

Small-Signal Model



$$g_m = \frac{dI_C}{dV_{BE}} = \frac{d}{dV_{BE}} I_S e^{\frac{V_{BE}}{VT}} = \frac{I_C}{V_T}$$

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

$$g_o = \frac{1}{r_0} = \frac{dI_C}{dV_{CE}} = \frac{d}{dV_{CE}}\left[I_S e^{\frac{V_{BE}}{V_T}} \left(1 + \frac{V_{CE}}{V_A}\right)\right] \cong \frac{I_C}{V_A}$$

Intrinsic Gain

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

- In the EE214 technology, the SiGe npn device has $V_A = 90\text{V}$, thus

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

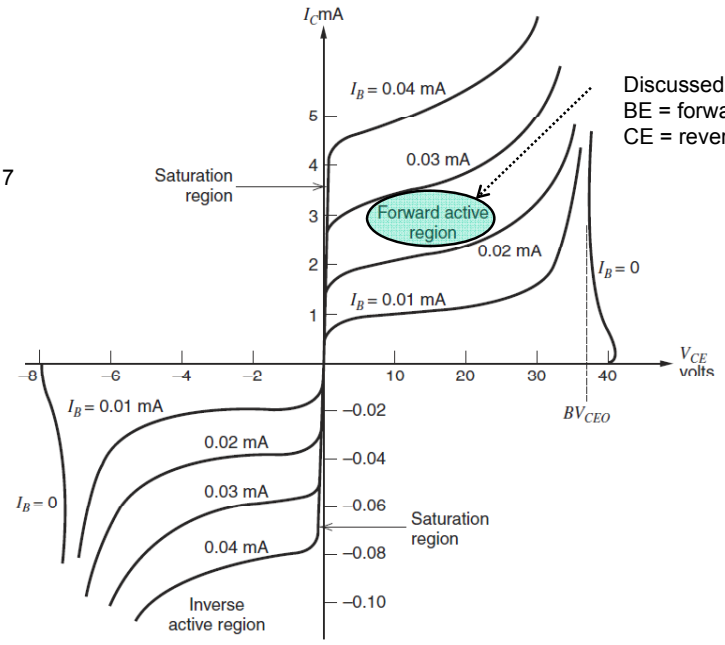
- Much larger than the intrinsic gain of typical MOSFET devices

Outline – Model Extensions and Technology

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

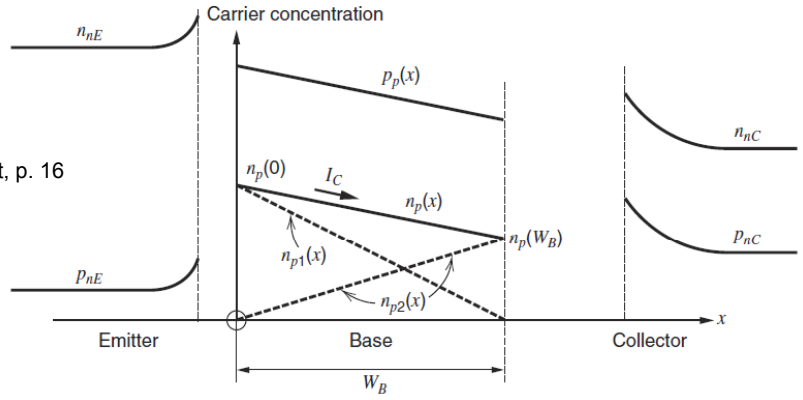
BJT Operating Regions

Text, p. 17



Carrier Concentrations in Saturation

Text, p. 16

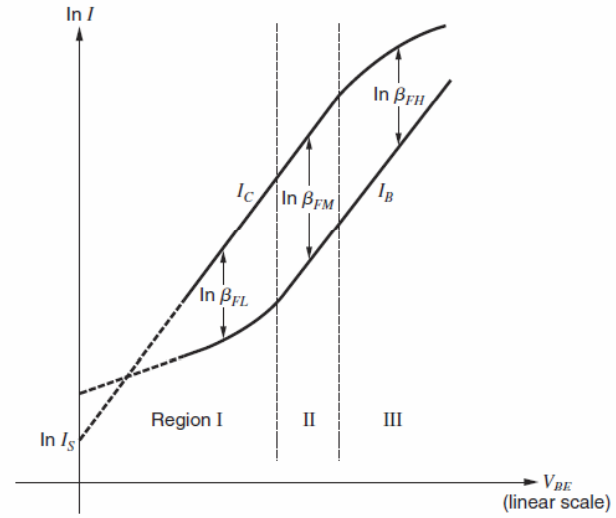


- Base-Collector junction is forward biased
- $n_p(W_B)$, and therefore also I_C, strongly depend on V_{BC}, V_{CE}
- $V_{CE(sat)}$ is the voltage at which the devices enters saturation
  - The difference between the two junction voltages, small ~0.05…0.3V

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 β_F is maintained (region II below)
- What happens in regions I and III?

Text, p. 24



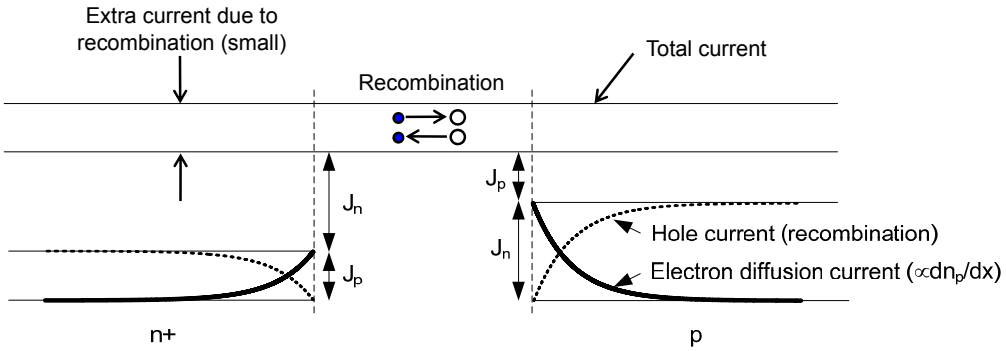
β_F Fall-Off

- 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

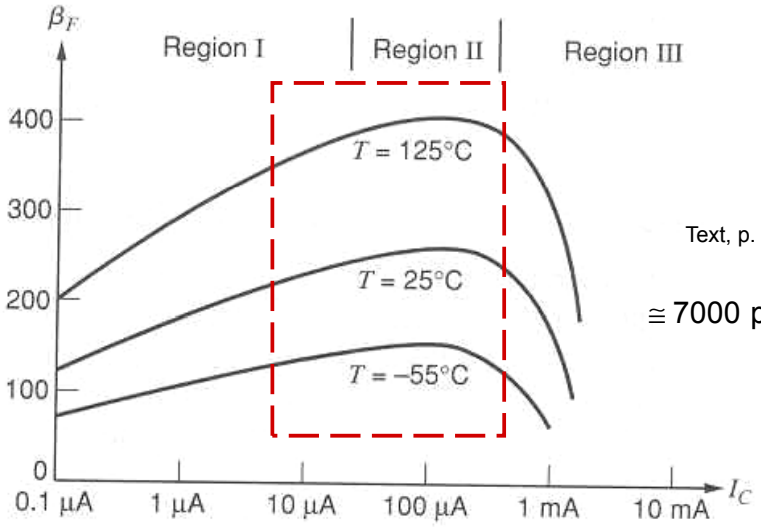
$$I_C = I_S e^{\frac{1}{2}\frac{V_{BE}}{V_T}}$$

- 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



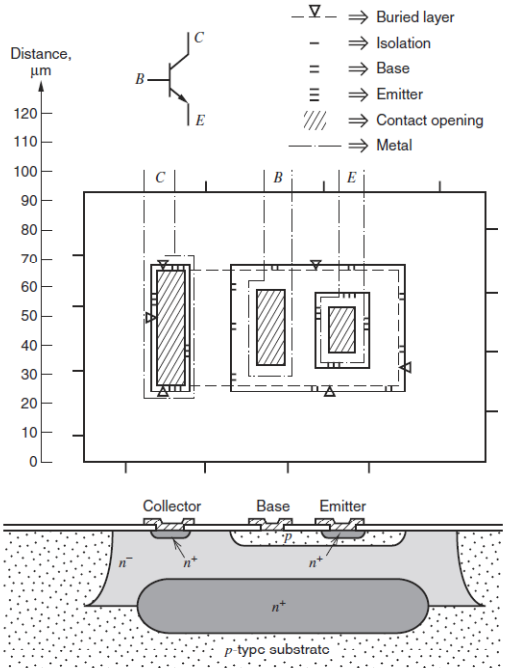
β_F vs. I_C and Temperature



Text, p. 24

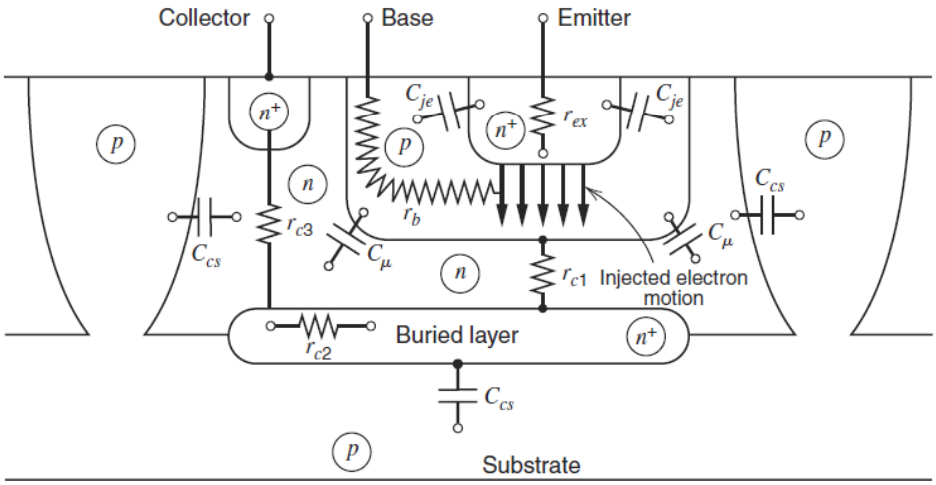
$\cong$ 7000 ppm / °C

Junction Isolated npn Transistor



Text, p. 97

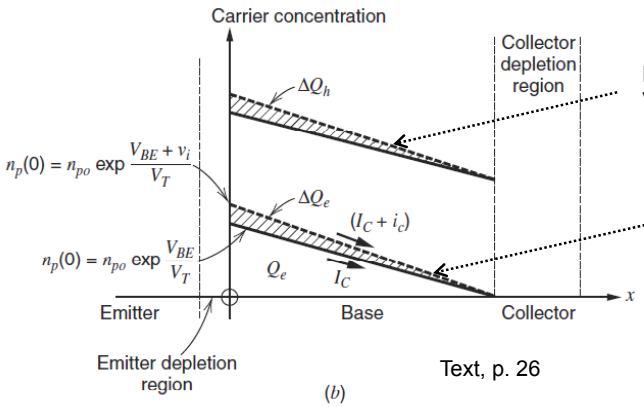
Device Capacitances and Resistances



- 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 terminal must supply charge for neutrality
Voltage dependent charge → capacitance (C_b)

ΔV_{BE} causes change in injected charge

Text, p. 26

Base Charging Capacitance

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

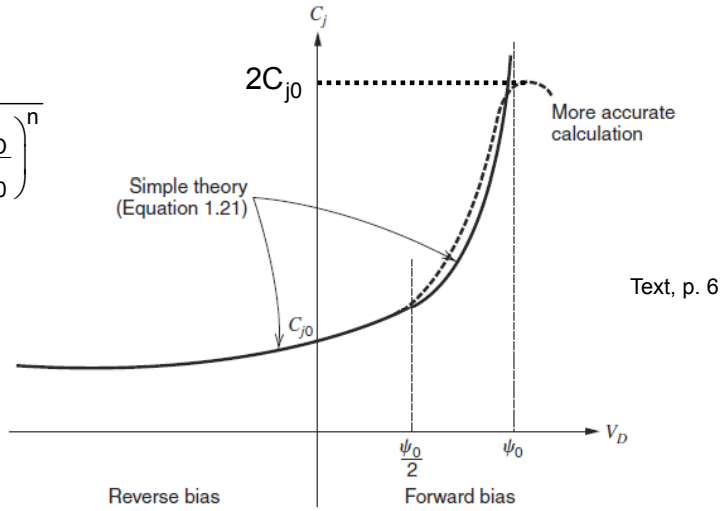
$$\tau_F = \frac{\partial Q_e}{\partial I_C} = \frac{\partial}{\partial I_C}\left(\frac{1}{2}n_p(0)W_B qA\right) \qquad I_C = \frac{qAD_n n_p(0)}{W_B}$$

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

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

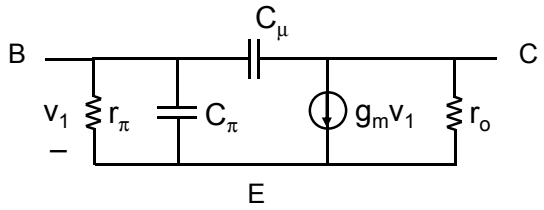
Junction Capacitance

$$C_j = \frac{C_{j0}}{\left(1 - \frac{V_D}{\psi_0}\right)^n}$$



Text, p. 6

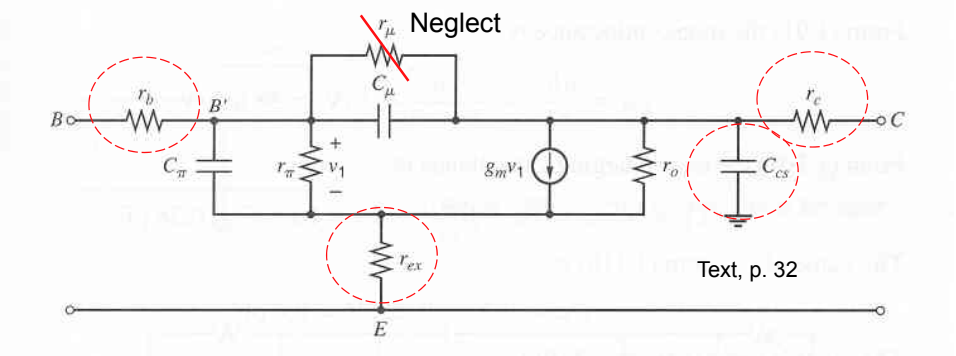
Small-Signal Model with Intrinsic Capacitances



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

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

Model with Additional Parasitics



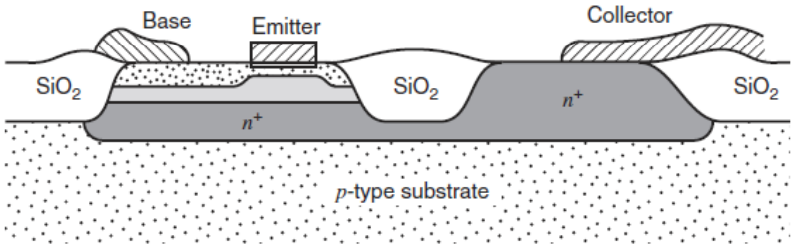
Text, p. 32

Range of numbers

r_e ~1-3Ω
r_b ~ 50-500Ω
r_c ~ 20-500Ω
C_{CS} ~ 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



Text, p. 107

- 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})

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

Added degree of freedom for HBT

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

Device Parameter Comparison

Parameter	Vertical *npn* Transistor with 2 μm² Emitter Area	SiGe npn HBT Transistor with 0.7μm² = 0.22μm x 3.2μm Emitter Area
β_F	120	**300**
β_R	2	**2**
V_A	35 V	**90**
I_S	6×10^{-18}A	**3.2x10^{-17}A** ← 3x smaller device, 5x bigger I_S
I_{CO}	1 pA	**1pA**
BV_{CEO}	8 V	**2.0V**
BV_{CBO}	18 V	**5.5V**
BV_{EBO}	6 V	**3.3V**
τ_F	10 ps	**0.56ps** ← 18x smaller τ_F
τ_R	5 ns	**10ps**
r_b	400 Ω	**25Ω** ← 16x smaller r_b
r_c	100 Ω	**60Ω**
r_{ex}	40 Ω	**2.5Ω**
C_{je0}	5 fF	**6.26fF**
ψ_{0e}	0.8 V	**0.8V**
n_e	0.4	**0.4**
$C_{\mu 0}$	5 fF	**3.42fF**
ψ_{0c}	0.6 V	**0.6V**
n_c	0.33	**0.33**
C_{cs0} (C_{bs0})	20 fF	**3.0fF** ← Oxide isolation vs. Junction isolation
ψ_{0s}	0.6 V	**0.6V**
n_s	0.33	**0.33**

BiCMOS Technology

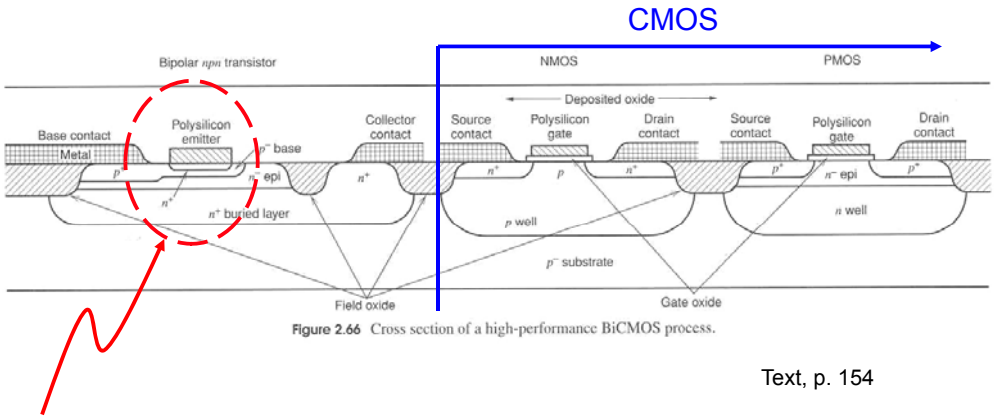


Figure 2.66 Cross section of a high-performance BiCMOS process.

Text, p. 154

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^2

- Qualified 1Q06
- In Production

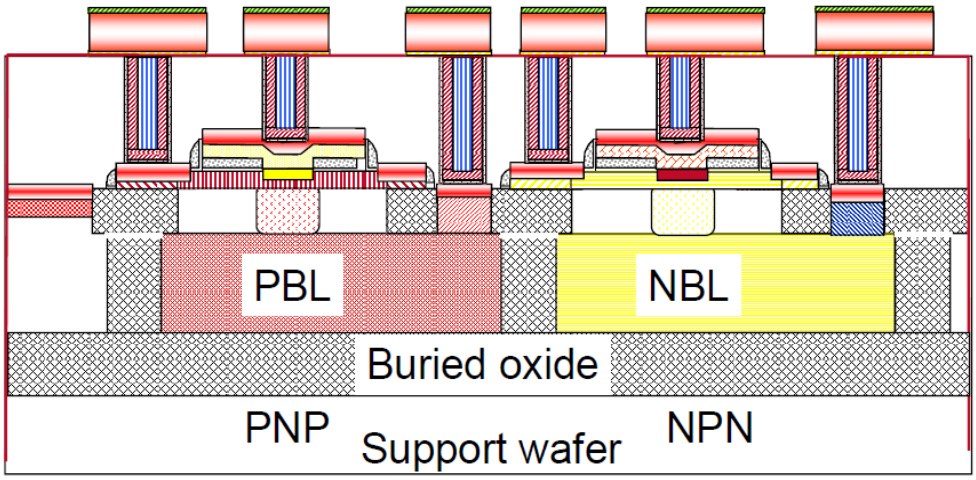
Component Set:

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

BiCom3 EDN innovation HONORING EXCELLENCE IN ELECTRONICS 2002 WINNER

[Texas Instruments]

Cross Section



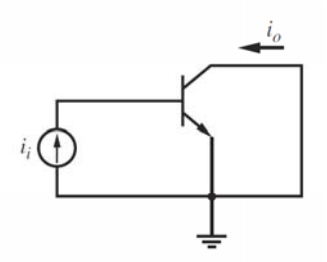
[Texas Instruments]

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 EE314
- 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 ~$f_T/5$
  - Therefore, f_T should be viewed as an extrapolated parameter, or simply as a proxy for device transconductance per capacitance

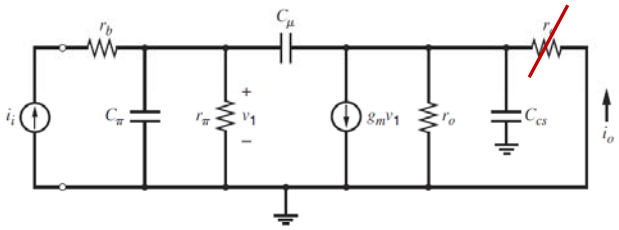
Transit Frequency Calculation (1)

Text, p. 35



(AC circuit; DC biasing not shown)

Ignore for simplicity

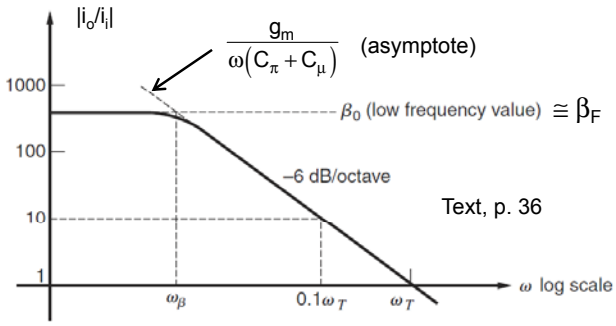


$$v_1 = \frac{r_\pi}{1 + r_\pi s\left(C_\pi + C_\mu\right)} i_i \qquad i_o = g_m v_1$$

$$\frac{i_o}{i_i} = \frac{g_m r_\pi}{1 + r_\pi j\omega\left(C_\pi + C_\mu\right)} \cong g_m r_\pi = \beta_F \quad \text{for} \quad \omega << \frac{1}{r_\pi\left(C_\pi + C_\mu\right)} = \omega_\beta$$

$$\frac{i_o}{i_i} \cong \frac{g_m}{j\omega\left(C_\pi + C_\mu\right)} \quad \text{for} \quad \omega >> \omega_\beta$$

Transit Frequency Calculation (2)



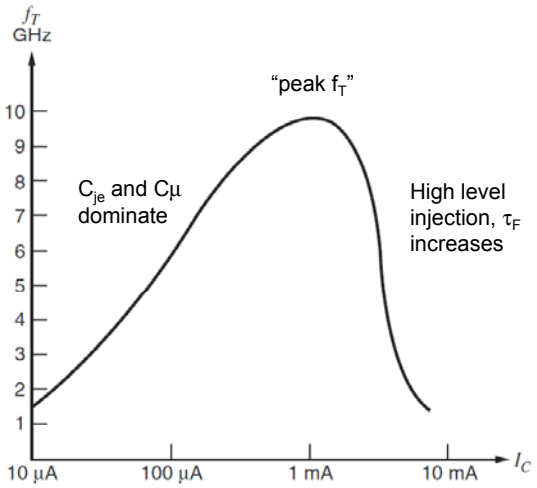
Text, p. 36

Note that r_π "matters" only for frequencies up to $\omega_\beta = \omega_T/\beta_F$

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

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

f_T versus I_C plot

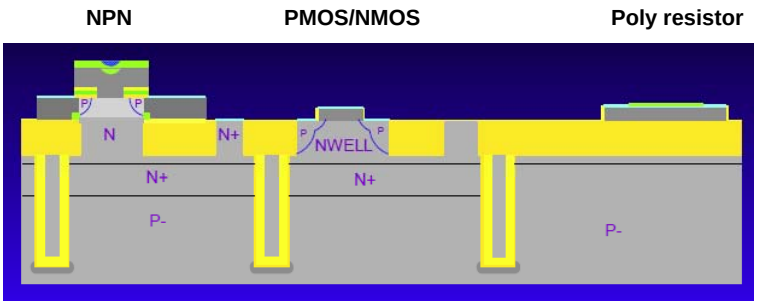


Text, p. 37

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

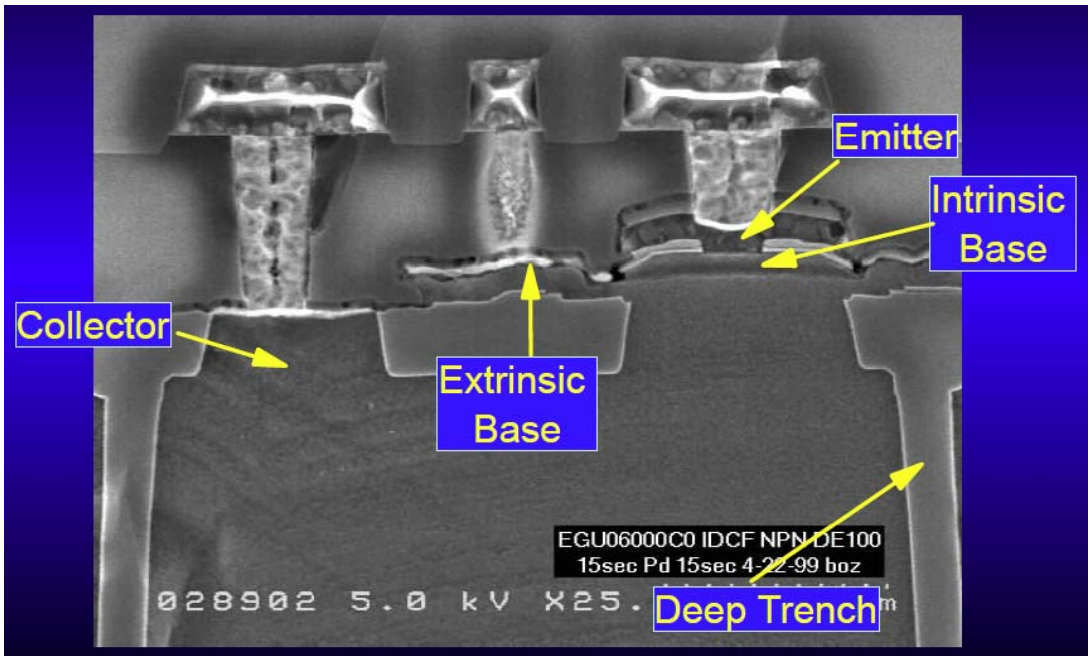
EE214 Technology

- Assumed to be similar to a 0.18-μm BiCMOS technology featuring a high-performance SiGe npn device
  - V_{CC} = 2.5V (BJT), V_{DD}=1.8V (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)

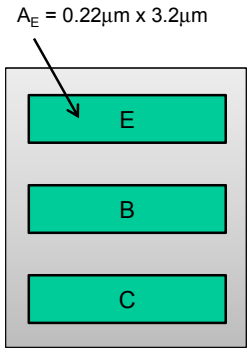


http://fuji.stanford.edu/events/spring01/slides/harameSlides.pdf

Cross Section of npn Device



EE214 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/ee214/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

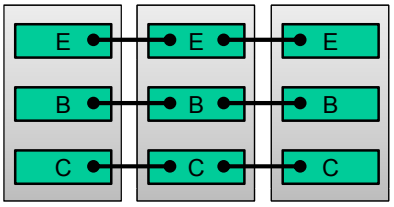
Parameter	Description
DC	BF, BR, IBC, IBE, IS, ISS, NF, NR, NS, VAF, VAR
beta degradation	ISC, ISE, NC, NE, IKF, IKR
geometric	SUBS, BULK
resistor	RB, RBM, RE, RC, IRB
junction capacitor	CJC, CJE, CJS, FC, MJC, MJE, MJS, VJC, VJE, VJS, XCJC
parasitic capacitance	CBCP, CBEP, CCSP
transit time	ITF, PTF,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

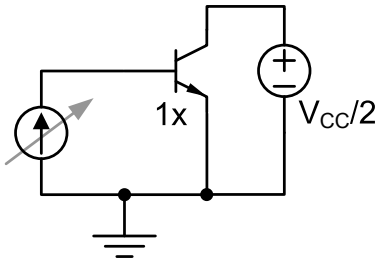


- For the unit device, there exists a practical upper bound for the collector current
  - Due to the onset of high level injection
- This means that the unit device can only deliver a certain maximum g_m
- If more g_m is needed, "m" unit devices can be connected in parallel
- Instantiation in a circuit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

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

npn Unit Device Characterization

```
* ee214 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/ee214/hspice/ee214_hspice.sp'
.end
```



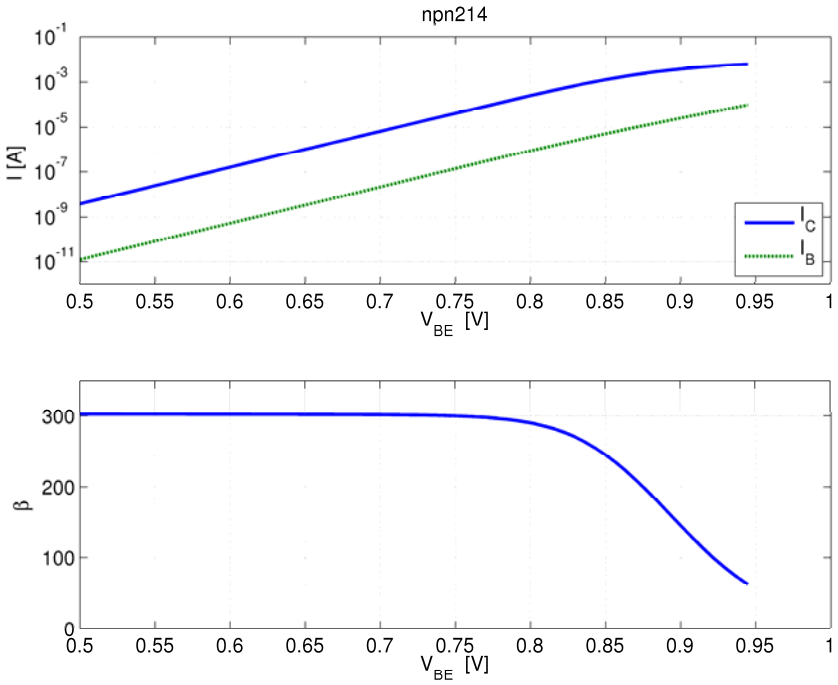
DC Operating Point Output

```
**** 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
```

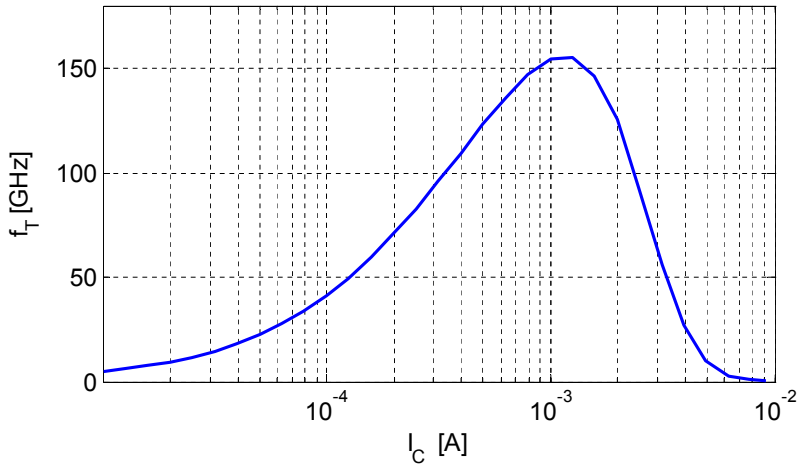
Table 5-18 BJT DC Operating Point Output

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

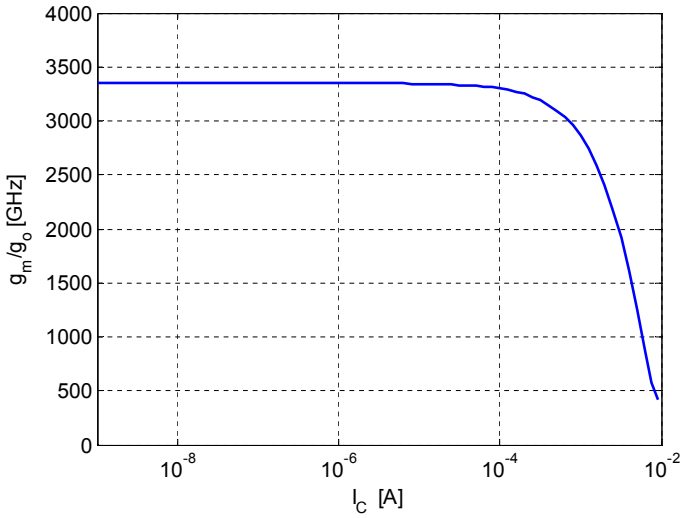
Gummel Plot



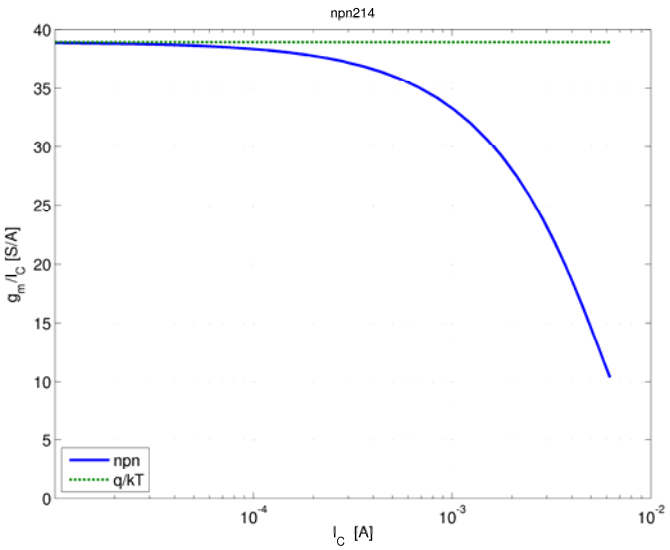
Transit Frequency



Intrinsic Gain

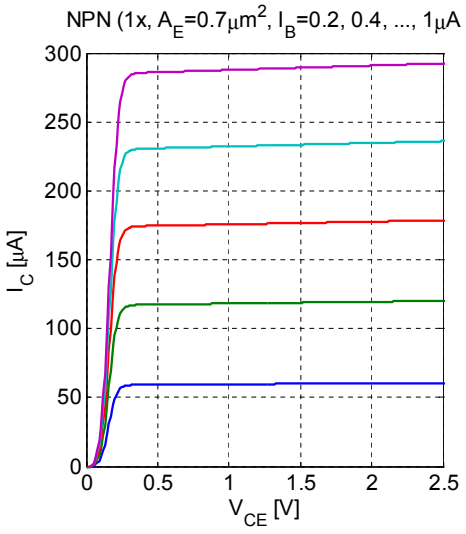


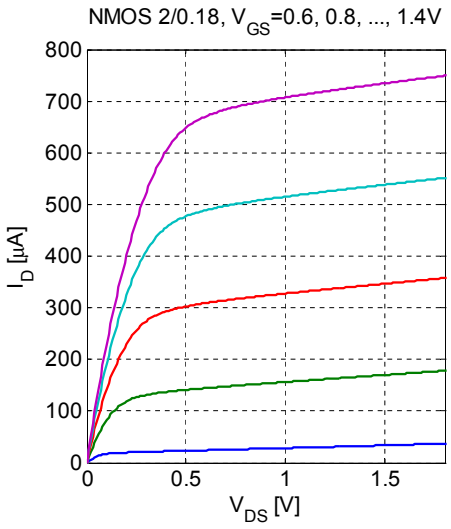
g_m/I_C



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

I-V Curves





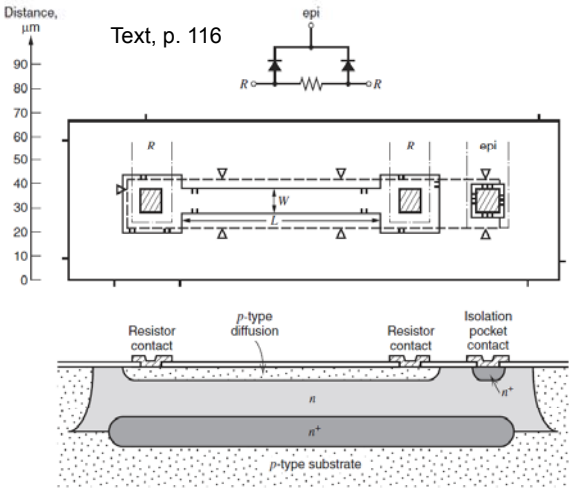
Passive Components (1)

Resistors	**Rs (O/Sq)**	**TCR (ppm/C)**
Subcollector	8.1	1430
N+ Diffusion	72	1910
P+ Diffusion	105	1430
P+ Polysilcon	270	50
P Polysilicon	1600	-1178
TaN	135	-750
Capacitors	**Cp (fF/um2)**	**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

Joseph et al., BCTM 2001

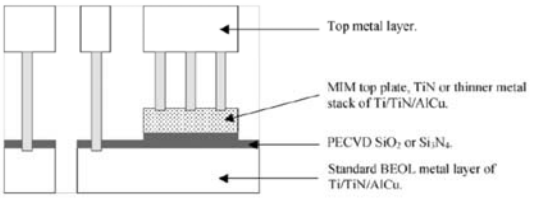
Passive Components (2)

Diffusion Resistor



Text, p. 116

MIM Capacitor



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

Chapter 3

MOS Transistor Modeling

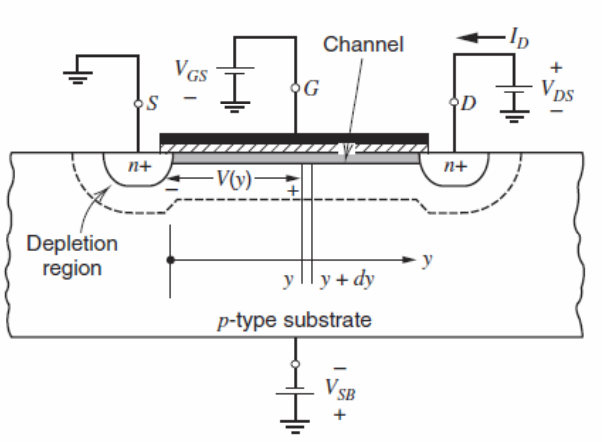
G_m/I_D-based Design

B. Murmann

Stanford University

Reading Material: Sections 1.5, 1.6, 1.7, 1.8

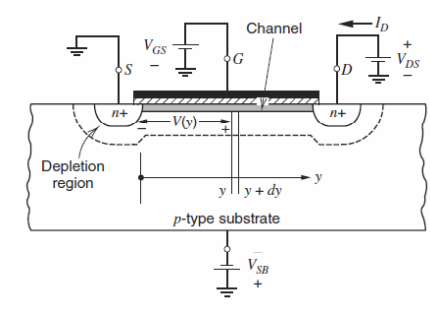
Basic MOSFET Operation (NMOS)



Text, p.41

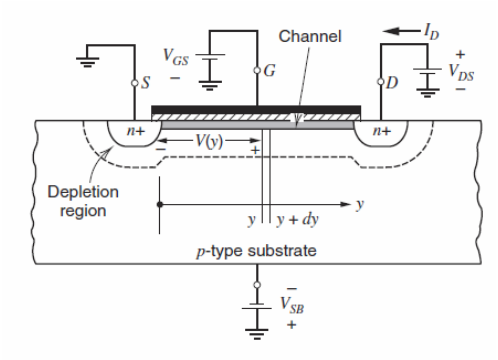
How to calculate drain current (I_D) current as a function of V_{GS}, V_{DS}?

Simplifying Assumptions



1) Current is controlled by the mobile charge in the channel. This is a very good approximation.
2) "Gradual Channel Assumption" - The vertical field sets channel charge, so we can approximate the available mobile charge through the voltage difference between the gate and the channel
3) The last and worst assumption (we will fix it later) is that the carrier velocity is proportional to lateral field ($v = \mu E$). This is equivalent to Ohm's law: velocity (current) is proportional to E-field (voltage)

Derivation of First Order IV Characteristics (1)



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

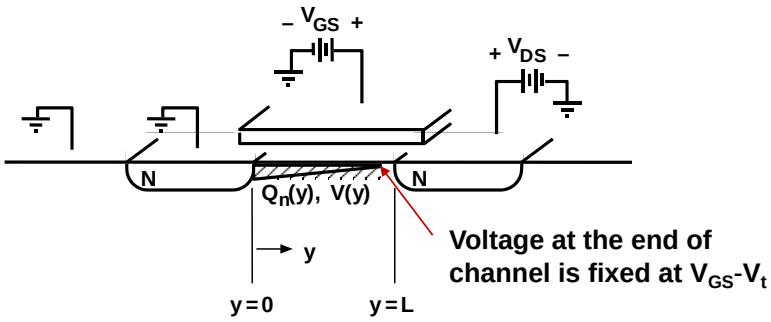
$$I_D = Q_n \cdot v \cdot W$$

$$v = \mu \cdot E$$

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

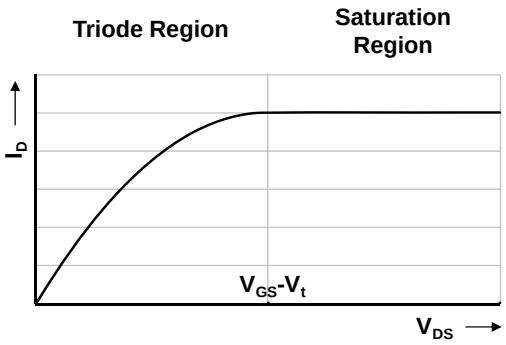
$$\boxed{I_D = \mu C_{ox} \frac{W}{L}\left[\left(V_{GS} - V_t\right) - \frac{V_{DS}}{2}\right] \cdot V_{DS}}$$

Pinch-Off



- Effective voltage across channel is V_{GS}-V_t
  - At the point where channel charge goes to zero, there is a high lateral field that sweeps the carriers to the drain
    - Recall that electrons are minority carriers in the p-region of a pn junction; they are being swept toward the n-region
  - The extra drain voltage drops across depletion region
- To first order, the current becomes independent of V_{DS}

Plot of Output Characteristic



Triode Region:

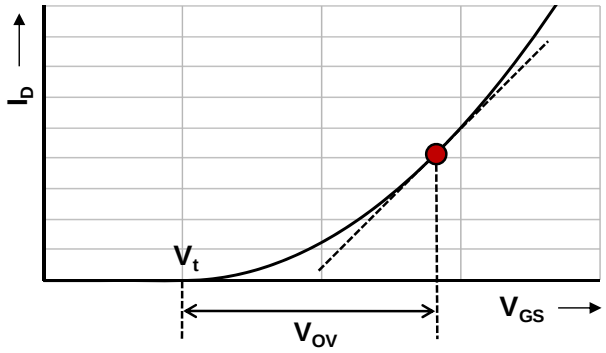
$$I_D = \mu C_{ox} \frac{W}{L}\left[(V_{GS} - V_t) - \frac{V_{DS}}{2}\right] \cdot V_{DS}$$

Saturation Region:

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

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

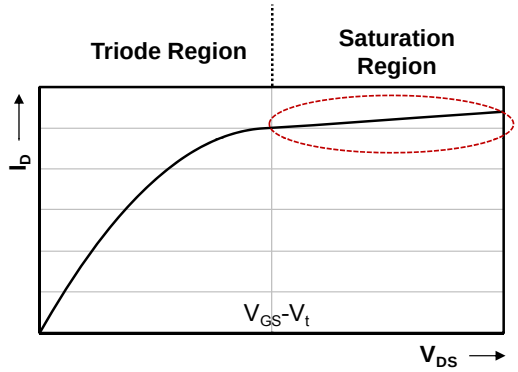
Plot of Transfer Characteristic (in Saturation)



$$g_m = \frac{dI_D}{dV_{GS}} = \mu C_{ox}\frac{W}{L}(V_{GS} - V_t) = \mu C_{ox}\frac{W}{L}V_{OV}$$

$$= \sqrt{2I_D \mu C_{ox}\frac{W}{L}} = \frac{2I_D}{V_{OV}}$$

Output Characteristic with "Channel Length Modulation"

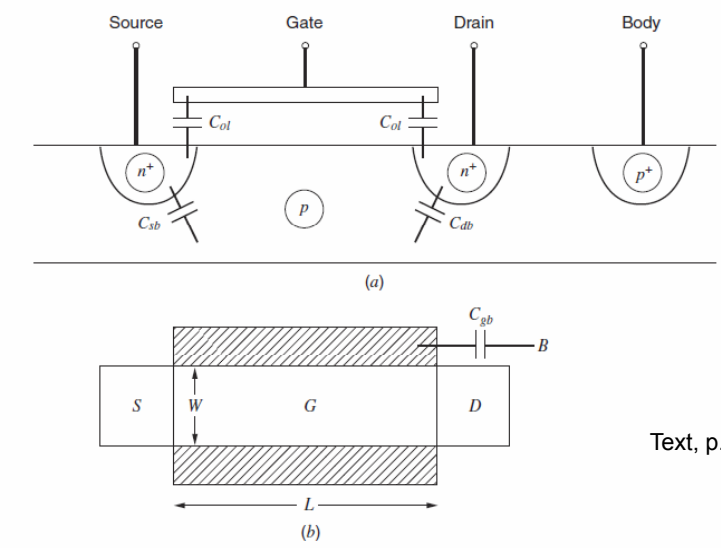


$g_o = dI_D/dV_{DS} \neq 0$

$$g_o = \frac{dI_D}{dV_{DS}} = \frac{d}{dV_{DS}}\left[\frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS} - V_t)^2(1+\lambda V_{ds})\right]$$

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

Capacitances



Text, p. 54

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

Capacitance Equations and Parameters

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$

$$C_{ol} = WC'_{ol}$$

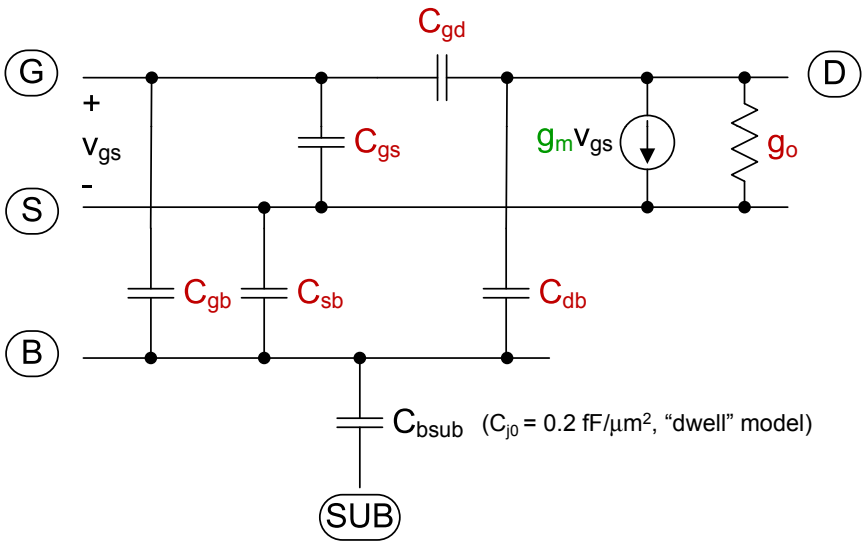
$$C_{db} = \frac{AD \cdot C_J}{\left(1 + \frac{V_{DB}}{PB}\right)^{MJ}} + \frac{PD \cdot C_{JSW}}{\left(1 + \frac{V_{DB}}{PB}\right)^{MJSW}}$$

$$AD = WL_{diff}$$

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

Parameter	EE 214 Technology (0.18µm)	
	NMOS	PMOS
C_{ox}	8.42 fF/µm²	8.42 fF/µm²
C'_{ol}	0.491 fF/µm	0.657 fF/µm
C_J	0.965 fF/µm²	1.19 fF/µm²
C_{JSW}	0.233 fF/µm	0.192 fF/µm
PB	0.8 V	0.8 V
MJ	0.38	0.40
MJSW	0.13	0.33
LDIF	0.64 µm	0.64 µm

Complete Small-Signal Model



$$C_{gg} \triangleq C_{gs} + C_{gb} + C_{gd} \qquad C_{dd} \triangleq C_{db} + C_{gd}$$

What are μC_{ox} ("KP") and λ ("LAMBDA") for our Technology?

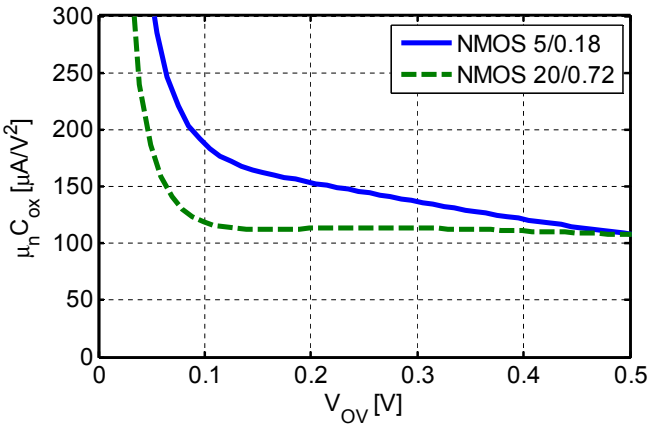
```
.MODEL nmos214 nmos
+acm      = 3              hdif     = 0.32e-6        LEVEL    = 49
+VERSION  = 3.1            TNOM     = 27             TOX      = 4.1E-9
+XJ       = 1E-7           NCH      = 2.3549E17      VTH0     = 0.3618397
+K1       = 0.5916053      K2       = 3.225139E-3    K3       = 1E-3
+K3B      = 2.3938862      W0       = 1E-7           NLX      = 1.776268E-7
+DVT0W    = 0              DVT1W    = 0              DVT2W    = 0
+DVT0     = 1.3127368      DVT1     = 0.3876801      DVT2     = 0.0238708
+U0       = 256.74093      UA       = -1.585658E-9   UB       = 2.528203E-18
+UC       = 5.182125E-11   VSAT     = 1.003268E5     A0       = 1.981392
+AGS      = 0.4347252      B0       = 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
+WR       = 1              WINT     = 0              LINT     = 1.617316E-8
+XL       = 0              XW       = -1E-8          DWG      = -5.383413E-9
+DWB      = 9.111767E-9    VOFF     = -0.0854824     NFACTOR  = 2.2420572
+CIT      = 0              CDSC     = 2.4E-4         CDSCD    = 0
+CDSCB    = 0              ETA0     = 2.981159E-3    ETAB     = 9.289544E-6
+DSUB     = 0.0159753      PCLM     = 0.7245546      PDIBLC1  = 0.1568183
+PDIBLC2  = 2.543351E-3    PDIBLCB  = -0.1           DROUT    = 0.7445011
+PSCBE1   = 8E10           PSCBE2   = 1.876443E-9    PVAG     = 7.200284E-3
+DELTA    = 0.01           RSH      = 6.6            MOBMOD   = 1
+PRT      = 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
+WWN      = 1              WWL      = 0              LL       = 0
+LLN      = 1              LW       = 0              LWN      = 1
+LWL      = 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              PVTH0    = -7.714081E-4   PRDSW    = -2.5827257
+PK2      = 9.619963E-4    WKETA    = -1.060423E-4   LKETA    = -5.373522E-3
+PU0      = 4.5760891      PUA      = 1.469028E-14   PUB      = 1.783193E-23
+PVSAT    = 1.19774E3      PETA0    = 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
- BSIM 3v3 model
- 110 parameters
- KP and LAMBDA nowhere to be found…

An Attempt to Extract μC_{ox}

- Bias MOSFET at constant $V_{DS}>V_{OV}$, sweep V_{GS} and plot μC_{ox} estimate

$$\mu C_{ox} = \frac{2I_D}{\frac{W}{L}V_{OV}^2}$$



- The extracted μC_{ox} depends on L and V_{OV} and cannot be viewed as a constant parameter

Questions

- Which physical effects explain the large deviation from the basic square law model?
- How can we design with such a device?
  - Is there another “simple” equation that describes its behavior?
- We will approach the above two questions by performing a systematic, simulation-based device characterization
  - And discuss the relevant physical phenomena that explain the observed behavior
- As a basis for this characterization, we consider three basic figures of merit that relate directly to circuit design

Figures of Merit for Device Characterization

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}

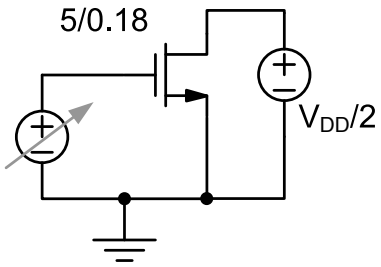
$$\boxed{\frac{g_m}{C_{gg}}} \cong \frac{3}{2}\frac{\mu V_{OV}}{L^2}$$

- Intrinsic gain
  - Want large g_m, but no g_o

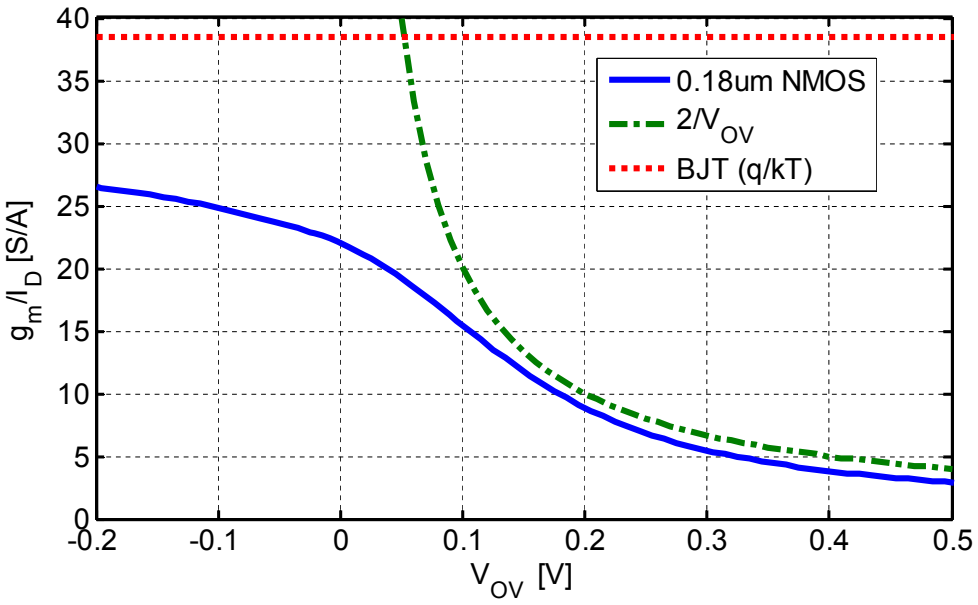
$$\boxed{\frac{g_m}{g_o}} \cong \frac{2}{\lambda V_{OV}}$$

Device Characterization

```
* NMOS characterization

.param  gs=0.7
.param  dd=1.8
vds     d 0       dc      'dd/2'
vgs     g 0       dc      'gs'
mn      d g 0 0   nmos214  L=0.18um  W=5um

.op
.dc gs 0.2V 1V 10mV
.probe ov      = par('gs-vth(mn)')
.probe gm_id   = par('gmo(mn)/i(mn)')
.probe ft      = par('1/6.28*gmo(mn)/cggbo(mn)')
.probe gm_gds  = par('gmo(mn)/gdso(mn)')

.options post brief dccap
.inc /usr/class/ee214/hspice/ee214_hspice.sp
.end
```



g_m/I_D Plot

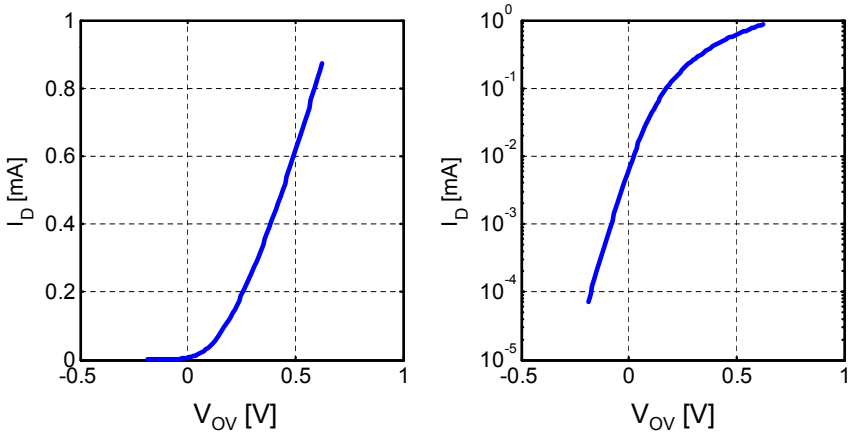


Observations

- Square law prediction is fairly close for $V_{OV} > 150mV$
- Unfortunately g_m/I_D does not approach infinity for $V_{OV} \rightarrow 0$
- It also seems that we cannot do better than a BJT, even though the square law equation would predict this for $0 < V_{OV} < 2kT/q \cong 52mV$
- For further analysis, it helps to identify three distinct operating regions
  - Strong inversion: $V_{OV} > 150mV$
    - Deviations due to short channel effects
  - Subthreshold: $V_{OV} < 0$
    - Behavior similar to a BJT, g_m/I_D nearly constant
  - Moderate Inversion: $0 < V_{OV} < 150mV$
    - Transition region, an interesting mix of the above

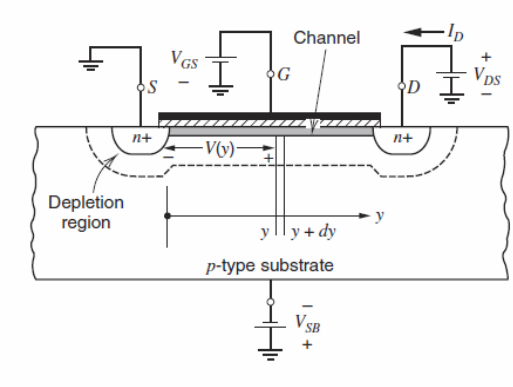
Subthreshold Operation

- A plot of the device current in our previous simulation



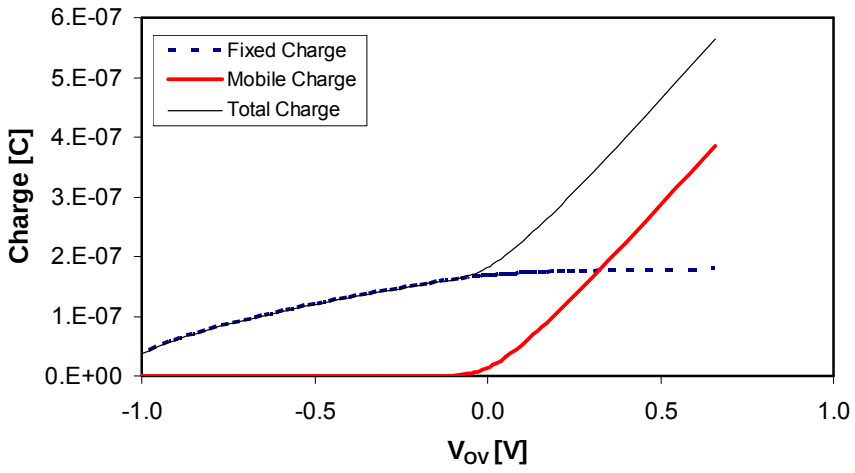
- Questions
  - What determines the current when $V_{OV} < 0$, i.e. $V_{GS} < V_t$?
  - What is the definition of V_t?

Definition of V_t



- V_t is defined as the V_{GS} at which the number of electrons at the surface equals the number of doping atoms
- Seems somewhat arbitrary, but makes sense in terms of surface charge control

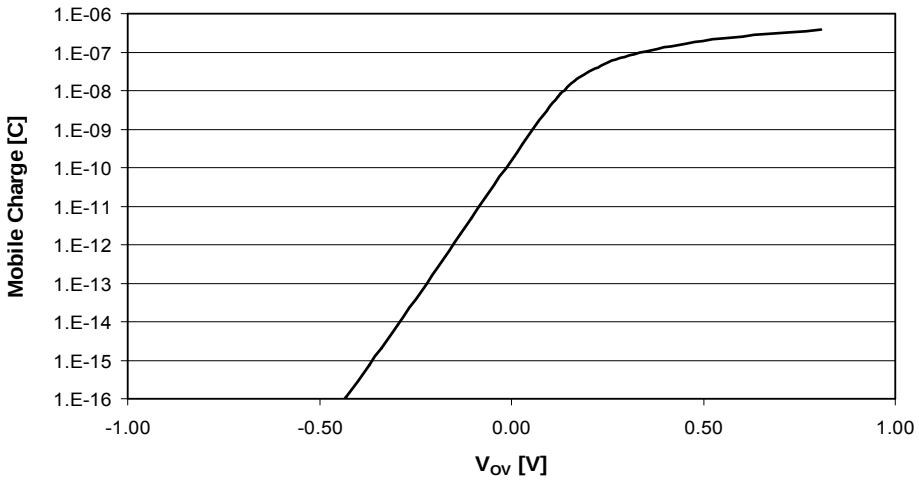
Mobile Charge versus V_{OV}



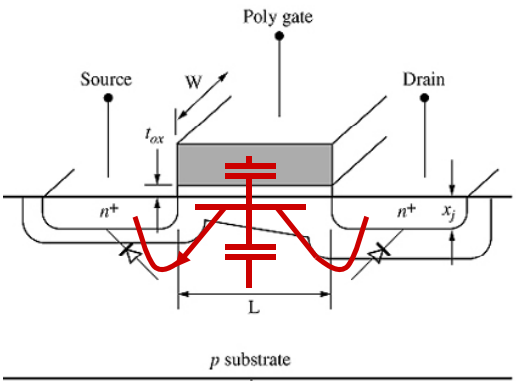
- Around $V_{GS}=V_t$ ($V_{OV}=0$), the relationship between mobile charge in the channel and gate voltage becomes linear ($Q_n \sim C_{ox}V_{OV}$)
  - Exactly what we assumed to derive the long channel model

Mobile Charge on a Log Scale

- On a log scale, we see that there are mobile charges **before** we reach the threshold voltage
  - Fundamental result of solid-state physics, not short channels



BJT Similarity



- We have
  - An NPN sandwich, mobile minority carriers in the P region
- This is a BJT!
  - Except that the base potential is here controlled through a capacitive divider, and not directly by an electrode

Subthreshold Current

- We know that for a BJT

$$I_C = I_S \cdot e^{\frac{V_{BE}}{V_T}} \qquad V_T = \frac{kT}{q}$$

- For the MOSFET in subthreshold we have

$$I_D = I_0 \cdot e^{\frac{V_{GS}-V_t}{nV_T}}$$

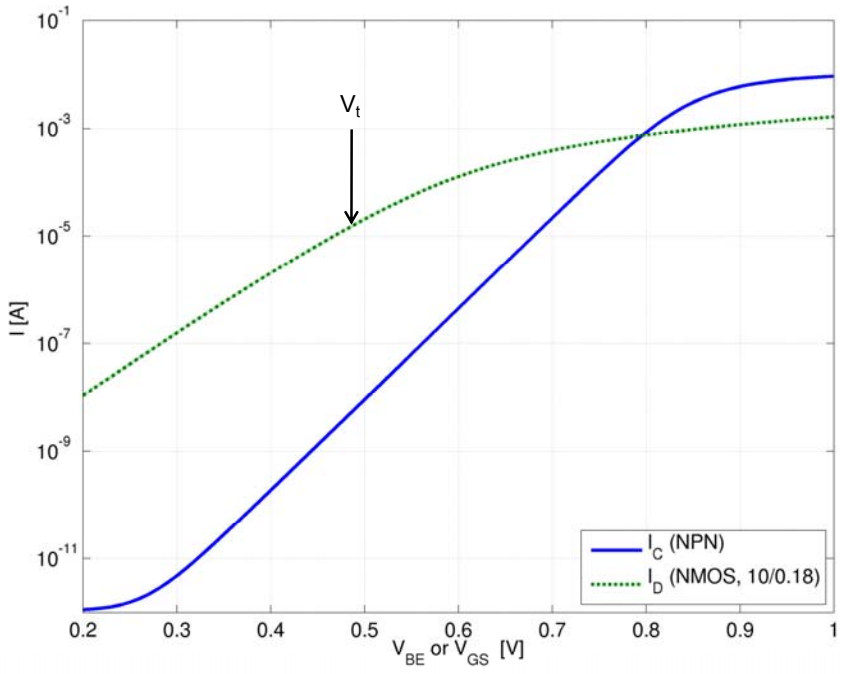
- n is given by the capacitive divider

$$n = \frac{C_{js} + C_{ox}}{C_{ox}} = 1 + \frac{C_{js}}{C_{ox}}$$

where C_{js} is the depletion layer capacitance

- In the EE214 technology n ≅ 1.5

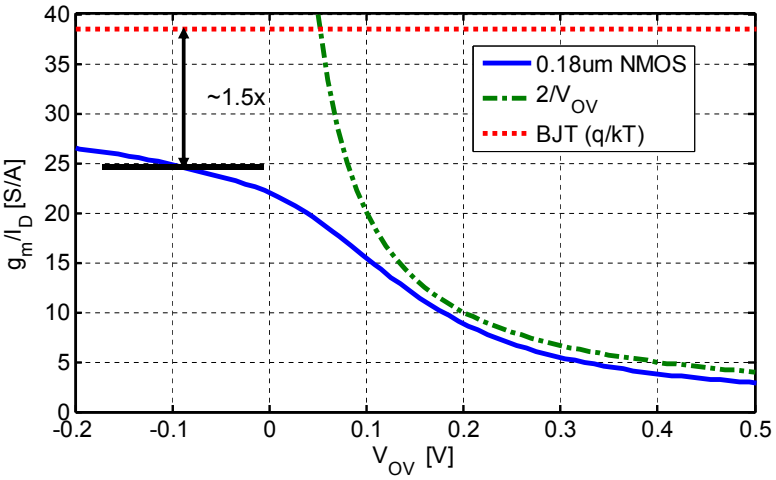
Comparison – NMOS versus NPN



Subthreshold Transconductance

$$g_m = \frac{dI_D}{dV_{GS}} = \frac{1}{n}\frac{I_D}{V_T} \qquad \frac{g_m}{I_D} = \frac{1}{nV_T}$$

- Similar to BJT, but unfortunately n (≅1.5) times lower



Moderate Inversion

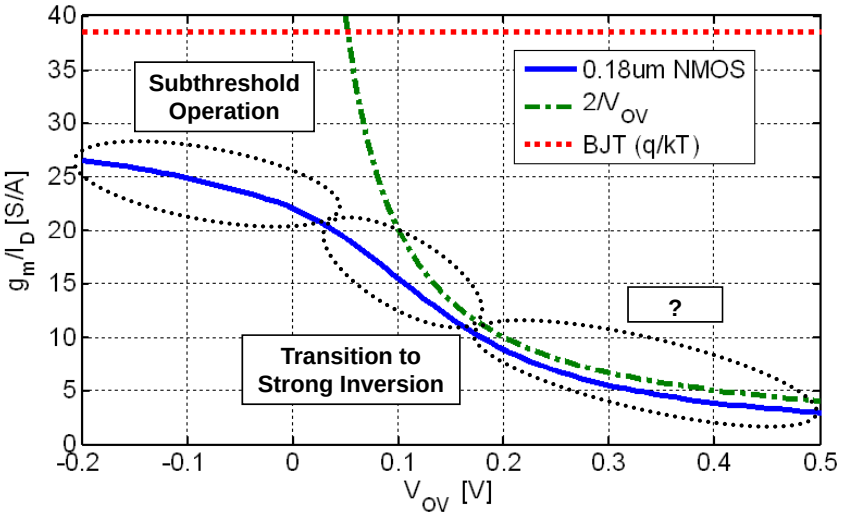
- In the transition region between subthreshold and strong inversion, we have two different current mechanisms

Drift (MOS) $\quad \nu = \mu E$

Diffusion (BJT) $\quad \nu = D\frac{dn}{dx} = \frac{kT}{q}\mu\frac{dn}{dx}$

- Both current components are always present
  - Neither one clearly dominates in moderate inversion
- Can show that ratio of drift/diffusion current ~$(V_{GS}-V_t)/(kT/q)$
  - MOS equation becomes dominant at several kT/q

Re-cap



- What causes the discrepancy between $2/V_{OV}$ and 0.18µm NMOS in strong inversion?

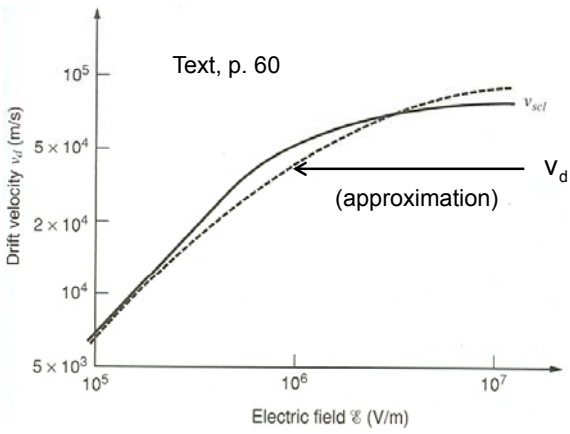
Short Channel Effects

- Velocity saturation due to high lateral field
- Mobility degradation due to high vertical field
- V_t dependence on channel length and width
- $V_t = f(V_{DS})$
- $r_o = f(V_{DS})$
- …

- We will limit the discussion in EE214 to the first two aspects of the above list, with a focus on qualitative understanding

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 ($v_{scl} \cong 10^5$ m/s)
  - At very high fields (high voltage drop across the conductive channel), the carrier velocity saturates



(approximation)

$$v_d(E) \cong \frac{\mu E}{1+\frac{E}{E_c}} \cong \mu E_c = v_{scl} \quad \text{for} \quad E >> E_c$$

$$= \frac{v_{scl}}{2} \quad \text{for} \quad E = E_c$$

Velocity Saturation (2)

- It is important to distinguish various regions in the above plot
  - Low field, the long channel equations still hold
  - Moderate field, the long channel equations become 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 long channel physics
- In saturation, the lateral field across the channel is

$$E = \frac{V_{OV}}{L} \quad \text{e.g.} \quad \frac{200\text{mV}}{0.18\mu\text{m}} = 1.11\cdot10^6 \frac{\text{V}}{\text{m}}$$

Field Estimates

- In our 0.18μm technology, we have for an NMOS device

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

Therefore

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

- This means that for V_{OV} on the order of 0.2V, the carrier velocity is somewhat reduced, but the impairment is relatively small
- The situation changes when much larger V_{OV} are applied, e.g. 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 (see text, p. 62)

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

$$\cong \frac{1}{2}\mu C_{ox} \frac{W}{L} V_{OV} \cdot \underbrace{\frac{E_c L \cdot V_{OV}}{(E_c L + V_{OV})}}$$

Think of this as a “parallel combination"

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

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

Modified g_m/I_D Expression

- Assuming $V_{OV} << E_cL$, we can show that (see text, pp. 63-64)

$$\frac{g_m}{I_D} \cong \frac{2}{V_{OV}} \cdot \frac{1}{\left(1+\frac{V_{OV}}{E_cL}\right)}$$

- E.g. for an NMOS device with V_{OV}=200mV

$$\frac{g_m}{I_D} \cong \frac{2}{V_{OV}} \cdot \frac{1}{\left(1+\frac{0.2}{1.2}\right)} = \frac{2}{V_{OV}} \cdot 0.86$$

- Means that the square law model in strong inversion (at $V_{OV} \cong 200mV$) should be off by about 15%
  - This prediction agrees well with the simulation data

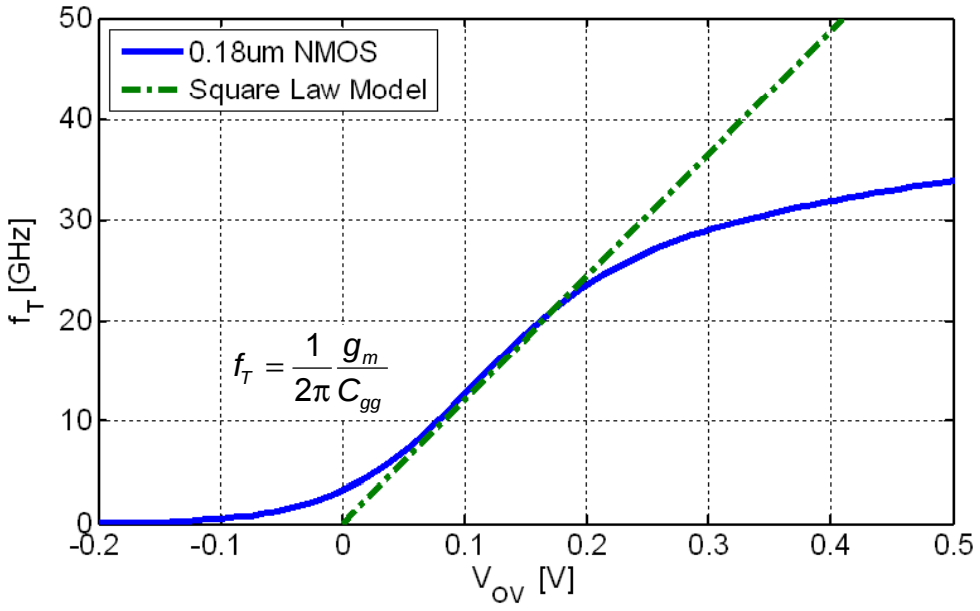
Mobility Degradation due to Vertical Field

- 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 increases and tries to pull the carriers closer to the "dirty" silicon surface
  - Imperfections impede movement and thus mobility
- This effect can be included by replacing the mobility term with an "effective mobility"

$$\mu_{eff} \cong \frac{\mu}{(1+\theta V_{OV})} \qquad \theta = 0.1...0.4\,\frac{1}{V}$$

- Yet another "fudge factor"
  - Possible to lump with E_cL parameter, if desired

Transit Frequency Plot

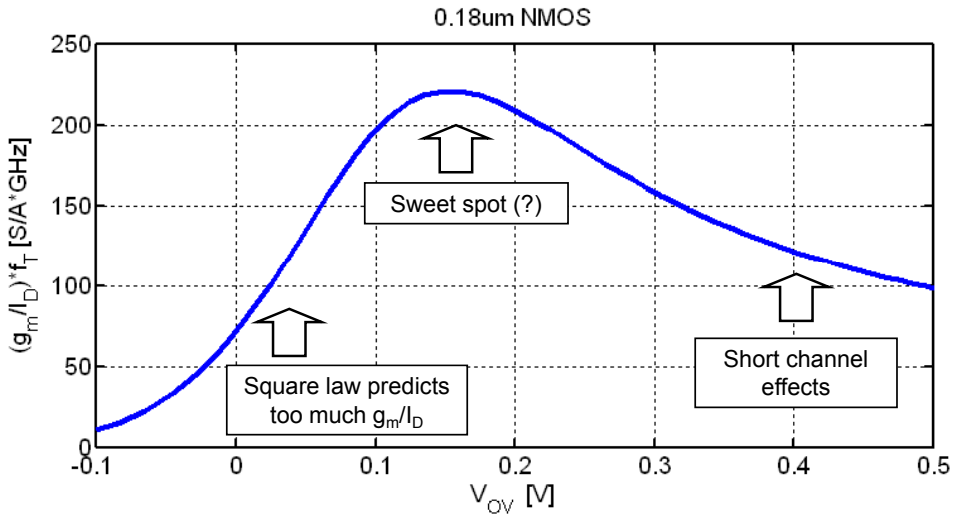


Observations - f_T

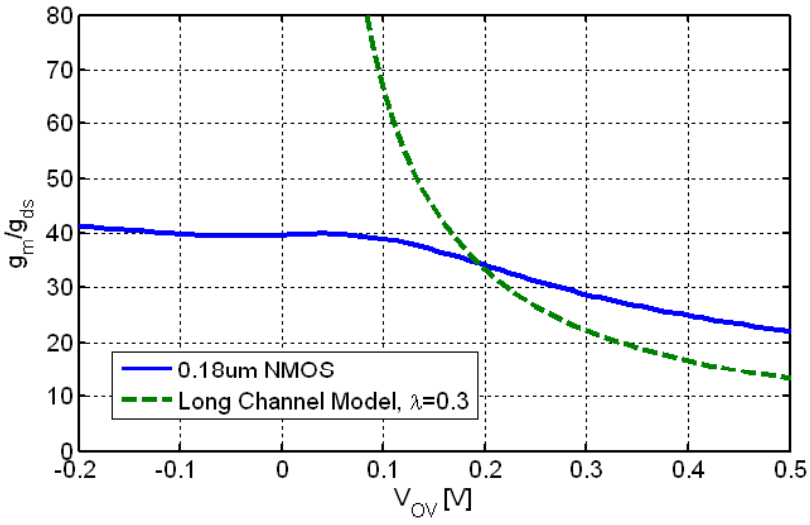
- Again the square-law model doesn't do a very good job
  - Large f_T discrepancy in subthreshold operation and in strong inversion (large V_{OV})
- The reasons for these discrepancies are exactly the same as the ones we came across when looking at g_m/I_D
  - Bipolar action in subthreshold operation and moderate inversion
  - Short channel effects at large V_{OV}
    - Less g_m, hence lower g_m/C_{gg}
- Same conclusion: we won't be able to make good predictions with a simple square law relationship

$g_m/I_D \cdot f_T$ Plot

Square Law: $\frac{g_m}{I_D} \cdot f_T \cong \frac{1}{2\pi} \frac{3\mu}{L^2}$

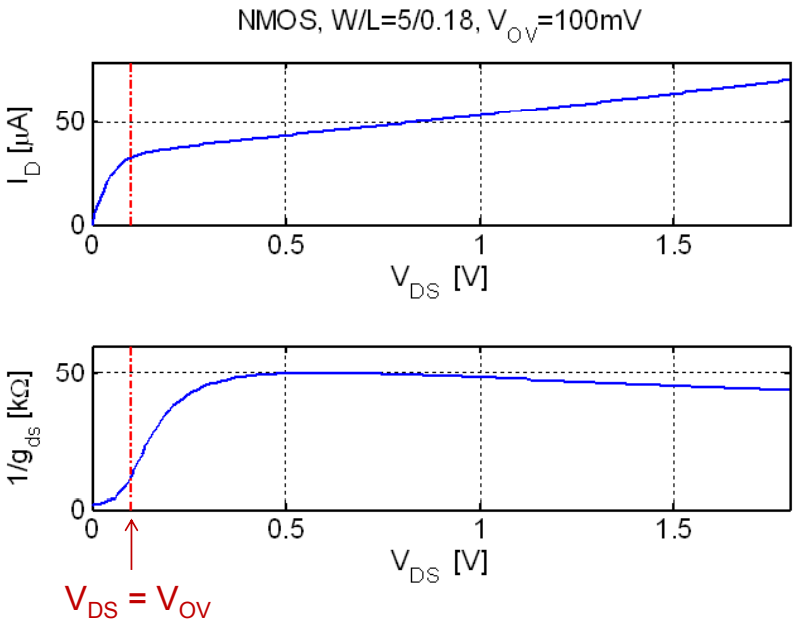


Intrinsic Gain Plot

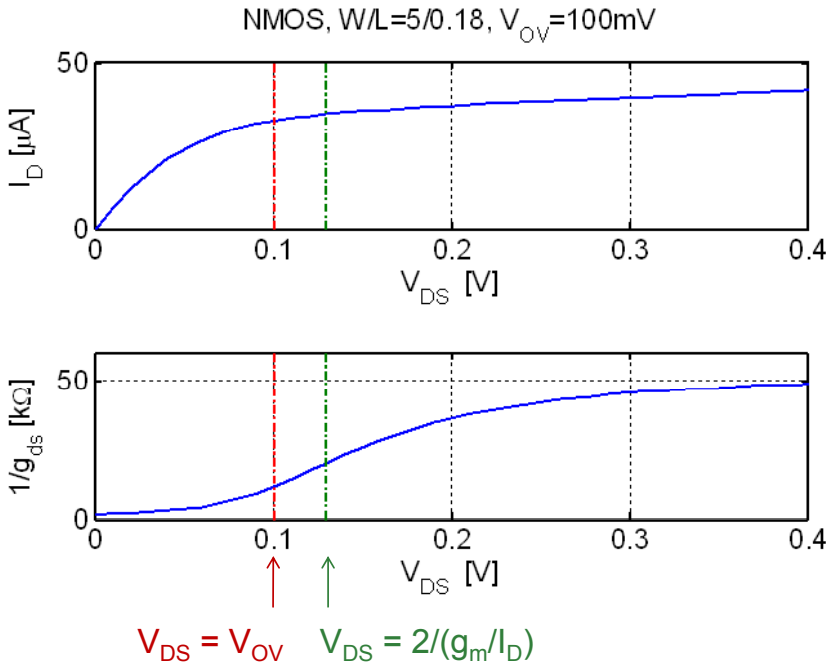


- Impossible to approximate with the "λ" model equation!

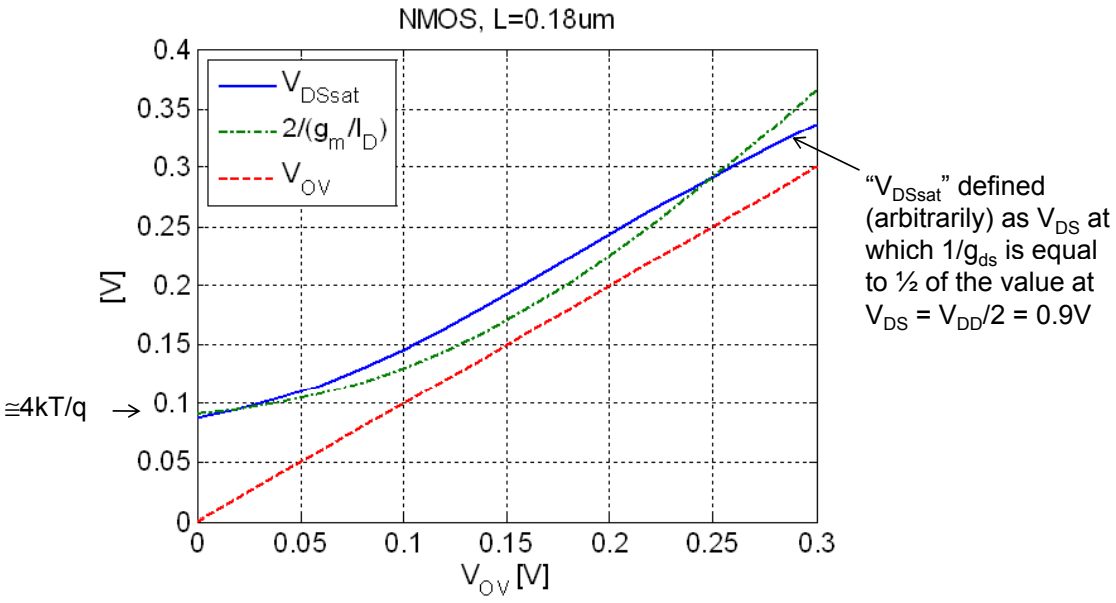
Gradual Onset of $1/g_{ds}$



Gradual Onset of $1/g_{ds}$ (Zoom)



"V_{DSsat}" Estimate Based on g_m/I_D



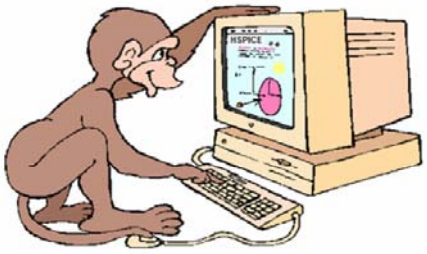
- $V^* = 2/(g_m/I_D)$ is a reasonable estimate of "V_{DSsat}"

Observations – Intrinsic Gain

- Device shows a rather gradual transition from triode to saturation
  - Square law predicts an abrupt change from small to large intrinsic gain at $V_{DS} = V_{OV}$
  - $V^* = 2/(g_m/I_D)$ provides a reasonable estimate for the minimum V_{DS} that is needed to extract gain from a device
    - Typically want to stay at least 100mV above this value in practical designs
- The physics that govern the behavior of $r_o=1/g_{ds}$ are complex
  - Channel length modulation
  - Drain induced barrier lowering (DIBL)
  - Substrate current induced body effect (SCBE)
    - Not present in all technologies and/or PMOS devices
- If you are interested in more details, please refer to EE316 or similar

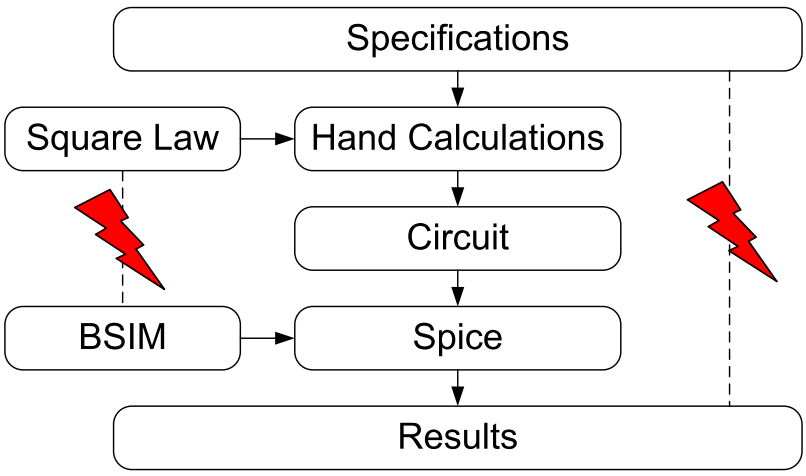
The Challenge

- Square-law model is inadequate for design in fine-line CMOS
  - But simulation models (BSIM, PSP, …) are too complex for hand-calculations
- This issue tends to drive many designers 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
- Our goal
  - Maintain a systematic design methodology in absence of a set of compact MOSFET equations
- Strategy
  - Design using look-up tables or charts

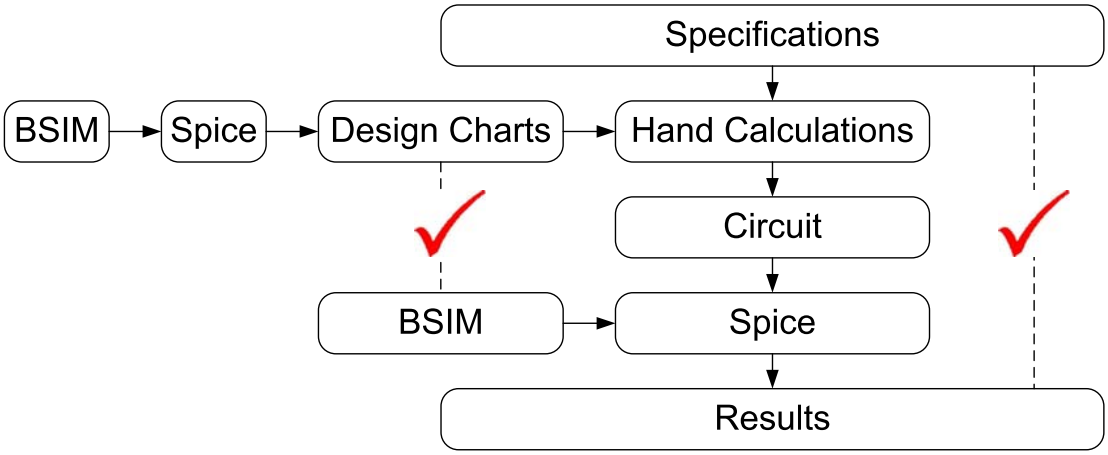


[Courtesy Isaac Martinez]

The Problem



The Solution

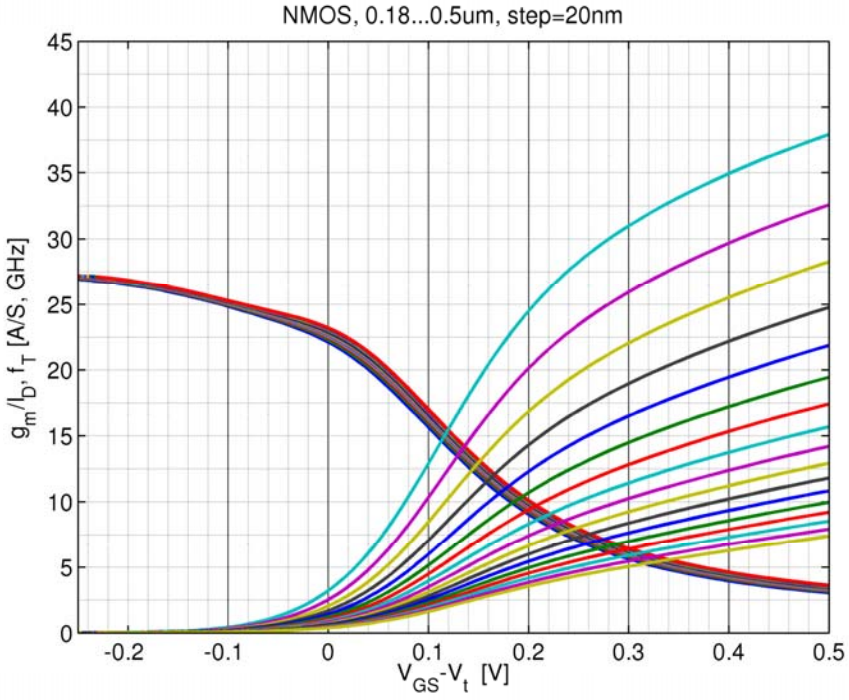


- Use pre-computed spice data in hand calculations

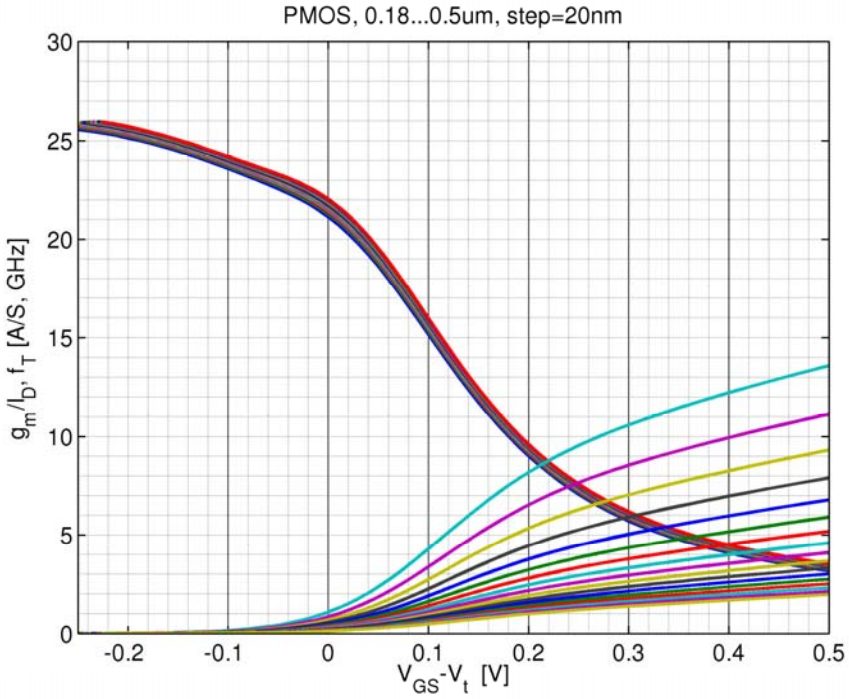
Technology Characterization for Design

- Plot the following parameters for a reasonable range of g_m/I_D and channel lengths
  - Transit frequency (f_T)
  - Intrinsic gain (g_m/g_{ds})
  - Current density (I_D/W)
- In addition, may want to tabulate relative estimates of extrinsic capacitances
  - C_{gd}/C_{gg} and C_{dd}/C_{gg}
- Parameters are (to first order) independent of device width
  - Enables "normalized design" and re-use of charts
  - Somewhat similar to filter design procedure using normalized coefficient tables
- Do hand calculations using the generated technology data
  - Can use Matlab functions to do table-look-up on pre-computed data

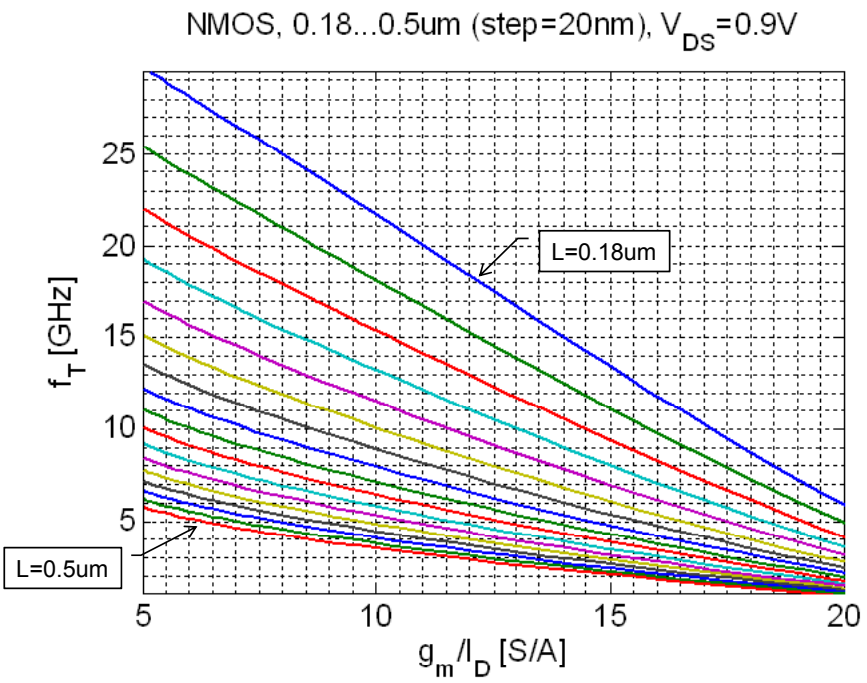
NMOS Simulation Data



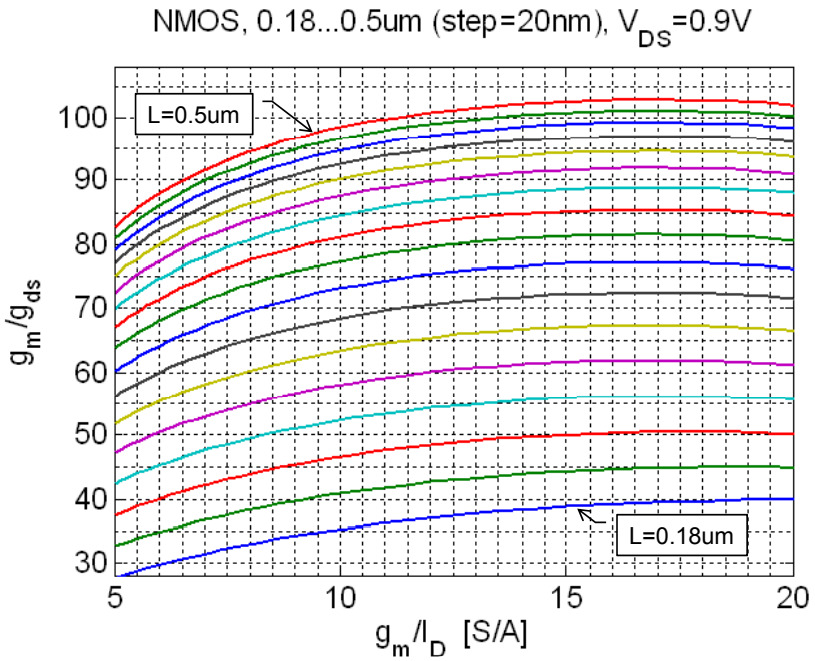
PMOS Simulation Data



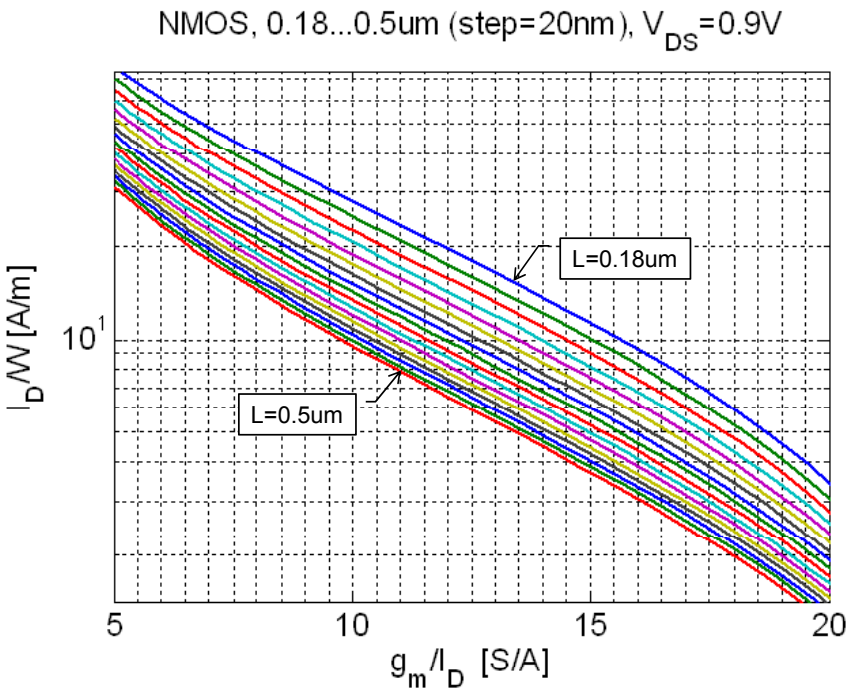
Transit Frequency Chart



Intrinsic Gain Chart



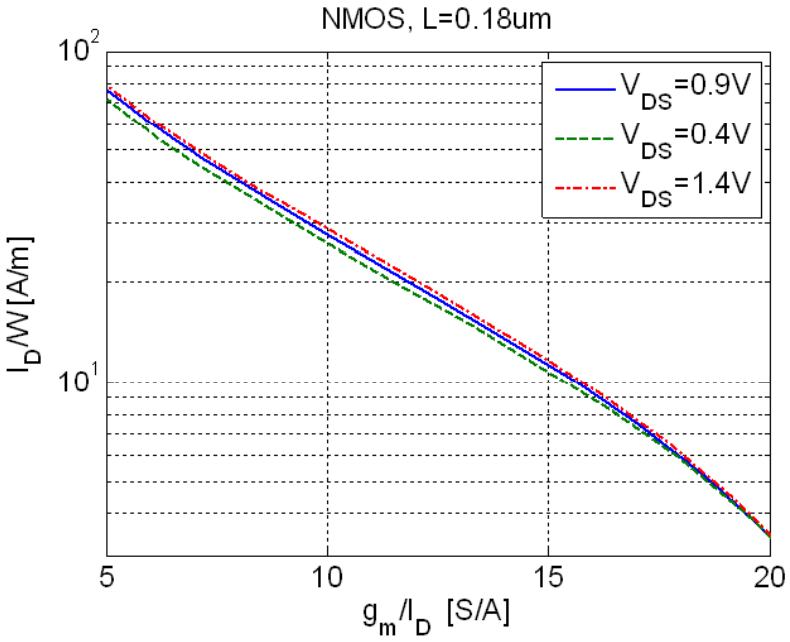
Current Density Chart



Lookup Functions in Matlab

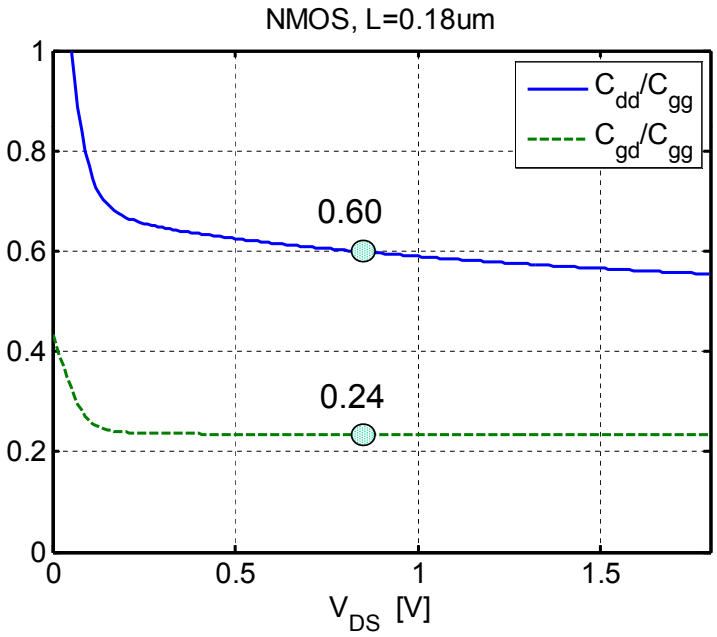
```
% Set up path and load simulation data (for VDS=0.9V)
addpath('/usr/class/ee214/matlab');
load techchar.mat;

% Lookup fT for NMOS, L=0.18um, at gm/ID=10S/A
lookup_ft(tech, 'n', 0.18e-6, 10)
ans =   2.2777e+10

% Lookup gm/ID for NMOS, L=0.18um, at fT=20GHz
lookup_gmid(tech, 'n', 0.18e-6, 20e9)
ans =   11.5367

% Lookup ID/W for NMOS, L=0.18um, at gm/ID=10S/A
lookup_idw(tech, 'n', 0.18e-6, 10)
ans =   29.3281
```

V_{DS} Dependence



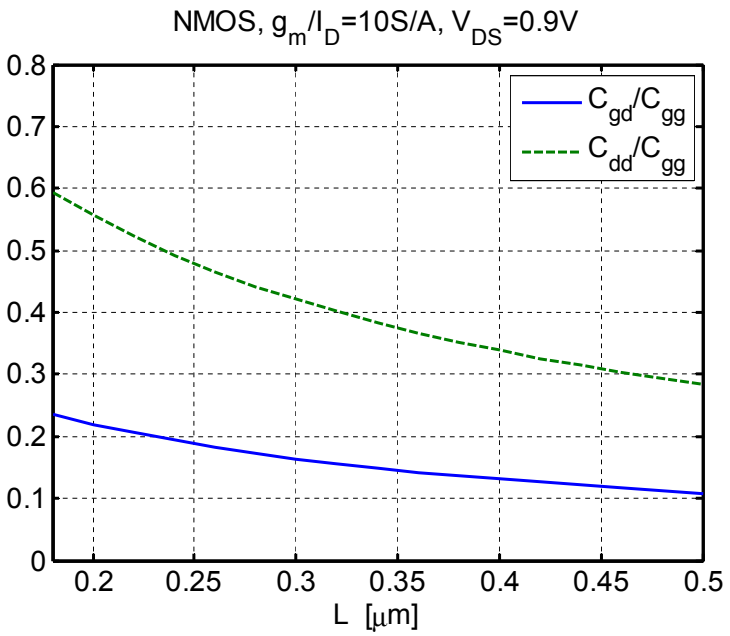
- V_{DS} dependence is relatively weak
- Typically OK to work with data generated for $V_{DD}/2$

Extrinsic Capacitances (1)

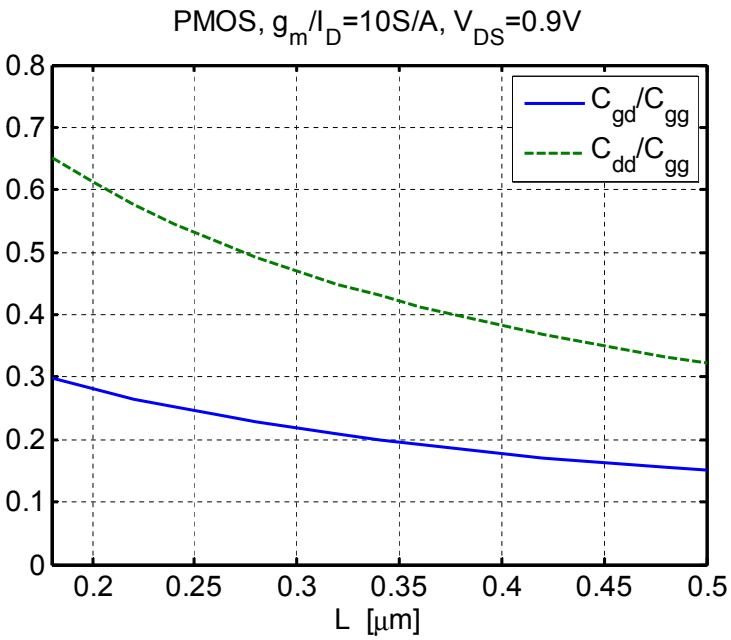


- Again, usually OK to work with estimates taken at $V_{DD}/2$

Extrinsic Capacitances (2)



Extrinsic Capacitances (3)

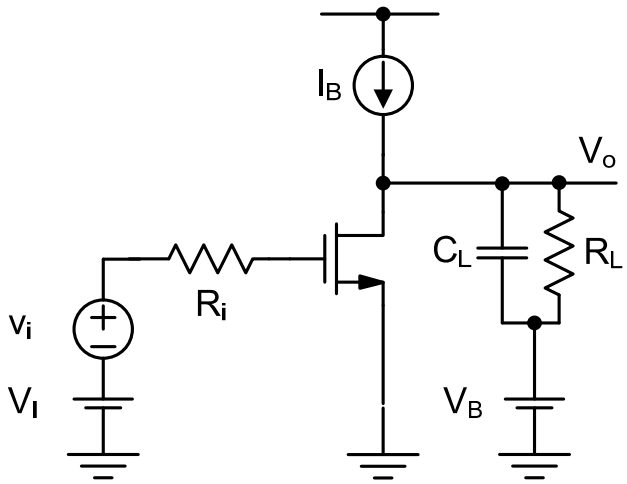


Generic Design Flow

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

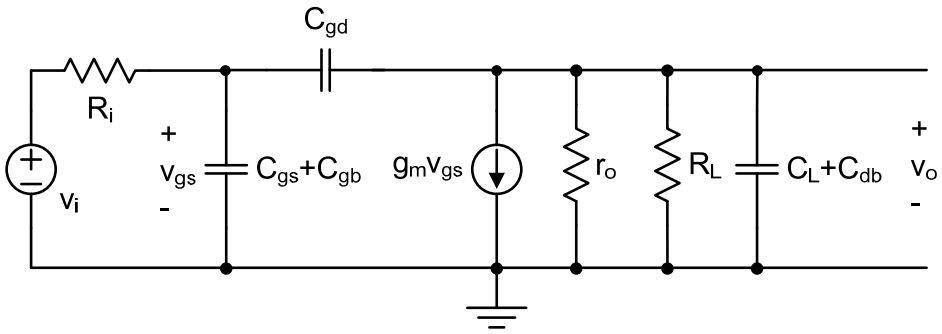
Many other possibilities exist (depending on circuit specifics, design constraints and objectives)

Basic Design Example



Given specifications and objectives
- 0.18µm technology
- DC gain = -4
- R_L=1k, C_L=50fF, R_i=10k
- Maximize bandwidth while keeping $I_B \leq 300$uA
  - Implies $L=L_{min}$=0.18um
- Determine device width
- Estimate dominant and non-dominant pole

Small-Signal Model



Calculate g_m and g_m/I_D

$$|A_v(0)| \cong g_m R_L = 4 \quad \Rightarrow \quad g_m = \frac{4}{1k\Omega} = 4mS$$

$$\boxed{\frac{g_m}{I_D} = \frac{4mS}{300\mu A} = 13.3\frac{S}{A}}$$

Why can we Neglect r_o?

$$|A_v(0)| = g_m (R_L \,||\, r_o)$$

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

$$\frac{1}{|A_v(0)|} = \frac{1}{g_m R_L} + \frac{1}{g_m r_o}$$

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

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

Zero and Pole Expressions

High frequency zero (negligible)

$$\omega_z \cong \frac{g_m}{C_{gd}} >> \omega_T$$

Dominant pole (see Chapter 4)

$$\omega_{p1} \cong \frac{1}{R_i\left[C_{gs} + C_{gb} + (1 + g_m R_L)\cdot C_{gd}\right]}$$

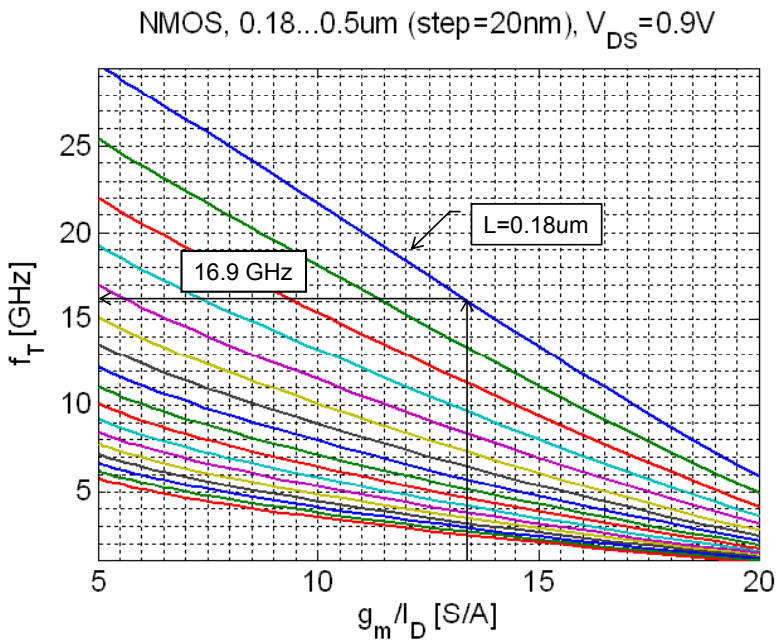
Nondominant pole (see Chapter 4)

$$\omega_{p2} \cong \frac{1}{\omega_{p1}} \frac{1}{R_i R_L\left(\left[C_{gs} + C_{gb}\right]C_L + \left[C_{gs} + C_{gd}\right]C_{db} + C_L C_{gd}\right)}$$

Calculation of capacitances from tabulated parameters:

$$C_{gs} + C_{gb} = C_{gg} - C_{gd} \qquad C_{db} = C_{dd} - C_{gd}$$

Determine C_{gg} via f_T Look-up



Find Capacitances and Plug in

$$C_{gg} = \frac{1}{2\pi}\frac{g_m}{f_T} = \frac{1}{2\pi}\frac{4\text{mS}}{16.9\text{GHz}} = 37.7\text{fF}$$

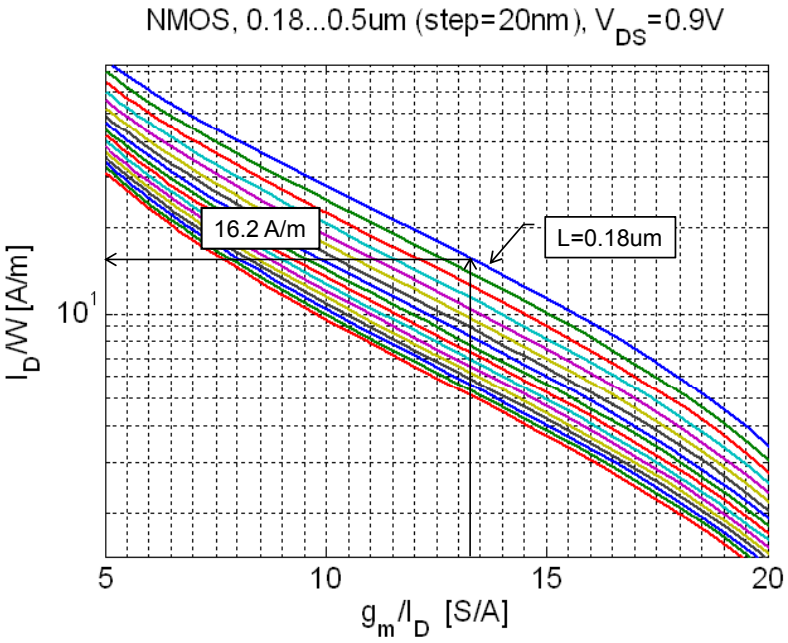
$$C_{gd} = \frac{C_{gd}}{C_{gg}}C_{gg} = 0.24 \cdot 37.7\text{fF} = 9.0\text{fF}$$

$$C_{dd} = \frac{C_{dd}}{C_{gg}}C_{gg} = 0.60 \cdot 39.4\text{fF} = 23.6\text{fF}$$

$\therefore f_{p1} \cong 196\ \text{MHz}$

$\therefore f_{p2} \cong 6.0\ \text{GHz}$

Device Sizing

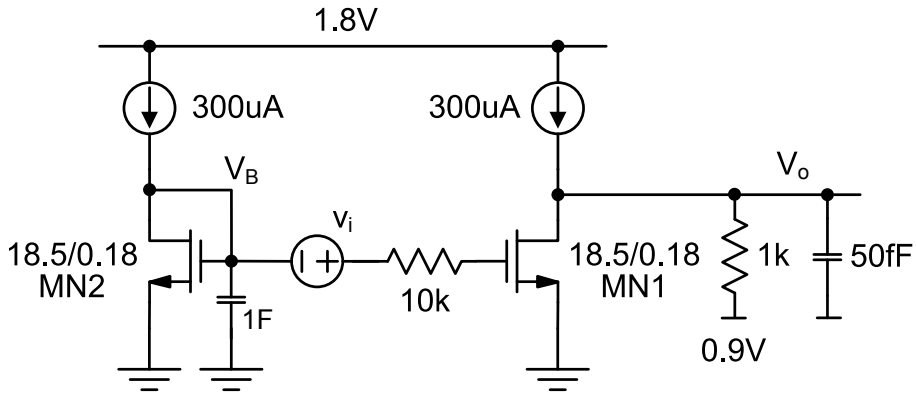


Circuit For Spice Verification

Device width

$$W = \frac{I_D}{\frac{I_D}{W}} = \frac{300\mu A}{16.2A/m} = 18.5\mu m$$

Simulation circuit



Simulated DC Operating Point

```
element   0:mn1
region    Saturati
 id       326.8330u
 vgs      624.9116m
 vds      873.1670m
 vdsat    113.7463m
 vod      138.5878m
 gm         4.1668m
 gds      108.2225u
 ...
 cdtot     21.8712f
 cgtot     37.6938f
 cgd        8.9163f
 ...
 gm/ID     12.8
```

Calculation

300 uA

4 mS

Good agreement!

23.6 fF
37.7 fF
9.0 fF

13.3 S/A

HSpice .OP Capacitance Output Variables

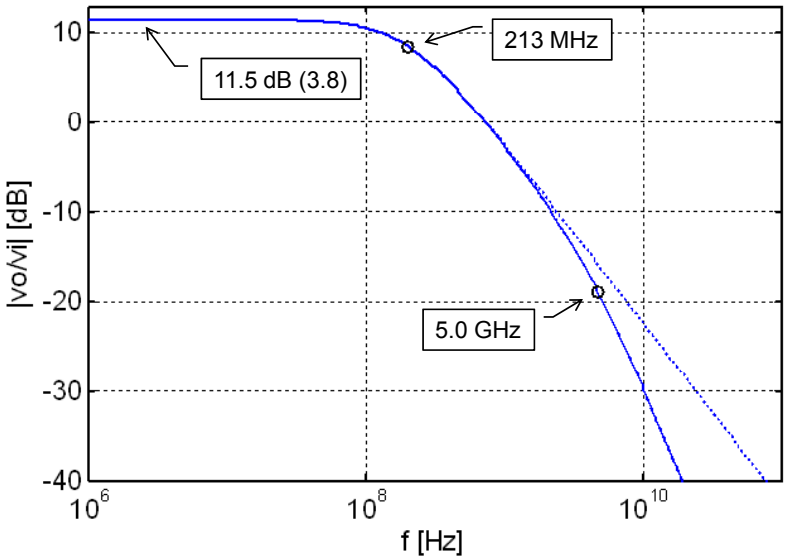
HSpice (.OP)

```
cdtot    21.8712f
cgtot    37.6938f
cstot    44.2809f
cbtot    34.9251f
cgs      26.7303f
cgd       8.9163f
```

Corresponding Small Signal Model Elements

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 $\equiv C_{gd}$

Simulated AC Response



- Calculated values: $|A_v(0)|$=12 dB (4.0), f_{p1}=196 MHz, f_{p2}=6.0 GHz

Using .pz Analysis

Netlist statement

```
.pz v(vo) vi
```

Output

```
*****************************************************
******   pole/zero analysis
input =  0:vi          output = v(vo)

       poles (rad/sec)                poles ( hertz)
 real            imag            real            imag
 -1.34190g       0.              -213.569x       0.
 -31.4253g       0.              -5.00149g       0.

       zeros (rad/sec)                zeros ( hertz)
 real            imag            real            imag
 458.247g        0.              72.9323g        0.
```


Observations

- The design and pole calculations essentially right on target!
  - Typical discrepancies are on the order of 10-20%, mostly 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 μC_{ox}, V_{OV}, etc.
    - Based on artificial parameters that do not exist or have no significance in the spice model

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.

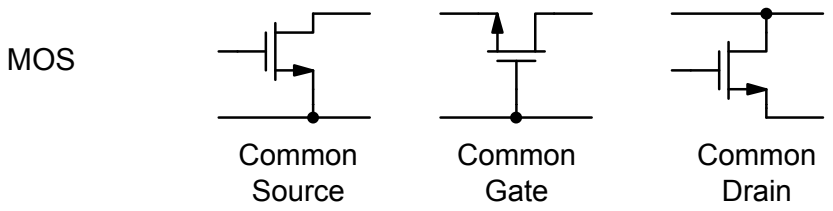
Chapter 4
Review of Elementary Circuit Configurations

B. Murmann

Stanford University

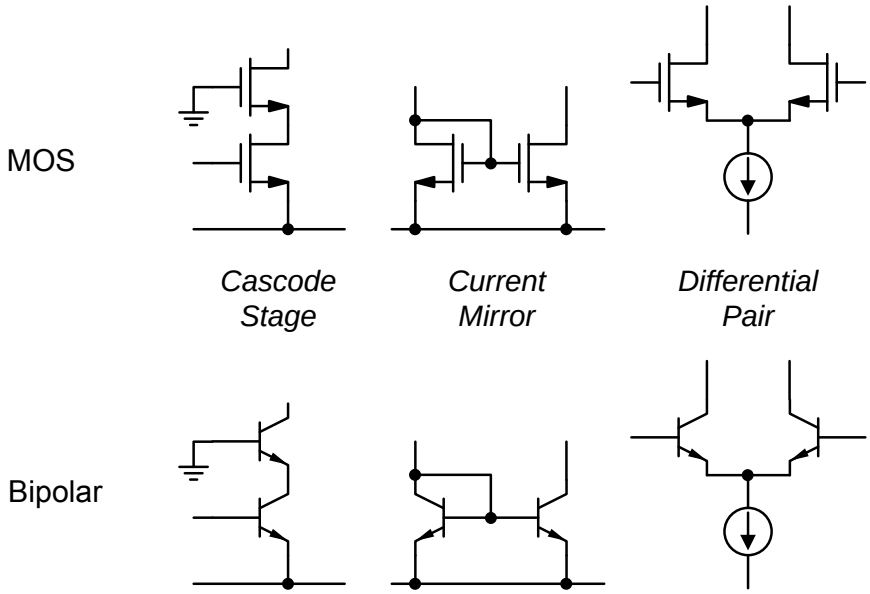
Reading Material: Sections 3.3.1, 3.3.3, 3.3.6, 3.3.8, 3.5, 7.2.3, 7.2.4.1, 7.3.2, 7.3.4, 4.2.2, 4.2.3, 4.2.4

Basic Single-Stage Amplifier Configurations



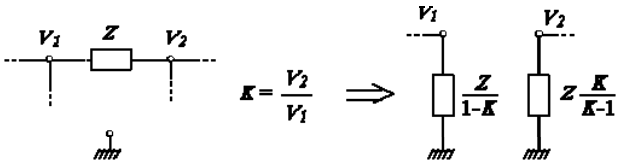
Transconductance Stage *Current Buffer* *Voltage Buffer*

Widely Used Two-Transistor Circuits



Analysis Techniques (1)

- Nodal analysis (KCL, KVL)
  - Write KCL for each node, solve for desired transfer function or port impedance
  - Most general method, but conveys limited qualitative insight and often yields high-entropy expressions
- Miller theorem



http://paginas.fe.up.pt/~fff/eBook/MDA/Teo_Miller.html

- Miller approximation
  - Approximate the gain across Z as frequency independent, i.e. K(s) ≅ K for the frequency range of interest
  - This approximation requires a check (or good intuition)

Analysis Techniques (2)

- Dominant pole approximation

$$\frac{1}{\left(1-\frac{s}{p_1}\right)\left(1-\frac{s}{p_2}\right)} = \frac{1}{1-\frac{s}{p_1}-\frac{s}{p_2}+\frac{s^2}{p_1p_2}} \cong \frac{1}{1-\frac{s}{p_1}+\frac{s^2}{p_1p_2}}$$

$$\text{Given } \frac{1}{1+b_1s+b_2s^2} \Rightarrow p_1 \cong -\frac{1}{b_1}, \quad p_2 \cong -\frac{b_1}{b_2}$$

- Zero value time constant analysis
  - The coefficient b_1 can be found by summing all zero value time constants in the circuit

$$b_1 = \sum \tau_i$$

- Generalized time constant analysis
  - Can also find higher order terms (e.g. b_2) using a sum of time constant products
  - A. Hajimiri, "Generalized Time- and Transfer-Constant Circuit Analysis," IEEE Trans. Circuits Syst. I, pp. 1105-1121, June 2010.

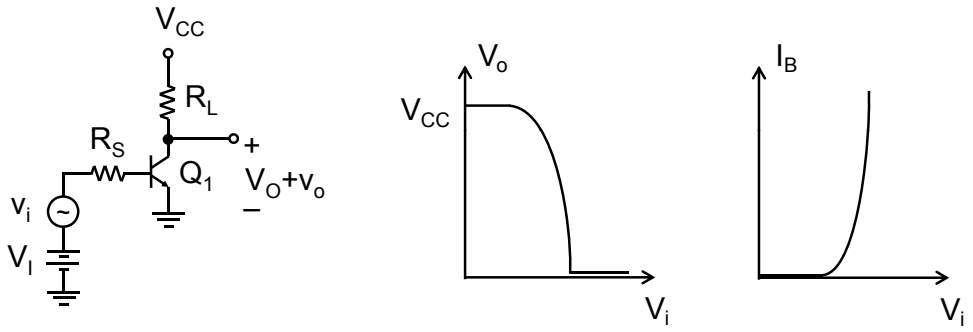
Analysis Techniques (3)

- Return ratio analysis
  - See text pp. 599-612
- Blackman's impedance formula
  - See text pp. 607-612
- Two-port feedback analysis
  - See text pp. 557-587
  - More later in this course

Chapter Overview

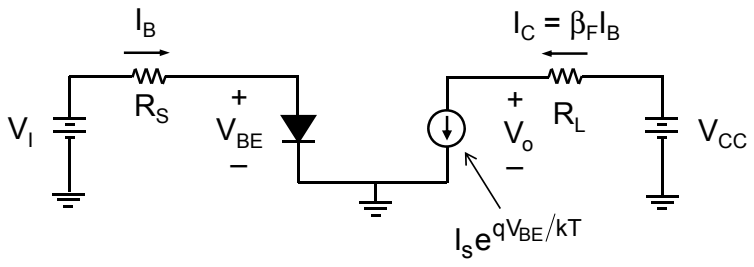
- BJT-centric review of elementary circuit configurations
  - Highlight differences to MOS circuits
- Single-stage amplifiers
  - Common emitter stage
  - Common emitter stage with source degeneration
  - Common collector stage
  - Common base stage
  - Common gate stage (discussion of bulk connection)
- BJT current mirrors
- BJT differential pair

Common-Emitter Stage



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

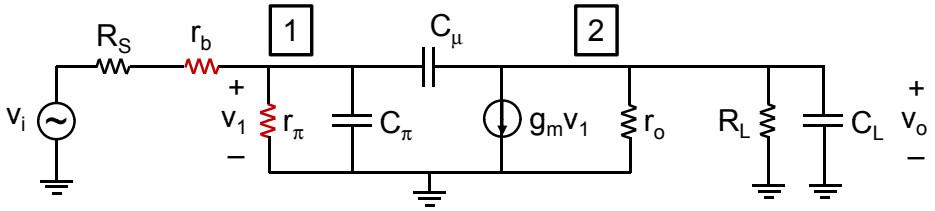
Bias Point Sensitivity



$$I_B \cong \frac{V_I - V_{BE(on)}}{R_S} \qquad V_O = V_{CC} - I_C R_L = V_{CC} - \beta_F I_B R_L = V_{CC} - \beta_F \frac{R_L}{R_S}\left(V_I - V_{BE(on)}\right)$$

- The dependence on β_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 circuits

Small-Signal Equivalent Circuit for CE Stage



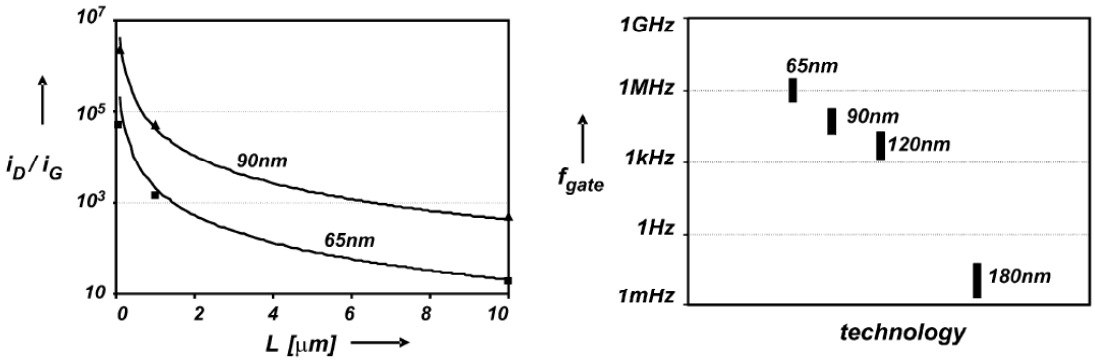
- For hand analysis, we will usually neglect r_c and r_e
- If significant, r_b can be included with R_S, i.e. $R_S^* = R_S + r_b$
- Resulting low-frequency gain

$$A_V(0) \triangleq \left.\frac{v_o}{v_i}\right|_{\omega=0} = -\underbrace{\left(\frac{r_\pi}{R_S^* + r_\pi}\right)}_{=1 \text{ for MOS}} \cdot g_m R_{Ltot} \qquad R_{Ltot} = r_o \,||\, R_L$$

Gate Tunnel Conductance for MOSFETS

- MOSFETs with extremely thin gate oxide draw a gate current due to direct tunneling
- This leads to a finite current gain and input resistance
  - Similar to BJT!

$$f_{\text{gate}} = \frac{g_{\text{tunnel}}}{2\pi C_{\text{in}}}$$



A. Annema, et al., "Analog circuits in ultra-deep-submicron CMOS," *IEEE J. Solid-State Circuits,* pp. 132-143, Jan. 2005.

Frequency Response

- Using nodal analysis, we find

$$\frac{v_o(s)}{v_i(s)} = -\left(\frac{r_\pi}{R_S^* + r_\pi}\right) \cdot g_m R_{Ltot} \frac{\left(1 - \frac{s}{z_1}\right)}{1 + b_1 s + b_2 s^2}$$

$$b_1 = R_S^*(C_\pi + C_\mu) + R_{Ltot}(C_L + C_\mu) + g_m R_S^* R_{Ltot} C_\mu$$

$$b_2 = R_S^* R_{Ltot}(C_\pi C_L + C_\pi C_\mu + C_L C_\mu)$$

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

- z_1 is a feedforward zero in the RHP. If $C_\pi >> C_\mu$, then

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

Dominant Pole Approximation

- If a dominant pole condition exists, we can write

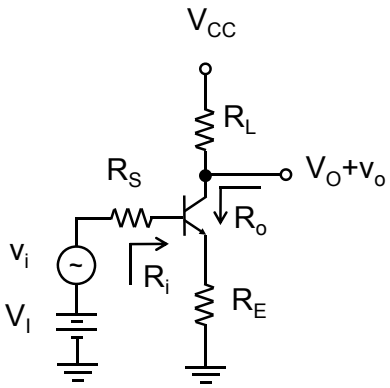
$$p_1 \cong -\frac{1}{b_1} = -\frac{1}{R_S^*(C_\pi + C_\mu) + R_{Ltot}(C_L + C_\mu) + g_m R_S^* R_{Ltot} C_\mu}$$
$$= -\frac{1}{R_S^*\left[C_\pi + (1 + g_m R_{Ltot})C_\mu\right] + R_{Ltot}(C_L + C_\mu)}$$

$$p_2 \cong -\frac{b_1}{b_2} = -\frac{R_S^*(C_\pi + C_\mu) + R_{Ltot}(C_L + C_\mu) + g_m R_S^* R_{Ltot} C_\mu}{R_S^* R_{Ltot}(C_\pi C_L + C_\pi C_\mu + C_L C_\mu)}$$

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

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

Emitter Degeneration

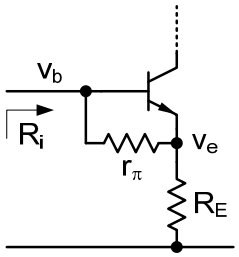


$$G_m \cong \frac{g_m}{1 + g_m R_E}$$
$$R_o \cong r_o(1 + g_m R_E)$$

- Degeneration resistor reduces the transconductance and increases the output resistance of the device
  - Same as in the MOSFET version of this circuit
- For the BJT version, R_E helps increase the input resistance

Calculation of R_i

- For the complete nodal analysis, see text p. 196
- A more intuitive way to find R_i is via the Miller theorem

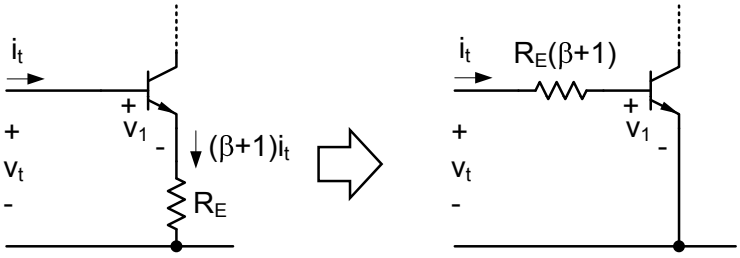


$$K = \frac{v_e}{v_b} \cong \frac{R_E}{\frac{1}{g_m} + R_E} = \frac{g_m R_E}{1 + g_m R_E}$$

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

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

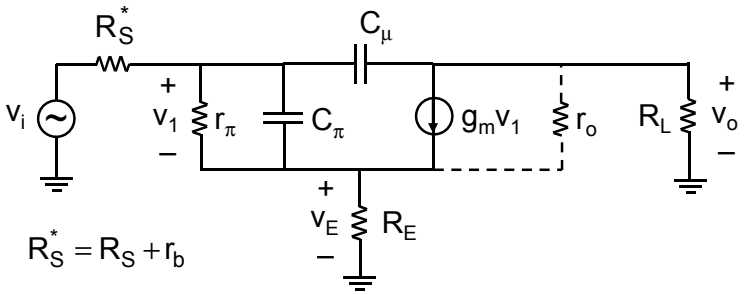
Alternative Calculation of R_i



$$R_i \cong r_\pi + R_E\left(1+\beta\right) = r_\pi + R_E\left(1+g_m r_\pi\right) \cong r_\pi\left(1+g_m R_E\right)$$

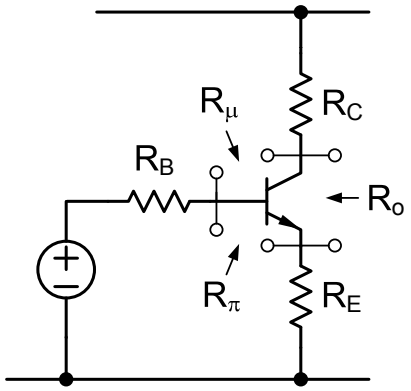
- Tricks of this kind are useful for reasoning about low frequency behavior
- More detailed analyses must be used be taken when investigating frequency dependence

Small-Signal Equivalent Circuit for Degenerated CE Stage



- Deriving the transfer function of this circuit requires solving a 3x3 system of equations
- In order to obtain an estimate of the circuit's bandwidth, it is more convenient (and intuitive) to perform a zero-value time constant analysis

Useful Expressions



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

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

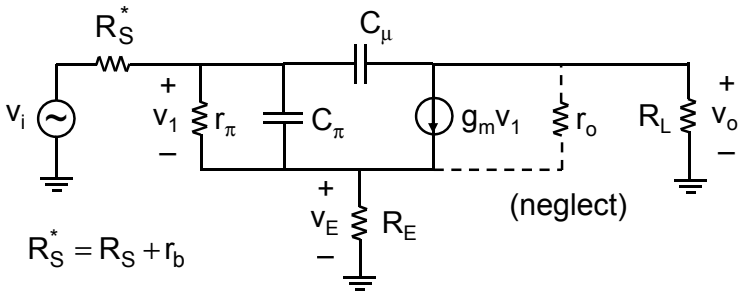
$$R_\mu = R_{left} + R_{right} + G_m R_{left} R_{right}$$

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

$$R_{right} \cong R_C$$

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

Bandwidth Estimate for Degenerated CE Stage (1)



$$R_\mu = R_S^* \| r_\pi (1 + g_m R_E) + R_C + G_m \left[R_S^* \| r_\pi (1 + g_m R_E) \right] R_C$$
$$\cong R_S^* + R_C + G_m R_C R_S^* = R_S^* \left(1 + |A_v(0)|\right) + R_C$$

$$R_\pi \cong r_\pi \| \frac{R_S^* + R_E}{1 + g_m R_E} \cong \frac{R_S^* + R_E}{1 + g_m R_E}$$

Bandwidth Estimate for Degenerated CE Stage (2)

$$\tau = \left[R_S^* \left(1 + |A_v(0)|\right) + R_C \right] C_\mu + \frac{1 + \frac{R_E}{R_S^*}}{1 + g_m R_E} R_S^* C_\pi \qquad \omega_{-3dB} \cong \frac{1}{\tau}$$

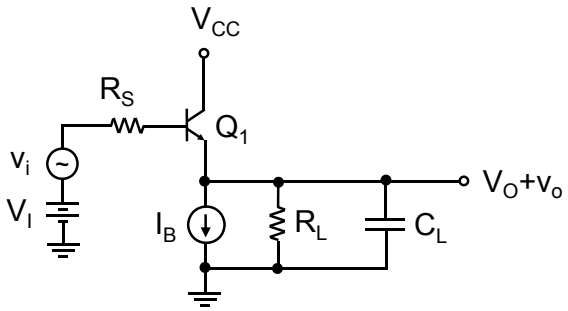
- Compare to the case of $R_E = 0$

$$\tau \cong \left[R_S^* \left(1 + |A_v(0)|\right) + R_C \right] C_\mu + R_S^* C_\pi \qquad \omega_{-3dB} \cong \frac{1}{\tau}$$

- 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 $A_v(0)$
- Consider another special case where $g_m R_E >> 1$ and the time constant due to C_μ is negligible

$$\tau \cong \left(1 + \frac{R_S^*}{R_E}\right) \frac{C_\pi}{g_m} \qquad \omega_{-3dB} \cong \omega_T \left(\frac{C_\pi + C_\mu}{C_\pi}\right) \left(\frac{R_E}{R_E + R_S^*}\right)$$

Common-Collector Stage (Emitter Follower)



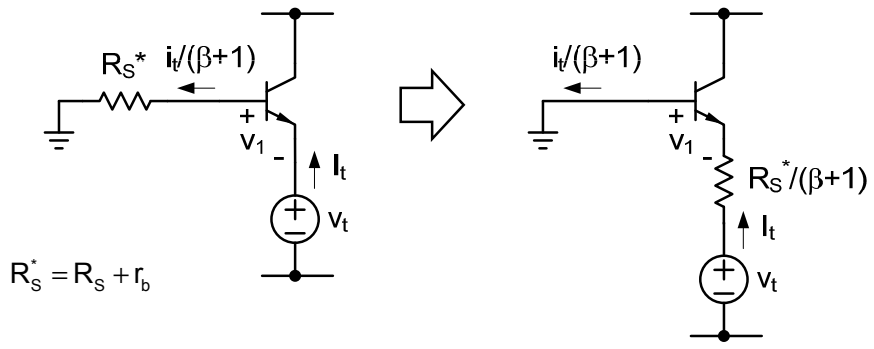
- Behavior is very similar to MOS common drain stage, except that
  - We do not need to worry about backgate effect
  - There is finite input resistance due to r_π
  - The output resistance depends on R_S (in addition to $1/g_m$)

Input and Output Resistance

- Input resistance (by inspection)

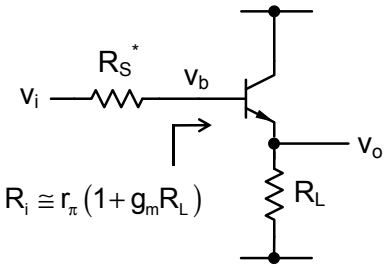
$$R_i \cong r_\pi \left(1+g_m R_L\right)$$

- Output resistance (using push-through trick)



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

Low Frequency Voltage Gain

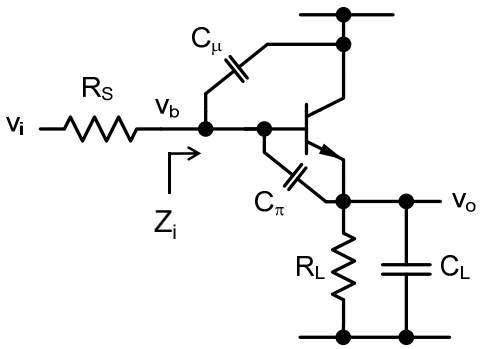


$$A_{v0} = \frac{v_o}{v_i} = \frac{v_b}{v_i}\frac{v_o}{v_b} \cong \frac{r_\pi(1+g_mR_L)}{r_\pi(1+g_mR_L)+R_s^*}\frac{g_mR_L}{1+g_mR_L}$$

$$\cong \frac{g_mR_L}{1+g_mR_L} \quad \text{for} \quad r_\pi(1+g_mR_L) >> R_s^*$$

$$\cong 1 \quad \text{for} \quad g_mR_L >> 1 \quad \text{and} \quad r_\pi(1+g_mR_L) >> R_s^*$$

Frequency Response



$$\frac{v_o}{v_i} = \frac{v_b}{v_i}\cdot\frac{v_o}{v_b} = \frac{Z_i}{Z_i+R_S}\cdot\frac{v_o}{v_b}$$

- Detailed analysis gives a very complex result for the general frequency response expression
- Must typically apply approximations based on given component values

Frequency Response

- Assuming that R_S is large (often the case, and the reason why the stage is used), we expect that the dominant pole is introduced at node v_b
- For the frequency range up until the dominant pole, we can therefore approximate

$$\frac{v_o}{v_b} \cong \frac{g_m R_L}{1+g_m R_L} = K \qquad Z_i \cong \frac{1}{s\left(C_\pi[1-K]+C_\mu\right)} = \frac{1}{sC_i}$$

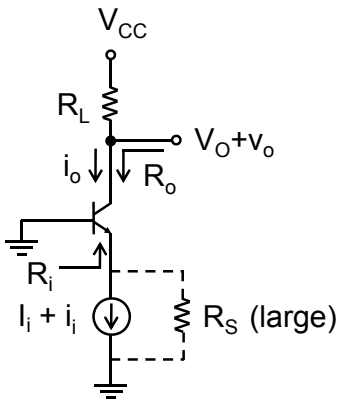
$$\omega_p \cong \frac{1}{\left(R_s \| R_{in}\right) C_i}$$

- See text, pp. 503 for a more detailed analysis, which also captures the feedforward zero introduced by C_π

Common Collector Output Impedance

- Detailed analysis of the common-collector output impedance shows potentially inductive behavior for large R_S
- The inductive behavior can lead to undesired “ringing” (or oscillations) during circuit transients
- In other cases, the inductive behavior is utilized for bandwidth extension (“inductive peaking”)
- The capacitance C_μ tends to reduce the inductive frequency range
  - C_μ appears in parallel with R_S, and creates a low impedance termination for high frequencies
  - Makes it difficult to use the circuit as a “good inductor”
- For a discussion on common collector output impedance and a detailed KCL-based analysis, see EE114 or section 7.2.3

Common-Base Stage



$$A_i = \frac{i_o}{i_i} \qquad A_i(0) = \frac{i_C}{i_E} = \frac{\beta}{\beta+1}$$

Neglecting r_b, r_c, r_e and r_π, we have

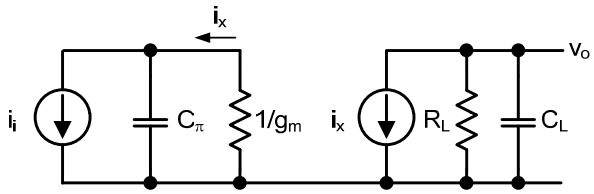
$$R_o \cong r_o(1+g_m R_S)$$

$$R_i \cong \frac{1}{g_m}\left(1+\frac{R_L}{r_o}\right) \cong \frac{1}{g_m}$$

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

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

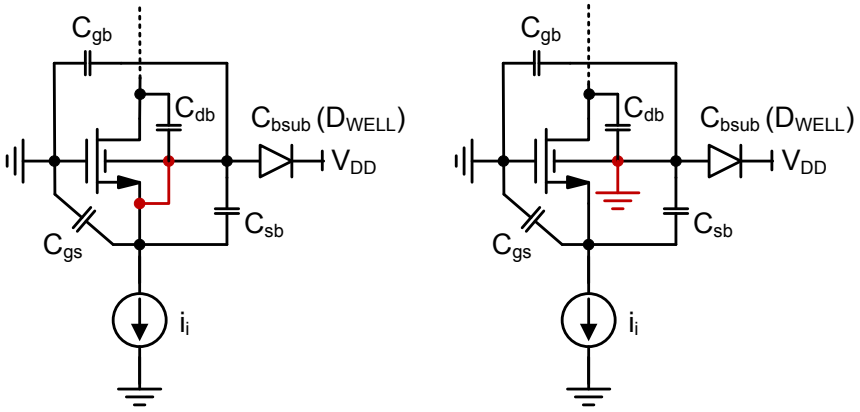
Simplified Small-Signal Model for High-Frequency Analysis



$$\frac{v_o}{i_i} = \frac{R_L}{\left(1-\frac{s}{p_1}\right)\left(1-\frac{s}{p_2}\right)} \qquad \omega_{p2} = \frac{g_m}{C_\pi} = \frac{C_\pi + C_\mu}{C_\pi}\omega_T \qquad \omega_{p1} = \frac{1}{R_L C_L}$$

- The time constant associated with the load usually dominates the frequency response, i.e. $\omega_{p1} < \omega_{p2}$
- Note, however, that ω_{p2} can be important in feedback circuits (phase margin)

Common Gate Stage – Bulk Connection Scenarios

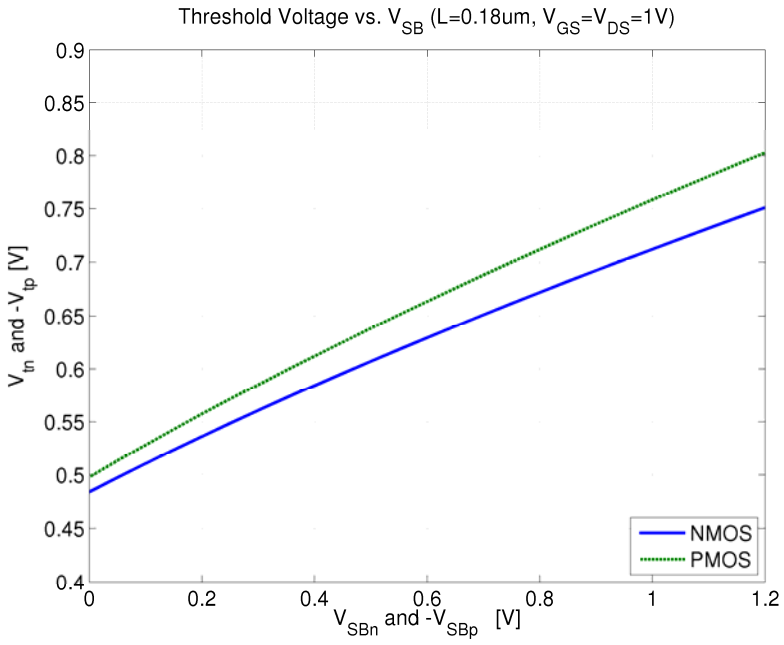


$$\omega_{p2,a} = \frac{g_m}{C_{gs} + C_{gb} + C_{bsub} + C_{db}[1 - K(s)]}$$

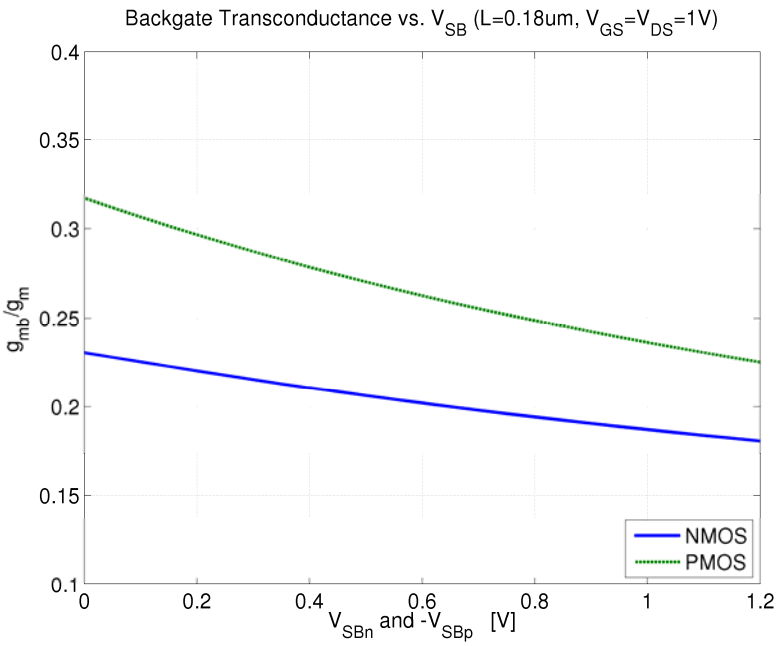
$$\omega_{p2,b} = \frac{g_m + g_{mb}}{C_{gs} + C_{sb}}$$

- $\omega_{p2,a}$ is always less than $\omega_{p2,b}$ → Usually a bad idea to connect source to bulk in a common gate stage

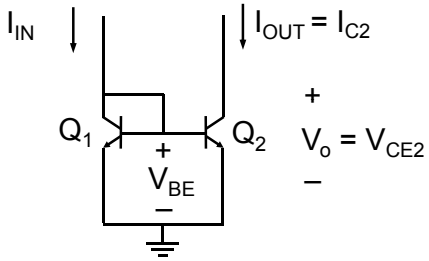
Backgate Effect in the EE214 Technology (1)



Backgate Effect in the EE214 Technology (2)



Basic BJT Current Mirror



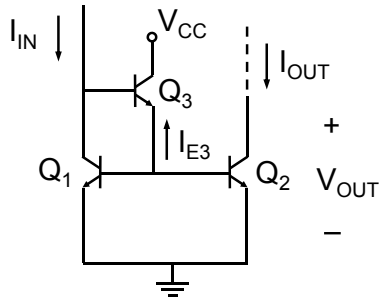
$$\frac{I_{C2}}{I_{C1}} = \frac{I_{S2} e^{\frac{V_{BE}}{V_T}} \left(1 + \frac{V_{CE2}}{V_A}\right)}{I_{S1} e^{\frac{V_{BE}}{V_T}} \left(1 + \frac{V_{CE1}}{V_A}\right)} \cong \frac{I_{S2}}{I_{S1}} \left(1 + \frac{V_{CE2}}{V_A} - \frac{V_{CE1}}{V_A}\right)$$

- Error due to base current

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

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

BJT Current Mirror with "Beta Helper"

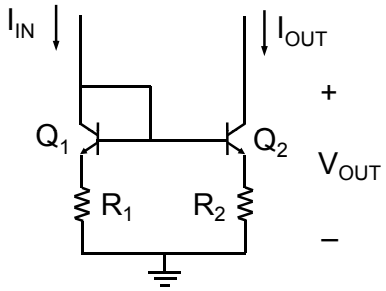


$$-I_{E3} = \frac{I_{C1}}{\beta} + \frac{I_{C2}}{\beta} \cong 2\frac{I_{C2}}{\beta} \quad \text{assuming} \quad I_{S1} = I_{S2}$$

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

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

BJT Current Mirror with Degeneration



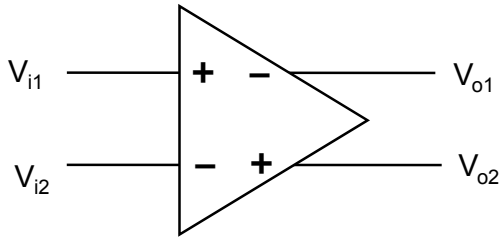
Neglecting base currents

$$V_{BE1} + I_{C1}R_1 = V_{BE2} + I_{C2}R_2$$

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

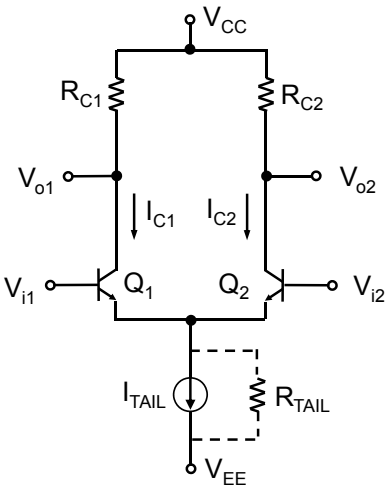
- Degeneration brings two benefits
  - Increased output resistance
  - Reduces sensitivity of mirror ratio to mismatches in I_S
- Minimum V_{OUT} for which Q_2 remains forward active is increased

Differential Circuits



- Enables straightforward biasing without AC coupling
- Information is carried in “differential” signals that are insensitive to “common-mode” perturbations, such as power supply noise
- Fully differential circuits are increasingly used for I/O and clock-and-data recovery (CDR) circuits to allow the use of small signals at high speeds

BJT Differential Pair



Differential Input Voltage

$$V_{id} \triangleq V_{i1} - V_{i2}$$

Differential Collector Current

$$I_{cd} \triangleq I_{c1} - I_{c2}$$

Differential Output Voltage

$$V_{od} \triangleq V_{o1} - V_{o2}$$

- 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

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

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

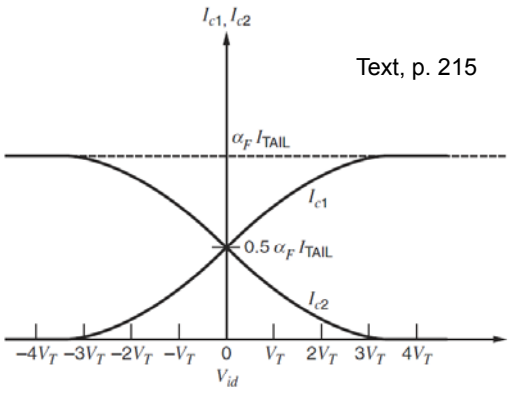
$$I_{TAIL} = -(I_{e1} + I_{e2}) = \frac{1}{\alpha}(I_{c1} + I_{c2}) \qquad \Rightarrow I_{c1} = \frac{\alpha I_{TAIL}}{1 + e^{-\frac{V_{id}}{V_T}}} \qquad I_{c2} = \frac{\alpha I_{TAIL}}{1 + e^{+\frac{V_{id}}{V_T}}}$$

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

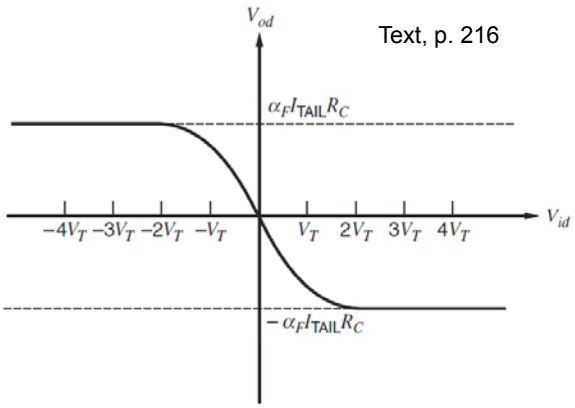
$$V_{od} = I_{od} R_L = \alpha I_{TAIL} R_L \tanh\left(\frac{V_{id}}{2V_T}\right)$$

Plot of Transfer Characteristics

Text, p. 215



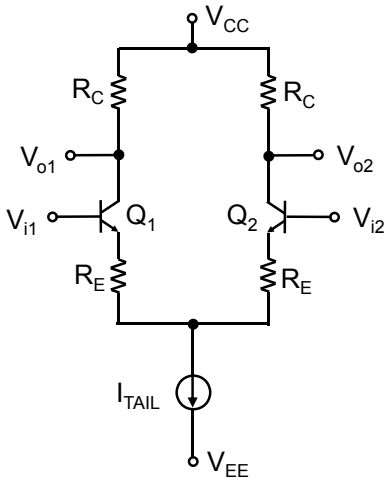
Text, p. 216

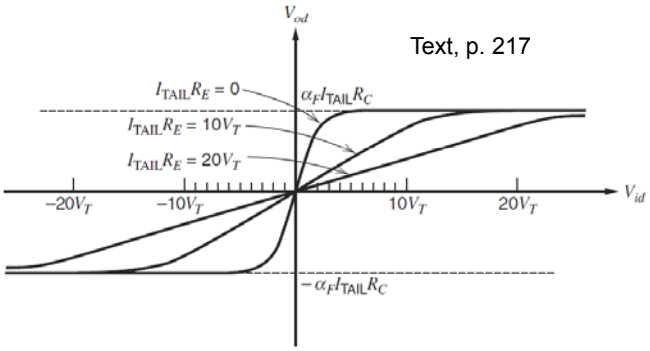


- Linear region in V_{od} vs. V_{id} characteristic is narrow compared to MOS
  - Recall that full steering in a MOS pair occurs for $V_{id} = \sqrt{2}V_{OV}$
- BJT differential pair is linear only for $|V_{id}| < V_T \cong 26$ mV

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 $I_{TAIL}R_E$





Voltage Decomposition

Common-Mode Voltages

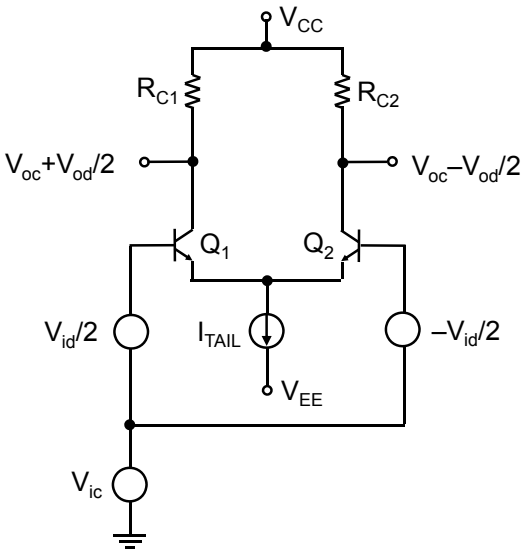
$$V_{ic} \triangleq \frac{1}{2}(V_{i1} + V_{i2})$$

$$V_{oc} \triangleq \frac{1}{2}(V_{o1} + V_{o2})$$

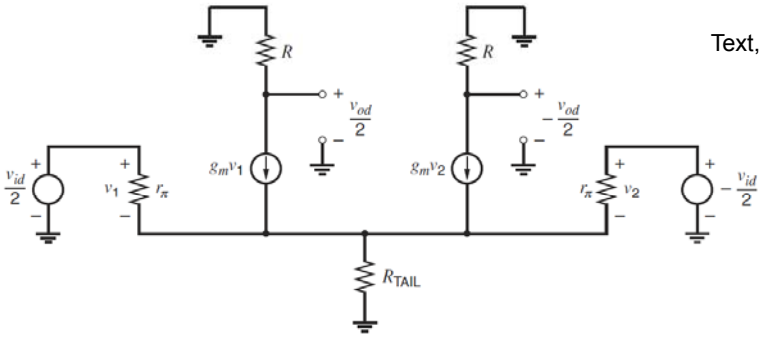
Inputs V_{i1} and V_{i2} can be decomposed into a combination of differential- and common-mode voltage sources

$$V_{i1} = V_{ic} + \frac{1}{2}V_{id}$$

$$V_{i2} = V_{ic} - \frac{1}{2}V_{id}$$



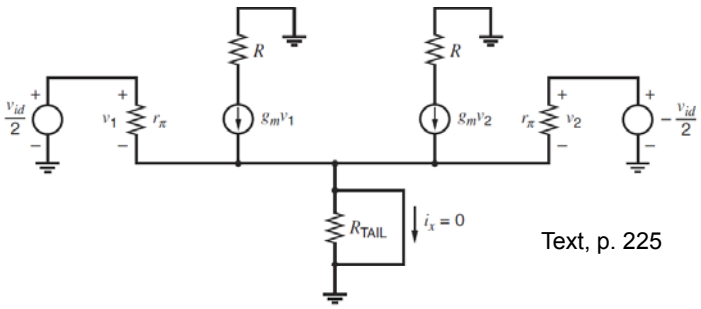
Small Signal Model for $v_{ic} = 0$



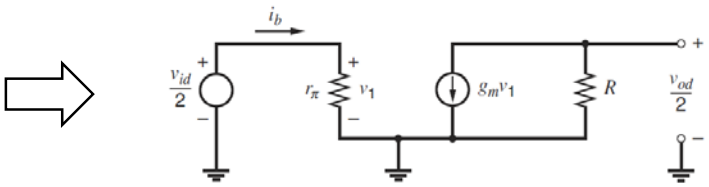
Text, p. 224

- Define differential mode gain as $A_{dm} \triangleq \left.\frac{v_{od}}{v_{id}}\right|_{v_{ic}=0}$

Differential Mode Half Circuit

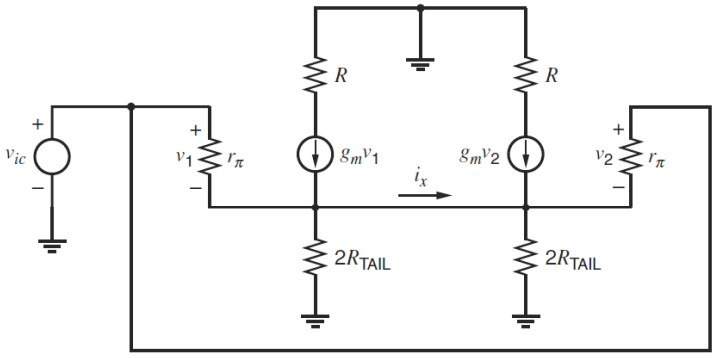


Text, p. 225



$$\frac{v_{od}}{2} = -g_m R \frac{v_{id}}{2} \qquad A_{dm} = \frac{v_{od}}{v_{id}} = -g_m R$$

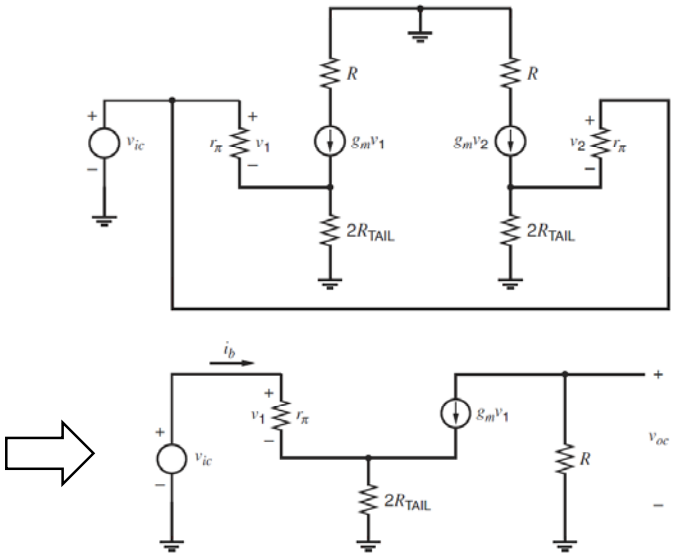
Small Signal Model for $v_{id} = 0$



Text, p. 227

- Define common mode gain as $A_{cm} \triangleq \left.\frac{v_{od}}{v_{id}}\right|_{v_{id}=0}$

Common Mode Half Circuit



Text, p. 227

$$v_{oc} = -G_m R v_{ic} \qquad A_{cm} = \frac{v_{oc}}{v_{ic}} = -G_m R = -\frac{g_m R}{1 + 2 g_m R_{TAIL}}$$

Interaction of Common Mode and Differential Mode

$$A_{cdm} \triangleq \left.\frac{v_{od}}{v_{ic}}\right|_{v_{id}=0} \quad \text{and} \quad A_{dcm} \triangleq \left.\frac{v_{oc}}{v_{id}}\right|_{v_{ic}=0}$$

- 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)
- See text, section 3.5.6.9 for a detailed analysis

Common-Mode Rejecton

For fully differential circuits, the common-mode rejection ratio (CMRR) is traditionally defined as

$$\text{CMRR} \triangleq \left|\frac{A_{dm}}{A_{cdm}}\right|$$

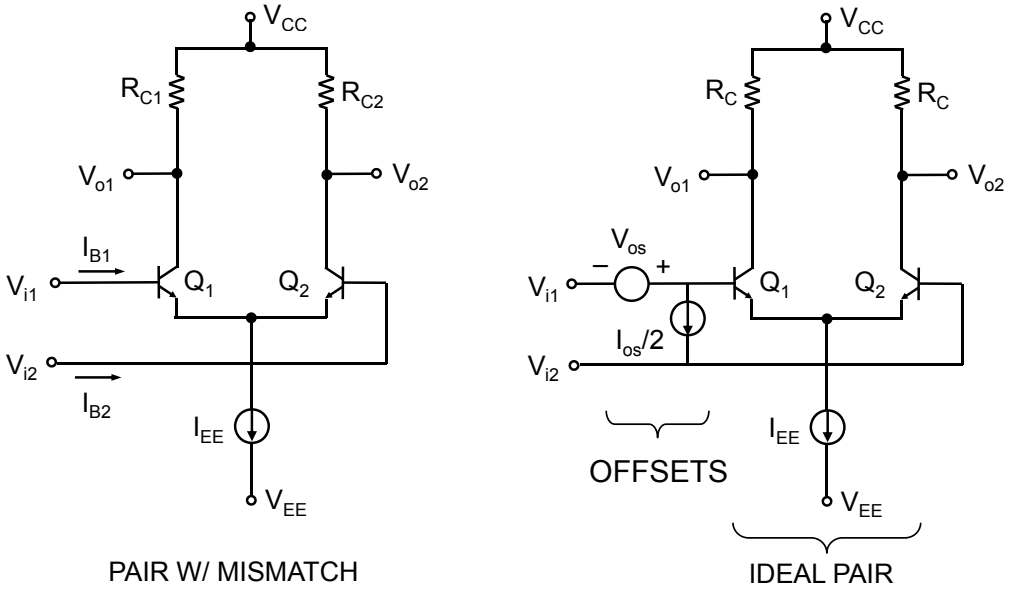
However, the text defines the ratio as

$$\text{CMRR}|_{\text{Text}} \triangleq \left|\frac{A_{dm}}{A_{cm}}\right|$$

This latter definition is appropriate for circuits with a differential input and single-ended output, such as operational amplifiers.

Input-Referred DC Offsets

- In a perfectly symmetric circuit, $V_{id} = 0$ yields $V_{od} = 0$
- Imbalances can be modeled as input referred offsets



Analysis

$$V_{os} - V_{BE1} + V_{BE2} = 0$$

$$\therefore\ V_{os} = V_T \ln\left(\frac{I_{C1}}{I_{S1}}\right) - V_T \ln\left(\frac{I_{C2}}{I_{S2}}\right)$$

$$= V_T \ln\left[\left(\frac{I_{C1}}{I_{C2}}\right)\left(\frac{I_{S2}}{I_{S1}}\right)\right], \qquad \text{where } V_T = \frac{kT}{q}$$

If $V_{od} = 0$, then

$$I_{C1}R_{C1} = I_{C2}R_{C2}$$

$$\therefore\ \frac{I_{C1}}{I_{C2}} = \frac{R_{C2}}{R_{C1}}$$

Thus

$$V_{os} = V_T \ln\left[\left(\frac{R_{C2}}{R_{C1}}\right)\left(\frac{I_{S2}}{I_{S1}}\right)\right]$$

Result

- For small mismatches $\Delta R_C << R_C$ and $\Delta I_S << I_S$, it follows that

$$V_{os} \cong V_T\left(-\frac{\Delta R_C}{R_C}-\frac{\Delta I_S}{I_S}\right)=\left(\frac{g_m}{I_D}\right)^{-1}\left(-\frac{\Delta R_C}{R_C}-\frac{\Delta I_S}{I_S}\right)$$

- And similarly

$$I_{os} \cong -\frac{I_C}{\beta}\left(\frac{\Delta R_C}{R_C}+\frac{\Delta\beta}{\beta}\right) \qquad \text{(see text, pp. 231)}$$

- Mismatch in I_S results primarily from mismatches in the emitter areas and the base doping
- Mismatch in β results primarily from mismatches in the base width
- The standard deviations of device-to-device variations in I_S and β is typically on the order of 5%

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 mV/300°K = 6.6 μV/°C
- For a MOS differential pair, the offset drift is less predictable and turns out to be a complex function of several process parameters

Comparison of V_{OS} for MOS and BJT Differential Pairs

$$V_{OS,BJT} \cong \left(\frac{g_m}{I_D}\right)^{-1}\left(-\frac{\Delta R}{R}-\frac{\Delta I_S}{I_S}\right)$$

$$V_{OS,MOS} \cong \underbrace{\Delta V_t}_{?} + \underbrace{\left(\frac{g_m}{I_D}\right)^{-1}}_{\text{Worse}}\underbrace{\left(-\frac{\Delta R}{R}-\frac{\Delta (W/L)}{(W/L)}\right)}_{\text{Similar to BJT}}$$

- The standard deviation of ΔV_t can be estimated using the following expression (Pelgrom, JSSC 10/1989)

$$\sigma_{\Delta V_t} \cong \frac{A_{Vt}}{\sqrt{WL}}$$

where $A_{vt} \cong 5\text{mV-}\mu\text{m}$ for a typical 0.18 μm process

Numerical Example

- Ignoring resistor mismatch for simplicity
- Assume $(g_m/I_D)_{MOS}$ = 10 S/A, W = 5 μm, L= 0.2 μm

$$\text{std}\left(V_{OS,BJT}\right) \cong \text{std}\left[\left(\frac{g_m}{I_D}\right)^{-1}\left(\frac{\Delta I_S}{I_S}\right)\right] = 26\text{mV}\cdot 5\% = 1.3\text{mV}$$

$$\text{std}\left(V_{OS,MOS}\right) \cong \text{std}\left[\Delta V_t + \left(\frac{g_m}{I_D}\right)^{-1}\left(\frac{\Delta (W/L)}{(W/L)}\right)\right]$$

$$\cong \sqrt{\left(\frac{5\text{mV}-\mu\text{m}}{\sqrt{5\mu\text{m}\cdot 0.2\mu\text{m}}}\right)^2 + \left(100\text{mV}\cdot 5\%\right)^2}$$

$$\cong \sqrt{\left(5\text{mV}\right)^2 + \left(5\text{mV}\right)^2} = 7.1\text{mV}$$

- MOS offset is typically 5-10 times worse than BJT

A_{Vt} Data

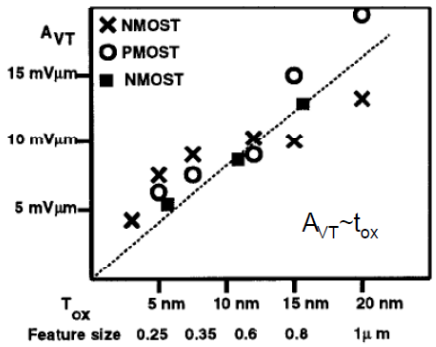
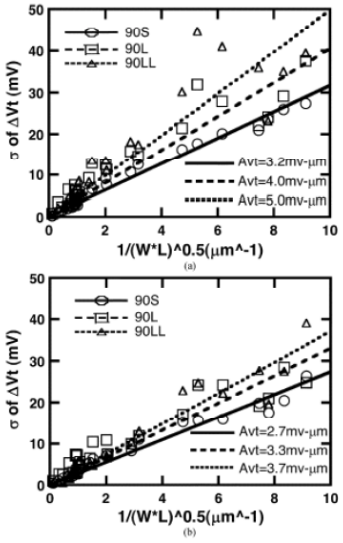


Fig. 3. Evolution of matching coefficient over process generation. Squares are derived from[4], the other measurements are by the authors.

[Pelgrom, IEDM 1998]



[Chang, TED 7/2005]

- A_{Vt} improves with technology scaling
- But, unfortunately, A_{Vt} scales not as fast as minimum device area
  - Hence V_t mismatch for minimum size devices worsens

Chapter 5

Two-Port Feedback Circuit Analysis

B. Murmann

Stanford University

Reading Material: Sections 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.8

Benefits and Costs of Negative Feeback

Negative feedback provides a means of exchanging gain for improvements in other performance metrics

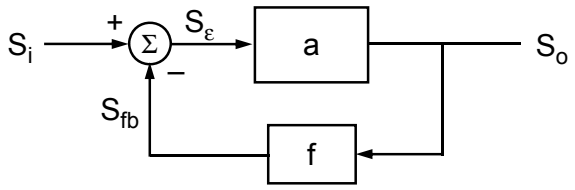
Benefits

- Reduced sensitivity (improved precision)
- Reduced distortion (more later)
- Scaling of impedance levels (up or down)
- Increased bandwidth

Costs

- Lower gain
- Potential instability

Ideal Feedback (1)



Assumptions for an ideal feedback system:

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

$$\left.\begin{aligned} S_o &= a \cdot S_\varepsilon \\ S_{fb} &= f \cdot S_o \\ S_\varepsilon &= S_i - S_{fb} \end{aligned}\right\} \Rightarrow \quad S_o = (S_i - S_{fb}) = a(S_i - f \cdot S_o)$$

Ideal Feedback (2)

$$\left.\begin{aligned} \textbf{Closed-Loop Gain:} \quad A &\triangleq \frac{S_o}{S_i} = \frac{a}{1+af} \\ \textbf{Loop Gain:} \quad T &\triangleq af = \frac{S_{fb}}{S_\varepsilon} \end{aligned}\right\} \Rightarrow \quad A = \frac{a}{1+T}$$

If **T >> 1**, then

$$A \cong \frac{a}{T} = \frac{1}{f}$$

The feedback loop acts to minimize the error signal, S_ε, thus forcing S_{fb} to track S_i. In particular,

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

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

Gain Sensitivity

- 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 relative, or fractional, gain error by the factor 1+T

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

- For a change δa in a

$$\delta a = \frac{dA}{da}\delta a = \frac{\delta a}{(1+T)^2}$$

$$\therefore \quad \frac{\delta A}{A} = \frac{\delta a}{(1+T)^2}\left(\frac{1+T}{a}\right) = \left(\frac{1}{1+T}\right)\frac{\delta a}{a}$$

The Two-Port Approach to Feedback Amplifier Design

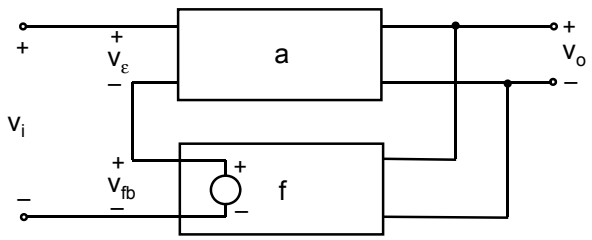
- A practical approach to feedback amplifier analysis and design is based on constructing two-port representations of the forward amplifier and feedback network
- The **two-port approach** relies on the following **assumptions**:
  - Loading effects can be incorporated in the two-port model for the forward amplifier
  - Transmission through the forward amplifier is nearly unilateral
  - Forward transmission through the feedback network is much less than that through the forward amplifier
- When the preceding assumptions break down, the two-port approach deviates from, and is less accurate than, the **return ratio** approach devised by Henrik Bode
- But, the two-port approach can provide better physical tuition, especially with respect to what happens in the frequency domain when the feedback loop is "closed"
  - It is basically an effort to model feedback amplifiers in the way that Harold Black conceived them

Feedback Configurations

- In the two-port approach to feedback amplifier analysis there are four possible amplifier configurations, depending on whether the two-port networks are connected in SHUNT or in SERIES at the input and output of the overall amplifier
- At the OUTPUT
  - A shunt connection senses the output voltage
  - A series connection senses the output current
- At the INPUT
  - A shunt connection feeds back a current in parallel with the input
  - A series connection feeds back a voltage in series with the input

- The four possible configurations are referred to as SERIES-SHUNT, SHUNT-SHUNT, SHUNT-SERIES, and SERIES-SERIES feedback
- The following pages illustrate these configurations using ideal two-port networks

Series-Shunt Feedback

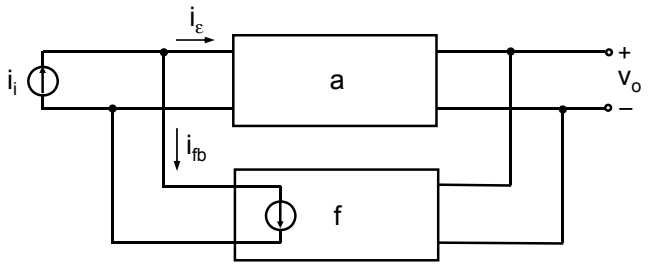


$$a = \frac{v_o}{v_\varepsilon}, \quad f = \frac{v_{fb}}{v_o}$$

$$A = \frac{v_o}{v_i} = \frac{a}{1+af} = \frac{a}{1+T}$$

In this case **a**, **A** and **f** are all **voltage gains**

Shunt-Shunt Feedback

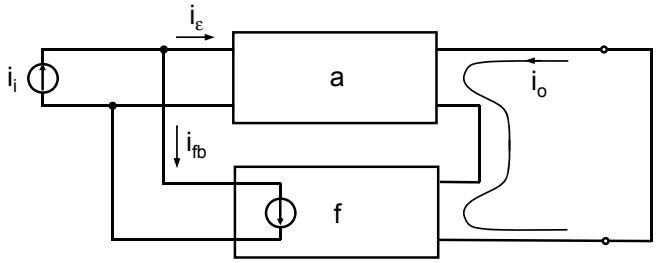


$$a = \frac{v_o}{i_\varepsilon}, \quad f = \frac{i_{fb}}{v_o}$$

$$A = \frac{v_o}{i_i} = \frac{a}{1+af} = \frac{a}{1+T}$$

a and **A** are **transimpedances**; **f** is a **transadmittance**

Shunt-Series Feedback

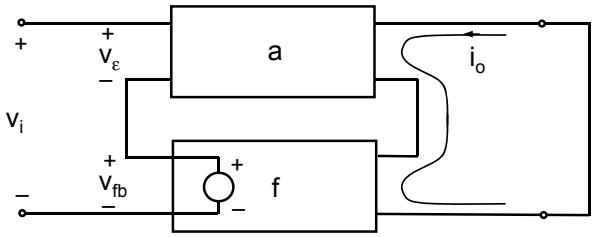


$$a = \frac{i_o}{i_\varepsilon}, \quad f = \frac{i_{fb}}{i_o}$$

$$A = \frac{i_o}{i_i} = \frac{a}{1+af} = \frac{a}{1+T}$$

a, **A**, and **f** are **current gains**

Series-Series Feedback



$$a = \frac{i_o}{v_\varepsilon}, \quad f = \frac{v_{fb}}{i_o}$$

$$A = \frac{i_o}{v_i} = \frac{a}{1+af} = \frac{a}{1+T}$$

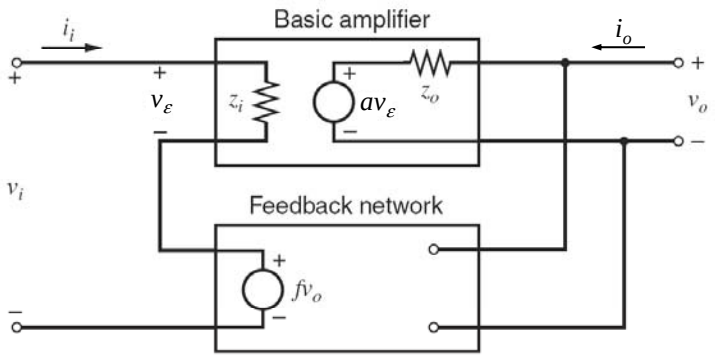
a and **A** are **transadmittances**; **f** is **transimpedance**

Note: T = af is always dimensionless

Closed Loop Impedances

- To illustrate the influence of feedback on the input and output impedances of an amplifier, consider the following:
  - Include finite input and output impedances in a simple, idealized two-port model of the forward amplifier
  - Assume that the feedback network has ideal input and output impedances so as not to load the forward amplifier
- Consider two examples, **series-shunt** and **shunt-series** amplifiers

SERIES-SHUNT



"Ideal" Series-Shunt Impedances

Input Impedance

$$Z_i \triangleq \left.\frac{v_i}{i_i}\right|_{i_o=0}$$

With $i_o = 0$

$$v_o = av_\varepsilon$$
$$v_i = v_\varepsilon + fv_o = (1+af)v_\varepsilon$$
$$= (1+T)v_\varepsilon$$
$$i_i = \frac{v_\varepsilon}{z_i} = \frac{1}{z_i}\left(\frac{1}{1+T}\right)v_i$$

$$\boxed{Z_i = \frac{v_i}{i_i} = (1+T)z_i}$$

Output Impedance

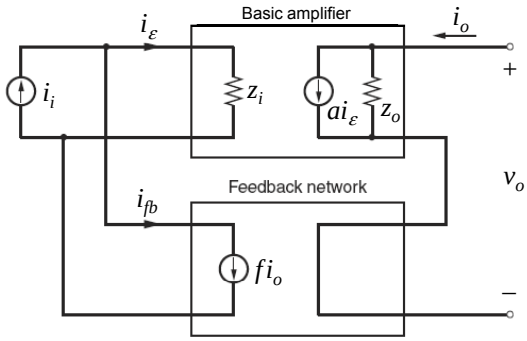
$$Z_o \triangleq \left.\frac{v_o}{i_o}\right|_{v_i=0}$$

With $v_i = 0$

$$v_\varepsilon + fv_o = v_i = 0$$
$$i_o = \frac{v_o - av_\varepsilon}{z_o} = \frac{1}{z_o}(1+af)v_o$$
$$= \frac{1}{z_o}(1+T)v_o$$

$$\boxed{Z_o = \frac{v_o}{i_o} = \frac{z_o}{1+T}}$$

"Ideal" Shunt-Series Impedances



Input Impedance

$$Z_i \triangleq \left.\frac{v_i}{i_i}\right|_{v_o=0}$$

With $v_o = 0$

$$i_o = ai_\varepsilon$$
$$i_i = i_\varepsilon + fi_o = (1+af)i_\varepsilon = (1+T)i_\varepsilon$$
$$v_i = i_\varepsilon z_i = \left(\frac{i_i}{1+T}\right)z_i$$

$$\Rightarrow \quad \boxed{Z_i = \frac{v_i}{i_i} = \frac{z_i}{(1+T)}}$$

Output Impedance

$$Z_o \triangleq \left.\frac{v_o}{i_o}\right|_{i_i=0}$$

With $i_i = 0$

$$i_\varepsilon + f i_o = 0$$
$$v_o = (i_o + a i_\varepsilon) z_o = (i_o + a f i_o) z_o$$
$$= i_o (1+T) z_o$$

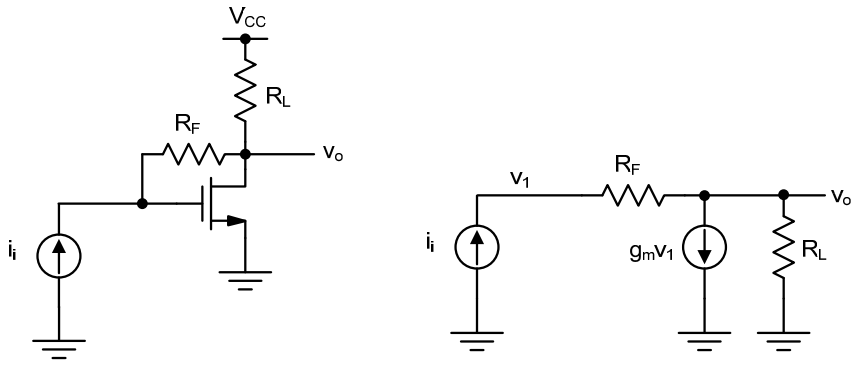
$$\Rightarrow \quad \boxed{Z_o = \frac{v_o}{i_o} = (1+T) z_o}$$

In general:

- Negative feedback connected in **series** **increases** the driving point impedance by (1+T)
- Negative feedback connected in **shunt** **reduces** the driving point impedance by (1+T)

Loading Effects – Introductory Example

- Consider the following feedback circuit



- Analysis methods
  - Closed loop transfer function using nodal analysis
  - Return ratio analysis (see EE114)
  - Two-port feedback circuit analysis

Nodal Analysis

Given

$$0 = \frac{v_1 - v_o}{R_F} - i_i$$

$$0 = \frac{v_o - v_1}{R_F} + g_m \cdot v_1 + \frac{v_o}{R_L}$$

$$\text{Find}(v_1, v_o) \rightarrow \begin{pmatrix} \frac{R_F \cdot i_i + R_L \cdot i_i}{R_L \cdot g_m + 1} \\ \frac{R_L \cdot i_i - R_F \cdot R_L \cdot g_m \cdot i_i}{R_L \cdot g_m + 1} \end{pmatrix} \qquad A = \frac{v_o}{i_i} = \frac{R_L - R_F \cdot R_L \cdot g_m}{R_L \cdot g_m + 1} = -R_F \cdot \left(\frac{1 - \frac{1}{g_m \cdot R_F}}{1 + \frac{1}{g_m \cdot R_L}} \right)$$

- No information about loop gain (which we need e.g. for stability analysis)

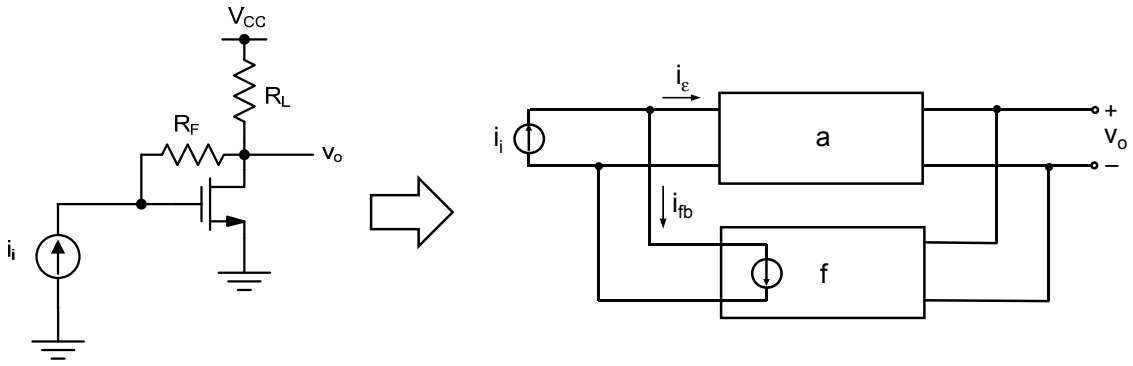
Return Ratio Analysis

$$A = A_{inf} \cdot \frac{RR}{1 + RR} + \frac{d}{1 + RR} \qquad A_{inf} = -R_F \qquad d = R_L \qquad RR = g_m \cdot R_L$$

$$A = -R_F \frac{g_m \cdot R_L}{1 + g_m \cdot R_L} + \frac{R_L}{1 + g_m \cdot R_L} = -R_F \left(\frac{g_m \cdot R_L - \frac{R_L}{R_F}}{1 + g_m \cdot R_L} \right) = -R_F \cdot \left(\frac{1 - \frac{1}{g_m \cdot R_F}}{1 + \frac{1}{g_m \cdot R_L}} \right)$$

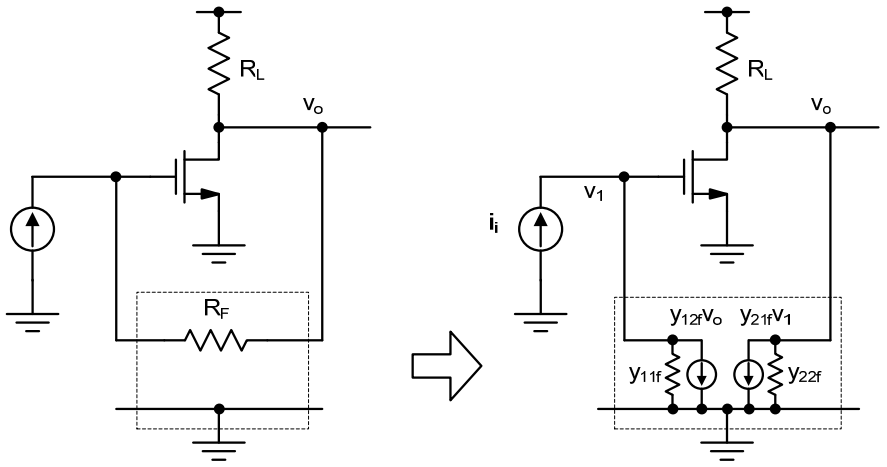
- The result for the closed loop gain (A) matches the nodal analysis perfectly
- In addition, we have determined (along the way) the loop gain (RR)
  - This is useful for stability analysis
- The return ratio method is accurate and general
- The two-port method aims to sacrifice some of this accuracy and generality in exchange for better intuition and less computational effort
  - The involved approximations follow from the typical design intent for each of the four possible approximations

Shunt-Shunt Feedback Model



- R_F feeds back a current that is proportional to the output voltage
  - The appropriate two-port model to use for this amplifier is therefore the “shunt-shunt” configuration
- We wish to identify proper “a” and “f “ blocks that model our circuit
- The key issue is that “a” and “f” are not perfectly separable without making any approximations
  - R_F is responsible for feedback, but also affects the behavior of “a”

y-Parameter Model for the Feedback Network (1)



y-Parameter Model for the Feedback Network (2)

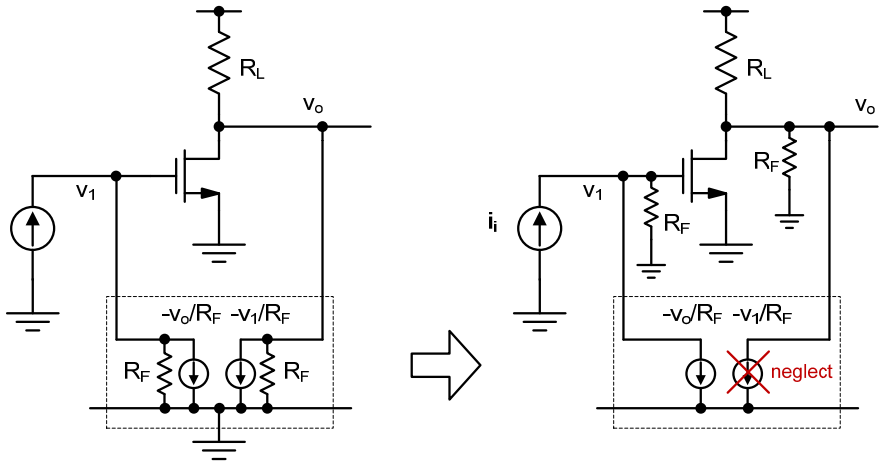
$$i_1 = y_{11}\, v_1 + y_{12}\, v_2$$
$$i_2 = y_{21}\, v_1 + y_{22}\, v_2$$

$$y_{11} = \left.\frac{i_1}{v_1}\right|_{v_2 = 0} \qquad y_{12} = \left.\frac{i_1}{v_2}\right|_{v_1 = 0}$$

$$y_{21} = \left.\frac{i_2}{v_1}\right|_{v_2 = 0} \qquad y_{22} = \left.\frac{i_2}{v_2}\right|_{v_1 = 0}$$

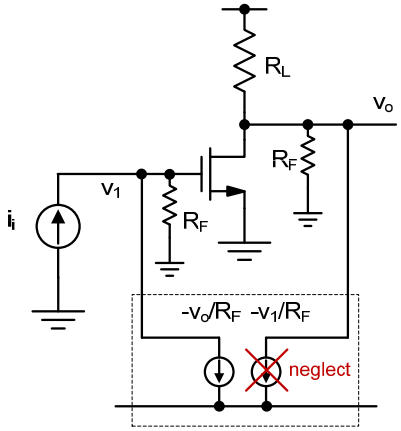
$$y_{11f} = \frac{1}{R_F} \qquad y_{12f} = -\frac{1}{R_F}$$
$$y_{21f} = -\frac{1}{R_F} \qquad y_{22f} = \frac{1}{R_F}$$

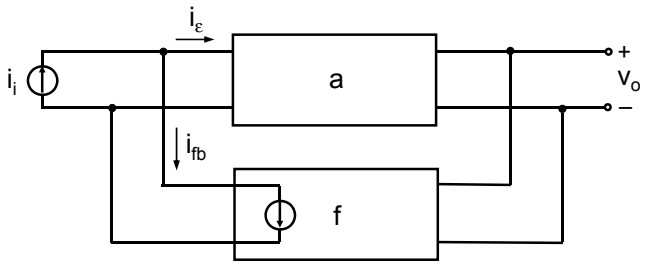
y-Parameter Model for the Feedback Network (3)



- Final steps
  - Absorb y_{11} and y_{22} into forward amplifier
  - Neglect feedforward through feedback network (y_{21})
    - This is justified by the usual design intent: we do not want the feedforward through the feedback network to be significant!

Identification of “a” and “f”





$$f = -\frac{1}{R_F} \qquad a = \frac{v_o}{v_1}\cdot\frac{v_1}{i_e} = -g_m\cdot\frac{R_L\cdot R_F}{R_L + R_F}\cdot R_F \qquad af = g_m\cdot\frac{R_L\cdot R_F}{R_L + R_F}$$

$$A = \frac{1}{f}\cdot\frac{1}{1+\frac{1}{af}} = R_F\,\frac{1}{1+\frac{1}{g_m}\cdot\frac{R_F + R_L}{R_F\cdot R_L}}$$

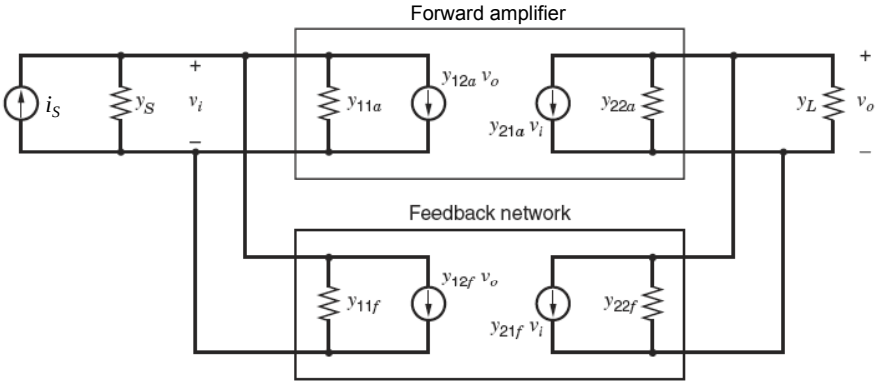
Comparison and Generalization

- Note that for $R_L << R_F$ (i.e. small feedforward)
  - The product of a and f approaches the true loop gain (RR)
  - The two-port closed-loop gain expression (A) approaches the nodal analysis and return ratio result
- The approach illustrated in the previous example can be generalized to cover all four feedback configurations
- For all four configurations, the following analysis steps apply
  - Identify input and output variables and feedback configuration
  - Find feedback function “f”
  - Add loading impedances to input and output port of the basic amplifier “a”
  - Perform calculations using ideal two-port feedback equations

Table 8.1 Summary of Feedback Configurations

Feedback Configuration	Two-Port Parameter Representation	Output Variable	Input Variable	Transfer Function Stabilized	Z_i	Z_o	To Calculate Feedback Loading: At Input	To Calculate Feedback Loading: At Output	To Calculate Feedback Function f
Shunt-shunt	y	v_o	i_s	$\frac{v_o}{i_s}$ Transresistance	Low	Low	Short circuit output feedback node	Short circuit input feedback node	Drive the feedback network with a voltage at its output and calculate the current flowing into a short at its input
Shunt-series	g	i_o	i_s	$\frac{i_o}{i_s}$ Current gain	Low	High	Open circuit output feedback node	Short circuit input feedback node	Drive the feedback network with a current and calculate the current flowing into a short
Series-shunt	h	v_o	v_s	$\frac{v_o}{v_s}$ Voltage gain	High	Low	Short circuit output feedback node	Open circuit input feedback node	Drive the feedback network with a voltage and calculate the voltage produced into an open circuit
Series-series	z	i_o	v_s	$\frac{i_o}{v_s}$ Transconductance	High	High	Open circuit output feedback node	Open circuit input feedback node	Drive the feedback network with a current and calculate the voltage produced into an open circuit

General Analysis for Shunt-Shunt Feedback



Summing currents at the input and output of the overall amplifier

$$i_S = (y_S + y_{11a} + y_{11f})v_i + (y_{12a} + y_{12f})v_o$$
$$0 = (y_{21a} + y_{21f})v_i + (y_L + y_{22a} + y_{22f})v_o$$

Define

$$y_i \triangleq y_S + y_{11a} + y_{11f}$$
$$y_o \triangleq y_L + y_{22a} + y_{22f}$$

Then

$$v_i = -\frac{y_o}{y_{21a} + y_{21f}} v_o$$
$$i_S = y_i\left(\frac{-y_o}{y_{21a} + y_{21f}}\right)v_o + (y_{12a} + y_{12f})v_o$$
$$= \left(\frac{1}{y_{21a} + y_{21f}}\right)\left[-y_i y_o + (y_{21a} + y_{21f})(y_{12a} + y_{12f})\right]v_o$$

$$\therefore \quad \frac{v_o}{i_S} = \frac{-(y_{21a} + y_{21f})}{y_i y_o - (y_{21a} + y_{21f})(y_{12a} + y_{12f})}$$
$$= \frac{-\left(\frac{y_{21a} + y_{21f}}{y_i y_o}\right)}{1 - \left(\frac{y_{21a} + y_{21f}}{y_i y_o}\right)(y_{12a} + y_{12f})}$$

Comparing this result with the ideal feedback equation

$$\frac{v_o}{i_s} = A = \frac{a}{1 + af}$$

it is apparent that a feedback representation can be used by defining

$$a \triangleq -\frac{y_{21a} + y_{21f}}{y_i y_o}$$
$$f \triangleq y_{12a} + y_{12f}$$

For the preceding definitions of a and f, it is difficult to establish equivalent circuits for these functions. However, the situation can be simplified by realizing that it may often be possible to neglect reverse transmission in the forward amplifier and forward transmission in the feedback network. That is, it can often be assumed that

$$|y_{12a}| << |y_{12f}|$$

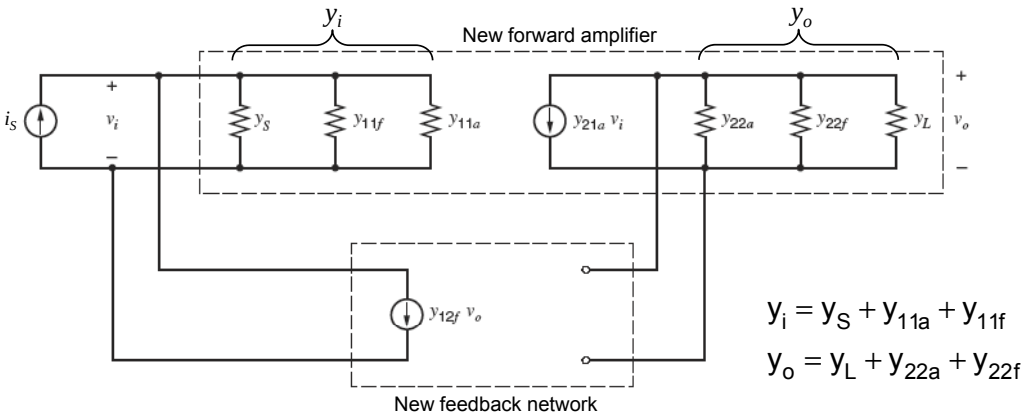
and

$$|y_{21a}| >> |y_{21f}|$$

in which case

$$A = \frac{v_o}{i_S} = \frac{\left(\frac{-y_{21a}}{y_i y_o}\right)}{1 + \left(\frac{-y_{21a}}{y_i y_o}\right) y_{12f}} \quad \Rightarrow \quad \boxed{a = -\frac{y_{21a}}{y_i y_o} \qquad f = y_{12f}}$$

For the simplified definitions, it is possible to identify separate equivalent circuits for a and f :

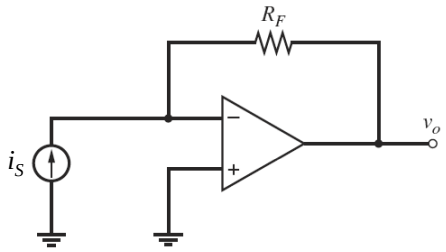


$$y_i = y_S + y_{11a} + y_{11f}$$

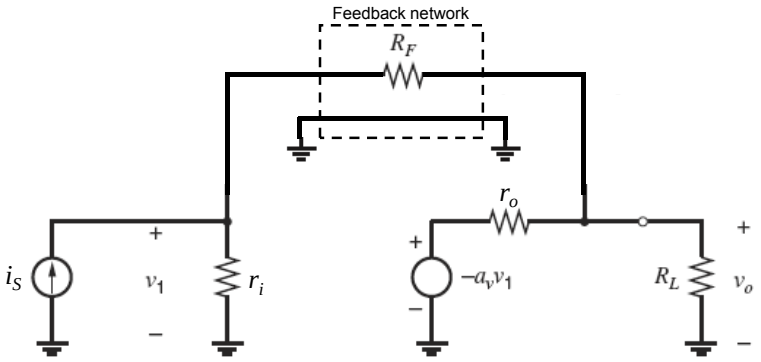
$$y_o = y_L + y_{22a} + y_{22f}$$

The admittances y_S, y_{11a}, y_{22f} and y_L have been "pulled" into the forward amplifier. Basically, the "loading" of the feedback network, as well as the source and load admittances, is absorbed in the equivalent forward amplifier.

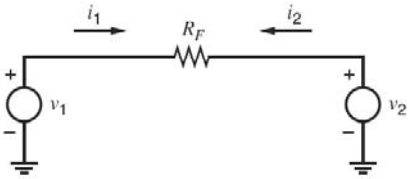
Example – Op Amp w/ Shunt-Shunt Feeback



Small-signal equivalent circuit:



In order to determine the loading of the feedback network on the forward amplifier, the y parameters of the feedback network are first determined using the following circuit :



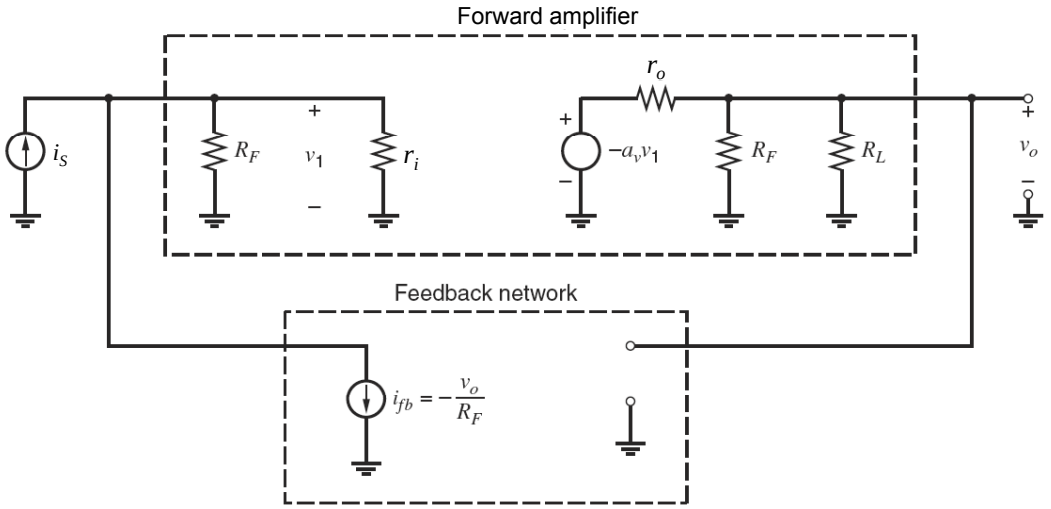
$$y_{11f} = \left.\frac{i_1}{v_1}\right|_{v_2=0} = \frac{1}{R_F}$$

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

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

Also, $y_{21f} = -1/R_F$, but it is neglected in comparison with y_{21a}

The overall circuit can now be separated into forward and feedback paths, with the loading of the feedback network included within the forward path



To determine the forward amplifier gain, a, break the feedback loop at the "output" of the feedback network by setting $i_{fb} = 0$.

Then

$$v_1 = (r_i \,\|\, R_F) i_S = \left(\frac{r_i R_F}{r_i + R_F} \right) i_S$$

$$v_o = -\left(\frac{R}{R + r_o} \right) a_v v_1 \qquad \text{where} \quad R \triangleq R_F \,\|\, R_L$$

Thus

$$a = \left. \frac{v_o}{i_S} \right|_{i_{fb}=0} = -a_v \left(\frac{r_i R_F}{r_i + R_F} \right) \left(\frac{R}{R + r_o} \right)$$

Since $f = -1/R_F$, the loop gain, $T = af$, is

$$T = a_v \left(\frac{r_i}{r_i + R_F} \right) \left(\frac{R_F \,\|\, R_L}{R_F \,\|\, R_L + r_o} \right) = a_v \left(\frac{r_i}{r_i + R_F} \right) \left(\frac{R_L R_F}{R_L R_F + r_o R_F + r_o R_L} \right)$$

The input and output impedances of the equivalent forward amplifier are

$$z_{ia} = R_F \,||\, r_i$$
$$z_{oa} = r_o \,||\, R_F \,||\, R_L$$

Since the feedback network is connected in shunt at both the input and output of the forward amplifier, the closed-loop input and output impedances are

$$Z_i = \frac{z_{ia}}{1+T} = \frac{R_F \,||\, r_i}{1+T}$$
$$Z_o = \frac{z_{oa}}{1+T} = \frac{r_o \,||\, R_F \,||\, R_L}{1+T}$$

Series-Series Feedback

In a series-series feedback amplifier, the forward amplifier and feedback network share common currents at the input and output. Therefore, open circuit impedance parameters (**z parameters**) are used to characterize the two-port networks.

$$v_1 = z_{11} i_1 + z_{12} i_2$$
$$v_2 = z_{21} i_1 + z_{22} i_2$$

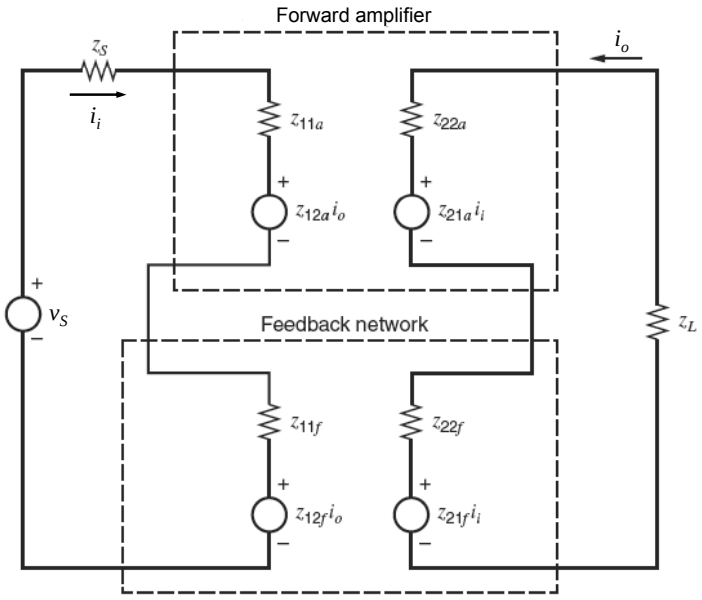
where

$$z_{11} \triangleq \left.\frac{v_1}{i_1}\right|_{i_2=0} \qquad z_{12} \triangleq \left.\frac{v_1}{i_2}\right|_{i_1=0}$$

$$z_{21} \triangleq \left.\frac{v_2}{i_1}\right|_{i_2=0} \qquad z_{22} \triangleq \left.\frac{v_2}{i_2}\right|_{i_1=0}$$

Series-Series Feedback, cont'd

The two-port representation of a series-series feedback amplifier is then



Series-Series Feedback, cont'd

Summing voltages at the input and output

$$v_S = (z_S + z_{11a} + z_{11f})i_i + (z_{12a} + z_{12f})i_o$$
$$0 = (z_{21a} + z_{21f})i_i + (z_L + z_{22a} + z_{22f})i_o$$

Define

$$z_i \triangleq z_S + z_{11a} + z_{11f}$$
$$z_o \triangleq z_L + z_{22a} + z_{22f}$$

Again, neglect reverse transmission through the forward amplifier and feedforward through the feedback network; that is, assume

$$|z_{12a}| << |z_{12f}|$$

and

$$|z_{21a}| >> |z_{21f}|$$

Series-Series Feedback Equations

Under the preceding assumptions, the series-series feedback amplifier can be represented by a two-port model with

$$A = \frac{i_o}{v_S} = \frac{\left(\frac{-z_{21a}}{z_i z_o}\right)}{1 + \left(\frac{-z_{21a}}{z_i z_o}\right) z_{12f}} = \frac{a}{1 + af}$$

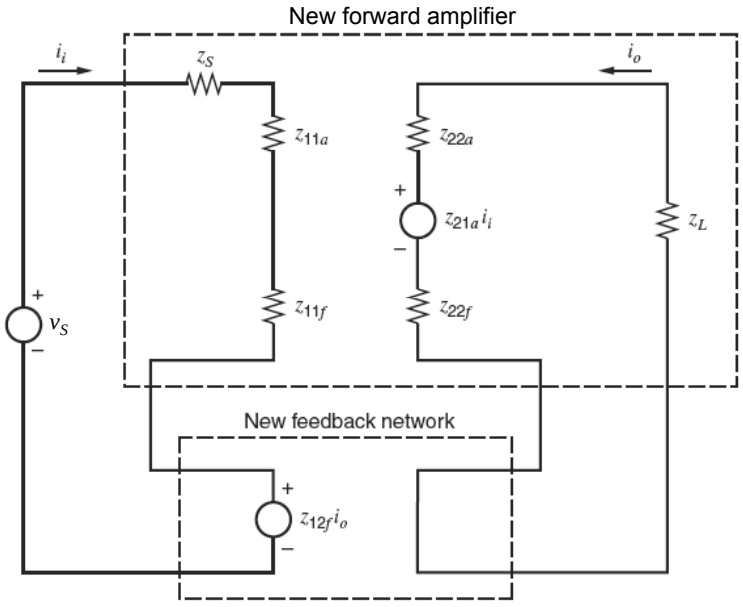
where

$$a = -\frac{z_{21a}}{z_i z_o}$$

$$f = z_{12f}$$

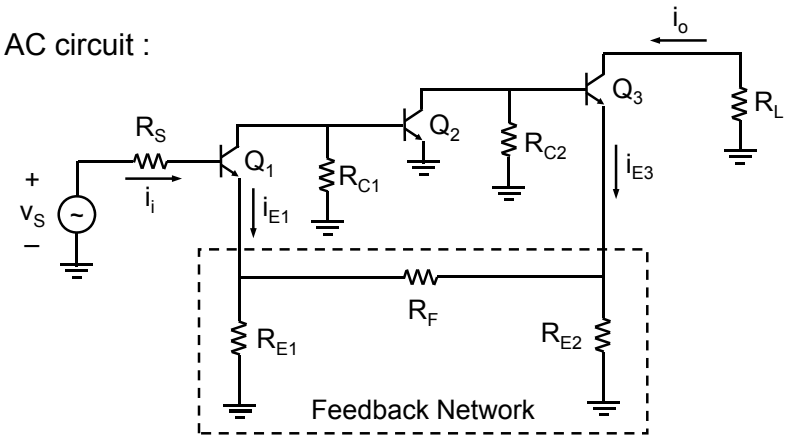
Series-Series Feedback Equivalent Two-Port Networks

Equivalent two-port networks can now be identified for **a** and **f** in which the loading of the feedback network is included in the forward amplifier



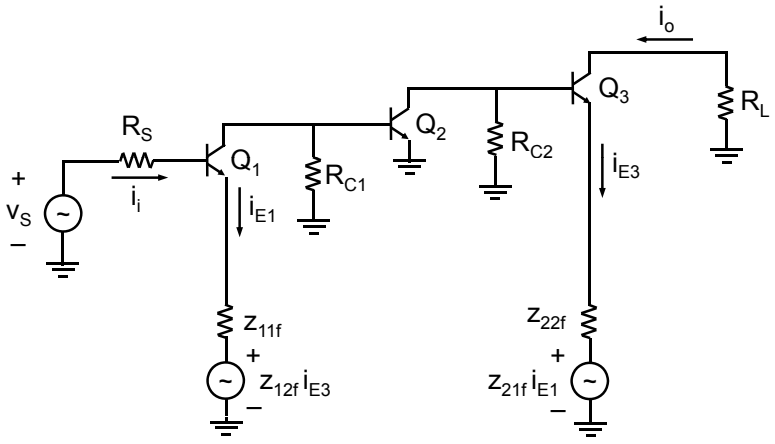
Series-Series Triple

AC circuit :



The forward amplifier in this circuit is **NOT** a two-port network because $i_{E1} \neq i_i$ and $i_{E3} \neq i_o$. However, because i_{E1} and i_{E3} are closely related to i_i and i_o, a two-port approach to feedback analysis can still be used.

Begin by representing the feedback network as a two-port network.

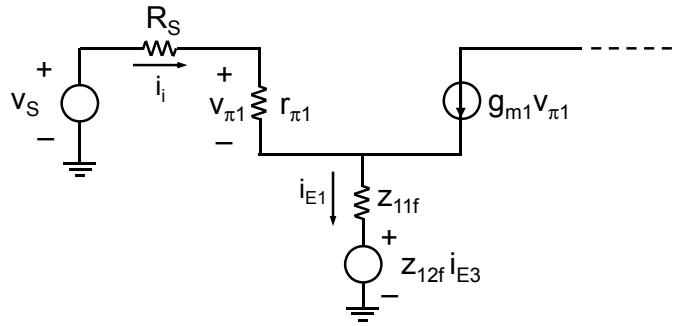


At the output, simply recognize that the feedback network "senses" i_{E3} rather than i_o, and that

$$i_{E3} = \frac{i_o}{\alpha_3}$$

In effect, α_3 is **outside the feedback loop**

At the input, consider the following small-signal equivalent circuit

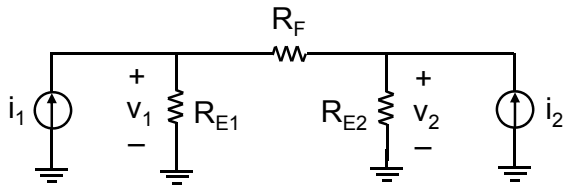


$$v_S = i_i R_S + v_{\pi 1} + i_{E1} z_{11f} + z_{12f} i_{E3}$$

$$\therefore\ v_S - z_{12f} \cdot \frac{i_o}{\alpha_3} = i_i R_S + v_{\pi 1} + i_{E1} z_{11f}$$

From this equation it is apparent that, although the z_{12f} feedback voltage generator is in the emitter of Q_1, it is still in series with, and **subtracts from**, the input voltage source, v_S. Thus, the z_{12f} generator can be moved from the emitter to the base of Q_1

Next, to determine the z parameters for the feedback network, consider the following circuit



$$z_{11f} = \left.\frac{v_1}{i_1}\right|_{i_2=0} = R_{E1} \,||\, (R_F + R_{E2})$$

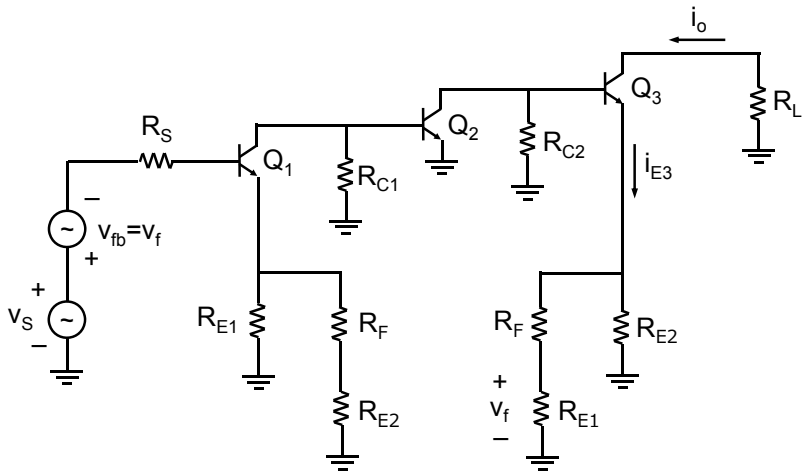
$$z_{22f} = \left.\frac{v_2}{i_2}\right|_{i_1=0} = R_{E2} \,||\, (R_F + R_{E1})$$

$$z_{12f} = \left.\frac{v_1}{i_2}\right|_{i_1=0} = \left(\frac{R_{E1}}{R_{E1}+R_F}\right)\left(\left.\frac{v_2}{i_2}\right|_{i_1=0}\right) = \left(\frac{R_{E1}}{R_{E1}+R_F}\right) z_{22f}$$

$$= \left(\frac{R_{E1}}{R_{E1}+R_F}\right)\left[\frac{R_{E2}(R_{E1}+R_F)}{R_{E1}+R_{E2}+R_F}\right] = \frac{R_{E1}R_{E2}}{R_{E1}+R_{E2}+R_F}$$

$z_{21f} = z_{12f}$, but z_{21f} is neglected in comparison with z_{21a}

Based on the preceding results, the series-series triple can be modeled with the following equivalent circuit



In this circuit we have managed to "break" the loop at the output of the feedback network. Thus, the circuit can be used to determine a, f and T.

Note that a and f for the series-series triple can be defined in terms of either i_o or i_{E3}. If

$$A = \frac{i_o}{v_S} = \alpha_3 \frac{i_{E3}}{v_S}$$

and a and f are defined with respect to i_o, then

$$f = \frac{z_{12f}}{\alpha_3} = \frac{1}{\alpha_3}\left(\frac{R_{E1}R_{E2}}{R_{E1}+R_{E2}+R_F}\right)$$

As T = af becomes large

$$A \cong \frac{1}{f} = \alpha_3\left(\frac{R_{E1}+R_{E2}+R_F}{R_{E1}R_{E2}}\right)$$

Thus, A is **not** desensitized to α_3 by the feedback loop

Series-Shunt Feedback

In a series-shunt feedback amplifier, the forward amplifier and feedback network share the same input current and output voltage. Therefore a hybrid **h-parameter** representation is used for the two-port networks.

$$v_1 = h_{11} i_1 + h_{12} v_2$$
$$i_2 = h_{21} i_1 + h_{22} v_2$$

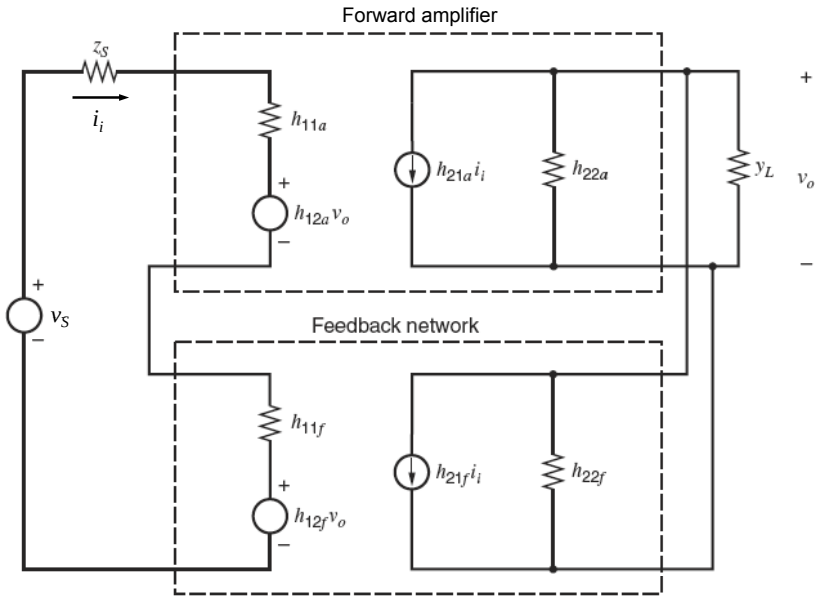
where

$$h_{11} \triangleq \left.\frac{v_1}{i_1}\right|_{v_2=0} \qquad h_{12} \triangleq \left.\frac{v_1}{v_2}\right|_{i_1=0}$$

$$h_{21} \triangleq \left.\frac{i_2}{i_1}\right|_{v_2=0} \qquad h_{22} \triangleq \left.\frac{i_2}{v_2}\right|_{i_1=0}$$

Series-Shunt Feedback, cont'd

Thus, the two-port model for a series-shunt feedback circuit is



Series-Shunt Feedback, cont'd

Summing voltages at the input and currents at the output,

$$v_S = (z_S + h_{11a} + h_{11f})i_i + (h_{12a} + h_{12f})v_o$$
$$0 = (h_{21a} + h_{21f})i_i + (y_L + h_{22a} + h_{22f})i_o$$

Define

$$z_i \triangleq z_S + h_{11a} + h_{11f}$$
$$y_o \triangleq y_L + h_{22a} + h_{22f}$$

and neglect reverse transmission through the forward amplifier and feedforward through the feedback network; that is, assume

$$|h_{12a}| \ll |h_{12f}|$$

and

$$|h_{21a}| \gg |h_{21f}|$$

Series-Shunt Feedback Equations

With the simplifying assumptions, the series-series feedback amplifier can be represented by a two-port model with

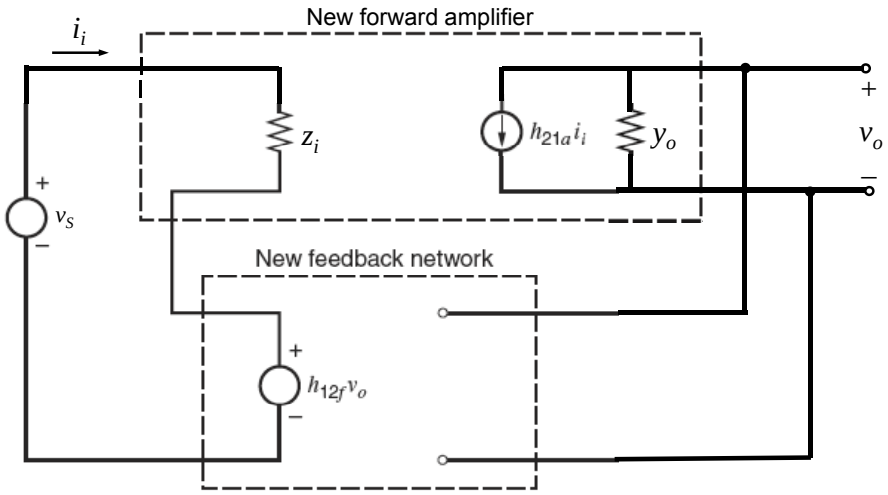
$$A = \frac{v_o}{v_S} = \frac{\left(\frac{-h_{21a}}{z_i y_o}\right)}{1 + \left(\frac{-h_{21a}}{z_i y_o}\right) h_{12f}} = \frac{a}{1 + af}$$

where

$$a = -\frac{h_{21a}}{z_i y_o}$$
$$f = h_{12f}$$

Series-Shunt Feedback Equivalent Two-Port Networks

Equivalent two-port circuits can now be identified for **a** and **f** in which the loading of the feedback network is included in the forward amplifier



Shunt-Series Feedback

In a shunt-series feedback circuit, the forward amplifier and feedback network share the same input voltage and output current. Thus, a hybrid **g-parameter** representation is used for the two-port networks.

$$i_1 = g_{11}v_1 + g_{12}i_2$$
$$v_2 = g_{21}v_1 + g_{22}i_2$$

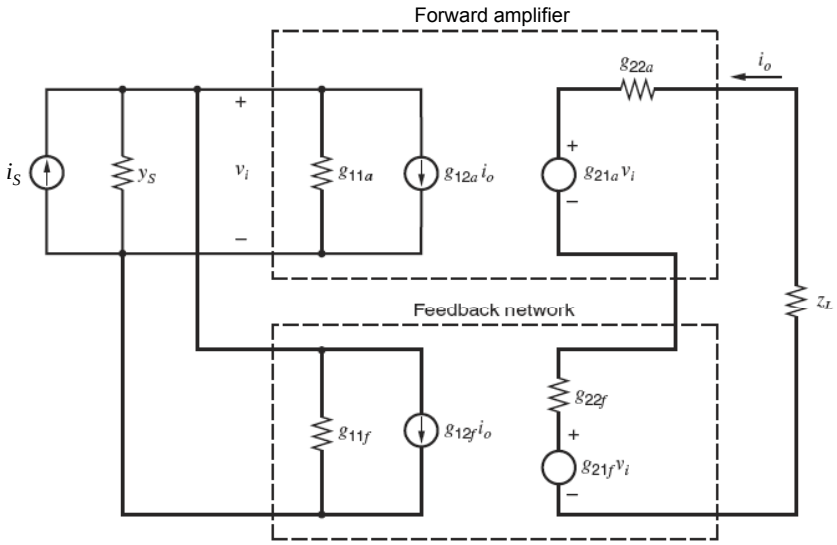
where

$$g_{11} \triangleq \left.\frac{i_1}{v_1}\right|_{i_2=0} \qquad g_{12} \triangleq \left.\frac{i_1}{i_2}\right|_{v_1=0}$$

$$g_{21} \triangleq \left.\frac{v_2}{v_1}\right|_{i_2=0} \qquad g_{22} \triangleq \left.\frac{v_2}{i_2}\right|_{v_1=0}$$

Shunt-Series Feedback, cont'd

The two-port model for a shunt-series feedback amplifier is thus



Shunt-Series Feedback, cont'd

Summing currents at the input and voltages at the output,

$$i_S = (y_S + g_{11a} + g_{11f})v_i + (g_{12a} + g_{12f})i_o$$
$$0 = (g_{21a} + g_{21f})v_i + (z_L + g_{22a} + g_{22f})i_o$$

Define

$$y_i \triangleq y_S + g_{11a} + g_{11f}$$
$$z_o \triangleq z_L + g_{22a} + g_{22f}$$

and neglect reverse transmission through the forward amplifier and feedforward through the feedback network; that is, assume

$$|g_{12a}| \ll |g_{12f}|$$

and

$$|g_{21a}| \gg |g_{21f}|$$

Shunt-Series Feedback Equations

With the simplifying assumptions the shunt-series feedback circuit can described in the form of the ideal feedback equation

$$A = \frac{i_o}{i_S} = \frac{\left(\frac{-g_{21a}}{y_i z_o}\right)}{1+\left(\frac{-g_{21a}}{y_i z_o}\right) g_{12f}} = \frac{a}{1+af}$$

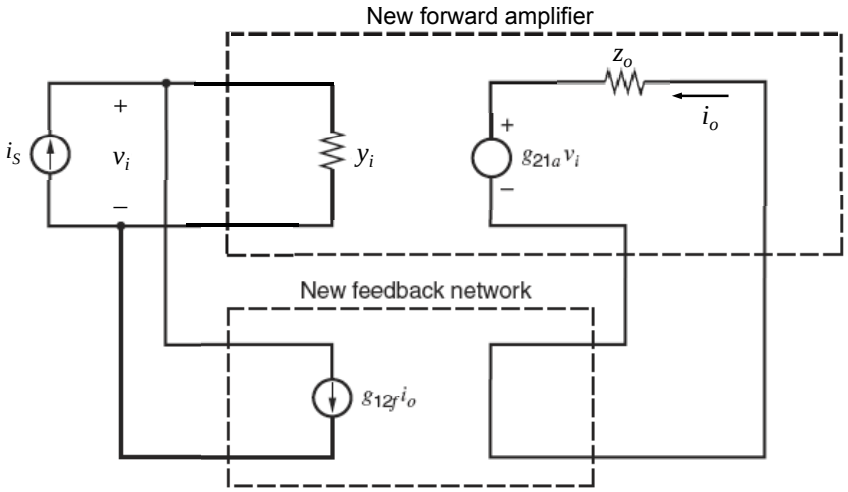
where

$$a = -\frac{g_{21a}}{y_i z_o}$$

$$f = g_{12f}$$

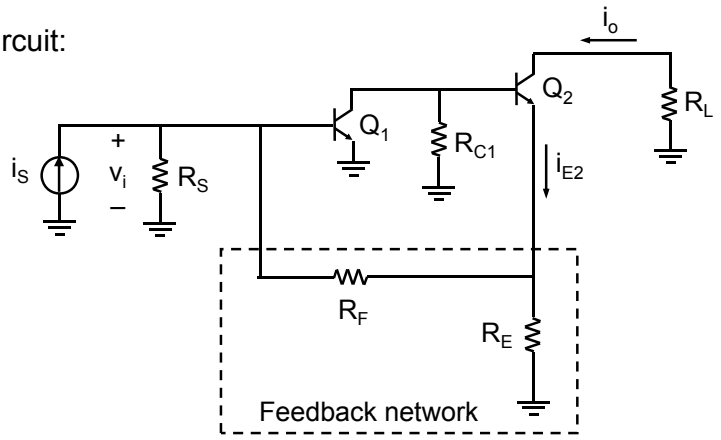
Shunt-Series Feedback Equivalent Two-Port Networks

Based on the above definitions and assumptions, equivalent two-port networks for a shunt-series amplifier can be defined as follows

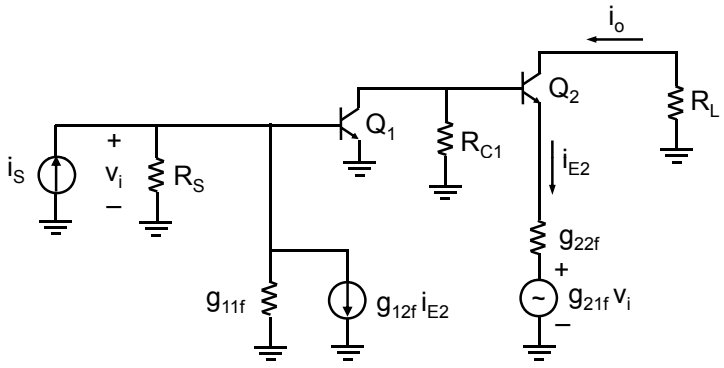


Shunt-Series Feedback Pair

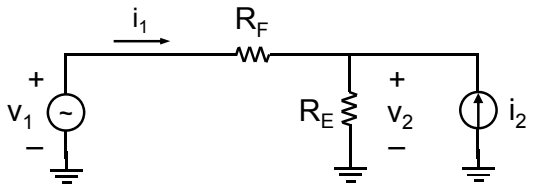
AC circuit:



The forward amplifier in this circuit is not a two-port network because $i_o \neq i_{E2}$. However, the circuit can be analyzed in a fashion similar to the series-series triple. The feedback network can be represented by a two-port characterized by g parameters.



The g parameters for the feedback network can be determined from the following circuit :



$$g_{11f} = \left.\frac{i_1}{v_1}\right|_{i_2=0} = \frac{1}{R_E + R_F}$$

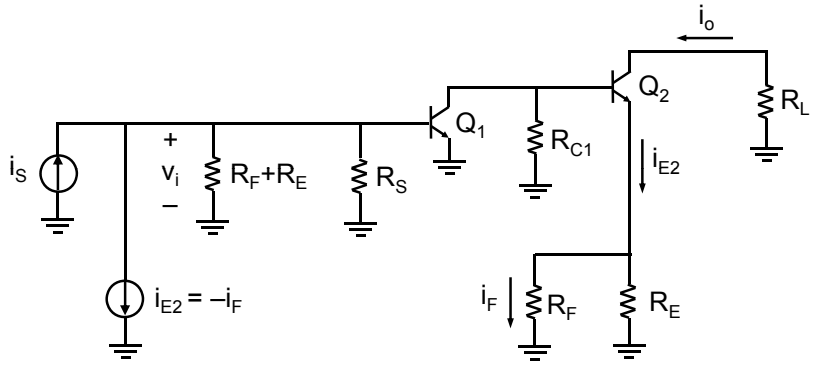
$$g_{22f} = \left.\frac{v_2}{i_2}\right|_{v_1=0} = R_E \,||\, R_F$$

$$g_{12f} = \left.\frac{i_1}{i_2}\right|_{v_1=0} = -\frac{R_E}{R_E + R_F}$$

$$g_{21f} = \left.\frac{v_2}{v_1}\right|_{i_2=0} = -g_{12f}$$

g_{21f} is neglected in comparison with g_{21a}

The shunt-series pair can then be redrawn as follows :



Again, we've managed to break the loop at the output of the feedback network. If a and f are defined in terms of i_o, then

$$A = \frac{i_o}{i_S} = \alpha_2 \frac{i_{E2}}{i_S}$$

$$f = \frac{i_{fb}}{i_o} = -\frac{i_F}{i_o} = \frac{g_{12f}}{\alpha_2} = -\frac{1}{\alpha_2}\left(\frac{R_E}{R_E + R_F}\right)$$

Chapter 6
Frequency Response of Feedback Amplifiers

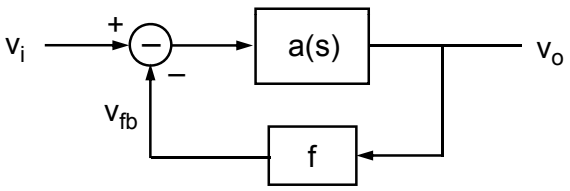
B. Murmann

Stanford University

Reading Material: Section 9.1, 9.2, 9.3, 9.4, 9.5

Gain and Bandwidth

Consider a feedback amplifier with a single pole in the response of the forward-path amplifier and feedback that is frequency independent.



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

$$A(s) \triangleq \frac{v_o(s)}{v_i(s)} = \frac{a(s)}{1 + a(s) \cdot f} = \frac{a(s)}{1 + T(s)}$$

where

$$T(s) \triangleq a(s) \cdot f = \text{"Loop Gain"}$$

$$A(s) = \frac{\dfrac{a_0}{1 - s/p_1}}{1 + \dfrac{a_0 f}{1 - s/p_1}} = \frac{a_0}{1 - \dfrac{s}{p_1} + a_0 f}$$

$$= \frac{a_0}{1 + a_0 f} \cdot \frac{1}{1 - \dfrac{s}{p_1}\left(\dfrac{1}{1 + a_0 f}\right)}$$

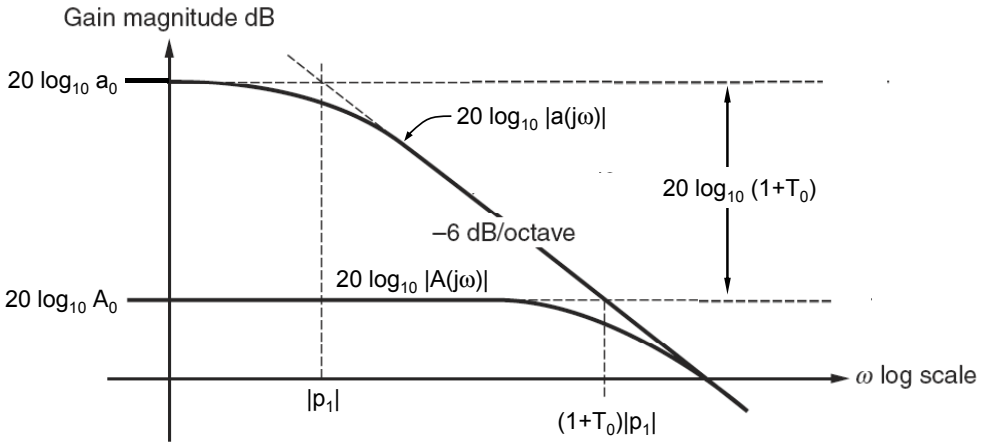
$$= A_o \cdot \frac{1}{\left[1 - \dfrac{s}{p_1(1 + T_0)}\right]}$$

where

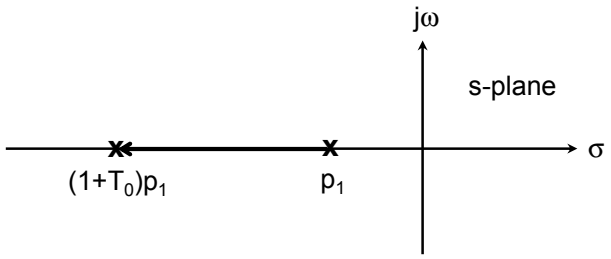
$$A_0 = A(0) = \frac{a_0}{1 + a_0 f}$$

$$T_0 = T(0) = a(0) \cdot f = a_0 f = \text{"Low Frequency Loop Gain"}$$

Thus, feedback reduces the gain by $(1+T_0)$ and increases the –3dB bandwidth by $(1+T_0)$ for a "one-pole" forward-path amplifier.
The Gain x Bandwidth (GBW) product remains constant.



Locus of the pole of A(s) in the s-plane:



Pole "moves" from p_1 to $(1+T_0)p_1$

Note that

$$20 \cdot \log_{10} a_0 - 20 \cdot \log_{10} A_0 = 20 \cdot \log_{10}\left(\frac{a_0}{A_0}\right)$$

$$= 20 \cdot \log_{10}(1+T_0)$$

$$\cong 20 \cdot \log_{10} T_0 \quad \text{when } T_0 >> 1$$

At the frequency $\omega_0 = (1+T_0)|p_1|$

$$\left|a(j\omega_0)\right| = A_0$$

and therefore

$$\left|T(j\omega_0)\right| = \left|a(j\omega_0)\cdot f\right| = A_0 f = \frac{a_0 f}{1+a_0 f}$$

$$\approx 1$$

Thus, ω_0 is the **unity-gain bandwidth** of the loop gain, T(s).

Instability

At a frequency where the phase shift around the loop of a feedback amplifier reaches ±180^o 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^o. 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

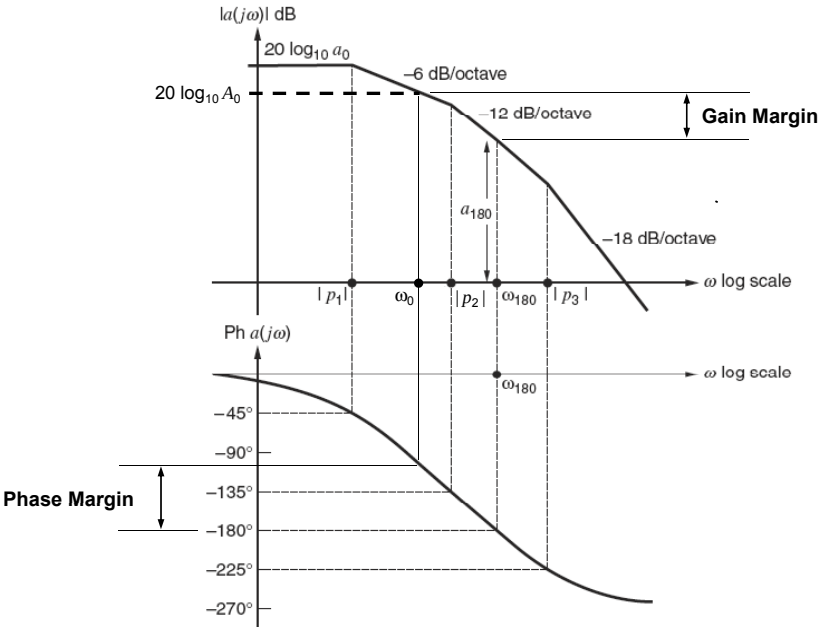
- A feedback system is unstable when $|T(j\omega)| > 1$ at the frequency where $\text{Phase}[T(j\omega)] = -180°$
- Phase margin
  - Defined at the frequency where $|T(j\omega)| = 1 \rightarrow \omega_0$

$$\boxed{PM = 180° + \text{Phase}\left[T(j\omega_0)\right]}$$

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

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

Example



Practical Phase Margins

- Practical circuits typically use phase margins greater 45°
  - For continuous time amplifiers, a common target is ~60°
  - For switched capacitor circuits, a phase margin of ~70° is desirable
    - See EE315A
- In order to see the need for phase margin >45°, investigate the closed-loop behavior of the circuit at $\omega = \omega_0$

$$\left|T(j\omega_0)\right| = \left|a(j\omega_0)\cdot f\right| = 1 \quad \Rightarrow \quad \left|a(j\omega_0)\right| = \frac{1}{f} \qquad \text{(assuming f is real)}$$

$$A(j\omega_0) = \frac{a(j\omega_0)}{1 + a(j\omega_0)/\left|a(j\omega_0)\right|}$$

$$= \frac{a(j\omega_0)}{1 + e^{j\phi[a(j\omega_0)]}} = \frac{a(j\omega_0)}{1 + e^{j(PM-180°)}}$$

PM = 45°

$$A(j\omega_0) = \frac{a(j\omega_0)}{1+e^{-j135^\circ}} = \frac{a(j\omega_0)}{1-0.7-0.7j}$$
$$= \frac{a(j\omega_0)}{0.3-0.7j}$$

$$\therefore \quad |A(j\omega_0)| = \frac{|a(j\omega_0)|}{0.76} = \frac{1.3}{f} \cong 1.3A_0$$

Thus, $|A(j\omega_0)|$ "peaks" at $\omega = \omega_0$, it's nominal –3dB bandwidth.

PM = 60°

$$A(j\omega_0) = \frac{a(j\omega_0)}{1+e^{-j120^\circ}} = \frac{a(j\omega_0)}{1-0.5-0.87j}$$
$$= \frac{a(j\omega_0)}{0.5-0.87j}$$

$$\therefore \quad |A(j\omega_0)| = |a(j\omega_0)| = \frac{1}{f} \cong A_0$$

PM = 90°

$$A(j\omega_0) = \frac{a(j\omega_0)}{1+e^{-j90^\circ}} = \frac{a(j\omega_0)}{1-j}$$

$$\therefore \quad |A(j\omega_0)| = \frac{|a(j\omega_0)|}{1.4} = \frac{0.7}{f} \cong 0.7A_0$$

Closed-Loop Response for Various Phase Margins

Text, p. 632

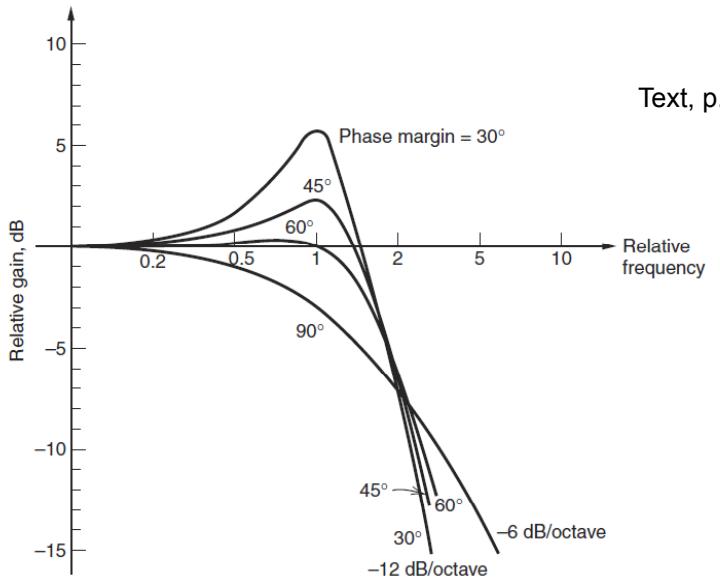


Figure 9.10 Normalized overall gain for feedback amplifiers versus normalized frequency for various phase margins. Frequency is normalized to the frequency where the loop gain is unity.

Frequency Compensation (1)

Frequency compensation refers to the means by which the frequency response of the loop gain in a feedback amplifier is altered so as to ensure adequate phase margin under the expected closed-loop conditions.

Because operational amplifiers must often be compensated for use with a variety of feedback networks, compensation is usually accomplished by modifying only the forward path of the loop (i.e. the op-amp itself). In special purpose wideband amplifier design, the response of the feedback network may also be altered.

Several types of frequency compensation are used in practice, e.g.

Narrowbanding (lag compensation)

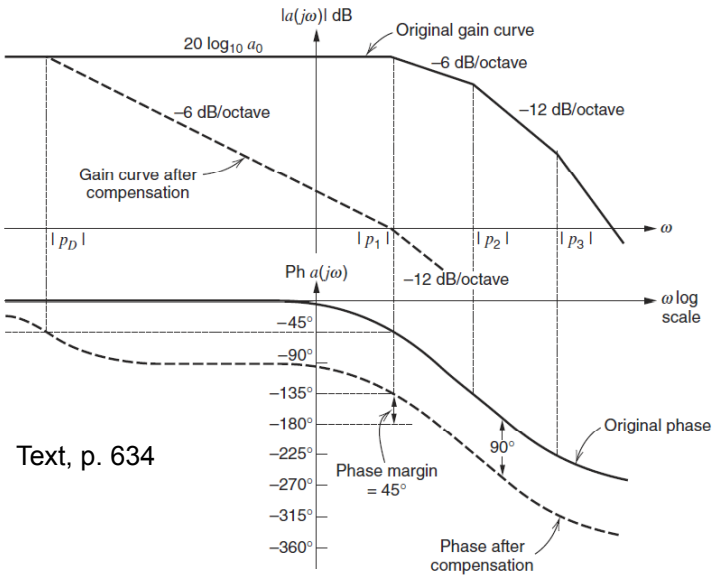
Feedforward (lead compensation)

Pole splitting (Miller compensation, cascode compensation)

Feedback (phantom) zero compensation

Narrowbanding (1)

- Create a dominant pole in a(s) to roll off |T(jω)| at a frequency low enough to ensure adequate phase margin

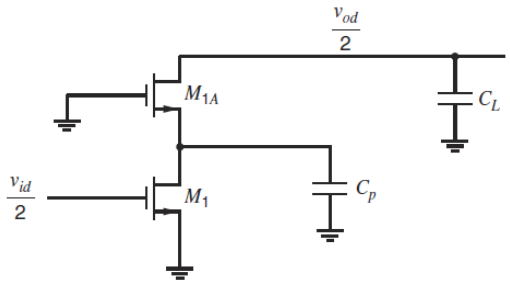


(f=1)

Text, p. 634

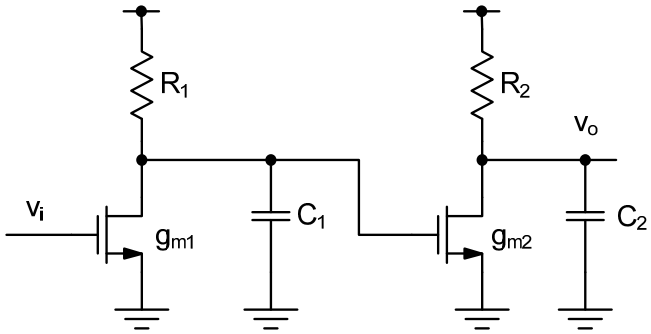
Narrowbanding (2)

- Note that in the example of the previous slide (with f=1, and PM=45°), the closed-loop bandwidth is limited to approximately ω_{p1}, the frequency of the closest non-dominant pole
- This can be acceptable if ω_{p1} is sufficiently large, as the e.g. case for the pole introduced by a cascode
- Consider e.g. the amplifier below
  - C_p introduces a non-dominant pole at high frequencies
  - C_L is adjusted until the circuit achieves the desired phase margin
    - Therefore, this type of narrowbanding is called “load compensation”

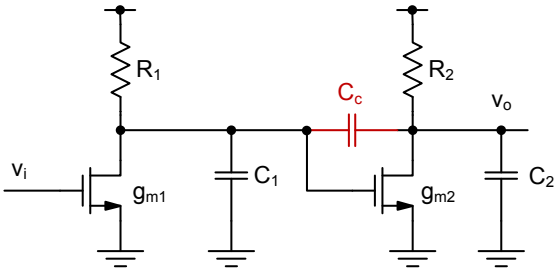


Two-Stage Amplifier

- But, what if we would like to stabilize an amplifier that has two poles at relatively low frequencies?
- Consider e.g. the two-stage amplifier shown below, and assume that $\omega_{p1}=1/R_1C_1$ and $\omega_{p2}=1/R_2C_2$ are comparable



Pole Splitting (1)



- Purposely connect an additional capacitor between gate and drain of M_2 (C_c = “compensation capacitor")
- Two interesting things happen
  - Low frequency input capacitance of second stage becomes large – exactly what we need for low ω_{p1}
  - At high frequencies, C_c turns M2 into a “diode connected device” – low impedance, i.e. large ω_{p2} !

Pole Splitting (2)

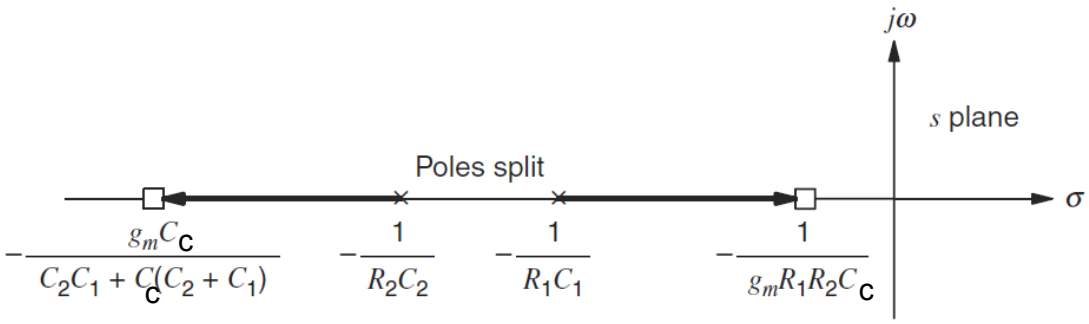
- From the general CS/CE stage analysis of Chapter 4, we can approximate resulting poles and zeros as follows
  - See also text, section 9.4.2

$$p_1 \cong -\frac{1}{g_{m2}R_2R_1C_C} \qquad p_2 \cong -\frac{g_{m2}C_C}{C_1C_2 + C_C(C_1 + C_2)} \qquad z = +\frac{g_{m2}}{C_C}$$

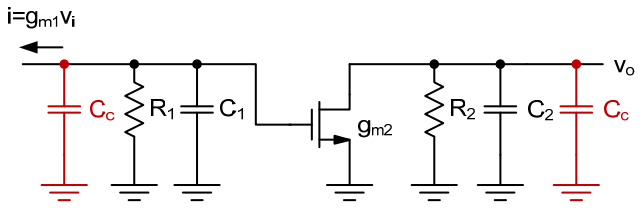
- Increasing C_c reduces ω_{p1}, and increases ω_{p2}
  - A very nice "knob" for adjusting the phase margin of the circuit
- The zero introduced due to C_c occurs at high frequencies and is not always troublesome
- In cases where the zero affects stability, several options exists to mitigate the problem
  - Nulling resistor, cascode compensation, etc.
  - See EE114, EE315A

Pole Splitting (3)

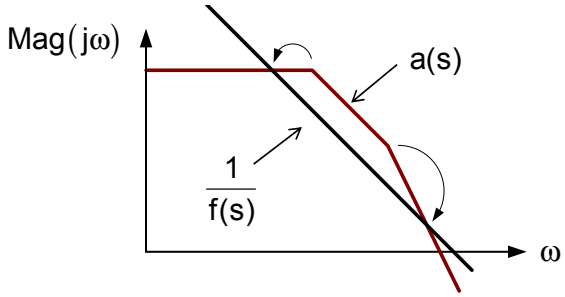
Text, p. 642



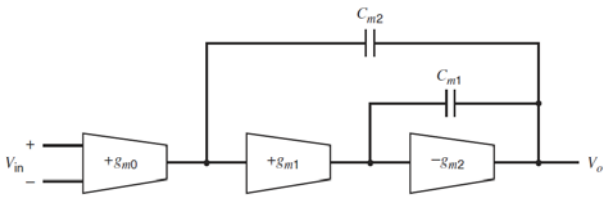
Intuitive Derivation of Pole Split Using Shunt-Shunt Feedback



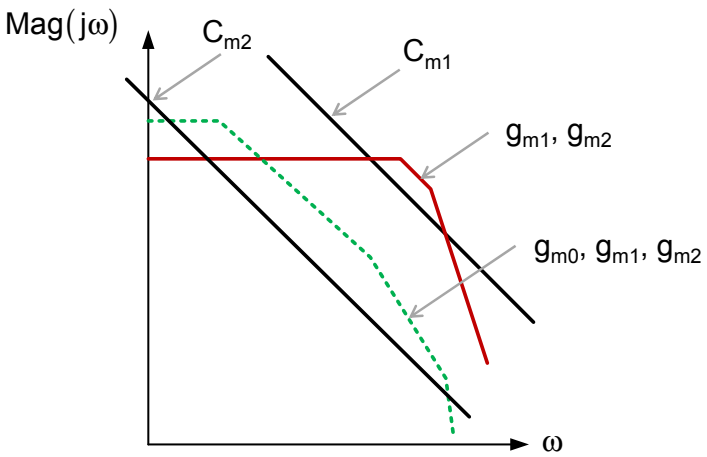
$$a(s) = -\frac{g_{m2}R_2R_1}{(1+sR_1[C_1+C_c])(1+sR_2[C_2+C_c])} \qquad f(s) = -sC_f$$



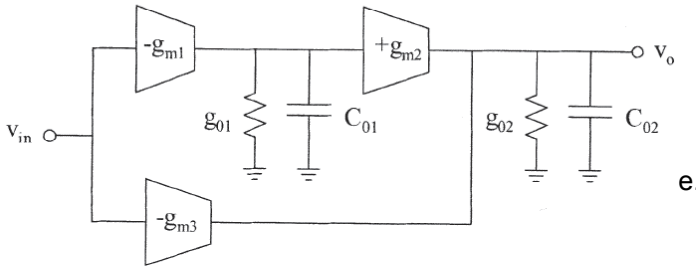
Nested Miller Compensation



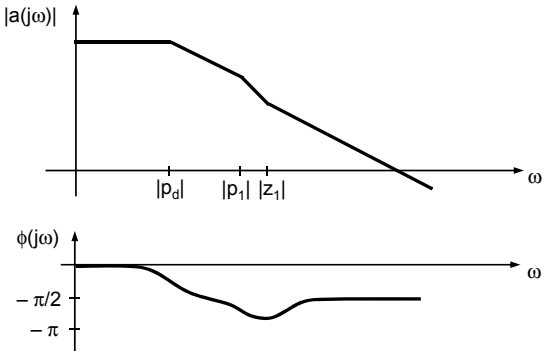
Text, p. 655



Feedforward Compensation



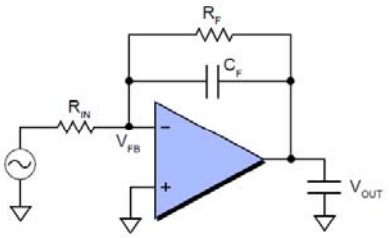
e.g. Thandri, JSSC 2/2003

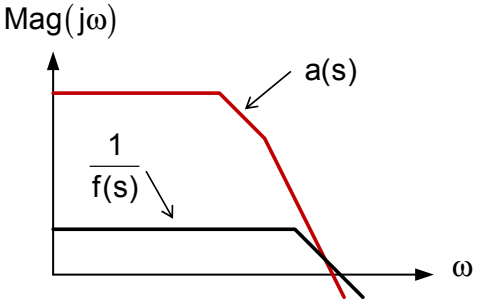


- Parallel path through g_{m3} dominates transfer function at high frequencies and returns the circuit behavior back to first order
- The doublet p_1, z_1 can make it difficult to achieve a fast transient response
  - See Kamath, JSSC 12/1974

Feedback Zero Compensation

Example:





- Due to the zero introduced in the feedback network, T(jω) behaves like a first order system near unity crossover
- Closed-loop bandwidth is approximately equal to the frequency of the zero in the feedback network
  - Can be far beyond second pole

Introduction to Root Locus Techniques: Three Pole Amplifier

- Consider a feedback network consisting of a forward amplifier with three identical poles, and a feedback network with a constant transfer function f

$$a(s)=\frac{a_0}{\left(1-\frac{s}{p_1}\right)^3} \qquad A(s)=\frac{a(s)}{1+a(s)f}=\frac{\frac{a_0}{\left(1-\frac{s}{p_1}\right)^3}}{1+a\frac{a_0}{\left(1-\frac{s}{p_1}\right)^3}f}=\frac{a_0}{\left(1-\frac{s}{p_1}\right)^3+T_0} \qquad T_0=a_0 f$$

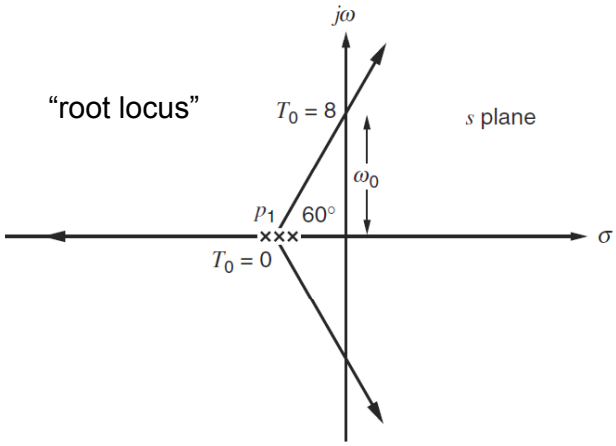
- The poles of A(s) are therefore the solution to

$$\left(1-\frac{s}{p_1}\right)^3+T_0=0$$

$$\left(1-\frac{s}{p_1}\right)^3=-T_0$$

$$\left(1-\frac{s}{p_1}\right)=\sqrt[3]{-T_0}=-\sqrt[3]{T_0} \quad \text{or} \quad \left(1-\frac{s}{p_1}\right)=\sqrt[3]{T_0}e^{j60^\circ} \quad \text{or} \quad \left(1-\frac{s}{p_1}\right)=\sqrt[3]{T_0}e^{-j60^\circ}$$

$$s_1=p_1\left(1+\sqrt[3]{T_0}\right)$$
$$s_2=p_1\left(1-\sqrt[3]{T_0}e^{j60^\circ}\right)$$
$$s_3=p_1\left(1-\sqrt[3]{T_0}e^{-j60^\circ}\right)$$

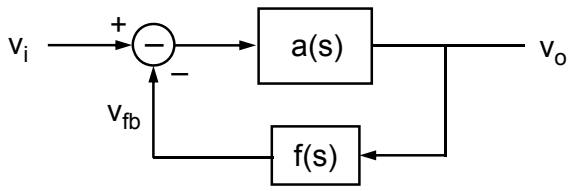


$$0=1-\mathrm{Re}\left(\sqrt[3]{T_0}e^{j60^\circ}\right)$$
$$0=1-\sqrt[3]{T_0}\cos(60^\circ)$$
$$\Rightarrow T_0=8$$

Root-Locus Method

We can generalize the above example to gain insight into the frequency response of any feedback amplifier by examining the "movement" of the closed-loop poles in the s-plane as a function of the low-frequency loop gain, T_0.

Consider a generalized feedback amplifier with both a and f dependent on frequency.



$$A(s) \triangleq \frac{v_o(s)}{v_i(s)} = \frac{a(s)}{1+a(s)f(s)} = \frac{a(s)}{1+T(s)}$$

The poles of A(s) are the roots of 1 + T(s) = 0.

In general,

$$a(s) = a_0 \frac{1+a_1 s + a_2 s^2 + \cdots}{1+b_1 s + b_2 s^2 + \cdots} = a_0 \frac{N_a(s)}{D_a(s)}$$

and

$$f(s) = f_0 \frac{1+c_1 s + c_2 s^2 + \cdots}{1+d_1 s + d_2 s^2 + \cdots} = f_0 \frac{N_f(s)}{D_f(s)}$$

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 ∞, the poles of A(s) move in the s-plane from the poles of T(s) to the zeros of T(s).

The **Root Locus** is the paths the roots of 1 + T(s) = 0 trace in the s-plane as T_0 varies from 0 to ∞.

The root-locus construction rules follow from the equation

$$1 + T(s) = 1 + T_0 \cdot \frac{N_a(s) \cdot N_f(s)}{D_a(s) \cdot D_f(s)} = 0$$

or equivalently

$$T_0 \cdot \frac{N_a(s) \cdot N_f(s)}{D_a(s) \cdot D_f(s)} = -1$$

Thus

$$T_0 \cdot \frac{(1 - s/z_{a1})(1 - s/z_{a2}) \cdots (1 - s/z_{f1})(1 - s/z_{f2}) \cdots}{(1 - s/p_{a1})(1 - s/p_{a2}) \cdots (1 - s/p_{f1})(1 - s/p_{f2}) \cdots} = -1$$

where

$z_{a1}, z_{a2}, \ldots$ = zeros of a(s)
$z_{f1}, z_{f2}, \ldots$ = zeros of f(s)
} zeros of T(s)

$p_{a1}, p_{a2}, \ldots$ = poles of a(s)
$p_{f1}, p_{f2}, \ldots$ = poles of f(s)
} poles of T(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,

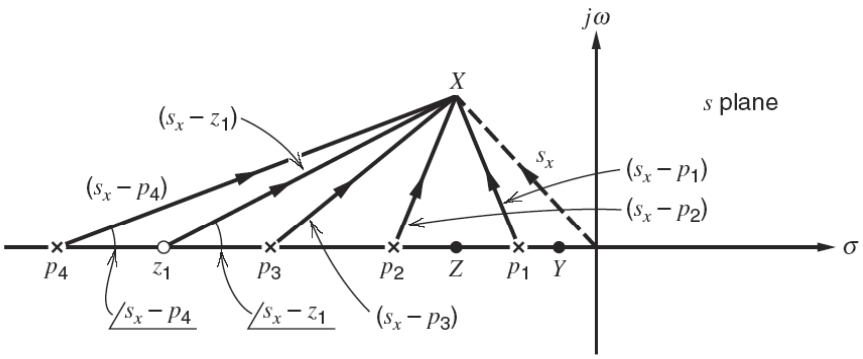
$$T_0 \cdot \left[\frac{|p_{a1}||p_{a2}|\cdots|p_{f1}||p_{f2}|\cdots}{|z_{a1}||z_{a2}|\cdots|z_{f1}||z_{f2}|\cdots}\right] \cdot \left[\frac{|s-z_{a1}||s-z_{a2}|\cdots|s-z_{f1}||s-z_{f2}|\cdots}{|s-p_{a1}||s-p_{a2}|\cdots|s-p_{f1}||s-p_{f2}|\cdots}\right] = +1$$

and

$$\begin{aligned}&\left[\angle(s-z_{a1})+\angle(s-z_{a2})+\cdots+\angle(s-z_{f1})+\angle(s-z_{f2})+\cdots\right]\\&\quad-\left[\angle(s-p_{a1})+\angle(s-p_{a2})+\cdots+\angle(s-p_{f1})+\angle(s-p_{f2})+\cdots\right]\\&\quad=2n\pi\end{aligned}$$

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:

The root locus is located along the real axis whenever there is an odd number of poles and zeros of T(s) between that portion of the real axis and the origin of the s-plane.

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:

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

$$(2n-1)\pi/(N_p - N_z),$$

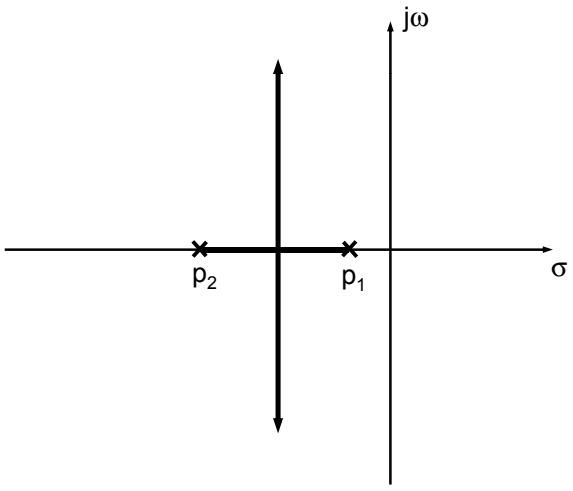
where N_p = # of poles of T(s) and N_z = # of zeros of T(s).

Rule 8:

The asymptotes of the branches terminating at infinity all intersect the real axis at a single point given by

$$\sigma_a = \frac{\Sigma[\text{poles of } T(s)] - \Sigma[\text{zeros of } T(s)]}{N_p - N_z}$$

2-Pole Example



3-Pole Example

Assume

$$T(s) = \frac{T_0}{(1 - s/p_1)(1 - s/p_2)(1 - s/p_3)}$$

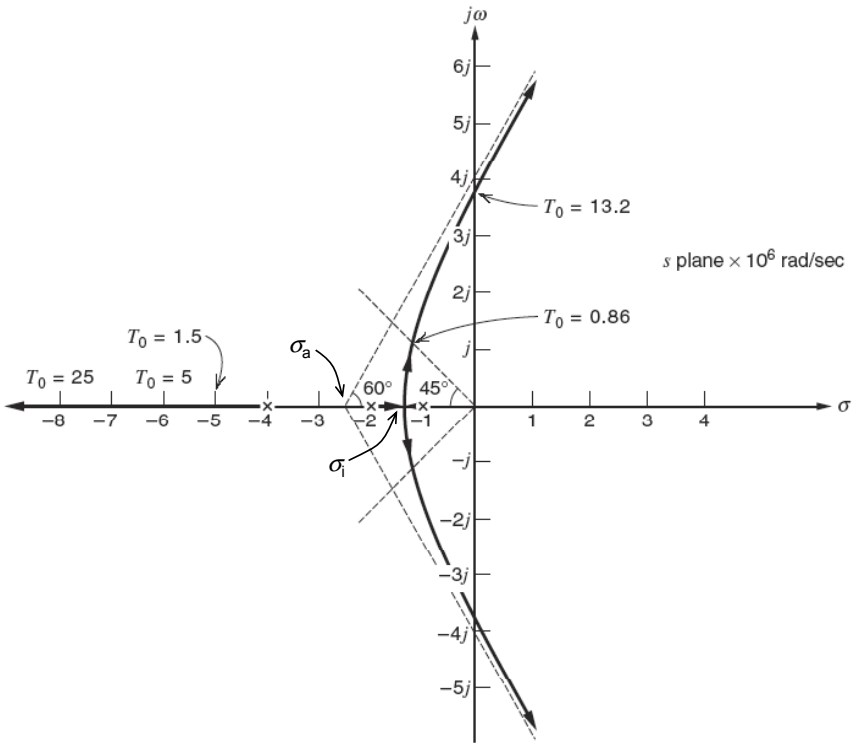
where

$$p_1 = -1 \times 10^6 \text{ sec}^{-1}$$

$$p_2 = -2 \times 10^6 \text{ sec}^{-1}$$

$$p_3 = -4 \times 10^6 \text{ sec}^{-1}$$

3-Pole Example, cont'd



Breakaway Point, σ_i

From **Rule 6**

$$\frac{1}{\sigma_i + 1} + \frac{1}{\sigma_i + 2} + \frac{1}{\sigma_i + 4} = 0$$

$$\therefore \; 3\sigma_i^2 + 14\sigma_i + 14 = 0$$

$$\sigma_i = -\frac{7}{3} \pm \sqrt{\left(\frac{7}{3}\right)^2 - \frac{14}{3}} = -\frac{1}{3}\left(7 \pm \sqrt{7}\right)$$

$$= -3.22 \text{ or } -1.45$$

Since σ_i lies between p_1 and p_2,

$$\sigma_i = -1.45 \times 10^6 \text{ sec}^{-1}$$

Asymptote Intercept, σ_a

From **Rule 8**

$$\sigma_a = \frac{(-1-2-4)-0}{3} = -2.33 \times 10^6 \text{ sec}^{-1}$$

To find the loop gain, T_0, at which the poles moving from p_1 and p_2 become complex, substitute the breakaway point, $s = \sigma_i = -1.45$, into the magnitude condition and solve for T_0.

$$T_0 \frac{|p_1||p_2||p_3|}{|s-p_1||s-p_2||s-p_3|} = 1$$

$$\therefore \; T_0 = \frac{|-1.45+1||-1.45+2||-1.45+3|}{(1)(2)(4)}$$

$$= \frac{(0.45)(0.55)(2.55)}{8} = 0.08$$

The complex poles enter the RHP at approximately the points where their asymptotes cross the jω axis. At these points

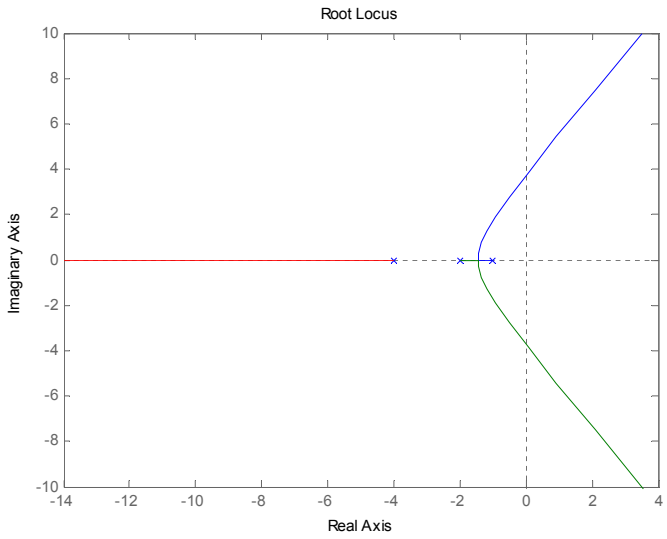
$$s = \pm j(2.33)\tan(60^\circ) = \pm 4j$$

Substituting into the magnitude condition and solving for T_0

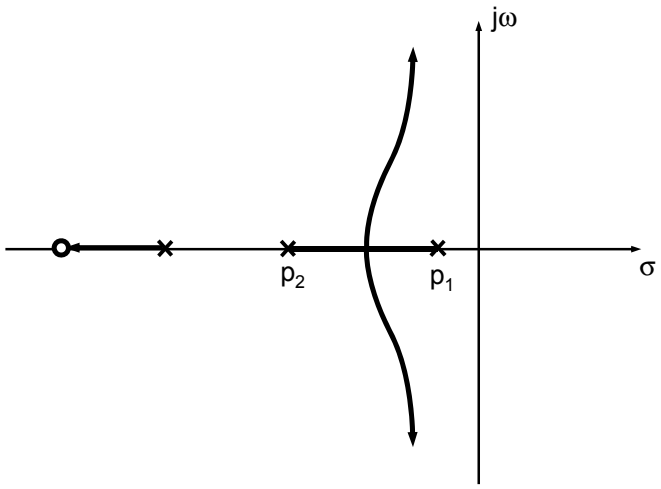
$$\begin{aligned} T_0 &= \frac{|s-p_1||s-p_2||s-p_3|}{|p_1||p_2||p_3|} \\ &= \frac{|4j+1||4j+2||4j+4|}{(1)(2)(4)} \\ &= \frac{\left(\sqrt{4^2+1}\right)\left(\sqrt{4^2+2^2}\right)\left(\sqrt{4^2+4^2}\right)}{8} = \frac{\sqrt{17}\cdot\sqrt{20}\cdot\sqrt{32}}{8} \\ &= \frac{(4.1)(4.5)(5.7)}{8} = 13.2 \end{aligned}$$

Plotting the Root Locus in Matlab

```
% root locus example
s = tf('s');
p1=-1; p2=-2; p3=-4;
a = 1 / [(1-s/p1)*(1-s/p2)*(1-/p3)]
rlocus(a)
```



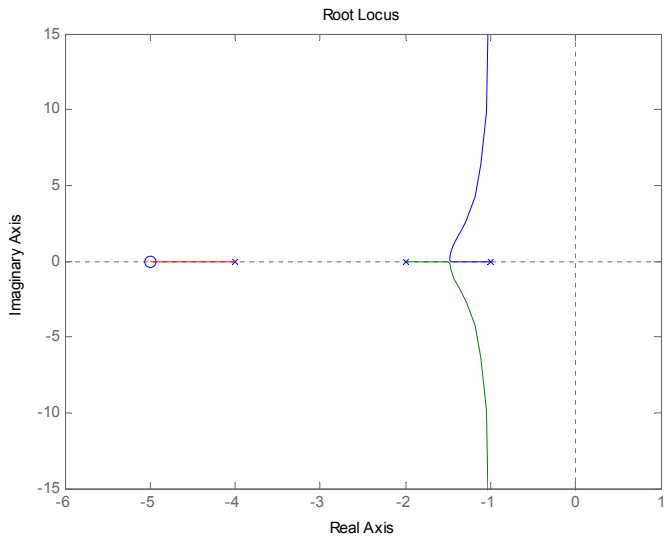
3 Poles and 1 Zero



Introducing a zero in T(s) can be used to stabilize a feedback amplifier. The use of zeros in the feedback path is an important aspect of wideband amplifier design.

Matlab Plot

```
% root locus example
s = tf('s');
z=-5; p1=-1; p2=-2; p3=-4;
a = (1-s/z) / [(1-s/p1)*(1-s/p2)*(1-s/p3)]
rlocus(a)
```



Chapter 7
Wideband Amplifiers

B. Murmann

Stanford University

Reading Material: Section 9.5.3, 9.5.4

Overview

- How to build amplifiers with large gain and bandwidth?

- Options we'll look at
  - Cascade of first order amplifier stages
  - Multi-stage amplifiers with global feedback
  - Multi-stage amplifiers with local feedback

Cascade of N First-Order Stages

Consider a cascade of N identical stages, each with the single-pole response

$$A(j\omega) = \frac{A_0}{1 + j\omega/\omega_B} \qquad GBW = A_0\omega_B$$

If there is no interaction among the stages, then the overall transfer function of the cascade is

$$A_N(j\omega) = \left(\frac{A_0}{1 + j\omega/\omega_B} \right)^N$$

In this case

$$A_{0N} = A_0^N$$

The –3dB bandwidth of the cascade is the frequency, ω_{BN}, at which

$$\left|A_N(j\omega_{BN})\right| = \frac{A_{0N}}{\sqrt{2}}$$

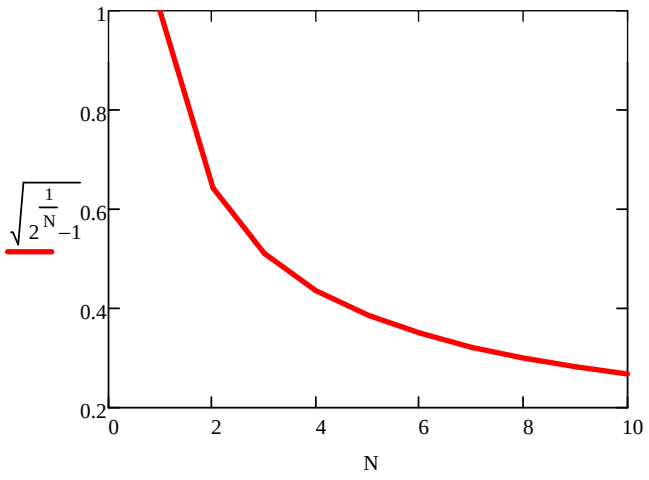
Thus,

$$\left[\frac{A_0}{\sqrt{1 + (\omega_{BN}/\omega_B)^2}} \right]^N = \frac{A_0^N}{\sqrt{2}}$$

$$1 + (\omega_{BN}/\omega_B)^2 = 2^{1/N}$$

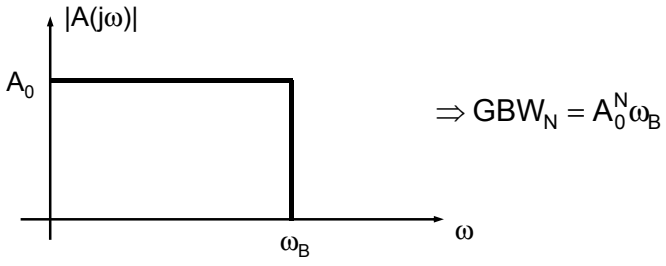
$$\omega_{BN} = \omega_B \cdot \sqrt{2^{1/N} - 1} \qquad GBW_N = A_0^N\omega_B \cdot \sqrt{2^{1/N} - 1} < A_0^N\omega_B$$

Bandwidth Shrinkage



Ideal Lowpass Response

Suppose each stage in the cascade of N identical stages had an ideal lowpass response; that is, a magnitude response with an abrupt band edge transition between the passband and the stopband:



In the design of wideband amplifiers it is therefore often desirable to have a much sharper band-edge transition than is obtained with a cascade of identical single-pole stages.

Maximally Flat Magnitude (MFM) Response

An approximation to the ideal lowpass response that is commonly used in broadband amplifier design is the **Maximally Flat Magnitude** (MFM), or **Butterworth**, response.

The MFM response provides a magnitude (gain) that is flat over as much of the passband as possible and decreases monotonically with frequency (i.e. no peaking). It is obtained by setting as many derivatives of the magnitude with respect to frequency as possible to zero at $\omega = 0$.

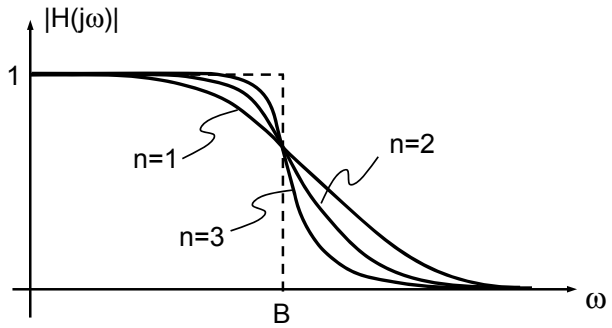
For an n^{th}-order MFM response,

$$\left.\frac{d^k\left|H(j\omega)\right|}{d\omega^k}\right|_{\omega=0} = 0 \quad \text{for} \quad 1 \le k \le 2n-1$$

The resulting magnitude response is

$$\left|H(j\omega)\right| = \frac{1}{\sqrt{1+(\omega/B)^{2n}}}$$

where n is the order of the response, B is the bandwidth of the ideal lowpass response, and the magnitude has been normalized to unity at $\omega = 0$.



There is no peaking in the MFM response, and as n increases the band edge transition becomes sharper. In the limit as n → ∞, the MFM response approaches the ideal lowpass response.

To determine the location of the poles of H(s), note that

$$\begin{aligned}\left|H(j\omega)\right|^2 &= H(j\omega)\cdot H(-j\omega)\\ &= H(s)\cdot H(-s)\Big|_{s=j\omega}\end{aligned}$$

Thus, for an MFM response

$$\begin{aligned}H(s)\cdot H(-s) &= \frac{1}{1+(s/jB)^{2n}} = \frac{j^{2n}}{j^{2n}+(s/B)^{2n}}\\ &= \frac{(-1)^n}{(-1)^n+(s/B)^{2n}}\end{aligned}$$

The poles of H(s)×H(–s) are the 2n roots satisfying

$$\left(\frac{s}{B}\right)^{2n} = -(-1)^n = (-1)^{n+1}$$

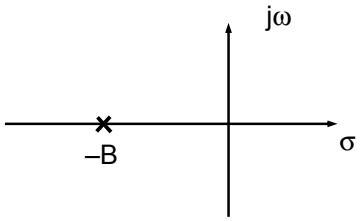
These roots lie equally spaced on a circle in the s-plane centered at the origin with radius B.

The LHP roots are taken to be the poles of H(s), while those in the RHP are regarded as the poles of H(–s).

MFM Pole Locations

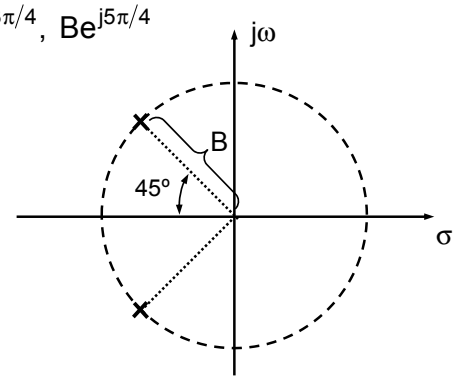
n = 1

$$p_1 = -B$$



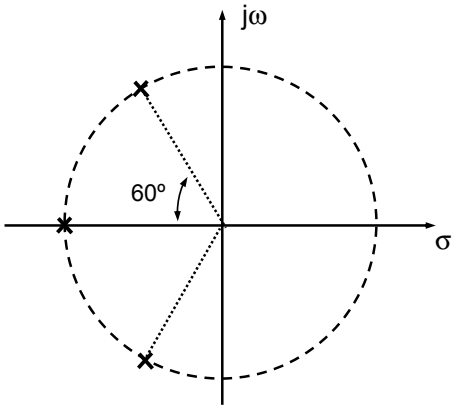
n = 2

$$p_i = Be^{j3\pi/4},\ Be^{j5\pi/4}$$



n = 3

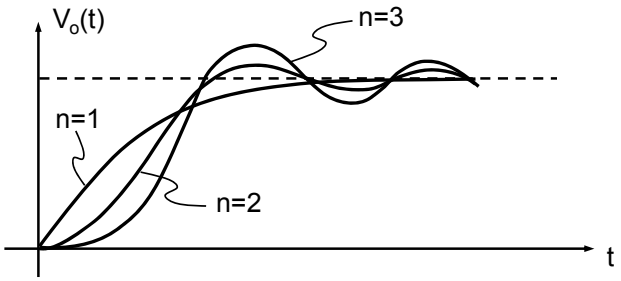
$$p_i = Be^{j2\pi/3},\ Be^{j\pi},\ Be^{j4\pi/3}$$



MFM Transient Response

A potential disadvantage of the MFM response is that there is an overshoot in the step response when n ≥ 2. This may be an important consideration in applications such as pulse amplifiers.

For a step input at t=0



n = 2 → 4% overshoot

n = 3 → 8% overshoot

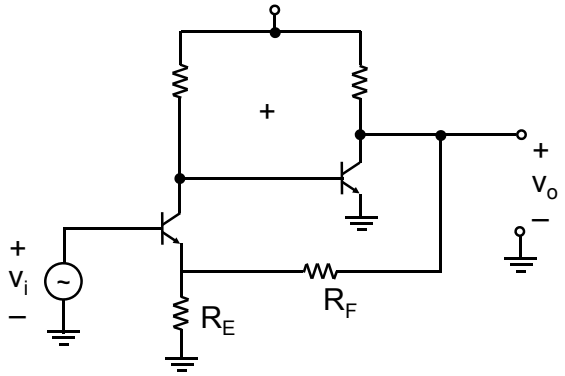
Comparison of 2-Pole Responses

	Angle of Poles	–3dB BW	Overshoot	10-90% Rise Time
Coincident poles	0°	0.64B	0	3.4/B
MFM (Butterworth)	45°	B	4%	2.2/B
Chebyshev (1-dB ripple)	60°	1.3B	16%	1.6/B
Bessel	30°	0.8B	0.4%	2.7/B

More in EE315A

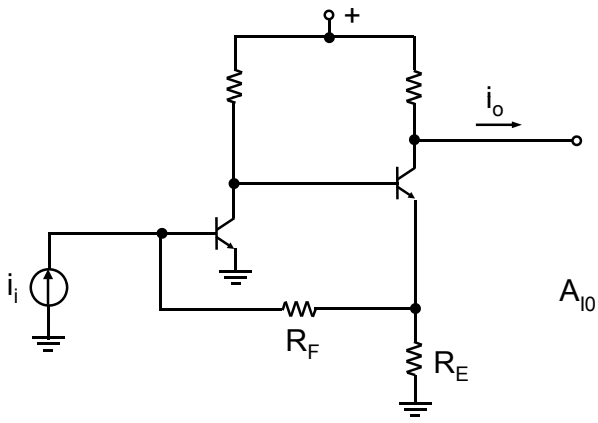
Wideband Amplification Using Multi-Stage Feedback

At most three gain stages can be effectively used within a single feedback loop. Therefore, there are four basic configurations.



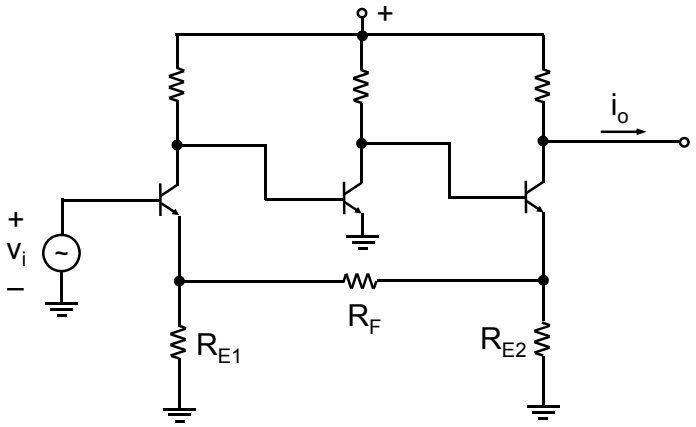
Series-Shunt Pair

$$A_{V0} = \frac{v_o}{v_i} \approx \frac{R_F}{R_E}$$



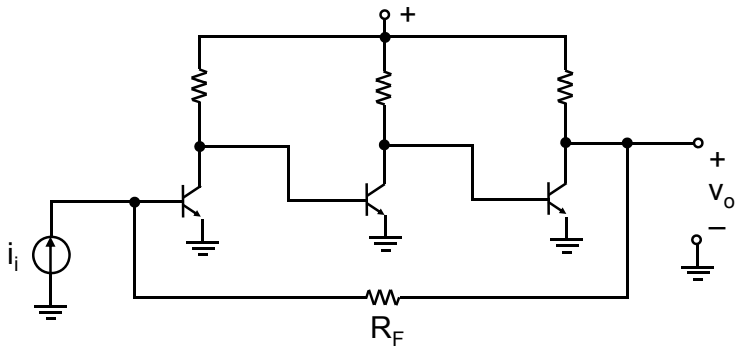
$$A_{I0} = \frac{i_o}{i_i} \approx \frac{R_F}{R_E}$$

Shunt-Series Pair



$$A_0 = \left|\frac{i_o}{v_i}\right| \approx \frac{R_F + R_{E1} + R_{E2}}{R_{E1} \cdot R_{E2}}$$

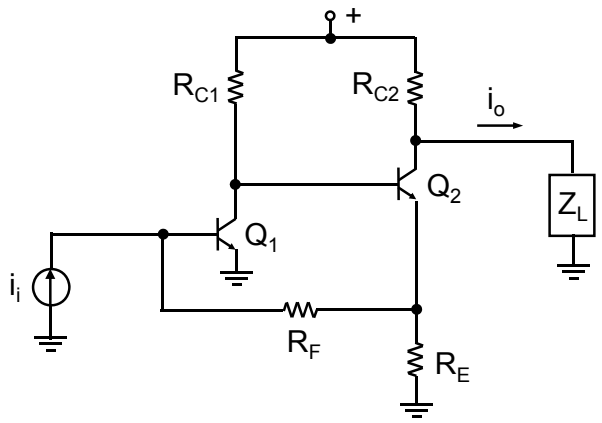
Series-Series Triple



$$A_0 = \left|\frac{v_o}{i_i}\right| \approx R_F$$

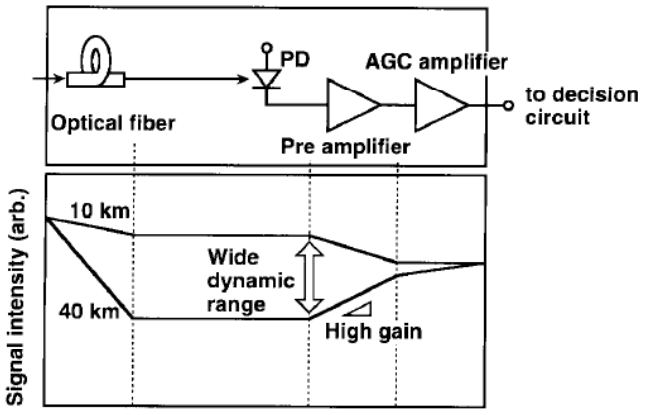
Shunt-Shunt Triple

Analysis of Shunt-Series Pair



Assume $Z_L \approx 0$ and $R_S \approx \infty$

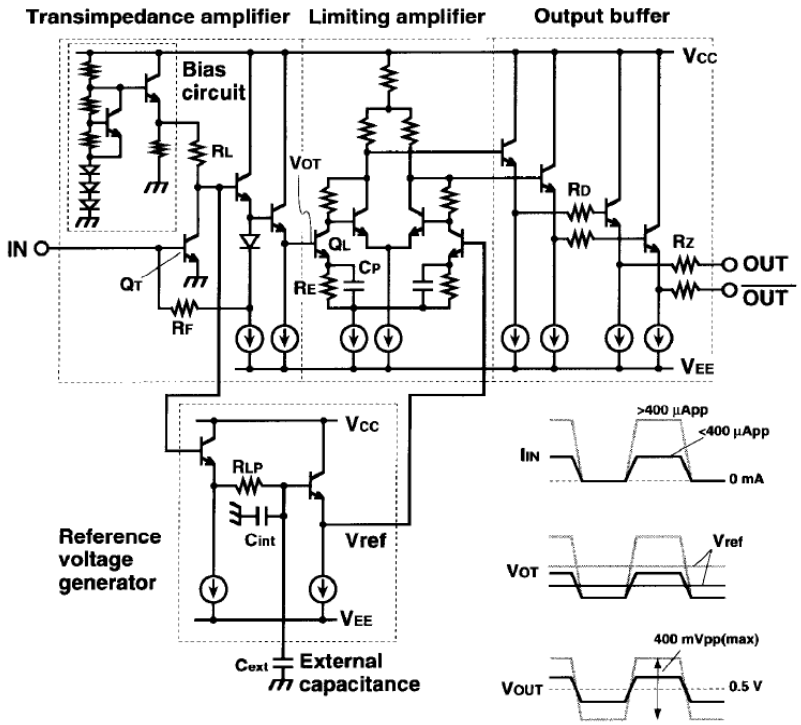
Application Example: Optical Fiber Links (1)



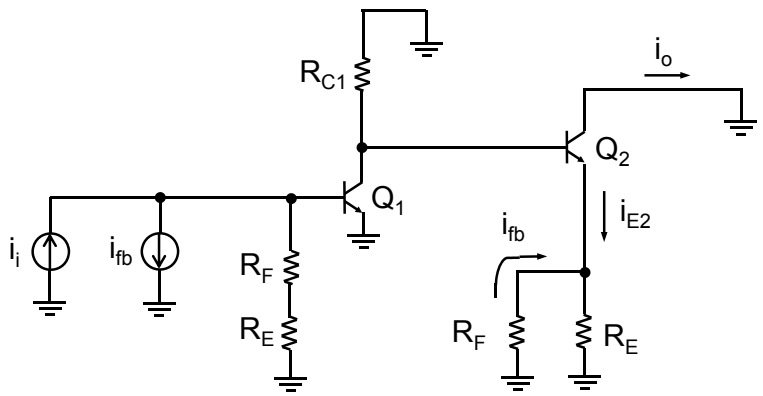
K. Ohhata et al., "A Wide-Dynamic-Range, High-Transimpedance Si Bipolar Preamplifier IC for 10-Gb/s Optical Fiber Links," JSSC 1/1999.

Circuit Implementation

Ohhata,
JSSC 1/1999



Small-signal equivalent circuit for evaluating the open-loop response of the shunt-series pair:



Note that in BJT amplifiers with a series output stage, the feedback network does not sample the output, i_o, directly, but rather the emitter current of Q_2, i_{E2}. Thus, a(s) and f(s) are defined as follows:

$$a(s) \triangleq \frac{i_{E2}}{i_i}$$

$$f(s) \triangleq \frac{i_{fb}}{i_{E2}}$$

For these definitions

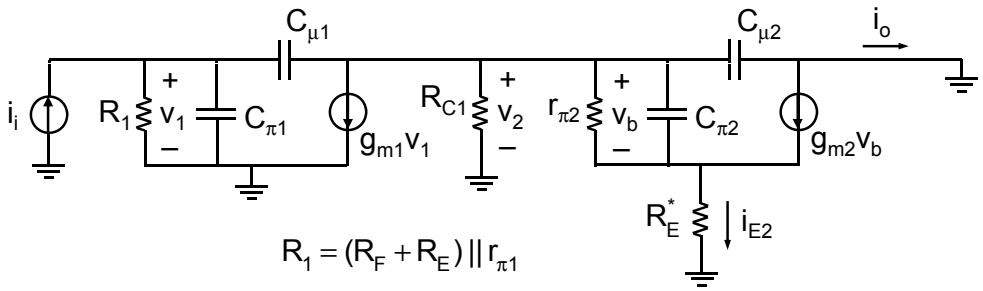
$$A(s) \triangleq \frac{i_o}{i_i} = -\alpha_2 \left(\frac{a(s)}{1 + a(s)f(s)} \right)$$

where α_2 is the common-base current gain of Q_2

The feedback response, f(s), for the shunt-series pair (w/o compensation) is simply

$$f(s) = \frac{i_{fb}}{i_{E2}} = -\frac{R_E}{R_E + R_F}$$

To estimate the forward path response, a(s), begin with the following small-signal equivalent circuit:



$$R_1 = (R_F + R_E) \,||\, r_{\pi 1}$$

$$R_E^* = R_E \,||\, R_F$$

First consider the transmission from the base to the emitter of the second stage

$$i_{E2} = (v_2 - i_{E2}R_E^*)(g_{m2} + Y_{\pi 2})$$

where

$$Y_{\pi 2} \triangleq g_{\pi 2} + sC_{\pi 2}$$

Thus,

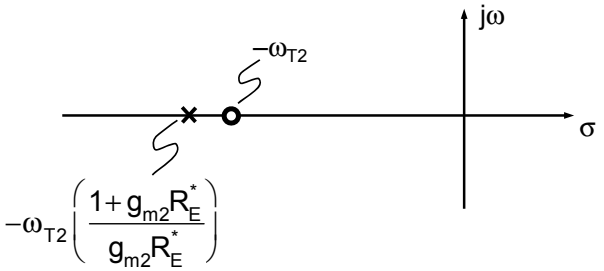
$$i_{E2}\left[1 + R_E^*(g_{m2} + Y_{\pi 2})\right] = v_2(g_{m2} + Y_{\pi 2})$$

$$\therefore \quad \frac{i_{E2}}{v_2} = \frac{g_{m2} + Y_{\pi 2}}{1 + R_E^*(g_{m2} + Y_{\pi 2})}$$

$$g_{m2} + Y_{\pi 2} = g_{m2} + \frac{g_{m2}}{\beta_0} + sC_{\pi 2} \cong g_{m2} + sC_{\pi 2}$$
$$= g_{m2}\left(1 + s\frac{C_{\pi 2}}{g_{m2}}\right) = g_{m2}\left(1 + s\tau_{T2}\right)$$

Therefore,

$$\frac{i_{E2}}{v_2} = \frac{g_{m2}(1 + s\tau_{T2})}{1 + g_{m2}R_E^* + sR_E^*C_{\pi 2}}$$
$$= \left(\frac{g_{m2}}{1 + g_{m2}R_E^*}\right)\left[\frac{1 + s\tau_{T2}}{1 + s\left(\frac{R_E^*C_{\pi 2}}{1 + g_{m2}R_E^*}\right)}\right] = \left(\frac{g_{m2}}{1 + g_{m2}R_E^*}\right)\left[\frac{1 + s\tau_{T2}}{1 + s\tau_{T2}\left(\frac{g_{m2}R_E^*}{1 + g_{m2}R_E^*}\right)}\right]$$

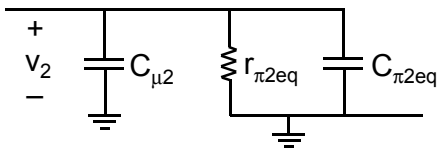


Thus, the second stage is very broadband, and the pole-zero pair associated with it can usually be neglected. In that case,

$$\frac{i_{E2}}{v_2} \cong \frac{g_{m2}}{1+g_{m2}R_E^*} \triangleq g_{m2eq}$$

Also, $i_o = -\alpha_2 i_{E2} \cong -i_{E2}$

The main influence of the series feedback stage on a(s) is the loading on Q_1. We can model the loading using the following approximation for the input impedance of the series feedback stage:

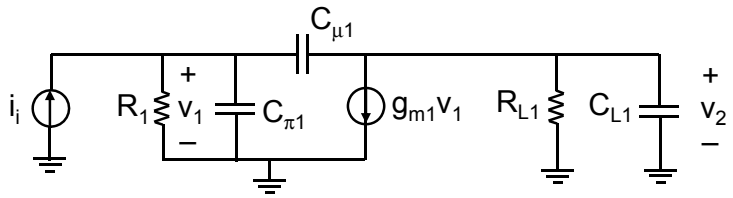


where

$$r_{\pi 2eq} = r_{\pi 2}(1+g_{m2}R_E^*) \cong r_{\pi 2}g_{m2}R_E^* = \beta_{02}R_E^*$$

$$C_{\pi 2eq} = \frac{C_{\pi 2}}{1+g_{m2}R_E^*} \cong \frac{C_{\pi 2}}{g_{m2}}\left(\frac{1}{R_E^*}\right) = \frac{\tau_{T2}}{R_E^*}$$

Thus, the equivalent circuit for determining the open-loop response $v_2(s)/i_i(s)$ is simply:



where

$$R_1 = r_{\pi 1} \,||\, (R_F + R_E)$$

$$R_{L1} = R_{C1} \,||\, r_{\pi 2eq} \cong R_{C1} \,||\, \beta_{02} R_E^*$$

$$C_{L1} = C_{cs1} + C_{\mu 2} + C_{\pi 2eq} \cong C_{cs1} + C_{\mu 2} + \frac{\tau_{T2}}{R_E^*}$$

Since

$$a(s) \triangleq \frac{i_{E2}}{i_i} = \left(\frac{i_{E2}}{v_2}\right)\left(\frac{v_2}{i_i}\right)$$

and

$$\frac{i_{E2}}{v_2} \cong g_{m2eq} = \frac{g_{m2}}{1 + g_{m2} R_E^*} \cong \frac{1}{R_E^*}$$

we find

$$a(s) = -\left[g_{m1} R_1 \left(\frac{R_{L1}}{R_E^*}\right)\right]\left[\frac{(1 - s/z_1)}{1 + b_1 s + b_2 s^2}\right]$$

Where (see Chapter 4)

$$b_1 = R_1(C_{\pi 1} + C_{\mu 1}) + R_{L1}(C_{L1} + C_{\mu 1}) + g_{m1} R_1 R_{L1} C_{\mu 1}$$

$$b_2 = R_1 R_{L1}(C_{\pi 1} C_{L1} + C_{\pi 1} C_{\mu 1} + C_{L1} C_{\mu 1})$$

$$z_1 = +\frac{g_{m1}}{C_{\mu 1}}$$

If a dominant pole condition exists, with $|p_1| << |p_2|$, then

$$p_1 \cong -\frac{1}{b_1}$$

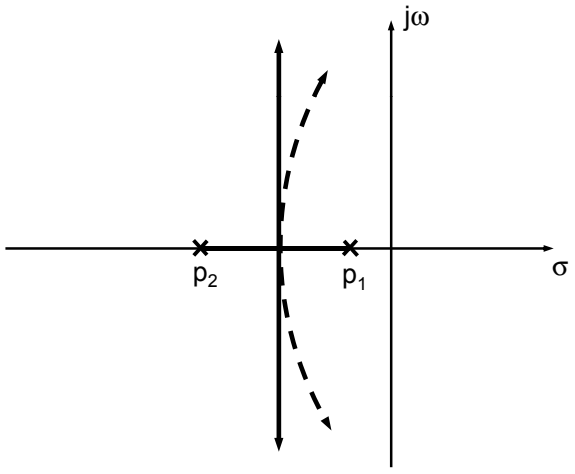
$$= -\frac{1}{R_1 C_{\pi 1} + R_{L1} C_{L1} + g_{m1} R_1 R_{L1} C_{\mu 1}}$$

and

$$p_2 = \frac{1}{b_2 p_1} \cong -\frac{R_1 C_{\pi 1} + R_{L1} C_{L1} + g_{m1} R_1 R_{L1} C_{\mu 1}}{R_1 C_{\pi 1} R_{L1} C_{L1}}$$

$$= -\left(\frac{1}{R_1 C_{\pi 1}} + \frac{1}{R_{L1} C_{L1}} + \frac{g_{m1} C_{\mu 1}}{C_{\pi 1} C_{L1}} \right)$$

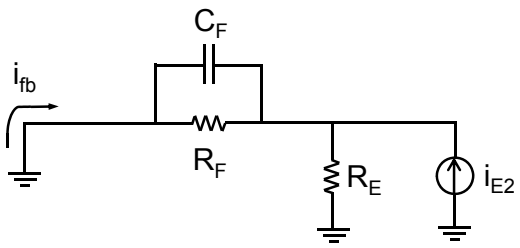
In the s-plane



Unfortunately, non-dominant poles (and RHP zeros) of T(s) tend to push the poles of A(s) toward the RHP

Feedback (Phantom) Zeros

The flat bandwidth and loop gain achievable in a shunt-series pair can be increased significantly by introducing a **feedback (phantom) zero**. This is accomplished by including a capacitor, C_F, in shunt with R_F; the feedback network then becomes:



Neglecting the loading of C_F at the input and the emitter of Q_2 in the forward path (since it appears in shunt with R_E), a(s) remains the same as w/o C_F. However,

$$f(s) = -\left(\frac{R_E}{R_E + R_F}\right)\left(\frac{1 - s/z_F}{1 - s/p_F}\right)$$

where

$$z_F = -\frac{1}{R_F C_F}$$

$$p_F = -\frac{1}{(R_F \,||\, R_E)C_F} = -\frac{R_F + R_E}{R_F R_E C_F}$$

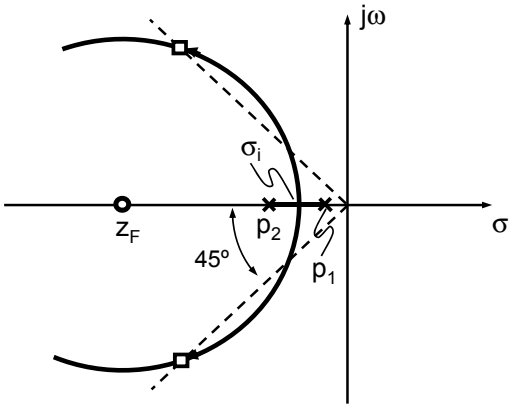
Since $A_0 \cong (R_F + R_E)/R_E$

$$|p_F| = A_0 |z_F|$$

Since C_F is chosen so that $|z_F|$ is on the order of the bandwidth of A(s), p_F can be ignored unless the low-frequency closed loop gain, A_0, is small. Thus,

$$f(s) \cong -\left(\frac{R_E}{R_E + R_F}\right)(1 - sR_F C_F)$$

C_F is chosen so as to obtain the desired closed-loop pole positions. For example, for an MFM response:



If z_F is reasonably remote from p_1 and p_2, then

$$\sigma_i = \frac{p_1 + p_2}{2}$$

The root locus is a circle of radius $|z_F - \sigma_i|$ centered at z_F. For each branch of the locus there are two intersections with the two lines at 45° to the negative real axis that correspond to an MFM response.

If $|z_F| >> |\sigma_i|$, then for an MFM response

$$B = \sqrt{2}\,|z_F|$$

and the closed-loop pole positions, s_1 and s_2, are

$$s_1, s_2 \cong |z_F|(-1 \pm j) = \frac{B}{\sqrt{2}}(-1 \pm j)$$

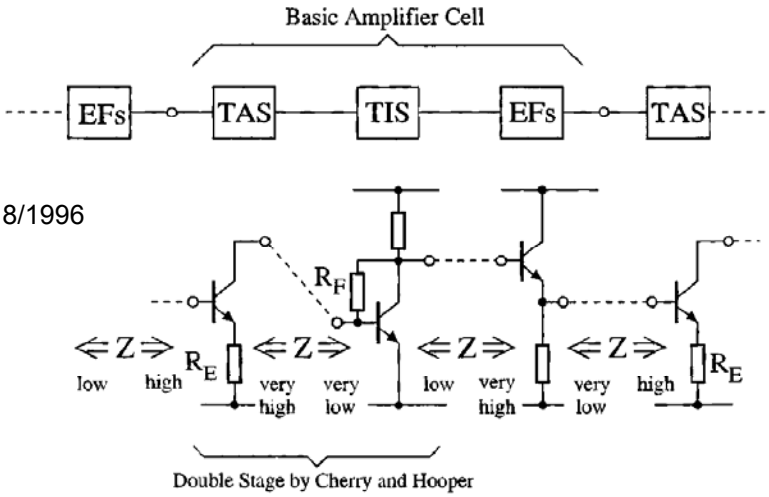
The loop gain, T_0, needed to place the closed-loop poles at these positions is

$$T_0 = \frac{|s_1 - p_1||s_2 - p_2|}{|p_1||p_2|} = \frac{B^2}{|p_1||p_2|}$$

$$= \frac{2|z_F|^2}{|p_1||p_2|} = \frac{2R_1C_{\pi 1}R_{L1}C_{L1}}{(R_FC_F)^2}$$

Broadband Amplifier Design with Local Feedback

Rein & Moller, JSSC 8/1996



In a cascade of amplifier stages, the interaction (loading) between adjacent stages can be reduced by alternating local series and shunt feedback circuits.

Advantages & Disadvantages of Local Feedback Cascades

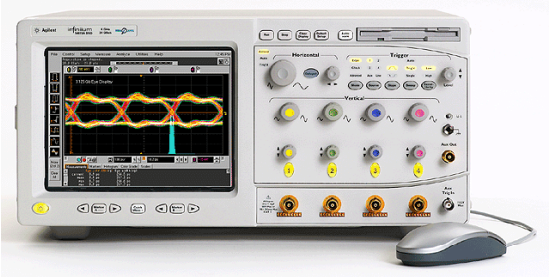
Advantages

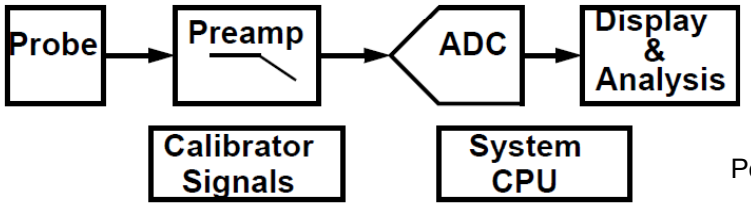
- No instability
- Lower gain sensitivity to component and device parameter variations than amplifiers without local feedback

Disadvantages

- Higher gain sensitivity (less loop gain) than multi-stage feedback amplifiers (i.e. multi-stage amplifiers with global feedback)
- Bandwidth obtainable is typically slightly less than that possible in multi-stage feedback amplifiers

Application Example (1)

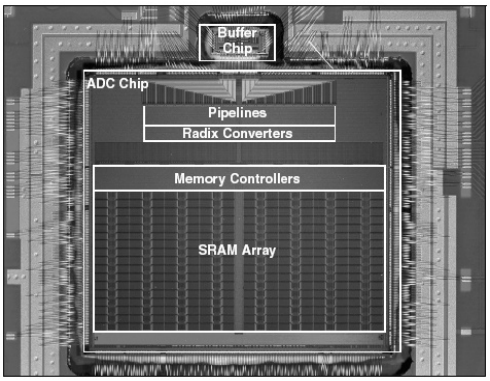




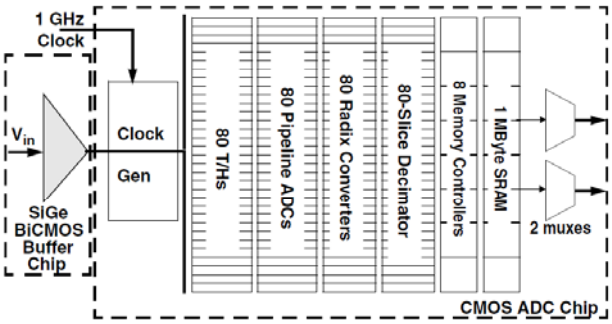
Poulton, ISSCC 2003

Simplified oscilloscope block diagram

Application Example (2)



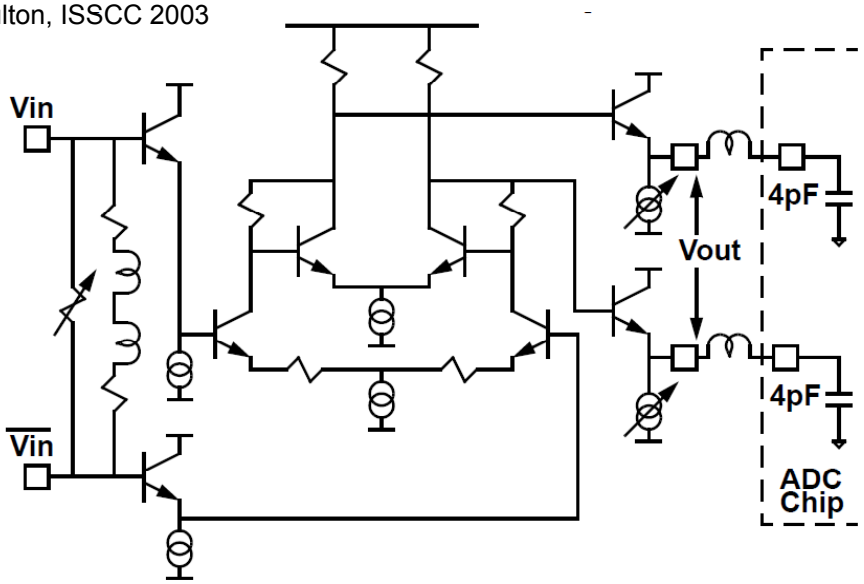
Poulton, ISSCC 2003



	Buffer Chip	ADC Chip
Input Capacitance	0.2 pF	4 pF
Power	1 W	9 W
Chip Size	1.2 x 2.6 mm	14 x 14 mm
Technology	40-GHz SiGe BiCMOS	0.18-μm CMOS
Transistors	1000	50M
Package	438-ball BGA	

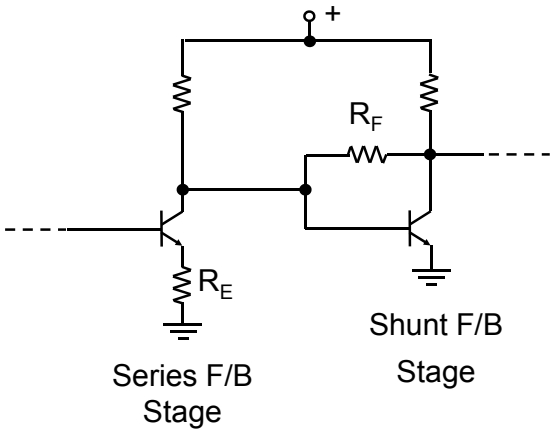
Application Example (3)

Poulton, ISSCC 2003



BW ~ 6 GHz

Cherry-Hooper Amplifier



E. Cherry and D. Hooper, Proc. IEE, Feb. 1963

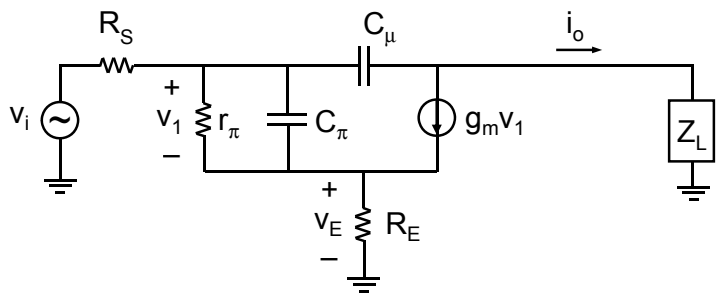
Analysis

$$A(0) = \frac{v_o}{v_{in}} \cong \left(-\frac{1}{R_E}\right)(-R_F) = \frac{R_F}{R_E}$$

- Thus, the low-frequency gain is approximately equal to the ratio of the two feedback resistors, and is therefore insensitive to component and device parameter values
- While the series feedback stage usually has a large bandwidth, there is considerable interaction between the input and output nodes of the shunt feedback stage. However this interaction can be controlled and can be used to position the poles to achieve a particular response, such as MFM.
- To see how the pole positions in a local feedback cascade can be controlled, we first consider the local feedback stages individually

Series Feedback Stage

Small-signal equivalent circuit:



Include transistor r_b in R_S

Neglect r_μ, r_o, r_e, and r_c

Assume that $Z_L \approx 0$; then there is no Millller effect and C_μ is just a (small) capacitance to ground which often has negligible impact.

Define

$$Y_\pi \triangleq \frac{1}{r_\pi} + sC_\pi$$

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

$$v_i - v_1\left[1 + (g_m + Y_\pi)R_E\right] = v_1 Y_\pi R_S$$

$$\therefore \quad \frac{v_1}{v_i} = \frac{1}{1 + (g_m + Y_\pi)R_E + Y_\pi R_S}$$

Since $i_o = -g_m v_1$

$$\frac{i_o}{v_i} = -\frac{g_m}{1+g_m R_E + (R_S+R_E)Y_\pi}$$

$$= -\left(\frac{g_m}{1+g_m R_E}\right)\left[\frac{1}{1+\left(\frac{R_S+R_E}{1+g_m R_E}\right)\left(\frac{1}{r_\pi}+sC_\pi\right)}\right]$$

$$= -\left(\frac{g_m}{1+g_m R_E}\right)\left[\frac{1}{1+\left(\frac{R_S+R_E}{r_\pi}\right)\left(\frac{1}{1+g_m R_E}\right)+sC_\pi\left(\frac{R_S+R_E}{1+g_m R_E}\right)}\right]$$

Usually,

$$\left(\frac{R_S+R_E}{r_\pi}\right)\left(\frac{1}{1+g_m R_E}\right) = \frac{g_m(R_S+R_E)}{\beta_0}\left(\frac{1}{1+g_m R_E}\right) \quad << \quad 1$$

Then

$$\frac{i_o}{v_i} \cong -g_{meq}\left(\frac{1}{1-s/p_1}\right)$$

where

$$g_{meq} = \frac{g_m}{1+g_m R_E}$$

$$p_1 = -\left(\frac{1+g_m R_E}{C_\pi}\right)\left(\frac{1}{R_E+R_S}\right)$$

If $g_m R_E >> 1$

$$g_{meq} \cong \frac{1}{R_E}$$

and

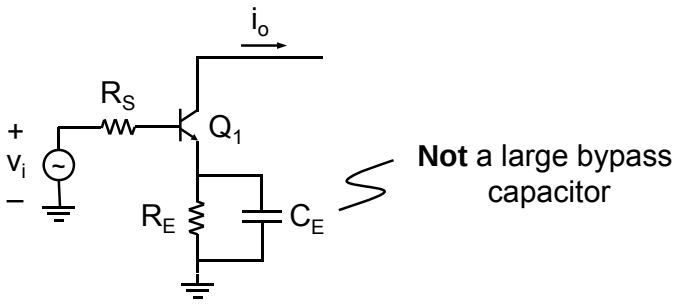
$$p_1 \cong -\frac{g_m}{C_\pi}\left(\frac{R_E}{R_E + R_S}\right) = -\omega_T\left(\frac{C_\pi + C_\mu}{C_\pi}\right)\left(\frac{R_E}{R_E + R_S}\right) \cong -\omega_T\left(\frac{R_E}{R_E + R_S}\right)$$

Note that this result corresponds to the dominant time constant found via ZVTC analysis in handout 7.

Thus, for $R_E >> R_S$, $p_1 \rightarrow -\omega_T$. In some cases, the bandwidth may then be actually limited by C_μ (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 $\tau_E \triangleq R_E C_E$ and assuming $\tau_T \cong \dfrac{C_\pi}{g_m}$

$$\frac{i_o}{v_i} = -\frac{g_m}{1+g_m Z_E + (R_S + Z_E)\left(\frac{1}{r_\pi} + sC_\pi\right)}$$

$$= -\frac{g_m}{1+\frac{R_S}{r_\pi} + sC_\pi R_S + \left(\frac{R_E}{1+s\tau_E}\right)\left(g_m + \frac{1}{r_\pi} + sC_\pi\right)}$$

$$\cong -\frac{g_m}{1+\frac{R_S}{r_\pi} + sC_\pi R_S + \left(\frac{g_m R_E}{1+s\tau_E}\right)\left(1 + s\frac{C_\pi}{g_m}\right)}$$

since $g_m >> \dfrac{1}{r_\pi}$

$$\therefore \quad \frac{i_o}{v_i} \cong -\frac{g_m}{1+\frac{R_S}{r_\pi} + sC_\pi R_S + g_m R_E\left(\frac{1+s\tau_T}{1+s\tau_E}\right)}$$

If C_E is chosen so that

$$\tau_E = \tau_T$$

and it is true that

$$g_m R_E >> \frac{R_S}{r_\pi} = \frac{g_m R_S}{\beta_0}$$

Then

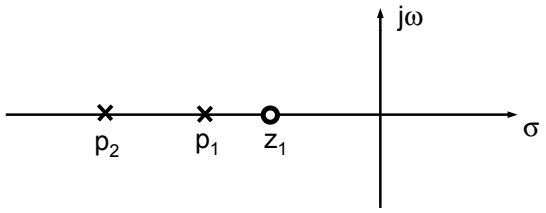
$$\frac{i_o}{v_i} \cong -\left(\frac{g_m}{1+g_m R_E}\right)\left[\frac{1}{1+sC_\pi\left(\frac{R_S}{1+g_m R_E}\right)}\right]$$

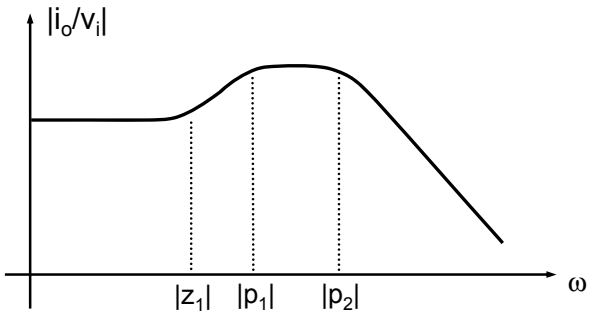
The result is a single-pole response with

$$p_1 = -\frac{1+g_m R_E}{R_S C_\pi} \cong -\omega_T \left(\frac{R_E}{R_S}\right)$$

In this case, if $R_E > R_S$, then $|p_1| > \omega_T$ (!)

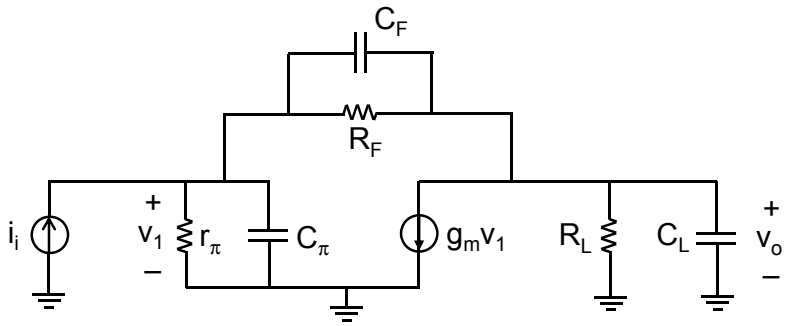
Note that $\tau_E > \tau_T$ results in a pole-zero separation that can cause peaking in the frequency response





Shunt Feedback Stage

Small-signal equivalent circuit



Include transistor C_μ in C_F

Neglect R_S or include it in r_π

Neglect r_b, r_c, r_e, and r_μ

Include r_o and C_{cs} in R_L and C_L

Define

$$\tau_F \triangleq R_F C_F$$
$$\tau_L \triangleq R_L C_L$$

and

$$Y_F = \frac{1}{R_F} + sC_F = \frac{1}{R_F}(1 + sC_F R_F) = \frac{1}{R_F}(1 + s\tau_F)$$
$$Y_L = \frac{1}{R_L} + sC_L = \frac{1}{R_L}(1 + s\tau_L)$$
$$Y_\pi = \frac{1}{r_\pi} + sC_\pi$$

From the equivalent circuit:

$$i_i + (v_o - v_1)Y_F = v_1 Y_\pi \qquad (1)$$

$$v_o Y_L + g_m v_1 + (v_o - v_1)Y_F = 0 \qquad (2)$$

Then from (2)

$$v_1 = -v_o\left(\frac{Y_L + Y_F}{g_m - Y_F}\right)$$

Substituting in (1)

$$i_i + v_o\left[Y_F + \left(\frac{Y_L + Y_F}{g_m - Y_F}\right)(Y_\pi + Y_F)\right] = 0$$

$$\therefore \quad \frac{v_o}{i_i} = -\frac{g_m - Y_F}{(g_m - Y_F)Y_F + (Y_L + Y_F)(Y_\pi + Y_F)}$$

Assuming that $g_m >> 1/R_F = G_F$

$$g_m - Y_F = g_m - (G_F + sC_F) \cong g_m - sC_F$$

Then

$$\frac{v_o}{i_i} \cong -\frac{g_m - sC_F}{(g_m - sC_F)(G_F + sC_F) + \left[(G_L + G_F) + s(C_L + C_F)\right]\left[(g_\pi + G_F) + s(C_\pi + C_F)\right]}$$

Thus

$$\frac{v_o}{i_i} \cong -\frac{g_m - sC_F}{a_0 + a_1 s + a_2 s^2}$$

where

$$a_0 = g_m G_F + (G_L + G_F)(g_\pi + G_F)$$
$$a_1 = (g_m - G_F)C_F + (G_L + G_F)(C_\pi + C_F) + (g_\pi + G_F)(C_L + C_F)$$
$$a_2 = C_L C_\pi + C_L C_F + C_\pi C_F$$

Further assuming that $g_m >> G_L$, $R_F >> r_\pi$, $C_\pi >> C_F$, and noting that $g_m >> g_\pi$

$$a_0 \cong g_m / R_F$$
$$a_1 \cong g_m C_F + (G_L + G_F) C_\pi + g_\pi (C_L + C_F)$$
$$a_2 \cong C_\pi (C_L + C_F)$$

Then

$$\frac{v_o}{i_i} \cong -\frac{R_F(1 - s/z_1)}{1 + b_1 s + b_2 s^2}$$

where

$$b_1 = a_1/a_0 \cong R_F C_F + \left(\frac{R_F + R_L}{R_L} \right) \tau_T + \frac{R_F}{\beta_0} (C_L + C_F)$$
$$b_2 = a_2/a_1 \cong R_F (C_L + C_F) \tau_T$$
$$z_1 = +g_m / C_F$$

Often $b_1 \cong R_F C_F$

z_1 is a feedforward zero that is located in the RHP.

Since it was assumed that $C_\pi >> C_F$

$$|z_1| >> \frac{g_m}{C_\pi} \cong \omega_T$$

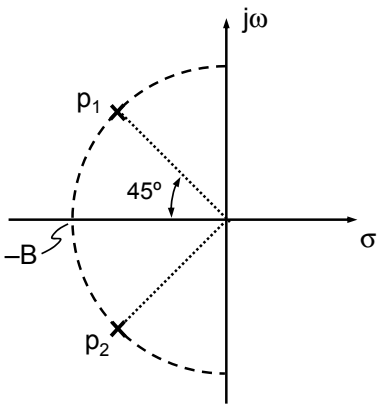
In these circumstances the feedforward zero can usually be neglected, and the shunt feedback stage has an approximate 2-pole response.

$$\frac{v_o}{i_i} = -\frac{R_F}{(1 - s/p_1)(1 - s/p_2)}$$

The poles p_1 and p_2 may be complex.

C_F can be used to alter the positions of the poles.

For example, C_F can be chosen to provide a 2-pole MFM response:



$$b_2 = \frac{1}{p_1 p_2} = \frac{1}{\left|p_1^2\right|} = \frac{1}{B^2}$$

Thus,

$$\omega_{-3dB} = \frac{1}{\sqrt{b_2}} = \frac{1}{\sqrt{R_F(C_F + C_L)\tau_T}}$$

If $C_F << C_L$

$$\omega_{-3dB} \cong \frac{1}{\sqrt{R_F C_L \tau_T}}$$

Further insight can be gained from this result by considering the input capacitance of the overall Cherry-Hooper amplifier, which is approximately $C_\pi/(g_m R_E)$. If we assume that $C_L \cong k \cdot C_\pi/(g_m R_E) = k \cdot \tau_T / R_E$

$$\omega_{-3dB} \cong \left(\sqrt{\frac{1}{k}\frac{R_E}{R_F}}\right)\frac{1}{\tau_T} = \left(\sqrt{\frac{1}{k}\frac{R_E}{R_F}}\right)\omega_T = \frac{\omega_T}{\sqrt{k \cdot A(0)}}$$

Since the overall low-frequency gain, A_{02}, of the cascade is

$$A_{02} = \frac{R_F}{R_E}$$

the per-stage gain-bandwidth product, GB_2, is

$$GB_2 = (A_{02})^{1/2} \cdot \omega_{-3dB}$$
$$= \left(\sqrt{\frac{R_F}{R_E}}\right) \cdot \left(\sqrt{\frac{1}{k}\frac{R_E}{R_F}}\right) \omega_T = \frac{\omega_T}{\sqrt{k}}$$

Thus, the amplifier approaches the ideal gain-bandwidth product limit for a cascade of identical stages (k=1); there is no bandwidth shrinkage.

The value of C_F that provides a 2-pole MFM response can be found as follows. Realize that

$$b_1 = -\left(\frac{1}{p_1} + \frac{1}{p_2}\right) = -\frac{p_1 + p_2}{p_1 p_2}$$

p_1 and p_2 are complex conjugates at 45° for an MFM response

$$p_1 = \sigma_1 + j\omega_1$$
$$p_2 = \sigma_1 - j\omega_1$$

where

$$-\sigma_1 = \omega_1 = \frac{B}{\sqrt{2}}$$

Thus

$$b_1 \cong R_F C_F = \frac{2\left(B/\sqrt{2}\right)}{B^2} = \frac{\sqrt{2}}{B}$$

From the earlier analysis, for $C_F << C_L$

$$B = \frac{1}{\sqrt{b_2}} \cong \frac{1}{\sqrt{R_F C_L \tau_T}}$$

Therefore

$$C_F = \frac{1}{R_F}\left(\frac{\sqrt{2}}{B}\right) = \frac{\sqrt{2R_F C_L \tau_T}}{R_F} = \sqrt{\frac{2C_L}{R_F} \cdot \tau_T}$$

Chapter 8

Noise Analysis

B. Murmann

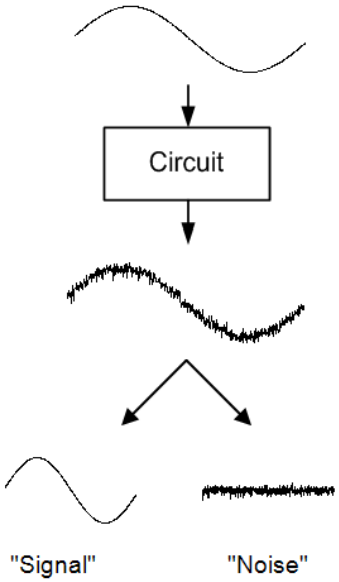
Stanford University

Reading Material: Sections 11.1, 11.2, 11.3, 11.4, 11.5, 11.6, 11.7, 11.9

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)



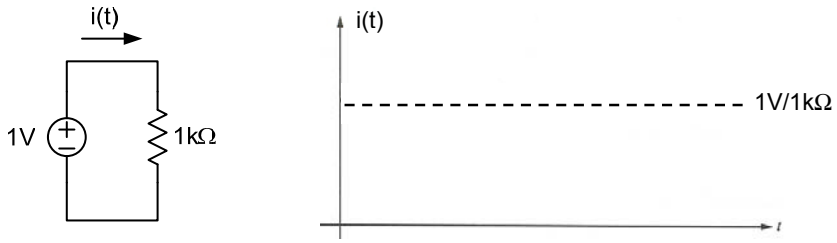
Signal-to-Noise Ratio

$$SNR = \frac{P_{signal}}{P_{noise}} \propto \frac{V_{signal}^2}{V_{noise}^2}$$

Significance of Electronic Noise (2)

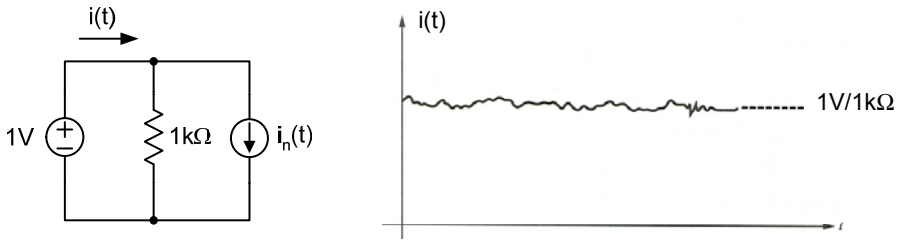
- The "fidelity" of electronic systems is often determined by their SNR
  - Examples
    - Audio systems
    - Imagers, cameras
    - Wireless and wireline transceivers
- Electronic noise directly trades with power dissipation and speed
- Noise has become increasingly important in modern technologies with reduced supply voltages
  - $SNR \sim V_{signal}^2/V_{noise}^2 \sim (\alpha V_{DD})^2/V_{noise}^2$
- 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 $i_n(t)$

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 constants of $\tau \cong 0.17$ps
  - Consequences:
    - Gaussian amplitude distribution
    - Knowing $i_n(t)$ does not help predict $i_n(t+\Delta t)$, unless Δt 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 ($1/\tau \cong 6{,}000$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

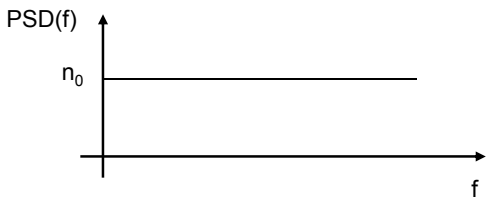
$$P_n = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} i_n^2(t)\cdot R\cdot dt$$

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

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

Thermal Noise Spectrum

- The so-called power spectral density (PSD) shows how much power a signal caries at a particular 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

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

Thermal Noise Power

- Nyquist showed that the noise PSD of a resistor is

$$PSD(f) = n_0 = 4 \cdot kT$$

- 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

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

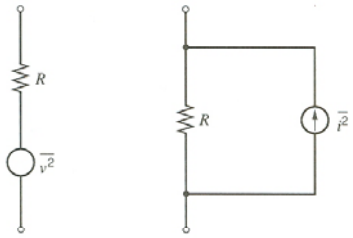
$$\overline{i_n^2} = \frac{P_n}{R} = 4kT \cdot \frac{1}{R} \cdot \Delta f$$

For $R = 1k\Omega$:

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

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

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



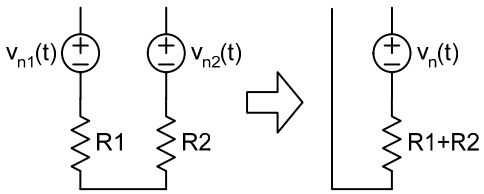
For $R = 1k\Omega$:

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

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

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

Two Resistors in Series



$$\overline{v_n^2} = \overline{\left(v_{n1} - v_{n2}\right)^2} = \overline{v_{n1}^2} + \overline{v_{n2}^2} - \overline{2 \cdot v_{n1} \cdot v_{n2}}$$

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

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

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

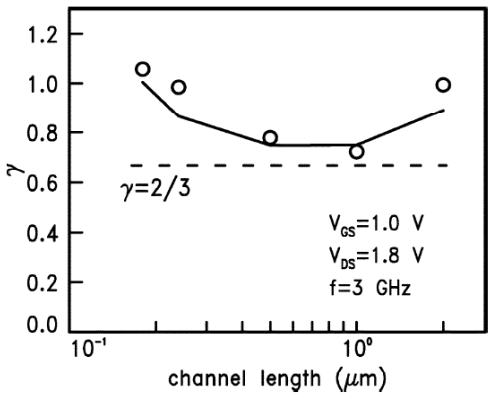
MOSFET Thermal Noise (1)

- 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 γ=2/3
- For the past 10-15 years, researchers have been debating the value of γ in short channels
- Preliminary (wrong) results had suggested that in short channels γ can be as high as 5 due to “hot carrier” effects

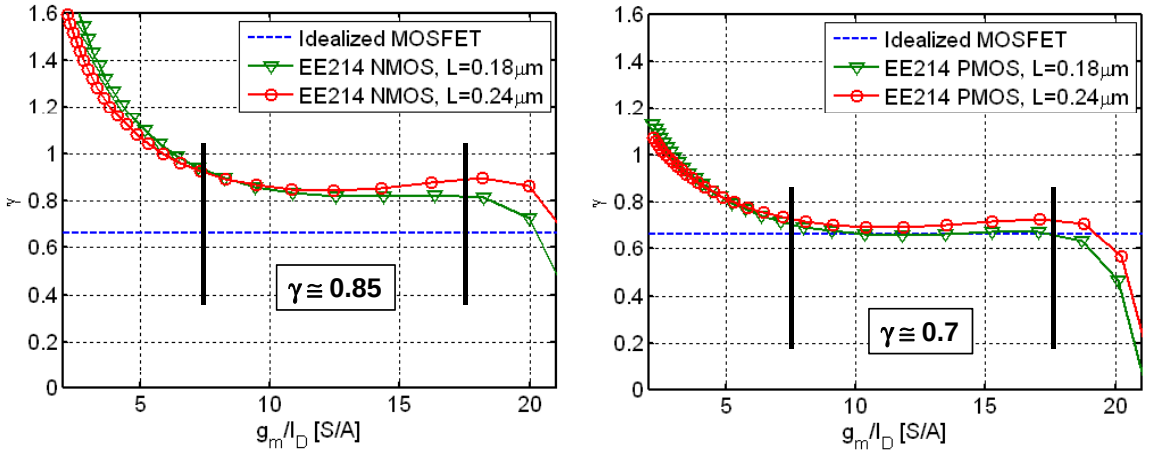
MOSFET Thermal Noise (2)



[Scholten]

- At moderate gate bias in strong inversion, short-channel MOSFETS have $\gamma \cong 1$
  - 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.

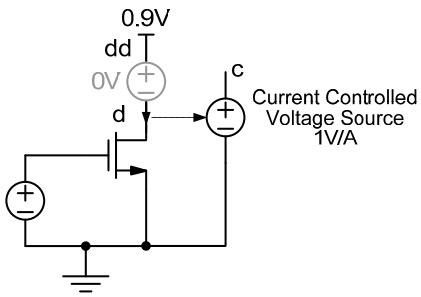
Thermal Noise in EE214 MOSFET Devices



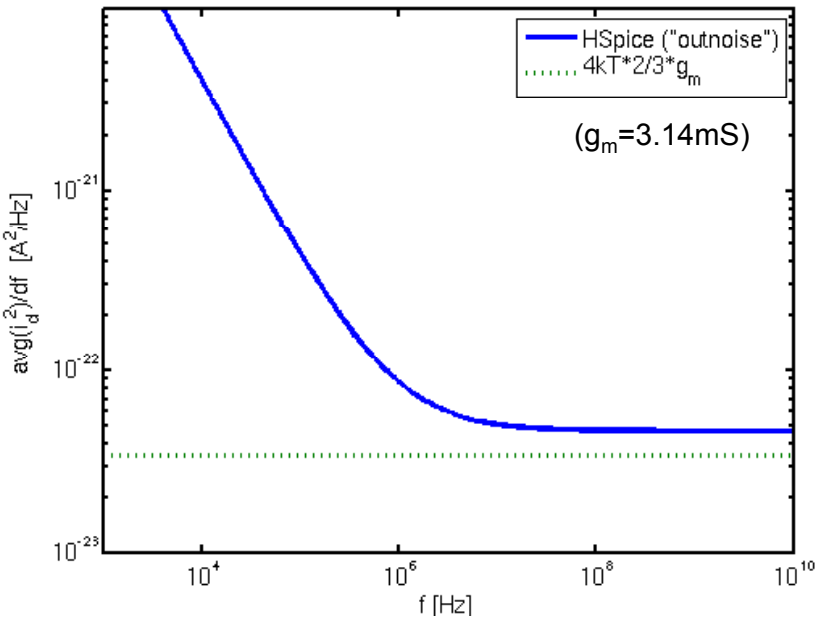
- Parameter γ depends on biasing conditions, but is roughly constant within a reasonable range of g_m/I_D used for analog design
- The EE214 HSpice models become inaccurate for sub-threshold operation

Spice Simulation (1)

```
* EE214 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   ccvs   vm   1

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

.options post brief
.inc './ee214_hspice.sp'
.end
```



Spice Simulation (2)



1/f Noise

- Also called "flicker noise" or "pink noise"
  - Caused by traps near Si/SiO_2 interface that randomly capture and release carriers
  - 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 EE214 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_f is strongly dependent on technology; numbers for EE214:
  - $K_{f,NMOS} = 0.5 \cdot 10^{-25}$ V^2F
  - $K_{f,PMOS} = 0.25 \cdot 10^{-25}$ V^2F

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_m/I_D (e.g. based on linearity considerations), the only way to achieve lower f_{co} is to use longer channel devices
  - In the above expression, both 1/L and I_D/W are reduced for increasing L
- Example
  - EE214 NMOS, L = 0.18µm, g_m/I_D = 12 S/A , $\Rightarrow I_D/W$ = 20 A/m
  $\Rightarrow f_{co}$ = 560 kHz
- In newer technologies, f_{co} can be on the order of 10 MHz

1/f Noise Contribution (1)

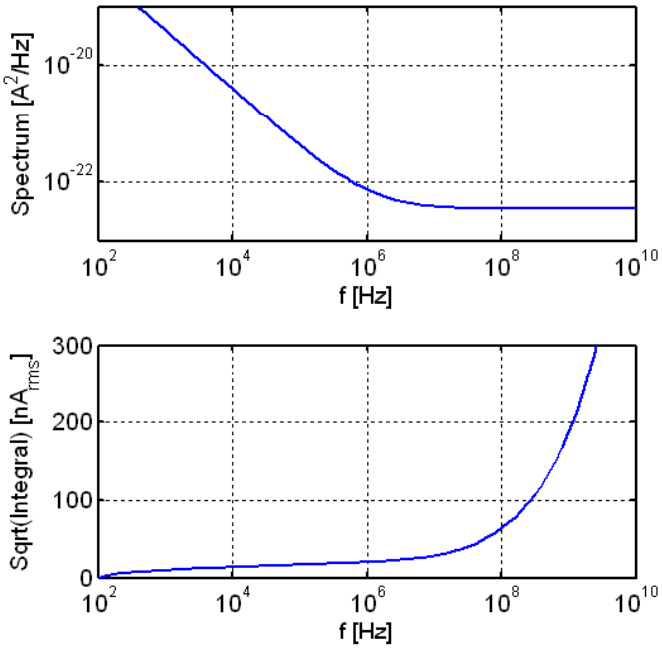
- Just as with white noise, the total 1/f noise contribution is found by integrating its power spectral density

$$\overline{i_{1/f,tot}^2} = \int_{f_1}^{f_2}\frac{K_f}{C_{ox}}\frac{g_m^2}{W\cdot L}\frac{\Delta f}{f}$$

$$= \frac{K_f}{C_{ox}}\frac{g_m^2}{W\cdot L}\ln\left(\frac{f_2}{f_1}\right) = \frac{K_f}{C_{ox}}\frac{g_m^2}{W\cdot L}2.3\log\left(\frac{f_2}{f_1}\right)$$

- 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)

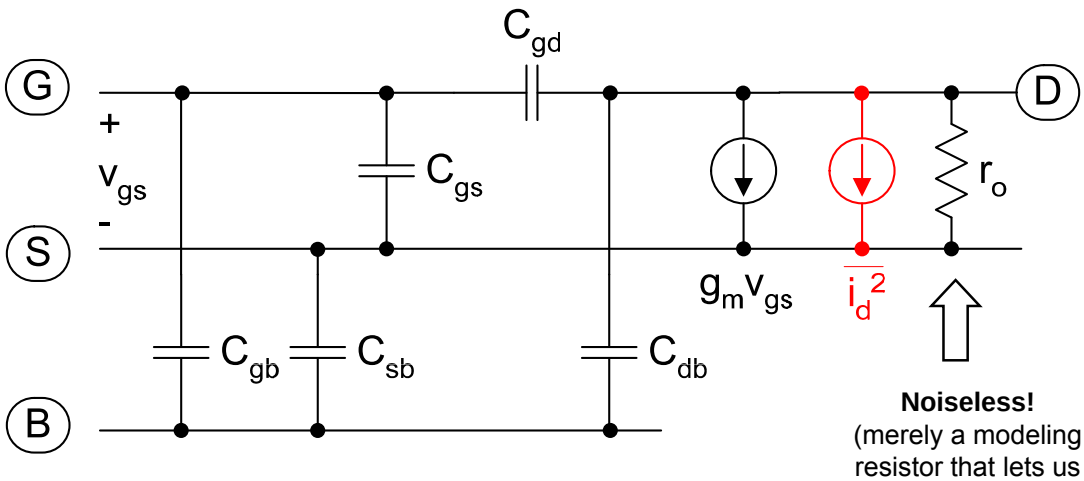


- In the example shown on the left, the noise spectrum is integrated from 100Hz to 10GHz
- The contribution of the flicker noise is relatively small, even though its PSD dominates at low frequencies
- For circuits with very large bandwidth, flicker noise is often insignificant

Lower Integration Limit

- Does the flicker noise PSD go to infinity for f → 0?
  - See e.g. E. Milotti, "1/f noise: a pedagogical review," available at http://arxiv.org/abs/physics/0204033
- Even if the PSD goes to infinity, do we care?
  - Say we are sensing a signal for a very long time (down to a very low frequency), e.g. 1 year ≅ 32 Msec, 1/year ≅ 0.03 μHz
  - Number of frequency decades in 1/year to 100Hz ≅ 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 → Only 50% more flicker noise!

MOS Model with Noise Generator



Noiseless!
(merely a modeling resistor that lets us account for finite dI_D/dV_{DS})

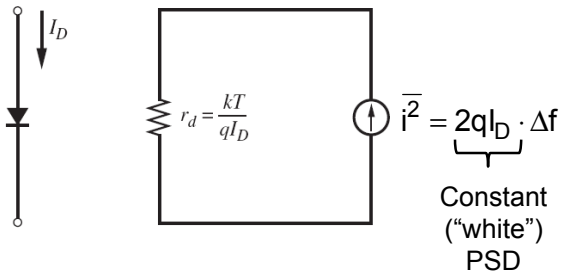
$$\frac{\overline{i_d^2}}{\Delta f} = 4kT \cdot \gamma \cdot g_m \cdot + \frac{K_f}{C_{ox}} \frac{g_m^2}{W \cdot L} \frac{1}{f}$$

Other MOSFET Noise Sources

- Gate noise
  - "Shot noise" from gate leakage current
  - Noise due to finite resistance of the gate material
  - Noise due to randomly changing potential/capacitance between the channel and bulk
    - Relevant only at very high frequencies
    - See EE314
- 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

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

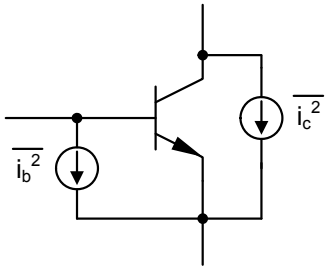


Shot Noise in a Bipolar Transistor

- 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

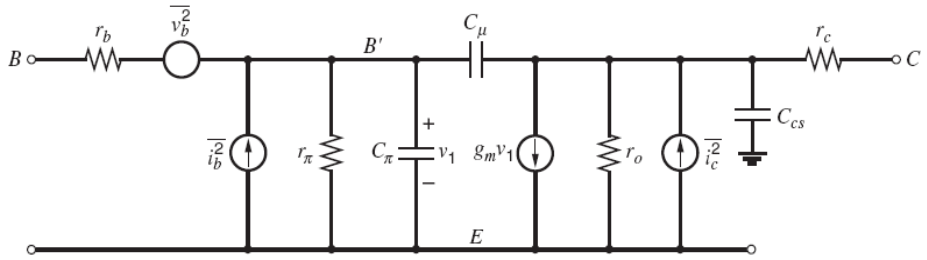
$$\overline{i_c^2} = 2qI_C\Delta f = 2kTg_m\Delta f$$

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



- 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: $\overline{v_b^2} = 4kTr_b\,\Delta f$

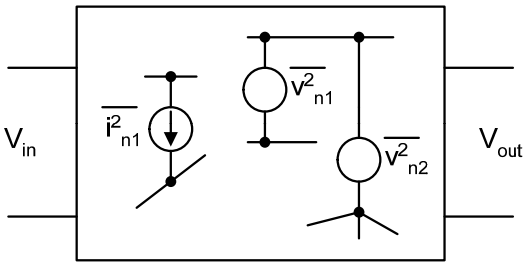
Collector shot noise: $\overline{i_c^2} = 2qI_C\,\Delta f$

Typically negligible

Base noise components:

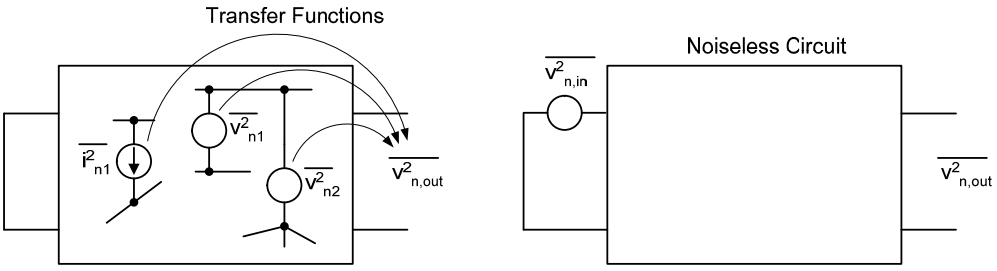
$$\overline{i_b^2} = \underbrace{2qI_B\,\Delta f}_{\text{Shot noise}} + \underbrace{K_1\frac{I_B^a}{f^c}\Delta f}_{\text{Flicker noise}} + \underbrace{K_2\frac{I_B^c}{1+\left(\frac{f}{f_c}\right)^2}\Delta f}_{\text{Burst noise}}$$

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

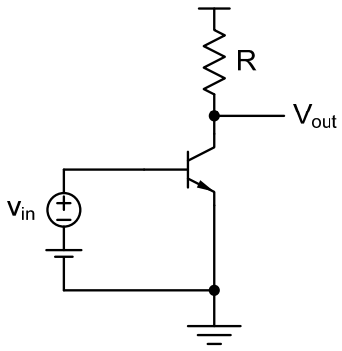
Noise in Circuits (2)



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

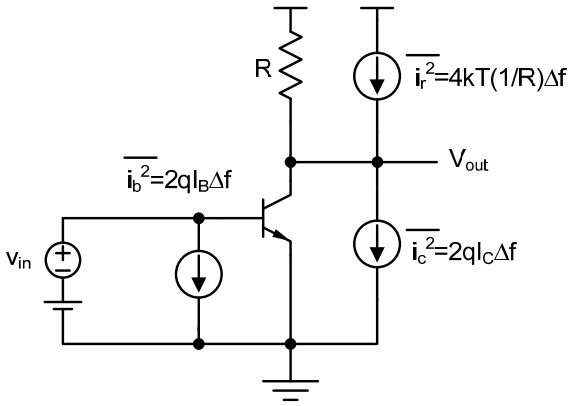
Input referred noise
- Represent total noise via a fictitious input source that captures all circuit-internal noise sources
- Useful for direct comparisons with input signal, or "general purpose" components in which the output noise depends on how the component is used

Circuit Example



- For simplicity, let's neglect
  - Source impedance
  - All capacitances
  - Burst and flicker noise
  - r_o, r_b, and r_μ

Output Referred Noise PSD



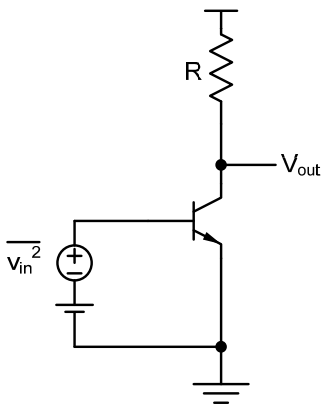
$$\overline{v_{out}^2} = \left(4kT\frac{1}{R}\Delta f + 2qI_C\Delta f\right)\cdot R^2$$

$$= 2kT\Delta f\left(2R + g_m R^2\right)$$

$$= 2kTg_m\Delta f\cdot R^2\left(\frac{2}{g_m R}+1\right)$$

- Shot noise due to base current is absorbed by the input source and does not contribute to noise at the output
- For large gain ($g_m R$), the collector shot noise dominates

Input Referred Noise PSD



- From the previous calculation, we know that

$$\overline{v_{out}^2} = 2kTg_m\Delta f\cdot R^2\left(\frac{2}{g_m R}+1\right)$$

- Since

$$v_{out} = A_v v_{in}$$

$$\overline{v_{out}^2} = A_v^2\overline{v_{in}^2} \quad \text{where} \quad A_v = g_m R$$

- We can write

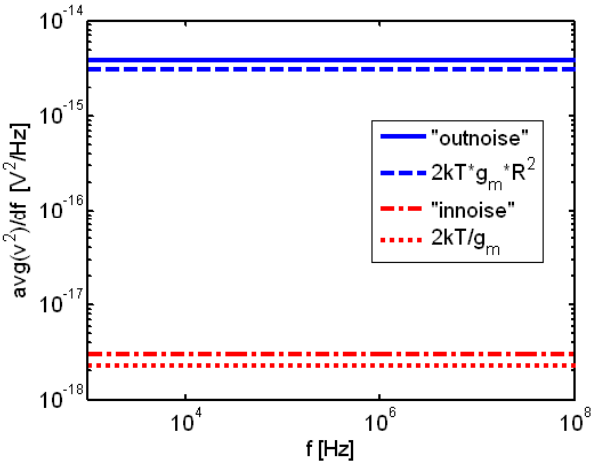
$$\overline{v_{in}^2} = \frac{2kTg_m\Delta f\cdot R^2\left(\frac{2}{g_m R}+1\right)}{(g_m R)^2} = 2kT\frac{1}{g_m}\Delta f\left(\frac{2}{g_m R}+1\right)$$

- Larger g_m translates into lower input referred voltage noise

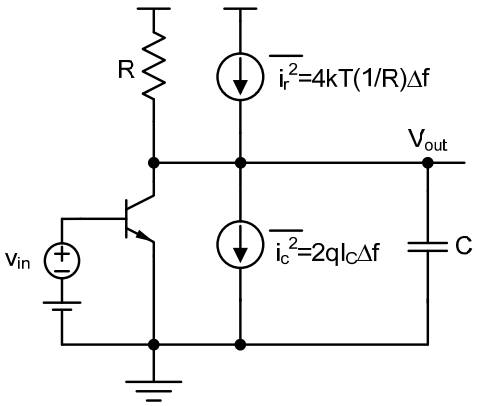
Spice Simulation

```
*** EE214 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=3.67mS, R=10k, A_v=36.7



Output Referred Noise PSD with Load Capacitance



$$\overline{v_{out}^2} = \left(4kT\frac{1}{R}\Delta f + 2qI_C\Delta f\right)\cdot\left|R \,||\, \frac{1}{j\omega C}\right|^2$$

$$= \left(4kT\frac{1}{R}\Delta f + 2qI_C\Delta f\right)\cdot R^2\left|\frac{1}{1+j\omega RC}\right|^2$$

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

- Same calculation as before, except that now the noise current drops into 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

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

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

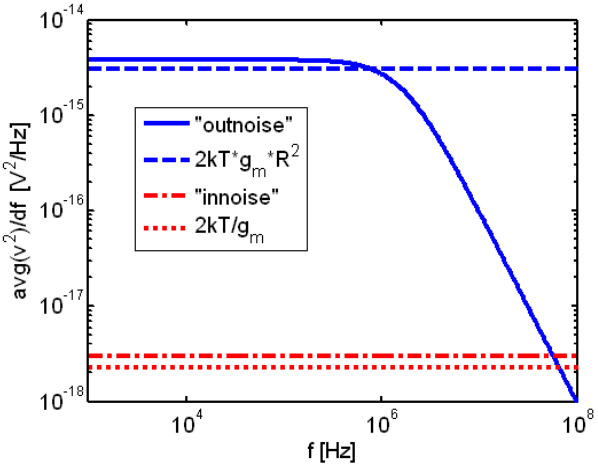
- Input referred noise is frequency independent, because the output noise and gain have the same frequency roll-off

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

Spice Simulation

```
*** EE214 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=3.67mS, R=10k, A_v=36.7, C=10pF



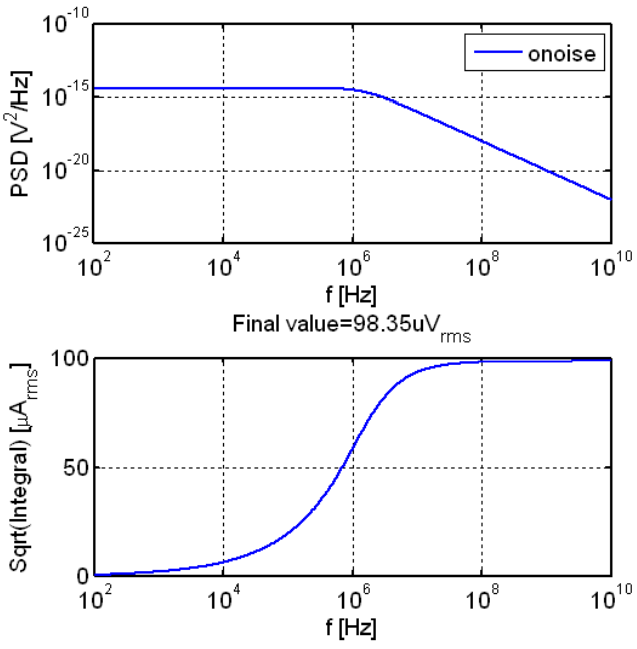
Signal-to-Noise Ratio

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

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{\frac{1}{2}\hat{v}_{out}^2}{\int_{f_1}^{f_2} \frac{\overline{v_{out}^2}}{\Delta f} \cdot df}$$

- 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

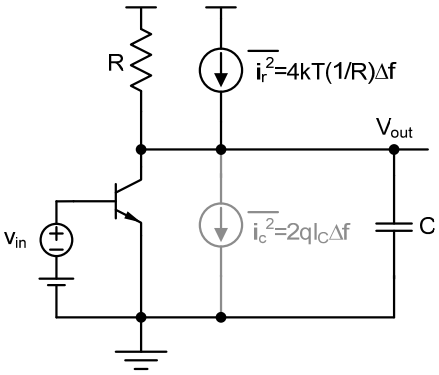
A Closer Look at The Circuit's Noise Integral



- 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 EE315A,B

Total Integrated Noise Calculation

- Let us first consider the noise from the resistor



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

$$= 4kTR \int_0^\infty \frac{df}{1+(2\pi fRC)^2}; \qquad \int \frac{du}{1+u^2} = \tan^{-1} u$$

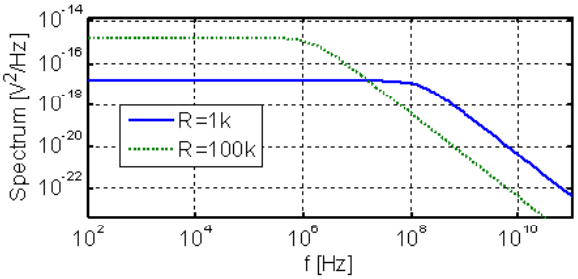
$$= 4kTR \cdot \frac{1}{4RC}$$

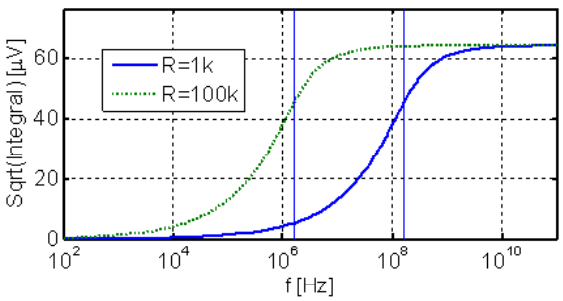
$$= \frac{kT}{C}$$

- 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 64μV_{rms}





Alternative Derivation

- The equipartition theorem of statistical mechanics says that each degree of freedom (or energy state) of a system in thermal equilibrium holds an average energy of kT/2
  - See e.g. EE248 for a derivation
- In our circuit, the quadratic degree of freedom is the energy stored on the capacitor

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

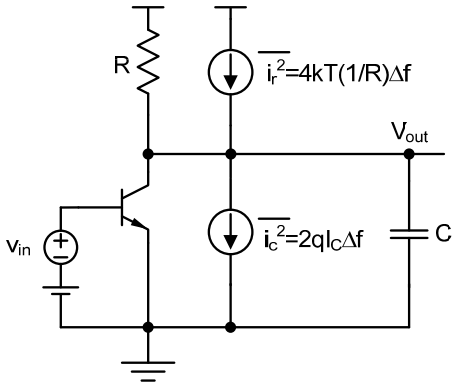
$$\overline{v_{out}^{\ 2}} = \frac{kT}{C}$$

Equivalent Noise Bandwidth

$$\left.\begin{aligned} \overline{v_{out,tot}^2} &= \frac{kT}{C} \\ \overline{v_{out,tot}^2} &\triangleq 4kTR \cdot \Delta f_{ENBW} \\ f_{-3dB} &= \frac{1}{2\pi RC} \end{aligned}\right\} \quad \Delta f_{ENBW} = \frac{\pi}{2} f_{-3dB}$$

- The equivalent noise bandwidth is generally defined as the bandwidth of a brick-wall filter which results in the same total noise power as the filter in question
- For a simple RC filter, the equivalent noise bandwidth is approximately 1.57 times its 3-dB corner frequency

Total Integrated Noise Calculation for the Complete Circuit



$$\overline{v_{out,tot}^2} = \int_0^{\infty} \underbrace{\left(4kTR + 2kTg_mR^2\right)}_{\text{Was } 4kTR \text{ in previous analysis}} \cdot \left|\frac{1}{1+j2\pi f \cdot RC}\right|^2 df$$

$$\overline{v_{out,tot}^2} = \frac{kT}{C} \cdot \frac{4kTR + 2kTg_mR^2}{4kTR}$$

$$= \frac{kT}{C}\left(1 + \frac{1}{2}g_mR\right)$$

- 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

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

$$SNR[dB] = 10\log\left(8.59 \cdot 10^6\right) = 69.3dB$$

- Typical system requirements
  - Audio: SNR ≅ 100dB
  - Video: SNR ≅ 60dB
  - Gigabit Ethernet Transceiver: SNR ≅ 35dB

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

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

- 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

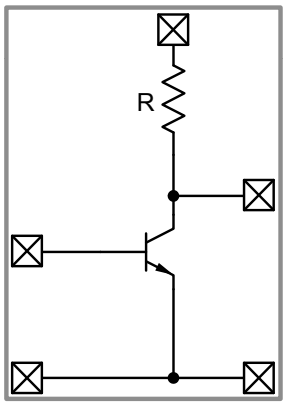
Does Thermal Noise Always Matter?

- 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	Hard to make such small capacitors…
40	0.000083	
60	0.0083	Designer will be concerned about thermal noise; component sizes often set by SNR
80	0.83	
100	83	A difficult battle with thermal noise …
120	8300	
140	830000	

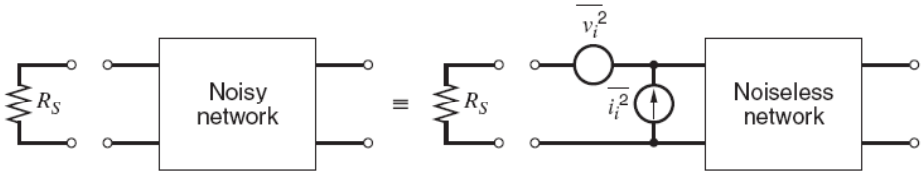
- Rules of thumb
  - Up to SNR ~ 30-40dB, integrated circuits are usually not limited by thermal noise
  - Achieving SNR >100dB is extremely difficult
    - Must usually rely on external components, or reduce bandwidth and remove noise by a succeeding filter
    - See e.g. oversampling ADCs in EE315B

More on Input Referred Noise



- Suppose you wanted to sell this amplifier as a "general purpose" building block
- How would you communicate information about its noise performance to the customer?
- None of the metrics that we have used so far will work to describe the circuit independent of the target application
  - The computed input and output referred voltage noise assumed that the circuit is driven by an ideal voltage source

Two-Port Representation 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
- Sometimes need to 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

Datasheet Example

ANALOG DEVICES

Ultraprecision Operational Amplifier

OP177

FEATURES

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

PIN CONFIGURATION

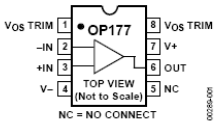


Figure 1. 8-Lead PDIP (P-Suffix), 8-Lead SOIC (S-Suffix)

Parameter	Symbol	Conditions	OP177F Min	OP177F Typ	OP177F Max	OP177G Min	OP177G Typ	OP177G Max	Unit
INPUT OFFSET VOLTAGE	V_{OS}			10	25		20	60	µV
LONG-TERM INPUT OFFSET[1] Voltage Stability	ΔV_{OS}/time			0.3			0.4		µV/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_O = 1 Hz to 100 Hz[2]		118	150		118	150	nV rms
INPUT NOISE CURRENT	i_n	f_O = 1 Hz to 100 Hz[2]		3	8		3	8	pA rms

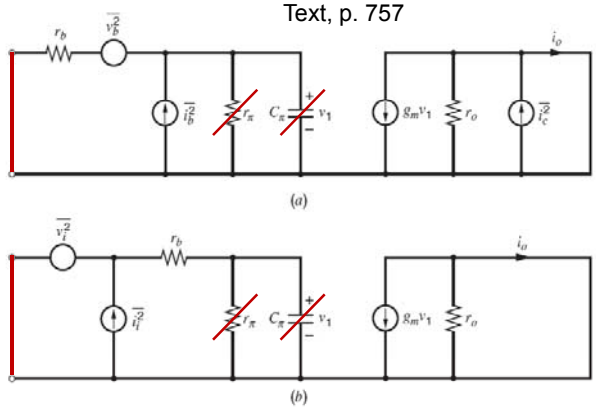
Examples

- Single BJT device
- Single MOS device
- CE stage
- CB stage

BJT Input Voltage Noise (1)

- To find input referred voltage generator, short the input of both circuit models and equate output noise
- Neglecting C_μ, r_c and r_e for simplicity

Text, p. 757



$$\overline{i_{o1}^2} \cong \overline{i_c^2} + g_m^2\left(\overline{i_b^2}\cdot r_b^2 + \overline{v_b^2}\right)$$

$$\overline{i_{o2}^2} \cong g_m^2\,\overline{v_i^2}$$

$$\overline{i_{o1}^2} = \overline{i_{o2}^2} \quad \Rightarrow \quad \overline{v_i^2} \cong \overline{v_b^2} + \frac{\overline{i_c^2}}{g_m^2} + \overline{i_b^2}\cdot r_b^2$$

BJT Input Voltage Noise (2)

- The PSD of the BJT input voltage noise generator is therefore

$$\frac{\overline{v_i^2}}{\Delta f} \cong 4kTr_b + \frac{2qI_C}{g_m^2} + 2qI_B r_b^2$$

$$\cong 4kTr_b + \frac{2qI_C}{g_m^2}\left[1 + \frac{g_m^2 r_b^2}{\beta}\right]$$

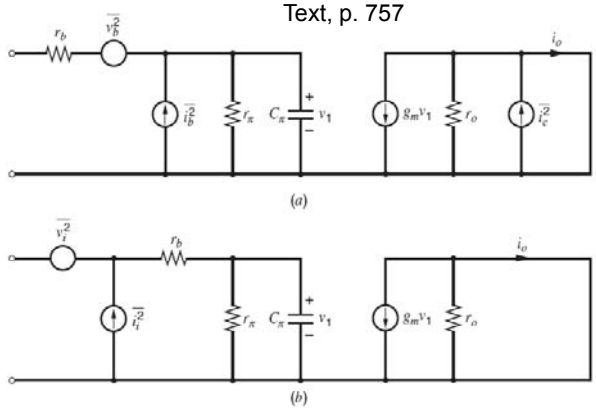
$$\cong 4kTr_b + \frac{2qI_C}{g_m^2}$$

$$\cong 4kT\left(r_b + \frac{1}{2g_m}\right)$$

BJT Input Current Noise (1)

- To find input referred current generator, open circuit the input of both circuit models and equate output noise

Text, p. 757



$$\overline{i_{o1}^2} = \overline{i_c^2} + g_m^2 \cdot \overline{i_b^2} \cdot |z_\pi|^2$$

$$\overline{i_{o2}^2} = g_m^2 \cdot \overline{i_i^2} \cdot |z_\pi|^2$$

$$\overline{i_{o1}^2} = \overline{i_{o2}^2} \quad \Rightarrow \quad \overline{i_i^2} = \overline{i_b^2} + \frac{\overline{i_c^2}}{g_m^2 |z_\pi|^2}$$

BJT Input Current Noise (2)

$$\frac{\overline{i_i^2}}{\Delta f} = 2qI_B + \frac{2qI_C}{g_m^2 |z_\pi|^2} \qquad \text{where } z_\pi = \frac{r_\pi \cdot \frac{1}{j\omega C_\pi}}{r_\pi + \frac{1}{j\omega C_\pi}} = \frac{r_\pi}{1 + j\omega r_\pi C_\pi} = \frac{\beta(j\omega)}{g_m}$$

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

$$= 2qI_B \left(1 + \frac{1 + \left(\frac{\omega}{\omega_\beta}\right)^2}{\beta_0} \right) \qquad \text{where } \omega_\beta = \frac{1}{r_\pi C_\pi} \cong -\frac{\omega_T}{\beta_0}$$

- The term due to I_C is negligible at low frequencies, but becomes comparable to the base current contribution at

$$\omega_b = \omega_\beta \sqrt{\beta_0} \cong \frac{\omega_T}{\sqrt{\beta_0}}$$

Plot of BJT Input Noise Current PSD

Text, p. 760

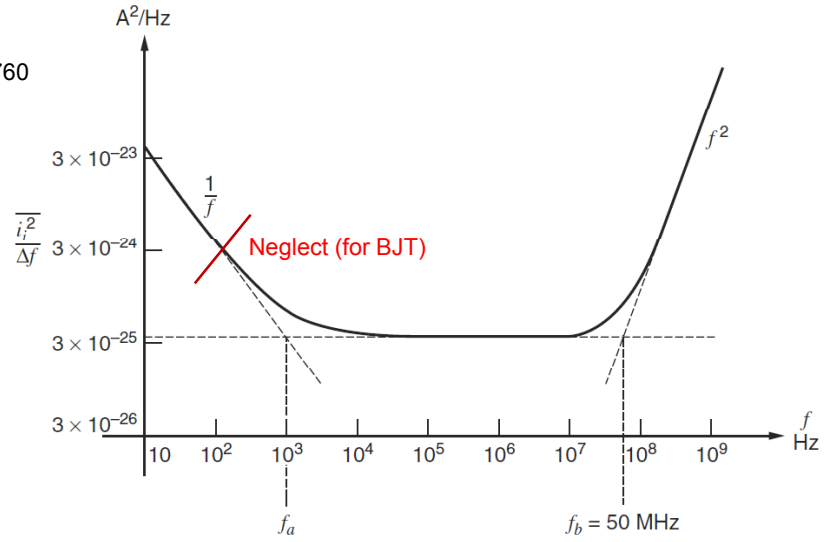
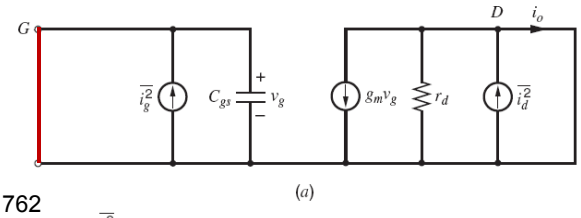


Figure 11.24 Equivalent input noise current spectral density of a bipolar transistor with $I_C = 100\ \mu A$, $\beta_0 = \beta_F = 100$, $f_T = 500$ MHz. Typical flicker noise is included.

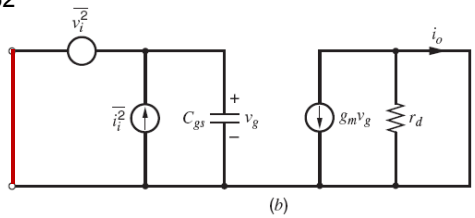
MOS Input Voltage Noise



(a)

$$\overline{i_{o1}^2} = \overline{i_c^2}$$

Text, p. 762

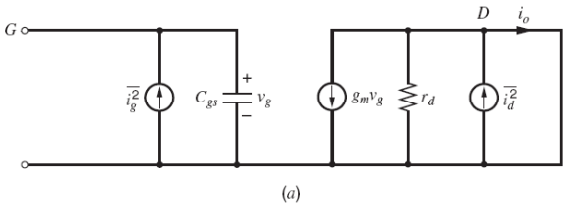


(b)

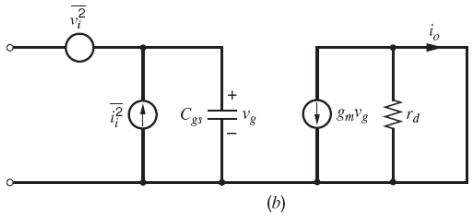
$$\overline{i_{o2}^2} = g_m^2 \overline{v_i^2}$$

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

MOS Input Current Noise



(a)



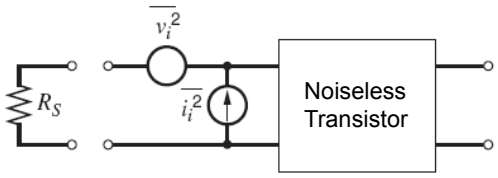
(b)

$$\overline{i_{o1}^2} = \overline{i_c^2} + g_m^2 \cdot \overline{i_g^2} \cdot \frac{1}{\omega^2 C_{gs}^2} \cong \overline{i_c^2} + \overline{i_g^2} \cdot \left(\frac{\omega_T}{\omega}\right)^2$$

$$\overline{i_{o2}^2} \cong \overline{i_i^2} \cdot \left(\frac{\omega_T}{\omega}\right)^2$$

$$\overline{i_i^2} = \overline{i_g^2} + \left(\frac{\omega}{\omega_T}\right)^2 \overline{i_c^2} \qquad \frac{\overline{i_i^2}}{\Delta f} \cong 2qI_G + \left(\frac{f}{f_T}\right)^2 \left(4kT\gamma g_m + \frac{K_f g_m^2}{WLC_{ox}}\frac{1}{f}\right)$$

BJT versus MOS (1)



- Consider low source impedance → voltage noise will dominate

$$\left[\frac{\overline{v_i^2}}{\Delta f}\right]_{BJT} \cong 4kT\left(r_b + \frac{1}{2g_m}\right) \qquad \left[\frac{\overline{v_i^2}}{\Delta f}\right]_{MOS} \cong 4kT\gamma\frac{1}{g_m} + \frac{K_f}{WLC_{ox}}\frac{1}{f}$$

- BJT is usually superior
  - Need less g_m for approximately same noise
  - g_m/I is higher, making it easier to achieve low noise at a given current budget

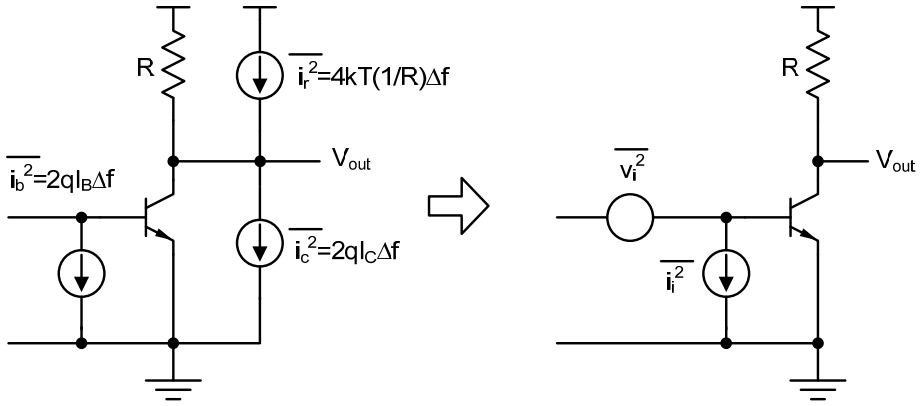
BJT versus MOS (2)

- Consider high source impedance → current noise will dominate

$$\left[\frac{\overline{i_i^2}}{\Delta f}\right]_{BJT} \cong 2q\left(I_B + \frac{I_C}{|\beta(j\omega)|^2}\right) \qquad \left[\frac{\overline{i_i^2}}{\Delta f}\right]_{MOS} \cong 2qI_G + \left(\frac{f}{f_T}\right)^2\left(4kT\gamma g_m + \frac{K_f g_m^2}{WLC_{ox}}\frac{1}{f}\right)$$

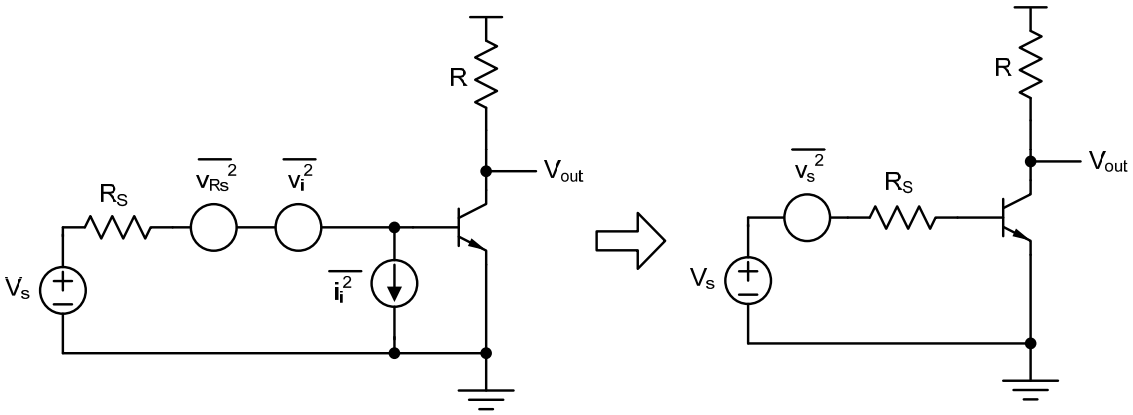
- MOS is usually superior
  - Gate leakage current (I_G) is typically much smaller than BJT base current
    - Unless the gate oxide becomes very thin, as e.g. in a 45-nm CMOS process that does not use high-k gate dielectrics

Example Revisited (With finite R_S)



$$\frac{\overline{v_i^2}}{\Delta f} \cong 4kT\left(r_b + \frac{1}{2g_m} + \frac{1}{g_m^2 R}\right) \qquad \frac{\overline{i_i^2}}{\Delta f} = 2q\left(I_B + \frac{I_C}{|\beta(j\omega)|^2}\right) + \frac{4kT\frac{1}{R}}{|\beta(j\omega)|^2}$$

Calculation of Equivalent Source Noise

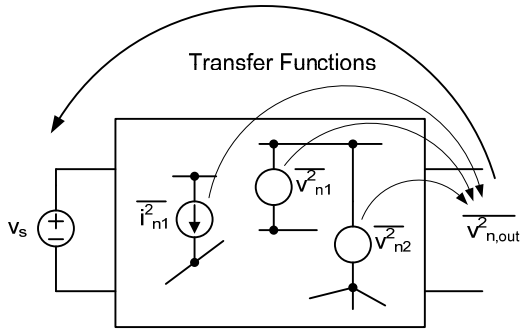


$$\overline{v_s^2} = \overline{v_{Rs}^2} + \overline{v_i^2} + R_s^2\overline{i_i^2}$$

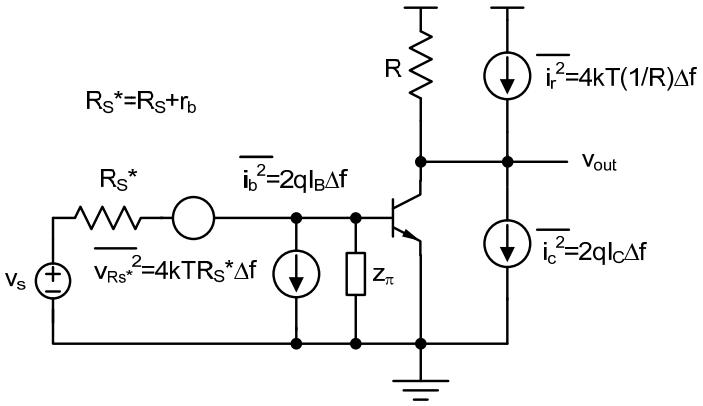
$$\frac{\overline{v_s^2}}{\Delta f} = 4kTR_S + 4kT\left(r_b + \frac{1}{2g_m} + \frac{1}{g_m^2 R}\right) + R_s^2\left[2q\left(I_B + \frac{I_C}{|\beta(j\omega)|^2}\right) + \frac{4kT\frac{1}{R}}{|\beta(j\omega)|^2}\right]$$

Alternative Method

- Draw complete circuit with all noise sources
- Compute transfer function from each noise source to the output
- Refer to output using squared transfer function
- Sum all output referred noise components
- Divided obtained sum by squared transfer function from input source to the output



Output Referred Noise PSD



$$H(s)=\frac{v_{out}}{v_s}=-g_mR\frac{z_\pi}{z_\pi+R_s}$$

$$=-\frac{g_mR}{1+\frac{g_mR_s}{\beta(j\omega)}}$$

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

(neglecting base shot noise for simplicity)

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

Source Referred Noise PSD

$$\frac{\overline{v_s^2}}{\Delta f} = 4kTR_S^* + \frac{\left(4kT\frac{1}{R} + 2qI_C\right)\cdot R^2\left|1+\frac{g_m R_s}{\beta(j\omega)}\right|^2}{(g_m R)^2}$$

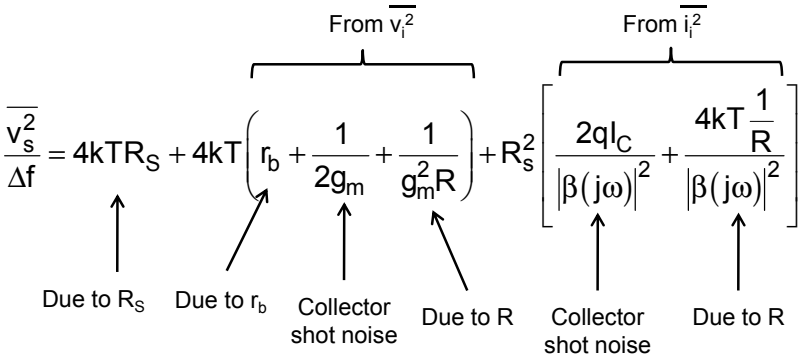
$$= 4kTR_S + 4kT\left(r_b + \left[\frac{1}{2g_m} + \frac{1}{g_m^2 R}\right]\cdot\left|1+\frac{g_m R_s}{\beta(j\omega)}\right|^2\right)$$

- This result <u>does not match</u> what we would expect from the analysis with BJT input referred generators:

$$\frac{\overline{v_s^2}}{\Delta f} = 4kTR_S + 4kT\left(r_b + \frac{1}{2g_m} + \frac{1}{g_m^2 R}\right) + R_s^2\left[\frac{2qI_C}{|\beta(j\omega)|^2} + \frac{4kT\frac{1}{R}}{|\beta(j\omega)|^2}\right]$$

Reason for Discrepancy

- It turns out that the result obtained using the BJT input referred generators is not quite correct

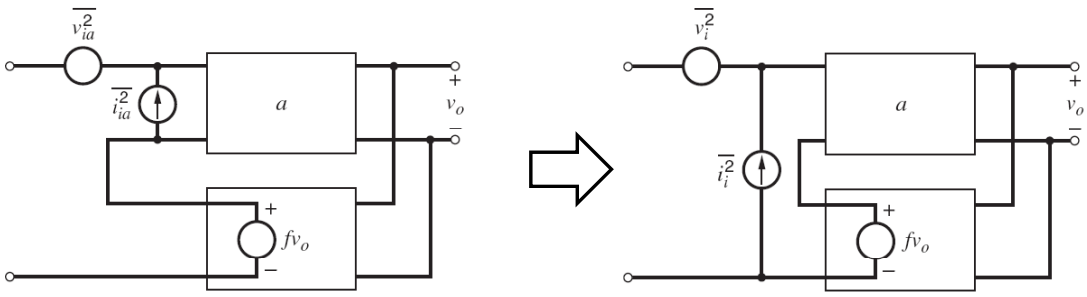


- In this expression, we are adding the powers of correlated noise currents without taking the correlation into account!
- This leads to an expression that overestimates the noise

Conclusion on Input Referred Noise Generators

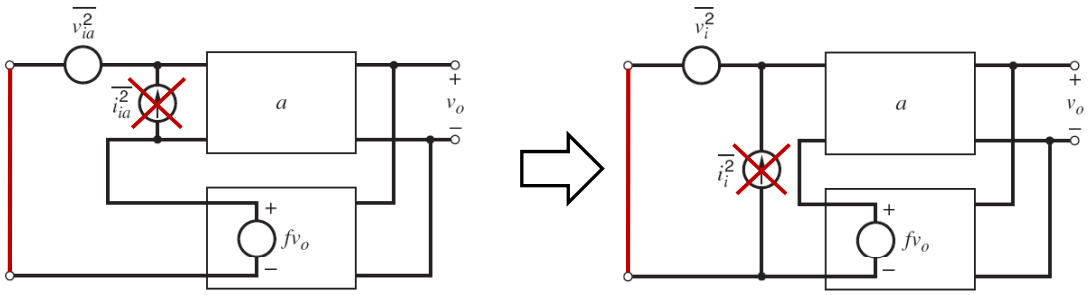
- Working with input referred noise generators saves time and works well if the input source can be modeled close to a an ideal voltage source (R_S is small) or an ideal current source (R_S is large)
  - In this case, either the input referred noise voltage or current will clearly dominate and only one of the two generators must be considered
  - Note that for R_S=0, the two expression on slide 65 match perfectly
- For any scenario in-between, working with input referred generators can still work as long as the input referred current and voltage generators are statistically independent, i.e. they are due physically distinct noise mechanisms
  - This is not the case for the expressions of slide 65
- When all of these conditions fail, it is a must to analyze the circuit from first principles, i.e. consider all noise sources at their root and refer them to the point(s) of interest via their individual transfer functions

Noise Performance of Feedback Circuits: Ideal Feedback



- Consider an amplifier with input referred noise sources placed into an ideal series-shunt feedback configuration (no loading effects)
- We can identify the equivalent input noise generators for the overall circuit in the same way we have done this previously
  - Short-circuit the input, find input noise voltage generator that yields the same output noise in both circuits
  - Open-circuit the input, find input noise current generator that yields the same output noise in both circuits

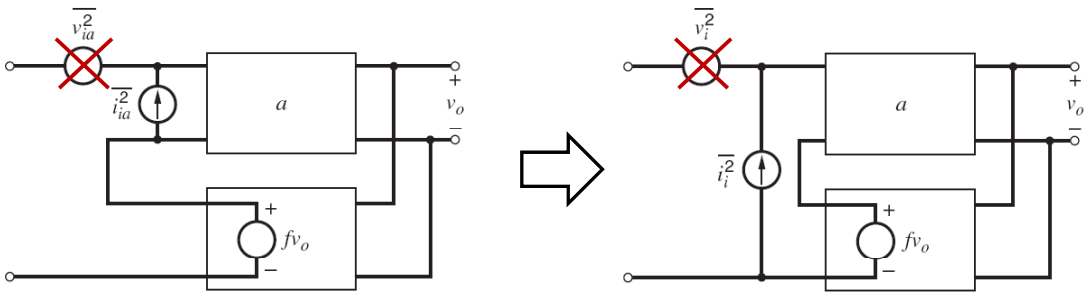
Input Shorted



- Equal output noise in both circuits is obtained for

$$\overline{v_i^2} = \overline{v_{ia}^2}$$

Input Open



- Equal output noise in both circuits is obtained for

$$\overline{i_i^2} = \overline{i_{ia}^2}$$

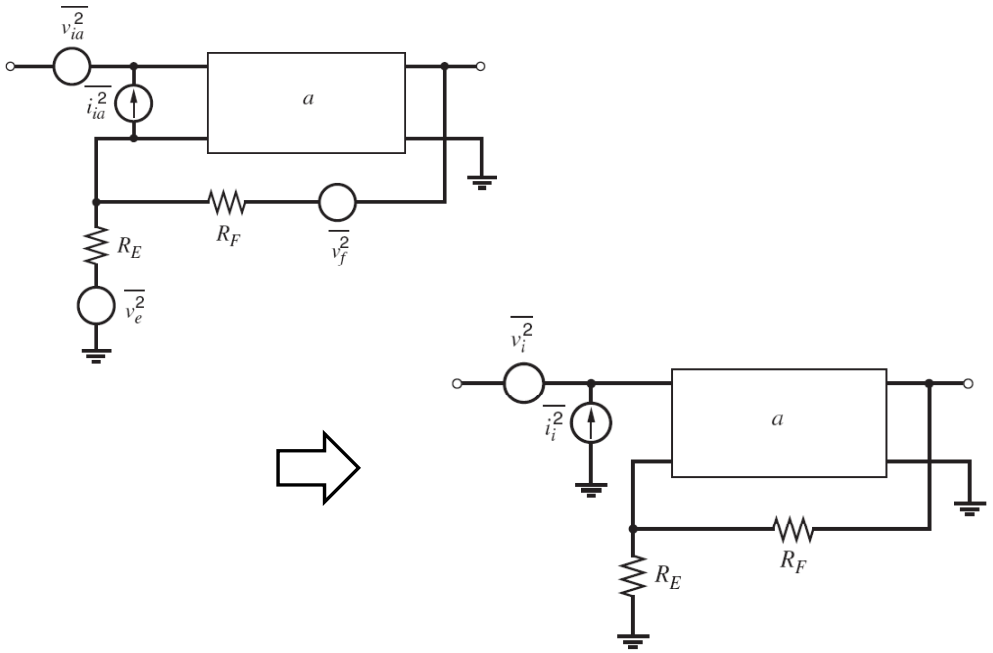
Observation

- For the idealized example studied on the previous slides, we see that the equivalent input noise generators of the amplifier can be moved *unchanged* outside the feedback loop
  - Applying feedback has no effect on the circuit's noise performance
  - Note that this is very different from the effect of feedback on distortion performance
- This results holds for all four possible (ideal) feedback configurations
  - Prove this as an exercise
- In practical feedback configurations, the input referred generators must be computed while taking loading effects into account
  - Loading makes the calculations more complicated, and generally worsens the noise performance
  - In the best possible design outcomes, the noise performance can approach (but not surpass) that of an idealized configuration

Including Loading and Noise from the Feedback Network

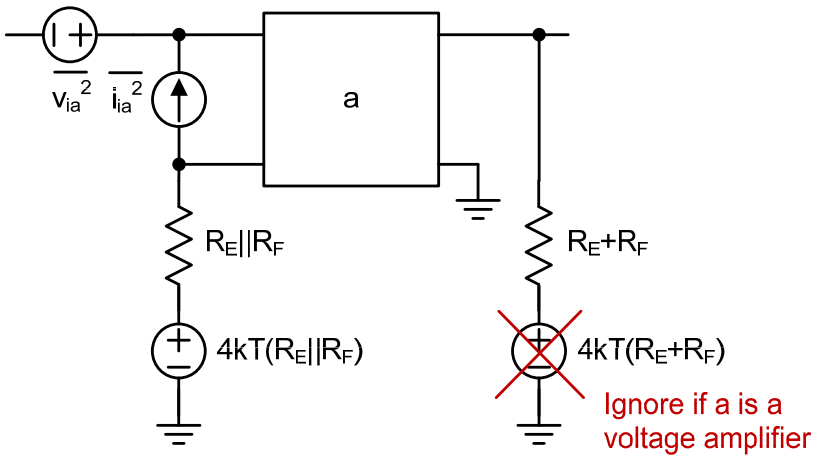
- The most general way of including loading and noise from the feedback network is to apply the same procedure as before
  - Short-circuit the input, find input noise voltage generator that yields the same output noise in both circuits
  - Open-circuit the input, find input noise current generator that yields the same output noise in both circuits
- A more convenient way to include loading and noise from the feedback network is to use the same two-port approximations we have already utilized for transfer function analysis
- Procedure
  - Absorb loading effects into the basic amplifier and work with ideal feedback network
  - Re-compute the input referred noise generators of the basic amplifier in presence of loading
  - Using the previous result, the re-computed noise generators of the basic amplifier can now be moved outside the feedback loop

Example: Practical Series-Shunt Circuit



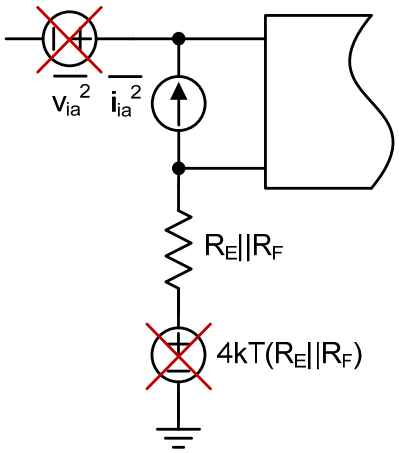
Basic Amplifier with Loading

- Step 1
  - Redraw the basic amplifier with loading included



Input Referred Noise Current (Open-Circuited Input)

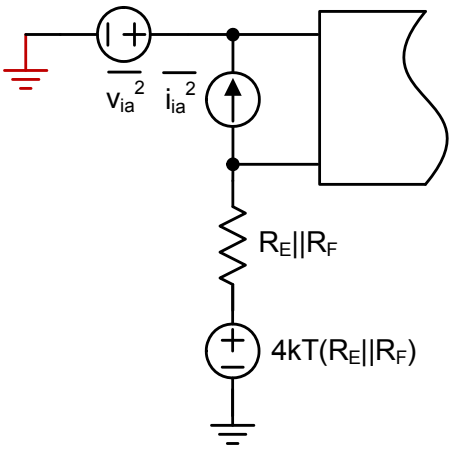
- Step 2
  - Compute input referred current noise
  - The result corresponds to the desired i_i



$$\overline{i_i^2} = \overline{i_{ia}^2}$$

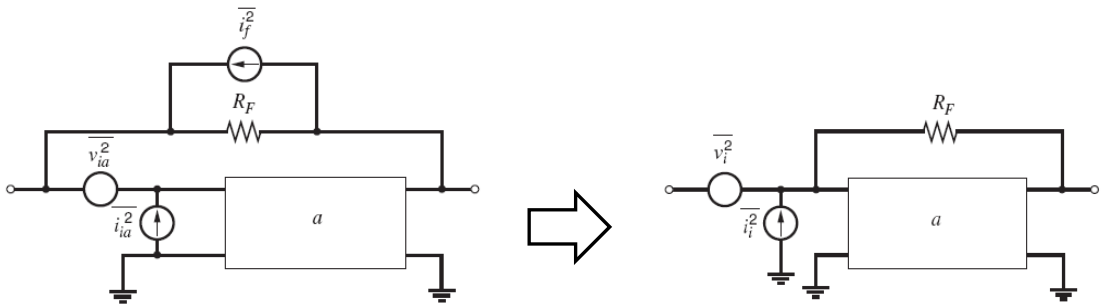
Input Referred Noise Voltage (Short-Circuited Input)

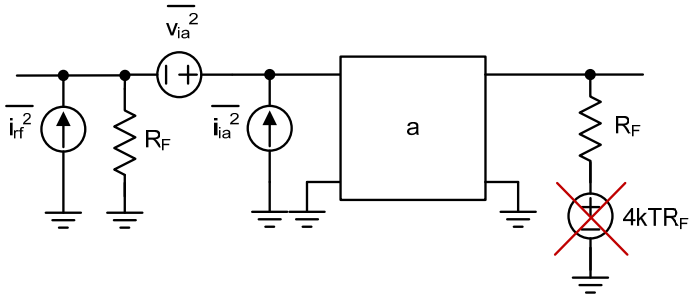
- Step 3
  - Compute input referred voltage noise
  - The result corresponds to the desired v_i



$$\overline{v_i^2} = \overline{v_{ia}^2} + \overline{i_{ia}^2} \cdot \left(R_E \,||\, R_F\right)^2 + 4kT\left(R_E \,||\, R_F\right)\Delta f$$

Example: Practical Shunt-Shunt Circuit





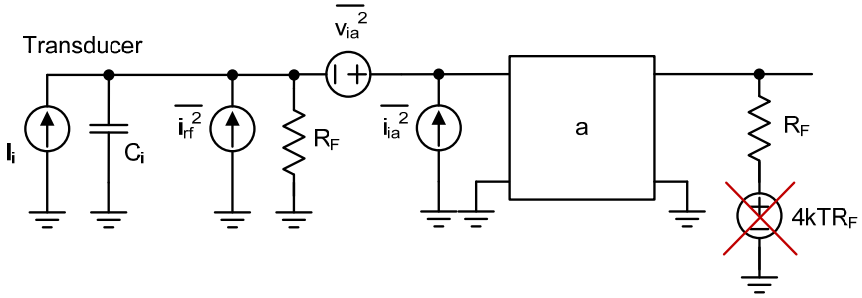
$$\overline{v_i^2} = \overline{v_{ia}^2}$$

$$\overline{i_i^2} = \overline{i_{ia}^2} + \frac{\overline{v_{ia}^2}}{R_F^2} + 4kT\frac{1}{R_F}\Delta f$$

(assuming i_{ia} and v_{ia} are uncorrelated)

High Frequency Issue

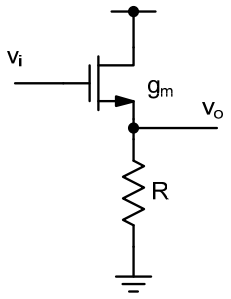
- At low frequencies, it is typically not hard to minimize the input noise current contribution due to v_{ia}
- However, at high frequencies, any shunt capacitance at the input tends to make the v_{ia} contribution more significant
- Means that v_{ia} must be minimized in high speed circuits, typically by increasing g_m of the input device → higher power dissipation

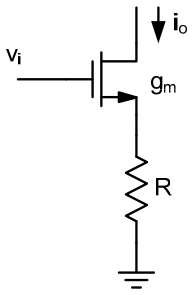


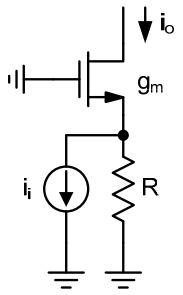
$$\overline{i_i^2} = \overline{i_{ia}^2} + \overline{v_{ia}^2}\left|\frac{1}{R_F} + j\omega C_i\right|^2 + 4kT\frac{1}{R_F}\Delta f$$

(assuming i_{ia} and v_{ia} are uncorrelated)

Local Feedback Circuit Examples





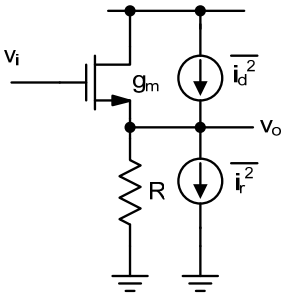


$$A_v = \frac{v_o}{v_i} = \frac{g_m R}{1 + g_m R} \qquad G_m = \frac{i_o}{v_i} = \frac{g_m}{1 + g_m R} \qquad A_i = \frac{i_o}{i_i} = \frac{g_m R}{1 + g_m R}$$

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

Source Follower

- The noise performance can be analyzed by treating the circuit as a series feedback stage (see text, sections 8.6.2 and 11.7.2)
  - We will do a direct analysis instead



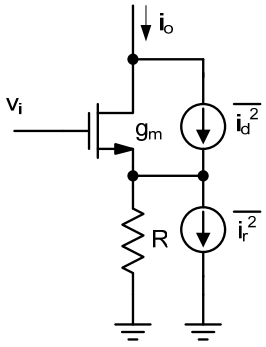
$$\overline{v_o^2} = \left(\overline{i_d^2} + i_r^2\right) \cdot \left(\frac{1}{g_m + \frac{1}{R}}\right)^2 = \frac{\overline{i_d^2} + i_r^2}{g_m^2} \cdot A_v^2$$

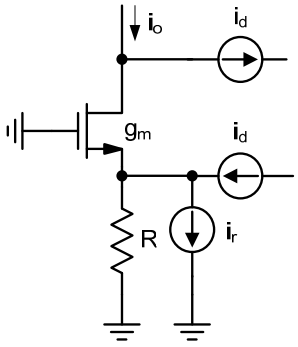
Often negligible

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

- 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

Degenerated Common Source Stage





$$i_o = i_d + (i_r - i_d)\frac{g_m}{g_m + \frac{1}{R}}$$

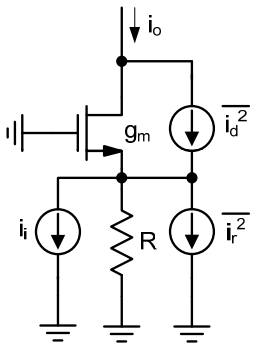
$$= i_r \frac{g_m R}{1 + g_m R} + i_d \frac{1}{1 + g_m R}$$

$$= i_r G_m R + i_d \frac{G_m}{g_m}$$

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

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

Common Gate Stage



$$i_o = i_r \frac{g_m R}{1 + g_m R} + i_d \frac{1}{1 + g_m R}$$

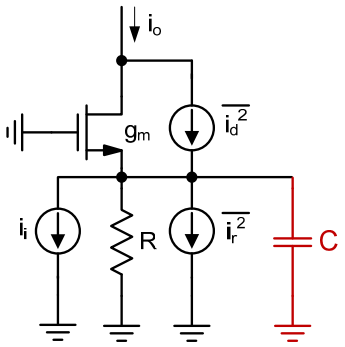
$$= i_r A_i + i_d \frac{A_i}{g_m R}$$

Often negligible

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

- 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



$$\overline{i_i^2} = 4kT\frac{1}{R}\Delta f + 4kT\gamma g_m \Delta f \frac{\left|\frac{1}{R} + j\omega C\right|^2}{g_m^2}$$

$$\cong 4kT\frac{1}{R}\Delta f + 4kT\gamma g_m \Delta f \left(\frac{\omega C}{g_m}\right)^2$$

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

Summary – Noise in Feedback Circuits

- Applying ideal feedback around an amplifier does not alter its input referred noise performance
- In practical circuits, loading and noise from the feedback network tend to deteriorate the circuit's overall noise performance
  - This is especially true at high frequencies, where parasitic capacitances can increase the noise transfer from sources that are typically negligible at low frequencies
- Loading effects can be considered by applying the same two-port approximation methods used in the transfer function analysis of practical feedback amplifiers
  - Absorb loading and feedback network noise sources into forward amplifier and work with idealized feedback result

Additional Topics in Noise Analysis

- Covered in EE314
  - RF-centric metrics
    - Noise figure
    - Receiver sensitivity
  - Phase noise in oscillators
- Covered in EE315A,B
  - 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 9

Distortion Analysis

B. Murmann

Stanford University

Reading Material: Sections 1.4.1, 5.3.2

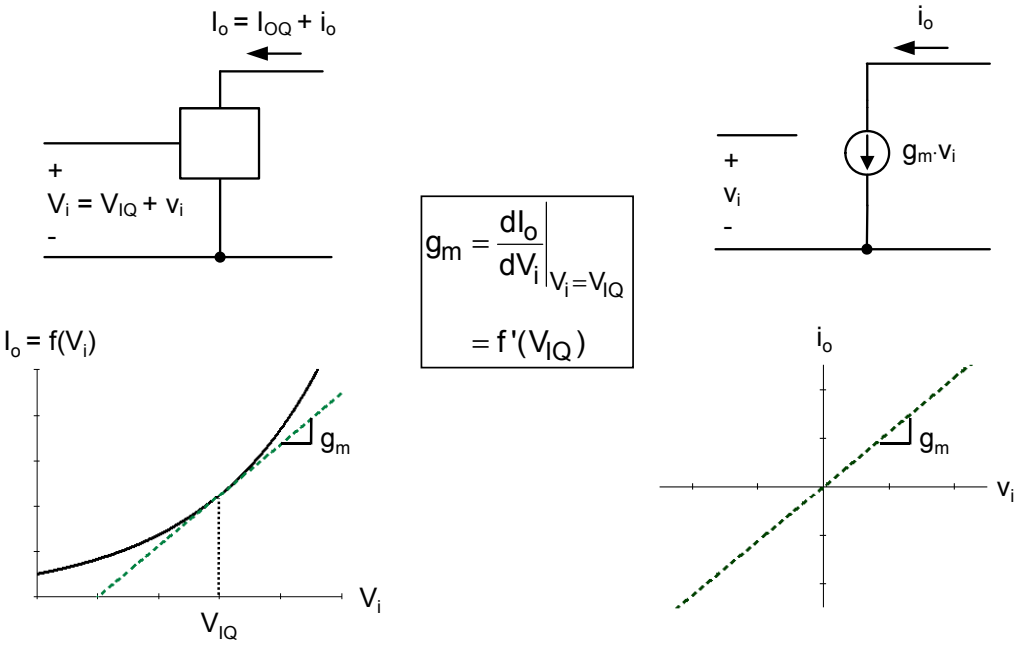
Overview

- Low frequency distortion analysis
- Effect of feedback on distortion
- High frequency distortion analysis

Introduction

- All electronic circuits exhibit some level of nonlinear behavior
  - The resulting waveform distortion is not captured in small-signal models
- In the first section of this chapter, we will begin by looking at the basic tools needed to analyze “memoryless” nonlinearities, i.e. nonlinearities that can be represented by a frequency independent model
  - Such models are valid in a frequency range where all capacitances and inductances in the circuit of interest can be ignored
- As a driving example, we will analyze the nonlinearity in the V-I transduction of BJTs and MOSFETs
- The general approach taken is to model the nonlinearities via a power series that links the input and output of the circuit
  - This approach is useful and accurate for the case of “small distortion” and cannot be used to predict the effect of gross distortion, e.g. due to signal clipping

Small-Signal AC Model



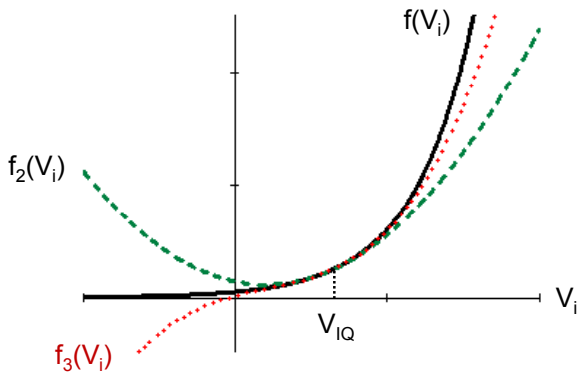
$$g_m = \left.\frac{dI_o}{dV_i}\right|_{V_i = V_{IQ}} = f'(V_{IQ})$$

Taylor Series Model

$$f(V_i) = f(V_{IQ}) + \frac{f'(V_{IQ})}{1!}(V_i - V_{IQ}) + \frac{f''(V_{IQ})}{2!}(V_i - V_{IQ})^2 + \frac{f^{(3)}(V_{IQ})}{3!}(V_i - V_{IQ})^3 + \ldots$$

$f_2(V_i)$

$f_3(V_i)$



Relationship Between Incremental Variables

- Using $v_i = V_i - V_{IQ}$ and $i_o = I_o - I_{OQ} = f(V_i) - f(V_{IQ})$

  we obtain $i_o = a_1 v_i + a_2 v_i^2 + a_3 v_i^3 + \ldots$

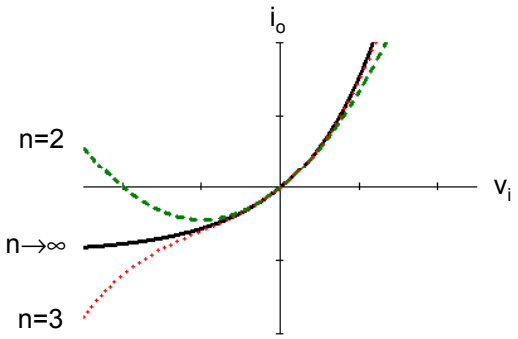
  where $a_m = \frac{f^{(m)}(V_{IQ})}{m!}$

- Note that $a_1 \equiv g_m$ $a_2 \equiv \frac{1}{2} g_m'$ $a_3 \equiv \frac{1}{6} g_m''$

- In practice, it is often sufficient to work with a truncated n^{th} order power series

$$i_o \cong a_1 v_i + a_2 v_i^2 + \ldots + a_n v_i^n$$

Graphical Illustration



- A model that relates the incremental signal components (v_i, i_o) though a nonlinear expression is sometimes called “large-signal AC model”
- The accuracy of a truncated power series model depends on the signal range and the curvature of the actual transfer function
  - Using a higher order series generally helps, but also makes the analysis more complex
  - As we will see, using a third order series is often sufficient to model the relevant distortion effects in practical, weakly nonlinear circuits

Harmonic Distortion Analysis

- Apply a sinusoidal signal and collect harmonic terms in the output signal

$$v_i = \hat{v}_i \cdot \cos(\omega t)$$

$$i_o = a_1 \hat{v}_i \cos(\omega t) + a_2 \left[\hat{v}_i \cos(\omega t)\right]^2 + a_3 \left[\hat{v}_i \cos(\omega t)\right]^3 + \ldots$$

$$\cos^2(\alpha) = \frac{1}{2}\left[\cos(2\alpha) + 1\right] \qquad \cos^3(\alpha) = \frac{1}{4}\left[\cos(3\alpha) + 3\cos(\alpha)\right]$$

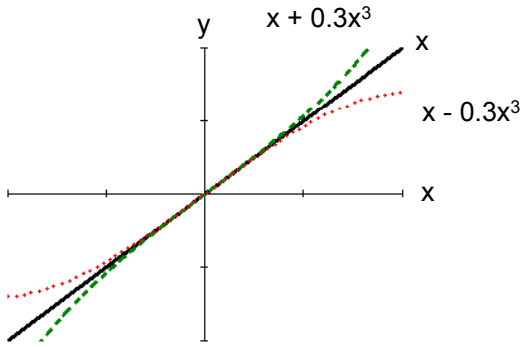
$$\therefore i_o = \left[\frac{1}{2} a_2 \hat{v}_i^2\right]$$ **DC shift**

$$+ \left[a_1 \hat{v}_i + \frac{3}{4} a_3 \hat{v}_i^3\right] \cos(\omega t)$$ **Fundamental**

$$+ \left[\frac{1}{2} a_2 \hat{v}_i^2\right] \cos(2\omega t) + \left[\frac{1}{4} a_3 \hat{v}_i^3\right] \cos(3\omega t) + \ldots$$ **Harmonics**

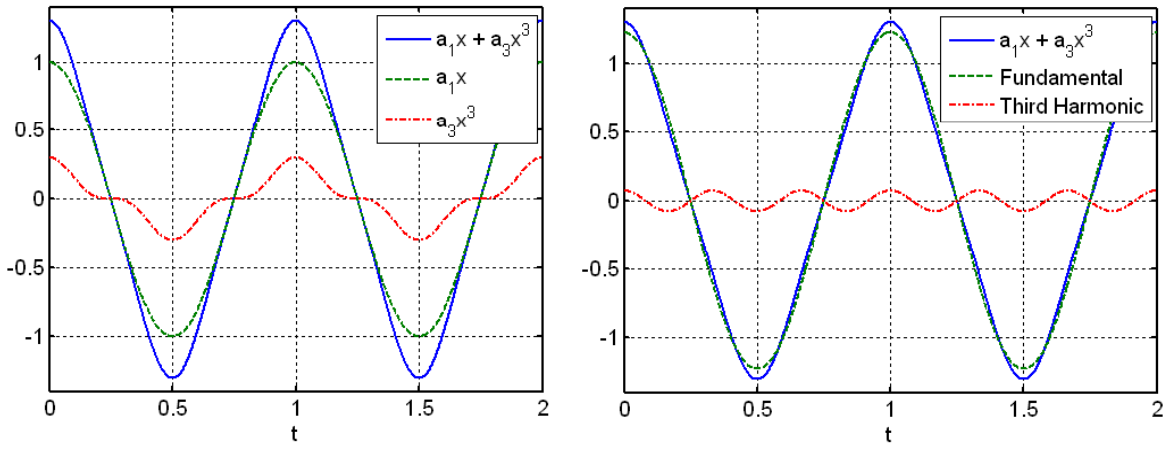
Observations

- The quadratic term (a_2) give rises to an undesired second harmonic tone and a DC shift
- The cubic term (a_3) give rises to an undesired third harmonic tone and it also modifies the amplitude of the fundamental
  - $a_3 < 0$ → "gain compression"
  - $a_3 > 0$ → "gain expansion"



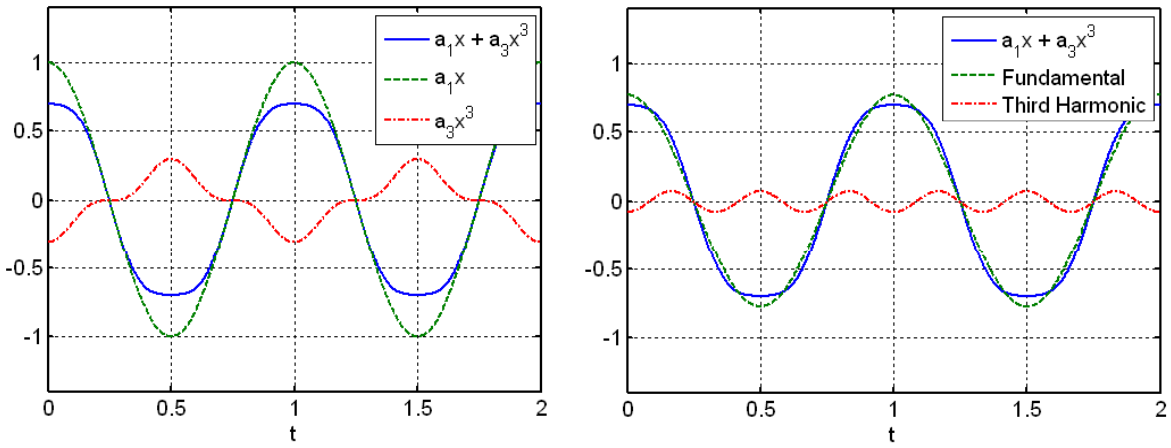
Waveforms with Gain Expansion

$$x = \cos(2\pi t) \qquad a_1 = 1 \qquad a_3 = 0.3$$



Waveforms with Gain Compression

$$x = \cos(2\pi t) \qquad a_1 = 1 \qquad a_3 = -0.3$$



Higher Order Terms

$$\cos^m(\alpha) = \frac{1}{2^m}\left(e^{j\alpha} + e^{-j\alpha}\right)^m = \frac{1}{2^m}\sum_{k=0}^{m}\binom{m}{k}e^{jk\alpha}e^{-j(m-k)\alpha}$$

$$\cos^4(\alpha) = \frac{1}{8}\left[\cos(4\alpha) + 4\cos(2\alpha) + 3\right]$$

$$\cos^5(\alpha) = \frac{1}{16}\left[\cos(5\alpha) + 5\cos(3\alpha) + 10\cos(\alpha)\right]$$

- Can show that
  - Terms raised to even powers of m affect the DC shift and even harmonics up to m
  - Terms raised to odd powers of m affect the fundamental and odd harmonics up to m

Inspection of 4th and 5th Order Contributions

$$\therefore i_o = \left[\frac{1}{2}a_2\hat{v}_i^2 + \frac{3}{8}a_4\hat{v}_i^4\right] + \left[a_1\hat{v}_i + \frac{3}{4}a_3\hat{v}_i^3 + \frac{5}{16}a_5\hat{v}_i^5\right]\cos(\omega t) + \left[\frac{1}{2}a_2\hat{v}_i^2 + \frac{1}{2}a_4\hat{v}_i^4\right]\cos(2\omega t) + \left[\frac{1}{4}a_3\hat{v}_i^3 + \frac{5}{16}a_5\hat{v}_i^5\right]\cos(3\omega t) + \left[\frac{1}{8}a_4\hat{v}_i^4\right]\cos(4\omega t) + \left[\frac{1}{16}a_5\hat{v}_i^5\right]\cos(5\omega t)$$

- As long as

$$\left|a_4\hat{v}_i^4\right| << \left|a_2\hat{v}_i^2\right| \quad \text{and} \quad \left|a_5\hat{v}_i^5\right| << \left|a_3\hat{v}_i^3\right|$$

or equivalently

$$\hat{v}_i << \sqrt{\left|\frac{a_2}{a_4}\right|} \quad \text{and} \quad \hat{v}_i << \sqrt{\left|\frac{a_3}{a_5}\right|}$$

the 4th and 5th order terms can be neglected

- This condition is usually met in practical, weakly nonlinear circuits

Fractional Harmonic Distortion Metrics

$$HD_2 = \frac{\text{amplitude of second harmonic distortion signal}}{\text{amplitude of fundamental}}$$

$$HD_3 = \frac{\text{amplitude of third harmonic distortion signal}}{\text{amplitude of fundamental}}$$

- Including only contributions from terms up to 3rd order, these quantities become

$$HD_2 \cong \left|\frac{\frac{1}{2}a_2\hat{v}_i^2}{a_1\hat{v}_i + \frac{3}{4}a_3\hat{v}_i^3}\right| \cong \frac{1}{2}\left|\frac{a_2}{a_1}\right|\hat{v}_i$$

$$HD_3 \cong \left|\frac{\frac{1}{4}a_3\hat{v}_i^3}{a_1\hat{v}_i + \frac{3}{4}a_3\hat{v}_i^3}\right| \cong \frac{1}{4}\left|\frac{a_3}{a_1}\right|\hat{v}_i^2$$

Total Harmonic Distortion

$$\text{THD} = \sqrt{\frac{\text{total power of distortion signals}}{\text{power of fundamental}}}$$

$$= \sqrt{HD_2^2 + HD_3^2 + HD_4^2 + \ldots}$$

- THD is often dominated by the HD_2 and/or HD_3 term
- Typical application requirements
  - Telephone audio: THD < ~10%
  - Video: THD < ~1%
  - RF low noise amplifiers: THD < ~0.1%
  - High quality audio: THD < ~0.01%

Intermodulation Distortion (1)

- Consider applying two tones to the nonlinear device

$$v_i = \hat{v}_{i1} \cdot \cos(\omega_1 t) + \hat{v}_{i2} \cdot \cos(\omega_2 t)$$

$$i_o = a_1\left[\hat{v}_{i1} \cdot \cos(\omega_1 t) + \hat{v}_{i2} \cdot \cos(\omega_2 t)\right]$$

$$+a_2\left[\hat{v}_{i1} \cdot \cos(\omega_1 t) + \hat{v}_{i2} \cdot \cos(\omega_2 t)\right]^2$$

$$+a_3\left[\hat{v}_{i1} \cdot \cos(\omega_1 t) + \hat{v}_{i2} \cdot \cos(\omega_2 t)\right]^3 + \ldots$$

- Inspect second-order term

$$a_2\left[\hat{v}_{i1} \cdot \cos(\omega_1 t) + \hat{v}_{i2} \cdot \cos(\omega_2 t)\right]^2 = a_2\left[\hat{v}_{i1} \cdot \cos(\omega_1 t)\right]^2 + a_2\left[\hat{v}_{i2} \cdot \cos(\omega_2 t)\right]^2 + 2a_2\left[\hat{v}_{i1}\hat{v}_{i2} \cdot \cos(\omega_1 t)\cos(\omega_2 t)\right]$$

Causes HD

New

Second-Order Intermodulation

$$2a_2\left[\hat{v}_{i1}\hat{v}_{i2}\cdot\cos(\omega_1 t)\cos(\omega_2 t)\right]=a_2\hat{v}_{i1}\hat{v}_{i2}\left[\cos\left(\{\omega_1+\omega_2\}t\right)+\cos\left(\{\omega_1-\omega_2\}t\right)\right]$$

- The output will contain tones at the sums and differences of the applied frequencies
- We define the fractional second-order intermodulation as

$$IM_2=\frac{\text{amplitude of second-order IM components (for } \hat{v}_{i1}=\hat{v}_{i2}=\hat{v}_i)}{\text{amplitude of fundamentals}}$$

$$=\left|\frac{a_2}{a_1}\right|\frac{\hat{v}_i^2}{\hat{v}_i}=\left|\frac{a_2}{a_1}\right|\hat{v}_i$$

$$=2HD_2$$

Third-Order Intermodulation (1)

$$a_3\left[\hat{v}_{i1}\cdot\cos(\omega_1 t)+\hat{v}_{i2}\cdot\cos(\omega_2 t)\right]^3=a_3\left[\hat{v}_{i1}\cdot\cos(\omega_1 t)\right]^3+a_3\left[\hat{v}_{i2}\cdot\cos(\omega_2 t)\right]^3$$

Causes HD

$$+3a_3\left[\hat{v}_{i1}\hat{v}_{i2}^2\cdot\cos(\omega_1 t)\cos^2(\omega_2 t)\right]$$

$$+3a_3\left[\hat{v}_{i1}^2\hat{v}_{i2}\cdot\cos^2(\omega_1 t)\cos(\omega_2 t)\right]$$

New

$$3a_3\left[\hat{v}_{i1}\hat{v}_{i2}^2\cdot\cos(\omega_1 t)\cos^2(\omega_2 t)\right]$$

$$=\frac{3}{4}a_3\hat{v}_{i1}\hat{v}_{i2}^2\left[2\cos(\omega_1 t)+\cos\left(\{2\omega_2-\omega_1\}t\right)+\cos\left(\{2\omega_2+\omega_1\}t\right)\right]$$

"Gain desensitization term" For $a_3<0$, large v_{i2} reduces fundamental tone due to v_{i1}

Third-order Intermodulation Products

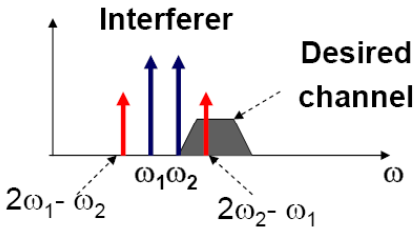
Third-Order Intermodulation (2)

- Third-order intermodulation products appear at $(2\omega_2 \pm \omega_1)$ and $(2\omega_1 \pm \omega_2)$
- We define the fractional third-order intermodulation as

$$IM_3 = \frac{\text{amplitude of third-order IM components (for } \hat{v}_{i1} = \hat{v}_{i2} = \hat{v}_i)}{\text{amplitude of fundamentals}}$$

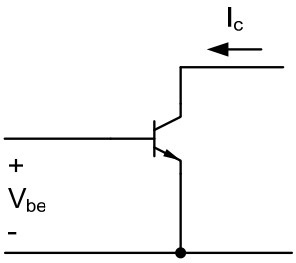
$$= \frac{3}{4}\left|\frac{a_3}{a_1}\right|\frac{\hat{v}_i^3}{\hat{v}_i} = \frac{3}{4}\left|\frac{a_3}{a_1}\right|\hat{v}_i^2 = 3HD_3$$

- Note that for $\omega_2 \cong \omega_1$, the third-order intermodulation products are close to the original frequencies and cannot be filtered out
  - This is a significant issue in narrowband systems



Distortion in a CE Stage (1)

$$I_c = I_s e^{\frac{V_{be}}{V_T}} \qquad V_T = \frac{kT}{q}$$



$$a_1 = \left.\frac{dI_c}{dV_{be}}\right|_{V_{be}=V_{BEQ}} = \left.\frac{I_s}{V_T}e^{\frac{V_{be}}{V_T}}\right|_{V_{be}=V_{BEQ}}$$

$$= \frac{I_{CQ}}{V_T} \equiv g_m$$

$$a_2 = \frac{1}{2}\left.\frac{d^2I_c}{dV_{be}^2}\right|_{V_{be}=V_{BEQ}} = \frac{1}{2}\frac{I_{CQ}}{V_T^2}$$

$$a_m = \frac{1}{m!}\frac{I_{CQ}}{V_T^m}$$

Distortion in a CE Stage (2)

$$HD_2 \cong \frac{1}{2}\left|\frac{a_2}{a_1}\right|\hat{v}_{be} = \frac{1}{4}\frac{\hat{v}_{be}}{V_T} \qquad HD_3 \cong \frac{1}{4}\left|\frac{a_3}{a_1}\right|\hat{v}_{be}^2 = \frac{1}{24}\left(\frac{\hat{v}_{be}}{V_T}\right)^2$$

- Low distortion in the collector current quires the B-E voltage excursion to be much smaller than $V_T \cong 26mV$
- Checking for the valid range of a third order model yields

$$\hat{v}_{be} << \sqrt{\left|\frac{a_2}{a_4}\right|} = \sqrt{12}V_T \quad \text{and} \quad \hat{v}_{be} << \sqrt{\left|\frac{a_3}{a_5}\right|} = \sqrt{20}V_T$$

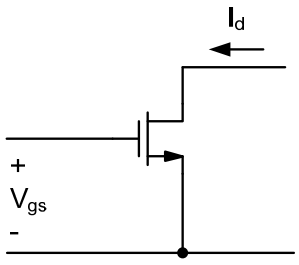
- For a B-E voltage swing of V_T, we have

$$HD_2 \cong 25\% \qquad HD_3 \cong 4.17\%$$

- Note that a typical application will demand much lower distortion

Distortion in a CS Stage (1)

$$I_d = \frac{1}{2}\mu C_{ox}\frac{W}{L}\left(V_{gs} - V_t\right)^2$$



$$a_1 = \left.\frac{dI_d}{dV_{gs}}\right|_{V_{gs}=V_{GSQ}} = \left.\mu C_{ox}\frac{W}{L}\left(V_{gs} - V_t\right)\right|_{V_{gs}=V_{GSQ}}$$

$$= \mu C_{ox}\frac{W}{L}V_{OV} = \frac{2I_{DQ}}{V_{OV}} \equiv g_m$$

$$a_2 = \left.\frac{1}{2}\frac{d^2I_d}{dV_{gs}^2}\right|_{V_{gs}=V_{GSQ}} = \frac{1}{2}\mu C_{ox}\frac{W}{L} = \frac{I_{DQ}}{V_{OV}^2}$$

$$a_3 = 0$$

Distortion in a CS Stage (2)

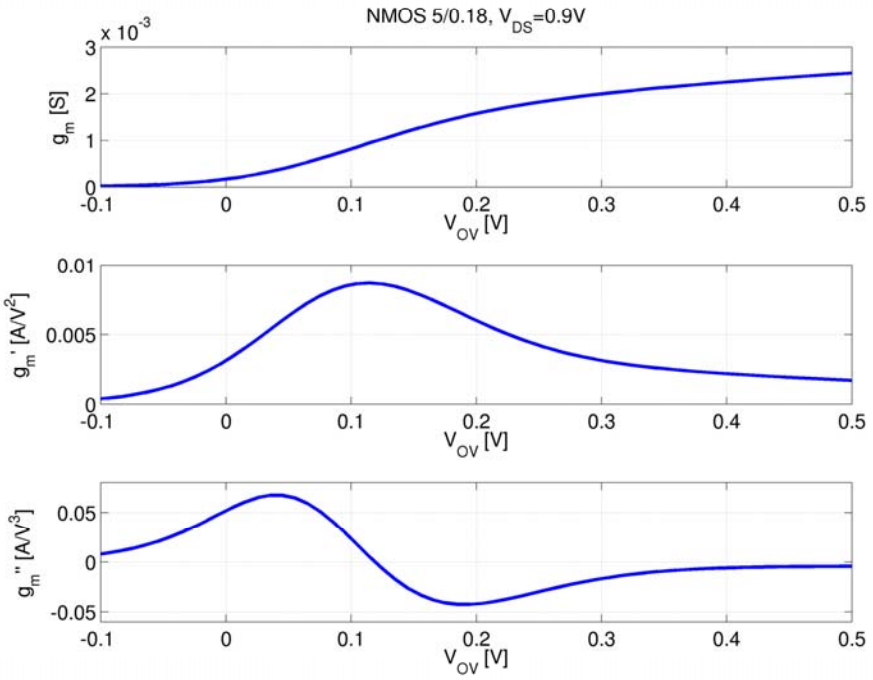
$$HD_2 \cong \frac{1}{2}\left|\frac{a_2}{a_1}\right|\hat{v}_{gs} = \frac{1}{4}\frac{\hat{v}_{gs}}{V_{OV}} \qquad HD_3 = 0$$

- Small second harmonic distortion in the drain current requires the G-S voltage excursion to be much smaller than the quiescent point gate overdrive (V_{OV}=V_{GS}-V_t)
- An idealized square-law device does not introduce high order distortion
  - However, this is not true for a real short-channel MOSFET
- Relevant effects
  - Velocity saturation, mobility reduction due to vertical field
  - Biasing in moderate or weak inversion
  - Nonlinearity in the device's output conductance
  - …

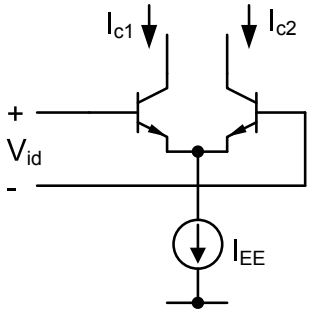
Distortion in Modern MOSFETS

- Distortion in modern MOSFET devices is generally hard to model accurately
- A basic approach for devices operating in strong inversion is to include short channel effects via basic extensions to the square law model
  - E.g. model velocity saturation as resistive source degeneration
  - See e.g. Terrovitis & Meyer, JSCC 10/2000
- Another approach is to extract the coefficients from "known-to-be-accurate" Spice models
  - Find coefficients a_m by simulating the derivatives of I-V curves
  - See e.g. Blaakmeer et al., JSSC 6/2008
- Unfortunately, generating accurate Spice models for MOSFET distortion is a very difficult task
  - See e.g. R. van Langevelde et al., IEDM 2000
- Never trust a Spice model blindly!

Distortion of EE214 NMOS Device



Distortion in a BJT Differential Pair (1)



$$I_{cd} = I_{c1} - I_{c2} = \alpha I_{EE} \tanh\left(\frac{V_{id}}{2V_T}\right)$$

$$\tanh(x) = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 - + \ldots$$

$$a_1 = \frac{\alpha I_{EE}}{2V_T} \equiv G_m \qquad a_3 = -\frac{\alpha I_{EE}}{24V_T^3} \qquad a_5 = \frac{\alpha I_{EE}}{240V_T^5}$$

- Third-order model is accurate for

$$\hat{v}_{id} << \sqrt{\left|\frac{a_3}{a_5}\right|} = \sqrt{10}V_T$$

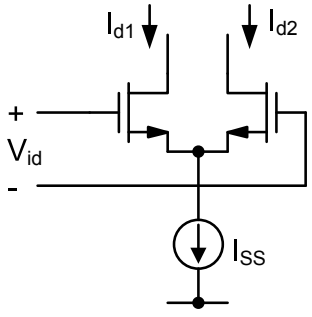
Distortion in a BJT Differential Pair (2)

- A differential pair with perfectly matched transistors does not generate any even-order distortion products
- Any mismatch (e.g. in I_s) will cause non-zero even-order terms
  - The resulting even order distortion products are typically smaller than the inherent odd-order distortion

- The HD3 performance of a BJT differential pair is better than that of a single BJT transistor

$$HD_{3,BJTdiff} \cong \frac{1}{4}\left|\frac{a_3}{a_1}\right|\hat{v}_{id}^2 = \frac{1}{48}\left(\frac{\hat{v}_{id}}{V_T}\right)^2 \qquad HD_{3,BJT} \cong \frac{1}{24}\left(\frac{\hat{v}_{be}}{V_T}\right)^2$$

Distortion in a MOS Differential Pair



$$I_{od} = I_{d1} - I_{d2} = I_{SS}\left(\frac{V_{id}}{V_{OV}}\right)\sqrt{1-\left(\frac{V_{id}}{2V_{OV}}\right)^2}$$

$$x\sqrt{1-\left(\frac{x}{2}\right)^2} = x - \frac{1}{8}x^3 - \frac{1}{128}x^5 + \ldots$$

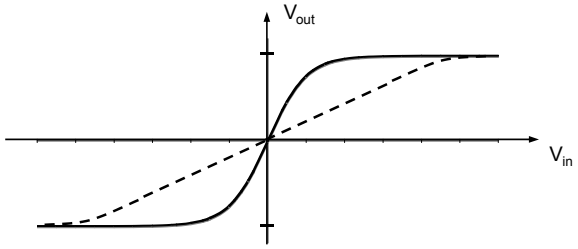
$$a_1 = \frac{I_{SS}}{V_{OV}} \equiv G_m \qquad a_3 = -\frac{I_{SS}}{8V_{OV}^3} \qquad a_5 = \frac{I_{SS}}{128V_{OV}^5}$$

- Third-order model is accurate for

$$\hat{v}_{id} << \sqrt{\left|\frac{a_3}{a_5}\right|} = 4V_{OV}$$

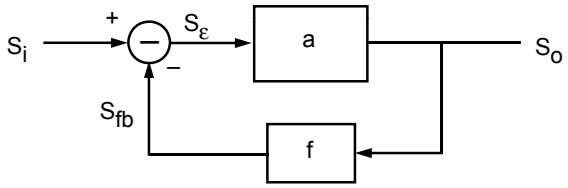
Feedback and Distortion

- Low-frequency distortion is a result of variations in the slope of an amplifier's transfer characteristic. Feedback reduces the relative variation, and thus the distortion, to the same extent that it reduces fractional changes in gain.



- Note that feedback reduces distortion without reducing the output voltage range. The gain is also reduced, but additional gain can be provided with a preamplifer the operates with smaller signal swings, and therefore less distortion.

Power Series Analysis (1)



$$S_o = a_1 S_\varepsilon + a_2 S_\varepsilon^2 + a_3 S_\varepsilon^3 + K$$

$$S_\varepsilon = S_i - f \cdot S_o$$

$$\therefore\ S_o = a_1(S_i - f \cdot S_o) + a_2(S_i - f \cdot S_o)^2 + a_3(S_i - f \cdot S_o)^3 + K$$

Power Series Analysis (2)

- Expressing S_o as a power series expansion in S_i

$$S_o = b_1 S_i + b_2 S_i^2 + b_3 S_i^3 + K$$

- Substitute this expression for S_o into the result on the previous page and compare coefficients to find b_i

$$b_1 S_i = a_1 (S_i - f \cdot b_1 S_i)$$

$$b_1 = \frac{a_1}{1 + a_1 f}$$

- Thus, the feedback reduces the coefficient of the fundamental term in the forward amplifier by 1 + af

Power Series Analysis (3)

- Second-order terms

$$b_2 S_i^2 = -a_1 f \cdot b_2 S_i^2 + a_2 (S_i - f \cdot b_1 S_i)^2$$

$$b_2 = \frac{a_2 (1 - b_1 f)^2}{1 + a_1 f} = \frac{a_2}{(1 + a_1 f)^3}$$

- Third-order terms

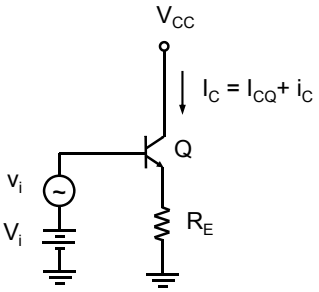
$$b_3 S_i^3 = -a_1 f \cdot b_3 S_i^3 - 2a_2 S_i^3 f \cdot b_2 (1 - f \cdot b_1) + a_3 (S_i - f \cdot b_1 S_i)^3$$

$$b_3 = \frac{a_3 (1 - b_1 f)^3 - 2a_2 f \cdot b_2 (1 - f \cdot b_1)}{1 + a_1 f} = \frac{a_3 (1 + a_1 f) - 2a_2^2 \cdot f}{(1 + a_1 f)^5}$$

Comments

- Large loop gain ($a_1 f$) leads to small nonlinearity
- b_3 contains a term due to a_2
  - This is due to signal interaction with the second-order term fed back to the input
- It is possible to obtain $b_3 = 0$ without large loop gain

An Interesting Example



$$S_o = i_C = a_1 S_\varepsilon + a_2 S_\varepsilon^2 + a_3 S_\varepsilon^3 + L$$
$$S_\varepsilon = v_{BE} = v_i - f \cdot i_C$$

$$f = R_E \qquad a_1 = g_m \qquad a_2 = \frac{1}{2}\frac{I_C}{V_T^2} \qquad a_3 = \frac{1}{6}\frac{I_C}{V_T^3}$$

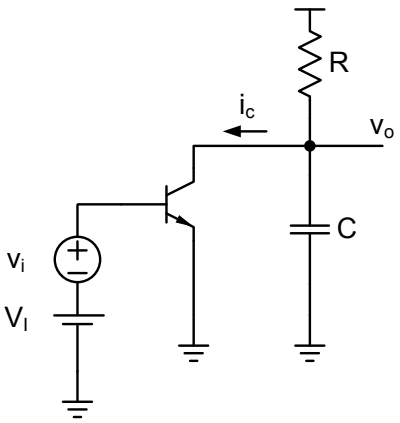
$$b_1 = \frac{a_1}{1 + a_1 f} = \frac{g_m}{1 + g_m R_E} \qquad b_2 = \frac{1}{2}\frac{I_C}{V_T^2}\frac{1}{(1 + g_m R_E)^3}$$

$$b_3 = \frac{\frac{1}{6}\frac{I_C}{V_T^3}(1 + g_m R_E) - \frac{1}{2}\frac{I_C}{V_T^3} g_m R_E}{(1 + g_m R_E)^5} \qquad g_m R_E = \frac{1}{2} \Rightarrow b_3 = 0 \Rightarrow HD_3 = 0$$

High Frequency Distortion Analysis

- The power series approach studied previously ignores any frequency dependence introduced by reactive elements
  - Sufficient for 90% of typical circuits, including some operating at RF
- Assuming weakly nonlinear behavior, the frequency dependence can be included using a Volterra Series model
  - Vito Volterra, 1887
- The purpose of this handout is to provide a few basic examples that will allow you to understand the general framework
- Examples
  - Memoryless nonlinearity followed by a filter
  - Memoryless nonlinearity preceded and followed by a filter
  - RC circuit with nonlinear capacitance
  - RC circuit with nonlinear resistance

Example 1



$$i_c = a_1 v_i + a_2 v_i^2 + a_3 v_i^3 + \ldots$$

$$a_1 = g_m \qquad a_2 = \frac{1}{2}\frac{I_{CQ}}{V_T^2} \qquad a_3 = \frac{1}{6}\frac{I_{CQ}}{V_T^3}$$

$$K(j\omega) = \frac{v_o}{i_c} = \frac{-R}{1 + j\omega RC}$$

- Ignoring device capacitances and finite output resistance for simplicity

Single-Tone Input

$$v_i = \hat{v}_i \cos(\omega t)$$

- Ignoring DC offset and gain expansion, we have

$$i_c = a_1 \hat{v}_i \cos(\omega t) + \frac{1}{2} a_2 \hat{v}_i^2 \cos(2\omega t) + \frac{1}{4} a_3 \hat{v}_i^3 \cos(3\omega t) + \ldots$$

- The output voltage consists of the same tones, with their magnitude and phase altered by the linear filter K(jω)

$$\begin{aligned} v_o = & \left|K(j\omega)\right| \cdot a_1 \hat{v}_i \cos(\omega t + \phi_\omega) \\ & + \left|K(2j\omega)\right| \cdot \frac{1}{2} a_2 \hat{v}_i^2 \cos(2\omega t + \phi_{2\omega}) \qquad \phi_{m\omega} = \angle K(m \cdot j\omega) \\ & + \left|K(3j\omega)\right| \cdot \frac{1}{4} a_3 \hat{v}_i^3 \cos(3\omega t + \phi_{3\omega}) + \ldots \end{aligned}$$

Two-Tone Input

$$v_i = \hat{v}_1 \cos(\omega_1 t) + \hat{v}_2 \cos(\omega_2 t)$$

- Substituting this input into the power series, and using the identities shown below, the complete expression for the collector current is most elegantly expressed as shown on the next slide

$$\cos(\alpha)\cos(\beta) = \frac{1}{2}\left[\cos(\alpha+\beta) + \cos(\alpha-\beta)\right]$$

$$\begin{aligned} \cos(\alpha)\cos(\beta)\cos(\gamma) = \frac{1}{4}\Big[& \cos(\alpha+\beta+\gamma) + \cos(\alpha+\beta-\gamma) \\ & + \cos(\alpha-\beta+\gamma) + \cos(\alpha-\beta-\gamma)\Big] \end{aligned}$$

Frequency Components

$$i_c = a_1\left[\hat{v}_1\cos(\omega_1 t)+\hat{v}_2\cos(\omega_2 t)\right] \longrightarrow \omega_1,\ \omega_2$$

$$+\frac{a_2}{2}\left[\hat{v}_1^2\cos\left([\omega_1\pm\omega_1]t\right)+\hat{v}_2^2\cos\left([\omega_2\pm\omega_2]t\right)\right. \longrightarrow 0,\ 2\omega_1,\ 2\omega_2$$

$$\left.+2\hat{v}_1\hat{v}_2\cos\left([\omega_1\pm\omega_2]t\right)\right] \longrightarrow \omega_1\text{-}\omega_2,\ \omega_1\text{+}\omega_2$$

$$+\frac{a_3}{4}\left[\hat{v}_1^3\cos\left([\omega_1\pm\omega_1\pm\omega_1]t\right)+\hat{v}_2^3\cos\left([\omega_2\pm\omega_2\pm\omega_2]t\right)\right. \longrightarrow \omega_1,\ \omega_2,\ 3\omega_1,\ 3\omega_2$$

$$+3\hat{v}_1^2\hat{v}_2\cos\left([\omega_1\pm\omega_2\pm\omega_2]t\right) \longrightarrow \omega_1,\ 2\omega_1\text{-}\omega_2,\ 2\omega_1\text{+}\omega_2$$

$$\left.+3\hat{v}_1\hat{v}_2^2\cos\left([\omega_1\pm\omega_1\pm\omega_2]t\right)\right]+\ldots \longrightarrow \omega_2,\ 2\omega_2\text{-}\omega_1,\ 2\omega_2\text{+}\omega_1$$

Filtered Output

$$v_o = a_1\left[\hat{v}_1\left|K(j\omega_1)\right|\cos\left(\omega_1 t+\phi_{\omega_1}\right)+\hat{v}_2\left|K(j\omega_2)\right|\cos\left(\omega_2 t+\phi_{\omega_2}\right)\right]$$

$$+\frac{a_2}{2}\left[\left|K(2j\omega_1)\right|\cdot\hat{v}_1^2\cos\left(2\omega_1 t+\phi_{2\omega_1}\right)+\hat{v}_1^2\left|K(0)\right|\right.$$

$$+\left|K(2j\omega_2)\right|\cdot\hat{v}_2^2\cos\left(2\omega_2 t+\phi_{2\omega_2}\right)+\hat{v}_2^2\left|K(0)\right|$$

$$+\left|K\left(j[\omega_1-\omega_2]\right)\right|\cdot 2\hat{v}_1\hat{v}_2\cos\left([\omega_1-\omega_2]t+\phi_{\omega_1-\omega_2}\right)$$

$$\left.+\left|K\left(j[\omega_1+\omega_2]\right)\right|\cdot 2\hat{v}_1\hat{v}_2\cos\left([\omega_1+\omega_2]t+\phi_{\omega_1+\omega_2}\right)\right]$$

$$+\frac{a_3}{4}[\ldots]$$

Short Hand Notation

$$v_o = \underbrace{a_1 K(j\omega_a)}_{H_1(j\omega_a)} \circ v_i + \underbrace{a_2 K(j\omega_a + j\omega_b)}_{H_2(j\omega_a + j\omega_b)} \circ v_i^2 + \underbrace{a_3 K(j\omega_a + j\omega_b + j\omega_c)}_{H_3(j\omega_a + j\omega_b + j\omega_c)} \circ v_i^3 + \ldots$$

- Operator "$\circ$" means
  - Multiply each frequency component in v_i^m by

$$\left|H_m(j\omega_a, j\omega_b, \ldots)\right|$$

and shift phase by

$$\angle H_m(j\omega_a, j\omega_b, \ldots)$$

- The arguments ω_a, ω_b, ω_c, … are auxiliary variables taking on all permutations of ω_1, $\pm\omega_2$, … $\pm\omega_m$

General Frequency Domain Volterra Series

$$\boxed{v_o = H_1(j\omega_a) \circ v_i + H_2(j\omega_a, j\omega_b) \circ v_i^2 + H_3(j\omega_a, j\omega_b, j\omega_c) \circ v_i^3 + \ldots}$$

- For the circuit example discussed previously, the coefficients are given as follows

$$H_1(j\omega_a) = \frac{-a_1 R}{1 + j\omega_a RC}$$

$$H_2(j\omega_a, j\omega_b) = \frac{-a_2 R}{1 + (j\omega_a + j\omega_b)RC}$$

$$H_3(j\omega_a, j\omega_b, j\omega_c) = \frac{-a_3 R}{1 + (j\omega_a + j\omega_b + j\omega_c)RC}$$

Distortion Metrics

	Taylor Series	Volterra Series
HD_2	$\frac{1}{2}\frac{a_2}{a_1}\hat{v}_i$	$\frac{1}{2}\frac{\lvert H_2(j\omega_1, j\omega_1)\rvert}{\lvert H_1(j\omega_1)\rvert}\hat{v}_i$
HD_3	$\frac{1}{4}\frac{a_3}{a_1}\hat{v}_i^2$	$\frac{1}{4}\frac{\lvert H_3(j\omega_1, j\omega_1, j\omega_1)\rvert}{\lvert H_1(j\omega_1)\rvert}\hat{v}_i^2$
IM_3	$\frac{3}{4}\frac{a_3}{a_1}\hat{v}_i^2$	$\frac{3}{4}\frac{\lvert H_3(j\omega_1, j\omega_1, -j\omega_2)\rvert}{\lvert H_1(j\omega_1)\rvert}\hat{v}_i^2$

- Volterra series model
  - Second and third order distortion still vary with square and cube of input amplitude, respectively
  - But, there is no fixed relationship between HD_2 and IM_2, and HD_3 and IM_3

Distortion Metrics for Example 1

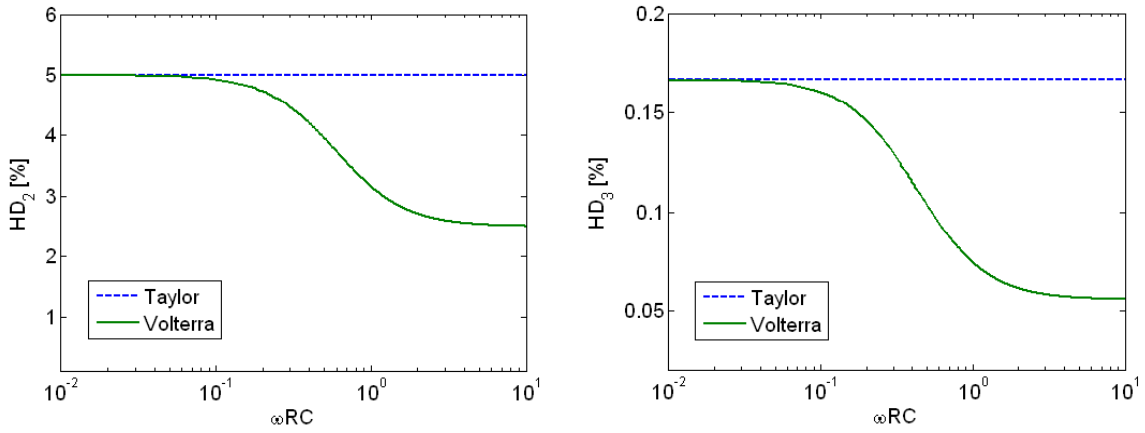
$$HD_2 = \frac{1}{2}\frac{\lvert H_2(j\omega, j\omega)\rvert}{\lvert H_1(j\omega)\rvert}\hat{v}_i = \frac{1}{2}\frac{a_2}{a_1}\frac{\left|\frac{1}{1+2j\omega RC}\right|}{\left|\frac{1}{1+j\omega RC}\right|}\hat{v}_i = \frac{1}{2}\frac{a_2}{a_1}\frac{\lvert 1+j\omega RC\rvert}{\lvert 1+2j\omega RC\rvert}\hat{v}_i$$

$$HD_3 = \frac{1}{4}\frac{\lvert H_3(j\omega, j\omega, j\omega)\rvert}{\lvert H_1(j\omega)\rvert}\hat{v}_i^2 = \frac{1}{4}\frac{a_3}{a_1}\frac{\lvert 1+j\omega RC\rvert}{\lvert 1+3j\omega RC\rvert}\hat{v}_i^2$$

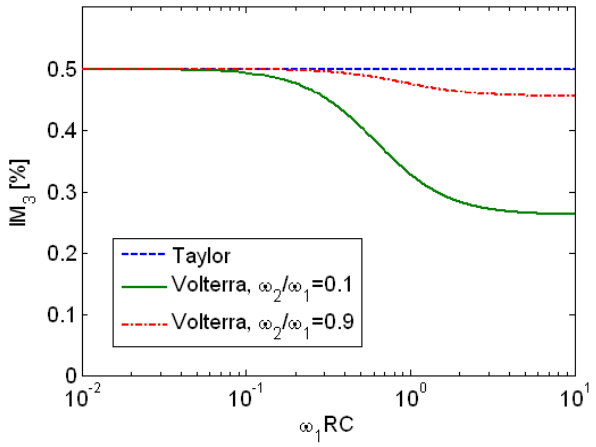
$$IM_3 = \frac{3}{4}\frac{\lvert H_3(j\omega_1, j\omega_1, -j\omega_2)\rvert}{\lvert H_1(j\omega_1)\rvert}\hat{v}_i^2 = \frac{1}{4}\frac{a_3}{a_1}\frac{\lvert 1+j\omega RC\rvert}{\lvert 1+j(2\omega_1-\omega_2)RC\rvert}\hat{v}_i^2$$

HD_2 and HD_3 Distortion Plots for Example 1

$$I_{CQ} = 1\text{mA} \qquad \hat{v}_i = \frac{V_T}{5}$$



IM_3 Distortion Plot for Example 1

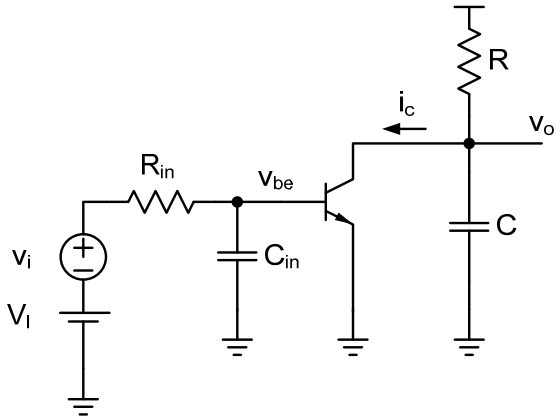


$$I_{CQ} = 1\text{mA}$$

$$\hat{v}_{i1} = \hat{v}_{i2} = \frac{V_T}{5}$$

- For $\omega_2 \to \omega_1$, the IM_3 distortion product is close to the fundamental tone ω_1, and therefore IM_3 becomes nearly frequency independent

Example 2



$$i_c = a_1 v_{be} + a_2 v_{be}^2 + a_3 v_{be}^3 + \ldots$$

$$K(j\omega) = \frac{v_o}{i_c} = \frac{-R}{1 + j\omega RC}$$

$$K_{in}(j\omega) = \frac{v_{be}}{v_i} = \frac{1}{1 + j\omega R_{in} C_{in}}$$

- Similar to example 1, but now including an additional filter at the input

Two-Tone Input

$$v_i = \hat{v}_1 \cos(\omega_1 t) + \hat{v}_2 \cos(\omega_2 t)$$

- The two input tones are now processed by a linear filter before being sent through the nonlinearity
- At the base of the BJT, we have

$$v_{be} = |K_{in}(j\omega_1)| \cdot \hat{v}_1 \cos(\omega_1 t + \psi_{\omega_1}) + |K_{in}(j\omega_2)| \cdot \hat{v}_2 \cos(\omega_2 t + \psi_{\omega_2})$$

$$\psi_{m\omega} = \angle K_{in}(m \cdot j\omega)$$

- In short hand notation, this can be written as

$$v_{be} = K_{in}(j\omega_a) \circ v_i$$

$$v_o = a_1 K(j\omega_a) \circ v_{be} + a_2 K(j\omega_a + j\omega_b) \circ v_{be}^2 + a_3 K(j\omega_a + j\omega_b + j\omega_c) \circ v_{be}^3 + \ldots$$

$$\begin{aligned} v_o = {} & a_1 K(j\omega_a) \circ \left[K_{in}(j\omega_a) \circ v_i \right] \\ & + a_2 K(j\omega_a + j\omega_b) \circ \left[K_{in}(j\omega_a) \circ v_i \right]^2 \\ & + a_3 K(j\omega_a + j\omega_b + j\omega_c) \circ \left[K_{in}(j\omega_a) \circ v_i \right]^3 \end{aligned}$$

- First order term

$$a_1 K(j\omega_a) \circ \left[K_{in}(j\omega_a) \circ v_i \right] = a_1 K(j\omega_a) K_{in}(j\omega_a) \circ v_i$$

- Second order term

$$a_2 K(j\omega_a + j\omega_b) \circ \left[K_{in}(j\omega_a) \circ v_i \right]^2 = ?$$

$$\begin{aligned} \left[K_{in}(j\omega_a) \circ v_i \right]^2 &= \left[K_{in}(j\omega_a) \circ \left\{ \hat{v}_1 \cos(\omega_1 t) + \hat{v}_2 \cos(\omega_2 t) \right\} \right]^2 \\ &= \left[\left| K_{in}(j\omega_1) \right| \cdot \hat{v}_1 \cos\left(\omega_1 t + \psi_{\omega_1} \right) + \left| K_{in}(j\omega_2) \right| \cdot \hat{v}_2 \cos\left(\omega_2 t + \psi_{\omega_2} \right) \right]^2 \\ &= \left| K_{in}(j\omega_1) \right|^2 \cdot \hat{v}_1^2 \cos\left(\{\omega_1 \pm \omega_1\} t + \psi_{\omega_1} \pm \psi_{\omega_1} \right) \\ &\quad + \left| K_{in}(j\omega_2) \right|^2 \cdot \hat{v}_2^2 \cos\left(\{\omega_2 \pm \omega_2\} t + \psi_{\omega_2} \pm \psi_{\omega_2} \right) \\ &\quad + \left| K_{in}(j\omega_1) \right| \left| K_{in}(j\omega_2) \right| \cdot \hat{v}_1 \hat{v}_2 \cos\left(\{\omega_1 \pm \omega_2\} t + \psi_{\omega_1} \pm \psi_{\omega_2} \right) \\ &= K_{in}(j\omega_a) K_{in}(j\omega_b) \circ v_i^2 \end{aligned}$$

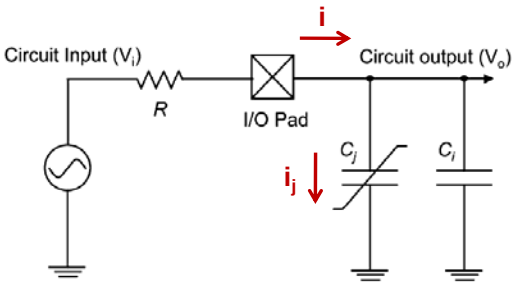
Coefficients

$$H_1(j\omega_a) = \frac{1}{1+j\omega_a RC}\frac{-a_1 R}{1+j\omega_a RC}$$

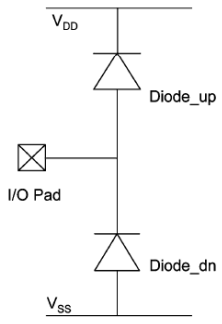
$$H_2(j\omega_a, j\omega_b) = \frac{1}{1+j\omega_a RC}\frac{1}{1+j\omega_b RC}\frac{-a_2 R}{1+(j\omega_a + j\omega_b)RC}$$

$$H_3(j\omega_a, j\omega_b, j\omega_c) = \frac{1}{1+j\omega_a RC}\frac{1}{1+j\omega_b RC}\frac{1}{1+j\omega_c RC}\frac{-a_3 R}{1+(j\omega_a + j\omega_b + j\omega_c)RC}$$

Example 3



[Chun & Murmann, JSSC 10/2006]



- C_j models the nonlinear capacitance of an electrostatic discharge (ESD) protection device (e.g. a basic diode structure as shown to the right)

Junction Capacitance Model

$$C_j = \frac{C_{j0}}{\left(1+\frac{V_{OQ}+v_o}{\psi_0}\right)^M} = \frac{C_{j0}}{\left(\frac{\psi_0+V_{OQ}}{\psi_0}\right)^M}\frac{1}{\left(\frac{\psi_0+V_{OQ}+v_o}{\psi_0+V_{OQ}}\right)^M}$$

- Using $C_{jQ} = \frac{C_{j0}}{\left(1+\frac{V_{OQ}}{\psi_0}\right)^M}$ $\qquad \frac{1}{(1+x)^M} = 1 - Mx + \frac{1}{2}\left(M+M^2\right)x^2 + \ldots$

$$V_R = V_{OQ} + \psi_0 \qquad b_1 = -\frac{M}{V_R} \qquad b_2 = \frac{M+M^2}{2V_R}$$

we can write $$C_j = \frac{C_{jQ}}{\left(1+\frac{v_o}{V_R}\right)^M} = C_{jQ}\left[1 + b_1 v_o + b_2 v_o^2 + \ldots\right]$$

Circuit Analysis

$$i_j = C_j\frac{dv_o}{dt} = C_{jQ}\left[1 + b_1 v_o + b_2 v_o^2 + \ldots\right]\frac{dv_o}{dt} = C_{jQ}\left[\frac{dv_o}{dt} + \frac{1}{2}b_1\frac{dv_o^2}{dt} + \frac{1}{3}b_3\frac{dv_o^3}{dt}\right]$$

$$i = \left(C_{jQ}+C_l\right)\frac{dv_o}{dt} + C_{jQ}\left[\frac{1}{2}b_1\frac{dv_o^2}{dt} + \frac{1}{3}b_3\frac{dv_o^3}{dt}\right]$$

$$v_i = v_o + i\cdot R = v_o + R\left(C_{jQ}+C_l\right)\frac{dv_o}{dt} + RC_{jQ}\left[\frac{1}{2}b_1\frac{dv_o^2}{dt} + \frac{1}{3}b_3\frac{dv_o^3}{dt}\right]$$

$$\boxed{v_i = v_o + RC_0\frac{d}{dt}v_o + RC_1\frac{d}{dt}v_o^2 + RC_2\frac{d}{dt}v_o^3}$$

$$C_0 = C_{jQ} + C_l \qquad C_1 = \frac{b_1}{2}C_{jQ} = -\frac{MC_{jQ}}{2V_R} \qquad C_2 = \frac{b_1}{3}C_{jQ} = -\frac{\left(M+M^2\right)C_{jQ}}{6V_R^2}$$

Volterra Series

- We are looking for a Volterra series representation of the form

$$v_o = H_1(j\omega_a) \circ v_i + H_2(j\omega_a, j\omega_b) \circ v_i^2 + H_3(j\omega_a, j\omega_b, j\omega_c) \circ v_i^3 + \ldots$$

- The coefficients H_1, H_2 and H_3 can be found by inserting the above series into the nonlinear differential equation shown on the previous slide and subsequently comparing the coefficients on the LHS and RHS of the equation

$$\begin{aligned} v_i = {} & H_1(j\omega_a) \circ v_i + H_2(j\omega_a, j\omega_b) \circ v_i^2 + H_3(j\omega_a, j\omega_b, j\omega_c) \circ v_i^3 \\ & + RC_0 \frac{d}{dt}\left[H_1(j\omega_a) \circ v_i + H_2(j\omega_a, j\omega_b) \circ v_i^2 + H_3(j\omega_a, j\omega_b, j\omega_c) \circ v_i^3 \right] \\ & + RC_1 \frac{d}{dt}\left[H_1(j\omega_a) \circ v_i + H_2(j\omega_a, j\omega_b) \circ v_i^2 + H_3(j\omega_a, j\omega_b, j\omega_c) \circ v_i^3 \right]^2 \\ & + RC_2 \frac{d}{dt}\left[H_1(j\omega_a) \circ v_i + H_2(j\omega_a, j\omega_b) \circ v_i^2 + H_3(j\omega_a, j\omega_b, j\omega_c) \circ v_i^3 \right]^3 \end{aligned}$$

Coefficient Comparison (1)

- First order

$$1 \circ v_i = \left[1 + RC_0 \frac{d}{dt} \right] H_1(j\omega_a) \circ v_i$$

$$1 = \left[1 + RC_0 \frac{d}{dt} \right] H_1(j\omega_a) = \left[1 + RC_0 j_{\omega_a} \right] H_1(j\omega_a)$$

$$\boxed{\therefore H_1(j\omega_a) = \frac{1}{1 + RC_0 j\omega_a}}$$

- This is just the linear transfer function as expected

Coefficient Comparison (2)

- Second order

$$0 \circ v_i^2 = \left[1 + RC_0 \frac{d}{dt}\right] H_2(j\omega_a, j\omega_b) \circ v_i^2 + RC_1 \frac{d}{dt}\left[H_1(j\omega_a) \circ v_i\right]^2$$

- Simplify using the following rules

$$\frac{d}{dt} H_2(j\omega_a, j\omega_b) = (j\omega_a + j\omega_b) H_2(j\omega_a, j\omega_b)$$

$$\left[H_1(j\omega_a) \circ v_i\right]^2 = H_1(j\omega_a) H_1(j\omega_b) \circ v_i^2$$

$$\frac{d}{dt}\left[H_1(j\omega_a) H_1(j\omega_b)\right] = (j\omega_a + j\omega_b) H_1(j\omega_a) H_1(j\omega_b)$$

Coefficient Comparison (3)

$$0 = \left[1 + RC_0(j\omega_a + j\omega_b)\right] H_2(j\omega_a, j\omega_b) + RC_1(j\omega_a + j\omega_b) H_1(j\omega_a) H_1(j\omega_b)$$

$$H_2(j\omega_a, j\omega_b) = \frac{-RC_1(j\omega_a + j\omega_b) H_1(j\omega_a) H_1(j\omega_b)}{1 + RC_0(j\omega_a + j\omega_b)}$$

$$\boxed{\therefore H_2(j\omega_a, j\omega_b) = -RC_1(j\omega_a + j\omega_b) H_1(j\omega_a) H_1(j\omega_b) H_1(j\omega_a + j\omega_b)}$$

- H_2 becomes zero for $\omega \to 0$; this makes intuitive sense

Coefficient Comparison (4)

- Third order

$$0 \circ v_i^3 = \left[1 + RC_0 \frac{d}{dt}\right] H_3(j\omega_a, j\omega_b, j\omega_c) \circ v_i^3$$
$$+ RC_1 \frac{d}{dt}\left[2\left(H_1(j\omega_a) \circ v_i\right)\left(H_2(j\omega_a, j\omega_b) \circ v_i^2\right)\right] \quad \left.\right\} \text{ Second-order Interaction Term}$$
$$+ RC_2 \frac{d}{dt}\left[H_1(j\omega_a) \circ v_i\right]^3$$

- Simplifying and neglecting the second-order interaction term yields

$$\therefore H_3(j\omega_a, j\omega_b, j\omega_c) = -RC_2(j\omega_a + j\omega_b + j\omega_c) H_1(j\omega_a) H_1(j\omega_b) H_1(j\omega_c) \cdot H_1(j\omega_a + j\omega_b + j\omega_c)$$

Harmonic Distortion Expressions (1)

$$HD_2 = \frac{1}{2}\frac{|H_2(j\omega, j\omega)|}{|H_1(j\omega)|}\hat{v}_i = \frac{1}{2}\frac{\left|RC_1(2j\omega)\left[H_1(j\omega)\right]^2 H_1(2j\omega)\right|}{|H_1(j\omega)|}\hat{v}_i = \omega RC_1 |H_1(j\omega)||H_1(2j\omega)|\hat{v}_i$$

$$\therefore HD_2 = \frac{1}{2} M\omega RC_{jQ} |H_1(j\omega)||H_1(2j\omega)|\left(\frac{\hat{v}_i}{V_R}\right)$$

- HD_2 improves for
  - Lower swing (v_i/V_R)
  - Lower frequency
  - Lower drive resistance (R)
  - Smaller capacitive nonlinearity (M→0) and smaller C_{jQ}
- Adding extra linear capacitance (C_l) will lower the corner frequency of H_1
  - Assuming that we cannot tolerate much signal attenuation, this won't help reduce the distortion all that much in the useable frequency range

Harmonic Distortion Expressions (2)

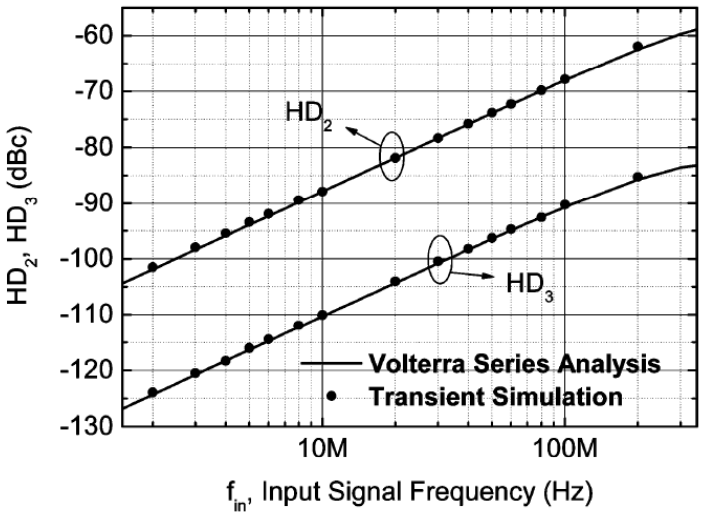
$$HD_3 = \frac{1}{4}\frac{\left|H_3(j\omega, j\omega, j\omega)\right|}{\left|H_1(j\omega)\right|}\hat{v}_i^2 = \frac{1}{4}\frac{\left|RC_2(3j\omega)\left(H_1(j\omega)\right)^3 H_1(3j\omega)\right|}{\left|H_1(j\omega)\right|}\hat{v}_i^2$$

$$= \frac{3}{4}\omega RC_2\left|H_1(j\omega)\right|^2\left|H_1(3j\omega)\right|\hat{v}_i^2$$

$$\therefore HD_3 = \frac{1}{8}\left(M + M^2\right)\omega RC_{jQ}\left|H_1(j\omega)\right|\left|H_1(3j\omega)\right|\left(\frac{\hat{v}_i}{V_R}\right)^2$$

- HD_3 behaves similar to H_{D2}
  - The distortion depends on similar quantities and cannot be improved by adding additional linear capacitance (C_l)
- HD_3 is often more critical than HD_2, since the latter can be improved significantly by employing a differential circuit topology

Plot



$\hat{v}_i = 0.5V$

$R = 25\Omega$

$C_l = 0$

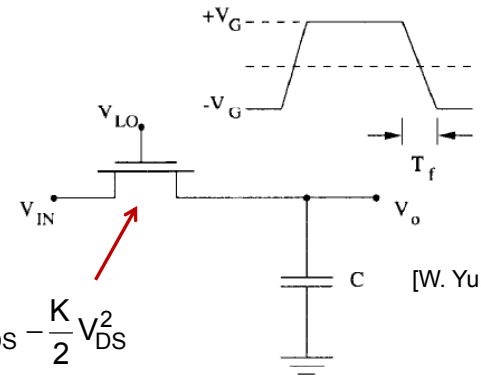
$C_{jQ} = 1pF$

$M = 0.3$

$V_R = V_{OQ} + \psi_0 = 2.2V$

Example 4

- An RC circuit with a linear capacitor, but nonlinear resistor
- The circuit below implements a track-and-hold circuit used in switched capacitor circuits (filters, A/D converters, etc.)
  - More in EE315A,B
- The MOSFET is used as a switch and operates in the triode region, exhibiting nonlinear resistive behavior



[W. Yu, IEEE TCAS II, 2/1999]

$$I_D \cong K\left(V_{GS} - V_t\right)V_{DS} - \frac{K}{2}V_{DS}^2$$

Analysis (1)

$$I_D \cong K\left(V_{GS} - V_t\right)V_{DS} - \frac{K}{2}V_{DS}^2$$

$$C\frac{dV_o}{dt} = K\left(\varphi - V_o - V_t\right)\left(V_i - V_o\right) - \frac{K}{2}\left(V_i - V_o\right)^2$$

- Solving for the coefficients of the Volterra series linking v_o and v_i, and evaluating HD_2 and HD_3 yields

$$HD_2 = \frac{1}{2}\omega RC\left(\frac{\hat{v}_i}{V_{GS} - V_t}\right)$$

$$HD_3 = \frac{1}{4}\omega RC\left(\frac{\hat{v}_i}{V_{GS} - V_t}\right)^2$$

where V_{GS} is the quiescent point gate-source voltage and R is the quiescent point resistance of the MOSFET, i.e. $K[V_{GS}\text{-}V_t]^{-1}$

Analysis (2)

- For low distortion
  - Make signal amplitude much smaller than V_{GS}-V_t
    - Low swing, which is usually undesired
  - Make 1/RC much larger than $2\pi \cdot f_{in}$
    - Small C or big MOSFET, which is usually undesired
- Example HD_3 calculation

$$\left.\begin{aligned} \omega &= \frac{1}{10}\frac{1}{RC} \\ \hat{v}_i &= 0.2V \\ V_{OQ} &= V_{IQ} = \frac{V_{DD}}{2} = 0.9V \\ V_{GS} - V_t &= V_{DD} - V_{OQ} - V_t \\ &= 1.8V - 0.9V - 0.4V = 0.45V \end{aligned}\right\} \quad HD_3 = \frac{1}{4}\frac{1}{10}\left(\frac{0.2}{0.45}\right)^2 = -46dB$$

Example 5

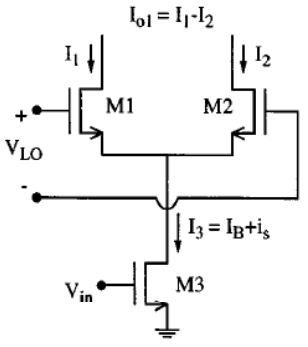


Fig. 2. A single-balanced active CMOS mixer.

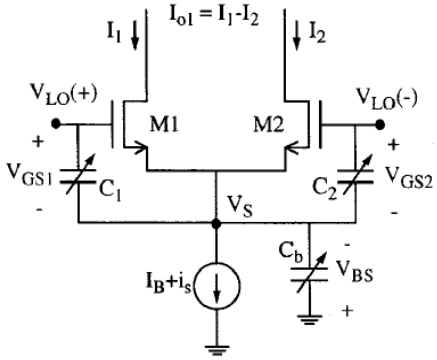


Fig. 11. Switching pair as considered in the high-frequency analysis.

[Terrovitis & Meyer, JSSC 10/2000]

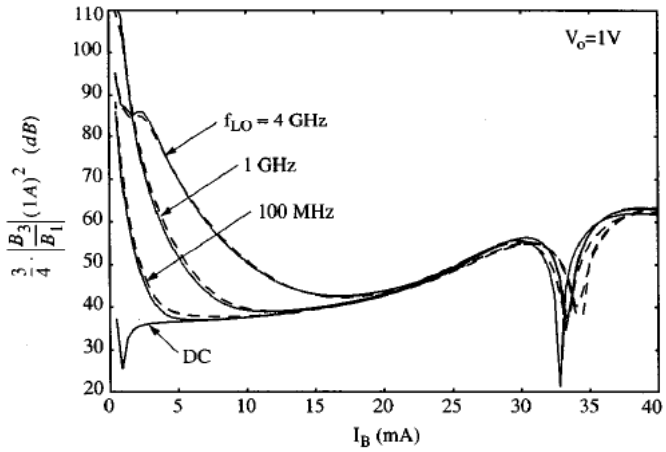


Fig. 12. High-frequency intermodulation prediction (solid line) and simulation with spectreRF (dashed line) for a switching pair operating as a downconverter, versus bias current and for several LO frequencies. The model corresponds to a 0.25 μm technology, the channel width is 100 μm, and the the LO amplitude is 1 V.

Additional Topics in Distortion Analysis

- Covered in EE314
  - RF-centric metrics
    - Intercept points
    - 1-dB gain compression point
- Other
  - Nonlinearities in passive components
  - Distortion cancelation techniques
  - Cascading nonlinearities
  - Series reversion
  - Distortion in clipped or amplitude limiting waveforms
    - E.g. in oscillators

References (1)

- D. O. Pederson and K. Mayaram, *Analog Integrated Circuits for Communication*, Springer, 1991
- P. Wambacq and W. M. C. Sansen, *Distortion Analysis of Analog Integrated Circuits*, Springer, 1998
- Hurst, Lewis, Gray & Meyer, 5th edition, pp. 355-359
  - Distortion analysis of a source follower
- B. Razavi, Design of Analog CMOS Integrated Circuits, McGraw-Hill, 2000, Chapter 13
- W. Sansen, “Distortion in elementary transistor circuits,” IEEE Trans. Circuits and Systems II, vol. 46, pp. 315-325, March 1999
- P. Wambacq, G. Gielen, and P. Kinget, “High-Frequency Distortion Analysis of Analog Integrated Circuits,” IEEE Trans. Circuits and Syst. II, pp. 335-345, March 1999

References (2)

- W. Rugh, *Nonlinear System Theory*, http://rfic.eecs.berkeley.edu/ee242/pdf/volterra_book.pdf
- R. G. Meyer and M. L. Stephens, "Distortion in variable-capacitance diodes," IEEE J. Solid-State Circuits, pp. 47-54, Feb. 1975
- K. L. Fong and R.G. Meyer, "High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages," IEEE J. Solid-State Circuits, pp.548-555, Apr. 1998
- M. T. Terrovitis and R. G. Meyer, "Intermodulation distortion in current-commutating CMOS mixers," IEEE J. Solid-State Circuits, pp.1461-1473, Oct. 2000
- J. Chun and B. Murmann, "Analysis and Measurement of Signal Distortion due to ESD Protection Circuits," IEEE J. Solid-State Circuits, pp. 2354-2358, Oct. 2006
- W. Yu, S. Sen and B. H. Leung, "Distortion analysis of MOS track-and-hold sampling mixers using time-varying Volterra series," IEEE Trans. Circuits and Syst. II, pp. 101-113, Feb. 1999

Chapter 10

Operational Amplifiers and Output Stages

B. Murmann

Stanford University

Reading Material: Sections 5.2, 5.5, 6.8, 7.4, 9.4.5

Philips Semiconductors — Product data

Matched quad high-performance low-voltage operational amplifier — NE/SA5234

DESCRIPTION

The NE/SA5234 is a matched, low voltage, high performance quad operational amplifier. Among its unique input and output characteristics is the capability for both input and output rail-to-rail operation, particularly critical in low voltage applications. The output swings to less than 50 mV of both rails across the entire power supply range. The NE/SA5234 is capable of delivering 5.5 V peak-to-peak across a 600 Ω load and will typically draw only 700 µA per amplifier. The bandwidth is 2.5 MHz and the 1% settling time is 1.4 µs.

PIN CONFIGURATION

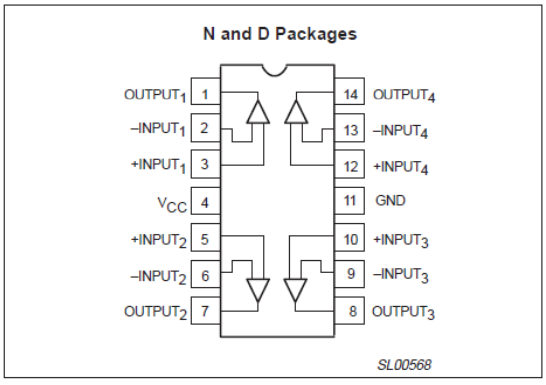


Figure 1. Pin configuration.

http://www.nxp.com/documents/data_sheet/NE5234_SA5234.pdf

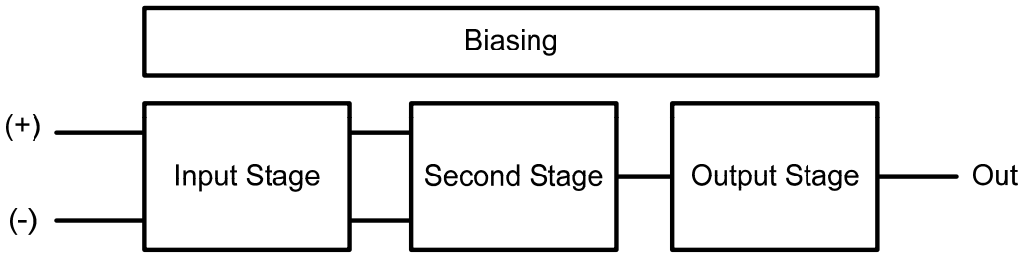
FEATURES

- Wide common-mode input voltage range: 250 mV beyond both rails
- Output swing within 50 mV of both rails
- Functionality to 1.8 V typical
- Low current consumption: 700 µA per amplifier
- ±15 mA output current capability
- Unity gain bandwidth: 2.5 MHz
- Slew rate: 0.8 V/µs
- Low noise: 25 nV/√Hz
- Electrostatic discharge protection
- Short-circuit protection
- Output inversion prevention

APPLICATIONS

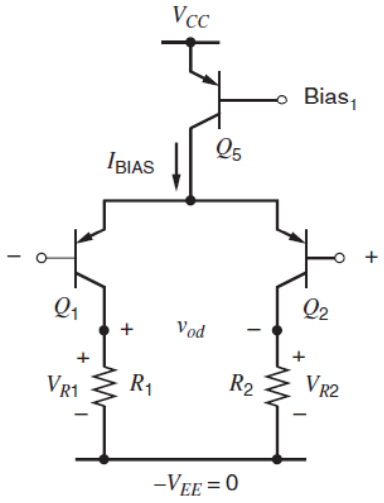
- Automotive electronics
- Signal conditioning and sensing amplification
- Portable instrumentation
  - Test and measurement
  - Medical monitors and diagnostics
  - Remote meters
- Audio equipment
- Security systems
- Communications
  - Pagers
  - Cellular telephone
  - LAN
  - 5 V Datacom bus
- Error amplifier in motor drives
- Transducer buffer amplifier

Architecture



- Provides differential input with large common mode range
- Provides some voltage gain, approximately independent of common mode
- Provides additional voltage gain
- Provides high current drive capability
- Output can (approximately) swing rail-to-rail

Basic PNP Differential Pair Input Stage



Text, p. 449

- Circuit works for input common mode down to V_{EE} (0V), as long as

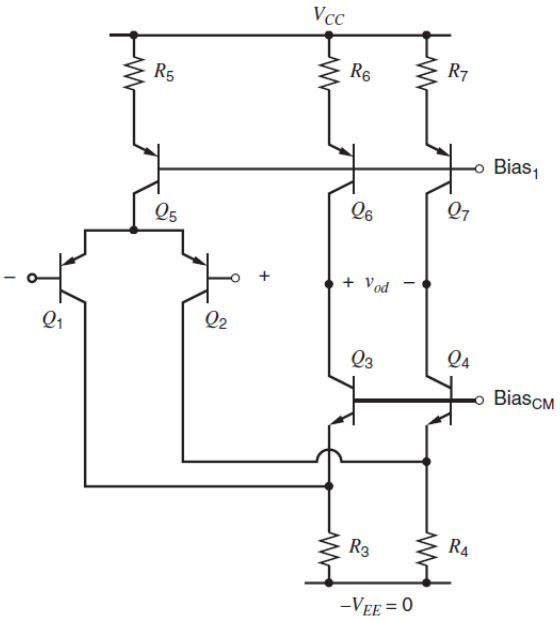
$$|V_{BE}| > |V_{CE(sat)}| + V_{R1}$$

$$V_{R1} < |V_{BE}| - |V_{CE(sat)}| \cong 0.8V - 0.2V = 0.6V$$

- This limits the gain we can extract from this stage

$$g_{m1}R_1 = \frac{I_{BIAS}}{2V_T}R_1 = \frac{V_{R1}}{V_T} = \frac{600mV}{26mV} = 23$$

PNP Folded Cascode Input Stage

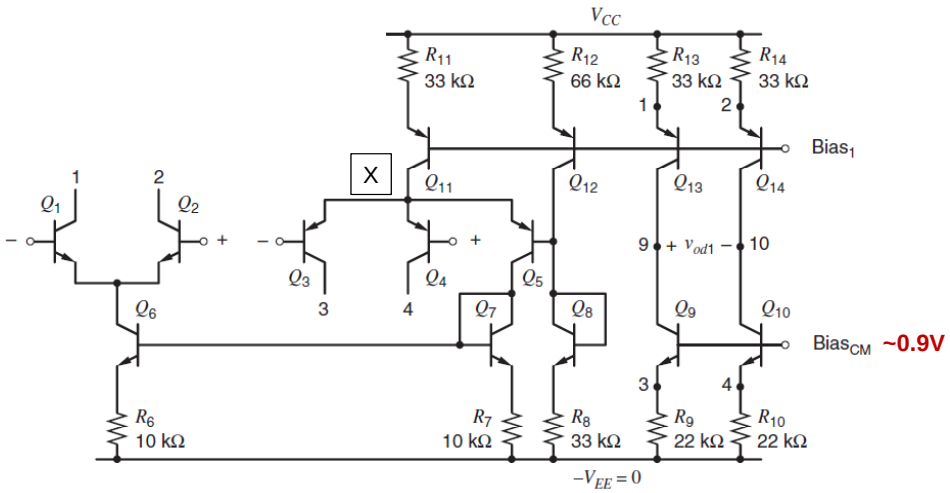


- Circuit still works for input common mode down to V_{EE} (0V), as before
- But, the low frequency voltage gain is much larger

$$\frac{v_{od}}{v_{id}} = g_{m1}\frac{R_3}{R_3 + \frac{1}{g_m}}R_o$$

$$R_o \cong r_{o6}\left(1 + g_{m6}R_6\right) \| r_{o3}\left(1 + g_{m3}R_3\right)$$

NE5234 Input Stage



- Provides rail-to rail common-mode input range by combining the currents from a PNP and NPN pair
- The NPN pair "steals" current from the PNP pair at node X whenever the input common mode is greater than $|V_{BE}|$
  - Total G_m is approximately constant

Second Stage

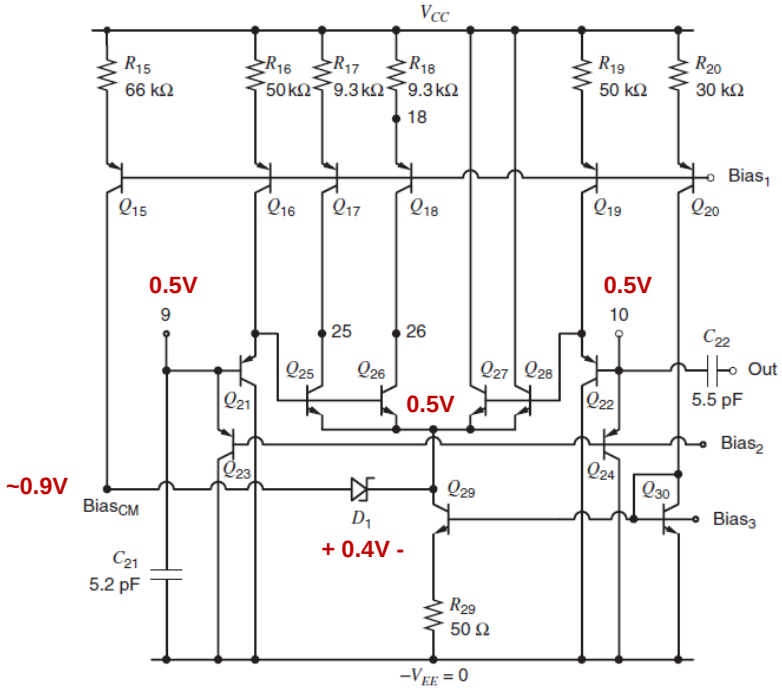


Figure 6.37 Schematic of the NE5234 second stage. Node Out is the output of the third stage (shown in Fig. 6.39).

Common Mode Operating Point of Stage 1 Output

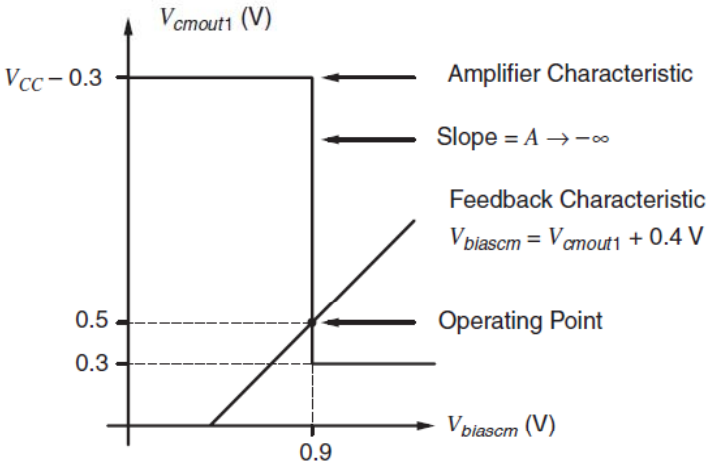


Figure 6.38 Plots of V_{cmout1} versus V_{biascm}. The amplifier characteristic comes from stage 1, and the feedback characteristic comes from stage 2.

Basic Emitter Follower ("Class-A") Output Stage

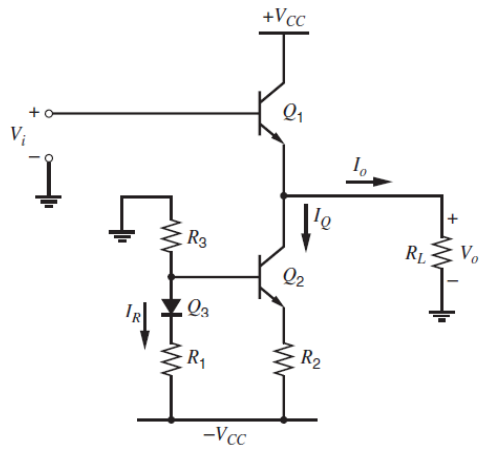
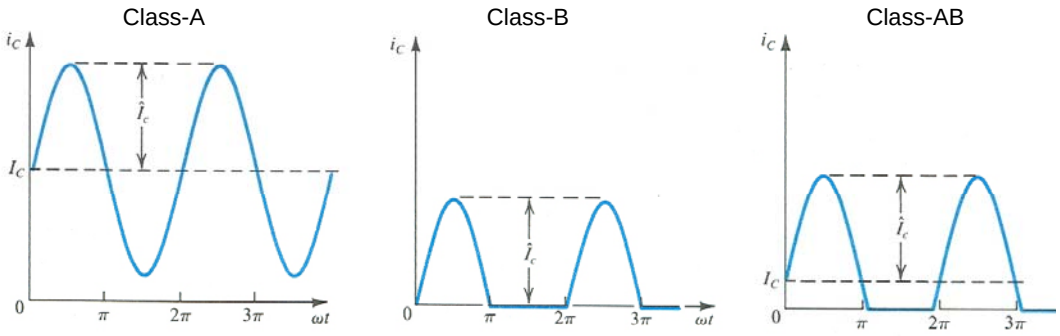


Figure 5.1 Emitter-follower output stage with current-mirror bias.

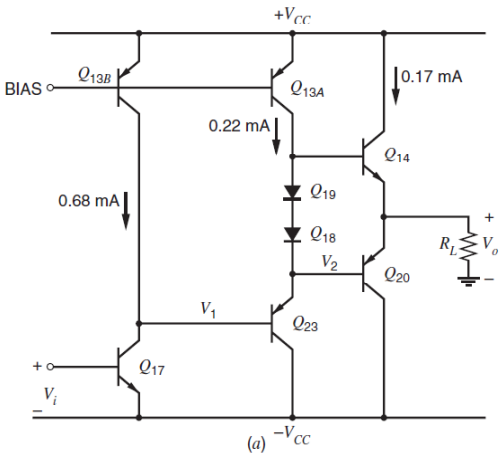
- Issues
  - Output voltage cannot go higher than $V_{CC} - V_{BE}$
  - Need large quiescent current I_Q to provide good current sinking capability
  - I_Q flows even if no signal is present, i.e. I_o=0

Output Stage Nomenclature

- Class-A
  - Output devices conduct for entire cycle of output sine wave
- Class-B
  - Output devices conduct for ≅50% of sine wave cycle
- Class-AB
  - Output devices conduct for >50%, but <100% of cycle



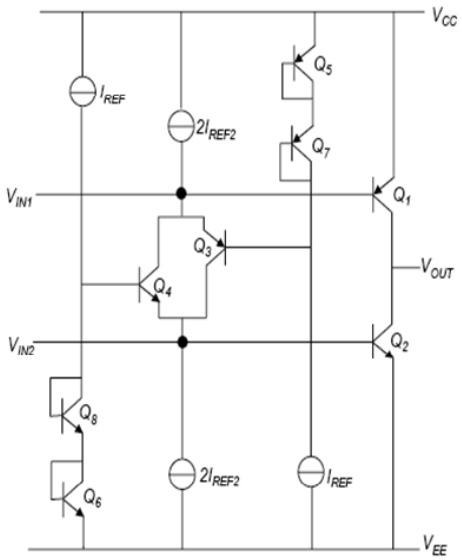
Class-AB Emitter Follower Output Stage



- Small quiescent current, but large drive capability
- Remaining issue
  - Output voltage swing is severely limited

Class-AB Common Emitter Output Stage

- Q1 and Q2 are large devices
- Want to set the quiescent point such that I_{C1} and I_{C2} are small when no signal is applied



Analysis

- Assuming $I_{C3} = I_{C4} = I_{REF}$ by symmetry, we can write

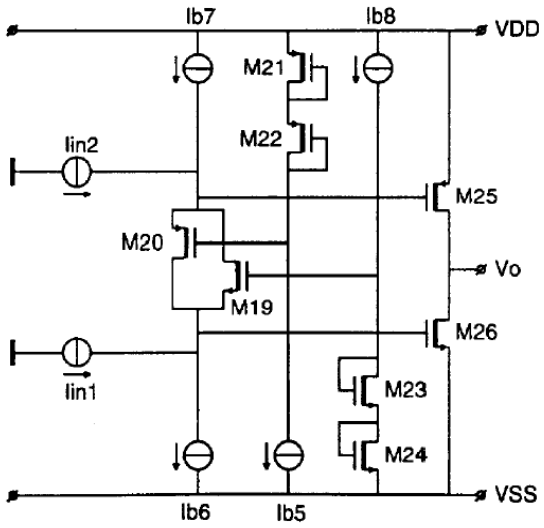
$$V_{BE6} + V_{BE8} = V_{BE4} + V_{BE2}$$

$$\frac{I_{REF}}{I_{s6}}\frac{I_{REF}}{I_{s8}} = \frac{I_{REF}}{I_{s4}}\frac{I_{C2}}{I_{s2}}$$

$$I_{C2} = I_{REF}\frac{I_{s2}}{I_{s6}}\frac{I_{s4}}{I_{s8}}$$

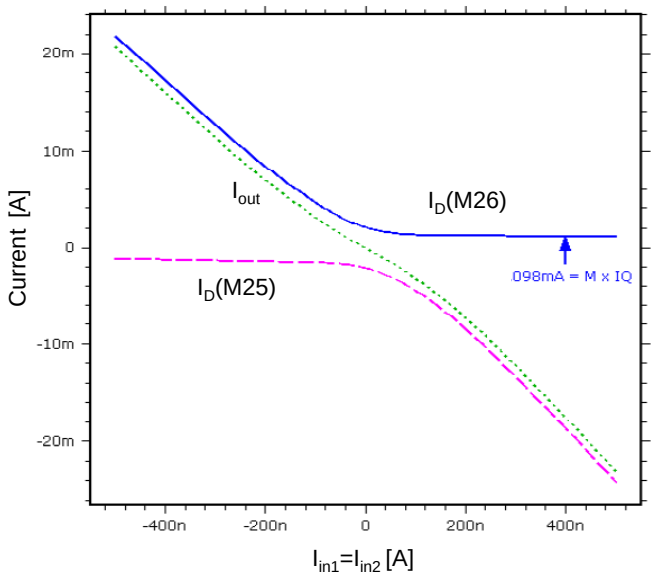
- Quiescent current is set by I_{REF} and emitter area ratios ($I_s \propto A_E$)
- The same principle applies to CMOS implementations
  - Original paper by Monticelli, JSSC 12/1986

CMOS Version



[Hogervorst, JSSC 12/1994]

Simulated Currents (V_o=0)



- Output transistors never turn off
- Quiescent current set by transistor ratios
- Large drive capability

NE 5234 Output Stage

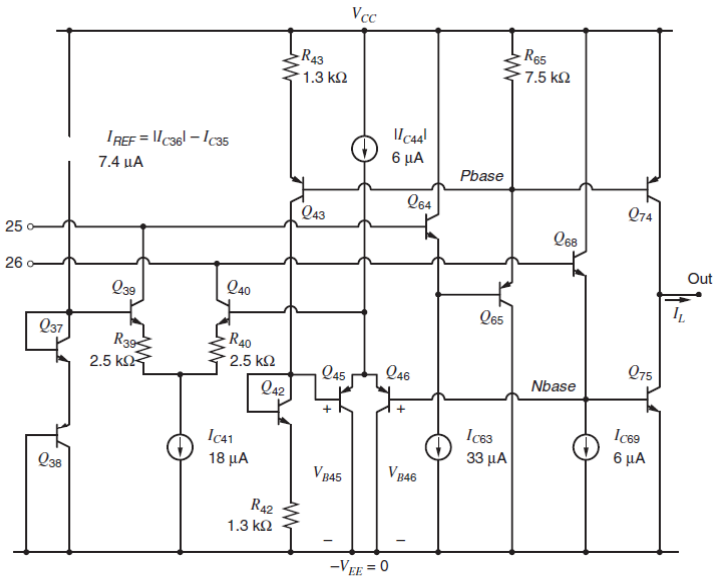


Figure 6.41 Simplified schematic of the NE5234 output stage with its bias circuit.

- Feedback through Q39-40 ensures that both transistors remain on at extreme output swings

Simplified AC Schematic of Complete OpAmp

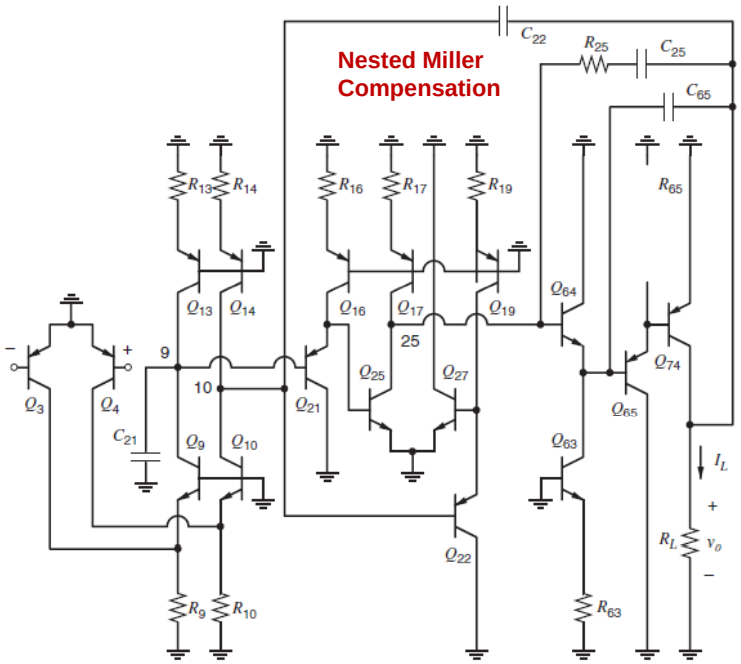


Figure 7.36 An ac schematic of the high-frequency gain path of the NE5234 assuming that V_{IC} is low enough that Q_1 and Q_2 in Fig. 6.36 are off.

Compensation Issue (1)

- The designer of a general purpose OpAmp knows nothing about the feedback network that the user will connect around the amplifier
- Therefore, general purpose OpAmps are typically compensated for the "worst case", i.e. unity feedback configuration
- This tends to be wasteful, since much less compensation may be needed for other feedback configurations
  - To overcome this issue, some general purpose OpAmps provide an external pin to let the user decide on the required compensation capacitor

Compensation Issue (2)

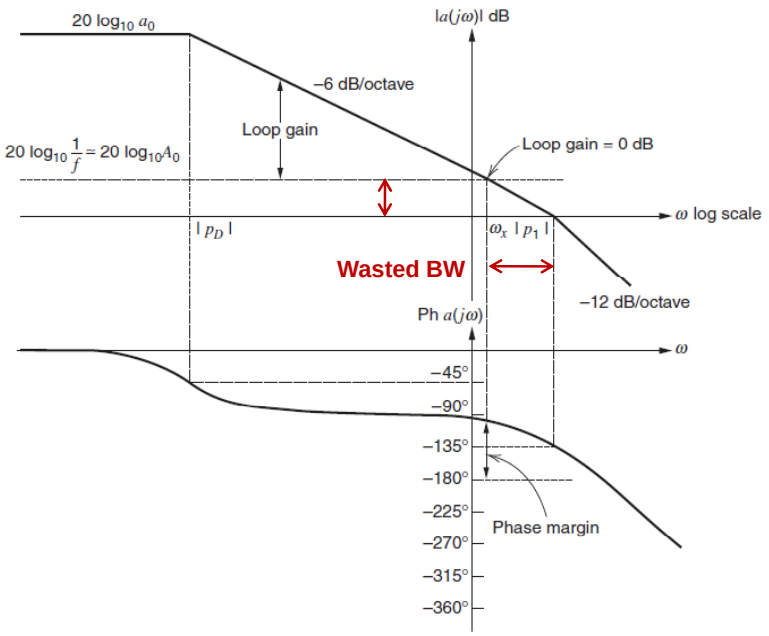


Figure 9.13 Gain and phase versus frequency for an amplifier compensated for use in a feedback loop with $f = 1$ and a phase margin of 45°. The phase margin is shown for operation in a feedback loop with $f < 1$.

Slewing

- Previous analyses have been concerned with the small-signal behavior of feedback amplifiers at high frequencies
- Sometimes the behavior with large input signals (either step inputs or sinusoidal signals) is also of interest
  - See e.g. switched capacitor circuits in EE315A
- Example using NE5234

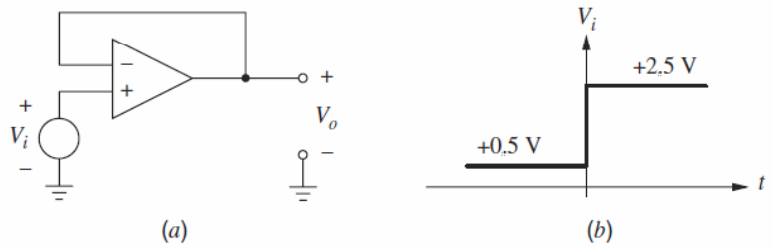


Figure 9.48 (*a*) Circuit and (*b*) input for testing slew-rate performance.

Output Response – Expected versus Actual

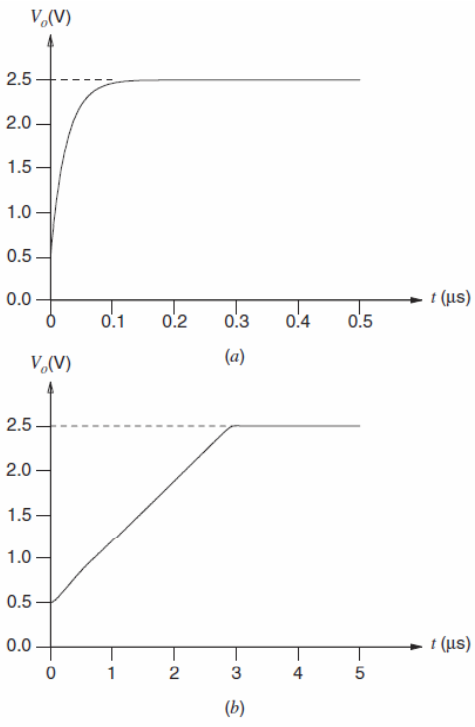


Figure 9.49 Response of the circuit of Fig. 9.48 when a 2-V step input is applied. (*a*) Response predicted by (9.125) for the NE5234 op amp. (*b*) Simulated response for the NE5234.

Investigation

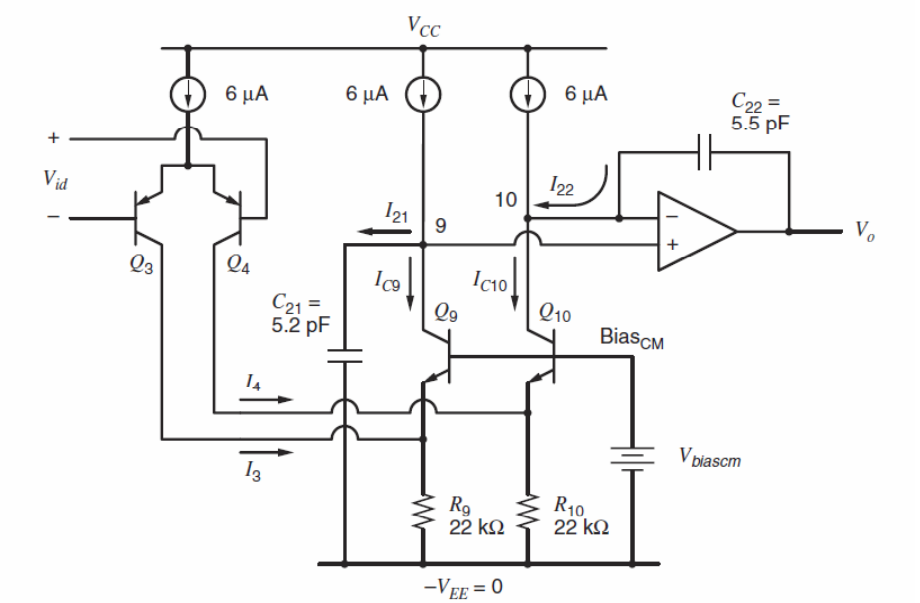


Figure 9.51 Simplified schematic of the NE5234 op amp.

General Analysis

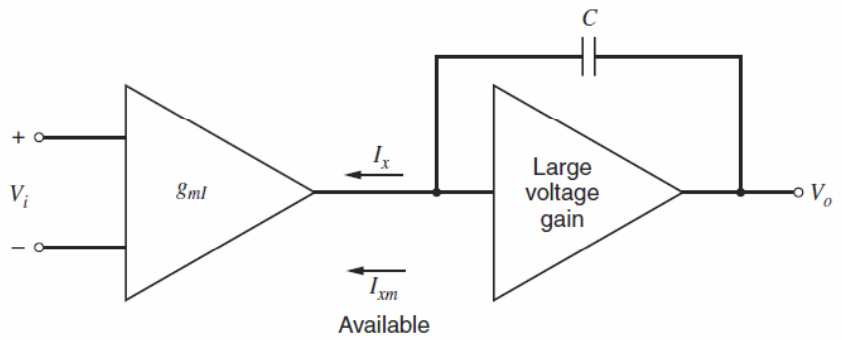
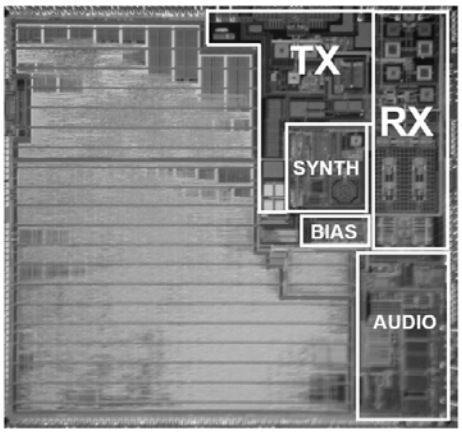


Figure 9.52 Generalized representation of an op amp for slew-rate calculations.

Slew Rate $$\frac{dV_o}{dt} = \frac{I_{xm}}{C}$$

"Internal" Amplifiers (1)

- Most amplifiers used in large systems-on-chip (SoC) are not as complex as a typical general purpose OpAmps
  - Typically, no output stage is needed, since the loads are either small and/or purely capacitive
- Operational Transconductance Amplifier (OTA) = OpAmp *minus* Output Stage
- More in EE315A (see also text, section 6.1.7, and chapter 12)



Mehta et al., "A 1.9GHz Single-Chip CMOS PHS Cellphone," ISSCC 2006.

"Internal" Amplifiers (2)

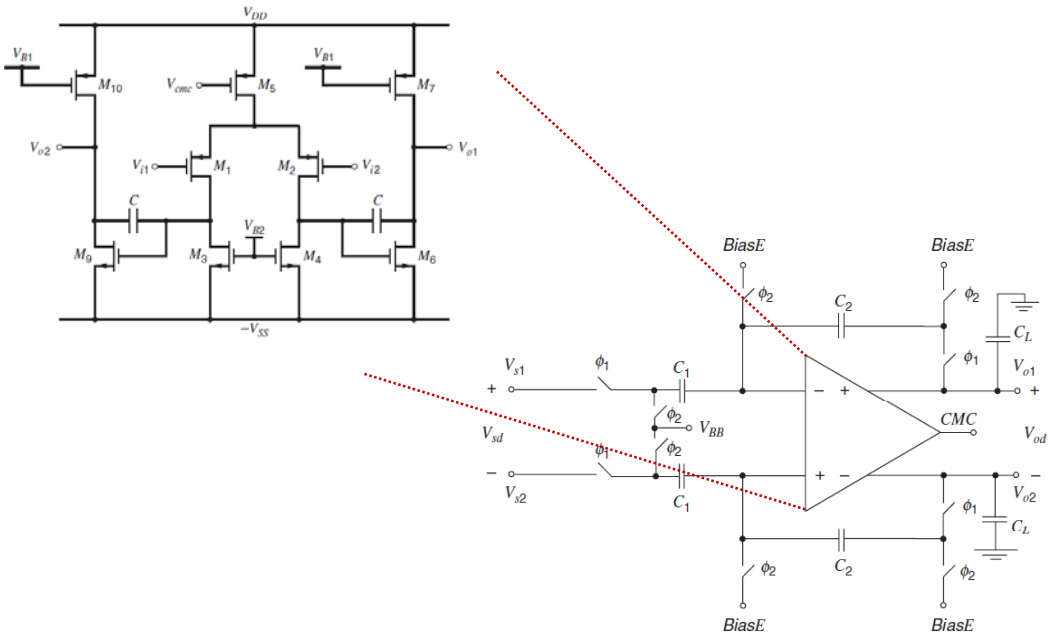


Figure 12.46 Switched-capacitor amplifier that uses the op amp in Figure 12.40. The CM sense circuitry is not shown.