EECS 242: MOS High Frequency Distortion

Professor Ali M Niknejad
Advanced Communication Integrated Circuits

University of California, Berkeley
MOS LNA Distortion

- Workhorse MOS LNA is a cascode with inductive degeneration
- Assume that the device is square law
- Neglect body effect (source tied to body)
Governing Differential Eq

\[ v_S = v_{R_s} + v_{L_g} + v_1 + v_{L_s} \]

\[ i_i = C_{gs} \frac{dv_1}{dt} \]

\[ v_{R_s} = R_s C_{gs} \frac{dv_1}{dt} \]

\[ v_{L_g} = L_g \frac{d}{dt} \left( C_{gs} \frac{dv_1}{dt} \right) = L_g C_{gs} \frac{d^2v_1}{dt^2} \]

\[ v_{L_s} = L_s \frac{d}{dt} \left( i_i + i_d \right) = L_s C_{gs} \frac{d^2v_1}{dt^2} + L_s \frac{d}{dt} \left( g_m v_1 + g_m v_1^2 + \cdots \right) \]

\[ v_S = R_s C_{gs} \frac{dv_1}{dt} + L_g C_{gs} \frac{d^2v_1}{dt^2} + v_1 + L_s C_{gs} \frac{d^2v_1}{dt^2} + L_s \frac{d}{dt} \left( g_m v_1 + g_m v_1^2 + \cdots \right) \]
Linear Analysis

- Let $v_1 = A_1(\omega_1) \circ v_s + A_2(\omega_1, \omega_2) \circ v_s^2 + A_3(\omega_1, \omega_2, \omega_3) \circ v_s^3 + \cdots$

- Equating linear terms

$$1 = C_{gs} R_s j \omega_1 A_1 - L_g C_{gs} \omega_1^2 + A_1 + L_s j \omega_1 g_{m1} A_1 - L_s C_{gs} \omega_1^2 A_1$$

$$A_1 = \frac{1}{1 - (L_g + L_s) C_{gs} \omega_1^2 + j \omega_1 (C_{gs} R_s + g_{m1} L_s)}$$

- By design, at resonance we have

$$A_1 = \frac{1}{1 - (L_g + L_s) C_{gs} \omega_0^2 + j \omega_0 C_{gs} (R_s + \omega_T L_s)} \frac{1}{j \omega_0 2 C_{gs} R_s} = -j Q_{net}$$
Relevant Time Constants

\[ \tau_1 = R_s C_{gs} \]

\[ \omega_0^2 = \frac{1}{(L_s + L_g)C_{gs}} \]

\[ \tau_2 = L_s g_m l = \frac{g_{m1}}{C_{gs}} C_{gs} L_s = \omega_T L_s C_{gs} = R_s C_{gs} = \tau_1 \]

\[ Q = \frac{1}{2R_s \omega_0 C_{gs}} = \frac{1}{2\tau_1 \omega_0} \]

\[ \tau_1 = \frac{1}{2Q \omega_0} \]

- Using these relations we simplify \( A_1 \) as

\[ A_1 = \frac{1}{1 - \left( \frac{\omega}{\omega_0} \right)^2 + j \frac{\omega}{\omega_0} \frac{1}{Q}} = \frac{1}{D(\omega)} \]
Second Order Terms

- Equate second-order terms

\[ 0 = C_{gs} R_s j(\omega_1 + \omega_2) A_2 - (L_g + L_s) C_{gs} (\omega_1 + \omega_2)^2 A_2 + A_2 + L_s j(\omega_1 + \omega_2) (g_{m1} A_2 + g_{m2} A_1(\omega_1) A_1(\omega_2)) \]

\[ A_2(\omega_1, \omega_2) = \frac{-A_1(\omega_1) A_1(\omega_2) j(\omega_1 + \omega_2) g_{m1} L_s}{1 - (\omega_1 + \omega_2)^2 (L_s + L_g) C_{gs} + j(\omega_1 + \omega_2) (C_{gs} R_s + L_s g_{m1})} \frac{1}{2(V_{gs} - V_t)} \]

\[ A_2(\omega_1, \omega_2) = \frac{-j(\omega_1 + \omega_2) \tau_1}{D(\omega_1) D(\omega_2) D(\omega_1 + \omega_2)} \frac{1}{2(V_{gs} - V_t)} \]
Third Order Terms

- If the MOS is truly square law, then there are no $A_3$ terms … but when we calculate the output current, we generate third order terms
- These are generated by the action of the feedback.

\[ i_d = g_{m1} \left( A_1(\omega_1) \circ v_s + A_2(\omega_1, \omega_2) \circ v_s^2 \right) + g_{m2} \left( A_1(\omega_1) \circ v_s + A_2(\omega_1, \omega_2) \circ v_s^2 \right)^3 \]

\[ i_d = B_1(\omega_1) \circ v_s + B_2(\omega_1, \omega_2) \circ v_s^2 + \cdots \]
Output Current Volterra Series

- Using Volterra algebra, we have

\[
B_1(\omega_1) = g_m A_1 = \frac{g_m}{D(\omega_1)}
\]

\[
B_2(\omega_1, \omega_2) = g_m A_2 + g_m A_1 A_1 = \left( \frac{-j(\omega_1 + \omega_2)\tau_1}{D(\omega_1)D(\omega_2)D(\omega_1 + \omega_2)} + \frac{1}{D(\omega_1)D(\omega_2)} \right) g_m
\]

\[
= \left( 1 - \frac{j(\omega_1 + \omega_2)\tau_1}{D(\omega_1 + \omega_2)} \right) \frac{g_m}{D(\omega_1)D(\omega_2)2(V_{gs} - V_t)}
\]

\[
B_3(\omega_1, \omega_2, \omega_3) = g_m^2 2A_1(\omega_1)A_2(\omega_1, \omega_2) = \left( \frac{-j(\omega_1 + \omega_2)}{D(\omega_1 + \omega_2)} + \frac{-j(\omega_1 + \omega_3)}{D(\omega_1 + \omega_3)} + \frac{-j(\omega_3 + \omega_2)}{D(\omega_3 + \omega_2)} \right) \frac{2g_m\tau_1}{2(V_{gs} - V_t)D(\omega_1)D(\omega_2)D(\omega_3)}
\]
The expression for IIP2 is now very simple! For a zero-IF system, we care about distortion at DC:

\[ V_{IIP2} = \frac{B_1(\omega_1)}{B_2(\omega_1, -\omega_2)} = \frac{D(-\omega_2)2(V_{gs} - V_t)}{D(-\omega_2)D(\omega_1 - \omega_2)} = \frac{D(-\omega_2)D(\omega_1 - \omega_2)2(V_{gs} - V_t)}{D(\omega_1 - \omega_2) - j(\omega_1 - \omega_2)\tau} \]
Compact Model Physics vs. Reality

- Even the most advanced compact models have very humble physical origins.
- Essentially a 1D core transistor model due to GCA (gradual channel approximation).
- Quantum effects necessarily ignored in core model.
- Small dimension and 2D effects treated in a perturbational manner as corrections to 1D core.
- Conclusion: Core model is important but any claims of “physical accuracy” should be taken with a grain of salt.
- Physical behavior and computational efficiency are therefore the key attributes.
- Need a simple design model to capture important effects!
Building on a strong foundation: BSIM4

Technology
- Short/Narrow Channel Effects on Threshold Voltage
- Non-Uniform Vertical and Lateral Doping Effects
- Mobility Reduction Due to Vertical Field
- Quantum Mechanic Effective Gate Oxide Thickness Model
- Carrier Velocity Saturation
- Channel Length Modulation (CLM)
- Substrate Current Induced Body Effect (SCBE)
- Unified current saturation model (velocity saturation, velocity overshoot, source end velocity limit)

Saturation
- Gate Dielectric Tunneling Current Model
- Gate Induced Drain Current Model (GIDL)
- Trap assisted tunneling and recombination current model

Leakage
- RF Model (Gate & Substrate Resistance Model)
- Unified Flicker Noise Model
- Holistic Thermal Noise Model

RF
- Asymmetric Layout-Dependent Parasitic Model
- Scalable Stress Effect Model
Important MOS Non-Linearity

- Square law $\rightarrow$ short channel effects
- Mobility degradation
  - Universal mobility curve
  - Velocity saturation
- Body effect
- Output impedance non-linearity
- $C_{gs}$ for high input swings
Velocity Saturation

- In triode, the average electric field across the channel is: \( E = \frac{V_{DS}}{L} \)
- The carrier velocity is proportional to the field for low field conditions \( v = \mu E = \mu \frac{V_{DS}}{L} \)
Mobility Reduction

- As we decrease the channel length from say 10 \( \mu \text{m} \) to 0.1 \( \mu \text{m} \) and keep \( V_{\text{DS}} = 1 \text{V} \), we see that the average electric field increases from \( 10^5 \text{ V/m} \) to \( 10^7 \text{ V/m} \). For large field strengths, the carrier velocity approaches the scattering-limited velocity.

- This behavior follows the following curve-fit approximation:

\[
\nu_d = \frac{\mu_n E}{1 + \frac{E}{E_c}}
\]

For \( E \ll E_c \), the linear behavior dominates

For \( E \gg E_c \), \( \nu_d \rightarrow \mu_n E_c = v_{\text{scl}} \)
Vertical Mobility Degradation

Experimentally, it is observed that $\mu$ is degraded due to the presence of a vertical field. A physical explanation for this is that a stronger $V_{GS}$ forces more of the carriers in the channel towards the surface where imperfections impede their movement.

$\mu_{eff} = \frac{\mu_n}{1 + \theta(V_{GS} - V_t)}$

$\theta \propto \frac{1}{T_{ox}}$

$T_{ox} = 100\text{Å}$, $\theta \approx 0.1V^{-1}$ to $0.4V^{-1}$
Unified Mobility Degradation

Model horizontal mobility degradation in saturation, modify low field equation to include velocity saturation.

Result looks like vertical field degradation.
Source Degeneration

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

\[ V_{GS} = V'_{GS} + I_D R_S \]

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

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

\[ = \frac{\mu C_{ox}}{2} \left( \frac{W}{L} \right) (V_{GS} - V_t)^2 \frac{1}{1 + \mu C_{ox} \left( \frac{W}{L} \right) R_S \left( V_{GS} - V_t \right) \theta} \]

- Equation has approximately the same form.
Simple Short Channel Model

- The short channel I-V model can be approximated by

\[ I_D = \frac{\mu C_{ox}}{2} \left( \frac{W}{L} \right) \frac{(V_{GS} - V_t)^2}{1 + \theta(V_{GS} - V_t)} \]

- \(\theta\) models:
  - Rsx parasitic
  - Vertical Field Mobility Degradation
  - Horizontal Field Mobility Degradation

\[ \theta = f(T_{ox}, L, R_{sx}) \]
Power Series Expansion

- Device is no longer square law...

\[ I_D = a_1 (V_{GS} - V_t) + a_2 (V_{GS} - V_t)^2 + a_3 (V_{GS} - V_t)^3 + ... \]

\[ a_1 = \frac{dI_D}{d(V_{GS} - V_t)} = \frac{\mu C_{ox}}{2} \left( \frac{W}{L} \right) \frac{(1 + \theta (V_{GS} - V_t))}{(1 + \theta (V_{GS} - V_t))^2} (V_{GS} - V_t) - \theta (V_{GS} - V_t) \]

\[ a_1 = \frac{\mu C_{ox}}{2} \left( \frac{W}{L} \right) (V_{GS} - V_t) \frac{2 + \theta (V_{GS} - V_t)}{(1 + \theta (V_{GS} - V_t))^2} \]

\[ a_2 = \frac{\mu C_{ox}}{2} \left( \frac{W}{L} \right) \frac{1}{(1 + \theta (V_{GS} - V_t))^3} \]

\[ a_3 = \frac{\mu C_{ox}}{2} \left( \frac{W}{L} \right) \frac{-\theta}{(1 + \theta (V_{GS} - V_t))^4} \]
Coulomb Scattering

- A more complete model should include the effects of low field mobility.
- At low fields, the low level of inversion exposes “Coulomb Scattering” sites (due to shielding effect of inversion layer)

Source: Physical Background of MOS Model 11 (Level 1101), R. van Langevelde, A.J. Scholten and D.B.M. Klaassen
Mobility with Coulomb Scattering

\[ \mu_{eff} = \frac{\mu_0}{1 + UA \cdot E_{eff} + UB \cdot E_{eff}^2 + UC \cdot \frac{Q_b}{Q_b + Q_{in}}} \]
Need for Single Equation I-V

- Our models up to now only include strong inversion. In many applications, we would like a model to capture the entire I-V range from weak inversion to moderate inversion to strong inversion
  - Hard switching transistor
  - Power amplifiers, mixers, VCOs
- Surface potential models do this in a natural way but are implicit equations and more appropriate for numerical techniques
Since we have good models for weak/strong inversion, but missing moderate inversion, we can “smooth” between these regions and hope we capture the region in between (BSIM 3/4 do this too)

$$I_{DS} = f(V_{GS} - V_T) = K \frac{X^2}{1 + \theta X}$$

$$X = 2\eta \frac{kT}{q} \ln \left(1 + e^{\frac{q(V_{GS} - V_T)}{2\eta kT}}\right)$$

As before $\theta$ models short-channel effects and $\eta$ models the weak-inversion slope.
Limiting Behavior

\[ I_{DS} = f(V_{GS} - V_T) = K \frac{X^2}{1 + \theta X} \]

- In strong inversion, the exponential dominates giving us square law
  \[ X = 2\eta \frac{kT}{q} \ln\left(1 + e^{\frac{q(V_{GS} - V_T)}{2\eta kT}}\right) \approx V_{GS} - V_T \]

- In weak inversion, we expand the \( \ln \) function
  \[ \ln(1 + x) \approx x \]
  \[ X \approx 2\eta \frac{kT}{q} e^{\frac{q(V_{GS} - V_T)}{2\eta kT}} \]

\[ I_{DS} = K \frac{X^2}{1 + \theta X} \approx KX^2 = K \left(2\eta \frac{kT}{q}\right)^2 e^{\frac{q(V_{GS} - V_T)}{\eta kT}} \]
I-V Derivatives

Use chain rule to differentiate the I-V relation:

\[ I_{DS} = f(V) = K \frac{X^2}{1 + \theta X} \]

\[ f_V = f_X \cdot X_V \quad f_{VV} = f_{XX} \cdot X_V^2 + f_X \cdot X_{VV} \]

\[ f_{VVV} = f_{XXX} \cdot X_V^3 + 3(f_{XX} \cdot X_V \cdot X_{VV}) + f_X \cdot X_{VVV} \]

\[ f_X = K \frac{X \cdot (2 + \theta X)}{(1 + \theta X)^2} \]

\[ f_{XX} = K \frac{2}{(1 + \theta X)^3} \]

\[ f_{XXX} = K \frac{-6\theta}{(1 + \theta X)^4} \]

\[ X_V = \frac{1}{1 + s^{-2}} \]

\[ X_{VV} = \frac{q}{2\eta kT} \cdot \frac{1}{s + s^{-1}} \]

\[ X_{VVV} = -\frac{q^2}{(2\eta kT)^2} \cdot \frac{(s - s^{-1})}{(s + s^{-1})} \]
Single Equation Distortion
IP3 “Sweet Spot”

- Notice “sweet spot” where $g_{m3} = 0$
- Notice the sign change in the third derivative. Can we exploit this?
Bias Point for High Linearity

- Assume $g_m$ nonlinearity dominates (current-mode operation and low output impedance)

$$i_{out} \approx g_m \times v_{gs} + g_m' \times v_{gs}^2 + g_m'' \times v_{gs}^3$$

- Sweet IIP3 point, $g_m'' \approx 0$, exists where the transistor transits from weak inversion (exponential law), moderate inversion (square law) to velocity saturation
Notice that we can use two parallel MOS devices, one biased in weak inversion (positive $g_{m3}$) and one in strong inversion (negative $g_{m3}$).

The composite transistor has zero $g_{m3}$ at bias point!

Source: A Low-Power Highly Linear Cascoded Multiple-Gated Transistor CMOS RF Amplifier With 10 dB IP3 Improvement, Tae Wook Kim et al., *IEEE MWCL*, Vol. 13, NO. 6, JUNE 2003
MGTR (Multi-Gated Transistor)

![MGTR Diagram]

- **Single Transistor**
- **Weak Inversion Bias**
- **Strong Inversion Bias**
- **Composite Transistor**

2nd derivative of gm vs. Vgs (V)

Vgs1 and Vgs2 represent the gate voltages for the composite transistor configuration.
Wideband Noise Cancellation

- Take advantage of amplifier topologies where the output thermal noise flows into the input (CG amplifier, shunt feedback amplifier, etc).
- Cancel thermal noise using a second feedforward path. Can we also cancel the distortion?
Noise Cancellation LNA

Motivated by [Bruccoleri, et al., ISSCC02]

Optimal choice subject to fewer design parameter variations
130nm LNA Prototype

- 130nm CMOS
- 1.5V, 12mA
- Employ only thin oxide transistors

Measured S-Parameters

Measured Noise Performance

Noise cancellation is clearly visible. This is also a “knob” for dynamic operation to save current.
Record linearity of +16 dBm for out of band blockers.

Linearity works over entire LNA band.
As we vary the bias of key transistor, we simulate the effects of process/temp variation. There is a 50 mV window where the performance is still acceptable.
LNA Distortion Analysis

- Let’s assume that the drain current is a power-series of the Vgs voltage:

\[ i_{ds} = g_m \times v_{gs} + \frac{g'_m}{2!} \times v_{gs}^2 + \frac{g''_m}{3!} \times v_{gs}^3 + \ldots \]

- The purpose of the “differential” input is to cancel the 2\textsuperscript{nd} order distortion of the first stage (to minimize 2\textsuperscript{nd} order interaction):

\[ i_{out} = i_{ds,n} + i_{ds,p} \]

\[ = \left( g_{m,N} \times v_{in} + \frac{g'_{m,N}}{2} \times v_{in}^2 + \frac{g''_{m,N}}{6} \times v_{in}^3 \right) - \left( g_{m,P} \times (-v_{in}) + \frac{g'_{m,P}}{2} \times (-v_{in})^2 + \frac{g''_{m,P}}{6} \times (-v_{in})^3 \right) \]

\[ = \left( g_{m,N} + g_{m,P} \right) \times v_{in} + \frac{\left( g'_{m,N} - g'_{m,P} \right)}{2} \times v_{in}^2 + \frac{\left( g''_{m,N} + g''_{m,P} \right)}{6} \times v_{in}^3 \]
Distortion Equivalent Circuit

- Assume $R_1/R_2$ and $RL$ are small so that $r_o$ non-linearity is ignored.

\[
V_x = A_1(s_1) \cdot V_s + A_2(s_1, s_2) \cdot V_s^2 + A_3(s_1, s_2, s_3) \cdot V_s^3 \\
V_1 = B_1(s_1) \cdot V_s + B_2(s_1, s_2) \cdot V_s^2 + B_3(s_1, s_2, s_3) \cdot V_s^3 \\
V_2 = C_1(s_1) \cdot V_s + C_2(s_1, s_2) \cdot V_s^2 + C_3(s_1, s_2, s_3) \cdot V_s^3
\]

\[
i_{m_1} + \frac{V_x - V_1}{r_{o1}} + i_{m_2} + \frac{V_x - V_2}{r_{o2}} + \frac{V_x}{Z_x(s)} = \frac{V_s - V_x}{Z_s(s)}
\]

\[
Z_1(s) = R_1 \parallel \frac{1}{sC_1} \\
Z_s(s) = R_s + \frac{1}{sC_s} \\
Z_x(s) = \frac{1}{sC_x}
\]
Drain Current Non-Linearity

- Assuming that the gates of the input transistors are grounded at RF:

\[ i_{m1} = -\left( g_{m1}(-V_x) + \frac{g'_{m1}}{2}(-V_x)^2 + \frac{g''_{m1}}{6}(-V_x)^3 \right) \]

\[ = g_{m1}V_x - \frac{g'_{m1}}{2}V_x^2 + \frac{g''_{m1}}{6}V_x^3 \]

\[ i_{m2} = g_{m2}V_x + \frac{g'_{m2}}{2}V_x^2 + \frac{g''_{m2}}{6}V_x^3 \]

- First-order Kernels are found from:

\[ g_{m1}A_1(s) + \frac{A_1(s) - B_1(s)}{r_{o1}} + g_{m2}A_1(s) + \frac{A_1(s) - C_1(s)}{r_{o2}} + \frac{A_1(s)}{Z_x(s)} = \frac{1 - A_1(s)}{Z_s(s)} \]

\[ g_{m1}A_1(s) + \frac{A_1(s) - B_1(s)}{r_{o1}} = \frac{B_1(s)}{Z_1(s)} + \frac{B_1(s) - C_1(s)}{Z_{12}(s)} \]

\[ g_{m2}A_1(s) + \frac{A_1(s) - C_1(s)}{r_{o2}} = \frac{C_1(s)}{Z_2(s)} + \frac{C_1(s) - B_1(s)}{Z_{12}(s)} \]
Simplified First-Order

- At the frequency of interest, \( Z_{12} \sim 0 \) and \( B_1 \sim C_1 \)

\[
A_1(s) = \frac{(Z_1(s) \parallel Z_2(s)) + (r_{o1} \parallel r_{o2})}{H(s)}
\]

\[
B_1(s) = \frac{Z_1(s) \parallel Z_2(s)}{(Z_1(s) \parallel Z_2(s) + (r_{o1} \parallel r_{o2}))} A_1(s)
\]

\[
C_1(s) = B_1(s)
\]

\[
H(s) = Z_s(s) \left( 1 + (g_{m1} + g_{m2})(r_{o1} \parallel r_{o2}) \right) + \left( (Z_1(s) \parallel Z_2(s)) + (r_{o1} \parallel r_{o2}) \right) \left( 1 + \frac{Z_s(s)}{Z_x(s)} \right)
\]
Second-Order Terms

- Retaining only 2\textsuperscript{nd} order terms in the KCL equations:

\[
g_{m1}A_2(s_1, s_2) - \frac{g'_{m1}}{2} A_1(s_1)A_1(s_2) + \frac{A_2(s_1, s_2) - B_2(s_1, s_2)}{r_{o1}} + \frac{A_2(s_1, s_2) - C_2(s_1, s_2)}{r_{o2}} + \frac{A_2(s_1, s_2)}{Z_x(s_1 + s_2)} = -\frac{A_2(s_1, s_2)}{Z_s(s_1 + s_2)}
\]

\[
g_{m1}A_2(s_1, s_2) - \frac{g'_{m1}}{2} A_1(s_1)A_1(s_2) + \frac{A_2(s_1, s_2) - B_2(s_1, s_2)}{r_{o1}} = \frac{B_2(s_1, s_2)}{Z_1(s_1 + s_2)} + \frac{B_2(s_1, s_2) - C_2(s_1, s_2)}{Z_{12}(s_1 + s_2)}
\]

\[
A_2(s_1, s_2) = \frac{1}{2}(-g'_{m1} + g'_{m2})(r_{o1} \parallel r_{o1})Z_s(s_1 + s_2)A_1(s_1)A_1(s_2) + \Delta A_2(s_1, s_2) \over H(s_1 + s_2) + \Delta H(s_1, s_2)
\]
Solving above equations we arrive at:

\[
A_2(s_1, s_2) = \frac{\frac{1}{2}(-g'_{m1} + g'_{m2})(r_{o1} \parallel r_{o1})Z_s(s_1 + s_2)A_1(s_1)A_1(s_2) + \Delta A_2(s_1, s_2)}{H(s_1 + s_2) + \Delta H(s_1, s_2)}
\]

\[
B_2(s_1, s_2) = -\frac{Z_1(s_1 + s_2)Z_2(s_1 + s_2)}{Z_x(s_1 + s_2)Z_s(s_1 + s_2)}\left(\frac{1}{2}(-g'_{m1} + g'_{m2})(r_{o1} \parallel r_{o1})Z_s(s_1 + s_2)A_1(s_1)A_1(s_2)\right) + \Delta B_2(s_1, s_2)
\]

\[
\Delta A_2(s_1, s_2) = \frac{1}{2}Z_{12}(s_1 + s_2)A_1(s_1)A_1(s_2)\frac{Z_s(s_1 + s_2)}{Z_1(s_1 + s_2) + Z_2(s_1 + s_2)} \times \left((g'_{m1} - g'_{m2})(r_{o1} \parallel r_{o2}) + \frac{g'_{m1}r_{o1}Z_2(s_1 + s_2) - g'_{m2}r_{o2}Z_1(s_1 + s_2)}{r_{o1} + r_{o2}}\right)
\]

\[
\Delta B_2(s_1, s_2) = -\frac{1}{2}Z_{12}(s_1 + s_2)A_1(s_1)A_1(s_2)\frac{Z_s(s_1 + s_2)}{Z_1(s_1 + s_2) + Z_2(s_1 + s_2)} \times \left(g'_{m1}r_{o1}\left(Z_2(s_1 + s_2) + r_{o2}\right)\left(1 + \frac{Z_s(s_1 + s_2)}{Z_x(s_1 + s_2)}\right) + Z_s(s_1 + s_2)\left(g'_{m2}r_{o2}(1 + g_{m1}r_{o1}) + g'_{m1}r_{o1}(1 + g_{m2}r_{o2})\right)\right)
\]

\[
\Delta H(s_1, s_2) = Z_{12}(s_1 + s_2)\frac{Z_s(s_1, s_2)}{Z_1(s_1, s_2) + Z_2(s_1, s_2)} \times \left(\frac{r_{o1} + Z_1(s_1 + s_2)}{Z_x(s_1 + s_2) || Z_s(s_1 + s_2)}\right) + \left((1 + g_{m1}r_{o1})\left(r_{o2} + Z_2(s_1 + s_2)\right) + (1 + g_{m2}r_{o2})\left(r_{o1} + Z_1(s_1 + s_2)\right)\right)
\]
Third-Order Terms

\[
g_{m1}A_3(s_1, s_2, s_3) + \frac{g''_{m1}}{6}A_1(s_1)A_2(s_2)A_1(s_3) - g'_{m1}A_1(s_1)A_2(s_2, s_3) + \frac{A_3(s_1, s_2, s_3) - B_3(s_1, s_2, s_3)}{r_{o1}}
\]

\[
+ g_{m2}A_3(s_1, s_2, s_3) + \frac{g''_{m2}}{6}A_1(s_1)A_2(s_2, s_3) + g'_{m2}A_1(s_1)A_2(s_2, s_3) + \frac{A_3(s_1, s_2, s_3) - C_3(s_1, s_2, s_3)}{r_{o2}}
\]

\[
= -\frac{A_3(s_1, s_2, s_3)}{Z_s(s_1 + s_2 + s_3)}
\]

\[
g_{m1}A_3(s_1, s_2, s_3) + \frac{g''_{m1}}{6}A_1(s_1)A_2(s_2, s_3) - g'_{m1}A_1(s_1)A_2(s_2, s_3) + \frac{A_3(s_1, s_2, s_3) - B_3(s_1, s_2, s_3)}{r_{o1}}
\]

\[
= \frac{B_3(s_1, s_2, s_3)}{Z_1(s_1 + s_2 + s_3)} + \frac{B_3(s_1, s_2, s_3) - C_3(s_1, s_2, s_3)}{Z_{12}(s_1 + s_2 + s_3)}
\]

\[
g_{m2}A_3(s_1, s_2, s_3) + \frac{g''_{m2}}{6}A_1(s_1)A_2(s_2, s_3) + g'_{m2}A_1(s_1)A_2(s_2, s_3) + \frac{A_3(s_1, s_2, s_3) - C_3(s_1, s_2, s_3)}{r_{o1}}
\]

\[
= \frac{C_3(s_1, s_2, s_3)}{Z_2(s_1 + s_2 + s_3)} + \frac{C_3(s_1, s_2, s_3) - B_3(s_1, s_2, s_3)}{Z_{12}(s_1 + s_2 + s_3)}
\]
Third-Order Kernel

- Assuming $Z_{12} \sim 0$ (at $s_1 + s_2 + s_3$):

\[
A_3(s_1, s_2, s_3) = \frac{-Z_s(r_{o1} \parallel r_{o2}) \left( - (g_{m1}^{'} + g_{m2}^{'})A_1(s_1)A_2(s_2, s_3) + \frac{1}{6}(g_{m1}^{''} + g_{m2}^{''})A_1(s_1)A_1(s_2)A_1(s_3) \right)}{H(s_1 + s_2 + s_3)}
\]

\[
B_3(s_1, s_2, s_3) = \frac{-Z_1(s_1 + s_2 + s_3)}{Z_x(s_1 + s_2 + s_3) \parallel Z_s(s_1 + s_2 + s_3)} \frac{A_3(s_1, s_2, s_3)}{A_3(s_1, s_2, s_3)}
\]
The simplified equations are summarized here:

\[
A_1(s) = \frac{Z_1(s) + r_{o1}}{H(s)}
\]

\[
A_2(s_1, s_2) = \frac{\frac{1}{2}g'_m r_{o1} Z_s(s_1 + s_2) A_1(s_1) A_1(s_2)}{H(s_1 + s_2)}
\]

\[
A_3(s_1, s_2, s_3) = \frac{-Z_s(s_1 + s_2 + s_3) r_{o1} \left( -g'_m A_1(s_1) A_2(s_2, s_3) + \frac{1}{6}g''_m A_1(s_1) A_1(s_2) A_1(s_3) \right)}{H(s_1 + s_2 + s_3)}
\]

\[
B_1(s) = \frac{Z_1(s) \times (1 + g_m r_{o1})}{Z_1(s) + r_{o1}} A_1(s)
\]

\[
B_2(s_1, s_2) = \frac{-Z_1(s_1 + s_2)}{Z_x(s_1 + s_2) \parallel Z_s(s_1 + s_2)} A_2(s_1, s_2)
\]

\[
B_3(s_1, s_2, s_3) = \frac{-Z_1(s_1 + s_2 + s_3)}{Z_x(s_1 + s_2 + s_3) \parallel Z_s(s_1 + s_2 + s_3)} A_3(s_1, s_2, s_3)
\]
Output Voltage

The output voltage is given by a new Volterra series. Assume for simplicity the following:

\[ V_{\text{out}} = \left( (g_{m3} \times V_1 + \frac{g'_{m3}}{2!} \times V_1^2 + \frac{g''_{m3}}{3!} \times V_1^3) \right) \times Z_L(s) \]

\[ + \left( (g_{m4} \times V_x + \frac{g'_{m4}}{2!} \times V_x^2 + \frac{g''_{m4}}{3!} \times V_x^3) \right) \times Z_L(s) \]

The fundamental and third-order output are therefore:

\[ V_{\text{out, fund}} = \left( (A_1(s) \circ V_s) \times g_{m4} + (B_1(s) \circ V_s) \times g_{m3} \right) \times Z_L(s) \]

\[ V_{\text{out, 3rd}} = \left( \left( (A_3(s_1, s_2, s_3) \circ V_s^3) \times g_{m4} + (B_3(s_1, s_2, s_3) \circ V_s^3) \times g_{m3} \right) \times Z_L(s) \]

\[ + \left( (A_1(s) \circ V_s)^3 \times \frac{g''_{m4}}{6} \right) \times Z_L(s) \]

\[ + \left( (A_1(s_1) A_2(s_2, s_3) \circ V_s^3) \times g'_{m4} + (B_1(s_1) B_2(s_2, s_3) \circ V_s^3) \times g'_{m3} \right) \times Z_L(s) \]
Focus on Third-Order Output

- At low frequencies:
  - $A_1/B_1 \sim R_{\text{in}}/R_1$
  - $A_2/B_2 \sim -R_s/R_1$
  - $A_3/B_3 \sim -R_s/R_1$

$V_{\text{out, 3rd}} = \left( \left( A_3(s_1, s_2, s_3) \circ V_s^3 \right) \times g_{m4} + \left( B_3(s_1, s_2, s_3) \circ V_s^3 \right) \times g_{m3} \right)$

$+ \left( \left( A_1(s) \circ V_s \right)^3 \times \frac{g_{m4}''}{6} + \left( B_1(s) \circ V_s \right)^3 \times \frac{g_{m3}''}{6} \right)$

$+ \left( \left( A_1(s_1)A_2(s_2, s_3) \circ V_s^3 \right) \times g_{m4}' + \left( B_1(s_1)B_2(s_2, s_3) \circ V_s^3 \right) \times g_{m3}' \right) \times Z_L(s)$

Cancels like thermal noise

New distortion gen at output. Cancel with MGTR.

Second-order interaction: Must use $g' = 0$
Two-Tone Spacing Dependence

- Because 2\textsuperscript{nd} order interaction is minimized by using a PMOS and NMOS in parallel, the capacitor $C_{12}$ plays an important role.
- When second order distortion is generated at low frequencies, $f_1 - f_2$, the capacitor $C_{12}$ has a high reactance and distortion cancellation does not take place.
- There is therefore a dependency to the two-tone spacing.
In RF systems, the supply ripple can non-linearity transfer noise modulation on the supply to the output.

This problem was recently analyzed by Jason Stauth: “Energy Efficient Wireless Transmitters: Polar and Direct-Digital Modulation Architectures,” Ph.D. Dissertation, U.C. Berkeley
Supply Noise Sources

- Supply ripple + thermal noise
- Digital current noise
- Noise current
- On/off chip magnetic coupling
- On/off chip electrostatic coupling
Multi-Port Memoryless Non-linearity

- The output voltage is a non-linear function of both the supply voltage and the input voltage. A two-variable Taylor series expansion can be used if the system is memory-less:

\[
S_{out}(S_{in}, S_{vdd}) = a_{10}S_{in} + a_{20}S_{in}^2 + a_{30}S_{in}^3 \cdots \\
+ a_{11}S_{in}S_{vdd} + a_{21}S_{in}^2S_{vdd} + \cdots \\
+ a_{01}S_{vdd} + a_{02}S_{vdd}^2 + a_{03}S_{vdd}^3 \cdots
\]
Supply Noise Sideband

- Assume the input is at RF and the supply noise is a tone. Then the output signal will contain a noise sideband given by:

\[ S_{in} = v_i \cos(\omega_0 t) \]
\[ S_{vdd} = v_s \cos(\omega_s t) \]
\[ v_{out}(\omega_0 \pm \omega_s) = \frac{1}{2} a_{11} v_i v_s \]

\[
\text{Sideband}(dBc) = dB \left( \frac{2a_{10}}{a_{11}} \cdot \frac{1}{v_s} \right) \\
\text{PSRR}(dBV) = dB \left( \frac{2a_{10}}{a_{11}} \right)
\]
Extending the concept of a Volterra Series to a two input-port system, we have

\[ v_{out}(t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} F_{mn}(v_1(t), v_2(t)) \]

\[ F_{mn}(v_1(t), v_2(t)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_{mn}(\tau_1, \ldots, \tau_{m+n}) v_1(t - \tau_1) \cdots v_1(t - \tau_m) v_2(t - \tau_{m+1}) \cdots v_2(t - \tau_{m+n}) d\tau_1 \cdots d\tau_{m+n} \]

\[ S_{out} = A_{10}(j\omega_a) \circ S_1 + A_{20}(j\omega_a, j\omega_b) \circ S_1^2 + A_{30}(j\omega_a, j\omega_b, j\omega_c) \circ S_1^3 + \cdots \]
\[ + A_{01}(j\omega_a) \circ S_2 + A_{02}(j\omega_a, j\omega_b) \circ S_2^2 + A_{03}(j\omega_a, j\omega_b, j\omega_c) \circ S_2^3 + \cdots \]
\[ + A_{11}(j\omega_a, j\omega_b) \circ S_1 S_2 + A_{21}(j\omega_a, j\omega_b, j\omega_c) \circ S_1^2 S_2 + A_{12}(j\omega_a, j\omega_b, j\omega_c) \circ S_1 S_2^2 + \cdots \]

\[ PSRR = dB \left| \frac{2A_{10}(j\omega_0)}{A_{11}(j\omega_0, j\omega_S)} \right| \]
Example

Several important terms:
- $g_m , g_o$ non-linearity is usual transconductance and output resistance terms
- $g_{mo}$ is the interaction between the input/output
- $C_j$ is the output voltage non-linear capacitance

$$i_d = g_{m1} v_{gs} + g_{m2} v_{gs}^2 + g_{m3} v_{gs}^3 + \cdots$$
$$- g_{mb1} v_{sb} - g_{mb2} v_{sb}^2 - g_{mb3} v_{sb}^3 + \cdots$$
$$+ g_{mo11} v_{ds} \cdot v_{gs} + g_{mo12} v_{ds} \cdot v_{gs}^2 + g_{mo21} v_{ds}^2 \cdot v_{gs} + \cdots$$
$$+ C_1 \frac{d}{dt} (v_{db}) + \frac{C_2}{2} \frac{d}{dt} (v_{db}^2) + \frac{C_3}{3} \frac{d}{dt} (v_{db}^3) + \cdots$$
First Order Terms

- **First-order transfer function:**

\[ A_{10}^1(j\omega_a) = -y_S(j\omega_a) \frac{gm_1}{K_0(j\omega_a)}, \text{ where} \quad (\text{RF Node Transfer}) \]

\[ K_0(j\omega_a) = (gm_1 + gmb_1 + y_1) \cdot (y_X + y_L) + y_S \cdot (y_X + y_1 + y_L). \]

\[ A_{01}^1(j\omega_a) = \frac{gm_1(y_X + y_L)}{K_0} \], \quad \text{(Supply Node Transfer)} \quad y_L = (R_L)^{-1} \]

\[ A_{10}^2(j\omega_a) = \frac{y_X(gm_1 + y_1 + gmb_1 + y_S)}{K_0}, \text{ and} \]

\[ A_{01}^2(j\omega_a) = \frac{y_X y_L}{K_0}. \]

\[ y_1(j\omega_a) = g_0 + j\omega_a C_1 \]

\[ y_X(j\omega_a) = (j\omega_a L_C)^{-1} \]

\[ y_S(j\omega_a) = (j\omega_a L_S)^{-1} \]
Mixing Product

- The most important term for now is the supply-noise mixing term:

\[
y_{out}(\omega_o \pm \omega_S) = A_{11}^1(j\omega_0, j\omega_S) \circ [Vi(\omega_0), Vs(\omega_S)],
\]

\[
A_{11}^1(j\omega_a, j\omega_b) = y_s \frac{gmo_{11}K_1 + 2y_2K_2 + 2gm_2K_3 - 2gmb_2K_4}{K_0},
\]

\[
K_1(j\omega_a, j\omega_b) = A_{01}^2(j\omega_b) \left[1 + A_{10}^1(j\omega_a) - 2A_{10}^2(j\omega_a)\right] - A_{01}^1(j\omega_b) \left[1 - A_{10}^2(j\omega_a)\right],
\]

\[
K_2(j\omega_a, j\omega_b) = A_{01}^2(j\omega_b) \left[A_{10}^1(j\omega_a) - A_{10}^2(j\omega_a)\right] + A_{01}^1(j\omega_b) \left[A_{10}^2(j\omega_a) - A_{10}^1(j\omega_a)\right],
\]

\[
K_3(j\omega_a, j\omega_b) = A_{01}^2(j\omega_b) \left[1 - A_{10}^1(j\omega_a)\right], \text{ and}
\]

\[
K_4(j\omega_a, j\omega_b) = A_{10}^2(j\omega_a) A_{01}^2(j\omega_b).
\]
PSSR Reduction

$PSRR = dB \left| \frac{g_{m1}}{g_{mo11}K_1 + 2y_2K_2 + 2g_{m2}K_3 - 2g_{mb2}K_4} \right|$ 

- Increase $g_{m1}$
- Reduce second order conductive non-linearity at drain ($g_{o2}$)
- Reduce the non-linear junction capacitance at drain
- Reduce cross-coupling term by shielding the device drain from supply noise (cascode)
Output Conductance Non-Linearity

- For short-channel devices, due to DIBL, the output has a strong influence on the drain current. A complete description of the drain current is therefore a function of $f(v_{ds}, v_{gs})$.

- This is especially true if the device is run close to triode region (large swing or equivalently high output impedance):

$$i_{ds}(v_{gs}, v_{ds}) = g_{m1}v_{gs} + g_{ds1}v_{ds} + g_{m2}v_{gs}^2 + g_{ds2}v_{ds}^2$$
$$+ x_{11}v_{gs}v_{ds} + g_{m3}v_{gs}^3 + g_{ds3}v_{ds}^3$$
$$+ x_{12}v_{gs}v_{ds}^2 + x_{21}v_{gs}^2v_{ds} + \ldots$$

$$g_{mk} = \frac{1}{k!} \frac{\partial^k I_{DS}}{\partial V_{GS}^k}; \quad g_{dk} = \frac{1}{k!} \frac{\partial^k I_{DS}}{\partial V_{DS}^k}; \quad x_{pq} = \frac{1}{p!q!} \frac{\partial^{p+q} I_{DS}}{\partial V_{GS}^p \partial V_{DS}^q}.$$
Total Distortion

- Including the output conductance non-linearity modifies the distortion as follows

\[ v_{ds} = c_1 v_{gs} + c_2 v_{gs}^2 + c_3 v_{gs}^3 + \ldots \]

\[
c_1 = -g_{m1} \cdot (R_{CS}||(1/g_{ds1}))
\]
\[
c_2 = -(g_{m2} + g_{ds2}c_1^2 + x_{11}c_1) \cdot (R_{CS}||(1/g_{ds1}))
\]
\[
c_3 = -(g_{m3} + g_{ds3}c_1^3 + 2g_{ds2}c_1c_2 + x_{11}c_2 + x_{12}c_1^2 + x_{21}c_1) \cdot (R_{CS}||(1/g_{ds1})).
\]

Example IIP Simulation

Contributions to $c_2$ are shown above.

For low bias, $g_{ds2}$ contributes very little but $x_{11}$ and $g_{m2}$ are significant. They also have opposite sign.
PA Power Supply Modulation

- When we apply a 1-tone to a class AB PA, the current drawn from the supply is constant.
- For when we apply 2-tones, there is a low-frequency component to the input:

$$V_{in} = A \sin(\omega_1 t) + A \sin(\omega_2 t)$$

$$= 2A \cos\left(\frac{\omega_1 - \omega_2}{2}\right) t \sin\left(\frac{\omega_1 + \omega_2}{2}\right)$$

$$= 2A \cos(\omega_c t) \sin(\omega_c t)$$

- This causes a low frequency current to be drawn from the supply as well, even for a differential circuit.

The supply current is a full-wave rectified sine.
Fourier Components of $i_s$

- Substitute the Fourier series for the “rectified” sine and cosine.

- Note that an on-chip bypass can usually absorb the higher frequencies ($2f_c$) but not the low frequency beat ($f_s$ and harmonics)

\[
i_s = k \left( \frac{2}{\pi} + \frac{4 \cos(2\omega_m t)}{\pi^3} - \ldots \right) \times \left( \frac{2}{\pi} - \frac{4 \cos(2\omega_c t)}{\pi^3} - \ldots \right) = k \left( \ldots - \frac{8 \cos(2\omega_c t)}{\pi^2} \right) + \frac{8 \cos(2\omega_m t)}{\pi^2} - \ldots \right) \]

\[
= k \left( \ldots - \frac{8 \cos(2\omega_c t)}{\pi^2} \right) + \frac{8 \cos(\omega_s t)}{\pi^2} - \ldots \right) \]
Supply Ripple Voltage

\[ V_{dd} = V_{DD} + A_2 \cdot \cos(\omega_s t) + \text{higher harmonics of } \omega_s. \]

- The finite impedance of the supply means that the supply ripple has the following form. Assuming a multi-port Volterra description for the transistor results in:

\[
S_o = F_1(\omega_a) \circ S_1 + F_2(\omega_a, \omega_b) \circ S_1^2
+ F_3(\omega_a, \omega_b, \omega_c) \circ S_1^3 + \ldots
+ G_1(\omega_a) \circ S_2 + G_2(\omega_a, \omega_b) \circ S_2^2
+ G_3(\omega_a, \omega_b, \omega_c) \circ S_2^3 + \ldots
+ H_{11}(\omega_a, \omega_b) \circ (S_1 \cdot S_2)
+ H_{12}(\omega_a, \omega_b, \omega_c) \circ S_1 S_2^2
+ H_{21}(\omega_a, \omega_b, \omega_c) \circ S_1^2 S_2 + \ldots
\]

\[ S(\omega_1 \pm \omega_s) = H_{11} \circ S_1 \cdot S_2 \]
Experimental Results

Even though the PA is fully balanced, the supply inductance impacts the linearity.

Measurements confirm the source of the IM3 at low offsets arising from supply modulation.

Fig. 14. Degradation in IM3 with increased supply inductance.

Fig. 15. Measured supply voltage ripple in a two-tone test with 100 MHz tone spacing.

Fig. 16. Measured IM3 degradation and supply voltage ripple in a two-tone test with different tone spacings.
C\textsubscript{gs}, C\textsubscript{u}, and C\textsubscript{db} all contribute to then non-linearity.

As expected, the contribution is frequency dependent and very much a strong function of the swing (drain, gate).

Gate cap is particular non-linear.
PMOS Compensation Technique

- Make an overall flat CV curve by adding an appropriately sized PMOS device.

General References