Chapter 3
Nonlinearity and Distortion in Electronic Devices

Nonlinearity and Distortion

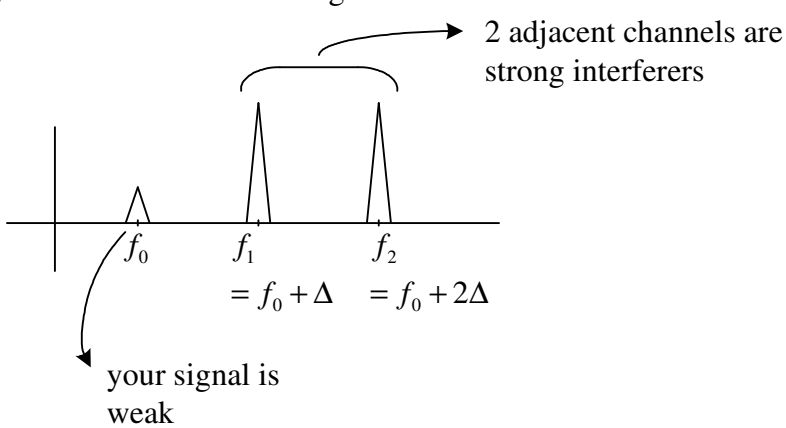
In this chapter we discuss the importance of nonlinearity in devices and how they are modeled. We take a simplified approach to understand nonlinearity. The approach helps us understand the mechanisms responsible for device nonlinear nature.

Note that in this chapter the nonlinearity is attributed to DC characteristics of the transistor. Therefore, the source of nonlinearity discussed here is the transconductance of the transistor. Transistors' input and output capacitances can also be very nonlinear. However, that type of nonlinearity is not captured by power series method used here. One needs to rely on Volterra Series, harmonic balance or time domain techniques to understand those nonlinearities.

Example of the effect of device nonlinearity on circuit performance

In the following example we assume that the RF transistor used as the input stage of a receiver (as the active circuit of a low noise amplifier (LNA)) is subjected to three signals coming from three adjacent channels:

your receiver receives 3 signals

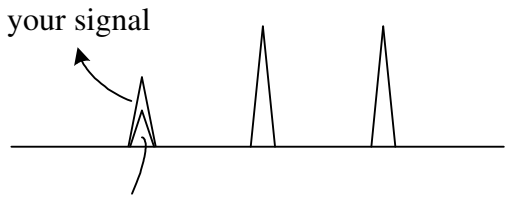


in your LNA you have nonlinearity so your strong interferers mix with each other due to LNA nonlinearity

This mixing generates frequencies such as

$$2f_1 - f_2 = 2(f_0 + \Delta) - (f_0 + 2\Delta) = f_0$$

This is what you get at the output of LNA



This distortion makes your reception difficult

→ we have to understand distortion and nonlinearity

Non-linearity
- **Weakly non-linear systems** (such as in receivers)
  LNA
- **Strongly non-linear systems** (such as in high efficiency transmitter)
  also clamping
  class B-E PA's

to deal with weakly non-linear system there are two

approaches
- **power series** → represent transfer function using a power series
- **volterra series** → power series with information about phase

for strongly non linear system → use **envelope analysis**

instead of frequency domain analysis

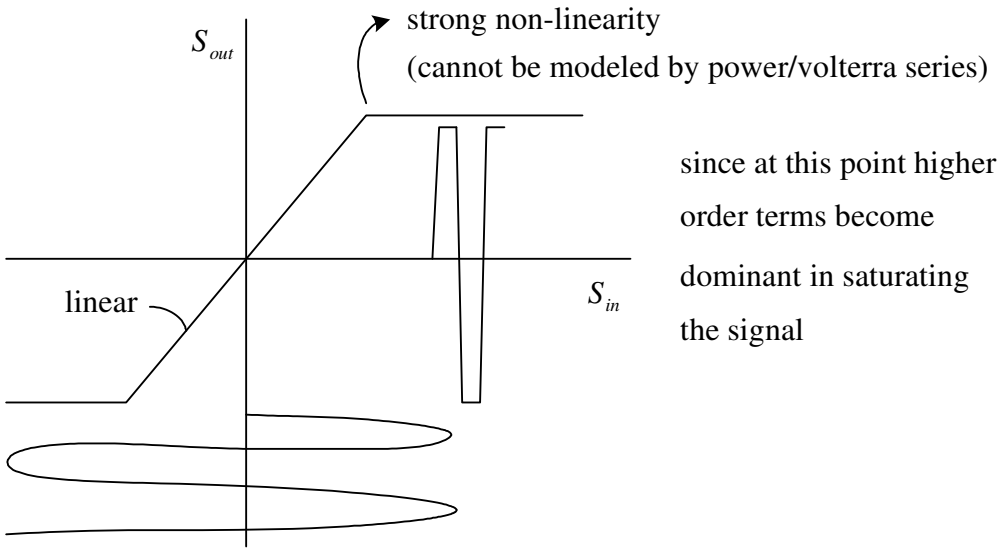
use time domain analysis

* At midband frequency

parasitic caps and inductors
bias and bypass caps
can be neglected in terms of the phase they introduce → use power series

* At low or high frequencies

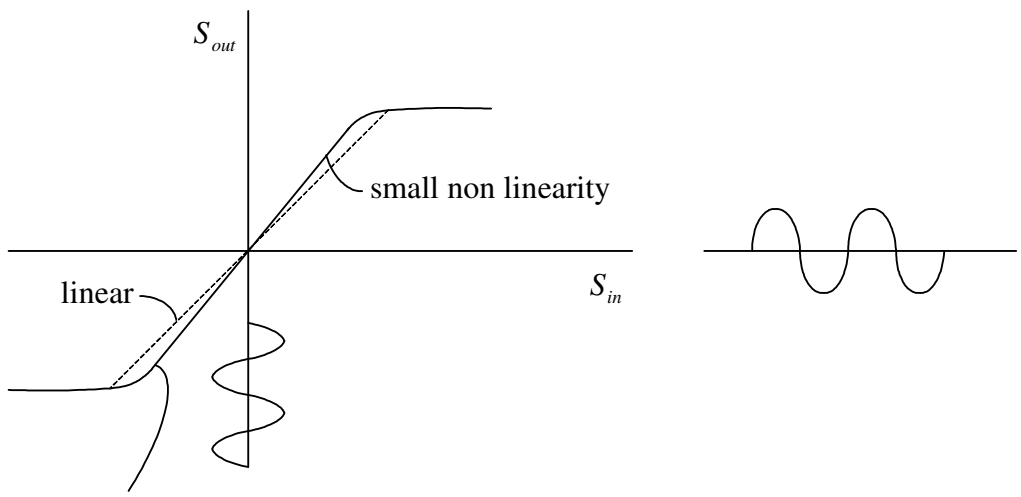
you need to know the phase of the signal that is subjected to non-linearily → use volterra series



Note that at low to midband frequencies, the impedances of device input and output capacitances are very high. Therefore, these capacitances do not have any effect on the signal and do not contribute to signal nonlinearities. For midband frequencies, it is not necessary to use volterra series. For low frequencies, transistor nonlinearities are combined with coupling and bypass capacitance effects. However, these capacitances are very linear and do not contribute to nonlinearities.

At very high frequencies, the existence of nonlinear capacitance as well as the combination of nonlinear effects with finite reactive elements changes the picture. The power series approach does not work anymore. Depending on the degree of nonlinearity, you need to use either Volterra series (weakly nonlinear systems) or other approaches.

* most often this is what you have



you can write this transfer function
in general by a power series

$$S_{out} = a_1 S_{in} + a_2 S_{in}^2 + a_3 S_{in}^3 + \cdots$$

where a_1, a_2, a_3, ,,, are constants

* they usually vary with signal amplitude

* they vary with the choice of bias

* they vary with frequency

* they vary with input/output matching

let's assume : they are constant (ignore variation with signal amplitude)
(fixed bias, small frequency range, fixed matching)

now apply an input function $S_{in} = f(\omega t)$

$$S_{out} = a_1 f(\omega t) + a_2 f^2(\omega t) + a_3 f^3(\omega t) + \cdots$$

$$let \quad f(\omega t) = m_1 \cos\omega_1 t + m_2 \cos\omega_2 t \qquad \omega_1 \sim \omega_2$$

$$S_{out} = a_1(m_1 \cos\omega_1 t + m_2 \cos\omega_2 t) + a_2(m_1 \cos\omega_1 t + m_2 \cos\omega_2 t)^2 + a_3(m_1 \cos\omega_1 t + m_2 \cos\omega_2 t)^3 + \cdots$$

use trigonometric expansions

$$S_{out} = a_1(m_1 \cos\omega_1 t + m_2 \cos\omega_2 t) + a_2\left(m_1^2 \frac{1+\cos 2\omega_1 t}{2} + m_2^2 \frac{1+\cos 2\omega_2 t}{2} + m_1 m_2 \frac{\cos(\omega_1+\omega_2)t + \cos(\omega_1-\omega_2)t}{2}\right) + a_3\left\{\left(m_1^3\left(\frac{\cos\omega_1 t}{2} + \frac{\cos\omega_2 t}{4} + \cdots\right)\right)\right\}$$

$$S_{out} = a_2\left(\frac{m_1^2+m_2^2}{2}\right) \qquad dc$$

$$+a_1(m_1 \cos\omega_1 t + m_2 \cos\omega_2 t) + a_3\left(\frac{3}{4}m_1^3 \cos\omega_1 t + \frac{3}{4}m_2^3 \cos\omega_2 t + \frac{3}{2}m_1 m_2^2 \cos\omega_1 t + \frac{3}{2}m_1^2 m_2^2 \cos\omega_2 t\right) \qquad \omega_1, \omega_2$$

$$+a_2\left(\frac{m_1}{2}^2 \cos 2\omega_1 t + \frac{m_2}{2}^2 \cos 2\omega_2 t\right) \qquad 2\omega_1,\ 2\omega_2$$

$$+a_2\left(\frac{m_1 m_2}{2}\cos(\omega_1+\omega_2)t + \frac{m_1 m_2}{2}\cos(\omega_1-\omega_2)t\right) \qquad \omega_1+\omega_2,\ \omega_1-\omega_2$$

$$+a_3\left(\frac{3}{4}m_1^2 m_2 \cos(2\omega_1-\omega_2)t + \cdots\right) \qquad 2\omega_1-\omega_2,\ 2\omega_2-\omega_1,\ 2\omega_1+\omega_2,\ 2\omega_2+\omega_1$$

$$+a_3\left(\frac{1}{4}m_1^3\cos(3\omega_1)t+\cdots\right)\Bigg\}\begin{matrix}3\omega_1\\3\omega_2\end{matrix}$$

Let's look at the important coefficient

① *coefficient of* $\cos\omega_1 t \Rightarrow$ $\left[a_1m_1+a_3\left(\frac{3}{4}m_1^3+\frac{3}{2}m_1m_2^2\right)\right]\cos\omega_1 t$

② *coefficient of* $\cos(\omega_1-\omega_2) \Rightarrow$ $\left[a_2\left(\frac{m_1m_2}{2}\right)\right]\cos(\omega_1-\omega_2)t$

③ *coefficient of* $\cos 2\omega_1 t \Rightarrow$ $\left[a_2\frac{m_1^2}{2}\right]\cos 2\omega_1 t$

④ *coefficient of* $\cos(2\omega_1-\omega_2)t \Rightarrow$ $\left[a_3\frac{3}{4}m_1^2m_2\right]\cos(2\omega_1-\omega_2)t$

⑤ *coefficient of* $\cos 3\omega_1 t \Rightarrow$ $\left[a_3\frac{1}{4}m_1^3\right]\cos 3\omega_1 t$

First non-linearity

① **Gain compression**

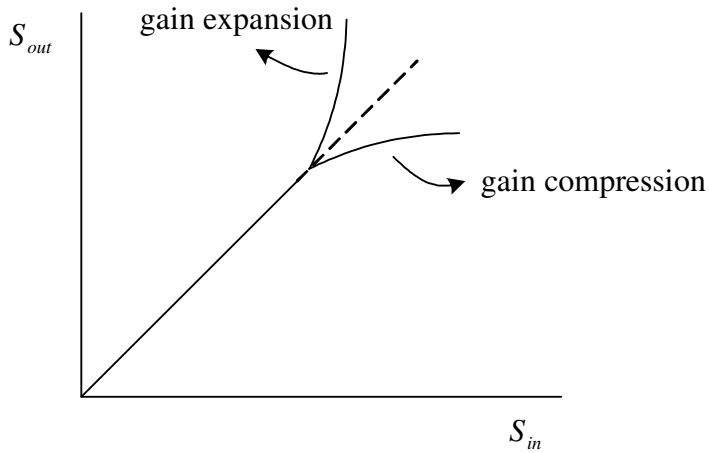
for only one tone : $f(\omega t)=m_1\cos\omega_1 t$

$$\left(a_1m_1+a_3\frac{3}{4}m_1^3\right)\cos\omega_1 t$$

linear gain

gain compression / expansion

as signal becomes stronger



usually we observe gain compression that means a_3 is usually negative

② **IM2**

$$a_2 \frac{m_1 m_2}{2} \cos(\omega_1 - \omega_2) + a_2 \frac{m_1 m_2}{2} \cos(\omega_2 - \omega_1) t = a_2 m_1^2 \cos(\omega_1 - \omega_2) t \qquad \text{good mixer}$$

$$m_1 = m_2$$

$$assume \qquad m_1 = m_2$$

$$\longrightarrow \quad IM2 = \frac{amplitude\ of\ \ 2nd\ \ order}{amplitude\ of\ \ fundamental} = \frac{a_2}{a_1} m_1$$

mixing comes from 2nd order (4th order inter modulation etc…) terms (IM2)

$\longrightarrow$ good news for MOSFETs since their input/output relationship is of 2nd order nature $\longrightarrow$ $i_d = const\ \ V_{gs}^2$

3rd order term also exists in MOSFETs due to two mechanisms

strong non-linearly gain clamping + body effect $\longrightarrow$ weakly non-linear effect

③ **HD2**

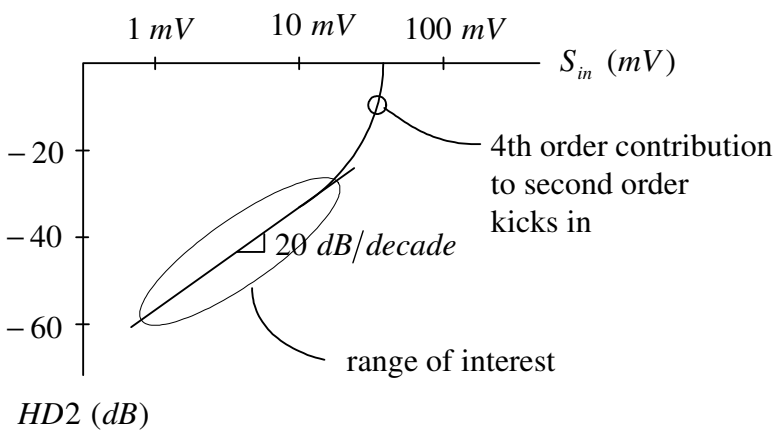
$$a_2 \frac{m_1^2}{2} \cos 2\omega_1 t \longrightarrow \text{2nd harmonic}$$

Define : 2nd harmonic distortion factor HD2

$$HD2 = \frac{second\ hamonic\ amplitude}{fundamental\ amplitude} = \frac{\frac{a_2}{2} m_1^2}{a_1 m_1}$$

assuming gain compression has not occurred yet

$$HD2 = \frac{1}{2}\frac{a_2}{a_1} m_1$$



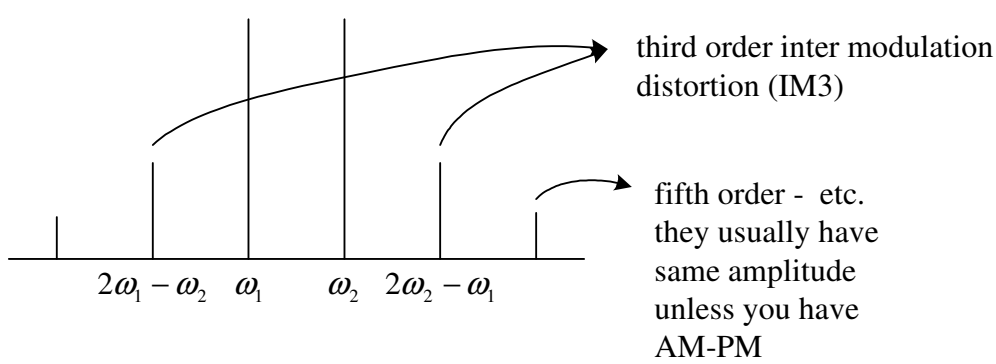
IM2 = 2HD2

2nd intermod

2nd harm distortion

④ **IM3** *assume* $m_1 = m_2$

$$a_3 \frac{3}{4} m_1^3 \cos(2\omega_1 - \omega_2)t = a_3 \frac{3}{4} m_1^3 \cos(2\omega_2 - \omega_1)t$$



IM3 = third order inter modulation

$m_1 = m_2 \longrightarrow$

$$IM3 = \frac{amplitude\ of\ third\ order\ \text{inter mod}}{amplitude\ of\ fundamental}$$

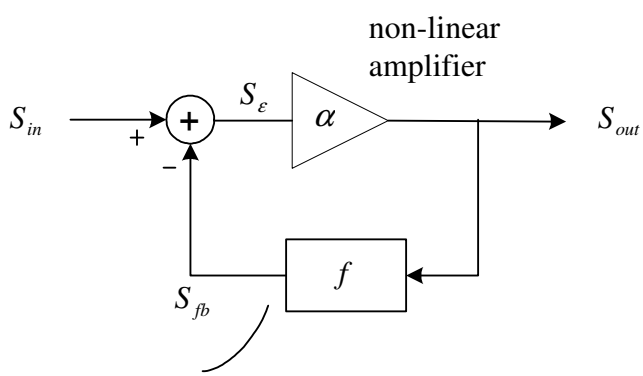
$$IM3 = \frac{a_3 \frac{3}{4} m_1^3}{a_1 m_1} = \frac{3}{4}\frac{a_3}{a_1} m_1^2$$

⑤ **<u>HD3</u>** : $\cos 3\omega_1 t$

$$HD3 = \frac{third\ order\ amplitude}{fundamental\ amplitude} = \frac{a_3 \times \frac{1}{4} m_1^3}{a_1 m_1} = \frac{1}{4}\frac{a_3}{a_1} m_1^2$$

$$\boxed{IM3 = 3HD3}$$

<u>Effect of feedback on distortion</u>



<u>assumption</u> : feedback network is linear

use power series :

$$\left.\begin{aligned} S_{out} &= a_1 S_\varepsilon + a_2 S_\varepsilon^2 + a_3 S_\varepsilon^3 + \cdots \\ S_\varepsilon &= S_{in} - f\, S_{out} \end{aligned}\right\} \longrightarrow$$

$$S_{out} = a_1\left(S_{in} - f\,S_{out}\right) + a_2\left(S_{in} - f\,S_{out}\right)^2 + a_3\left(S_{in} - f\,S_{out}\right)^3 + \cdots$$

substitute

$$S_{out} = b_1 S_{in} + b_2 S_{in}^2 + b_3 S_{in}^3$$

we are looking for b_1, b_2, b_3 coefficients

$$b_1 S_{in} + b_2 S_{in}^2 + b_3 S_{in}^3 = a_1 S_{in} - a_1 f\, b_1 S_{in} - a_1 f\, b_2 S_{in}^2 - a_1 f\, b_3 S_{in}^3$$
$$+ a_2\left(S_{in} - f\, b_1 S_{in} - f\, b_2 S_{in}^2 - f\, b_3 S_{in}^3\right)^2$$
$$+ a_3\left(S_{in} - f\, b_1 S_{in} - f\, b_2 S_{in}^2 - f\, b_3 S_{in}^3\right)^3$$

1st order terms $\longrightarrow$ $b_1 \cancel{S_{in}} = a_1 \cancel{S_{in}} - a_1 f\, b_1 \cancel{S_{in}}$

$$\boxed{b_1 = \frac{a_1}{1 + a_1 f}}$$

standard feedback equation for a linear system

2nd order terms

$$b_2 \cancel{S_{in}^2} = -a_1 f\, b_2 \cancel{S_{in}^2} + a_2\left(1 - f\, b_1\right)^2 \cancel{S_{in}^2}$$

$$b_2\left(1 + a_1 f\right) = a_2\left(1 - f\, b_1\right)^2 = a_2 \frac{1}{\left(1 + a_1 f\right)^2} \longrightarrow \boxed{b_2 = \frac{a_2}{\left(1 + a_1 f\right)^3}}$$

3rd order terms

$$b_3 \cancel{S_{in}^3} = -a_1 f\, b_3 \cancel{S_{in}^3} - 2a_2(1 - f\, b_1) f\, b_2 \cancel{S_{in}^3} + a_3(1 - f\, b_1)^3 \cancel{S_{in}^3}$$

$$b_3(1 + a_1 f) = a_3 \frac{1}{(1 + a_1 f)^3} - 2a_2 \frac{f}{1 + a_1 f} \cdot \frac{a_2}{(1 + a_1 f)^3}$$

$$\boxed{b_3 = \frac{a_3(1 + a_1 f) - 2a_2^2 f}{(1 + a_1 f)^5}}$$

Now let's look at different distortion coefficients

no feedback $\quad HD2 = \frac{1}{2} IM2 = \frac{a_2}{2a_1} m_1$

with feedback $\quad HD2_{feedback} = \frac{b_2}{2b_1} m_1 = \frac{a_1/(1 + a_1 f)}{2a_2/(1 + a_1 f)^3} \cdot m_1$

$$HD2_{feedback} = \frac{a_1}{2a_2} m_1 \Big/ (1 + a_1 f)^2 = \boxed{\frac{HD2_{no-feedback}}{(1 + a_1 f)^2}}$$

→ Dramatic reduction of HD2 (IM2) with feedback when measured based on same input power (m_1)

if HD2 is measured based on output power amplitude

(since total gain is reduced by $\frac{1}{1 + a_1 f}$,

therefore $HD2_{feedback} = \frac{HD2_{no-feedback}}{1 + a_1 f}$ at the same output power)

no feedback $HD3 = \frac{1}{3} IM3 = \frac{1}{4}\frac{a_3}{a_1} m_1^2$

with feedback

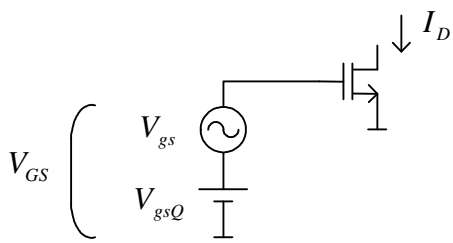
$$HD3_{feedback} = \frac{1}{4}\frac{b_3}{b_1} m_1^2 = \frac{1}{4}\frac{\dfrac{a_3(1+a_1 f) - 2a_2^2 f}{(1+a_1 f)^5}}{\dfrac{a_1}{1+a_1 f}} m_1^2$$

$$= \frac{1}{4}\frac{\dfrac{a_3}{a_1}(1+a_1 f) - 2\dfrac{a_2^2}{a_1} f}{(1+a_1 f)^4} m_1^2$$

you can make $HD3_{feedback}$ *or* $IM3_{feedback} = 0$

by using right amount of feedback!

Distortion in MOS transistors



assume : ideal MOSFET

(no body effect)

body effect introduces third order terms!

$$I_D = \frac{\mu C_{OX}}{2}\frac{W}{L}(V_{GS} - V_{Th})^2$$

$$V_{GS} = V_{gsQ} + V_{gs}$$

$$I_D = \frac{\mu C_{OX}}{2}\frac{W}{L}\left[(V_{gsQ} - V_{Th})^2 + V_{gs}^2 + 2V_{gs}(V_{gsQ} - V_{Th})\right]$$

$$I_D = I_{DQ} + \frac{\mu C_{OX}}{2}\frac{W}{L}V_{gs}^2 + \underbrace{\frac{\mu C_{OX} W}{L}\left(V_{gsQ} - V_{Th}\right)}_{g_m} \cdot V_{gs}$$

$$I_D = I_{DQ} + g_m V_{gs} + \frac{\mu C_{OX}}{2}\frac{W}{L}V_{gs}^2$$

$$\left.\begin{aligned} a_1 &= g_m = \frac{\mu C_{OX} W}{L}\left(V_{gsQ} - V_{Th}\right) \\ a_2 &= \frac{\mu C_{OX}}{2}\frac{W}{L} \\ a_3 &= 0 \end{aligned}\right\} \text{non-linear coefficients}$$

$$V_{gs} = \hat{V}_{gs}\cos\omega t$$

$$HD2 = \frac{1}{2}IM2 = \frac{a_2}{2a_1}m_1 = \frac{\frac{\mu C_{OX}}{2}\frac{W}{L}\hat{V}_{gs}}{\frac{2\mu C_{OX} W}{L}\left(V_{gsQ} - V_{Th}\right)} = \frac{\hat{V}_{gs}}{4\left(V_{gsQ} - V_{Th}\right)}$$

$$\boxed{HD2 = \frac{\hat{V}_{gs}}{4\left(V_{gsQ} - V_{Th}\right)}}$$ in terms of input signal

in terms of output signal $\hat{i}_D = g_m \hat{V}_{gs} = \hat{V}_{gs}\frac{\mu C_{OX} W}{L}\left(V_{gsQ} - V_{Th}\right)$

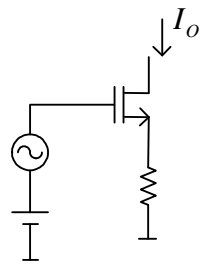
$$\boxed{HD2 = \frac{1}{8}\frac{\hat{i}_D}{I_{DQ}}}$$

→ ideal MOSFET does not show

* gain compression

* IM3, HD3 as long as it is in saturation!

MOSFET with Rs



$$I_O = I_{DQ} + i_d$$

bias signal

+ feedback

$$g_m R_S = f$$

*** ignoring body effect**

$$i_d = b_1 V_{gs} + b_2 V_{gs}^2 + b_3 V_{gs}^3 + \cdots$$

$$b_1 = \frac{a_1}{1 + a_1 f} = \boxed{\frac{g_m}{1 + g_m R_S}}$$

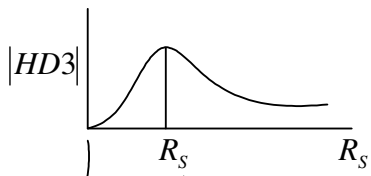
$$b_2 = \frac{a_2}{(1 + a_1 f)^3} = \boxed{\frac{\mu C_{OX} W/L}{2} \times \frac{1}{(1 + g_m R_S)^3}}$$

$$b_3 = \frac{\overset{0}{\cancel{a_3}}(1 + a_1 f) - 2a_2^2 f}{(1 + a_1 f)^5} = \frac{-2a_2^2 f}{(1 + a_1 f)^5} = \boxed{\frac{-2\left(\frac{\mu C_{OX} W/L}{2}\right)^2 R_S}{(1 + g_m R_S)^5}}$$

$$HD3 = \frac{1}{4}\frac{b_3}{b_1}\left|\hat{V}_{gs}\right|^2$$

$$= \frac{1}{4}\ \frac{-2\left(\frac{\mu C_{OX}}{2} W/L\right)^2 R_S \Big/ (1 + g_m R_S)^{\cancel{5}}}{g_m / \cancel{(1 + g_m R_S)}}\left|\hat{V}_{gs}\right|^2$$

$$HD3 = -\frac{1}{2}\frac{\left(\frac{\mu C_{OX}}{2}W/L\right) g_m R_S \left|\hat{V}_{gs}\right|^2}{(1+g_m R_S)^4}$$



$$\frac{\partial HD3}{\partial R_S} = 0$$

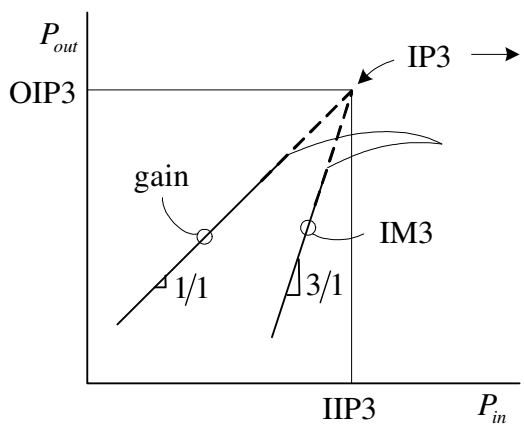
that gives worst HD3

HD3 is zero
for ideal MOS

$$\boxed{R_S = \frac{1}{3g_m}}$$

Rs helps reducing HD2 but worsens HD3 !

Intercept Point



* IP3 is typically
10 ~ 15 dB higher than
1 dB compression point