Integrated Circuits for Communication

Berkeley

Distortion Analysis

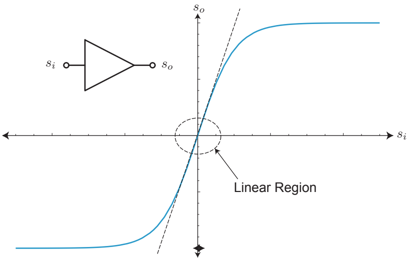
Prof. Ali M. Niknejad

U.C. Berkeley

March 20, 2025

The Origin of Distortion

Introduction to Distortion

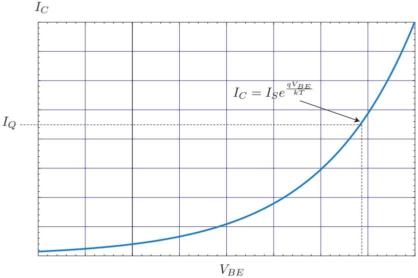


- Up to now we have treated amplifiers as small-signal linear circuits. Since transistors are non-linear, this assumption is only valid for extremeley small signals.
- Consider a class of memoryless non-linear amplifiers. In other words, let's neglect energy storage elements.
- This is the same as saying the output is an instantaneous function of the input. Thus the amplifier has no memory.

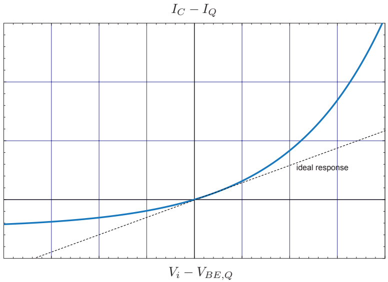
- We also assume the input/output description is sufficiently smooth and continuous as to be accurately described by a power series

$$s_o = a_1 s_i + a_2 s_i^2 + a_3 s_i^3 + \dots$$

For instance, for a BJT (Si, SiGe, GaAs) operated in forward-active region, the collector current is a smooth function of the voltage V_{BE}

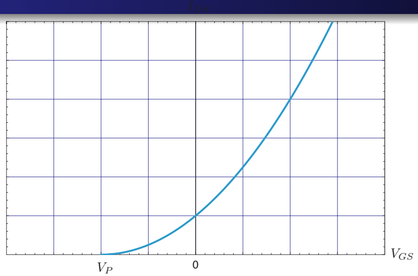


BJT Distortion



- We shift the origin by eliminating the DC signals, $i_o = I_C - I_Q$. The input signal is then applied around the DC level $V_{BE,Q}$.
- Note that an ideal amplifier has a perfectly linear line.

JFET Distortion

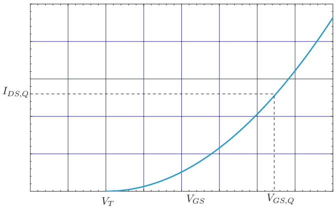


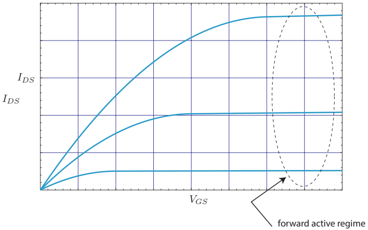
- JFETs are sometimes used in RF circuits. The I-V relation is also approximately square law

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2$$

- The gate current (junction leakage) is typically very small $I_G \sim 10^{12}$A. So for all practical purposes, $R_i = \infty$.

MOSFET Distortion





- The long-channel device also follows the square law relation (neglecting bulk charge effects)

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

- This is assuming the device does not leave the forward active (saturation) regime.

MOSFET Model

- Short-channel devices are even more difficult due to velocity saturation and field dependent mobility. A simple model for a transistor in forward active region is given by (neglecting output resistance)

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

- Note that the device operation near threshold is not captured by this equation.

Single Equation MOSFET Model

- The I-V curve of a MOSFET in moderate and weak inversion is easy to describe in a "piece-meal" fashion, but difficult to capture with a single equation. One approximate single-equation relationship often used is given by

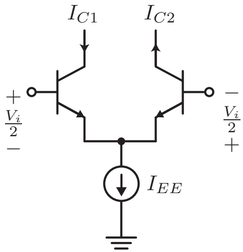
$$I_D = \tfrac{1}{2}\mu C_{ox}\frac{W}{L}\frac{X^2}{1+\theta X}$$

where X is given by

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

- If the exponential term dominates, then $X = V_{GS} - V_t$, which is true for operation in strong inversion. Otherwise, $\ln(1+a) \approx a$, which makes the model mimic the weak-inversion "bipolar" exponential characteristic.

Differential Pair



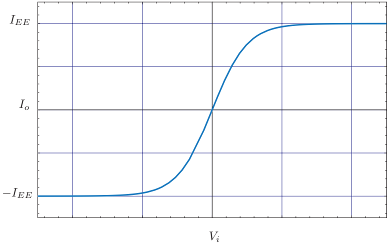
The differential pair is an important analog and RF building block.

- For a BJT diff pair, we have $V_i = V_{BE1} - V_{BE2}$

$$I_{C1,2} = I_S e^{\frac{qV_{BE1,2}}{kT}}$$

- The sum of the collector currents are equal to the current source $I_{C1} + I_{C2} = I_{EE}$

BJT Diff Pair



- The ideal BJT diff pair I-V relationship (neglecting base and emitter resistance) is give by

$$I_o = I_{C1} - I_{C2} = \alpha I_{EE} \tanh \frac{qV_i}{2kT}$$

- Notice that the output current saturates for large input voltages

Modeling Amplifiers with a Power Series

Power Series Relation

- For a general circuit, let's represent this behavior with a power series

$$s_o = a_1 s_i + a_2 s_i^2 + a_3 s_i^3 + \dots$$

- a_1 is the small signal gain
- The coefficients a_1,a_2,a_3,... are independent of the input signal s_i *but* they depend on bias, temperature, and other factors.

Harmonic Distortion

- Assume we drive the amplifier with a time harmonic signal at frequency ω_1

$$s_i = S_1 \cos \omega_1 t$$

- A linear amplifier would output $s_o = a_1 S_1 \cos \omega_1 t$ whereas our amplifier generates

$$s_o = a_1 S_1 \cos \omega_1 t + a_2 S_1^2 \cos^2 \omega_1 t + a_3 S_1^3 \cos^3 \omega_1 t + \dots$$

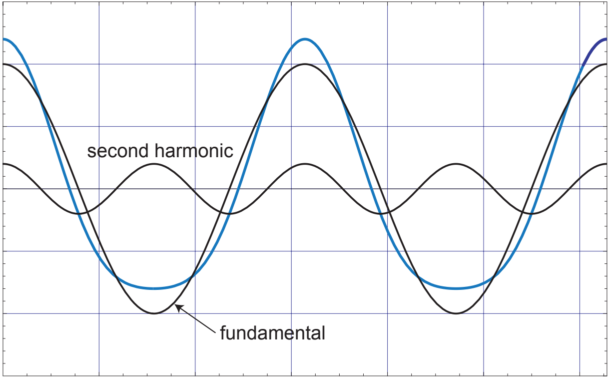
or

$$s_o = a_1 S_1 \cos \omega_1 t + \frac{a_2 S_1^2}{2} (1 + \cos 2\omega_1 t) + \frac{a_3 S_1^3}{4} (\cos 3\omega_1 t + 3 \cos \omega_1 t) + \dots$$

Harmonic Distortion (cont)

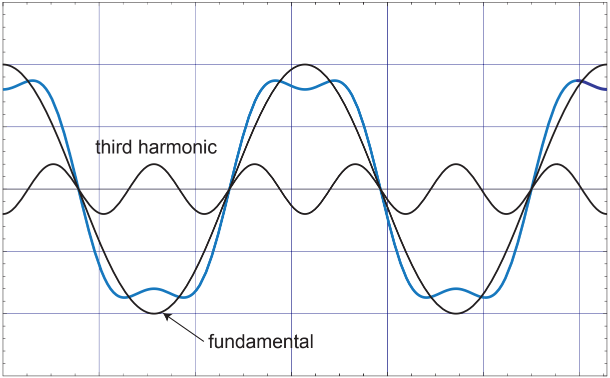
- The term $a_1 s_1 \cos \omega_1 t$ is the wanted signal.
- Higher harmonics are also generated. These are unwanted and thus called "distortion" terms. We already see that the second-harmonic $\cos 2\omega_1 t$ and third harmonic $\cos 3\omega_1 t$ are generated.
- Also the second order non-linearity produces a DC shift of $\frac{1}{2} a_2 S_1^2$.
- The third order generates both third order distortion and more fundamental. The sign of a_1 and a_3 determine whether the distortion product $a_3 S_1^3 \frac{3}{4} \cos \omega_1 t$ adds or subtracts from the fundamental.
- If the signal adds, we say there is gain expansion. If it subtracts, we say there is gain compression.

Second Harmonic Disto Waveforms



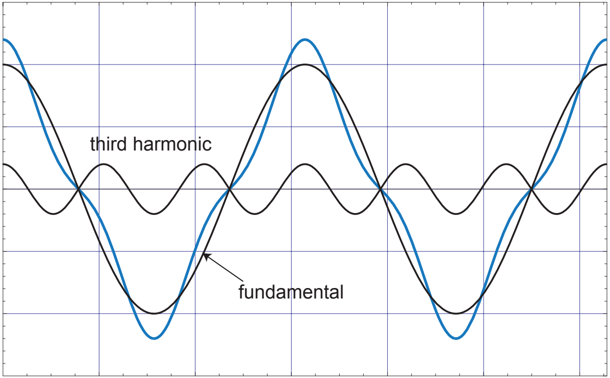
- The figure above demonstrates the waveform distortion due to second harmonic only.

Third Harmonic Distortion Waveform



- The above figure shows the effects of the third harmonic, where we assume the third harmonic is in phase with the fundamental.

Third Harmonic Waveform (cont)



- The above figure shows the effects of the third harmonic, where we assume the third harmonic is out of phase with the fundamental.

General Distortion Term

- Consider the term $\cos^n \theta = \frac{1}{2^n}\left(e^{j\theta} + e^{-j\theta}\right)^n$. Using the Binomial formula, we can expand to

$$= \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} e^{jk\theta} e^{-j(n-k)\theta}$$

- For $n = 3$

$$= \quad \frac{1}{8}\left(\binom{3}{0} e^{-j3\theta} + \binom{3}{1} e^{j\theta} e^{-j2\theta} + \binom{3}{2} e^{j2\theta} e^{-j\theta} + \binom{3}{3} e^{j3\theta}\right)$$

$$= \frac{1}{8}\left(e^{-j3\theta} + e^{j3\theta}\right) + \frac{1}{8} 3 \left(e^{j\theta} + e^{-j\theta}\right) = \frac{1}{4}\cos 3\theta + \frac{3}{4}\cos\theta$$

General Distortion Term (cont)

- We can already see that for an odd power, we will see a nice pairing up of positive and negative powers of exponentials
- For the even case, the middle term is the unpaired DC term

$$\binom{2k}{k} e^{jk\theta} e^{-jk\theta} = \binom{2k}{k}$$

- So only even powers in the transfer function can shift the DC operation point.
- The general term in the binomial expansion of $(x + x^{-1})^n$ is given by

$$\binom{n}{k} x^{n-k} x^{-k} = \binom{n}{k} x^{n-2k}$$

General Distortion Term (cont)

- The term $\binom{n}{k} x^{n-2k}$ generates every other harmonic.
- If n is even, then only even harmonics are generated. If n is odd, likewise, only odd harmonics are generated.
- Recall that an "odd" function $f(-x) = -f(x)$ (anti-symmetric) has an odd power series expansion

$$f(x) = a_1 x + a_3 x^3 + a_5 x^5 + \dots$$

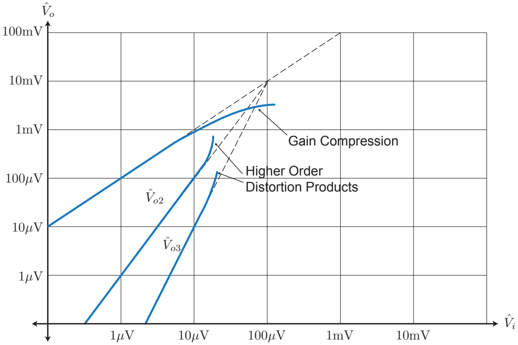
- Whereas an even function, $g(-x) = g(x)$, has an even power series expansion

$$g(x) = a_0 + a_2 x^2 + a_4 x^4 + \dots$$

Output Waveform

- In general, then, the output waveform is a Fourier series

$$v_o = \hat{V}_{o1} \cos \omega_1 t + \hat{V}_{o2} \cos 2\omega_1 t + \hat{V}_{o3} \cos 3\omega_1 t + \ldots$$



Harmonic Distortion Metrics

Fractional Harmonic Distortion

- The fractional second-harmonic distortion is a commonly cited metric

$$HD_2 = \frac{\text{ampl of second harmonic}}{\text{ampl of fund}}$$

- If we assume that the square power dominates the second-harmonic

$$HD_2 = \frac{a_2 \frac{S_1^2}{2}}{a_1 S_1}$$

or

$$HD_2 = \frac{1}{2}\frac{a_2}{a_1}S_1$$

Third Harmonic Distortion

- The fractional third harmonic distortion is given by

$$HD_3 = \frac{\text{ampl of third harmonic}}{\text{ampl of fund}}$$

- If we assume that the cubic power dominates the third harmonic

$$HD_3 = \frac{a_3 \frac{S_1^3}{4}}{a_1 S_1}$$

or

$$HD_3 = \frac{1}{4} \frac{a_3}{a_1} S_1^2$$

Output Referred Harmonic Distortion

- In terms of the output signal S_{om}, if we again neglect gain expansion/compression, we have $S_{om} = a_1 S_1$

$$HD_2 = \frac{1}{2}\frac{a_2}{a_1^2} S_{om}$$

$$HD_3 = \frac{1}{4}\frac{a_3}{a_1^3} S_{om}^2$$

- On a dB scale, the second harmonic increases linearly with a slope of one in terms of the output power whereas the thrid harmonic increases with a slope of 2.

Signal Power

- Recall that a general memoryless non-linear system will produce an output that can be written in the following form

$$v_o(t) = \hat{V}_{o1} \cos \omega_1 t + \hat{V}_{o2} \cos 2\omega_1 t + \hat{V}_{o3} \cos 3\omega_1 t + \dots$$

- By Parseval's theorem, we know the total power in the signal is related to the power in the harmonics

$$\int_T v^2(t) dt = \int_T \sum_j \hat{V}_{oj} \cos(j\omega_1 t) \sum_k \hat{V}_{ok} \cos(k\omega_1 t) dt$$

$$= \sum_j \sum_k \int_T \hat{V}_{oj} \cos(j\omega_1 t) \hat{V}_{ok} \cos(k\omega_1 t) dt$$

Power in Distortion

- By the orthogonality of the harmonics, we obtain Parseval's Them

$$\int_T v^2(t)dt = \sum_j \sum_k \tfrac{1}{2}\delta_{jk} \hat{V}_{oj} \hat{V}_{ok} = \tfrac{1}{2} \sum_j \hat{V}_{oj}^2$$

- The power in the distortion relative to the fundamental power is therefore given by

$$\frac{\text{Power in Distortion}}{\text{Power in Fundamental}} = \frac{V_{o2}^2}{V_{o1}^2} + \frac{V_{o3}^2}{V_{o1}^2} + \cdots$$

$$= HD_2^2 + HD_3^2 + HD_4^2 + \cdots$$

Total Harmonic Distortion

- We define the *Total Harmonic Distortion* (*THD*) by the following expression

$$THD = \sqrt{HD_2^2 + HD_3^2 + \cdots}$$

- Based on the particular application, we specify the maximum tolerable *THD*
- Telephone audio can be pretty distorted ($THD < 10\%$)
- High quality audio is very sensitive ($THD < 1\%$ to $THD < .001\%$)
- Video is also pretty forgiving, $THD < 5\%$ for most applications
- Analog Repeaters $< .001\%$. RF Amplifiers $< 0.1\%$

Intermodulation and Crossmodulation Distortion

Intermodulation Distortion

- So far we have characterized a non-linear system for a single tone. What if we apply two tones

$$S_i = S_1 \cos \omega_1 t + S_2 \cos \omega_2 t$$

$$S_o = a_1 S_i + a_2 S_i^2 + a_3 S_i^3 + \cdots$$

$$= a_1 S_1 \cos \omega_1 t + a_1 S_2 \cos \omega_2 t + a_2 (S_i)^2 + a_3 (S_i)^3 + \cdots$$

- The second power term gives

$$a_2 S_1^2 \cos^2 \omega_1 t + a_2 S_2^2 \cos^2 \omega_2 t + 2 a_2 S_1 S_2 \cos \omega_1 t \cos \omega_2 t$$

$$= a_2 \frac{S_1^2}{2} (\cos 2\omega_1 t + 1) + a_2 \frac{S_2^2}{2} (\cos 2\omega_2 t + 1) + a_2 S_1 S_2 \left(\cos(\omega_1 + \omega_2) t - \cos(\omega_1 - \omega_2) t \right)$$

Second Order Intermodulation

- The last term $\cos(\omega_1 \pm \omega_2)t$ is the second-order intermodulation term
- The intermodulation distortion IM_2 is defined when the two input signals have equal amplitude $S_i = S_1 = S_2$

$$IM_2 = \frac{\text{Amp of Intermod}}{\text{Amp of Fund}} = \frac{a_2}{a_1} S_i$$

- Note the relation between IM_2 and HD_2

$$IM_2 = 2HD_2 = HD_2 + 6\mathrm{dB}$$

Practical Effects of IM_2

- This term produces distortion at a lower frequency $\omega_1 - \omega_2$ and at a higher frequency $\omega_1 + \omega_2$
- Example: Say the receiver bandwidth is from $800\mathrm{MHz} - 2.4\mathrm{GHz}$ and two unwanted interfering signals appear at $800\mathrm{MHz}$ and $900\mathrm{MHz}$.
- Then we see that the second-order distortion will produce distortion at $100\mathrm{MHz}$ and $1.7\mathrm{GHz}$. Since $1.7\mathrm{GHz}$ is in the receiver band, signals at this frequency will be corrupted by the distortion.
- A weak signal in this band can be "swamped" by the distortion.
- Apparently, a "narrowband" system does not suffer from IM_2? Or does it ?

Low-IF Receiver

- In a low-IF or direct conversion receiver, the signal is down-converted to a low intermediate frequency f_{IF}
- Since $\omega_1 - \omega_2$ can potentially produce distortion at low frequency, IM_2 is very important in such systems
- Example: A narrowband system has a receiver bandwidth of 1.9GHz - 2.0GHz. A sharp input filter eliminates any interference outside of this band. The IF frequency is 1MHz
- Imagine two interfering signals appear at $f_1 = 1.910\text{GHz}$ and $f_2 = 1.911\text{GHz}$. Notice that $f_2 - f_1 = f_{IF}$
- Thus the output of the amplifier/mixer generate distortion at the IF frequency, potentially disrupting the communication.

Cubic IM

- Now let's consider the output of the cubic term

$$a_3 s_i^3 = a_3(S_1 \cos \omega_1 t + S_2 \cos \omega_2 t)^3$$

- Let's first notice that the first and last term in the expansion are the same as the cubic distortion with a single input

$$\frac{a_3 S_{1,2}^3}{4} (\cos 3\omega_{1,2} t + 3 \cos \omega_{1,2} t)$$

- The cross terms look like

$$\binom{3}{2} a_3 S_1 S_2^2 \cos \omega_1 t \cos^2 \omega_2 t$$

Third Order IM

- Which can be simplified to

$$3\cos\omega_1 t\cos^2\omega_2 t = \frac{3}{2}\cos\omega_1 t(1+\cos 2\omega_2 t) =$$

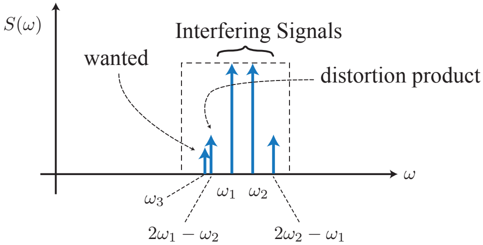
$$= \frac{3}{2}\cos\omega_1 t + \frac{3}{4}\cos(2\omega_2 \pm \omega_1)$$

- The interesting term is the intermodulation at $2\omega_2 \pm \omega_1$
- By symmetry, then, we also generate a term like

$$a_3 S_1^2 S_2 \frac{3}{4}\cos(2\omega_1 \pm \omega_2)$$

- Notice that if $\omega_1 \approx \omega_2$, then the intermodulation $2\omega_2 - \omega_1 \approx \omega_1$

Inband IM3 Distortion



- Now we see that even if the system is narrowband, the output of an amplifier can contain in band intermodulation due to IM_3.
- This is in contrast to IM_2 where the frequency of the intermodulation was at a lower and higher frequency. The IM_3 frequency can fall in-band for two in-band interferer

Definition of IM_3

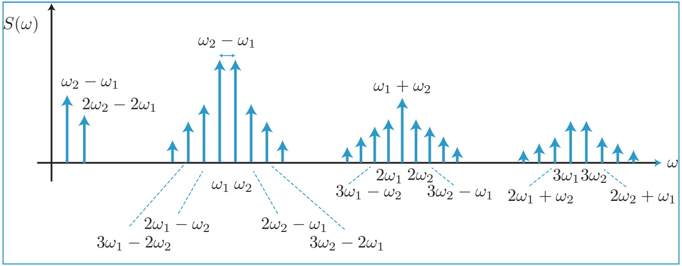
- We define IM_3 in a similar manner for $S_i = S_1 = S_2$

$$IM_3 = \frac{\text{Amp of Third Intermod}}{\text{Amp of Fund}} = \frac{3}{4}\frac{a_3}{a_1}S_i^2$$

- Note the relation between IM_3 and HD_3

$$IM_3 = 3HD_3 = HD_3 + 10\mathrm{dB}$$

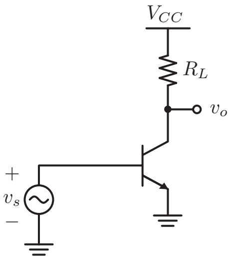
Complete Two-Tone Response



- We have so far identified the harmonics and IM_2 and IM_3 products
- A more detailed analysis shows that an order n non-linearity can produce intermodulation at frequencies $j\omega_1 \pm k\omega_2$ where $j + k = n$
- All tones are spaced by the difference $\omega_2 - \omega_1$

Examples

Distortion of BJT Amplifiers



- Consider the CE BJT amplifier shown. The biasing is omitted for clarity.

- The output voltage is simply

$$V_o = V_{CC} - I_C R_C$$

- Therefore the distortion is generated by I_C alone. Recall that

$$I_C = I_S e^{qV_{BE}/kT}$$

BJT CE Distortion (cont)

- Now assume the input $V_{BE} = v_i + V_Q$, where V_Q is the bias point. The current is therefore given by

$$I_C = \underbrace{I_S e^{\frac{V_Q}{V_T}}}_{I_Q} e^{\frac{v_i}{V_T}}$$

- Using a Taylor expansion for the exponential

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots$$

$$I_C = I_Q(1 + \frac{v_i}{V_T} + \frac{1}{2}\left(\frac{v_i}{V_T}\right)^2 + \frac{1}{6}\left(\frac{v_i}{V_T}\right)^3 + \cdots)$$

BJT CE Distortion (cont)

- Define the output signal $i_c = I_C - I_Q$

$$i_c = \frac{I_Q}{V_T} v_i + \frac{1}{2}\left(\frac{q}{kT}\right)^2 I_Q v_i^2 + \frac{1}{6}\left(\frac{q}{kT}\right)^3 I_Q v_i^3 + \cdots$$

- Compare to $S_o = a_1 S_i + a_2 S_i^2 + a_3 S_i^3 + \cdots$

$$a_1 = \frac{qI_Q}{kT} = g_m$$

$$a_2 = \frac{1}{2}\left(\frac{q}{kT}\right)^2 I_Q$$

$$a_3 = \frac{1}{6}\left(\frac{q}{kT}\right)^3 I_Q$$

Example: BJT HD2

- For any BJT (Si, SiGe, Ge, GaAs), we have the following result

$$HD_2 = \frac{1}{4}\frac{q\hat{v}_i}{kT}$$

- where $\hat{v}_i$ is the peak value of the input sine voltage
- For $\hat{v}_i = 10\mathrm{mV}$, $HD_2 = 0.1 = 10\%$
- We can also express the distortion as a function of the output current swing $\hat{i}_c$

$$HD_2 = \frac{1}{2}\frac{a_2}{a_1^2}S_{om} = \frac{1}{4}\frac{\hat{i}_c}{I_Q}$$

- For $\frac{\hat{i}_c}{I_Q} = 0.4$, $HD_2 = 10\%$

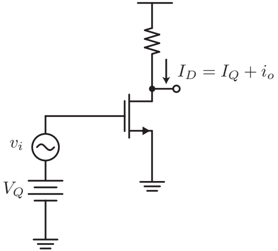
Example: BJT IM3

- Let's see the maximum allowed signal for $IM_3 \leq 1\%$

$$IM_3 = \frac{3}{4}\frac{a_3}{a_1}S_1^2 = \frac{1}{8}\left(\frac{q\hat{v}_i}{kT}\right)^2$$

- Solve $\hat{v}_i = 7.3\mathrm{mV}$. That's a pretty small voltage. For practical applications we'd like to improve the linearity of this amplifier.

Example: Disto in Long-Ch. MOS



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

$$i_o+I_Q = \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_Q+v_i-V_T)^2$$

- Ignoring the output impedance we have

$$= \frac{1}{2}\mu C_{ox}\frac{W}{L}\left\{(V_Q - V_T)^2 + v_i^2 + 2v_i(V_Q - V_T)\right\}$$

$$= \underbrace{I_Q}_{\text{dc}} + \underbrace{\mu C_{ox}\frac{W}{L}v_i(V_Q - V_T)}_{\text{linear}} + \underbrace{\frac{1}{2}\mu C_{ox}\frac{W}{L}v_i^2}_{\text{quadratic}}$$

Ideal Square Law Device

- An ideal square law device only generates 2nd order distortion

$$i_o = g_m v_i + \tfrac{1}{2}\mu C_{ox}\frac{W}{L}v_i^2$$

$$a_1 = g_m$$

$$a_2 = \tfrac{1}{2}\mu C_{ox}\frac{W}{L} = \tfrac{1}{2}\frac{g_m}{V_Q - V_T}$$

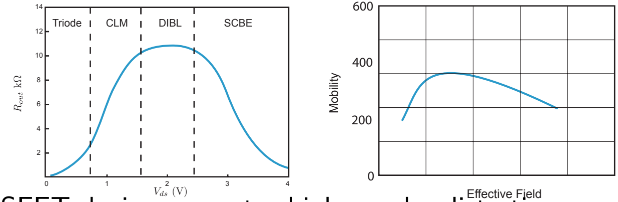
$$a_3 \equiv 0$$

- The harmonic distortion is given by

$$HD_2 = \frac{1}{2}\frac{a_2}{a_1}v_i = \frac{1}{4}\frac{g_m}{V_Q - V_T}\frac{1}{g_m}v_i = \frac{1}{4}\frac{v_i}{V_Q - V_T}$$

$$HD_3 = 0$$

Real MOSFET Device



- The real MOSFET device generates higher order distortion
- The output impedance is non-linear. The mobility μ is not a constant but a function of the vertical and horizontal electric field
- We may also bias the device at moderate or weak inversion, where the device behavior is more exponential
- There is also internal *feedback*