EE 538: Low-Noise Analog Circuit Design

Spring 2021

Instructor: Jason Silver

Announcements

- Assignment 5 will be posted Saturday, May 8
- Midterm exam will be posted Friday, May 7
  - Take-home, due Sunday, May 16

Week 6

- EE538 lecture notes
- Analysis and Design of Analog Integrated Circuits, 5^{th} Edition
- The Design of CMOS Radio-Frequency Integrated Circuits, 2^{nd} Edition

Overview

- Last time...
  - Noise in common-emitter stages
  - Active load and current mirror noise
  - Cascode and parallel stages
  - Differential amplifier noise
- Today...
  - Nonlinear device characteristics
  - Distortion in amplifiers
  - Total harmonic distortion
  - Intermodulation distortion
  - Effect of feedback on nonlinearity
  - Dynamic range

Python packages/modules

In [3]:

```
import matplotlib as mpl
from matplotlib import pyplot as plt
from matplotlib import ticker, cm
import numpy as np
from scipy import signal
from scipy import integrate
from scipy.fft import fft
#%matplotlib notebook

mpl.rcParams['font.size'] = 14
mpl.rcParams['legend.fontsize'] = 'large'

def plot_xy(x, y, xlabel, ylabel):
    fig, ax = plt.subplots(figsize=(10.0, 7.5));
    ax.plot(x, y, 'b')
    ax.grid()
    ax.set_xlabel(xlabel)
    ax.set_ylabel(ylabel)

def plot_2xy(x, y1, y2, xlabel, ylabel, y1label, y2label):
    fig, ax = plt.subplots(figsize=(10.0,7.5))

    ax.plot(x, y1, 'b', label=y1label)
    ax.plot(x, y2, 'r', label=y2label)
    ax.set_ylabel(ylabel)
    ax.set_xlabel(xlabel)
    ax.grid()

    ax.legend()
    ax.legend(loc='upper center', ncol=2, fancybox=True,
              shadow=True, bbox_to_anchor=(0.5,1.13) )

def plot_xy2(x1, y1, x1label, y1label, x2, y2, x2label, y2label):
    fig, ax = plt.subplots(2, figsize = (10.0, 7.5));
    ax[0].plot(x1, y1, 'b')
    ax[0].set_ylabel(y1label)
    ax[0].grid()

    ax[1].plot(x2, y2, 'b')
    ax[1].set_xlabel(x1label)
    ax[1].set_xlabel(x2label)
    ax[1].set_ylabel(y2label)
    ax[1].grid()

    fig.align_ylabels(ax[:])

def plot_xy3(x, y1, y2, y3, xlabel, y1label, y2label, y3label):
    fig, ax = plt.subplots(3, figsize=(10.0,7.5))

    ax[0].plot(x, y1)
    ax[0].set_ylabel(y1label)
    ax[0].grid()

    ax[1].plot(x, y2)
    ax[1].set_ylabel(y2label)
    ax[1].grid()

    ax[2].plot(x, y3)
    ax[2].set_ylabel(y3label)
    ax[2].set_xlabel(xlabel)
    ax[2].grid()

def plot_logxy3(x, y1, y2, y3, xlabel, y1label, y2label, y3label):
    fig, ax = plt.subplots(3, figsize=(10.0,7.5))

    ax[0].semilogx(x, y1)
    ax[0].set_ylabel(y1label)
    ax[0].grid()

    ax[1].semilogx(x, y2)
    ax[1].set_ylabel(y2label)
    ax[1].grid()

    ax[2].semilogx(x, y3)
    ax[2].set_ylabel(y3label)
    ax[2].set_xlabel(xlabel)
    ax[2].grid()

def plot_logxy(x, y, xlabel, ylabel):
    fig, ax = plt.subplots(figsize=(10.0, 7.5))
    ax.semilogx(x, y, 'b')
    ax.grid();
    ax.set_xlabel(xlabel)
    ax.set_ylabel(ylabel)

def plot_log3xy(x, y1, y2, y3, xlabel, ylabel, y1label, y2label, y3label):
    fig, ax = plt.subplots(figsize=(10.0,7.5))

    ax.semilogx(x, y1, 'b', label=y1label)
    ax.semilogx(x, y2, 'r', label=y2label)
    ax.semilogx(x, y3, 'g', label=y3label)
    ax.set_ylabel(ylabel)
    ax.set_xlabel(xlabel)
    ax.grid()

    ax.legend()
    ax.legend(loc='upper center', ncol=3, fancybox=True,
              shadow=True, bbox_to_anchor=(0.5,1.13) )

def plot_log2xy(x, y1, y2, xlabel, ylabel, y1label, y2label):
    fig, ax = plt.subplots(figsize=(10.0,7.5))

    ax.semilogx(x, y1, 'b', label=y1label)
    ax.semilogx(x, y2, 'r', label=y2label)
    ax.set_ylabel(ylabel)
    ax.set_xlabel(xlabel)
    ax.grid()

    ax.legend()
    ax.legend(loc='upper center', ncol=2, fancybox=True,
              shadow=True, bbox_to_anchor=(0.5,1.13) )

def plot_loglog(x, y, xlabel, ylabel):
    fig, ax = plt.subplots(figsize=(10.0, 7.5))
    ax.loglog(x, y, 'b')
    ax.grid();
    ax.set_xlabel(xlabel)
    ax.set_ylabel(ylabel)


def plot_loglog2(x, y1, y2, xlabel, ylabel, y1label, y2label):
    fig, ax = plt.subplots(figsize=(10.0,7.5))

    ax.loglog(x, y1, 'b', label=y1label)
    ax.loglog(x, y2, 'r', label=y2label)
    ax.set_ylabel(ylabel)
    ax.set_xlabel(xlabel)
    ax.grid()

    ax.legend()
    ax.legend(loc='upper center', ncol=2, fancybox=True,
              shadow=True, bbox_to_anchor=(0.5,1.13) )

def plot_xlogy(x, y, xlabel, ylabel):
    fig, ax = plt.subplots(figsize=(10.0, 7.5));
    ax.semilogy(x, y, 'b');
    ax.grid();
    ax.set_xlabel(xlabel);
    ax.set_ylabel(ylabel);

def read_ltspice_ac(file_name):
    with open(file_name, 'r') as data:
        x = []
        y = []
        z = []
        next(data) # skip header line
        for line in data:
            p = line.split()
            x.append(float(p[0]))
            complex = p[1].split(",")
            y.append(float(complex[0]))
            z.append(float(complex[1]))

    return x, y, z

def plot_logxy2(x1, y1, x2, y2, x1label, y1label, x2label, y2label):
    fig, ax = plt.subplots(2, figsize = (10.0, 7.5));
    ax[0].semilogx(x1, y1, 'b');
    ax[0].set_ylabel(y1label)
    ax[0].grid()

    ax[1].semilogx(x2, y2, 'b');
    ax[1].set_xlabel(x1label)
    ax[1].set_xlabel(x2label);
    ax[1].set_ylabel(y2label);
    ax[1].grid();

    fig.align_ylabels(ax[:])

def plot_noise_bandwidth(f, mag):
    fig, ax = plt.subplots(2, figsize=(10.0,7.5))
    ax[0].semilogx(f, RC_mag)
    ax[0].set_xscale("log")
    ax[0].set_xlim(f[0], f[-1])
    ax[0].set_xticks(np.logspace(0.1,4,5))
    ax[0].set_xticklabels([])
    ax[0].set_ylabel('Magnitude [V/V]')
    ax[0].set_title('Equivalent Noise Bandwidth')
    ax[0].grid()

    ax[1].hlines(1, 0, f_enb, color='tab:blue')
    ax[1].hlines(0, f_enb, f[-1], color='tab:blue')
    ax[1].vlines(f_enb, 0, 1, color='tab:blue')
    ax[1].set_xlim(f[0], f[-1])
    ax[1].set_xscale("log")
    ax[1].set_xticks(np.logspace(0.1,4,5))
    ax[1].set_xticklabels([r'$10^0$',r'$10^1$', r'$10^2$', r'$10^3$', r'$10^4$'])
    ax[1].set_ylabel('Magnitude [V/V]')
    ax[1].set_xlabel('Frequency [Hz]')
    ax[1].grid()

def noise_hist(vnoise, vn_rms, bins):
    fig = plt.figure( figsize=(10.0,7.5) )
    vn_norm = vnoise/ vn_rms
    ax = fig.add_subplot(111)
    n, bins, rectangles = ax.hist(vn_norm, bins, density=True, range=(-3, 3),
                                  color='b')
    ax.set_xlabel(r'Sample Voltage [$v_{n(rms)}$]')
    ax.set_ylabel('Probability Density')
    ax.grid()
    fig.canvas.draw()

def plot_NF_vs_Rs(en_vals, in_vals, Rs_min, Rs_max, T_in_K):
    fig, ax = plt.subplots(figsize=(10.0, 7.5))
    k = 1.38e-23
    Rs = np.logspace(np.log10(Rs_min), np.log10(Rs_max), num=200)
    F1 = 1 + (en_vals[0]**2+Rs**2*in_vals[0]**2)/(4*k*T_in_K*Rs)
    F2 = 1 + (en_vals[1]**2+Rs**2*in_vals[1]**2)/(4*k*T_in_K*Rs)
    F3 = 1 + (en_vals[2]**2+Rs**2*in_vals[2]**2)/(4*k*T_in_K*Rs)
    ax.semilogx(Rs, 10*np.log10(F1), 'b', label=r'$e_{n1}$, $i_{n1}$')
    ax.semilogx(Rs, 10*np.log10(F2), 'r', label=r'$e_{n2}$, $i_{n2}$')
    ax.semilogx(Rs, 10*np.log10(F3), 'g', label=r'$e_{n3}$, $i_{n3}$')
    ax.grid();
    ax.set_xlabel(r'Source Resistance $R_s$ [\Omega]$')
    ax.set_ylabel(r'Noise Figure $NF$ [$dB$]')

    ax.legend()
    ax.legend(loc='upper center', ncol=3, fancybox=True,
              shadow=True, bbox_to_anchor=(0.5,1.13) )

def plot_noise_curve(e_n, i_n, Rs_min, Rs_max):
    fig, ax = plt.subplots(figsize=(10.0, 7.5))
    Rs = np.logspace(np.log10(Rs_min), np.log10(Rs_max), num=200)
    e_ni_2 = 4*k*T*Rs + e_n**2 + i_n**2*Rs**2
    ax.loglog(Rs, np.sqrt(e_ni_2), 'b', label='Total Noise')
    ax.loglog(Rs, np.sqrt(4*k*T*Rs), 'r', label=r'$\sqrt{4kTR_s}$')
    ax.loglog(Rs, e_n*np.ones(np.size(Rs)), 'g', label=r'$e_n$')
    ax.loglog(Rs, i_n*Rs, 'y', label=r'$i_n R_s$')
    ax.grid();
    ax.set_xlabel(r'Source Resistance $R_s$ [\Omega]$')
    ax.set_ylabel(r'Equivalent Input Noise [$V/\sqrt{Hz}$]')

    ax.legend()
    ax.legend(loc='upper center', ncol=4, fancybox=True,
              shadow=True, bbox_to_anchor=(0.5,1.13) )

def plot_bjt_NF(beta, r_bb, Rmin, Rmax, Imin, Imax):
    fig, ax = plt.subplots(figsize=(10.0, 7.5))
    k = 1.38e-23
    T = 300
    q = 1.602e-19
    V_T = k*T/q
    rs = np.logspace(np.log10(Rmin), np.log10(Rmax), num = 100)
    ic = np.logspace(np.log10(Imin), np.log10(Imax), num = 100)
    I_C, R_S = np.meshgrid(ic, rs)
    e_n_2 = 4*k*T*(V_T/2/I_C + r_bb)
    i_n_2 = 2*q*I_C/beta_0
    NF = 1 + (e_n_2 + i_n_2*R_S**2)/(4*k*T*R_S)
    cp = ax.contourf(I_C, R_S, 10*np.log10(NF), levels=np.linspace(0,15, num=16))
    plt.xscale('log')
    plt.yscale('log')
    plt.ylabel(r'Source Resistance $R_s$ [$\Omega$]')
    plt.xlabel(r'Collector Current $I_C$ [A]')
    fig.colorbar(cp)

def fftnoise(f):
    f = np.array(f, dtype='complex')
    Np = (len(f) - 1) // 2
    phases = np.random.rand(Np) * 2 * np.pi
    phases = np.cos(phases) + 1j * np.sin(phases)
    f[1:Np+1] *= phases
    f[-1:-1-Np:-1] = np.conj(f[1:Np+1])
    return np.fft.ifft(f).real

def band_limited_noise(min_freq, max_freq, samples=1024, samplerate=1):
    freqs = np.abs(np.fft.fftfreq(samples, 1/samplerate))
    f = np.zeros(samples)
    idx = np.where(np.logical_and(freqs>=min_freq, freqs<=max_freq))[0]
    f[idx] = 1
    return fftnoise(f)

def fft_mag(x, N, T, t):
    fft_sig  = fft(x, N)
    freqs = np.linspace(0.0, 1.0/(2.0*T), N//2)
    mags = 2.0/N * np.abs(fft_sig[0:N//2]) # single-sided FFT

    return freqs, mags

def plot_fft_dB(freqs, mags, fmin, fmax):
    fig, ax = plt.subplots(figsize = (10.0,7.5))
    ax.plot(1e-3*freqs, 20*np.log10(mags), 'b')
    ax.set_xlim(fmin, fmax)
    ax.set_xlabel('Frequency [kHz]')
    ax.set_ylabel('Magnitude [dB]')
    ax.grid()
```

Lecture 6 - Nonlinearity and Distortion

Nonlinear characteristics of semiconductor devices

- Semiconductor devices (i.e. diodes and transistors), like most physical systems, are by nature nonlinear in their operation
- For example, the collector current of a BJT depends exponentially on the base-emitter voltage

$$I_C = I_S \exp\left(\frac{V_{be}}{V_T}\right) \tag{1}$$

- Similarly, the drain current of a MOSFET in strong inversion depends quadratically on the gate-source voltage:

$$I_D = \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{gs} - V_{th})^2 \tag{2}$$

- In analog design, we often assume "small-signal" operation, and linearize our model about an operating point
- What are the bounds of the "small-signal approximation," and what happens when we exceed them?

Taylor series expansion

- A Taylor series is the expression of a function as an infinite sum of its derivatives about a single point
- Given a function $f(x)$ and a point $x = a$ (and assuming certain requirements like infinite differentiability at a), $f(x)$ can be expressed as

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots \tag{3}$$

$$= \sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(x-a)^n \tag{4}$$

- The series is typically truncated using an appropriate number of terms to express the function as a finite polynomial
- The McLaurin series is a special case of the Taylor series, i.e. that of $a = 0$

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots \tag{5}$$

MOSFET nonlinearity

- We can evaluate the nonlinearity of the MOSFET by expanding the drain current equation and assessing the magnitude of the nonlinear term(s)

$$I_d(v_{gs}) = \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS} - V_{th} + v_{gs})^2 = \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{ov} + v_{gs})^2 \tag{6}$$

$$= \frac{1}{2}\mu C_{ox}\frac{W}{L}[V_{ov}^2 + 2V_{ov}v_{gs} + v_{gs}^2] \tag{7}$$

$$= \frac{1}{2}\mu C_{ox}\frac{W}{L}V_{ov}^2 + \mu C_{ox}\frac{W}{L}V_{ov}v_{gs} + \frac{1}{2}\mu C_{ox}\frac{W}{L}v_{gs}^2 \tag{8}$$

$$= I_{D0}\left(1 + \frac{2}{V_{ov}}v_{gs} + \frac{1}{V_{ov}^2}v_{gs}^2\right) \tag{9}$$

$$= a_0 + a_1 v_{gs} + a_2 v_{gs}^2 \tag{10}$$

- In this case obtaining the Taylor series is straightforward, since the drain current expression is already in the form of a polynomial
- Here, $a_0 = I_{D0}$ (DC current), $a_1 = g_m$ (transconductance), and a_2 is the second-order nonlinear coefficient

BJT nonlinearity

- The nonlinearity of the BJT can be evaluated with the aid of a McLaurin series expansion of its characteristic about an operating point
- First express the collector current as a combination of bias and small-signal terms

$$I_c = I_S \exp\left(\frac{V_{BE0} + v_{be}}{V_T}\right) = I_S \exp\left(\frac{V_{BE0}}{V_T}\right)\exp\left(\frac{v_{be}}{V_T}\right) = I_{C0}\exp\left(\frac{v_{be}}{V_T}\right) \tag{11}$$

- Then use the McLaurin series to determine coefficients

$$I_c(v_{be}) = I_{C0}\left(1 + \frac{1}{V_T}v_{be} + \frac{1}{2V_T^2}v_{be}^2 + \frac{1}{6V_T^3}v_{be}^3 + \cdots\right) \tag{12}$$

$$I_c(v_{be}) = a_0 + a_1 v_{be} + a_2 v_{be}^2 + a_3 v_{be}^3 + \cdots \tag{13}$$

- Here again, a_1 represents the DC bias current I_{C0}, a_1 the transconductance g_m, and a_2 and a_3 the second- and third-order nonlinear coeffcients

Common-emitter nonlinearity

- Given the nonlinear relationship between the base-emitter voltage and the collector current of the BJT, we can determine a nonlinear expression for the output voltage of a common-emitter amplifier with load resistance R_C

$$V_o = V_{CC} - I_c(v_i)R_C \tag{14}$$

$$= V_{CC} - [a_o + a_1 v_i + a_2 v_i^2 + a_3 v_i^3 + \cdots]R_C \tag{15}$$

- Substituting in the expressions for a_0, a_1, $\cdots$ gives

$$V_o = V_{CC} - I_c(v_i)R_C \tag{16}$$

$$= V_{CC} - I_{C0}\left[1 + \frac{1}{V_T}v_i + \frac{v_i}{2V_T^2}v_i + \frac{v_i^2}{6V_T^3}v_i + \cdots\right]R_C \tag{17}$$

- While the linear (transconductance) term is independent of the amplitude of v_i, the higher-order terms aren't (definition of nonlinearity)
- If we ignore the higher-order terms we obtain the small-signal approximation

$$V_o \approx V_{CC} - I_{C0}R_C - \frac{I_{C0}}{V_T}R_C v_i = V_{CC} - V_0 - g_m v_i R_C \tag{18}$$

Harmonic distortion

- An amplifier's nonlinearity is commonly specified by *harmonic distortion*, which can be assessed by determining the amplitudes of frequency components at integer multiples (i.e. harmonics) of the input frequency $(2\omega_0, 3\omega_0, \cdots)$
- Given a nonlinear transfer function expressed as

$$v_{out} = a_0 + a_1 v_{in} + a_2 v_{in}^2 + a_3 v_{in}^3 + \cdots \tag{19}$$

- The output voltage given the nonlinear gain characteristic for an input signal given by $v_{in} = A\sin\omega_0 t$ is given approximately by

$$v_{out} \approx a_0 + a_1 A\sin\omega_0 t + a_2 A^2 \sin^2\omega_0 t + a_3 A^3 \sin^3\omega_0 t \tag{20}$$

- Here we have ignored higher order terms above the $\sin^3$ terms, though these may have an impact depending on the system under consideration

- The output voltage can be expressed as

$$v_{out} \approx a_0 + a_1 A\sin\omega_0 t + \frac{a_2 A^2}{2}(1 - \cos 2\omega_0 t) + \frac{a_3 A^3}{4}(3\sin\omega_0 t - \sin 3\omega_0 t) \tag{21}$$

- Combining terms with like frequencies gives

$$v_{out} \approx a_0 + \frac{a_2 A^2}{2} + \left(a_1 A + \frac{3a_3 A^3}{4}\right)\sin\omega_0 t - \frac{a_2 A^2}{2}\cos 2\omega_0 t - \frac{a_3 A^3}{4}\sin 3\omega_0 t \tag{22}$$

- Second-harmonic distortion is defined as the ratio of the amplitude of the output component at $2\omega_0$ to the amplitude of the first harmonic (or fundamental) at ω_0. Assuming $a_1 A \gg a_3 A^3/4$,

$$HD_2 = \frac{a_2 A^2}{2}\frac{1}{a_1 A} = \frac{1}{2}\frac{a_2}{a_1}A \tag{23}$$

- Thus, HD_2 varies linearly with the input signal amplitude

- Similarly, third-harmonic distortion is given by

$$HD_3 = \frac{a_3 A^3}{4}\frac{1}{a_1 A} = \frac{1}{4}\frac{a_3}{a_1}A^2 \tag{24}$$

- So, HD_3 varies quadratically with signal amplitude, meaning that very small amplitude inputs produces little third harmonic distortion, but this grows substantially with increasingly large amplitude signals
- HD_2 and HD_3 are essentially expressions of the second and third harmonic amplitudes as percentages of the fundamental

Common-emitter distortion

- The output voltage of the common-emitter amplifier given a sinusoidal input v_i is again given by

$$V_o = V_{CC} - I_c(v_i)R_C \tag{25}$$

$$= V_{CC} - I_{C0}\left[1 + \frac{1}{V_T}v_i + \frac{1}{2V_T^2}v_i^2 + \frac{1}{6V_T^3}v_i^3 + \cdots\right]R_C \tag{26}$$

- Here, $a_1 = g_m R_C$ and $a_2 = I_{C0}R_C/2V_T^2$. HD_2 is thus given by

$$HD_2 = \frac{1}{2}\frac{a_2}{a_1}A = \frac{1}{4}\frac{1}{V_T}A \tag{27}$$

- That is, if the input signal amplitude is equal to V_T the ratio of the fundamental to second harmonic is about $1/4$, or $-12dB$
- From this it is clear that the input signal amplitude should be kept well below V_T if we want to minimize distortion

- For small input amplitudes the second harmonic will dominate distortion due to its dependence on the square of the amplitude
- However, as the input amplitude increases the higher-order terms will contribute substantially to the total distortion
- Taking the common-emitter amplifier as an example, for an input signal with amplitude $V_T/2$, the second and third distortion components are given by

$$HD_2 = \frac{1}{4}\frac{1}{V_T}\frac{V_T}{2} = \frac{1}{8} \qquad HD_3 = \frac{1}{24}\frac{1}{V_T^2}\frac{V_T^2}{4} = \frac{1}{96} \tag{28}$$

- However, for an input amplitude of $2V_T$, the situation is somewhat different

$$HD_2 = \frac{1}{4}\frac{1}{V_T}2V_T = \frac{1}{2} \qquad HD_3 = \frac{1}{24}\frac{1}{V_T^2}4V_T^2 = \frac{1}{6} \tag{29}$$

Common-source distortion

- As a comparison, distortion in a common-source amplifier is only second order (assuming strong inversion)
- Second-harmonic distortion in a common-source amplifier is given by

$$HD_2 = \frac{1}{2}\frac{a_2}{a_1}A = \frac{1}{4}\frac{1}{V_{ov}}A \tag{30}$$

- This is similar to the HD_2 for a common-emitter amplifier, except that it is inversely proportional to V_{ov} instead of V_T
- This reveals a tradeoff between (transconductance) gain and linearity in the common source amplifier, along with the opportunity to improve linearity by reducing g_m (any potential issues with this?)

Differential pair distortion

- The nonlinearity of a differential pair can be assessed using the Taylor series expansion of the hyperbolic tangent function:

$$V_{od} = V_{op} - V_{om} = \alpha I_{TAIL}R_C \tanh\left(\frac{-V_{id}}{2V_T}\right) \tag{31}$$

$$= \alpha I_{TAIL}R_C\left[\frac{V_{id}}{2V_T} - \frac{1}{3}\left(\frac{V_{id}}{2V_T}\right)^3 + \frac{2}{15}\left(\frac{V_{id}}{2V_T}\right)^5 - \frac{17}{315}\left(\frac{V_{id}}{2V_T}\right)^7 - \cdots\right] \tag{32}$$

- An interesting aspect of this expression is the absence of even-order terms, which is due to the symmetry of the the differential amplifier and the fact that even-order harmonics are effectively common-mode
- Third harmonic distortion in the BJT differential pair is given by

$$HD_3 = \frac{1}{4}\frac{a_3}{a_1}V_{id}^2 = \frac{1}{4}\frac{2V_T}{24V_T^3}V_{id}^2 = \frac{V_{id}^2}{48V_T^2} \tag{33}$$

- How much distortion is present when the input signal equals, say, V_T?
- For $V_{id} = V_T$ the third harmonic distortion is

$$HD_3 = \frac{V_{id}^2}{48V_T^2} = \frac{V_T^2}{48V_T^2} \approx 0.02 \tag{34}$$

- This is the amplitude of the third harmonic component of the output signal relative to that of the fundamental, which equates to a distortion percentage of about 2%

Intermodulation distortion

- Distortion can also result from the interaction between signals at different frequencies, a phenomenon generally referred to as *intermodulation distortion*, important in communication systems
- Consider the input to a nonlinear amplifier as the sum of two sinusoidal components

$$v_{in} = A[\cos(\omega_1 t) + \cos(\omega_2 t)] \tag{35}$$

- The amplifier output (current) can be described as

$$i_{out} \approx c_0 + c_1 v_{in} + c_2 v_{in}^2 + c_3 v_{in}^3 \tag{36}$$

- Combining the above equations enables expression of the fundamental and DC components of the output as

$$[c_0 + c_2 A^2] + [c_1 A + \tfrac{9}{4}c_3 A^3][\cos(\omega_1 t) + \cos(\omega_2 t)] \tag{37}$$

- From this expression we see that the quadratic term (c_2) contributes a DC term that adds to the output bias, and that the cubic factor (c_3) contributes to distortion of the fundamental component

- The second and third harmonic terms are expressed as

$$(\tfrac{1}{2}c_2 A^2)[\cos(2\omega_1 t) + \cos(2\omega_2 t)] + (\tfrac{3}{4}c_3 A^3)[\cos(3\omega_1 t) + \cos(3\omega_2 t)] \tag{38}$$

- Applying appropriate trigonometric relations to the third harmonic term results in

$$(\tfrac{3}{4}c_3 A^3)[\cos(\omega_1 + 2\omega_2)t + \cos(\omega_1 - 2\omega_2)t + \cos(2\omega_1 + \omega_2)t + \cos(2\omega_1 - \omega_2)t] \tag{39}$$

- The sum terms in the above expression are typically out of band and thus attenuated, but the difference terms can be significant, particularly in the case where ω_1 and ω_2 are close in frequency (as with an interferer in an adjacent channel in radio-frequency applications)
- The input amplitude at which the amplitude of the intermodulation difference terms ($\frac{3}{4}c_3 A^3$)equal that of the fundamental ($c_1 A$) is given by

$$A^2 = \frac{4}{3}\left|\frac{c_1}{c_3}\right| \rightarrow A = \sqrt{\frac{4}{3}\left|\frac{c_1}{c_3}\right|} \tag{40}$$

Negative feedback

- Negative feedback is used to reduce the sensitivity of *closed-loop* gain to variability in open-loop gain due to, for example, temperature and manufacturing variations
- Feedback also reduces distortion by reducing the amplitude of the signal, s_e, applied to the open-loop gain element A
- This signal is effectively an *error* signal, the difference between the actual output signal and the desired one

- Taking the expression for the closed-loop gain to be $G = A/(1 + \beta A)$, we can differentiate with respect to deviations in the open-loop gain dA

$$\frac{dG}{dA} = \frac{(1 + \beta A) - \beta A}{(1 + \beta A)^2} = \frac{1}{(1 + \beta A)^2} \tag{41}$$

- If A changes by δA for a fractional change of $\delta A/A$ the resulting change in G is

$$\delta G = \frac{\delta A}{(1 + \beta A)^2} \tag{42}$$

- The fractional change in G is

$$\frac{\delta G}{G} = \frac{1 + \beta A}{A}\frac{\delta A}{(1 + \beta A)^2} = \frac{\frac{\delta A}{A}}{1 + \beta A} \tag{43}$$

- That is, the sensitivity of the *closed-loop* gain to fractional changes in the *open-loop* gain is reduced by the factor $1 + \beta A$

Non-inverting amplifier

- In the context of the standard non-inverting amplifier, the improved linearity can be understood as the minimization of the *error voltage* given by $v_e = v^+ - v^-$
- In the closed loop, the error voltage can be approximated as

$$v_e = v^+ - v^- \approx V_i - \beta V_o = V_i - \frac{\beta A}{1 + \beta A}V_i = \boxed{\frac{V_i}{1 + \beta A}} \tag{44}$$

- That is, feedback reduces distortion by minimizing the signal applied to the nonlinear open-loop element A

Emitter degeneration

- In the case of the emitter-degenerated common-emitter amplifier the error signal is the base-emitter voltage, while the output signal is the collector current

$$v_{err} = v_{be} \approx \frac{v_{in}}{1 + g_m R_E} \tag{45}$$

- If $g_m R_E \gg 1$, the base-emitter voltage seen by the transistor will be much smaller than the actual input signal
- Assuming the product $g_m R_E$ is relatively constant, the collector current distortion can be approximated as

$$I_c \approx I_{C0}\left[1 + \frac{1}{V_T}\frac{v_i}{1 + g_m R_E} + \frac{1}{2V_T^2}\frac{v_i^2}{(1 + g_m R_E)^2} + \frac{1}{6V_T^3}\frac{v_i^3}{(1 + g_m R_E)^3} + \cdots\right] \tag{46}$$

$$\tag{47}$$

$$\approx c_0' + c_1' v_{in} + c_2' v_{in}^2 + c_3' v_{in}^3 + \cdots \tag{48}$$

Total harmonic distortion

- *Total harmonic distortion* (THD) is a measure of, you guessed it, the total harmonic distortion in a signal relative to the fundamental
- Strictly speaking, THD is the RMS sum of all the harmonic components present in a signal (over the signal bandwidth), relative to the RMS signal level
- Given an amplifier transfer function described as $a_0 + a_1 v_{in} + a_2 v_{in}^2 + \cdots$, and assuming the only distortion present is that due to the intrinsic nonlinearity of the amplifier, the THD of the output voltage can be expressed as

$$THD = \frac{\sqrt{(\frac{a_2 v_{in}^2}{2})^2 + (\frac{a_3 v_{in}^3}{3})^2 + \cdots}}{a_1 \cdot v_{in}} \tag{49}$$

- In the case of the common-emitter amplifier, the collector current THD is given by

$$THD = \frac{\sqrt{(\frac{g_m v_{in}^2}{4V_T})^2 + (\frac{g_m v_{in}^3}{24V_T^2})^2 + \cdots}}{g_m \cdot v_{in}} \tag{50}$$

- For reasonably small input amplitudes the THD is limited by the second harmonic term, and the THD can be approximated as

$$THD \approx \frac{v_{in}}{4V_T} \tag{51}$$

- For moderate signal amplitudes, the input amplitude that results in a specified level of THD in the common-emitter amplifier can be determined by

$$v_{in,max} = THD_{max} \cdot 4V_T \tag{52}$$

Dynamic range

- Whereas noise limits amplifier performance for small input signals, low distortion is critical for performance at large signal amplitudes
- The "space between" the two extremes is often referred to as the *dynamic range*, which is (roughly) defined as the ratio of the largest to smallest "useful" signal levels
- Different metrics are used to quantify dynamic range for different applications, one of which is the ratio of signal amplitude for a specified maximum level of THD to RMS noise:

$$DR = \frac{v_{in,max,rms}}{v_{n,rms}} \tag{53}$$

- Assuming shot noise only and a noise bandwidth of f_{enb}, the RMS input-referred noise of a common-emitter amplifier is

$$v_{n,rms} = \sqrt{\frac{2kT}{g_m}f_{enb}} \tag{54}$$

- For a maximum distortion of 5% (assuming only second order distortion is significant), the maximum signal amplitude is

$$v_{in,max} = 2 \cdot 0.05 \cdot \frac{a_1}{a_2} = 4V_T \cdot 0.05 \approx 5mV \tag{55}$$

- Taking the ratio of the maximum (set by linearity) to minimum (set by noise) signal amplitudes gives us the dynamic range

$$DR = \frac{0.2V_T/\sqrt{2}}{\sqrt{\frac{2kT}{g_m}f_{enb}}} \tag{56}$$

- For example, assuming a signal bandwidth of $10kHz$, and a collector current of $1mA$, the dynamic range is ~$98dB$ (approximately 16 bits ENOB, or effective number of bits)

- For a common-source amplifier, assuming only drain current thermal noise over a bandwidth of f_{enb} we have

$$v_{n,rms} = \sqrt{\frac{4kT\gamma}{g_m}f_{enb}} \tag{57}$$

- For a maximum distortion of 5% and assuming $g_m/I_D = 10$ ($g_m = 2I_D/V_{ov}$), the maximum signal amplitude is

$$v_{in,max} = 2 \cdot 0.05 \cdot \frac{a_1}{a_2} = 4V_{ov} \cdot 0.05 = 800mV \cdot 0.05 = 40mV \tag{58}$$

- The dynamic range is thus

$$DR = \frac{0.2V_{ov}/\sqrt{2}}{\sqrt{\frac{4kT\gamma}{g_m}f_{enb}}} \tag{59}$$

- Again assuming a signal bandwidth of $10kHz$ and a drain current of $1mA$ (and a γ value of $2/3$), the dynamic range is ~$108dB$ (about 18 bits)

Opamp slew rate distortion

- For small input/output signals an opamp's behavior is assumed to be linear, exhibiting a settling characteristic in response to a step input given by

$$V_{out}(t) = G_{CL}\left[1 - \exp\left(\frac{t}{\tau_{CL}}\right)\right]V_{step}(t) \tag{60}$$

- As an example, a non-inverting amplifier has closed-loop gain and time constant given by

$$G_{CL} = \frac{A_0}{1 + \beta A_0} \qquad \tau_{CL} \approx \frac{1}{\beta\omega_t} \tag{61}$$

- A_0 is the open-loop gain of the opamp, ω_t is its transit frequency, and β is the feedback factor of the closed-loop amplifier, typically taking a form similar to $\beta = R_2/(R_1 + R_2)$
- However, as the amplitude of the opamp output increases, the rate of change of the output voltage becomes limited by the internal bias currents of the opamp, making the response slew-rate limited

- Opamps can become slew-rate limited even for sinusoidal signals
- The maximum possible signal amplitude an opamp is capable of delivering is limited by the supply rails, i.e. $2V_{max} = V_{CC} - V_{EE}$ (lower than this for opamps without rail-to-rail output capability)
- Assuming a sinusoidal output with amplitude V_{max} and frequency f_{max}, the maximum slope is given by

$$\frac{d}{dt}[V_{max} \cdot \sin(2\pi \cdot f_{max} \cdot t)] = V_{max} \cdot 2\pi \cdot f_{max}\cos(2\pi \cdot f_{max} \cdot t) \tag{62}$$

- This reaches a maximum value when $\cos 2\pi \cdot f_{max} \cdot t = 1$, requiring a slew rate of

$$SR \geq V_{max} \cdot 2\pi \cdot f_{max} \tag{63}$$

- The slew rate requirement scales with the maximum frequency (assuming full-scale signals are expected at this frequency)

- Taking a typical value of V_{max} to be $36V$ ($\pm 18V$), this limits the RMS value of a sinusoidal signal to be $12.7V_{rms}$
- Using the audio frequency range as an example, the maximum required frequency of operation is $20kHz$, and a $12.7V_{rms}$, $20kHz$ sinusoid has a maximum slope of

$$\frac{d}{dt}[18V \cdot \sin(2\pi \cdot 20kHz \cdot t)] = 18V \cdot 2\pi \cdot 20kHz\cos(2\pi \cdot 20kHz \cdot t) \tag{64}$$

- The required opamp slew rate is given by

$$SR \geq 18V \cdot 2\pi \cdot 20kHz = 2.26V/\mu s \tag{65}$$

Summary

- While amplifier noise sets a lower bound on precision for small signals, nonlinearity/distortion limits precision as signal amplitude grows
- Transistors are intrinsically nonlinear in converting voltage to current, and produce significant distortion in open-loop operation
- Distortion increases with input signal amplitude, and can be assessed by determining the ratio of harmonic component amplitudes to that of the fundamental
- Due to symmetry even order harmonics are common-mode and thus suppressed in differential amplifiers
- Total harmonic distortion (THD) is often used as a measure of the distortion in a signal
- Dynamic range is a measure of the useful signal range of an amplifier, and is taken as the ratio of the maximum signal amplitude that results in a specified level of distortion to the amplifier's noise