Effects of Nonlinearity

All real electronic systems, including linear amplifiers, possess nonlinear characteristics. The i/o characteristics of an amplifier can be modeled as

$$y(t) = a_o + a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) + \dots$$

in general, where x(t) is the input and y(t) is output of the amplifier. Here, a_o and $a_1 x(t)$ are the offset and the *linear amplification* terms, respectively. $a_2 x^2(t) + a_3 x^3(t) + \dots$ are the terms modeling the inherent nonlinearity. There are few important outcomes of these nonlinearities in communication systems.

At the Receiver:

1. ***Harmonics:***

If $x(t) = A \cos\omega t$ (a narrowband signal)

$$y(t) = a_o + a_2 A^2/2 + (a_1 A + 3a_3 A^3/4) \cos\omega t + (a_2 A^2/2) \cos 2\omega t + (a_3 A^3/4) \cos 3\omega t + \dots$$

considering only first four terms. Extra signal components at 0, 2ω, 3ω are produced. We are only interested in fundamental term but it now contains a new additive component of amplitude $3a_3A^3/4$. A distortion! a_1A is usually much lager than all other terms.

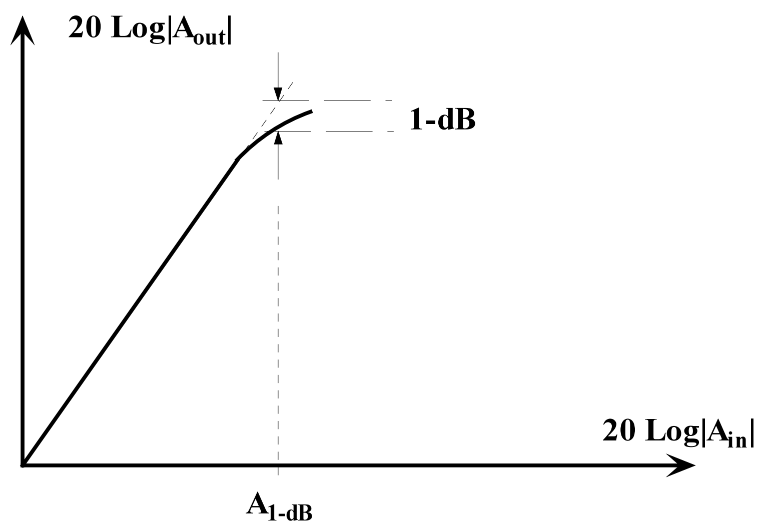
Operation of a service is almost always such that it uses a bandwidth less than an octave:
i.e. $\omega_I < \omega < \omega_{II}$ where $\omega_{II} < 2\omega_I$.

Then all harmonics can be filtered out at the o/p by means of a low pass filter.

2. ***Gain Compression:***

a_3 is usually the dominant nonlinearity term (particularly in "linear amplifiers" or mixers). Therefore the distortion in electronic systems is defined on the effect of third order terms.

If we plot the amplitude of the fundamental component $A_{out} = a_1 A + 3a_3 A^3/4$ wrt $A_{in} = A$, on a log-log scale, we obtain:



1-dB compression point is thus defined as

$20\log | a_1A_{1\text{-dB}} + 3a_3A^3_{1\text{-dB}}/4| = 20\log | a_1A_{1\text{-dB}}| -1$ dB, yielding

$$A_{1\text{-dB}} = \sqrt{0.145\left|\frac{a_1}{a_3}\right|}.$$

In typical front-end RF amplifiers, 1-dB compression is around –20 to –25 dBm i/p signal level.

3. *Desensitization:*

Assume there are two signals in the passband:

$x(t) = A_1\cos\omega_1 t + A_2\cos\omega_2 t$

Now also assume that the wanted signal is $A_1\cos\omega_1 t$ and that $A_1 << A_2$. Then

$y(t) \approx (a_1A_1 + 3a_3A_1^3/4 + 3a_3A_1A_2^2/2)\cos\omega_1 t + \ldots$

Although $a_3 << a_1$, since $A_2 >> A_1$, the distortion term $3a_3A_1A_2^2/2$ can be much larger than a_1A_1. This problem is more aggravated if ω_1 is only slightly different than ω_2 (e.g. channels near by or adjacent channel)

RF receiver amplifiers are expected to withstand blocking signals 60-70 dB greater than wanted signal.

4. *Cross Modulation:*

Consider the case with two modulated signals

$x(t) = A_1(t)\cos\omega_1 t + A_2(t)\cos\omega_2 t$

where

$A_1(t) = A_1[1+m_1(t)]$

and

$A_2(t) = A_2[1+m_2(t)]$.

Then

$y(t) \approx \{a_1A_1[1+m_1(t)] + (3a_3A_1^3/4)[1+m_1(t)]^3 + (3a_3A_1A_2^2/2)[1+m_1(t)][1+m_2(t)]^2\}\cos\omega_1 t$

$\approx \{a_1A_1[1+m_1(t)] + (3a_3A_1^3/4)[1+m_1(t)]^3 + (3a_3A_1A_2^2/2)[1+m_1(t)][1+m_2(t)]^2\}\cos\omega_1 t$

Hence the signal at ω_1 contains modulation components of $m_1(t)^2$, $m_1(t)^3$, $m_2(t)$, $m_1(t)m_2(t)$, $m_2(t)^2$ and $m_2(t)^3$. Again, depending on the relative values of a_1 and a_3 and/or A_1 and A_3, these distortion components can be very large.

If the bandwidth allocated to a single channel is B Hz, then we can assume that m(t) has a the same bandwidth at most. Then any second and higher order term occupies a bandwidth more than B Hz. This effect aggravates interference, particularly adjacent channel interference (ACI).

5. *Intermodulation:*

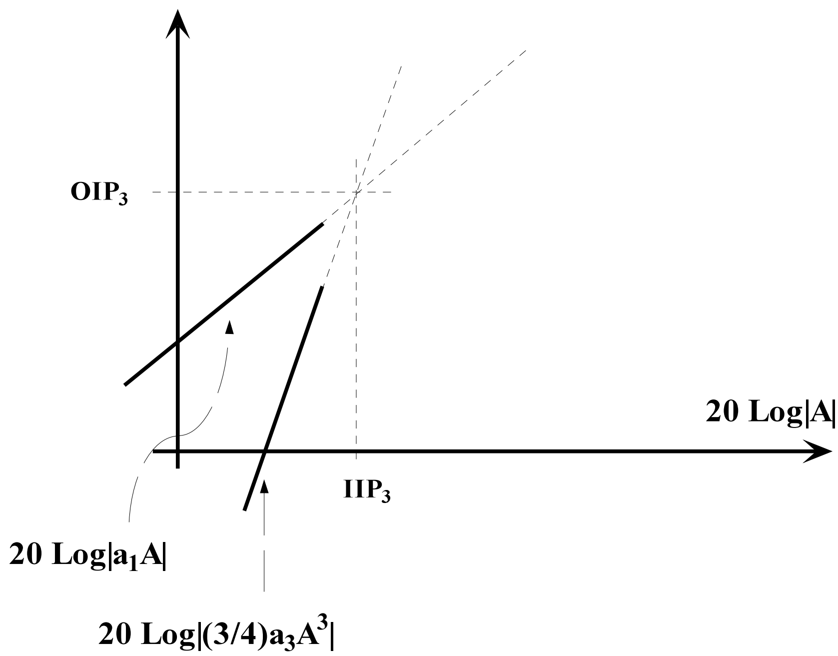
When two signals of different frequency, ω_1 and ω_2, are applied to input of a system with nonlinearity, output contains components that are not harmonics of ω_1 or ω_2. This is called intermodulation (IM):

$$x(t) = A_1\cos\omega_1 t + A_2\cos\omega_2 t$$

The output contains IM products at

$\omega = \omega_1 \pm \omega_2$	:	$a_2A_1A_2\cos(\omega_1 + \omega_2)t + a_2A_1A_2\cos(\omega_1 - \omega_2)t$
$\omega = 2\omega_1 \pm \omega_2$	:	$(3/4)a_3A_1^2A_2\cos(2\omega_1 + \omega_2)t + (3/4)a_3A_1^2A_2\cos(2\omega_1 - \omega_2)t$
$\omega = 2\omega_2 \pm \omega_1$	:	$(3/4)a_3A_1A_2^2\cos(2\omega_2 + \omega_1)t + (3/4)a_3A_1A_2^2\cos(2\omega_2 - \omega_1)t$
$\omega = \omega_1$	:	$[a_1A_1 + (3/4)a_3A_1^3 + (3/2)a_3A_1A_2^2]\cos\omega_1 t$
$\omega = \omega_2$	:	$[a_1A_2 + (3/4)a_3A_2^3 + (3/2)a_3A_1^2A_2]\cos\omega_2 t$

Components at $2\omega_1 \pm \omega_2$ and $2\omega_2 \pm \omega_1$ are important, because they may be close to ω_1 and/or ω_2 (if ω_1 and ω_2 are close to each other). This type of distortion/interference is defined for $A_1=A_2=A$, as follows:



IP3 is the "third order intercept point". IIP3 is "input IP3" and OIP3 is the "output IP3". IIP3 is the value of the input amplitude where

$|a_1|A_{IP3} = (3/4)|a_3|A_{IP3}^3$
i.e.

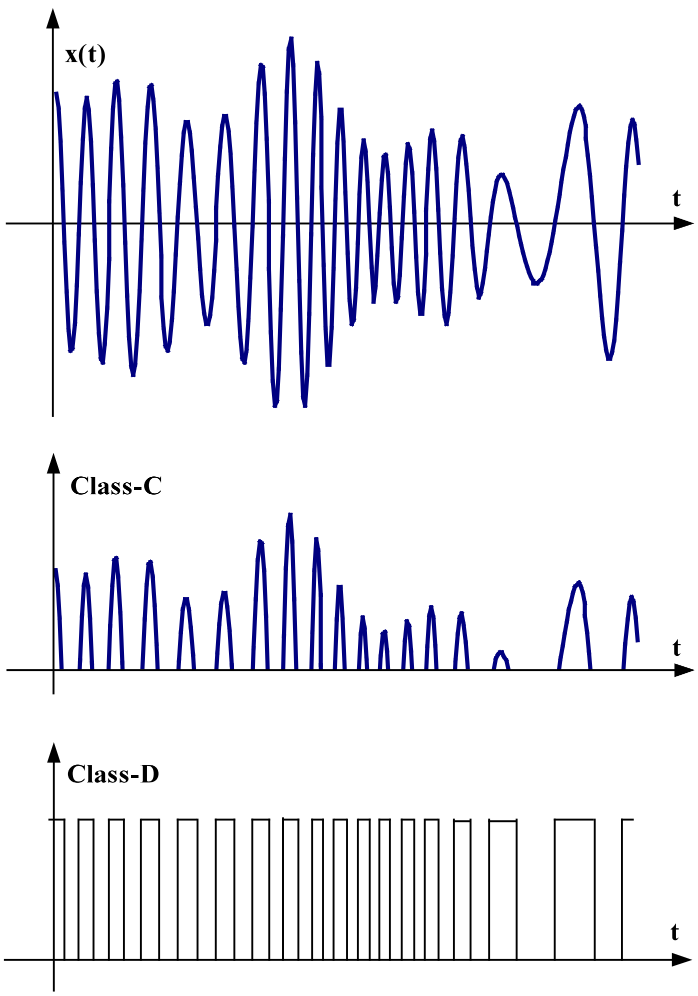
$$A_{IP3} = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|} \text{ and } OIP3 = |a_1|A_{IP3}.$$

Why constant amplitude modulation techniques in mobile communications?

Class-C or Class-D amplifiers are preferred as the RF output amplifier in mobile communications, because these amplifiers have better efficiency and thus lower energy consumption. For a given general narrowband input signal

$$x(t) = A(t)\cos[\omega_o t + \theta(t)]$$

Class-C amplifier amplifies only for a fraction of the period, less than 50%, determined by ω_o. Class-D amplifiers produce a hard limited version of input signal. The input signal and respective amplifier output signals are depicted in the figure below.



The nonlinearity in Class-C and Class-D amplifiers are severe. Coefficients $|a_2|$ and $|a_3|$ can be as large as $|a_1|$, sometimes larger than $|a_1|$. The relevant part of output becomes

$$y(t) \approx a_2 A(t)^2/2 + [a_1 A(t) + 3a_3 A(t)^3/4]\cos[\omega_o t + \theta(t)] + [a_2 A(t)^2/2]\cos[2\omega_o t + 2\theta(t)] + [a_3 A(t)^3/4]\cos[3\omega_o t + 3\theta(t)] + \ldots$$

The output signals of these amplifiers are filtered so that all harmonics are eliminated. Hence the signal that reaches to the antenna is only

$$y(t) \approx [a_1 A(t) + 3a_3 A(t)^3/4]\cos[\omega_o t + \theta(t)]$$

$$\approx a_1[A(t) + 3A(t)^3/4]\cos[\omega_o t + \theta(t)],$$

assuming $a_2 \approx a_3 \approx a_1$.

Let us consider two cases

1. Amplitude modulation: $\theta(t) = \theta_o$ a constant phase:

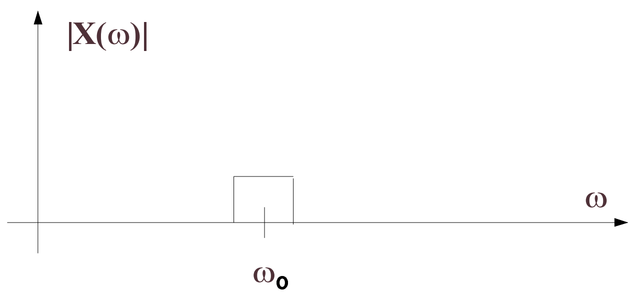
The output

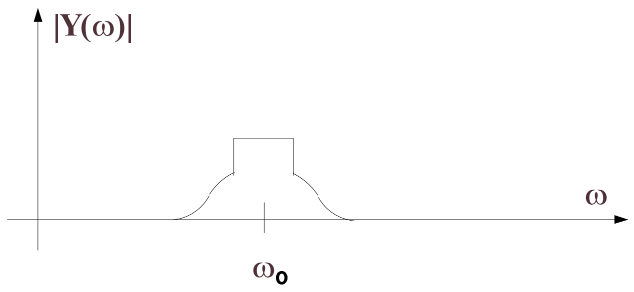
$$y(t) \approx a_1[A(t) + 3A(t)^3/4]\cos[\omega_o t + \theta_o]$$

is severely distorted compared to the input

$$x(t) = A(t)\cos[\omega_o t + \theta_o].$$

Also, the term $A(t)^3$ occupies a bandwidth three times as large as $A(t)$, which also creates interference in near by channels, as depicted in the following figure.





Hence it is not possible to use these types of amplifiers in systems which employ modulation schemes where the information is mounted on the envelope.

2. Frequency modulation: $A(t)= A_o$ a constant amplitude:

As far as the information content is concerned, the output signal

$y(t) \approx a_1 [A_o+ 3A_o^3/4] \cos[\omega_o t+\theta(t)]$

and the input

$x(t) = A_o \cos[\omega_o t +\theta(t)]$

are exactly the same. Information is contained in $\theta(t)$ and it is well preserved. The amplitude is altered, but it is again constant and does not cause any interference in adjacent channel.