# Basic concepts

## Big picture

How have we been able to make cell phones more affordable while their functionality has exploded (cellular, WiFi, Bluetooth, GPS, computing, storage, digital camera, user-friendly interface)?

* Integration: how much functionality can be placed on a single chip (or how few components are left off-chip)
* Integration is due to (1) scaling of VLSI processes – CMOS technology – (2) innovations in RF architectures, circuits, devices.

These disciplines are all required, to some degree, for an RF designer:

* Communication theory
* Random signals
* Transceiver architectures
* IC design
* CAD tools
* Wireless standards
* Multiple access
* Signal propagation
* Microwave theory

RF design hexagon:

* Noise
* Power
* Frequency
* Gain
* Supply voltage
* Linearity
* … back to noise

Each metric trades off with its two adjacent metrics. For example, to lower the noise of a front-end amplifier, we must consume more power or sacrifice linearity.

Generic RF transceiver architecture:

* Receiver: antenna 🡪 LNA 🡪 downconverter (driven by oscillator, generated by frequency synthesizer) 🡪 ADC 🡪 digital baseband processor
* Transmitter: digital baseband processor 🡪 DAC 🡪 upconverter 🡪 PA 🡪 antenna

## Units in RF design

Voltage gain:

Power gain:

**Voltage and power gain are equal in dB if and only if the input and output voltages appear across equal impedances.** For example, the gain of an amplifier with an input resistance of and driving a load of is

and are rms values.

Powers are expressed in dBm:

For example, if we deliver a power of 0dBm across a 50Ohm load for a sinusoidal signal, what is the peak-to-peak voltage swing?

Another example, a GSM receiver senses a narrowband modulated signal having a level of -100dBm. If the front-end amplifier has a voltage gain of 15dB, what is the peak-to-peak voltage swing at the amplifier output?

We assumed that

* The input impedance of the front-end amplifier is 50Ohm
* A narrowband signal has roughly the same peak-to-average power relationship as a sinusoid

In most integrated RF systems, we prefer voltage quantities to power quantities since

* Input and output impedances of cascade stages may be unequal, so voltage gain and power gain are not equal
* Impedances may be largely capacitive or inductive, in which case there is no “real” (active) power

However, we still sometimes use dBm at interfaces that do not necessarily entail power transfer. If we drive a purely-capacitive load, the delivered average power is zero, but we can still calculate dBm as if we were driving a 50Ohm load with our voltage signal.

## Time variance

A system is linear if and only if it satisfies the principle of superposition.

Systems with nonzero initial conditions or DC offsets are technically nonlinear, but we often relax the rule to accommodate these two effects (in this case, the system is incrementally linear).

A system is time invariant if a time-shift in the input causes the same time-shift in the output.

A system that changes with time is time variant.

Take the ideal switch for example. Let drive the control s.t. the switch is on if , and let drive the input.

If we look at as the input and as part of the system, then the system is both nonlinear and time variant. The output is independent of the amplitude of , and the system varies over time with .

If we look at as the input and as part of the system, then the system is linear but time variant. The output scales with , but the system varies over time with . In this case, the input-output relationship is

is a square wave toggling between 0 and 1 with frequency .

The output spectrum consists of copies of at

**A linear system can generate frequency components that don’t exist in the input signal if the system is time variant.**

## Nonlinearity

A system is “memoryless” or “static” if its output does not depend on past values of the input and/or output.

A memoryless linear system is given by

A memoryless nonlinear system is given by

are functions of time if the system is time variant.

When for even , the nonlinear system has odd symmetry, which means that is an odd function of , i.e.

This kind of system is also called balanced. One example: a differential pair.

A system is “memory” or “dynamic” if its output depends on past values of input and/or output.

An LTI dynamic system is represented by

This is the convolution integral.

If a dynamic system is linear but time variant, its impulse response depends on the time origin. Then

If a system is both nonlinear and dynamic (what about time variant?), then its impulse response can be approximated by a Volterra series.

**Example: square-law MOS transistor operating in saturation**

For a differential pair, the transfer curve is

Factoring out from the square root and assuming , and using , we get

The first term is the small signal gain . Due to symmetry, there is no even-order nonlinearity. Square-law devices generate a third-order term.

## Effects of memoryless nonlinearity

Model:

is the small signal gain of the system.

We will analyze the effect of nonlinearity on sinusoidal inputs.

### Harmonic distortion (signal only)

Let the input be a single real sinusoid.

Even-order nonlinearity introduces the DC offset. Ideally, even-order nonlinearity vanishes in balanced circuits, but random mismatches corrupt the symmetry, yielding finite even-order harmonics.

th-order harmonic amplitude is proportional to .

**RF harmonics are typically less critical because they fall way outside the frequencies of interest.**

**However, you need to be careful if mixing occurs.**

For example, let’s say are the two inputs to an analog multiplier (mixer). The ideal multiplier is modeled as

is a constant.

Let .

The ideal output has frequency components .

What happens if experiences third-order nonlinearity at the input?

Then the output will have spurious frequency components . Let . Let’s say the desired output frequency component is . , which can fall inband – this is a problem.

### Gain compression (signal + 1 inteferer)

Nonlinearity compresses small signal gain, as seen in the fundamental component

Where and have opposite signs for compressive behavior.

The 1dB compression point – the point at which gain drops by 1dB – is one standard way of characterizing nonlinearity.

In general,

Note that is the peak value of the sinusoidal input.

**Often, gain compression is not dominated by your desired signal but by a large interferer.**

Let

If you multiply out , you will get a term .

Then the fundamental is

Since ,

As interferer amplitude increases, gain decreases and can even become 0.

### Cross modulation (signal + 1 inteferer)

What if the interferer has amplitude modulation, i.e. ?

is a constant. Then

The desired signal suffers from amplitude modulation at and .

Cross modulation commonly occurs in systems that must simultaneously process multiple independent channels.

**In a memoryless nonlinear system, cross modulation does not occur if the interferer has phase modulation, e.g.**.

### Intermodulation (signal + 2 interferers)

The two interferers mix inside the nonlinear system.

If are close in frequency, then the 3rd-order products at are close to .

Let’s say we have a signal at , and it happens that . The IM3 product falls directly on the desired signal and corrupts it.

**For IM3, we typically model narrowband signals by condensing them into unmodulated tones.**

**Even if gain is not significantly compressed, IM3 products may still severely corrupt the desired signal.**

Example: we have a LNA with a gain of 10 and input impedance of 50Ohms. The LNA senses a desired signal level of -80dBm @ 2.41GHz and two interferers of equal power at 2.42 and 2.43GHz. Assume the LNA drives a 50Ohm load.

What value of yields a 1dB compression point of -30dBm? -30dBm is average power.

If each interferer is 10dB below , what is the corruption of the desired signal?

The IM3 at the LNA output is given by

The signal at the LNA output is , so the IM product is as large as the signal even though the LNA does not experience significant compression.

#### IP3 and two-tone test

The two-tone test is very useful for characterizing the nonlinearity of the system since the frequency difference can be made arbitrarily small to ensure the IM3 product lands inband (unlike HD).

We apply two sinusoids of equal amplitude, representing the interferers, and normalize the IM3 amplitude to the fundamental amplitude at the system output. This gives us relative IM:

When increases by 6dB (2x), IM3 increases by 18dB and relative IM3 increases by 12dB.

IP3 is a measure of IM3 nonlinearity that is independent of . The idea is that as increases, at a (theoretical) point, relative IM3 becomes 0dBc. In terms of amplitude, fundamental increases 20dB/decade, while IM3 increases 60dB/decade.

This point is

is peak input amplitude. At this (theoretical) point, the output amplitude of one IM3 tone is equal to the output amplitude of one fundamental tone.

Note: is a theoretical quantity – it cannot be directly measured. It is 9.6dB higher than , which means we cannot say that the output fundamental is because of gain compression. Furthermore, may be higher than the supply voltage. When measuring , we must ensure we measure in the region where 1dB increase in results in 3dB increase in IM3.

Example: LNA senses -80dBm signal at 2.41GHz and two -20dBm interferers at 2.42 and 2.43GHz. What is the value of IP3 that results in IM3 being 20dB below the desired signal? Assume 50Ohm interfaces.

Typically, we measure IIP3 using extrapolation. We measure IM3 at different input amplitudes and extrapolate out to the point where IM3 and fundamental intersect.

However, there is a shortcut that provides an estimate (true IIP3 may differ if there is dynamic nonlinearity).

Let be the power of one input tone. is the power of one output fundamental tone, and is the power of one output IM product.

Theoretically, every 1dB backoff from equals 2dB backoff in relative IM3.

This equation requires only one measurement.

It’s often useful to calculate the input-referred IM3 level when you know .

### Cascaded IM3

Let’s say you have two cascaded nonlinear stages

Expanding and keeping only first and third-order terms,

Then the IIP3 of the cascade is

At this point, let’s examine how the IM3 are generated.

1. IM3 generated in first stage
2. IM3 generated in second stage
3. First stage generates HD2, then second stage mixes fundamental and HD2 to form IM3.

In real systems, this last term will be very small because (1) 2nd-order nonlinearity is suppressed by odd symmetry (2) HD2 is suppressed in a narrowband system.

Then we can write

That is, referred to the input, is reduced by the gain of the first stage.

For stages,

This means that IIP3 becomes more important as you progress along a chain of amplifiers because the IIP3 of a given stage, referred to the overall input, is scaled down by the amplification of the previous stages.

In this analysis, we assumed that the IM3 products of each stage add in-phase (that is, there is no destructive cancellation). In a real system that has memory, there will be phase shifts that may result in finite cancellation. However, since are similar frequencies, we expect them to experience similar phase shifts.

**Example:** LNA with IIP3 of -10dBm and a gain of 20dB is followed by a mixer with IIP3 of +4dBm. Which stage limits the overall IIP3?

The contribution of the second stage is

The second stage limits IIP3.

### AM/PM conversion (APC) – TBD include first-order example

Phase shift depends on signal amplitude.

APC doesn’t occur in LTI systems because in LTI systems, phase shift is a function only of frequency.

APC doesn’t occur in a memoryless nonlinear system because phase shift is 0.

**Therefore, APC occurs only in a dynamic, nonlinear system.**

## Noise

### Noise as a random process

Noise is random, which means we cannot predict the instantaneous value. However, its statistics are quantifiable.

For example, if is the noise waveform, average noise power is

This is in analogy with periodic power signals. This is energy over time. To measure , depends on the frequencies of interest. must be long enough to capture several cycles of the lowest frequency component. In practice, we make a guess for and measure . Keep increasing until converges.

### Noise spectrum

In time domain, the only measure of noise we can get is average power.

Knowing the noise spectrum – average power for each frequency component – provides much more information.

Finding the spectrum or PSD of any signal (or noise) is useful.

Conceptually, if you want to know the power at frequency , you pass the signal through a 1Hz-bandwidth brickwall filter at and measure average power.

Do this for all frequency components and you get PSD of , , which has units of or . is power in a 1Hz bandwidth for all frequencies. Conceptually, this is how spectrum analyzers work.

In signals and systems, the PSD is defined as the Fourier transform of the autocorrelation of a signal. These two views are equivalent.

(Parseval’s theorem) The total area under represents the average power of :

Note: this definition is for a single-sided PSD and assumes that is real, because if is real, then PSD is symmetric (PSD is always real). If PSD is symmetric, then it can either be defined as one-sided or two-sided. The two-sided PSD is the same as one-sided except scaled by ½ and symmetric around DC.

**Example: a resistor of value generates a noise voltage whose one-sided PSD is given by**

is absolute temperature (which is in Kelvin). is the Boltzmann constant and is equal to 1.38e-23 .

This flat PSD is called white noise because, like white light, it contains all frequencies with equal power levels.

The area under the PSD is infinite, implying that noise power is infinite. In reality,

If you were to integover all frequency, you would get infinite power. In reality, begins to fall off at , i.e. thermal noise is not truly white.

The dimension of is .

In fact, if we denote the noise voltage by , we can write

denotes the average power of in 1Hz (aka spot noise).

Then denotes average noise power in bandwidth.

For a 50Ohm resistor at room temperature, noise power in a 1Hz bandwidth is

Noise voltage is

has no physical meaning.

### Effect of transfer function on noise

Noise is filtered just like any other signal. can be a desired signal, or noise, or interference. If we apply to an LTI system with transfer function , then the output spectrum is

### Device noise

We model the noise of electronic devices by voltage and current sources. This allows us to analyze circuit noise using standard circuit analysis techniques.

#### Thermal noise of resistors

Ambient thermal energy leads to random agitation of charge carriers in resistors and hence noise.

Noise is modeled in 2 ways:

1. Thevenin equivalent: voltage source in series with the resistor with PSD of
2. Norton equivalent: current source in parallel with the resistor with PSD of

If a resistor converts ambient heat to a noise voltage or current, can we extract energy from the resistor? In particular, let’s analyze this circuit.

A diagram of a circuit

AI-generated content may be incorrect.

If and are at the same temperature, no net energy is transferred between them because the power that delivers to is the same as the power that delivers to .

What if is at 0 K? Then draws thermal energy from its environment, converting it to noise and delivering it to .

The power delivered to is