

ECE 4880: RF Systems

Fall 2016

Chapter 4: Noises and Noise Figure

Reading Assignments:

- 1. W. F. Eagan, Practical RF System Design, Wiley, 2003, Chap. 3.
- 2. T. H. Lee, *The Design of CMOS Radio Frequency Integrated Circuits*, 2nd Ed, Cambridge, 2004. Chap. 11

Game Plan for Chap. 4: Gain and Reflection

- 1. Definition in noise factor and noise figure in functional modules
- 2. Noise sources
- 3. Noise calculation in module cascade
- 4. Image noise in mixers
- 5. Constant noise circles in the Smith Chart

4.1 Noise Factor and Noise Figure

We will first treat the noise influence to the signal amplitude in this Chapter. The phase noise originates from the same noise sources to be introduced in this Chapter, but its influence in view of nonlinearity and frequency synthesis will be treated later. The amplitude and phase noises are fundamentally related by the total noise power, which will be clear in later chapters.

To isolate the noise contribution in each functional module and to avoid double-counting in module cascade, the noise factor f (unitless) of a given module is defined by the signal-to-noise ratio (SNR) of the input to output, as shown in Fig. 4.1.

$$f = \frac{SNR_{in}}{SNR_{out}} = \frac{\left(\frac{P_{signal,in}}{P_{noise,in}}\right)}{\left(\frac{P_{signal,out}}{P_{noise,out}}\right)} = \frac{\left(\frac{P_{noise,out}}{P_{noise,in}}\right)}{\left(\frac{P_{signal,out}}{P_{signal,out}}\right)}$$
(4.1)

The noise figure F (in unit of dB) is the decibel expression for the noise factor

$$F \equiv 10\log_{10} f \tag{4.2}$$

The noise factor f and the noise figure F denote the degradation in SNR after the module processing. Notice that although RF engineers like express "ratio" in dB, the noise and signal are combined by addition, not by multiplication. Therefore, for many calculations with noise we need to go back and forth between f and F.

Fig. 4.1. Definition of noise factor (f) and noise figure (F) for an independent module.

4.2 Noise Sources

There are six major noise sources, characterized by the physical effects and the origin. Before we begin, the noise can be described by the "unusable power" AND by the "random noise voltage/current". This implies that for a module with total energy conservation, such as genuine lossless capacitors and inductors where the stored electrostatic and magnetostatic energy can be fully retrieved in later time, there is NO noise. A resistor or transistor, on the other hand, with power dissipated in the thermal bath, will have the noise term. In circuit representation, for a noise RMS voltage source \tilde{e}_n , to deliver the most noise to the load *R*, the noise source will also have a source resistance of *R*, as shown in Fig. 4.2.



Fig. 4.2. The relation between the noise voltage and power.

The noise power and noise RMS voltage is therefore related by:

$$p_{noise} = \frac{\left(\frac{\widetilde{e}_n}{2}\right)^2}{R} = \frac{\widetilde{e}_n^2}{4R}$$
(4.3)

The noises sources have two categories. First, noises from the electronic component in the module include thermal noise, shot noise and Flicker noise, all of which relate to the quantized charge of the conducting electrons. Second, noises from the external ambient to the module under consideration include black-body radiation noise, cosmic microwave background noise, and power line and ground noise.

4.2.1 Thermal noise

Thermal noise is also called the Johnson noise, which originates from the random walk of the quantized electrons in a conductor (a resistor or a MOSFET channel). You may derive that this will be very different noise when the conducting species are ions in the electrolyte, although the noise power will be similar. The thermal noise power is

$$p_{n,thermal} = kTB = \frac{\widetilde{e}_{n,thermal}^2}{4R}$$
(4.4)

where *k* is the Boltzmann constant (J/K), *T* is the temperature (K) and *B* is the noise bandwidth (Hz). At room temperature of 273K, *kT* is about 4×10^{-21} J. For a 1Hz *B*, $p_{n,thermal}$ is 4×10^{-21} W or -174dBm. As thermal noise has a "white background", which means the power density is independent of the frequency, we can estimate that the thermal noise with a 1MHz bandwidth (or between 1.000GHz and 1.001GHz) at room temperature will be -174dBm + 60dB = -114dBm. Based on this rationale, the spectral noise power density $\hat{p}_{n,thermal} = p_{n,thermal} / B$ is also often used, which carries a unit of W/Hz.

From Eq. (4.4), we can also derive the thermal noise in a resistor *R* with an open-circuit RMS thermal voltage noise $\tilde{e}_{n,thermal}$ (in unit of V) by:

$$\tilde{e}_{n,thermal} = \sqrt{4kTRB} \tag{4.5}$$

Similar to the spectral noise power density, we can define the spectral noise voltage density as

$$\hat{e}_{n,thermal} = \tilde{e}_{n,thermal} / \sqrt{B} = \sqrt{4kTR}$$
(4.6)

which carries a unit of $V/(Hz)^{1/2}$. For a 50 Ω resistor, $\hat{e}_{n,thermal}$ is about $1nV/(Hz)^{1/2}$, while for a 1k Ω resistor, $\hat{e}_{n,thermal}$ is about $4nV/(Hz)^{1/2}$ (a good number to remember). At 1k Ω resistor and 1MHz bandwidth, $\tilde{e}_{n,thermal}$ is about $4\mu V$. Notice the unit difference between $\hat{e}_{n,thermal}(V/(Hz)^{1/2})$ and $\tilde{e}_{n,thermal}(V)$. For sure we can transfer Fig. 4.2 to a Norton noise current source as well,

$$\widetilde{i}_{n,thermal} = \sqrt{\frac{4kT}{R}B}$$

$$\widehat{i}_{n,thermal} = \frac{\widetilde{i}_{n,thermal}}{\sqrt{B}} = \sqrt{\frac{4kT}{R}}$$
(4.7)

Notice the unit difference between $\hat{i}_{n,thermal}$ (A) and $\tilde{i}_{n,thermal}$ (A/(Hz)^{1/2}).

4.2.2 Shot noise

Shot noise, also called Schottky noise", originates from quantized electrons going over a potential barrier such as a Schottky diode or a forward-biased p-n junction or in a vacuum tube. The name "shot" originates from the old radio, where the shot noise in the vacuum tube caused audible "shots" (a packet of strong noises on top of the more uniform thermal noise). The phenomenon is similar to sizable floating objects flowing through a narrow channel when the objects seem to go through "shots". The shot noise also has a white spectrum because the arrival time for those "shot of electrons" is random. The shot noise is most often expressed as a noise current source (RMS):

$$\widetilde{i}_{n,shot} = \sqrt{2qI_{DC}B}$$

$$\widehat{i}_{n,shot} = \frac{\widetilde{i}_{n,shot}}{\sqrt{B}} = \sqrt{2qI_{DC}}$$
(4.8)

For $I_{DC} = 1$ mA (a typical transistor current), $\hat{i}_{n,shot}$ is about 18 pA/(Hz)^{1/2}. From the functional dependence, the shot noise can be viewed as disturbances of electron flows over a barrier.

4.2.3 Flicker noise

The flicker noise is also called the 1/f noise or the pink¹ noise. The physical origin is complicated and inconclusive. However, in electronic circuits, this often relates to quantized electrons flowing near surfaces. The larger the surface, the smaller the \tilde{e}_n from the averaging effects. Loosely speaking, flicker noise has an empirical $1/f^{\alpha}$ power spectrum as shown in Fig. 4.3, where α is between 0 and 2 or even higher. The intersection of flicker noise and white noise is called the corner frequency f_c . The higher the f_c , the more flicker noise. As the flicker noise is related to electrons flowing near surfaces, MOSFET (f_c between 10 kHz to 1MHz) often has much higher f_c than BJT (f_c between 10Hz to 1kHz). The 1/f noises cannot be reduced by "averaging" as in the case for white noises, as the power density increases without limit when frequency decreases, and the noise power increases just as fast as the averaging integral.



Fig. 4.3. The pink noise power spectrum on top of the white noise in a log-log plot.

There are several most common showing of Flicker noises in circuits:

1. In thin-film resistors, the Flicker noise can be expressed as:

$$\widetilde{e}_n^2 = \frac{K}{f} \cdot \frac{R_s^2}{A} \cdot V^2 B \tag{4.9}$$

where R_S is the sheet resistance, A is the thin-film resistor area, V is the voltage across the resistor, and K is a technology and material parameter.

¹ "Pink" here refers to the additional red (low frequency) components to a white spectrum. There is a "red noise" as well, which comes from the random Brownian motion. Red noise is confusingly also called "Brown noise", not for color, but for Brownian motion (Brownian motion is named after Robert Brown). Red noise is "redder" than common pink noise of 1/f, and often has a frequency dependence of $1/f^2$. A noise whose power density increases with frequency (say f^2) is called "violet" or "purple noise, which originates from taking differential signals out of sources dominated by white noise.

2. In MOSFET transistors, the first kind of flicker noise has a *1/f* behavior similar to the thin-film resistors (MOSFET can be thought as a variable thin-film resistor):

$$\widetilde{i}_n^2 = \frac{K}{f} \cdot \frac{g_m^2}{WLC_{ax}} \cdot B \tag{4.9}$$

where g_m is the transconductance, W is the transistor width, L is the transistor length, C_{ox} is the unitarea gate capacitance.

- 3. In MOSFET transistors, there is another Flicker component which has a $1/f^2$ behavior. This term is important in digital electronics, as it appears as sudden step-like transitions between discrete voltage and current levels, which originates from the channel carrier being trapped or released from the interface traps. This noise has many names: random telegraph noise (as it is similar to the random noises caused by the static charges near an old telegraph line), popcorn noise, impulse noise, bi-stable noise, and burst noise. As the trapped and released charge is discrete, the electrostatic repulsion shows the two-level characteristics.
- 4. The flicker noise also contributes fundamentally to the phase noise in oscillators, which we will treat later.

4.2.4 Black-body radiation noise

The thermal, shot and Flicker noises are from the internal sources in the RF components. Now we will look at noises from the external factors.

Below 10GHz, the black-body radiation noise is often very small (but large in THz and infrared frequencies), unless we have a very hot source nearby, such as nuclear explosion. The general form of the black-body radiation has a $f^3 \exp(-f/f_0)$ functional form which relates to the quantum mechanics:

$$p_n = \frac{2hf^3}{c^2} \cdot \frac{1}{e^{hf/kT} - 1}$$
(4.10)

where h is the Plank constant and c is the speed of light.



Fig. 4.4. The power density of a black-body radiation.

4.2.4.1 Cosmic microwave background (CMB) noise

Cosmic microwave background (CMB) noise is a signature of the remnant radiation left by the Big Bang. Our universe is sparse in "matters", but between all matters of stars, black holes, etc, there is a microwave glow everywhere at 2.7260±0.0013°K as a black body radiation (with peak radiation frequency at 160.2GHz) at the present time. Notice the number of significant digits that can be picked up by the satellite. This precise measurement makes the "red shift" (expansion of the universe) also possible. Since 1896, this radiation has been predicted to be between 3K and 50K until Penzias and Wilson (Nobel Physics Prize 1978) measured it with a microwave equipment at Bell Labs, Holmdel, New Jersey in 1964. Several Nobel Physics Prizes had been awarded to the perfection of the theory and measurement ever since.



Fig. 4.5. Cosmic microwave background noise fitted at a black-body radiation of 2.7260±0.0013°K at the present time (including red shift). The peak frequency is at 160.2GHz. (a) The COBE (Explorer 66) satellite for the CMB measurement (Nobel Physics Prize 2006); (b) The COBE measurement and the associated black-body radiation fit.

Cosmic microwave background radiation with its measurable red shift is a direct evidence to the Big Bang Theory, and simultaneous validation for the quantum mechanics (for the black-body spectrum). For sure, it has left a philosophical puzzle behind as well. Our universe and the scientific laws as we know today has a starting point (or as S. Hawkings put it: the time has a start). Logic will derive that our "mother nature" is not the ultimate cause or Alpha. Logic will also derive that our universe should have an end (although very far away) according to the 2nd law of thermodynamics and the continuous expansion (with more stars dying and less and less blackhole formation). What is before t = 0? What is behind the 2nd law of thermodynamics to define time? We will leave those to the physicists and philosophers for now. If you are an ECE Ph.D. student, feel a bit at ease that your eventual degree has "Philosophy" in it, especially after you know how CMB can be measured today in specially designed radios.

4.2.5 Power line and ground noise

For many small mobile systems, there is no true "electrical" ground, but just a large decoupling capacitance (often including the casing) between the power and ground lines. Therefore, the load current and the ambient static charge can affect the potentials on the electrical circuits. If we treat the power source or the battery has a source resistance R_s and the decoupling capacitance C_{dec} , it will provide a low-pass filter behavior to the circuit network. A watch battery has R_s around 1 Ω and a small C_{dec} for a Bluetooth unit may be around 0.1mF, which will give a corner frequency only around 10kHz, depending on the load situation.

The nonstationary load current also has a major effect on the power line noise, especially for nearby digital circuits only shielded by a small C_{dec} , which tend to switch all together at the clock edge. Long power line also has an inductive and radiation effect, which further complicates the problem at hand.

Nonvolatile charge in the ambient does not only present a potential shift, but the electrostatic discharge (ESD) gives a broad radio spectrum (fortunately, many of the discharge paths have a very poor antenna), which was used as the first radio transmitter (spark plug) when radio is perceived. Human fingers typically carry a static charge between $1 - 20\mu$ C from tribology, and can interfere your radio transmitter and receiver in many transient ESD forms (in addition to burn your integrated circuits). However, we will leave this topic to your future courses.

4.3 Noise Figure in Module Cascade

To cascade function modules in the transceiver chain, we will first need to isolate the noise contribution in each module. We first consider a single module with a known source and load, which ideally would output a noise level that is the product of the input noise and the gain. This ideal situation will give the noise factor f to be 1 and noise figure F to be 0dB (no additional noise is added by this module). If we compare to the background thermal noise, this model will also has a noise temperature of 0° K.

For sure, no module is that ideal (although passive LC network is pretty close to ideal, i.e., noise temperature of 0°K). We now define the contribution of additional noise power by module k as the difference between its output noise $p_{n,k+1}$ and the assumed known source noise kT_0B multiplied by the module gain g_k (notice T_0 here is not necessarily the room temperature, depending on where your system or the preceding stage collects its first noise source).

$$\Delta p_{n,out,k} = p_{n,k+1} - kT_0 Bg_k = kBT_k g_k \tag{4.11}$$

where T_k is defined as the noise temperature for the module k, and B is again the bandwidth where noise will be collected. If we lump this additional output noise as a source noise at the input of module k, we then have:

$$\Delta p_{n,in,k} = \frac{p_{n,k+1}}{g_k} - kT_0 B = kBT_k$$
(4.12)

The contribution to the noise factor is the ratio to the original known source noise at kT_0B .

$$\Delta f_k = \frac{\Delta p_{n,in,k}}{kT_0 B} = \frac{\frac{p_{n,k+1}}{g_k} - kT_0 B}{kT_0 B} = \frac{T_k}{T_0} = f_k - 1$$
(4.13)

Now at a module cascade, if we are referring to all of gain leading to module *k* to be lumped at the source, we would have:

$$\Delta f_{source,k} = \frac{f_k - 1}{\prod_{i=1}^{k-1} g_i} = \frac{T_k}{T_0 \prod_{i=1}^{k-1} g_i}$$
(4.14)

We can thus define the noise factor for a cascade f_{cas} to be treated as a big functional module as:

$$f_{cas} = f_1 + \sum_{k=2}^{N} \frac{f_k - 1}{\prod_{i=1}^{k-1} g_i}$$
(4.15)

Or for two modules,

$$f_{cas} = f_1 + \frac{f_2 - 1}{g_1} \tag{4.16}$$

The noise temperature in the system is most often expressed as

$$T_{sys} = T_0 + T_{cas} = T_0 + \sum_{k=1}^{N} \frac{T_k}{T_0 \prod_{i=1}^{k-1} g_i}$$
(4.17)

where T_0 is often taken at the temperature of the receiving antenna.

When all of the module noises are lumped to the source (for ease of calculation of SNR and bit error rate in the system), we have achieved the purpose of separating the noise contribution from the signal: now that all modules are free of noise after we combine noise contribution to f_{cas} or T_{sys} . The noise and gain interplay of Eq. (4.15) in the cascade gives a very important principle for gain module cascade: the noise is especially important in the beginning stage as it will be amplified later. For example, consider two amplifiers A₁ and A₂, one with the noise figure $F_1 = 3$ dB and $g_1 = 6$ dB, and the other with $F_2 = 10$ dB and $g_2 = 20$ dB, if we put A₁ first and A₂ second in the signal flow, the noise factor for the cascade will be (notice that we cannot use the noise figure *F* directly in Eqs. (4.15) and (4.16) due to the subtraction):

$$f_{casA1A2} = f_1 + \frac{f_2 - 1}{g_1} = 10^{3/10} + \frac{10^{10/10} - 1}{10^{6/10}} = 4.26$$
(4.17)

Or $F_{casAIA2} = 6.3$ dB. If instead we put A₂ first and then A₁, the noise factor f_{cas} now becomes:

$$f_{casA2A1} = f_1 + \frac{f_2 - 1}{g_1} = 10^{10/10} + \frac{10^{3/10} - 1}{10^{20/10}} = 10.01$$
(4.18)

Or $F_{casA2AI} = 10$ dB. From this simple example we can derive two conclusions:

- 1. For an RF receiver with a gain cascade, the first amplifier is at the carrier frequency with the most important noise figure. The noise factors in the following stage will be reduced by the gain in the preceding state. Now you understand why there is an LNA (low-noise amplifier) which is to give some moderate gain to reduce the noises introduced in the later stage!
- 2. As long as the first few stages provide a reasonable gain say 15 to 20 dB, the noise in the following stages does NOT matter much, and we can use that to trade with power consumption, linearity, tunability, etc.

4.4 Noise factor in other RF components

There are three components that worth special notice in noise treatment: attenuator, mixer and local oscillator (LO). Attenuator is straightforward, mixer needs some treatments for additional bandwidth, and the phase noise in LO needs special discussion later.

4.4.1 Noise factor in attenuators

We will first treat the easy noise case of the attenuator. Why attenuation? As the transmitter-receiver has unknown distance or path in a general wireless network, to make sure the receiver can make out the signal defined by its RF input sensitivity, when the receiver is under direct line-of-sight and in close proximity, the RF input power can be too strong to saturate the behavior. This is similar to a football field. When the broadcast is audible for the people 300 m away, it will cause a major problem for people sitting besides the loudspeaker. The RF system has much more dynamic ranges than the sound broadcast systems!!!

Attenuation is typically achieved by passive elements that introduce loss is signal strength. It does not need to use resistor ladders, as reactance works equally well in attenuating the signal without introducing further noises. If we assume the attenuator does not introduce any further noises (its own noise temperature is zero), according to Eqs. (4.1) and (4.11), the noise factor for attenuators will be:

$$f = \frac{1}{g}$$
 (4.19) (attenuators)

where g < 1 for attenuators. Or when it precedes another module with noise factor f_2 , the total cascade noise will be:

$$f_{cas} = \frac{1}{g_1} + \frac{f_2 - 1}{g_1} = \frac{f_2}{g_1}$$
(4.20) (Attenuators)

That is, the attenuator will magnify the noise factor in the cascade by its attenuation factor! For sure, this makes sense as noise factor is defined on signal-to-noise ratio.

Exercise: When amplifiers and attenuators are cascaded (most often the attenuators are for variable gains), for noise consideration, should we put the amplifier or the attenuator at the beginning of the cascade?

4.4.2 Noise factor in mixers from additional noise bandwidth

In addition to phase noise in LO, the mixer introduces another noise possibility from the function of frequency conversion. Let's use a simple receiver with f_{LO} , f_{RF} and f_{IF} as an example, where f_{LO} is the local oscillator frequency to demodulate the data with channel selection, f_{RF} is where the received signal with spread spectrum that carries the data (by definition, it has to carry a bandwidth, instead of single-tone like LO), and f_{IF} is the intermediate frequency in the superheterodyne scheme. Let's look at the case where f_{RF} spreads only on one side of the f_{LO} . This is called the single sideband (SSB) modulation, which is often preferred due to its spectral efficiency over double sideband (DSB) modulation, as shown in Fig. 4.6. This means: $f_{IF} = f_{LO} - f_{RF}$.



Fig. 4.6. Illustration of noise from image frequency in the mixer module.

We immediately notice that f_{IF} will not only contain the "noise" in the bandwidth f_{RF} , but also contain the noise in the bandwidth of the image frequency f_{IM} , where $f_{IF} = f_{IM} - f_{LO}$. Even if we have some way that the interference signal in f_{IM} is filtered out before entering the mixer, just at the mixer, the noise within f_{IM} will enter f_{IF} , as an enlarged bandwidth! If we have ONLY white noises, then the mixer will give two times larger noise factor by f_{IM} , or 3dB in noise figure!

4.4.3 Noise factor in LO: A preview

As a preview before formal analysis, LO generates monotone sinusoidal functions to modulate and demodulate signals, and the uncertain shift in zero-crossing timing defines a unique phase noise. Phase noise has the same origins as the amplitude noise that modulates the waveforms (superposition in magnitude in any time point for desired signals and unintended noise/interference), but the nature of the phase noise is specifically on timing through a *multiplicative* process (modulation). LO is the main functional module to perform channel selection, as the filter alone is very difficult to provide the needed sharp transition between the passband and the stopband as the channel is often less than 0.1% of the carrier. Also, it is often expensive to make very high-order filters to be switchable to satisfy the system-specified gain in pass and stop bands by filtering alone, although channel-select filters can be made by switch-capacitance circuits to enhance the channel reception. The phase noise and frequency uncertainties in LO are hence fundamental to many modern RF systems. Two distinctive features for phase noise from amplitude noise are:

- 1) Phase noise is multiplicative as in signal modulation, and therefore it is NOT useful to increase the signal power to have a better understanding of the signals through higher SNR. On the other hand, amplitude noise is *additive*, and boosting the signal strength under constant noise is a straightforward way to boost the symbol rate or decrease the bit-error rate. If phase noise is less controllable (or no good LO can be generated due to power constraint such as in RFID tags), the only possible modulation scheme is incoherent amplitude modulation, which has a bit rate to be only half of the symbol rate (for derivation, this will be ECE 4670).
- 2) Phase noise affects timing, so it is more serious in affecting the phase encoding and less on the amplitude encoding. From this simple picture, we can already see the gain factor in mixers by the amplitude of LO is useful to boost the SNR in magnitude, but useless to improve the phase noise.

We will treat more about the influence of LO phase noise to the signal modulation and demodulation in later chapters.

4.5 Constant noise curves in the Smith Chart

Noises can be represented in the Smith Chart for visualization as well. Remember that the Smith chart is the polar coordinates of the complex reflection coefficient $\Gamma = \Gamma_L e^{2jkz}$, and then the lines represent the constant loci of the impedance Z as defined by

$$Z(z) = \frac{V(z)}{I(z)} = Z_o \frac{1 + \Gamma_L e^{2jkz}}{1 - \Gamma_L e^{2jkz}}$$
(4.21)

Assume a module has minimal noise factor of f_0 under an optimal source impedance of Y_0 . When the source impedance is at Y_S , the corresponding noise factor is f_1 . We can treat that the additional noise $f_1 - f_0$ as an uncertainty in the reflection coefficient, i.e., $f_1 = f_0$ when $Y_S = Y_0$, and

$$\left|\Gamma - \Gamma_0(f_1)\right| = f_1 - f_0 \tag{4.22}$$

where Γ_0 is the reflection coefficient with the module has no additional noise (notice that in the general case Γ_0 is a function of f_1). This is represented in Fig. 4.7. We will treat more on the use of the constant circles in the Smith Chart when we treat the stability of the power amplifier. For now, we will leave at the conclusion that during source impedance mismatch, there can be larger noise factors contributed by the module due to the uncertainties in the reflection coefficient.



Fig. 4.7. Constant noise circles for modules with impedance mismatch.