Noise effects and signal-to-noise ratio

In this course we have seen how to deal with radiation in terms of the average power transmitted through free space between a target and a sensor. We have also shown how to deal with the fluctuations of this radiation that occur whether it is incoherent (e.g., thermal) or coherent (sinusoidal). For passive RF systems, the radiation propagation through free space is generally handled by the antenna theorem and effective source brightness function. For active systems, the radiation propagation is generally handled with Friis' transmission formula.

The received power is generally very weak in sensor systems, typically orders-of-magnitude weaker than it is in communications systems. So an important issue with any sensor system is "masking" of the signal by fluctuations in the power (i.e., the "noise") in the receiver. The "noise" is the totality of all the electronic mechanisms, in addition to the radiation fluctuations. Such noise is always present, even in the absence of electronic noise, so it is important to define a metric for the sensor performance in the presence of noise. A useful metric for all types of sensors is the power signal-to-noise ratio (SNR).

$$\frac{S}{N} = \frac{< P >}{\sqrt{< (\Delta P)^2 >}} = \frac{< P >}{S_P \cdot B_{ENB}} \tag{1}$$

where S_P is the power spectral density and B_{ENB} is the equivalent noise bandwidth at that point in the sensor

$$B_{ENB} = (G_{\max})^{-1} \int_0^\infty G(f) df$$

where G(f) is the sensor gain function vs frequency, and G_{max} is the maximum value of this gain B_{ENB} is generally dictated by sensor phenomenology, such as the resolution requirements and measurement time.

<u>Noise from Electronic Components</u>

Within every sensor system, particularly at the front end, are components that contribute significant noise to the detection process and therefore degrade the ultimate detectability of the signal. The majority of this noise usually comes from electronics, particularly the first device, which is often a mixer or direct detector. After this there is generally a low-level amplifier that contributes comparable noise. The majority of noise from such devices falls in two classes: (1) thermal noise, and (2) shot noise. Thermal noise in semiconductors is caused by the inevitable fluctuations in voltage or current associated with the resistance in and around the active region of the device. This causes fluctuations in the voltage or current in the device by the same mechanism that causes resistance – the Joule heating that couples energy to and from electromagnetic fields. The form of the thermal noise is very similar to that for free-space blackbody radiation. And the Rayleigh-Jeans approximation is generally valid for room temperature operation, so that the Johnson-Nyquist theorem applies. However, one must account for the fact that the device is coupled to a transmission line circuit, not to a free space mode, and the device may not be in equilibrium with the radiation as assumed by the blackbody model.

All of these issues are addressed by Nyquist's generalized theorem

$$\Delta V_{rms} = [4k_B T_D \mathrm{Re}\{Z_D\}\Delta f]^{1/2}$$

where T_D, Z_D , and Δf are the temperature, differential impedance, and bandwidth of the device. Even this generalized form has limitations since it is not straightforward to define the temperature of the device if it is well away from thermal equilibrium.

Shot noise is a ramification of the device being well out of equilibrium. It is generally described as fluctuations in the current arriving at the collector (or drain) of a three-terminal device caused by fluctuations in the emission time of these same carriers over or through a barrier at or near the emitter (or source) of the device. The mean-square current fluctuations are given by

$$<(\Delta i)^2> = 2e\Gamma I \cdot \Delta f$$

where Γ is a numerical factor for the degree to which the random Poissonian fluctuations of emission times is modified by the transport between the emitter (or source) and collector (or drain). If $\Gamma = 1$, the transport has no effect and the terminal current has the same rms fluctuations. When $\Gamma < 1$, the transport reduces the fluctuations, usually through some form of degenerative feedback mechanism, and the shot noise is said to be suppressed. When $\Gamma > 1$, the transport increases the fluctuations, usually through some form of regenerative feedback mechanism, and the shot noise is said to be enhanced.

Linear Components and Noise Factor

While at first appearing to add insurmountable complexity to sensor analysis, a great simplification results from the fact that radiation noise and two forms of physical noise discussed above are, in general, statistically Gaussian (the shot noise becomes Gaussian in the limit of large samples, consistent with the central limit theorem). A very important fact is that *any Gaussian noise passing through a linear component or network remains statistically Gaussian*. Hence, the output power spectrum S in terms of electrical variable X (current or voltage) will be white and will satisfy the important identity

$$<(\Delta X)^2>=\int_{f_0}^{f_0+\Delta f} S_X(f)\cdot df \approx S_X(f)\cdot\Delta f$$

where Δf is the equivalent-noise bandwidth. Then one can do circuit and system analysis on noise added by that component at the output port by translating it back to the input port. In the language of linear system theory, the output and input ports are connected by the system transfer function $H_X(f)$, so the power spectrum referenced back to the input (reference) port becomes

$$S_X(in)=\frac{S_X(out)}{|H_X(f)|^2}\ .$$

Because the different Gaussian mechanisms are statistically independent, the total noise at the reference point can be written as the uncorrelated sum

$$<(\Delta X_{TOT})^2>=\sum_{j=1}^{N}<(\Delta X_i)^2> \text{ or } <(\Delta P_{TOT})^2>=\sum_{j=1}^{N}<(\Delta P_i)^2>$$

As in any RF system, it is the signal-to-noise ratio *after detection* that matters most. And there is always several components between the sensor input and the detector that mask the signal by an amount that depends on the gain “ahead” of the component. This leads to a figure of merit that combines the noise contribution and gain together. It is called the noise figure (or factor)

$$F=\left(\frac{(S/N)_{IN}}{(S/N)_{OUT}}\right)\ ;\ 1<F<\infty$$

In other words, the noise figure quantifies the degradation in SNR as a signal passes through a component in a linear chain. It can be combined with the other noise figures in the chain to get

$$F_{TOT} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 \cdot G_2} + \ldots . \frac{F_n - 1}{\prod_{i=1}^{n-1} G_i}$$

where G_i is the power gain of the ith element. Note that this gain accounts for impedance mismatch between elements. Intuitively, this means that components located further down a chain tend to be less important if the earlier components have high gain. Note that this gain accounts for any impedance mismatch between elements, so that when the impedance match is poor the gain will suffer too. Hence, for linear components such as unsaturated amplifiers, we have $G(f) \approx |H_X(f)|^2$

Physically, the noise figure represents the degree to which the linear component degrades the SNR at the input through the introduction of its own noise mechanisms. Since linear components generally maintain the statistics of the input noise (i.e., Gaussian noise stays Gaussian), the SNR at the output can be no better than the SNR at the input. This means that the noise factor can be no less than unity, which is why F generally lies in the range $1 < F < \infty$. Note that noise figure, which is more commonly used in systems engineering than noise factor, is just defined by

$$NF = 10 * \log_{10}(F)$$

System engineers love their decibel units, and don't forget that they almost always refer to relative power levels, not signal strength.

A convenient way to represent the noise figure more intuitively is to assume that the noise at input and output are both additive Gaussian. The noise contributed by linear component can then be represented as an equivalent fictitous noise temperature T_N corresponding to $k_B T_N B_N G$ noise power at the output, or $k_B T_N B_N$ at the input. The noise factor is then given by

$$\boldsymbol{F = \left[\frac{(S/N)_{IN}}{(S/N)_{OUT}} \right] = \frac{S/(k_B T_0 B_N)}{GS/[G(k_B T_0 + k_B T_N)B_N]}}$$

After several cancellations we get the simple result

$$F = 1 + T_N/T_0$$

where T_0 is the ambient temperature. Decades ago a convention was established of $T_0 = 290$ K – an average room temperature around the world.

As an example, let's take a modern low-noise amplifier (LNA) which at low enough frequency (up to ~6 GHz) can have a noise *figure* as low as 1.0 dB. And it can provide this noise figure under room temperature operation. The corresponding noise factor and temperature are given by

$$F = 10^{NF/10} = 1.26 \text{ and } T_N = 290(F-1) = 290\,(1.26-1) = 75.4 \text{ K}$$

On first glance, this might appear to violate the laws of thermodynamics. How can a device produce a noise temperature lower than its operating temperature ? Fortunately, there is no violation of thermodynamics because the amplifier is categorically not in thermodynamic equilibrium. This is why it can have gain ! And so consistent with definition, the noise temperature is a fictitious measure of just how "clean" this gain is. It is also the reason so many RF and THz engineers work on amplifiers and worry so much about noise figure. Quoting an old systems engineer, "gain is a wonderful thing."

More General Sensitivity Metrics

Unfortunately, the noise factor concept is not applicable or particularly useful to all components or THz systems since it presupposes linearity. As we shall see shortly, there are very good reasons to use nonlinear elements in RF and THz systems, not the least of which is the fact that the extraction of information from or making a decision based on the incoming radiation ultimately requires power or energy detection. And we know from simple circuit theory that measurement of power or energy is inherently a nonlinear process, usually quadratic in the signal levels. Hence, it is the SNR *at the point of detection or decision* that is generally the most important quantity in system performance. And because of reasons discussed later in statistical detection theory, generally this SNR must generally be greater than or equal to unity to have a reliable detection or decision. Therefore, a very useful metric for sensor performance is to fix the "after detection" SNR , SNR_{AD} , at unity and then solve for the signal power a*t the input to the system* sensor that achieves this. The resulting metric is the noise-equivalent power spectral density, NEP, which most simply put, is the *input* signal power to the sensor required to achieve a SNR of unity at the *output*.

A simple example may be helpful at this point. Suppose we have an ideal "square-law" detector, which we will address in more detail later because of its great utility in RF and THz systems both. A good and very old example is the Schottky-diode rectifier. By definition, the square- law detector has a circuit transfer function of

$$X_{out} = \Re_X P_{in}$$

where X_{out} is the output signal (usually current or voltage), P_{in} is the input power, and $\Re_X$ is the "responsivity". The noise at the output of the detector is minimally given by the Nyquist generalized theorem, which for a Schottky rectifier is

$$P_N \propto \langle(\Delta I)^2\rangle = 4k_B T_0 \mathrm{Re}\{Y_D\}\Delta f \approx 4k_B T_0 \Delta f / R_D$$

where Δf is the post-detection bandwidth. The corresponding signal power at the output is just $(X_{out})^2$, so that the output SNR (with X chosen as I) is

$$SNR_{AD} = (\Re_I P_{in})^2 / 4k_B T_0 \Delta f / R_D$$

Setting this SNR to unity and solving for P_{in}, we get $NEP_{AD} = (4k_B T_0 \Delta f / R_D)^{1/2} / \Re_I$

For the purpose of comparing different sensor technologies, it is conventional to divide out the post-detection bandwidth effect (or equivalently, setting it equal to 1 Hz). This yields the normalized NEP_{AD} [in units of $W/(Hz)^{1/2}$], given by

$$NEP'_{AD} = NEP_{AD}\sqrt{1/\Delta f} \quad NEP'_{AD} = (4k_B T_0 / R_D)^{1/2} / \Re_I$$

As a useful example, we consider a garden variety Schottky diode used in RF and THz rectification and detection. For reasons having to do with the solid-state physics of Schottky barriers, we find $\Re_I \approx 25$ A/W or less. And at room temperature, $R_D \sim 1$ KΩ. Choosing these values, we get $NEP'_{AD} \approx 1.6\text{x}10^{-13}$ $W/(Hz)^{1/2}$ – an excellent sensitivity not ever achieved in practice because of power coupling. With a differential resistance of 1 KW, the power delivery to the device is not very efficient because of impedance mismatch. And this problem is pervasive in the THz region where getting adequate device speed and good impedance matching are very difficult to achieve together. So we introduce an external power coupling factor, η, analogous to the external coupling efficiency in photonic devices and define it by

$$P_{abs} = \eta\, P_{inc}$$

where P_{inc} is the power incident on the device (or equivalently, the "available" power). Then the NEP is re-calculated for the incident power that achieves a SNR of unity, or

$$NEP'_{AD} = (4k_B T_0 / R_D)^{1/2} / (\eta \Re_I)$$

For the 1 KΩ Schottky diode, we expect $\eta \approx (1 - R)$ where R is the power reflection coefficient R $= [(1000 - R_A)/(1000+R_A)]^2$, where R_A is the a antenna resistance. Choosing a typical value of R_A = 100 Ω, we get R = 0.67, η = 0.33 and $NEP'_{AD} \approx 5\text{x}10^{-13}$ $W/Hz^{1/2}$ – a value that has been achieved at microwaves, but not yet at THz frequencies because the η is even lower than 0.33 largely because of reactive impedance mismatch that we have ignored. Nevertheless, this example should be illuminating to students trying to understand NEP - one of the most bewildering of all system metrics.

We deduce the generic expression for a square- law detector including the external power coupling factor:

$$NEP'_{AD} = (<(\Delta X)^2>/\Delta f)^{1/2}/(\eta \Re X)$$

Although simple looking, there is a lot of interesting physics buried in this expression !
A useful feature of the NEP is its additivity. If there are N mechanisms contributing to the noise at a given node in the receiver, if the mechanisms are uncorrelated to the signal and to each other, if they obey Gaussian statistics, then the total NEP is the uncorrelated sum

$$NEP_{TOT}^2 = NEP_1^2 + NEP_2^2 + ...NEP_N^2$$

This property applies to any node, pre- or post-detection, and will be used explicitly when we discuss the contribution from electronic noise. In reality there are cases where a noise mechanism is correlated to the signal (e.g., radiation noise) or to another noise mechanism (current and voltage noise in transistors), so one must be careful in applying this addition formula. In such cases, one can always fall back on the SNR as a useful measure of overall system performance.

Noise Equivalent Delta Temperature

For radiometric and thermal imaging systems, it is sometimes convenient to express the sensitivity in terms of the change of temperature of a thermal source at the input that produces a post-detection SNR of unity. The resulting metric is the noise equivalent delta temperature, or NEΔT, which is given by

$$NE\Delta T = \frac{NEP_{AD}}{dP_{inc}/dT\,|_{P_B}} \quad \text{or} \quad NE\Delta T' = \frac{NEP'_{AD}}{dP_{inc}/dT\,|_{P_B}}$$

where P_{inc} is the incident power. We assume that the receiver accepts just one spatial mode with power coupling η, and that the thermal source satisfies the Rayleigh-Jeans limit and fills the field-of-view of the sensor. In this case, $P_{inc} = k_B T_B \Delta\nu$ and $P_{abs} = \eta k_B T_B \Delta\nu$ so the NEΔT becomes

$$NE\Delta T' = \frac{NEP'_{AD}}{k_B \Delta\nu}$$

where Δν is the spectral bandwidth, not to be confused with the electrical bandwidth Δf.