2. **Noise**

By the end of this section you will be able to:

- Describe some of the sources of electrical noise and compute their magnitude.
- Compute noise in multistage amplifiers and know how to minimise it.
- Describe several methods for eliminating interference, drift and other artefacts in measurements.

2.1. Interference

"Any unwanted sound is **noise**." This is a common but unsatisfactory definition, since it does not distinguish between the avoidable and the unavoidable, the fundamental and the artefact. We will not define noise, but we will describe it.

Noise in electronic systems includes contributions from:

- Mechanical interference
- Electrical interference
- Fluctuations in the electromagnetic field
- Fluctuations in electron movement.

Mechanical Interference refers to vibrations. Sensitive measurements require sound isolation, e.g. solid buildings, –higher floors in building oscillate noticeably! - anechoic chambers, shock absorbent mountings for instruments and sensors, temperature and humidity conditioning of the laboratory, among others.

Electrical Interference refers to the interaction of an electronic system with ambient electromagnetic fields. The latter may be due to radio broadcasts, power lines, motors, electromagnets, and in general the operation of other electrical and electronic devices. It is straightforward –if expensive- to eliminate electrical interference from measurements, by grounding and shielding (screen rooms) and meticulous elimination of sources of electromagnetic radiation (e.g. fluorescent lights, switching power supplies, AC power lines, electric motors, etc). Most of the effort in grounding and shielding is devoted in the elimination of **ground loops.** Ground loops are networks assumed to be at ground potential, but:

- Current flows through some or all branches of the network , so that ohmic voltages are developed
- Comprise loops threaded by time-changing magnetic flux, so that voltages are induced, and the ground terminal instead of being a reliable reference potential becomes a noisy antenna.

Grounding and shielding is an important subject in measurement science and technique. Even though it has a solid foundation on Maxwell's equations, it is considered by many to be a black art, which takes years of laboratory experience to master. An excellent book on the subject is "Grounding and Shielding Techniques" by Ralph Morrison (4^{th} ed., J. Wiley).

Interference, once characterised, is modelled as an additional signal source, or even as an extra noise contribution.

2.2. Thermal noise

2.2.1. Fluctuations and Information

Fluctuations are a consequence of a fundamental theorem of thermodynamics, the **equipartition theorem.** This theorem states that:

Each degree of freedom of any system fluctuates with an energy equal to $1/2k_BT = 2{\bullet}10^{-21}\,J$ at the standard temperature of 290 K**. (k_B =1.38 x 10^{-23} J/K , is the Boltzmann constant and T the absolute temperature).** This means that each degree of freedom contains $1/2k_BT$ free energy. A degree of freedom is any quantity free to change. An electron will wander around its average trajectory by 3/2kT in 3 dimensions, for example, and a lattice ion will vibrate with the same vibration energy.

The fluctuations described by the equipartition theorem cause electrical noise as we know it. Fluctuations are also at the heart of the definition of **information**. "Information" is intimately related to disorder, which a physicist calls the **Entropy**. Entropy is defined as:

$$S = k_B \ln\left(\Omega_i\right)$$

Ω_i is the number of different ways of putting together a many degree-of-freedom system so that the same measurements emerge. This difficult concept can be illustrated by a simple example: in a system we have N identical particles, and we know N-1 of them are not moving while 1 is moving. Clearly the single moving particle can be any one of the N, so that the entropy of this system is $k\ln\left(N\right)$. If Ω is the total number of ways the system can be put together, $P_i = \frac{\Omega_i}{\Omega}$ is the probability the system is in a particular state. Ordered systems have lower entropy than disordered systems (there are fewer ways to rearrange their constituents), and for this reason Entropy is often loosely defined as "the measure of disorder in a system".

Information, on the other hand, is defined as:

$$I = -\ln\left(P_i\right) = -\ln\left(\Omega_i/\Omega\right) = \ln\left(\Omega\right) - \ln\left(\Omega_i\right) = \text{constant} - S/k_B$$

When we pass information from one place to another we reduce the entropy of the receiver. This definition reveals that information is, up to an additive constant, the negative of entropy.

2.2.2. The power-SNR tradeoff

We know from thermodynamics that to change the entropy of something, its energy must change. A fundamental relation in thermodynamics connects the change of Entropy to the change in energy of a system:

$$\delta E = T\delta S$$

This is true in equilibrium, but for small steady state changes we can attempt to write:

$$\frac{\partial E}{\partial t} = P = T\left(\frac{\partial S}{\partial t}\right) = -k_B T\frac{\partial I}{\partial t}$$

Where P is power flow. Information flow is clearly related to power flow, assuming the system remains the same. The transmitter of information acts like a refrigerator, and causes the receiver to lose energy:

$$\frac{\partial I}{\partial t} \propto -\frac{P}{T}$$

Of course, refrigeration cannot be accomplished without the transmitter dissipating power, and indeed, it can never be accomplished at a 100% efficiency. We then know that the power dissipated at the transmitter must exceed:

$$P_{diss} > kT\frac{\partial I}{\partial T} \Rightarrow P_{diss} = c\frac{\partial I}{\partial T}$$

We can now introduce the Shannon formula which connects the Signal-to-Noise Ratio (SNR) to analogue bandwidth and information flow:

$$\partial I/\partial T\,(bits/\sec) = B\ln_2\left(S/N+1\right)$$

Tthe SNR is understood as the ratio of signal **power** to noise **power**. We conclude that:

$$S/N \propto \left(2^{P/BkT}\right)^{1/c}$$

The constant c is system dependant and not known. This equation quantifies the "Power-S/N-Bandwidth conflict". This conflict is fundamental. Even though we can approach the limit for c=1 we will never be able to exceed it. Today's state of the art digital circuits operate at $c = O\left(10^2 - 10^3\right)$. The lowest power analogue circuits published operate with $c = O\left(10^3 - 10^4\right)$. Digital CMOS is still a significantly lower power technology than analogue micropower.

Note: As we already mentioned, strictly speaking, an electronic circuit in which current flows is not in equilibrium, and the argument we just presented does not apply. What we just computed is the lowest amount of energy necessary to perform a signal processing task, i.e. the dissipation of a quantum computer. There is some disagreement in the literature as to what is the lowest energy per bit allowed by the laws of nature. In the electronics literature there appears to be an agreement with Vittoz who considering RC filters showed that $c \geq 8$ for each real pole in the filter. The physics and mathematics literature converges on that $c \geq \ln 2 = 0.694$, i.e. over an order of magnitude lower, and indeed lower than the intuitive c=1 value!

2.2.3. Johnson noise

Every resistor generates noise with an independent of frequency power spectral density. The RMS open circuit noise voltage noise voltage of the resistor is:

$$V_N = \sqrt{4kTRB} \tag{2.1}$$

The RMS short circuit current noise is:

$$I_N = \sqrt{4kTB/R} \tag{2.2}$$

These statements together suggest that a noisy resistor can be modelled as a Thevenin or Norton circuit consisting of a noiseless resistance R together with a Thevenin source (2.1) or a Norton source (2.2).

This is a very powerful theorem which makes no reference to the internal workings of the components. This means that the formulas are valid even if the resistance is not a physical resistance, for example for a Thevenin resistance! The proof of the Johnson noise formula is rather simple, and is reproduced below to disperse any residual doubts on its validity.

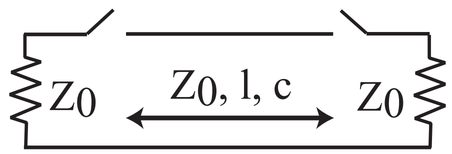

Figure 1: Circuit to calculate the Johnson noise. A length of transmission line l is terminated, through ideal switches to is characteristic Z_0.

Proof: In Figure 1, we show a transmission line of length L, which supports travelling waves in both directions. The phase velocity on the line is c and the characteristic impedance is Z_0. The line is kept at temperature T. According to the equipartition theorem, all modes of this line will be excited, each with energy $1/2k_BT$. Assume now that the line is so long that the wave does not have time to reach the end of the line within the observation time. We will only consider the travelling waves which are transverse electromagnetic modes (TEM). (It is difficult to show, but these are ALL the propagation modes in certain types of waveguides such as lossless coaxial cables, as well as being all the modes that can couple to the current flow inside the resistor). The transmission line is like a flute. Its modes have nodes at the line endpoints, so that the nth mode is at a frequency:

$$f_n = \frac{nc}{2L} \qquad \textbf{(2.3)}$$

A TEM line supports two modes at each frequency, one for each polarisation (for example horizontal or vertical) of the electric field. According to the equipartition theorem each mode is excited with energy:

$$W_n = k_BT \qquad (2.4)$$

The energy stored in the travelling waves in a range of frequencies **B** between modes **n** and **m** is:

$$W_B = (n-m)k_BT = \frac{2L}{c}(f_n - f_m)k_BT = \frac{2L}{c}Bk_BT \qquad (2.5)$$

This is the thermal energy stored in the line when the two end switches are open.

We can then close the two switches. The line is now terminated at both ends with resistors equal to its characteristic impedance Z_0 held at the temperature T of the waveguide. From transmission line theory we know there will be no reflections at the end points; all the energy contained in a travelling wave incident to one of the terminations will be absorbed by that termination. The terminated line, like the open one, is in equilibrium with its surroundings, and the power absorbed by one of the terminations **must** be generated by the termination at the opposite end.

This means that the energy stored in the line modes is the sum of the power generated by both resistors over the time it take for radiation to travel the length of the line. Symbolically:

$$W_B = 2P(Z_0)\frac{L}{c} \Rightarrow P(Z_0) = Bk_BT \qquad (2.6)$$

This is the most general statement of the **Johnson noise** formula, stating that the available noise power at the terminals of any resistor is simply Bk_BT **.**

Between a termination resistor and the transmission line we have set up a voltage divider of magnitude ½ . The RMS noise voltage between the terminals of the resistor is therefore given by the familiar Johnson noise formula $V_N = \sqrt{4kTRB}$.

The spectrum of Johnson noise is **white,** that is, the power spectral density is independent of frequency. Of course, the spectral density must decrease beyond some frequency. Because of the quantised nature of light, modes of frequency such that the mode photon energy hf is comparable to kT are less likely to be excited, and there should be an exponential reduction of the Johnson noise power spectral density at high enough frequencies. The roll off starts roughly at $f = k_BT / h \simeq 6THz$ (h is Plank's constant, $h = 6.636 \bullet 10^{-34} Js$).

If the quantised energy of a photon is taken into account the Johnson noise formula reads:

$$V_n^2 = 4RB\frac{hf}{e^{hf/k_BT}-1}$$

If the frequency is very small, $hf \ll k_BT$, then

$$V_n^2 \simeq 4k_BTRB\left(1-\frac{1}{2}\frac{hf}{k_BT}-\frac{1}{6}\left(\frac{hf}{k_BT}\right)^2-\cdots\right)$$

In practice the Johnson noise formula breaks down at slightly lower frequencies: when the assumptions we made to derive it are no longer valid. We assumed that the resistor is dimensionless. So we expect some deviation from the Johnson formulae when the wavelength is small enough that the resistor can support internal electromagnetic modes, i.e. when the resistor maximum dimension is similar to half the wavelength λ. This can happen at frequencies of 10s to 100s of GHz.

2.2.4. Shot Noise

Current is due to the transport of discrete electrons. Fluctuation in their kinetic energy will modulate the arrival rate of electrons at a node where current is measured. A counting argument leads to a formula for the shot noise, which is also spectrally white:

Assume a current I flows into a node. This means that over a time τ $n = \frac{I}{e}\tau$ electrons must arrive at the node. The current is then $I = ne/\tau$

This number n is subject to fluctuations governed by Poisson statistics, so it is uncertain by:

$$\Delta n \propto \sqrt{n} \tag{2.7}$$

Electrons have spin and the arrivals are correlated in pairs, do that the square variance is double:

$$\Delta n = \sqrt{2n} \tag{2.8}$$

It follows that the RMS fluctuation of current over a bandwidth B is:

$$\Delta I^2 = 2eIB \tag{2.9}$$

Shot noise has a constant power spectral density. Like Johnson Noise, shot noise is spectrally white, up to frequencies of the order of the electron arrival rate, when the averaging argument breaks down.

$$f_0 \simeq \frac{I}{e} \tag{2.10}$$

For a current of 1nA this frequency is a few GHz.

Although we assume shot noise is modelled by a pure current source, we have to observe conservation of energy, which implies that the internal impedance of this source needs to be finite, even if it turns out to be very big. To calculate the correct Norton impedance of a shot noise source, we can use the Eistein relation. This states that the ratio of power to current carried by an electron gas needs to be (the numerical value for room temperature):

$$\frac{P}{I} = \frac{kT}{e} = 25mV$$

This would be half the open loop voltage of the noise source, so, the Norton resistance of a shot noise source I_N is simply:

$$R_N = \frac{2k_B T}{eI_N}$$

A numerical estimate for the Norton impedance of the shot noise source associated with a 1A current source over a 1GHz bandwidth is a few $k\Omega$. At current levels encountered in microelectonics this impedance is several $M\Omega$.

2.2.5. Flicker Noise

A fundamental source of noise in any physical system containing a large number of degrees of freedom has a power spectral density :

$$P(f) \propto f^{-k} \tag{2.11}$$

The constant k is ideally k=1, and in practice it is close to unity. This type of noise is called **1/f Noise, flicker noise** or **pink noise. 1/f** noise is found in the frequency spectra of waves, winds, movements of heavenly objects, the flicker of the light intensity of a candle flame, and indeed in voltage and current noise sources in electronics. There are numerous arguments in the literature deriving pink noise in all kinds of physical systems, especially transistors. Most of these arguments are plausible. However, we need to remember that this is a universal phenomenon whose explanation is well beyond derivations based on charge transport in MOSFETs.

The general explanation of the origins of flicker noise remains one of the great unsolved problems in theoretical physics, and many believe that such an explanation is worth a Nobel Prize. The power spectral density of pink noise increases as frequency is reduced, and at low enough frequencies it dominates Johnson and shot noise. This is the reason that sensitive DC measurements at the level of nV and pA require enormous skill and very specialised equipment to perform.

The absolute magnitude of pink noise is an empirical constant. It is usually presented as the **corner frequency,** the frequency at which the PSD of pink noise equals that of white noise:

$$P_{pink}(f_C) = \frac{\alpha}{f_C} = kT \Rightarrow \alpha = k_B T f_C \tag{2.12}$$

The usual model for pink noise is then:

$$P_{pink}(f) = \frac{k_B T f_C}{f} \tag{2.13}$$

Pink noise is largely responsible for the "**phase noise"** in oscillators. As Rohde has shown in his book on microwave synthesisers the output of an oscillator is effectively DC power mixed (upconverted) into the output frequency. To model pink noise, we add it as an additive, uncorrelated correction to Johnson or shot noise. Since the power spectral density this mechanism can deliver to a load is P_{pink} , the voltage amplitude we must employ to model its contribution through an impedance R is:

$$V_{pink} = \sqrt{2P_{pink}R}$$

2.3. Modelling noise

Most electronic systems include noise contributions from several devices. To correctly calculate the noise we need to account for both the system's deterministic electrical behaviour and its noise contribution.

Any noisy device can be represented as a noiseless version of itself properly connected to suitable noise voltage and current sources. A resistor, for example, is represented as a Thevenin circuit of a Johnson voltage source in series with its resistance, or a Johnson current source in parallel with its resistance. Current sources (whether they are independent or dependent sources) may be modelled as shunt connections of their nominal sources with the shot noise sources. Pink noise is modelled by a voltage or current source connected to the circuit through a low pass filter.

2.3.1. Addition of noise sources.

Two noise sources in a network may depend on each other, but most of the time they are statistically independent. We say that two sources f and g are independent if their cross-correlation is zero for all times:

$$c_{f,g}(\tau) = \int_{-\infty}^{\infty} f^*(t)\, g(t+\tau)\, dt \qquad (2.14)$$

We use the correlation theorem which relates the Fourier transforms of c, f and g:

$$C(\omega) = F^*(\omega) G(\omega)$$

If the cross-correlator does not vanish we say the sources are correlated, which in practice means that one is proportional to the other. This now looks like the magnitude of the projection of f on g. We can normalise the correlator to get the "coherence function", essentially a projection operator P so that P=1 for perfectly correlated sources, and P=0 for uncorrelated sources:

$$P(F,G,\omega) = \frac{F^*(\omega)G(\omega)}{|F(\omega)||G(\omega)|}$$

The amplitudes of statistically independent noise sources add like orthogonal vectors. If V_{1n}, V_{2n} are the voltage amplitudes of two voltage noise sources connected in series, the amplitude of the total noise voltage is:

$$V_n^2 = V_{1n}^2 + V_{2n}^2$$

Similarly, if I_{1n}, I_{2n} the amplitudes of two current noise sources connected in parallel, the amplitude of the total noise current is:

$$I_n^2 = I_{1n}^2 + I_{2n}^2$$

When the sources are correlated, then Kirchhoff's laws apply as usual and the amplitudes of series voltage sources (or the amplitudes of parallel current sources) add, eg.

$$V_n = V_{1n} + V_{2n} \qquad (2.15)$$

For two partially correlated sources we can write an expression similar to the cosine addition rule for vectors in a plane:

$$|V_n|^2(\omega) = |V_{1n}|^2(\omega) + |V_{2n}|^2(\omega) + 2\,\mathrm{Re}\left(V_{1n}(\omega)V_{2n}^*(\omega)\right)$$

It is easy to verify that this expression takes the correct values if the two voltages are equal or orthogonal to each other.

2.3.2. Excess noise

A physical device always contributes more noise power than predicted by fundamental considerations. To give an example, material defects inside a real resistor act as additional noise generators and contributing noise power beyond that predicted by the Johnson formula. It is impossible to predict theoretically the actual noise of a real device, so it is common to lump all noise contributions beyond Johnson, shot and flicker noise into a single empirical constant, the **Excess Noise Ratio (ENR)**. ENR is defined as the ratio of actual to theoretical noise power:

$$ENR = \frac{N_{actual}}{N_{ideal}} \quad (2.16)$$

ENR is always greater than unity. In very low noise instrumentation engineering it is a good idea to check on the ENR values of components used. These are rarely stated by component manufacturers, but estimates have been published for different types of components. It is, for example known that carbon resistors are noisier than metal film resistors of the same value.

2.3.3. Added noise, noise factor, noise figure and noise temperature

Amplifiers always generate noise internally. The excess noise contributed by an amplifier, gives rise to the concept of the **Noise Factor.** The nose factor is defined as the ratio of the total noise power at the output of the amplifier to the output noise power due to the noise injected by the signal source.

$$F = \frac{\text{Total output noise power}}{\text{Output noise power due to source}} \quad (2.17)$$

By dividing by the output signal power we get:

$$F = \frac{P_{N,out}}{GP_{N,in}} = \frac{GS_{IN}}{S_{out}}\frac{P_{N,out}}{GP_{N,in}} = \frac{SNR_{in}}{SNR_{out}} \quad (2.18)$$

So the noise factor is the **Signal-to-Noise Ratio (SNR)** of the amplifier at its input divided by the SNR at its output. The amplifier is assumed to be driven by a signal source which contains a Thevenin impedance R. R is assumed to be held at the "standard room temperature", namely 290K (17^{o}C), and contribute Johnson noise. It appears the strange temperature of 290K has been chosen as the standard because at 290K, $kT = 4 \cdot 10^{-21}\,J$ i.e. a nice round number.

The noise factor is symbolised by **F**, the output noise power added by the amplifier by $\mathbf{N_a}$ and $\mathbf{S_i}$, $\mathbf{S_o}$, $\mathbf{N_i}$, $\mathbf{N_o}$, **G** , are the input and output signal power, and input and output noise power and power gain of the amplifier respectively. The noise factor is defined as:

By definition, the noise factor F is given by:

$$F = \frac{SNR_i}{SNR_o} \quad (2.19)$$

But we can also take into account the function of the amplifier:

$$SNR_o = \frac{S_o}{N_o} = \frac{GS_i}{GN_i + N_a} \quad (2.20)$$

Where S_i and N_i are the input signal and noise powers, S_o and N_o the output signal and noise powers, and N_a the output referred noise power added by the amplifier. Substituting (2.20) into (2.19) we get:

$$F = 1 + \frac{N_a}{GN_i} \quad (2.21)$$

The noise factor is clearly always greater than one. Usually the term **Noise Figure** is used to describe the deterioration of the SNR of a signal as it goes through an amplifier:

$$N = 10\log F \quad (2.22)$$

The expression 2.22 for the noise factor gives rise to another way of modelling a noisy amplifier. The **noise temperature** of an amplifier is the temperature of a hot source connected at the input of a noiseless copy of the amplifier which would result into the same output noise as the noisy amplifier. We define then the added noise as $N_a = Gk_BT_N$ and the input reference noise as $N_i = k_BT_0$. The noise factor F and noise temperature T_N are related by:

$$F = 1 + \frac{T_N}{T_0} \Rightarrow T_N = T_0\left(F - 1\right)$$

The noise temperature is used to describe extremely low noise amplifiers. Noise temperatures of 50K are quite common in low noise applications. Note that the noise temperature is not the physical temperature of the amplifier.

These equations give a first hint of why a low noise figure is considered important, and why a high gain is often associated with a low system noise figure. Noise figure is, however, overstated in its importance in a low noise system design. It is the correct balance between noise figure and gain that is important rather than a low noise figure itself. It may not yet be obvious, but an amplifier with a noise figure (in dB) numerically greater than its gain (in dB) is not useful at all. Indeed, it can be proven that there exists a combination of passive components which would perform better than an amplifier with lower gain than noise factor. The most obvious, of many passive devices which would outperform, at the system level, such an amplifier, is an impedance matching transformer.

There is another weakness in the concept of noise factor. There is no mention of how well is the source impedance matched to the input impedance of the amplifier in question. Usually F is specified at some source impedance level, for RF circuits 50 Ohms (75 ohms for VHF), and depends strongly on the actual source impedance. We will discuss this point in some detail shortly.

2.3.4. The Friis cascade formula

Assume now we have a collection of amplifiers, with power gains $G_1, G_2, \ldots, G_n$ and noise factors $F_1, F_2, \ldots, F_n$. The overall noise figure of a **cascade** of these amplifiers, with G_1 connected to the input and driving G_2, etc, is given by the Friis formula:

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1G_2} + \cdots + \frac{F_n - 1}{G_1G_2\cdots G_{n-1}} \tag{2.23}$$

Or, the cascade of N amplifiers has a noise factor related to a cascade of N-1 amplifiers and the Nth as follows:

$$F_{(N)} = F_{(N-1)} + \frac{F_N - 1}{G_{(N-1)}} \tag{2.24}$$

The parentheses in the subscript signify a collection of amplifiers up to, and including the one indicated. For example, $G_{(N-1)}$ is the gain of the chain of the first N-1 amplifiers, but G_N is the gain of the Nth amplifier.

To prove the Friis formula we need to use induction:

- If there is only one amplifier, the statement is obviously true, $F_{(1)} = F_1$.
- Assume the formula is true for N-1 amplifiers.
- Adding the nth amplifier to the cascade means that:

  The output signal power of the chain is: $S_{o(N)} = S_iG_{(N)} = S_iG_{(N-1)}G_N$

  The output noise power of the chain is: $N_{o(N)} = N_iG_{(N)} + G_NN_{a(N-1)} + N_{aN}$

We can relate the noise added by the first N-1 amplifiers to their noise factor and the input noise power:

$$N_{a(N-1)} = \left(F_{(N-1)} - 1\right) G_{(N-1)} N_i \tag{2.25}$$

The output SNR of the N amplifier chain is:

$$SNR_{out(N)} = \frac{S_i G_{(N-1)} G_N}{N_i G_{(N)} + G_N \left(F_{(N-1)} - 1\right) G_{(N-1)} N_i + N_{aN}} = \frac{S_i G_{(N)}}{G_{(N)} F_{(N-1)} N_i + N_{aN}} \Rightarrow$$

$$F_{(N)} = \frac{SNR_{in}}{SNR_{out(N)}} = \frac{G_{(N)} N_i F_{(N-1)} + N_{aN}}{S_i G_{(N)}} \frac{S_i}{N_i} = F_{(N-1)} + \frac{N_{aN}}{G_{(N-1)} N_i G_N} = F_{(N-1)} + \frac{\left(F_N - 1\right) G_N N_i}{G_{(N-1)} N_i G_N} \Rightarrow$$

$$F_{(N)} = F_{(N-1)} + \frac{F_N - 1}{G_{(N-1)}} \text{ since } N_{aN} = \left(F_N - 1\right) G_N N_i$$

This proves the Friis formula.

We did not need to make any assumptions about terminal impedances and impedance matching. If the amplifiers are mismatched the same mismatch factor applies to both signal and noise, so that in the SNR calculation this mismatch factor cancels. The noise factor is, therefore, the same as it would be if each amplifier in the entire chain was impedance matched, and the gain of each individual stage was its maximum available gain for unconditionally stable stages, and the maximum stable gain for conditionally stable stages. This is called the “associated gain” in RF circuit design. As we shall see, the noise factor used for each stage in the calculation depends on the source impedance each stage sees, and has a minimum value for a particular optimum source impedance. The minimum noise factor of each stage can be used to obtain the minimum noise factor for the chain, and then the source impedance correction can be used over the entire amplifier chain treated as one.

2.3.5. The Noise Measure

The noise factor of the first amplifier in the chain appears to dominate the sum. This is not exactly so. In practical radio applications we may need a cascade of several amplifiers to realise typical gains of 60-100dB. In long cascades of identical amplifiers the noise factor rapidly converges to the **Noise Measure**, namely the noise factor of an infinitely long cascade of identical amplifiers:

$$M = F + \frac{F-1}{G} + \frac{F-1}{G^2} + \cdots = 1 + \left(F - 1\right) \sum_{n=0}^{\infty} \frac{1}{G^n} = \cdots = \frac{FG-1}{G-1} \tag{2.26}$$

it follows that

$$FG - 1 = M\left(G - 1\right) \tag{2.27}$$

We can now compare the noise factors of the two permutations of two amplifiers, $F_{12} = F_1 + \frac{F_2 - 1}{G_1}$ and $F_{21} = F_2 + \frac{F_1 - 1}{G_2}$:

$$\begin{aligned} F_{12} - F_{21} &= \frac{F_1 G_1 G_2 - F_2 G_1 G_2 + F_2 G_2 - G_2 - F_1 G_1 + G_1}{G_1 G_2} = \\ &= \frac{G_2\left(F_1 G_1 - 1\right) - G_1\left(F_2 G_2 - 1\right) + \left(F_2 G_2 - 1\right) - \left(F_1 G_1 - 1\right)}{G_1 G_2} = \\ &\frac{\left(G_2 - 1\right)\left(F_1 G_1 - 1\right) - \left(G_1 - 1\right)\left(F_2 G_2 - 1\right)}{G_1 G_2} = \frac{\left(G_1 - 1\right)\left(G_2 - 1\right)}{G_1 G_2}\left(M_1 - M_2\right) \end{aligned} \tag{2.28}$$

We have just proven that the noise factor of a cascade of **two** amplifiers is minimised if we put the one with the lowest noise measure first. Clearly, the amplifier with the lowest noise measure is not necessarily the one with the lowest noise figure, nor the one with the largest gain! As we shall see later, the noise factor depends on the source impedance. The minimum noise factor is almost never obtained for an impedance matched source. Source impedance matching only results in maximum gain. A different source impedance leads to minimum noise factor, and yet a different one minimises the noise measure!

Finally, a very useful lemma for high frequency measurements is that the noise figure of an attenuator (or a lossy connection, for that matter) is numerically equal (in dB) to the attenuation.

The proof is simple and depends on the assumption that the noise factor is defined for a device driven by a thermal resistive source, and that impedance matching is maintained both at the input and at the output of the Device Under Test (DUT: acronym introduced by Agilent, now in universal use in electrical engineering).

Assume that the source is of resistance R, and the attenuator is perfectly matched so its output impedance is R again. Then,

$$F = \frac{SNR_i}{SNR_o} = \frac{S_i / k_B TB}{AS_i / k_B TB} = \frac{1}{A} \tag{2.29}$$

A thought experiment confirms that this result is correct. If we enclose the source and the attenuator in a black box, the thermal noise power of the combination will only depend on the temperature, and will have to be equal to the noise power of the source. The signal power, on the other hand, is attenuated.

2.4. Measurement of Noise

Measurement of noise is not simple. The signal levels that need to be measured are of the order of the Johnson noise of a resistor into a matched load is -174 dBm/Hz (dBm is dB referred to a power level of 1mW). For example, the noise power that needs to be measured in order to characterise a device with noise factor F=10 and power gain of 1000 (30dB) over a 1MHz bandwidth is only 40pW!

2.4.1. Spectrum Analyser

One would be tempted to think that the noise factor can easily be measured with a spectrum analyser. This is not the case, except when the noise factor is extremely high. The noise floor of most spectrum analysers is so high (the best instruments may have an input referred added noise of -140dBm/Hz, i.e. a noise factor of 10^3) that it usually masks the noise we are trying to measure, leading to big errors.

We can model the spectrum analyser as a noisy amplifier followed by an ideal spectrum analyser (i.e. an ideal narrow band power meter).

The total noise factor of the amplifier we try to measure and the spectrum analyser will be:

$$F = F_{DUT} + \frac{F_{SPECAN} - 1}{G_{DUT}}$$

Spectrum analysers have typical noise factors of $10^3 - 10^4$ (30-40dB). To make a 1% accurate measurement the DUT needs to have a gain greater than 50-60dB. Unless we are measuring an entire radio front end such large gains are not common. The measurement gets even more difficult if, for example we try to measure the noise factor of a low noise GaAs HEMT which can easily be as low as F<1.1 (N<0.35dB), with an associated gain of, perhaps, 20dB. In such a case the spectrum analyser can contribute over 20dB to a 0.3 dB measurement!

2.4.2. Noise Figure Meter

Common ways of measuring noise in high frequency electronics are based on a simple observation: The noise added by a device cannot be correlated to the noise power supplied to its input. We therefore know how to compute the sum of any noise applied to its input and the DUT's added noise.

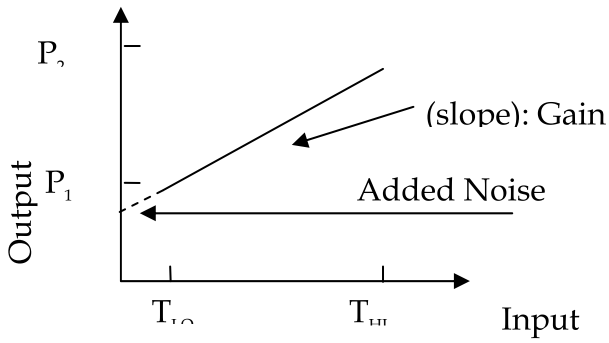

Figure 2: Noise Factor measurement. A "cold" and a "hot" source of the same output impedance are used as inputs to the device under test.

Then, by successively applying a low source noise power N_{LO} and a high source noise power N_{HI} to the DUT input with no other signal present, as in fig. 2.2, the output noise powers will be respectively:

$$P_o^{HI} = GN_{HI} + N_a\,, \quad P_o^{LO} = GN_{LO} + N_a \tag{2.30}$$

From these two measurements both $\mathbf{N_a}$ and **G** are easily determined. The noise source is usually an avalanche diode operated "cold" i.e. is equivalent to a resistor at room temperature (little or no bias current) and "hot" i.e. well into the avalanche regime, where its noise power is equivalent to that of a resistor at a temperature of 30000 K.

Despite the apparent simplicity of this method and the availability of dedicated automatic equipment engineered to perform it, this is an extremely difficult measurement. Any noise measurement procedure measures the cascade of the measurement system, cables, connectors and the Device Under Test. Cables, fixtures and connectors are unknown and often irreproducible attenuators. As such they contribute to the overall system noise factor and undermine the sensitivity and resolution of the measurement.

2.5. Theory of noise in amplifiers

2.5.1. Noise model of an amplifier

We can combine the results of section 2.3 with some circuit analysis to analyse general amplifiers. A noisy amplifier can be modelled as a noiseless amplifier with two noise sources, a voltage and a current noise source connected to its input as in figure 3. These two sources are not necessarily uncorrelated with each other.

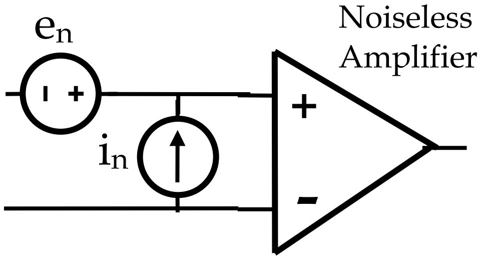

Figure 3: Noise model for an amplifier. The noise contributions can be modelled as two uncorrelated noise sources at the input of an ideal noiseless amplifier.

We shall prove shortly (see T Lee, chapter 13) that the noise factor of an amplifier depends on the source impedance driving it. To do this we draw the driven amplifier as in Figure 4. The amplifier is driven by a Norton circuit describing the Johnson noise of the Y_s source. The noise power of the source is assumed to be uncorrelated to either of the amplifier internal noise sources, e_n and i_n. Please observe that we have not made any claims about e_n and i_n being uncorrelated.

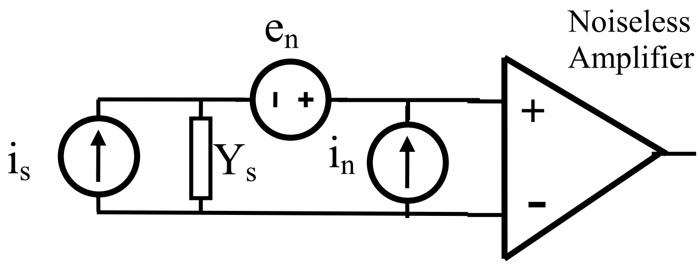

Figure 4: Noise model for an amplifier driven by a noisy Johnson source Y_s

2.5.2. Noise factor of an amplifier

To analyse this network we need a Norton-to-Thevenin transformation of i_s, Y_S then add e_n to the Thevenin source i_S / Y_N, and then a Norton-to Thevenin transformation. The input noise power is proportional to $\left|i_n + Y_s e_n\right|^2$, while the input signal power is proportional to $\left\langle \left|i_s\right|^2 \right\rangle$.

The noise factor is by definition (note that the input impedance of the amplifier cancels!):

$$F = \frac{\left\langle \left|i_s\right|^2 \right\rangle + \left\langle \left|i_n + Y_s e_n\right|^2 \right\rangle}{\left\langle \left|i_s\right|^2 \right\rangle} \tag{2.31}$$

In this equation $\langle x \rangle$ is the expectation value of x.

We can then express i_n as the sum of two sources, $i_c = Y_c e_n$ correlated to e_n by a correlation admittance Y_c and i_u uncorrelated to e_n. Then, we can write :

$$i_n = i_c + i_u = Y_c e_n + i_u \tag{2.32}$$

We can rewrite (2.31) using (2.32) as:

$$F = 1 + \frac{\left\langle \left|i_u + \left(Y_c + Y_s\right) e_n\right|^2 \right\rangle}{\left\langle \left|i_s\right|^2 \right\rangle} = 1 + \frac{\left\langle \left|i_u\right|^2 \right\rangle + \left|Y_c + Y_s\right|^2 \left\langle \left|e_n\right|^2 \right\rangle}{\left\langle \left|i_s\right|^2 \right\rangle} \tag{2.33}$$

we can model the noise sources as equivalent Johnson sources. To do this, we define:

$$Y_C = G_c + jB_c$$

$$Y_s = G_s + jB_s$$

$$\left\langle \left|e_n\right|^2 \right\rangle = 4kT\Delta f R_n$$

$$\left\langle \left|i_u\right|^2 \right\rangle = 4kT\Delta f G_u \tag{2.34}$$

$$\left\langle \left|i_s\right|^2 \right\rangle = 4kT\Delta f G_s$$

By plugging into the definition for F (2.30):

$$F = 1 + \frac{R_n\left(\left(G_c + G_s\right)^2 + \left(B_c + B_s\right)^2\right) + G_u}{G_s} \tag{2.35}$$

Y_c is a characteristic of the amplifier we are considering, and F is a quadratic surface (paraboloid) in Y_s. F has a local minimum in the complex Y_s plane at an optimum admittance:

$$Y_{opt} = G_{opt} + jB_{opt}\,.$$

2.5.3. Minimum F and optimum source admittance

The noise factor has a minimum for an optimum value of $B_s = B_{opt}$

$$B_{opt} = -B_c \tag{2.36}$$

and an optimum value of $G_s = G_{opt}$ such that

$$\left.\frac{dF}{dG_s}\right|_{G_s = G_{opt}} = 0 \Rightarrow \cdots \Rightarrow G_{opt} = \sqrt{G_c^2 + G_u / R_n} \tag{2.37}$$

By substitution, this minimum noise factor is:

$$F_{\min} = 1 + \frac{R_n\left(G_c + G_{opt}\right)^2 + G_u}{G_{opt}} = 1 + 2R_n\left(\sqrt{G_c^2 + G_u / R_n} + G_c\right) \tag{2.38}$$

From (2.37) we note that

$$G_c^2 = G_{opt}^2 - G_u / R_n \tag{2.39}$$

So that (2.35) can be rewritten as :

$$F = F_{\min} + \frac{R_n}{G_s}\left[\left(G_s - G_{opt}\right)^2 + \left(B_s - B_{opt}\right)^2\right] \tag{2.40}$$

The subscript S denotes the components of the source complex admittance Y. The amplifier's noise behaviour is completely characterised by

- the optimum source admittance Y_{opt} which results to a minimum noise factor $F_{\min}$,
- the minimum of the noise factor, $F_{\min}$, and
- the **"noise resistance"** R_N describing how rapidly the noise factor increases for a deviation from the optimum source impedance.

Even though they can, in principle be calculated from the amplifier model, $F_{\min}, R_N$ and Y_{opt} are measured and no effort is made to calculate them. Eq 2.39 is then used to predict the noise figure for a given source admittance.

Equation (2.40) describes a family of constant noise factor circles on the source admittance plane. The optimum noise factor is usually achieved at a different source admittance than the maximum gain is obtained. The optimum noise measure is obtained at yet a different source admittance.

2.5.4. Noise factor of ideal electronics

A lot of the discussion in elementary electronics revolves around the use of ideal signal sources and ideal amplifiers. Realistic sources with finite source impedance of admittance are treated as something of a nuisance; the finite source impedance treated as "parasitic" implying that it makes engineering more difficult than it would otherwise be.

Careful examination of the noise factor of an amplifier (any amplifier!) driven by an ideal voltage source shows that:

$$\lim_{G_s\to\infty,B_s\to\infty} F = \lim_{G_s\to\infty,B_s\to\infty}\left[F_{\min} + \frac{R_n}{G_s}\left[\left(G_s - G_{opt}\right)^2 + \left(B_s - B_{opt}\right)^2\right]\right] = \infty \qquad (2.41)$$

This means that an ideal voltage source driving an amplifier makes the amplifier operate with an infinite noise factor. As a result the SNT at the amplifier output vanishes!. Something similar happens if we drive an amplifier with an ideal current source:

$$\lim_{G_s\to 0,B_s\to 0} F = \lim_{G_s\to 0,B_s\to 0}\left[F_{\min} + \frac{R_n}{G_s}\left[\left(G_s - G_{opt}\right)^2 + \left(B_s - B_{opt}\right)^2\right]\right] = \infty \qquad (2.42)$$

We can now consider the output of the amplifier. This acts as the source to whatever it dives, and consequently, an amplifier with an ideal voltage (or current) output stage forces the noise factor of the instrument it drives to be infinite. The Friis cascade formula suggests that unless the amplifier has infinite gain, which it cannot have, the system noise factor will be infinite.

A final question remains concerning idealities: What if the amplifier has an ideal input and is driven by a finite impedance source? The definition of the noise factor:

$$F = \frac{SNR_i}{SNR_o} \qquad (2.43)$$

Gives the answer. The input circuit is a voltage divider between a Thevenin source and the amplifier input impedance. The input signal power to the amplifier is:

$$P_{s,in} = \operatorname{Re} iv^* = |I|^2 \operatorname{Re}\frac{Z_{in}}{|Z_{in} + Z_s|^2} = |V|^2 \operatorname{Re}\frac{Y_{in}}{|Y_{in} + Y_s|^2} \qquad (2.44)$$

Clearly the input signal power becomes zero when the input resistance ($\operatorname{Re} Z_{in}$) or admittance ($\operatorname{Re} Y_{in}$) of the amplifier vanishes. A purely capacitive input admittance (e.g. the gate of an idealised FET) is not any better. The output signal power of the amplifier must be proportional to the real (not reactive!) power absorbed from the source, and is zero in all these ideal cases. The output SNR, as a result, vanishes. Then we can write:

$$\lim_{SNR_o\to 0} F = \lim_{SNR_o\to 0}\frac{SNR_i}{SNR_o} = \infty \qquad (2.45)$$

This discussion explains why all electronic amplifiers engineered to have near ideal terminal characteristics are guaranteed to have very poor noise performance. The physics, possibly obscured by the mathematical derivation, is that the noise factor is optimised when the signal power transferred to the amplifier is maximised relative to the noise power coupled from the source into the input of the amplifier. Signal power, we have already argued, is the vehicle on which the information contained in the signal is carried.

2.6. Optimising the SNR of a measurement

Any signal processing (such as amplification, sampling, filtering) adds noise to a measurement. The highest signal to noise ratio in a system is found, therefore, at the sensor terminals.

Yet, it is possible to reduce the apparent noise in a measurement. Although thermal noise cannot be eliminated, the total noise power is proportional to the measurement bandwidth. By restricting the measurement bandwidth to what is absolutely essential the signal to noise ratio can be optimised. If S is the total signal power and B the measurement bandwidth,

$$SNR \propto \frac{S}{kTB}$$

We recall that the SNR is a ratio of powers, so we can conclude that the voltage noise amplitude scales as:

$$V_N \propto \frac{1}{\sqrt{B}}$$

A low noise measurement apparently requires a long time, $\Delta\tau \propto 1/B$, to perform. This is only true for low pass measurements. In a bandpass measurement of bandwidth B centred at a carrier frequency f_c the characteristic time required for a measurement is much smaller:

$$\Delta\tau \propto \frac{B}{f_c^2}$$

For this reason, the most sensitive heterodyne receivers perform upconversion, rather than downconversion at their first stage. Extremely narrowband band-pass filters are difficult to engineer at relatively low frequencies (less than 1GHz). At higher frequencies it is relatively easy to exploit molecular resonances in certain materials to make extremely narrowband filters.

The need to reduce the measurement bandwidth is the theoretical foundation of all modulated measurements, including chopper amplifiers, and interferometers. It is also the foundation of telecommunications, since a typical communications link is a modulator-demodulator pair. A radio receiver is simply an extremely sensitive, noise optimised, voltage meter!

A more damaging kind of “noise”, usually of much higher power than thermal noise, is interference. Interference can also be reduced by modulation and filtering, especially when the interference is known to be band limited.

Finally, signal and noise are always statistically independent. Statistical correlation techniques can be used to reduce the noise in the output of a measurement. Much of instrumentation science is concerned with the realisation, in hardware, of statistical recipes.

A simple example of exploiting statistics is the use of two identical instruments to measure a signal. The two instruments will show at their output respectively:

$$V_1 = GS_i + GN_i + N_{a1}$$
$$V_2 = GS_i + GN_i + N_{a2}$$

The SNR of V_1 or V_2 alone is $SNR_1 = SNR_2 = \dfrac{|GS_i|^2}{|GN_i|^2 + |N_a|^2}$

While the SNR of $V_1 + V_2$ is:

$SNR_{1+2} = \dfrac{4|GS_i|^2}{4|GN_i|^2 + 2|N_a|^2} = \dfrac{|GS_i|^2}{|GN_i|^2 + |N_a|^2 / 2} < SNR_1$ because N_{a1} and N_{a2} are uncorrelated!

This technique is used in low noise preamplifiers where a number of low noise transistors connected in parallel are used for the input stage to reduce noise.

If a signal is slowly varying, all the noise of two successive sampled measurements will be uncorrelated. The average of two such measurements will have an enhanced SNR! In the “correlated double sampling” technique, two measurements are performed in rapid succession and subtracted: One is a null measurement, measuring only source noise, while the other is the required measurement. CDS has the advantage of reducing autocorrelated noise.

We will study a number of noise reduction techniques in more detail later in the course.