Microwave noise in semiconductor electronics

Austin Minnich, California Institute of Technology
Spring 2020

Measurement and microwave electronics

Experimental physics and astrophysics relies on the ability to measure very small electrical signals. For various reasons, often these signals are in the microwave frequency domain (e.g. GHz - hundreds of GHz). One example:

This image was taken using an array of radio telescopes over the world.

An enabling technology for advancing science is therefore active electronic devices that amplify these weak signals. A workhorse device for microwave electronics is the low noise amplifier, based on various architectures of transistors.

You can buy them in a package like this:

The signal you care about goes in, and an amplified signal comes out. However, some amount of noise is unavoidably added.

Spring 2020
This class is about fluctuations in semiconductor microwave that impede the accurate measurements of weak signals. Thanks to Carlton Caves (see here) we know that quantum mechanics requires some amount of noise to be added even to a perfect amplifier.

In practice, transistor microwave amplifiers add noise that is at least 5X that of the quantum limit. The goal of this class is to understand the microscopic mechanisms that lead to this extra noise and how they can be mitigated.

Noise mechanisms

Definition: Noise is defined as the fluctuation of a quantity of interest (e.g. voltage or current) from its mean value. It can be described by the theory of stochastic processes as we will discuss.

The fluctuations have various underlying microscopic origins:

• **1/f noise** - observed in many physical systems; its origin is still debated. The frequencies at which 1/f noise is important are much lower than microwave frequencies and so we won’t discuss this mechanism in detail.

• **Generation-recombination noise** - random transitions between the conduction band and trap states lead to fluctuations in the density of mobile electrons, thereby modulating the conductance of the semiconductor. This noise source is also only relevant below microwave frequencies and will not be discussed in detail.

• **Thermal noise** - arises from the thermal motion of mobile electrons. The Nyquist theorem relates the spectral noise power to the absolute temperature and the dissipative part of its conductance.

• **Shot noise** - originating from the fact that electric charge is carried by discrete particles (electrons). Its spectral intensity is proportional to the current.

• **Hot electron noise** - It is found that applying a field to an electron gas alters the spectral noise power from the zero field value (Johnson noise). The generic term for this noise mechanism is hot electron noise. We will extensively discuss this mechanism as it is quite important for transistor amplifiers.
Mechanism of amplification and noise sources of a HEMT LNA
Mechanism of amplification and noise sources of a HBT LNA

Actual layout:
Noise characterization
[Hartnagel 1.5]

Fluctuating currents lead to emission of electromagnetic waves into a load. In experiment, it is easy to measure the power incident on a load. Therefore, the **noise power** is a key quantity. If the output impedance of the device and load impedance are matched, noise power -> available noise power.

**Equivalent noise temperature**

The noise power in a band can be compared to the power emitted by a blackbody at a temperature. We can define the noise temperature as the noise power per unit bandwidth about a frequency emitted into a matched load.

Mathematically:

In **equilibrium**, the noise temperature is just the physical temperature of the noisy element. The Nyquist theorem gives the relation between the spectral density of voltage fluctuations and temperature:

where is the AC impedance of the noise source around a DC bias.

We will prove this later in the class.

Under non-equilibrium conditions, we can define a noise temperature that in general will depend on frequency:
Another way to characterize noise is the noise figure $F$, defined as:

$$F = 10 \log \left( \frac{P_{in}}{P_{out}} \right)$$

Note that the output noise power is referred to input, that is divided by the gain so that

$$F = -10 \log \left( \frac{G}{1} \right)$$

The noise figure depends on the relative values of the output impedance of the circuit and the load impedance. If the optimal load and bias are used, we get $F = 0$, which is independent of external considerations. For non-optimal circuits, four independent parameters are required to describe the two-port circuit. We will discuss these parameters later.

### Modeling of noise in semiconductor devices

A common way to understand the noise performance of a device is to construct an equivalent circuit containing lumped two-terminal elements representing resistances, capacitances, controlled current sources, and so on.

Many of the necessary parameters are extracted from small-signal response experiments (S-parameters). Noise is represented as voltage and current generators at particular locations of the circuit.

Microscopic physical models help to understand the origin of the noise and the appropriate values for their lumped element description. Often, these models are based on a Boltzmann kinetic description (accurate provided all relevant device dimension are greater than a de Broglie wavelength). Noise can be computed using a microscopic theory of fluctuations for a weakly interacting many-particle system, just as response properties are computed for such a system using the Boltzmann equation.

Note that the calculation of noise spectra in the non-equilibrium case is an independent problem from the calculation of response properties. Fluctuational properties provide new physical information about the system of interest.

Drift-diffusion models are simplified equations obtained from the Boltzmann equation. When modified they can account for effects like electron and lattice non-equilibrium in...
a phenomenological way. Noise can be computed from this approach by introducing Langevin forces.

This class will extensively focus on the Boltzmann-level treatment of noise in semiconductors and examine the role of band structure, intervalley scattering, and other factors in setting the noise.

**An introductory example: shot noise**  
[Ambrozy chap 1]

Consider electrons at a p-n junction. Applying a bias leads to electron injection across a barrier and a current.

![Diagram of energy gap in a vacuum tube and a p-n junction](image)

Figure 1-1 (a) Energy gap in a vacuum tube. (b) Energy gap in a p-n junction. ($W$, electron energy; $W_C$, bottom of the conduction band; $W_F$, Fermi level; $W_V$, top of the valence band).

For a given number of electrons that have enough energy to go over the barrier (determined by temperature and DC bias), there are different probabilities for different fractions to actually do so. Therefore there are fluctuations in the current from this random process.

We aim to calculate the spectral noise density of this random process. Start by considering an interval of time $\Delta t$. Take the number of electrons passing across the
barrier in this time to be $\ldots$. For this thought experiment, we take the number of electrons in the starting side to also be $\ldots$, and that they are distinguishable.

Divide the time interval into equal intervals of $\ldots$. The number of electrons in intervals $\ldots$ is $\ldots$. Note that $\ldots$ can be equal or different.

Now consider the probability for electrons 1, 17, and 23 (for example) to pass over the barrier, but no others, in a given time interval $\ldots$? The probability is given by whether those electrons do go over while all others do not. Since we have independent events, overall probability is probability those 3 electrons go over while all others do not.

$P \text{ (go over)} \quad P \text{ (do not)}$

Overall probability:

But actually electrons are not distinguishable. From binomial theory, there are ways to select $\ldots$ electrons out of $\ldots$ total. Therefore, we get the binomial distribution:

We now take the limit of but holding fixed

Spring 2020
gives the average number of electrons passing in time$t$. The distribution becomes:

Second factor:

Third factor:

Fourth factor:

Overall result:

This is the Poisson distribution. It tells us the probability for electrons to pass over the barrier in time interval$t$. Note that it is not continuous in$t$since electrons come in integer numbers. Also the key approximation here is that. It is not always true!
Now let’s compute the fluctuation in the value of \( \ldots \).

The mean of \( \ldots \) is obtained by:

The variance of \( \ldots \) is:

We get the well-known result that for the Poisson distribution, the mean and variance are equal.

Let’s now see how these statistical considerations translate to current noise. The instantaneous current in time interval \( \ldots \) is:

The mean current value is:

The mean square of current fluctuations, directly proportional to noise power, is:
To turn this into the conventional Schottky formula, use the handwaving argument that is on the order of the bandwidth. Then we get:

\[
\text{where the factor of 2 accounts for positive and negative frequencies.}
\]

In practice, physical processes (like transit time of electrons through a vacuum gap, etc) limit the bandwidth to some finite value, even assuming infinite bandwidth electronics.

**Some terminology**

**Amplitude density**

The instantaneous current flowing through a device can be written as:

\[
\text{This decomposition is not that useful for stochastic processes since we generally don’t know the exact in advance. We can instead compute probabilities for to be between to get the amplitude density function:}
\]
Its dimension is \( \_ \_ \_ \_ \_ \_ \_ \_ \_ \). We can also get an amplitude density function for voltage, \( \_ \_ \_ \_ \_ \_ \_ \_ \_ \), with units \( \_ \_ \_ \_ \_ \_ \_ \_ \_ \) .

For shot noise, \( \_ \_ \_ \_ \_ \_ \_ \_ \_ \) is also a Poisson function since \( \_ \_ \_ \_ \_ \_ \_ \_ \_ \) is proportional to \( \_ \_ \_ \_ \_ \_ \_ \_ \_ \) : 

Exercise: compute the amplitude density function for \( \sin(x) \).

**Power density**

Since we have the mean square of current fluctuations, we can define a noise power:

The power density is the power per unit bandwidth. For shot noise, we have:
We see that for shot noise the spectrum is white, or frequency-independent. In general, we often calculate \[ S(f) \], and leave them in that form. Therefore we define the spectral density of current fluctuations:

\[
S(f) = \frac{1}{2\pi} \frac{dP(f)}{df}
\]

We can also define rms noise current and voltage:

Note that we can only draw conclusion about the power spectra from the above discussions, not the noise or current spectra.

**Elements of probability - distributions and density functions**

[Ambrozy Chap 2]

In general, we don’t know the exact signal originating from a stochastic process. However, we can predict the probabilities of instantaneous values from an understanding of the microscopic origin.

Some terms:

A random trial is an event with a non-unique outcome.

Elementary events = possible outcomes. Mathematically, a possible event is a subset of a set containing the elementary events, which is known as an event space.

Kolmorogov axioms: a probability function is defined on subsets of if:
if are mutually exclusive (either finite or infinite series is ok).

A random variable can be defined as a real function over the set .

Calling the elementary event , we call the real number characterizing the elementary event as . Interpretation: tells us the numerical result of a trial.

Example: consider electrons passing over a barrier. The i-th event is electrons passing over the barrier. Here the random variable is also .

But, if two barriers are connected in parallel and we measure the total current, the relevant random variable is , and various combinations of elementary events (individual ) are relevant for the random variable .

**Distributions**

Now let’s define a stochastic process: a single parameter assembly of random variables, , where the parameter continuously covers the time set.

Example: (defined in the previous example) is a function of time as a parameter and is composed of a random variable defined on two event spaces.

We can now define a distribution function of a random variable .

It gives the probability of having a value less than .

Properties:

1. Monotonically increasing
2. Limiting values of 0 and 1:
3. Continuous from the left:

From the definition, the probability that

Discrete and continuous distribution functions can be defined.

**Discrete:** the possible values of are a finite or infinite series.

Define so that the distribution function is:

As , the sum of tends to unity.

**Continuous:** a distribution is continuous if a function exists so that
We can define discrete probabilities and density function as:

### Discrete

### Continuous

Although the density function of a discrete function is not strictly defined, it can be approximated by a continuous function for sufficiently large measurement numbers:

Here \( \Delta \) is the uniform increment of the random variable (voltage, current, etc.).

We actually already did this conversion when we derived the Poisson distribution.

**Expected value and standard deviation**

For a discrete distribution, we have values with probabilities .

Expected value is:

Notation:

For a continuous distribution,

where we require that

is often referred to as the first-order moment of the distribution.

Some properties:

Homogeneity

Additivity

Spring 2020
If we have two statistically independent random variables, then

This condition is satisfied if the joint probability, defined as:

has the density function that separates:

We can also characterize the fluctuations of the random variable about the mean. We introduce the standard deviation:

It is also described as the square root of the expectation value of .

We can rewrite the definition in a more convenient form:

Hence the variance, , is also the difference between the expectation value of the random variable squared, and the squared expected value.

Side note: in practice, using the first formula can be numerically preferable. Say we have some error in our knowledge of the true expected value. Then,

On the other hand, using the other formula,
The error is higher in this case since

If we know the density function, we can compute the variance:

Discrete:

Continuous:

Writing the continuous formula in another way:

Here is the second-order moment.

Some properties:

Effect of scaling and shift.

So standard deviation is multiplied by but additive term leaves it unchanged.

Variance of mutually independent random variables (proof to do in HW):

From this result, we see that if independent random variables have equal variance, then
Higher-order moments can be defined as

Higher-order moments tell us about asymmetrical features of the density function. The third-order moment is tells us about skewness, fourth-order tells us about kurtosis (nature of the tail of the distribution).

**Characteristic functions and sum distributions**

The characteristic function is the Fourier transform of the density function. Definition:

Expected value:

Continuous distribution:

We can Fourier transform back using:

Some properties:

1. If , since . If , one can show that
2. If is even, is real. Otherwise,
Derivatives:

So we can get moments of the density function from derivatives of .

With the characteristic function, we can compute the sum distribution of random variables. It is useful because in devices several noise sources are present simultaneously.

Define the random variable of the sum distribution as:

where are mutually independent. Then,

due to mutual independence. So, the Fourier transform of sum distribution density function is the product of the individual characteristic functions.

Consider a two-term example:

Define the density functions as
The characteristic function for is just .

Using the properties of Fourier transforms to go back to a density function, we get

Another way to see it: the joint probability for independent events is given by the project . But since we care about the value of the sum random variable, defined as , we have . We then need to add up all the possible ways to get , i.e. integrate for continuous density functions.

Example: say we have two dice, and the sum random variable is the sum of the numbers that we see. The probability of any number on one dice is .

Probability to get 2 or 12?

Probability to get 5? Rewriting convolution equation for a discrete probability:

HW: try the calculation for 3 dice.

Central limit theorem

The noise we measure is often a sum of random phenomena with the same distribution. Can we figure out the sum distribution, even if the number of components is large and perhaps not known?

The central limit theorem tells us that the sum distribution follows a normal distribution as the number of independent random variables tends to infinity.

Derivation: define as mutually independent random variables with expected value = 0, std dev = , skewness . (not required but helpful)
The sum random variable is:

Expected values of sum:

Variance of sum:

We also have:

Define a normalized sum random variable as:

Then:

Now let’s write the characteristic function of the normalized sum distribution:

We can expand the function using the definition of characteristic function:

Since and , we have
The sum characteristic function is given as a product of \( \ldots \). Let’s take a log of it to turn the products into sums. We need the logs of individual characteristic functions, and let’s take the limit as \( \ldots \), to get

Therefore,

The Fourier transform of sum distribution density function is a Gaussian. Therefore, the original density function is a Gaussian (normal) distribution! Applying an inverse FT, and using evenness of Gaussian functions:

We have thus obtained the normal distribution. In the general form, we get:

Example: say we have a random variable \( \ldots \) that depends on time. The density function gives the probability of individual instantaneous values and should give a normal distribution.

Exercise: plot the amplitude density function of equal-amplitude, incoherent sine waves - you should get a normal distribution.
Note 1: the instantaneous range of values for many independent random processes tends to a normal distribution. This statement says nothing about the spectral power (i.e. frequency content) of the time-varying signal. If sample are uncorrelated in time and are independent random processes, we get Gaussian white noise.

Note 2: a white power spectrum does not necessarily imply instantaneous values following a normal distribution.

Note 3: the distribution of a sum of two random variables with equal expected values and normal distributions is also normal!

**Normal distribution**

Density function:

```
```

Distribution function:

```
```

Probability to fall within limits
Expected value: since a Gaussian is even.

Expected value of half distribution:

Including the absolute value of the negative axis, we get:

We can calculate higher-order moments by differentiation of the characteristic function.

Here is the Fourier transform of a normal distribution with zero mean and arbitrary : 

We get the derivatives and hence moments as:
A useful property of the normal distribution:

The previously derived properties for independent random variables still holds for normally distributed random variables, and the sum variable is also normal.

**Discrete distributions:**

We will examine the binomial and Poisson distributions in more detail.

First, consider two sets of events that have exactly two, mutually exclusive outcomes.

Example: emission or non-emission of an electron from a vacuum-tube cathode.

Example: which of two electrodes receives an emitted electron.

Call \( p \) the probability of the event occurring (say, electrode A is hit by the electron). Probability of not occurring is \( 1 - p \).

We can define the values of random variable \( X \) as \( 1 \) in the first case, and \( 0 \) in the second case.

Expected value:

Variance:
Now repeat the experiment times. The probability of the event occurring times and not occurring times is the product of individual probabilities if experiments are independent:

Since we don’t care about the particular order of the experiments, we count all the ways events can occur from total events as:

Thus, the probability is

which is the binomial distribution.

In terms of a sum random variable , we have

As the experiments are independent, we have

Mean Variances

If and , then

But, if and , then

Limiting cases: if and so that , we get the Poisson distribution:
We can get moments of the Poisson distribution from the characteristic function:

First derivative:

Second derivative:

Since \( n \to \infty \), we get:

Let’s consider a limiting case of the Poisson distribution for \( \lambda \to \infty \). We use the Stirling formula:

Substitute it into the Poisson distribution and take the log:

Use the identity

and note that if \( n \) is close to the mean \( \lambda \), then \( n! \approx \lambda^n \sqrt{2\pi\lambda} \), so that:

We can then write:
Since \( N \) is large enough, we can neglect the middle two terms to get:

Taking the continuous limit, we see the above is the density function of a Gaussian distribution! Therefore, if \( N \) is sufficiently large, i.e.

then the Poisson distribution can be accurately approximated by the normal distribution in the vicinity of \( N \). The condition is that \( N \) is sufficiently large which means a sufficiently large average current. Typically in electronic devices these conditions are satisfied, and so the instantaneous current values for shot noise follow a normal distribution.

Binomial distribution: we now consider the case where \( N \) is not small, in the vicinity of \( N \). We have
We will obtain the relative probability of . If , then we need the extra terms.

Similarly, decreases by the terms.

Extract a factor of from each term in the top, and a factor of in the bottom. These terms and the fraction term yield unity:

The remaining equation is:

Take the log of this equation.

Use the assumption that and the approximation of log:

We can perform these algebraic series to get:
Note that and that the numerator .

(True exactly if ).

We thus get:

We can get with the help of the Stirling formula:

and so finally we conclude that for sufficiently high mean , the normal distribution can be used in place of the binomial distribution. This approximation is nearly always accurate in electronic devices.

**Continuous distributions**

Since electrons are discrete particles, strictly we should use discrete distributions. However, due to the large number of charge carriers that compose typical electrical
currents, we can approximate discrete distributions with continuous ones. Let’s examine a few of these distributions.

Consider a **uniform distribution**. A continuous random variable has a uniform distribution if in the interval the density function is:

Outside, .

Note: the density function typically has a physical dimension.

Example: voltage distribution would have units.

Consider a uniform distribution with zero expected value, limits .

Density function:

Note that this function satisfies the requirement:

Variance and std dev:

Fourth-order moment:

Kurtosis:
The next distributions have relevance for methods of measuring stochastic signals.

Example: in quadratic detection, the instantaneous voltage or current value is squared. We would therefore like to know the distribution of instantaneous values, squared, as well as the k-term sum distribution, obtained from the former distribution.

Start with a normal distribution of zero expected value, std dev \( \sigma \).

The probability that \( X \) is:

Since normal distributions are symmetric about 0, we have:

The density function is:

This is known as a chi-square distribution with one degree of freedom (\( \chi^2 \)).

If we have a random variable that is a sum of squared random variables each having the same normal distribution:

we get a \( \chi^2 \) distribution with \( k \) degrees of freedom. The density function is:
This result can be derived from the characteristic function.

is the generalization of the factorial function:

Expected value and std dev of the distribution are obtained in the usual way:

We see the second- and fourth-order moments. Since has a normal distribution,

and

The relative std dev of the distribution is

If is large, the distribution tends to the normal distribution.
For full-wave, piecewise linear detection, the distribution of the measured output signal follows the distribution of \( \cdot \). If \( X \) has a normal distribution, density function of has a distribution as:

Expected value:

Variance:

We usually measure an averaged signal, and hence we care about the sum distribution:

e.g. a \( k \)-term sum. For a small number of terms, the distribution is complicated. For a large number of terms, we get a normal distribution with

Expected value:

Variance:

Relative std dev:
A final problem: consider narrow-band noise with a low relative bandwidth. We want the distribution of the envelope curve. We can consider the signal as a carrier of frequency, which is stochastically modulated in amplitude and phase. The random variable is:

The instantaneous values of envelope and phase angle are and . We can write as:

where and are mutually independent random variables with normal distributions (if the original distribution is also normal).

Let and have the common density function, . Since they are independent and normal, we have:

Transforming to polar coordinates so the density function depends on ,

The function doesn’t depend on , so

Therefore, the single variable density function of the envelope curve is:

Spring 2020
which is a Rayleigh distribution.

The cumulative density function is

so that the probability of an envelope amplitude higher than $x$ is

Expected value:

Variance:

Relative std dev of a k-term sum:

Last distribution: the exponential distribution, defined by the density function:

The distribution parameter $\lambda$ is an arbitrary positive number.

Distribution function:

Moments:
Interesting property: assume follows the exp distribution and represents a time interval. Say we’ve waited 1 second for the event to occur. What is the probability it takes another second? The value doesn’t depend on how long we’ve waited due to the exponential function and definition of conditional probability. Exp is memoryless.