Microwave noise in semiconductor electronics

Austin Minnich, California Institute of Technology
Spring 2020

Measurement and microwave electronics

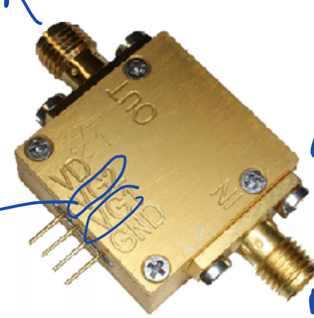
Experimental physics and astrophysics relies on the ability to measure very small electrical signals. For various reasons, often these signals are in the microwave frequency domain (e.g. GHz - hundreds of GHz). One example:

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: output

gate voltages



input

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

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.

$$\underbrace{A}_{\text{added photon number}} \geq \frac{1}{2}\left|1 - \underbrace{G^{-1}}_{\text{gain}}\right| \approx \frac{1}{2} = \frac{1}{2} \text{ photon} \qquad T_N \sim \left(\frac{f}{20\,\text{GHz}}\right) K$$

In practice, transistor microwave amplifiers add noise that is ~~at least 5X~~ 3-5X 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

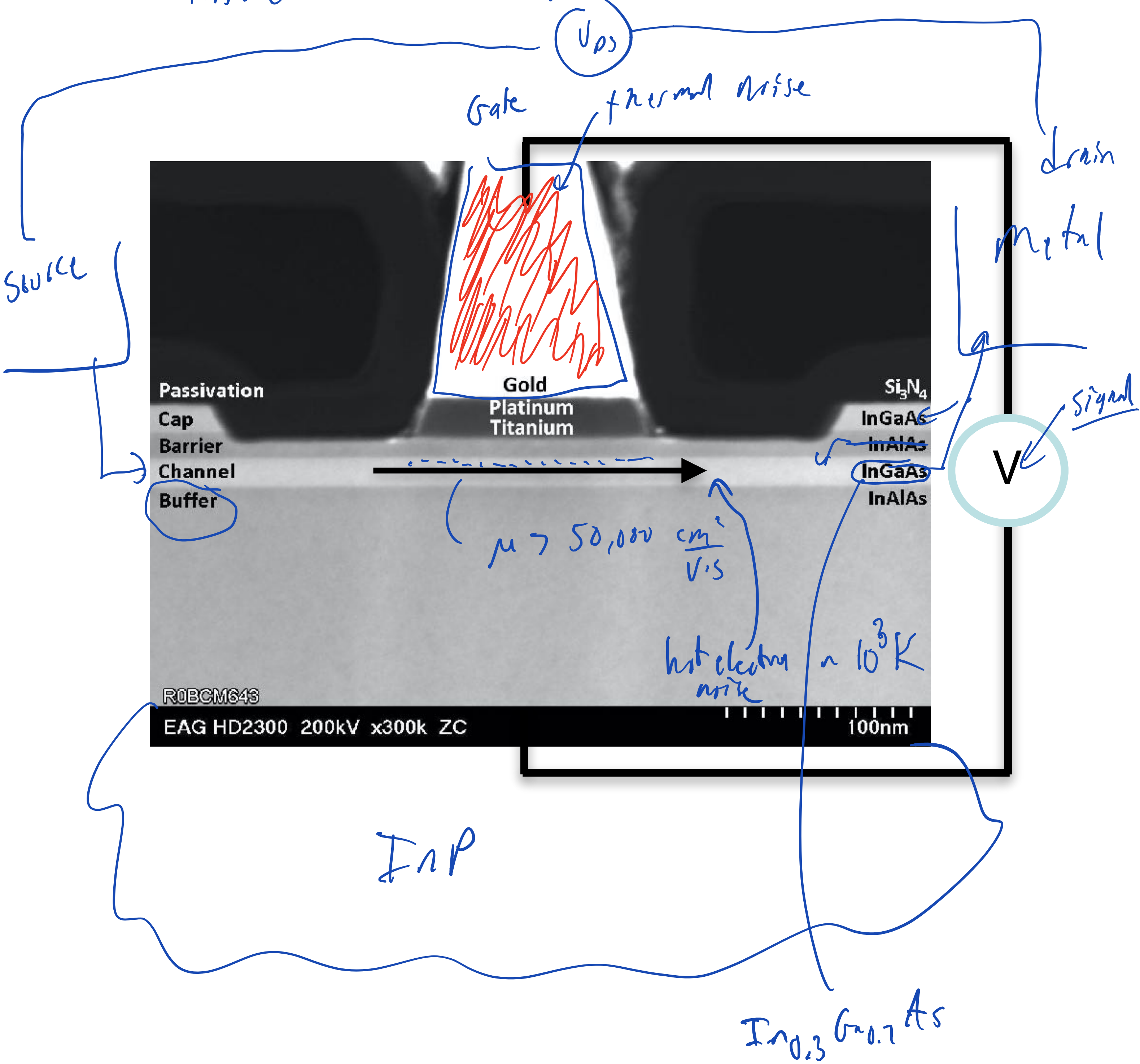
Definition: Noise is defined as the fluctuation of a quantity of interest (e.g. voltage or current) from its mean value. It can be described by the theory of stochastic processes as we will discuss.

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. ~ KHz
- **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. trap
- **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. $\rightarrow$ resistance
- **Shot noise** - originating from the fact that electric charge is carried by discrete particles (electrons). Its spectral intensity is proportional to the current. absence of scattering
- **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 microwave amplifiers.

Mechanism of amplification and noise sources of a HEMT LNA

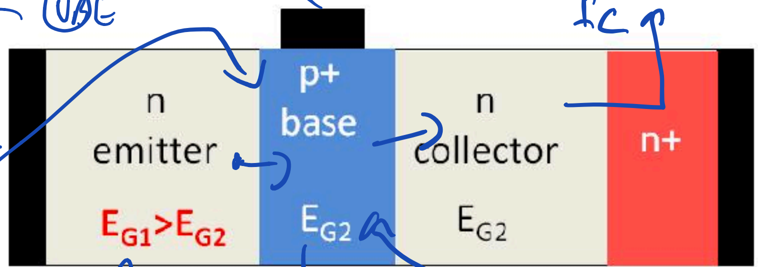
High electron mobility transistor



Mechanism of amplification and noise sources of a HBT LNA

Heterojunction bipolar transistor

V_{BE} I_C



shot noise

[Image: Alam/Purdue]

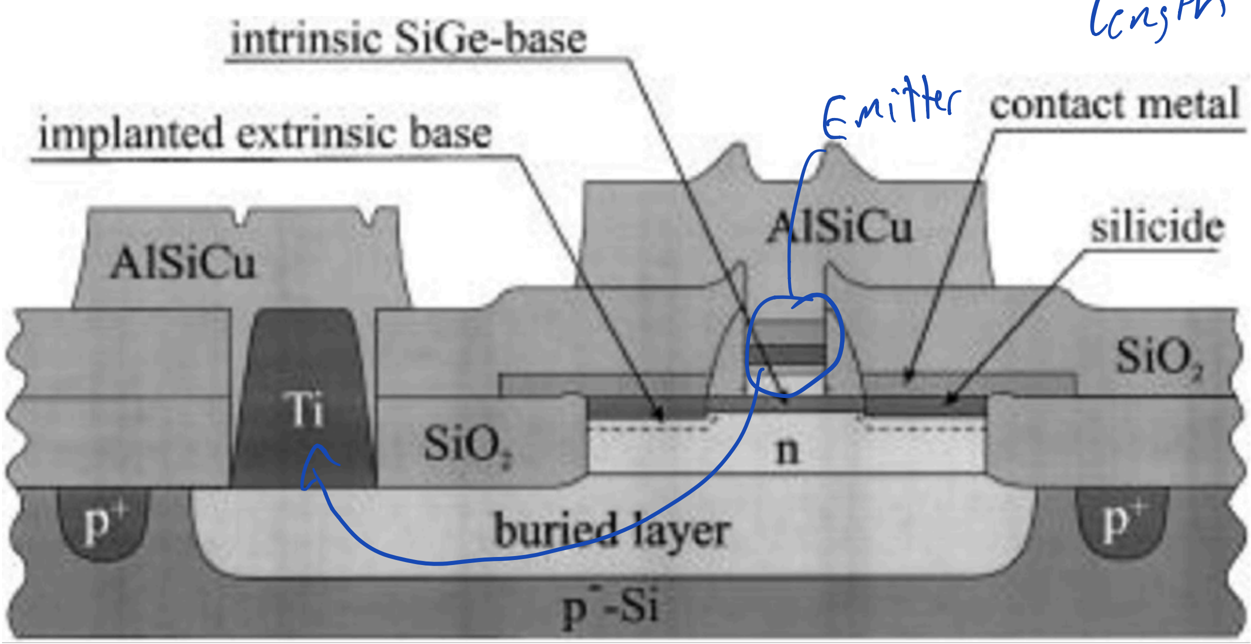
I_B np pn resistance → thermal noise

$\frac{I_C}{I_B} = \beta \gg 1$

$l \sim$ tens of nm $\ll$ Minority carrier diffusion length

Actual layout:

Emitter



Heterojunction ⟺ SiGe base

[ecsdl.org]

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

Mathematically:

$$T_n(f) = \frac{\Delta P_n(f)}{k_B \Delta f} \leftarrow \text{experiment}$$

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 or current fluctuations and temperature:

$$S_v(f) = 4k_B T_0 \, \mathrm{Re}[Z(f)] \sim 4k_B T_0 R$$

$$S_I(f) = 4k_B T_0 \, \mathrm{Re}[Z^{-1}(f)] \sim 4k_B T_0 R^{-1}$$

where $Z(f)$ 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: $T_0 \to T_n(f)$

$$S_v(f) = 4k_B T_n(f) \, \mathrm{Re}[Z(f)]$$

$$= \Delta P_n(f) / \Delta f$$

Another way to characterize noise is the noise figure $NF(f)$, defined as:

$$NF(f) \equiv 10 \log_{10} \frac{T_n(f)}{T_0}$$

Note that the output noise power is referred to input, that is divided by the gain G so that $T_n \to T_n/G$.

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 NF_{min}, 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~~. sources

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.

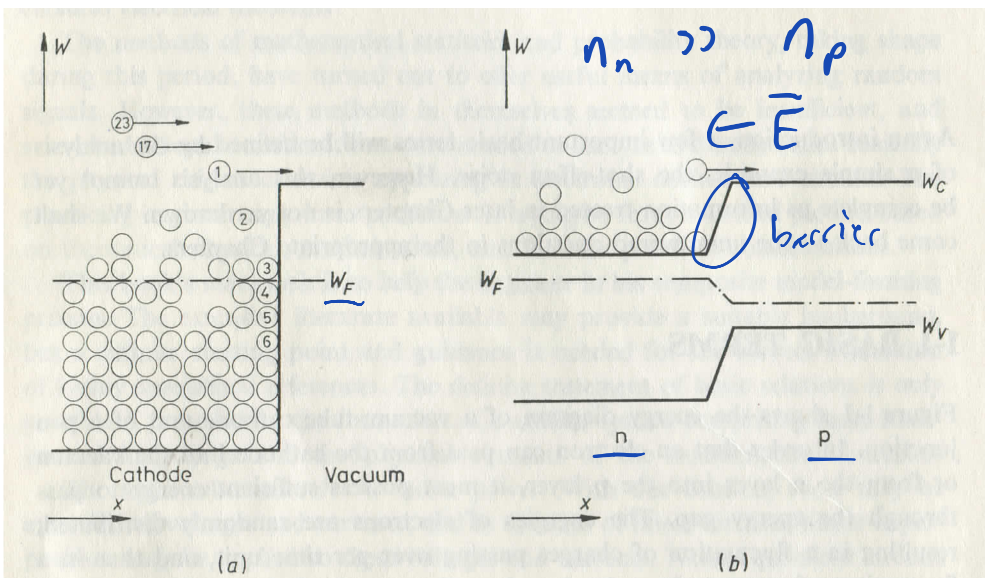


Figure 1-1 *(a)* Energy gap in a vacuum tube. *(b)* Energy gap in a p–n junction. (W, electron energy; W_C, bottom of the conduction band; W_F, Fermi level, W_V, top of the valence band).

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 T . Take the number of electrons passing across the

not temperature!

barrier in this time to be N. For this thought experiment, we take the number of electrons in the starting side to also be N, and that they are distinguishable.

Divide the time interval into equal intervals of Δt. The number of electrons in intervals $\Delta t_1 \ldots \Delta t_i$ is $n_1 \ldots n_i$. Note that $n_1 \ldots n_i$ 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 Δt? 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 (n_i) go over while all others do not.

P (go over) $\left(\frac{\Delta t}{T}\right)^{n_i}$ P (do not) $\Big(\underbrace{1 - \frac{\Delta t}{T}}_{\text{P of single } e^- \text{ not going over}}\Big)^{N-n_i}$

Overall probability:

$$\left(\frac{\Delta t}{T}\right)^{n_i}\left(1 - \frac{\Delta t}{T}\right)^{N-n_i}$$

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

$$f(n_i) = \binom{N}{n_i}\left(\frac{\Delta t}{T}\right)^{n_i}\left(1 - \frac{\Delta t}{T}\right)^{N-n_i}$$

We now take the limit of $T \to \infty, \quad N \to \infty, \quad \Delta t / T \to 0$

but holding fixed

$$\lambda \equiv \frac{N \Delta t}{T} = \text{some finite number}$$

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

per unit time.

λ gives the average number of electrons ~~passing in time~~ Δt .

The distribution becomes:

$$f(n) = \binom{N}{n}\left(\frac{\lambda}{N}\right)^n\left(1-\frac{\lambda}{N}\right)^{N-n}$$

$$= \frac{\lambda^n}{n!}\,\frac{(N(N-1)\ldots(N-n+1))}{N^n}\left(1-\frac{\lambda}{N}\right)^{-n}\left(1-\frac{\lambda}{N}\right)^{N}$$

Second factor:

$$N\to\infty, \quad \frac{N(N-1)\ldots(N-n+1)}{N^n}\to 1$$

Third factor:

$$\lim_{N\to\infty}\left[\left(1-\frac{\lambda}{N}\right)^{-n}\right] = 1$$

Fourth factor:

$$\lim_{N\to\infty}\left[\left(1-\frac{\lambda}{N}\right)^{N}\right] = e^{-\lambda}$$

→ Overall result:

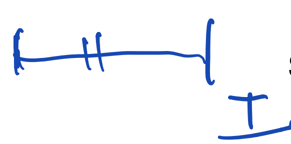
$$\lim_{\substack{N\to\infty \\ \Delta t/t\to 0}} f(n) = \frac{\lambda^n}{n!}e^{-\lambda}$$

This is the Poisson distribution. $f(n)$ tells us the probability for n electrons to pass over the barrier in time interval dt .

Note that it is not continuous in n since electrons come in integer numbers.

Also the key approximation here is that $\Delta t/t \ll 1$. It is not always true!

$$\frac{\Delta t}{T} \ll 1$$



Now let's compute the fluctuation in the value of n.

The mean of n, $M(n)$ is obtained by:

$$\bar{n} = M(n) = \sum_{n=0}^{\infty} n f(n) = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=1}^{\infty} \lambda \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda}$$

$$\bar{n} = \lambda \underbrace{\sum_{n=0} \frac{\lambda^n}{n!} e^{-\lambda}}_{\sum_{n=0}^{\infty} f(n) = 1} \longrightarrow M(n) = \lambda$$

The variance of n is:

$$\sigma^2(n) = \overline{(n-\bar{n})^2} = \overline{n^2} - (\bar{n})^2$$

$$\overline{n^2} = \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=1}^{\infty} n \lambda \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} = \sum_{n=0}^{\infty} (n+1)\lambda \frac{\lambda^n}{n!} e^{-\lambda}$$

$$= \lambda \cdot \underbrace{\lambda}_{\bar{n}} + \lambda \cdot 1 = \lambda^2 + \lambda \Rightarrow \sigma^2(n) = \lambda^2 + \lambda - \lambda^2 = \lambda = \bar{n}$$

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 T is: i_T = total instantaneous current

The mean current value is:

$$I = \overline{i_T} = \frac{\lambda q T}{T} = \lambda q$$

The mean square of current fluctuations, directly proportional to noise power, is:

$$\overline{i^2} \equiv \overline{(i_T - I)^2} = \text{var}\left(\frac{qn}{T}\right) = \frac{q^2}{T^2} \text{var}(n)$$

$$= \frac{q^2 \lambda T}{T^2} = \frac{q}{T} \cdot q\lambda = \frac{qI}{T}$$

To turn this into the conventional Schottky formula, use the handwaving argument that T^{-1} is on the order of the bandwidth Δf. Then we get:

$$\boxed{\overline{i^2} = 2qI\Delta f}$$

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:

$$\underbrace{i_T(t)}_{\text{total}} = I + i(t)$$

This decomposition is not that useful for stochastic processes since we generally don't know the exact $i_T(t)$ in advance. We can instead compute probabilities for $i_T(t)$ to be between I_0 & $I_0 + \Delta I$.

to get the amplitude density function:

$$f(I_0)\,\Delta I = P\left(I_0 \le i_T(t) \le I_0 + \Delta I \right)$$

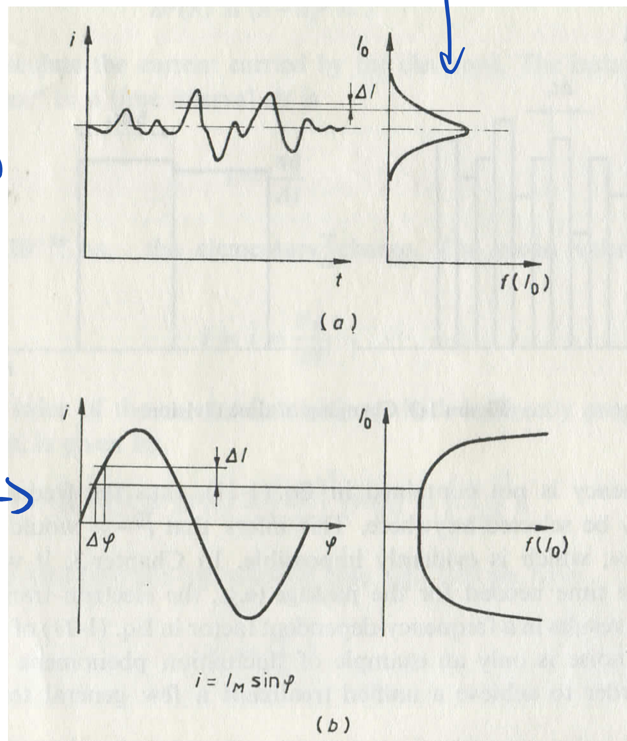


Figure 1-4 Time function and amplitude density function. *(a)* Stochastic signal. *(b)* Sine-wave signal.

sin

Its dimension is A^{-1}. We can also get an amplitude density function for voltage, $f(U_0)$, with units V^{-1}.

For shot noise, $f(I_0)$ is also a Poisson function since $i_T(t)$ is proportional to n_i

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:

$$P = \underbrace{\overline{i^2}}_{\text{fluctuation}} R = 2qIR\underbrace{\Delta f}_{\text{bandwidth}}$$

The power density is the power per unit bandwidth. For shot noise, we have:

$$p = \frac{P}{\Delta f} = 2qIR \quad \left[\frac{W}{Hz}\right]$$

We see that for shot noise the spectrum is white, or frequency-independent. In general,

$$\rho = \rho(\underset{\text{frequency}}{f}) \quad ; \quad \rho = \rho(\omega) \qquad \Bigg| \qquad \rho = \int_0^\infty \rho(f)\, df = \text{total noise}$$

We often calculate $\overline{i^2}$, $\underset{\text{voltage}}{\overline{u^2}}$, and leave them in that form. Therefore we define the spectral density of current fluctuations:

$$S = S(f) \quad [A^2/Hz \,;\, V^2/Hz]$$
$$S = S(\omega) \quad [A^2/\text{rad}\, s^{-1} \,;\, V^2/\text{rad}\, s^{-1}]$$

We can also define rms noise current and voltage:

$$i_{rms} = \sqrt{\overline{i^2}} \quad ; \quad u_{rms} = \sqrt{\overline{u^2}}$$

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. roll a dice

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

Kolmorogov axioms: a probability function $P(A)$ is defined on subsets of Ω if:

$$0 \le P(A) \le 1 \qquad P\left(\sum_k A_k\right) = \sum_k P(A_k)$$
$$P(\Omega) = 1$$

if $A_1 \ldots A_k$ are mutually exclusive (either finite or infinite series is ok).

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

Calling the elementary event ω , we call the real number characterizing the elementary event as ξ . Interpretation: $\xi(\omega)$ tells us the numerical result of a trial.

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

But, if two barriers are connected in parallel and we measure the total current, the relevant random variable is $n_i' = n_{i1} + n_{i2}$, and various combinations of elementary events (individual n_i) are relevant for the random variable n_i' .

Distributions

Now let's define a stochastic process: a single parameter assembly of random variables, $\xi(t)$, where the parameter t continuously covers the time set.

Example: n_i' (defined in the previous example) is a function of time as a parameter and is composed of a $\underbrace{\text{random variable}}_{\xi}$ defined on two event spaces.

We can now define a distribution function $F(x)$ of a random variable ξ .

$$F(x) = P(\xi < x)$$

It gives the probability of ξ having a value less than x .

Properties:

1. Monotonically increasing
2. Limiting values of 0 and 1: $\lim_{x \to -\infty} F(x) = 0$

$$\lim_{x \to \infty} F(x) = 1$$

3. Continuous from the left:

(right) $\lim_{x \to x_0 - \epsilon} F(x) = F(x_0)$

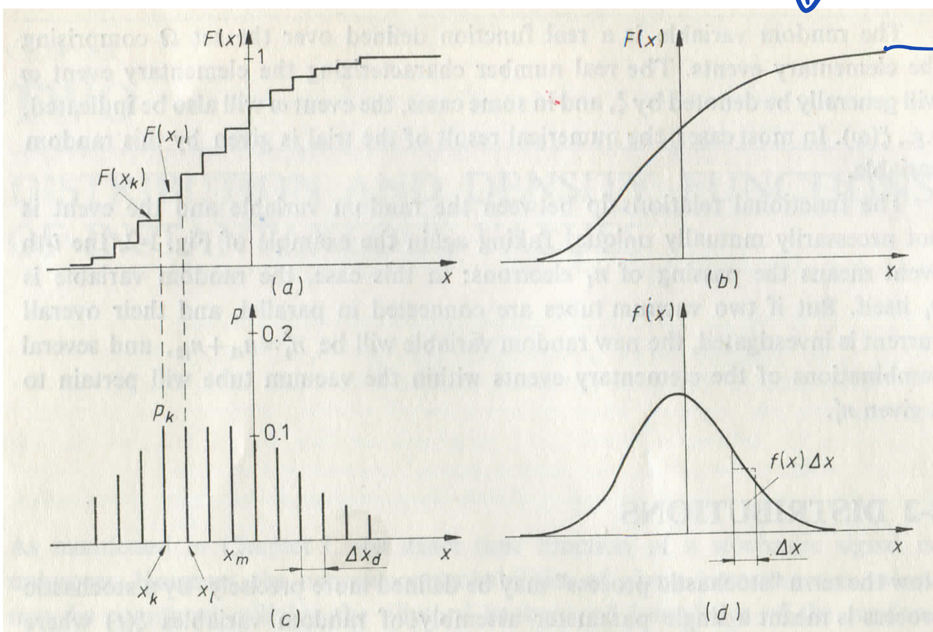


Figure 2-1 *(a)* Discrete distribution. *(b)* Continuous distribution. *(c)* Probabilities of discrete distribution *(d)* Density function of continuous distribution.

From the definition, the probability that $P(x_k \le x \le x_l) = F(x_k) - F(x_l)$

Discrete and continuous distribution functions can be defined.

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

Define $p_i = P(\xi = x_i)$ so that the distribution function is:

$$F(x) = P(\xi < x) = \sum_{x_i < x} p_i$$

As $x \to \infty$, the sum of p_i tends to unity.

Continuous: a distribution is continuous if a function $f(x) \ge 0$ exists so that

$$F(x_2) - F(x_1) = P(x_1 \le x \le x_2)$$

$$= \int_{x_1}^{x_2} f(x)\,dx$$

We can define discrete probabilities p_i and density function $f(x)$ as:

Discrete

$$p_i = F(x_{i+1}) - F(x_i)$$

Continuous

$$f(x) = \frac{dF}{dx}$$

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:

$$p_i = f(x)\,\Delta x_d \qquad f(x) = \frac{p_i}{\Delta x_d}$$

Here Δx_d 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 $x_1 \ldots x_i$ with probabilities $p_1 \ldots p_i$.

Expected value is: $M(\xi) = \sum_{i=1}^{\infty} x_i p_i$

$$p_i = \lim_{N\to\infty} \frac{n_i}{N}$$

n_i – # observations of x_i

N – total # of trials

Notation: $\langle \xi \rangle = M(\xi)$

For a continuous distribution,

$$M(\xi) = \int_{-\infty}^{\infty} x f(x)\,dx$$

where we require that

$$\int_{-\infty}^{\infty} |x| f(x)\,dx < \infty$$

$M(\xi)$ is often referred to as the first-order moment of the distribution.

Some properties:

Homogeneity

$$M(c\xi) = cM(\xi)$$

Additivity

$$M(\xi_1 + \ldots + \xi_n) = M(\xi_1) + \ldots + M(\xi_n)$$

If we have two statistically independent random variables, then

$$M(\xi_1 \xi_2) = M(\xi_1) M(\xi_2)$$

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

$$p(\xi_1 < x_1, \xi_2 < x_2) = G(x_1, x_2)$$

has the density function that separates:

$$g(x_1, x_2) = f_1(x_1) f_2(x_2)$$

We can also characterize the fluctuations of the random variable about the mean. We introduce the standard deviation: $D(\xi) = \sqrt{M[\xi - M(\xi)]^2}$

It is also described as the square root of the expectation value of $(\xi - M(\xi))^2$.

We can rewrite the definition in a more convenient form: $D^2(\xi) = M(\xi^2) - M(\xi)^2$

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

HW

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,

$$M(\xi) \to M(\xi) + \epsilon$$

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

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

Discrete: $$D^2(\xi) = \sum_i (x_i - M(\xi))^2 p_i$$

Continuous: $$D^2(\xi) = \int_{-\infty}^{\infty} (x - M(x))^2 f(x)\,dx$$

Writing the continuous formula in another way:

$$D^2(\xi) = \mu_2 - \mu_1^2$$

Here μ_2 is the second-order moment. $\mu_2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx$

Some properties:

Effect of scaling and shift.

$$D^2(a\xi + b) = a^2 D^2(\xi)$$

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

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

$$D^2(\xi_1 + \dots + \xi_n) = D^2(\xi_1) + \dots + D^2(\xi_n)$$

From this result, we see that if independent random variables have equal variance, then

$$D^2(\xi_1 + \dots + \xi_n) = n D^2(\xi)$$

Higher-order moments can be defined as

$$\mu_n = \int_{-\infty}^{\infty} x^n f(x)\,dx$$

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:

$$\phi(v) = M\left(e^{j\xi v}\right) = M(\cos \xi v) + j\,M(\sin \xi v)$$

Expected value: (discrete):

$$\phi(v) = \sum_i e^{j x_i v} p_i$$

Continuous distribution:

$$\phi(v) = \int_{-\infty}^{\infty} e^{jxv} f(x)\,dx$$

We can Fourier transform back using:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jvx}\,\phi(v)\,dv$$

Some properties:

1. If $v=0$, $\phi(v)=1$, since $\int_{-\infty}^{\infty} f(x)\,dx = 1$. If $v \neq 0$, one can show that $|\phi(v)| \leq 1$.
2. If $f(x)$ is even, $\phi(v)$ is real. Otherwise, $\phi(-v) = \phi^*(v)$.

complex conjugate

Derivatives: $\frac{d\phi}{dv} = j\int_{-\infty}^{\infty} e^{jvx}\, \underbrace{x f(x)}\, dx \xrightarrow{v=0} jM(\xi) = j\mu_1$

$$\frac{d^2\phi}{dv^2} = -\int_{-\infty}^{\infty} e^{jvx} x^2 f(x)\, dx \xrightarrow{v=0} -\mu_2$$

etc...

So we can get moments of the density function from derivatives of $\phi(v)$.

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:

$$\eta = \xi_1 + \ldots + \xi_n$$

where $\xi_1 \ldots \xi_n$ are mutually independent. Then,

$$\phi_\eta(v) = M(e^{jv\eta}) = M(e^{jv\xi_1} \ldots e^{jv\xi_n})$$
$$= M(e^{jv\xi_1}) \ldots M(e^{jv\xi_n})$$

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

$$\phi_\eta(v) = \phi_1(v) \ldots \phi_n(v)$$

Consider a two-term example: $\eta = \xi_1 + \xi_2$

Define the density functions as

$$f_1(x_1),\quad f_2(x_2),\quad \text{sum}: f(y)$$

The characteristic function for $\phi_y(v)$ is just $\phi_1(v)\phi_2(v)$.

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

$$f(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jvy}\,\phi_1(v)\,\phi_2(v) = \int_{-\infty}^{\infty} f_1(x_1)\,f_2(y-x_1)\,dx_1$$

Another way to see it: the joint probability for independent events is given by the ~~project~~ product $f_1(x_1) f_2(x_2)$. But since we care about the value of the sum random variable, defined as $y = x_1 + x_2$, we have $x_2 = y - x_1$. We then need to add up all the possible ways to get y, 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 $1/6$.

Probability to get 2 or 12? $\frac{1}{6}\cdot\frac{1}{6} = 1/36$

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

$$p(5) = \sum_{i=1}^{4} p_i\, p_{5-i} = 4\cdot\frac{1}{36} = 1/9$$

HW: try the calculation for 3 dice.

Central limit theorem

process

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 <u>normal distribution</u> as the number of independent random variables tends to infinity.

<u>Derivation</u>: define $z_1 \ldots z_n$ as mutually independent random variables with expected value = 0, std dev = σ_i, skewness $\mu_3 = 0$. (not required but helpful)

The sum random variable is: $\eta = \sum_{i=1}^{n} \xi_i$

Expected values of sum: $$M(\eta) = \sum_{i=1}^{n} M(\xi_i) = n M(\xi_i) = 0$$

Variance of sum: $$\sigma^2 = D^2(\eta) = n D^2(\xi_i) = n \sigma_i^2$$

We also have:
$$\mu_4(\eta) = M(\eta^4)$$
$$\mu_4(\xi) = M(\xi^4)$$

Define a normalized sum random variable as:
$$z \equiv \frac{\eta}{\sigma} = \frac{\xi_1 + \dots + \xi_n}{\sigma}$$

Then: $$\frac{D^2(\xi_i)}{\sigma^2} = \frac{\sigma_i^2}{\sigma^2} = \frac{1}{n} ; \quad \frac{\mu_4(\xi_i)}{\sigma^4} = \frac{\mu_4(\xi_i)}{n^2 \sigma_i^4}$$

Now let's write the characteristic function of the normalized sum distribution:
$$\underline{\phi(v)} = M(e^{j \overset{\sum \xi_i/\sigma}{z} v}) = \prod_{i=1}^{n} M(e^{j v \xi_i/\sigma}) = \left(\phi_i\left(\frac{v}{\sigma}\right)\right)^n$$

We can expand the function $\phi_i(v/\sigma)$ using the definition of characteristic function:
$$\phi_i\left(\frac{v}{\sigma}\right) = M(e^{jv/\sigma \cdot \xi_i}) = M\left(1 + j\left(\frac{v}{\sigma}\right)\xi_i - \frac{1}{2!}\left(\frac{v}{\sigma}\right)^2 \xi_i^2 + \dots\right)$$
$$= 1 + j\left(\frac{v}{\sigma}\right) M(\xi_i) - \frac{1}{2!}\left(\frac{v}{\sigma}\right)^2 M(\xi_i^2) + \dots$$

Since $M(\xi_i) = 0$ and $\mu_3 = 0$, we have
$$\phi_i\left(\frac{v}{\sigma}\right) = 1 - \frac{1}{2}\frac{v^2}{n} + \frac{1}{4!}\frac{v^4}{n^2}$$
$$\frac{M(\xi_i^2)}{\sigma^2} = \frac{1}{n}$$

The sum characteristic function is given as a product of ϕ_i. 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 $n \to \infty$ $(y \to 0)$, to get $\ln(1+y) \simeq y$

Therefore,

$\ln \phi = n \ln(\phi_i)$

$$\ln\left[\phi_i\left(\frac{v}{\sigma}\right)\right] \simeq \ln\left(1 - \frac{v^2}{2n}\right) \simeq -\frac{v^2}{2n} \leftarrow$$

$$\hookrightarrow \quad \phi(v) = e^{-v^2/2} = \text{Gaussian}$$

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:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-v^2/2} \cos vx \, dx = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

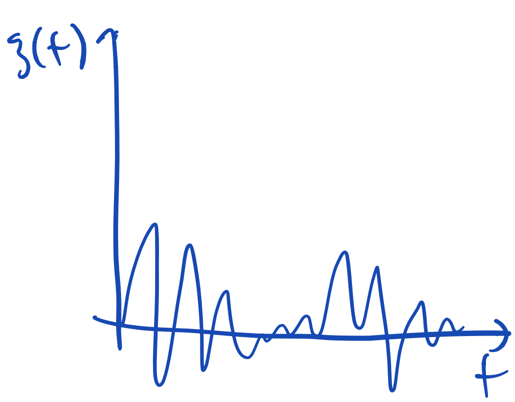
We have thus obtained the normal distribution. In the general form, we get:

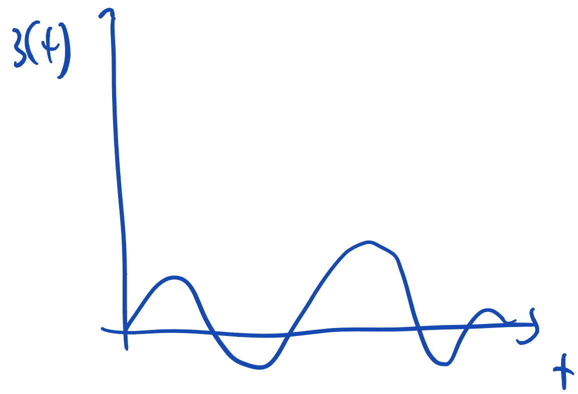
$$\rightarrow f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma^2}$$

Example: say we have a random variable ξ that depends on time. The density function gives the probability of individual instantaneous values and ~~should~~ often 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: $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma^2}$$

Distribution function: $$F(x) = \int_{-\infty}^{x} f(u)\,du = \Phi\left(\frac{x-\mu_1}{\sigma}\right)$$

Probability to fall within limits $x = \pm k\sigma$?

1σ: 0.683

2σ: 0.954

3σ: 0.9972

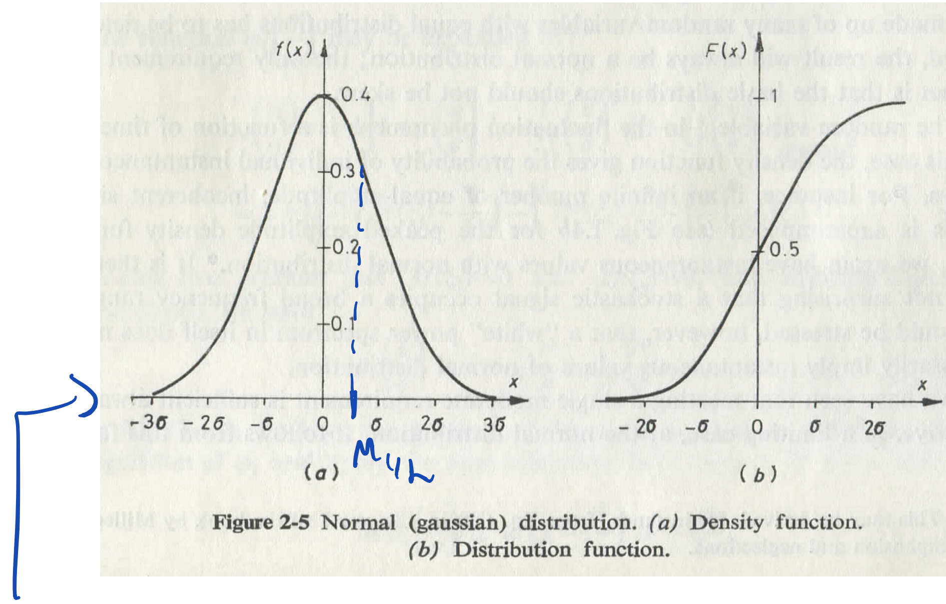


Figure 2-5 Normal (gaussian) distribution. *(a)* Density function. *(b)* Distribution function.

Expected value: $M(\xi) = 0$ since a Gaussian is even.

Expected value of half distribution:

$$M_{1/2}(\xi) = \int_0^\infty x f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty x e^{-x^2/2\sigma^2}\,dx$$

$$= \frac{\sigma}{\sqrt{2\pi}} \doteq 0.4\sigma$$

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

$$M(|\xi|) = 2 M_{1/2} \doteq \underline{0.8\sigma}$$

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 σ :

$$\phi(v) = e^{-\sigma^2 v^2/2}$$

We get the derivatives and hence moments as:

$$\left.\frac{d\phi}{dv}\right|_{v=0} = -\sigma^2 v\, e^{-\sigma^2 v^2/2}\Big|_{v=0} = \underline{0} = j M_1$$

$$\left.\frac{d^2\phi}{dv^2}\right|_{v=0} = -\sigma^2 = -M_2$$

$$\frac{d^3\phi}{dv^3} = 0 = -j\mu_3$$

$$\frac{d^4\phi}{dv^4} = 3\sigma^4 = \mu_4$$

kurtosis

A useful property of the normal distribution: $\mu_4/\mu_2^2 = 3.$

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 $P(A) = p_1$ the probability of the event occurring (say, electrode A is hit by the electron). Probability of not occurring is $p_2 = 1 - p_1$.

We can define the values of random variable ξ as $\xi = x_1 = 1$ in the first case, and $\xi = x_2 = 0$ in the second case.

Expected value: $M(\xi) = \sum_{i=1}^{2} p_i x_i = 1 \cdot p_1 + 0 \cdot p_2 = p_1$

Variance:

$$D^2(\xi) = M(\xi^2) - M(\xi)^2 = \sum_{i=1}^{2} p_i x_i^2 - M(\xi)^2$$

$$= 1 \cdot p_1 + 0 \cdot p_2 - p_1^2 = p_1(1 - p_1) = p_1 p_2$$

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

$$\prod_n p_1 = p_1^n \quad ; \quad \prod_{N-n} p_2 = p_2^{N-n}$$

Since we don't care about the particular order of the experiments, we count all the ways n events can occur from N total events as: $\binom{N}{n}$

Thus, the probability is

$$P(\eta = n) = \binom{N}{n} p_1^n p_2^{N-n} \leftarrow \text{binomial}$$

which is the binomial distribution.

In terms of a sum random variable η, we have

$$\eta = \zeta_1 + \ldots + \zeta_N \qquad (\zeta_i = 0, 1)$$

As the experiments are independent, we have

Mean

$$M(\eta) = N M(\zeta) = N p_1$$

Variance

$$\sigma^2(\eta) = N \sigma^2(\zeta) = N p_1 p_2$$

If $p_1 \ll 1$ and $p_2 \simeq 1$, then $\sigma^2(\eta) \simeq M(\eta)$

But, if $p_1 = \frac{1}{2}$ and $p_2 = \frac{1}{2}$, then $\sigma^2(\eta) = \frac{M(\eta)}{2} = \frac{N}{4}$

Limiting cases: if $N \to \infty$ and $p \to 0$ so that $Np = \lambda$, we get the Poisson distribution:

$$P(\eta = n) = \frac{\lambda^n}{n!} e^{-\lambda} = f(n)$$

λ = rate x unit time

We can get moments of the Poisson distribution from the characteristic function:

$$\phi(v) = \sum_{n=0}^{\infty} e^{jvn} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^{jv})^n}{n!}$$

$$= e^{-\lambda} e^{\lambda e^{jv}} = e^{\lambda(e^{jv}-1)}$$

First derivative: $$\left.\frac{d\phi}{dv}\right|_{v=0} = j\lambda e^{jv} e^{\lambda(e^{jv}-1)}\Big|_{v=0} = j\lambda = j\mu(1)$$

Second derivative: $$\left.\frac{d^2\phi}{dv^2}\right|_{v=0} = -\lambda^2 - \lambda = -\mu_2$$

Since $\sigma^2(1) = \mu_2 - \mu_1^2$, we get:

$$\sigma^2(1) = \mu_2 - \mu_1^2 = \lambda \quad \checkmark$$

Let's consider a limiting case of the Poisson distribution for large $\lambda = \bar{n}$, $n \sim \lambda$. (actual #: n)

We use the Stirling formula: $$n! \simeq n^n e^{-n} \sqrt{2\pi n}$$

Substitute it into the Poisson distribution and take the log:

$$\rightarrow \ln P = -n \ln\left(\frac{n}{\lambda}\right) + n - \lambda - \frac{1}{2}\ln n - \frac{1}{2}\ln 2\pi$$

Use the identity $$n \equiv \lambda\left(1 + \frac{n-\lambda}{\lambda}\right) = \lambda\left(1 + \frac{\Delta n}{\lambda}\right) = \lambda(1+\alpha)$$

and note that if n is close to the mean λ, then $\alpha \ll 1$, so that: $\ln(1+\alpha) \simeq \alpha - \frac{\alpha^2}{2} + \ldots$

We can then write:

$$\ln P = -\lambda(1+\alpha)\left(\alpha - \frac{\alpha^2}{2} + \ldots\right) + \lambda\alpha - \frac{1}{2}\ln\lambda$$

$$-\frac{1}{2}\left(\alpha - \frac{\alpha^2}{2} + \ldots\right) - \frac{1}{2}\ln 2\pi$$

$$= -\frac{1}{2}\lambda\alpha^2 + \frac{1}{2}\lambda\alpha^3 - \frac{1}{2}\alpha + \frac{1}{4}\alpha^2 - \frac{1}{2}\ln(2\pi\lambda)$$

$$\frac{\Delta n}{\lambda}\cdot\Delta n$$

$$\rightarrow = -\frac{1}{2}\frac{\Delta n^2}{\lambda} + \frac{1}{2}\frac{\Delta n^2}{\lambda^2}\left(\Delta n + \frac{1}{2}\right) - \frac{1}{2}\frac{\Delta n}{\lambda} - \frac{1}{2}\ln 2\pi\lambda$$

Since $\Delta n \ll \lambda$, $\Delta n \sim 0,1,2..$, we can neglect the middle two terms to get:

$$\underbrace{P(\eta = n)} = \frac{1}{\sqrt{2\pi\lambda}} e^{-\Delta n^2/2\lambda} = \frac{1}{\underbrace{D(\eta)}_{=\sqrt{\lambda}}\sqrt{2\pi}} e^{-\Delta n^2/2D^2(\eta)}$$

Taking the continuous limit, we see the above is the density function of a Gaussian distribution! Therefore, if $D(\eta) \ll M(\eta)$, i.e. $(\eta \sim \lambda)$

$$\frac{D(\eta)}{M(\eta)} = \frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}} \ll 1 \qquad (\lambda \gg 1)$$

then the Poisson distribution can be accurately approximated by the normal distribution in the vicinity of $M(\eta)$. The condition is that λ 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 p_1 is not small, in the vicinity of $n = Np_1 = M(\eta)$. We have

$$P(\eta = Np_1) = P_0 = \binom{N}{Np_1} p_1^{Np_1} p_2^{N-Np_1}$$

$$= \frac{N!\, p_2^N}{(Np_1)!\,(N-Np_1)!}\left(\frac{p_1}{1-p_1}\right)^{Np_1}$$

We will obtain the relative probability of $\eta = Np_1 + k$ (integer). If $k \ll Np_1$, then we need the extra terms $(Np_1)! \to (Np_1+k)!$

Similarly, $(N-Np_1)!$ decreases by the terms $(N-Np_1)\ldots(N-Np_1-(k-1))$.

$$\frac{P_k}{P_0} = \frac{P(\eta = Np_1+k)}{P(\eta = Np_1)} = \frac{(N-Np_1)\ldots(N-Np_1-(k-1))}{(Np_1+1)\ldots(Np_1+k)}\left(\frac{p_1}{1-p_1}\right)^k$$

Extract a factor of $(N-Np_1)^k$ from each term in the top, and a factor of $(Np_1)^k$ in the bottom. These terms and the fraction term yield unity:

$$\frac{(N-Np_1)^k}{(Np_1)^k}\cdot\frac{p_1^k}{(1-p_1)^k} = \frac{N^k}{N^k} = 1$$

The remaining equation is:

$$\frac{P_k}{P_0} = \frac{(1-1/(N-Np_1))\ldots(1-(k-1)/(N-Np_1))}{\left(1+\frac{1}{Np_1}\right)\ldots\left(1+\frac{k}{Np_1}\right)}$$

Take the log of this equation

$$\ln P_k = \ln P_0 + \sum_{i=0}^{k-1}\ln\left(1-\frac{i}{N-Np_1}\right) - \sum_{i=1}^{k}\ln\left(1+\frac{i}{Np_1}\right)$$

Use the assumption that $k \ll Np_1$ and the approximation of log: $\ln(1+x) \doteq x$

We can perform these algebraic series to get:

$$\ln P_k = \ln P_0 - \frac{k}{2N}\left(\frac{k-1}{1-p_1} + \frac{k+1}{p_1}\right)$$

$$= \ln P_0 - \frac{k(k+1-2p_1)}{2Np_1(1-p_1)}$$

Note that $2Np_1(1-p_1) = 2D^2(\eta)$ and that the numerator $\sim k^2$.

(True exactly if $p_1 = \frac{1}{2}$).

We thus get:

$$p_k \cong p_0\, e^{-k^2/2D^2(\eta)}$$

We can get p_0 with the help of the Stirling formula:

$$N! \cong N^N e^{-N}\sqrt{2\pi N}$$

$$(Np_1)! = N^{Np_1}\, p_1^{Np_1}\, e^{-Np_1}\sqrt{2\pi N p_1}$$

$$(N-Np_1)! = N^{N(1-p_1)}\,(1-p_1)^{N(1-p_1)}\, e^{-N(1-p_1)}\sqrt{2\pi N(1-p_1)}$$

$$p_0 = \frac{N!\,(1-p_1)^N}{(Np_1)!\,(N-Np_1)!}\left(\frac{p_1}{1-p_1}\right)^{Np_1} \cong \frac{1}{\underbrace{D(\eta)}_{\sqrt{Np_1(1-p_1)}}\sqrt{2\pi}}$$

and so finally we conclude that for sufficiently high mean $M(\eta)$, 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 ξ has a uniform distribution if in the interval (a, b) the density function is:

$$f(x) = \frac{1}{b-a} \qquad a < x < b$$

Outside, $f(x) = 0$.

Note: the density function typically has a physical dimension.

Example: voltage distribution would have units V^{-1}.

Consider a uniform distribution with zero expected value, limits $(-a, a)$.

Density function: $$f(x) = \frac{1}{2a} \qquad -a < x < a$$

Note that this function satisfies the requirement: $$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Variance and std dev:

$$D^2(\xi) = \int_{-a}^{a} x^2 \cdot \frac{1}{2a}\,dx = \frac{a^2}{3}$$

Fourth-order moment: $$\mu_4 = \int_{-a}^{a} x^4 \frac{1}{2a}\,dx = \frac{a^4}{5}$$

Kurtosis:

$$\frac{\mu_4}{\mu_2^2} = \frac{9}{5} < \underbrace{3}_{\text{Gaussian}}$$

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 σ.

The probability that $\xi^2 < y$ is:

$$F(y) = P(\xi^2 < y) = P(-\sqrt{y} < \xi < \sqrt{y})$$

Since normal distributions are symmetric about 0, we have:

$$F(y) = \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\sqrt{y}} \underbrace{e^{-u^2/2\sigma^2}}_{\text{Gaussian}} du = 2\left(\Phi\left(\frac{\sqrt{y}}{\sigma}\right) - \Phi(0)\right)$$

The density function is:

$$f(y) = \frac{dF}{dy} = \frac{d}{d\sqrt{y}} \cdot 2\Phi\left(\frac{\sqrt{y}}{\sigma}\right) \cdot \frac{dy^{1/2}}{dy} = \frac{e^{-y/2\sigma^2}}{\sigma\sqrt{2\pi y}}$$

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

If we have a random variable that is a sum of squared random variables $\xi_1 \ldots \xi_k$ each having the same normal distribution: $\chi^2 = \xi_1^2 + \ldots + \xi_k^2$

we get a χ^2 distribution with k degrees of freedom. The density function is:

$$f_k(y) = \frac{y^{(k/2)-1} e^{-y/2\sigma^2}}{2^{k/2}\Gamma(k/2)\sigma^k} = \frac{\chi^{k-2} e^{-\chi^2/2\sigma^2}}{2^{k/2}\Gamma(k/2)\sigma^k} = \underset{\text{original var}}{f_k(\chi^2)}$$

This result can be derived from the characteristic function.

$\Gamma(v)$ is the generalization of the factorial function: $\Gamma(v) = \int_0^\infty x^{v-1} e^{-x}\, dx$

Expected value and std dev of the χ^2 distribution are obtained in the usual way:

$$M(\chi^2) = M(\xi_1^2) + \dots + M(\xi_k^2) = k\, M(\xi^2)$$

$$D^2(\chi^2) = k\, D^2(\xi^2) = k\left(M(\xi^4) - M^2(\xi^2)\right)$$

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

$$D^2(\chi^2) = k\left(3\sigma^4 - \sigma^4\right) = 2k\sigma^4$$

and

$$M(\chi^2) = k\sigma^2$$

The relative std dev of the distribution is

If K is large, the χ^2 distribution tends to the normal distribution.

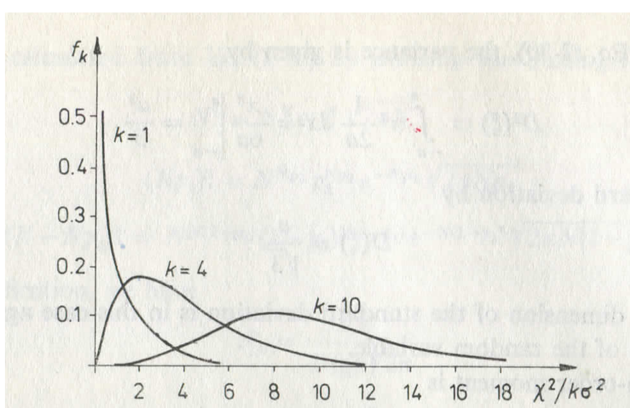


Figure 2-8 χ^2 distributions for squared sums having different number of terms.

For full-wave, piecewise linear detection, the distribution of the measured output signal follows the distribution of $|\xi|$. If ξ has a normal distribution, density function of has a distribution as:

$$f(x) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\, e^{-x^2/2\sigma^2} \qquad (0 < x < \infty)$$

Expected value: $M(|\xi|) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} \int_0^\infty x\, e^{-x^2/2}\, dx = \sqrt{\frac{2}{\pi}}\,\sigma$

Variance: $D^2(|\xi|) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} \int_0^\infty x^2 e^{-x^2/2\sigma^2}\, dx - \frac{2}{\pi}\sigma^2 = \left(1 - \frac{2}{\pi}\right)\sigma^2$

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

$$\eta = |\xi_1| + \dots + |\xi_k|$$

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: $M(\eta) = k\, M(|\xi|) = k\sqrt{\frac{2}{\pi}}$

Variance: $D^2(\eta) = k\, D^2(|\xi|)$
$= k\left(1 - \frac{2}{\pi}\right)\sigma^2$

Relative std dev:

$$\frac{D(\eta)}{M(\eta)} = \sqrt{\frac{\pi}{2} - 1}\,\frac{1}{\sqrt{k}}$$

$$\simeq \frac{0.75}{\sqrt{k}}$$

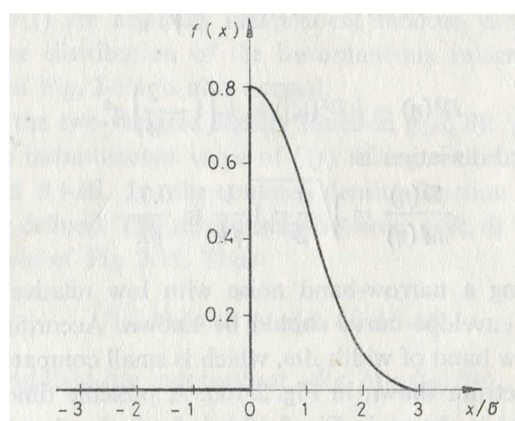


Figure 2-9 Distribution of absolute value

A final problem: consider narrow-band noise with a low relative bandwidth $\Delta\omega$. We want the distribution of the envelope curve. We can consider the signal as a carrier of frequency ω_0, which is stochastically modulated in amplitude and phase. The random variable is:

$$\zeta(t) = \rho(t)\cos(\omega_0 t + \phi(t))$$

The instantaneous values of envelope and phase angle are $\rho(t)$ and $\phi(t)$.

We can write $\zeta(t)$ as:

$$\zeta(t) = \alpha(t)\cos(\omega_0 t) + \beta(t)\sin\omega_0 t$$

where $\alpha(t)$ and $\beta(t)$ are mutually independent random variables with normal distributions (if the original distribution is also normal).

Let $\alpha(t)$ and $\beta(t)$ have the common density function, $f(a,b)$. Since they are independent and normal, we have: $f(a,b) = f(a)f(b)$

$$f(a)f(b) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^2 e^{-(a^2+b^2)/2\sigma^2}$$

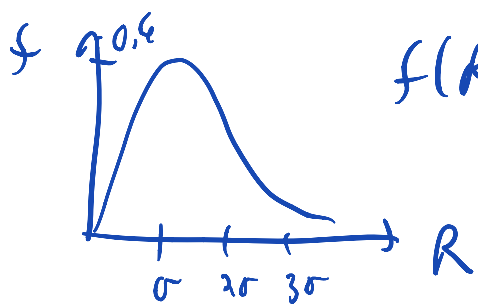
Transforming to polar coordinates so the density function depends on R, θ,

$$g(R,\theta) = \frac{R}{2\pi\sigma^2} e^{-R^2/2\sigma^2}$$

The function doesn't depend on θ, so

$$f(R) = 2\pi g(R,\theta) = \boxed{\frac{R}{\sigma^2} e^{-R^2/2\sigma^2} = f(R)}$$

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



$f(R)$

which is a Rayleigh distribution.

The cumulative density function is (distribution function)

$$F(R) = P(\rho < R) = 1 - e^{-R^2/2\sigma^2}$$

so that the probability of an envelope amplitude higher than R is $e^{-R^2/2\sigma^2}$.

Expected value: $$M(\rho) = \int_0^\infty R \cdot \frac{R}{\sigma^2} e^{-R^2/2\sigma^2}\, dR = \sqrt{\frac{\pi}{2}}\,\sigma \simeq 1.25\sigma$$

Variance: $$D^2(\rho) = \mu_2 - \mu_1^2 = 2\sigma^2 - \frac{\pi}{2}\sigma^2 \simeq 0.43\,\sigma^2$$

Relative std dev of a k-term sum:

$$\frac{D\left(\sum_k \rho\right)}{M\left(\sum_k \rho\right)} = \frac{\sqrt{0.43}}{1.25}\frac{1}{\sqrt{k}} = \frac{0.525}{\sqrt{k}}$$

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

$$f(x) = \begin{cases} 0 & x \le 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases}$$

The distribution parameter λ is an arbitrary positive number. (same λ as Poisson).

Distribution function:

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

Moments:

$$M(\xi) = \lambda^{-1}$$

$$M(\xi^2) = 2\lambda^{-2}$$

$$D^2(\xi) = \lambda^{-2}$$

Interesting property: assume ξ 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.

$$\frac{e^{-(s+t)}}{e^{-s}} = e^{-t}$$