Chapter 3

Probability, Random Variables, Random Processes, Signal Bandwidth

3.1 Introduction

Noise is an important aspect of what limits the performance of communication systems. As such it is important to understand the statistical properties of noise. Noise at the input of a receiver will affect the performance of a communication system. Because receivers filter the received signal (desired signal plus noise) it is important to be able to characterize the noise out of a linear system, i.e. a filter. To this end in this chapter we discuss probability, random variables, and random processes. Of particular importance is determining the statistical properties (e.g. distribution, the mean, the variance) of the output of a filter when the input is noise. In addition, since the data in a communication system is random it is important to be able to characterize the spectral properties of the transmitted signal, which is a random process. The spectral properties determine the bandwidth that the random signal occupies. This is an abbreviated treatment of probability, random variables and random processes. There are many texts that treat this topic in depth. See for example [2], [3] or [4]. As such, most proofs are left out here.

Communications systems have some amount of noise. Because we can not know the value of the noise ahead of time we can only characterize it by its statistical properties such as average values. Using probability we can determine likelihoods of certain events. Electrical systems have random thermal noise due to motion of electrons because the system is not at absolute zero temperature ($-273°$ C). This noise is usually modeled as a random process, that is an infinite collection of random variables, one for each value of time. The noise is also typically modeled as having the same power at all frequencies. This is called white noise. However, in reality at extremely high frequencies (i.e. in the optical range of frequencies) the noise power decreases. Since only the frequency band of the transmitted signal is of interest, the noise outside this band is not important. The distribution of the random samples also generally has a Gaussian (bell-shaped) distribution with zero mean. As such the model widely used for thermal noise is that of zero mean white Gaussian noise.

This chapter begins with a review of probability and random variables. Then random processes are described and characterized with a focus on white Gaussian noise. The statistical relation between the input and the output of a linear system when the input is a random process is described. The output of a linear system when the input is white Gaussian noise is characterized by the variance of the output. Finally the bandwidth of a digitally modulated signal is derived.

3.2 Probability

The foundations of probability starts with an experiment that has multiple possible outcomes. For example flipping a coin can have two outcomes, heads or tail. Flipping a coin 10 times can have $2^{10} = 1024$ possible outcomes that are essentially impossible to know before hand. The outcomes consist of all possible sequences of length 10 of heads or tails. Events are a collection of possible outcomes of an experiment. Consider two events A and B. We denote the event that the outcome of the experiment is not in B by $\bar{B}$ and the event that the outcome is in A and B as $A \cap B$, the intersection of A and B. The event that the outcome is either in A or B is denoted as $A \cup B$, the union of A and B. The set of all possible outcomes, denoted by S is called the sample space. Probability is a mapping from a set of outcomes (events) of an experiment to numbers in the interval $[0,1]$. The axioms of probability include the following.

- The probability of any event is non-negative: $P(A) \geq 0$ for an event A.
- The probability of the event of all outcomes is 1: $P(S) = 1$.
- The probability of a union of disjoint events is the sum of the probabilities: for $A \cap B = \phi, P(A \cup B) = P(A) + P(B)$ where ϕ is the empty set.

Using these axioms we can derive many other properties of probability. We will summarize a few of these below.

1. The probability of the complement of an event $P(\bar{A}) = 1 - P(A)$.
2. The probability of an arbitrary union (A and B are not necessarily disjoint) is $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. We can decompose the probability of an event A into two probabilities.

$$P(A) = P(A \cap B) + P(A \cap \bar{B}).$$

More generally

$$P(A) = \sum_i P(A \cap B_j), \text{ if } B_i \cap B_j = \phi \text{ for } i \neq j \text{ and } \cup_i B_i = S.$$

This is called the *law of total probability*.

Two events, A and B, are independent if $P(A \cap B) = P(A)P(B)$. For example in flipping a coin twice we would naturally assume that the outcomes of the first flip and the second flip are independent. So if in every flip the probability of heads is p then the probability of getting two heads in a row is p^2. The probability that the first flip is tails and the second flip is heads is $(1-p)p$. The probability of the event one head and one tail in an experiment consisting of two flips of a coin is $2p(1-p)$ because the possible outcomes for the event one head and one tail are 1) first head then tail and 2) first tail then head and these two events are disjoint. If we flip a coin N times there are $\binom{N}{k} = N!/(k!(N-k)!)$ possible outcomes with k heads and $N-k$ tails and each of these distinct outcomes has probability $p^k(1-p)^{N-k}$ so that the probability of k heads in N flips is $\binom{N}{k} p^k (1-p)^{N-k}$.

If in an experiment you have some knowledge about a certain event occurring then the probability of another event occurring is called the conditional probability. For example, the probability that in flipping a coin 10 times the number of heads is greater than 5 given that at least 3 heads occurred is a conditional probability. The conditional probability of event A given event B for an event B with nonzero probability is denoted by $P(A|B)$ and defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

For the example of flipping a coin 10 times and A being the event that the number of heads is greater than 5 and B being the event that the number of heads is at least 3 the event $A \cap B$ is the same as event A because A is a subset of B. The probability of A is

$$P(A) = \sum_{k=6}^{10} \binom{10}{k} (p)^k (1-p)^{10-k}$$

where p is the probability of heads on each toss. Similarly

$$P(B) = \sum_{k=3}^{10} \binom{10}{k} (p)^k (1-p)^{10-k}.$$

So

$$P(A|B) = \frac{\sum_{k=5}^{10} \binom{10}{k} (p)^k (1-p)^{10-k}}{\sum_{k=3}^{10} \binom{10}{k} (p)^k (1-p)^{10-k}}.$$

Using the definition of conditional probability we can express the law of total probability as

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}).$$

or more generally as

$$P(A) = \sum_i P(A|B_j)P(B_j), \text{ if } B_i \cap B_j = \phi \text{ for } i \neq j \text{ and } \cup_i B_i = S.$$

If two events A and B are independent then knowledge of whether or not the outcome is in B does not change the probability of A. That is, for independent events $P(A|B) = P(A)$. In the coin tossing experiment we assumed the event H_1, that the first toss is a heads, is independent of the event H_2, that the second toss is a heads. Thus, the event that the first two tosses are both heads $H_1 \cap H_2$ is p^2.

$$P(H_1 \cap H_2) = P(H_1)P(H_2) = p^2.$$

Two events A and B are *conditionally independent* given an event C if $P(A \cap B|C) = P(A|C)P(B|C)$. As an example suppose that a hat contains two coins. One coin has probability of heads p_1 while the second coin has probability of heads p_2. A coin from the hat is drawn with each coin being selected with equal probability (1/2). The coin selected is tossed twice with independent outcomes on each toss. The event that the first toss is heads and the event that the second toss is heads are not independent. To see this consider that $p_1 = 0.99$ while $p_2 = 0.01$. If the first toss is heads it is very likely that coin 1 was selected and then the second toss will likely be heads. However, if we know that coin 1 was selected then the probability of heads is independent from toss to toss. We can use the law of total probability to calculate the unconditional probability of two heads in two tosses as follows. Let C_1 be the event the first coin is selected and let $C_2 = \bar{C}_1$ be the event the second coin is selected. Then

$$\begin{aligned} P(H_1 \cap H_2) &= P(H_1 \cap H_2|C_1)P(C_1) + P(H_1 \cap H_2|C_2)P(C_2) \\ &= P(H_1|C_1)P(H_2|C_1)P(C_1) + P(H_1|C_2)P(H_2|C_2)P(C_2) \\ &= p_1^2(1/2) + p_2^2(1/2). \end{aligned}$$

Here we have used the fact that given coin 1 (or 2) was selected the event of heads on the first toss is conditionally independent of heads on the second toss. Alternatively, we can consider a modified experiment where a coin is selected from the hat and tossed once and then put back into the hat. A coin is again selected independently of the first selection from the two coins in the hat and tossed. The probability of two heads in this case is

$$P(H_1 \cap H_2) = P(H_1)P(H_2) = (\frac{p_1 + p_2}{2})^2$$

which is different from the probability of two heads in the case one coin is selected and that one coin is tossed twice.

3.3 Random Variables

Random variables are mappings from outcomes of an experiment to real numbers. As a simple example, in the experiment of tossing a coin 10 times, the number of heads that occurs X, is a random variable. The outcome $HHTHTHTTHH$ is mapped to the real number 6. There are 210 different sequences of 10 heads and tails (outcomes) that result in 6 heads that are mapped to $X = 6$. Each of these possible outcomes has probability 2^{-10} for the case of a fair coin $p = 1/2$. So the probability of getting 6 heads is 210×2^{-10}. This is an example of a discrete random variable. A random variable is characterized as a *discrete random variable* if there are a countable number of possible values for the random variable. For example, consider flipping a coin until a head occurs and let X be the number of flips. The possible values of X are 1,2,3,... which is a countable number of possible values (it can be mapped to the integers). The number of heads that appear in tossing a coin 10 times can be 0,1,...,10 which is a finite number of possible values (also countable). For a discrete random variable a probability mass function (pmf) is defined as $p_X(x) = P(X = x)$ that is the probability of the event $\{X = x\}$. Consider counting the number of heads in only three flips. The probability of no heads in three flips is $1/8$. The probability of 1 head in three flips is $3/8$ since there are three ways of flipping a coin three times and getting one head and each of the three possible outcomes (HTT, HTH, TTH) has probability 1/8. Similarly there are three outcomes with two heads and one tail. Finally there is only one way of getting three heads. So the probability mass function is

$$p_X(x) = \begin{cases} .125, & x = 0 \\ .375, & x = 1 \\ .375, & x = 2 \\ .125, & x = 3 \\ 0, & \text{otherwise.} \end{cases}$$

For any random variable X the cumulative probability distribution function (cdf) $F_X(x)$ is defined as

$$F_X(x) = P(X \leq x).$$

For the example of flipping a coin three times the distribution function of the random variable X that counts the number of heads would be

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.125, & 0 \leq x < 1 \\ 0.5, & 1 \leq x < 2 \\ 0.875, & 1 \leq x < 2 \\ 1.000, & x \geq 3. \end{cases}$$

A random variable is a *continuous random variable* if there is a function, known as the probability density function (pdf), such that

$$P(X \in C) = \int_{u \in C} f_X(u) du.$$

Note that for a continuous random variable

$$F_X(x) = \int_{u \leq x} f_X(u) du = \int_{u=-\infty}^{x} f_X(u) du.$$

For a continuous random variable the density function is the derivative of the distribution function.

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

Note that the density function must integrate to 1. As an example consider a random variable representing the phase of a sinusoidal signal. A possible density function for this random variable is

$$f_X(x) = \begin{cases} \frac{1}{2\pi}, & 0 \leq x \leq 2\pi \\ 0, & \text{otherwise.} \end{cases}$$

This density function corresponds to a random variable uniformly distributed over the interval $[0, 2\pi]$ and would be a reasonable model when no knowledge of the phase of a sinusoid is available. The corresponding distribution function is

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2\pi}, & 0 \leq x \leq 2\pi \\ 1, & x > 2\pi. \end{cases}$$

The expected value or average value or mean value of a discrete random variable X, denoted by $\mathbb{E}[X]$ to distinguish it from energy, is defined as

$$\mathbb{E}[X] = \sum_x x p_X(x)$$

while for a continuous random variable the expected value is

$$\mathbb{E}[X] = \int_x x f_X(x) dx.$$

Sometimes it is important to know the statistical properties (e.g. mean) of a function of a random variable. If X is a random variable and $g(x)$ is a real function, then $Y = g(X)$ is also a random variable since Y is a mapping from the set of possible outcomes to real numbers. The probability distribution of Y can often be calculated from the probability distribution of X. The expected value of Y can be calculated from the probability mass or density of X. For example, if X is a continuous random variable and $Y = g(X)$ then the expected value of Y is

$$\mathbb{E}[Y] = \int g(x) f_X(x) dx.$$

This result is sometimes called the law of the unconscious statistician. As a simple result of this if a is a constant (not random) then $\mathbb{E}[aX] = a\mathbb{E}[X]$. Also $\mathbb{E}[X + a] = \mathbb{E}[X] + a$. Often the expected value or mean of a random variable X is denoted μ_X. The variance of a random variable is a measure of the spread of the random variable. A random variable that can vary widely will have a large variance. The variance of a random variable (discrete or continuous), denoted by σ_X^2, is

$$\sigma_X^2 = \mathrm{Var}[X] = \mathbb{E}[(X-\mu_X)^2] = \mathbb{E}[X^2] - \mu_X^2.$$

If a is a constant then the variance of aX is $\mathrm{Var}[aX] = a^2\mathrm{Var}[X] = a^2\sigma_X^2$. However, $\mathrm{Var}[X+a] = \mathrm{Var}[X]$. The standard deviation, denoted by σ_X, of a random variable is the square root of the variance.

Example 1: Flipping a coin

Consider flipping a coin 10 times. Of course, until the experiment is done we do not know the outcome of the experiment. However, we can say something about the probability of certain events, like there being more than 8 heads. Suppose that each flip has probability of heads 1/2. In this example every outcome (10 flips) will have the same probability, namely $1/(2^{10})$ since there are 2^{10} possible outcomes. The probability of getting $HTHTTTHTHHT$ will be $P(HTHTTTHTHHT) = 2^{-10}$. In this example the number of heads that occurred is a discrete random variable that is a mapping from experimental outcomes to an integer between 0 and 10. Random variables are a way to describe many different events. The event more than 8 heads would be written as $\{X > 8\}$. We can calculate probabilities involving random variables. For example, $P(X=8) = \binom{10}{8}(1/2)^8(1/2)^2$. There are $\binom{10}{8} = 45$ ways of flipping a coin 10 times and getting exactly 8 heads. Each of those outcomes has probability 2^{-10}. The probability mass function of the random variable X is just the probability that $X = k$.

$$p_X(k) = P\{X=k\} = \binom{10}{k}(0.5)^k(0.5)^{n-k},\ k = 0,1,...,10.$$

The average value of the number of heads or the expected number of heads is

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{10} k p_X(k) \\
&= \sum_{k=0}^{10} k\binom{10}{k}(1/2)^k(1/2)^{10-k} \\
&= 5.
\end{aligned}$$

The expected value of the square of the number of heads is

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{10} k^2 P\{X=k\} \\
&= \sum_{k=0}^{10} k^2\binom{10}{k}(1/2)^k(1/2)^{10-k} \\
&= 27.5.
\end{aligned}$$

The variance of the number of heads is

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = 27.5 - 25 = 2.5.$$

Flipping a coin and counting the number of heads is an example of a discrete random variable. There are a finite (or countably infinite) number of possible values of the random variable for a discrete random variable.

Example 2: Gaussian Random Variables

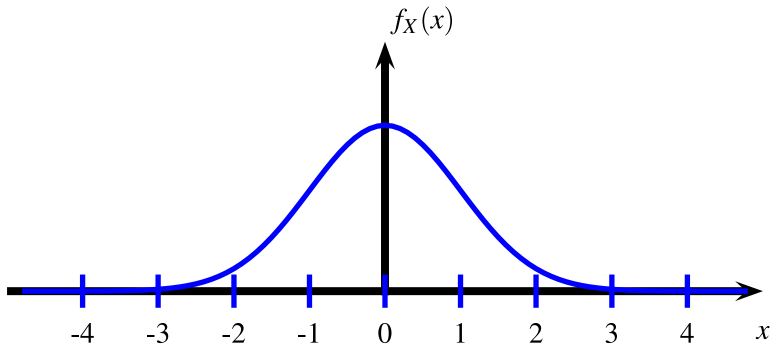

Figure 3.1: Gaussian density function (mean 0, variance 1).

A Gaussian random variable X is a continuous random variable and has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

where μ is the mean and σ^2 is the variance. The density is shown in Figure 3.1 for $\mu = 0$ and $\sigma^2 = 1$. For a zero mean, variance 1 Gaussian random variable X the distribution function $F_X(x)$ is

$$F_X(x) = P\{X \leq x\} = \Phi(x) = \int_{u=-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

The special function, $\Phi(x)$, is called the standard Gaussian distribution function or standard normal distribution function and is the distribution function of a zero mean, variance 1 Gaussian random variable. The complement of this probability distribution function is another special function called the Q function.

$$P\{X > x\} = Q(x) = \int_{u=x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

The cumulative distribution function and complementary distribution function are the areas under the curve of the density function from $-\infty$ to x and from x to ∞ as seen in Figure 3.2. The distribution of a

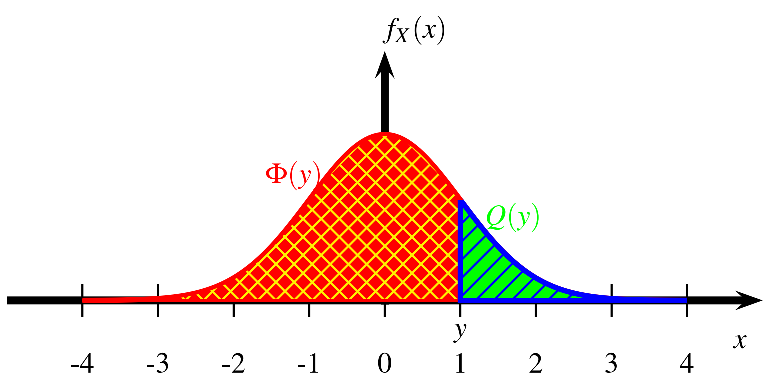

Figure 3.2: Cumulative distribution function and complementary cumulative distribution function

standard (mean 0, variance 1) Gaussian random, $\Phi(x)$ is shown in Figure 3.3. The complement of the distribution, $Q(x)$ is shown in Figure 3.4. The Q function is tabulated in Table 3.1 and the inverse of the Q function is tabulated in Table 3.2. Note that $Q(x) + \Phi(x) = 1$ and $\Phi(-x) = Q(x) = 1 - \Phi(x)$.

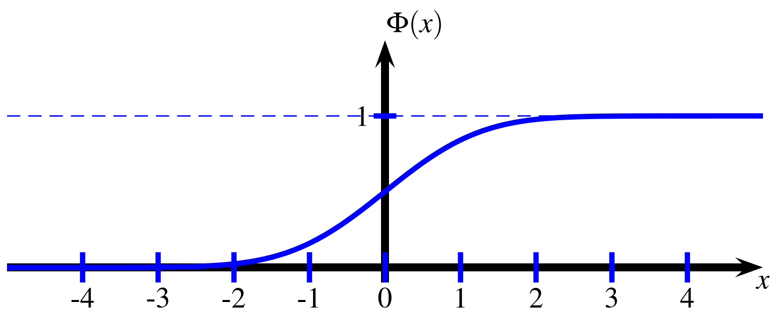

Figure 3.3: Standard Gaussian distribution function, $\Phi(x)$

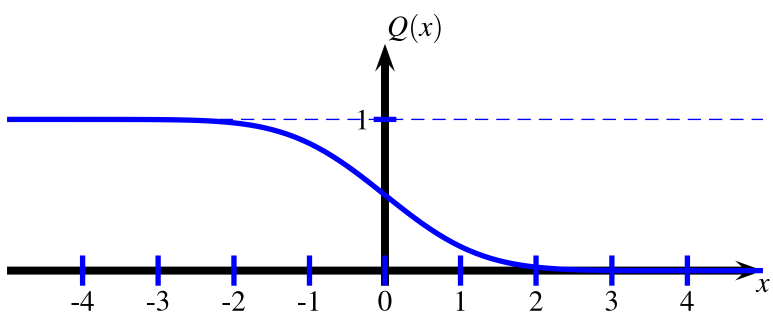

Figure 3.4: Standard Gaussian complementary distribution function, $Q(x)$

x	$Q(x)$	x	$Q(x)$	x	$Q(x)$	x	$Q(x)$
0.00	0.500	1.30	0.097	2.60	0.00467	3.90	0.0000481
0.10	0.460	1.40	0.0817	2.70	0.00347	4.00	0.0000317
0.20	0.421	1.50	0.0668	2.80	0.00256	4.10	0.0000207
0.30	0.382	1.60	0.0548	2.90	0.00187	4.20	0.0000133
0.40	0.345	1.70	0.0446	3.00	0.00135	4.30	0.00000854
0.50	0.309	1.80	0.0360	3.10	0.000968	4.40	0.00000541
0.60	0.274	1.90	0.0287	3.20	0.000687	4.50	0.00000340
0.70	0.242	2.00	0.0228	3.30	0.000483	4.60	0.00000211
0.80	0.212	2.10	0.0179	3.40	0.000337	4.70	0.00000130
0.90	0.184	2.20	0.0139	3.50	0.000233	4.80	0.000000793
1.00	0.159	2.30	0.0107	3.60	0.000159	4.90	0.000000479
1.10	0.136	2.40	0.00820	3.70	0.000108	5.00	0.000000287
1.20	0.115	2.50	0.00621	3.80	0.0000723	5.10	0.000000170

Table 3.1: Q function

$Q(x)$	x
10^{-1}	1.2816
10^{-2}	2.3264
10^{-3}	3.0903
10^{-4}	3.7190
10^{-5}	4.2649
10^{-6}	4.7534
10^{-7}	5.1993
10^{-8}	5.6120
10^{-9}	5.9978
10^{-10}	6.3614

Table 3.2: Inverse Q function

Suppose that Y is a Gaussian random variable but it has mean μ and variance σ^2. Then the distribution of Y is related to the standard Gaussian distribution function.

$$F_Y(y) = P\{Y \le y\} = P\{\frac{(Y-\mu)}{\sigma} \le \frac{(y-\mu)}{\sigma}\}.$$

Since $(Y-\mu)/\sigma$ is a zero mean Gaussian random variable with variance 1, the distribution function of Y can be written as

$$F_Y(y) = P\{\frac{(Y-\mu)}{\sigma} \le \frac{(y-\mu)}{\sigma}\} = \Phi(\frac{y-\mu}{\sigma}).$$

Similarly $P\{Y > y\} = Q(\frac{y-\mu}{\sigma})$.

3.3.1 Pairs of Random Variables

Often we have more than one random variable. For a pair of random variables (X,Y), we define the joint distribution function as the probability that $X \leq x$ and $Y \leq y$. We write this as

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y).$$

For continuous random variables the joint density function is defined as

$$f_{X,Y}(x,y) = \frac{d^2 F_{X,y}(x,y)}{dxdy}.$$

The probability that the pair of random variables is in a certain subset of the two dimension plane is

$$P\{(X,Y) \in C\} = \int\int_{(x,y)\in C} f_{X,Y}(x,y)dxdy.$$

The expected value of a function of two random variables is similar to the expected value of a function of one random variable.

$$\mathbb{E}[g(X,Y)] = \int_x \int_y g(x,y) f_{X,Y}(x,y)dydx.$$

From the joint density of two random variables the density of one of the random variables can be found by integrating out the other. That is,

$$f_X(x) = \int_y f_{X,Y}(x,y)dy, \quad f_Y(y) = \int_x f_{X,Y}(x,y)dx.$$

Two random variables are independent if either of these conditions is true.

$$(i) F_{X,Y}(x,y) = F_X(x)F_Y(y)$$
$$(ii) f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

Consider two random variables, X and Y with joint density $f_{X,Y}(x,y)$. Let $Z = X+Y$. Since both X and Y are random variables, so is Z. The distribution of Z can be found as follows.

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(X+Y \leq z) \\ &= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z-y} f_{X,Y}(x,y)dxdy. \end{aligned}$$

The density of Z is the derivative of the distribution function.

$$\begin{aligned} f_Z(z) &= \frac{\partial F_Z(z)}{\partial z} \\ &= \int_{y=-\infty}^{\infty} f_{X,Y}(z-y,y)dy. \end{aligned}$$

If X and Y are independent then the density of the sum is the convolution of the densities.

$$f_Z(z) = \int_{y=-\infty}^{\infty} f_X(z-y) f_Y(y)dy.$$

The expected value of a function of a pair of random variables where the function is a sum is easy to compute. Let $g(X,Y) = g_1(X) + g_2(Y)$. Then

$$\begin{aligned}
\mathbb{E}[g(X,Y)] &= \int_{x,y} g(x,y) f_{X,Y}(x,y)dydx \\
&= \int_{x,y} (g_1(x) + g_2(y)) f_{X,Y}(x,y)dydx \\
&= \int_{x,y} g_1(x) f_{X,Y}(x,y) + \int_{x,y} g_2(y)) f_{X,Y}(x,y)dydx \\
&= \int_x g_1(x) \in_y f_{X,Y}(x,y)dydx + \int_y g_2(y)) \int_x f_{X,Y}(x,y)dxdy \\
&= \int_x g_1(x) f_X(x) + \int_y g_2(y) f_Y(y)dy \\
&= \mathbb{E}[g_1(X)] + E[g_2(Y)].
\end{aligned}$$

In particular, if $g_1(x) = x$ and $g_2(y) = y$ this shows that the expected value of a sum of random variables is the sum of the expected values. This is true even if X and Y are related (not independent). The correlation of two random variables is defined as

$$r_{X,Y} = \mathbb{E}[XY] = \int_x \int_y xy f_{X,Y}(x,y)dydx.$$

The covariance of two random variables is defined as

$$c_{X,Y} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X\mu_Y = r_{X,Y} - \mu_X\mu_Y.$$

The correlation coefficient of two random variables is

$$\rho_{X,Y} = \frac{c_{X,Y}}{\sigma_X\sigma_Y} = \frac{r_{X,Y} - \mu_X\mu_Y}{\sigma_X\sigma_Y}.$$

The correlation coefficient is bounded between -1 and 1; $-1 \le \rho_{X,Y} \le 1$. If $|\rho_{X,Y}| = 1$ then the random variables are such that Y can be predicted from X in a linear fashion resulting in a mean squared error of 0. If two random variables are independent then $E[XY] = E[X]E[Y]$ so that $c_{X,Y} = \rho_{X,Y} = 0$.

For any pair of random variables (X,Y) with means μ_X and μ_Y the mean of the weighted sum $Z = aX + bY$ is always the weighted sum of the means.

$$\mu_Z = \mathbb{E}[Z] = \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] = a\mu_X + b\mu_Y. \tag{3.1}$$

The variance of Z is

$$\begin{aligned}
\sigma_Z^2 &= \mathbb{E}[(Z - \mu_Z)^2] \\
&= \mathbb{E}[(aX + bY - a\mu_X - b\mu_Y)^2] \\
&= \mathbb{E}[(a(X - \mu_X) + b(Y - \mu_Y))^2] \\
&= a^2\mathbb{E}[(X - \mu_X)^2] + 2ab\mathbb{E}[(X - \mu_X)(Y - \mu_Y)] + b^2\mathbb{E}[(Y - \mu_Y))^2] \\
&= a^2\sigma_X^2 + 2abc_{X,Y} + b^2\sigma_Y^2.
\end{aligned} \tag{3.2}$$

Example 3: Two random variables are jointly Gaussian if

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\{-\frac{(\tilde{x}^2 - 2\rho\tilde{x}\tilde{y} + \tilde{y}^2)}{2(1-\rho^2)}\} \tag{3.3}$$

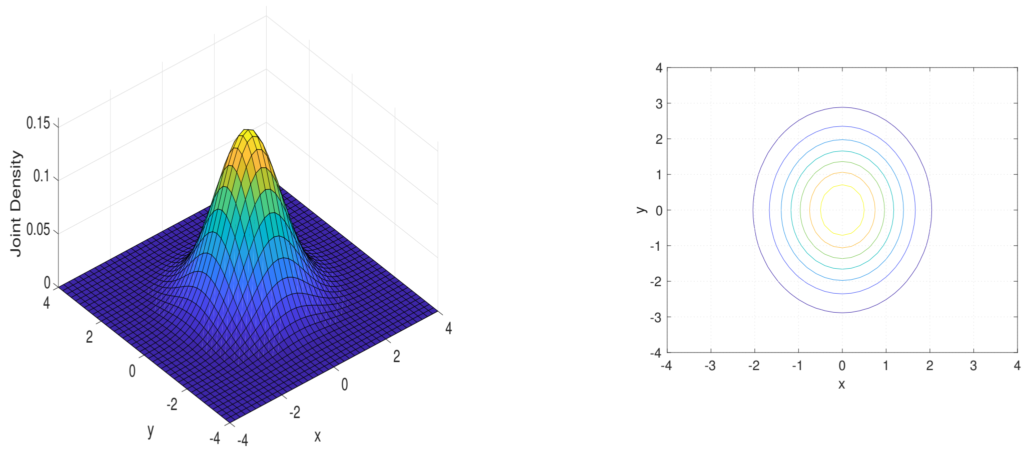

Figure 3.5: Jointly Gaussian: $\mu_X = 0, \mu_Y = 0, \sigma_X = 1, \sigma_Y = 1, \rho = 0$

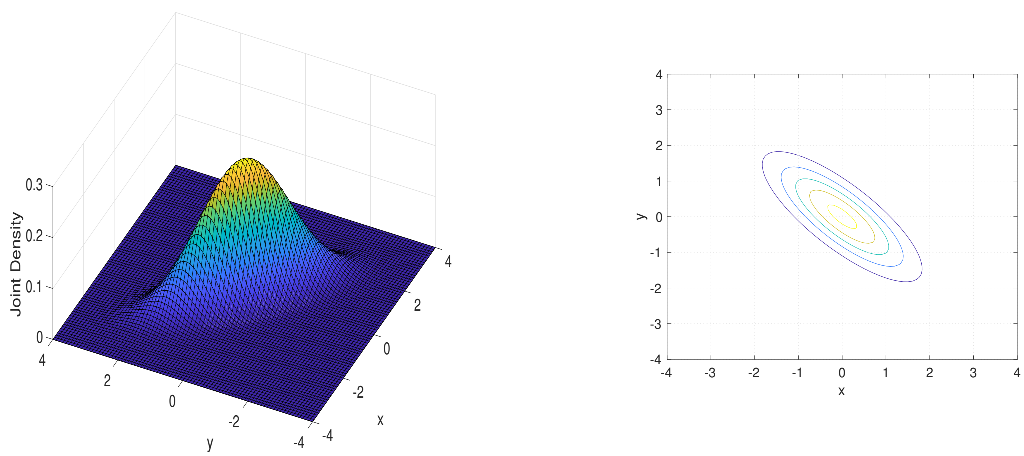

Figure 3.6: Jointly Gaussian: $\mu_X = 0, \mu_Y = 0, \sigma_X = 1, \sigma_Y = 1, \rho = -.8$

where $\tilde{x} = (x - \mu_X)/\sigma_X$ and $\tilde{y} = (y - \mu_Y)/\sigma_Y$. For two jointly Gaussian random variables with density given by (3.3) the means of X and Y are μ_X and μ_Y, the variances are σ_X^2, and σ_Y^2 and the correlation coefficient is ρ. Note that for two jointly Gaussian random variables the density is determined solely by the means, the variances and the correlation coefficient. Note that for two jointly Gaussian random variables if the correlation coefficient is 0 then the joint density factors into the density of X alone and the density of Y alone. Thus uncorrelated ($\rho = 0$) and jointly Gaussian (real) random variables are independent. This is not true necessarily for other jointly distributed random variables. The joint density of two independent, identically distributed Gaussian random variables is circularly symmetric. Figure 3.5 shows the two-dimension density function for the case $\rho = 0$, $\sigma_X = \sigma_Y = 1$ and the corresponding contour plot of constant density. Because the variances are equal and the correlation coefficient is 0 the contours are just circles. Figure 3.6 shows the two-dimension density function for the case $\rho = -0.8$, $\sigma_X = \sigma_Y = 1$ and the corresponding contour plot of constant density. If X and Y are jointly Gaussian and $Z = aX + bY$ then Z is also Gaussian random variable with mean and variance given by (3.1) and (3.2).

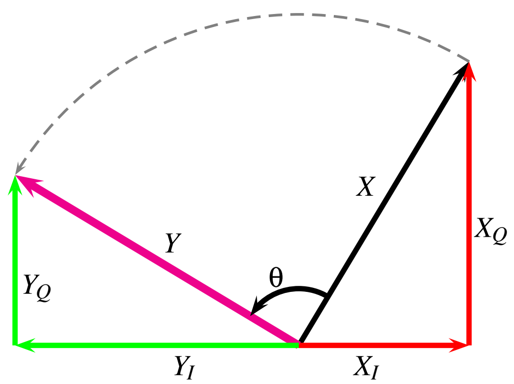

Figure 3.7: Coordinate Transformation

3.3.2 Coordinate Transformation

Consider a pair of correlated random variables X_I and X_Q, zero mean and jointly Gaussian each with variance σ^2. The correlation coefficient is ρ so that $\mathbb{E}[X_I X_Q] = \rho\sigma^2$. So $\mathbb{E}[X_I^2] = E[X_Q^2] = \sigma^2$. Now consider a transformation of those random variables that is just a rotation, $X \to Y$, as shown in Figure 3.7. That is if $X = (X_I + jX_Q)$ and $Y = Xe^{j\theta}$ then

$$\begin{aligned}
Y = (Y_I + jY_Q) &= (X_I + jX_Q)(\cos(\theta) + j\sin(\theta)) \\
&= X_I\cos(\theta) - X_Q\sin(\theta) + j(X_Q\cos(\theta) + X_I\sin(\theta)) \\
Y_I &= X_I\cos(\theta) - X_Q\sin(\theta) \\
Y_Q &= X_Q\cos(\theta) + X_I\sin(\theta).
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[Y_I] &= \mathbb{E}[X_I\cos(\theta) - X_Q\sin(\theta)] \\
&= \mathbb{E}[X_I]\cos(\theta) - \mathbb{E}[X_Q]\sin(\theta) \\
&= 0 \\
\mathbb{E}[Y_Q] &= \mathbb{E}[X_Q\cos(\theta) + X_I\sin(\theta)] \\
&= \mathbb{E}[X_Q]\cos(\theta) + \mathbb{E}[X_I]\sin(\theta) \\
&= 0 \\
\mathrm{Var}[Y_I] &= \mathrm{Var}[X_I\cos(\theta) - X_Q\sin(\theta)] \\
&= \mathbb{E}[(X_I\cos(\theta) - X_Q\sin(\theta))^2] \\
&= \mathbb{E}[X_I^2\cos^2(\theta)] - 2E[X_IX_Q]\cos(\theta)\sin(\theta) + \mathbb{E}[X_Q^2\sin^2(\theta)] \\
&= E[X_I^2]\cos^2(\theta) + E[X_Q^2]\sin^2(\theta) - \rho\sigma^2\sin(2\theta) \\
&= \sigma^2\cos^2(\theta) + \sigma^2\sin^2(\theta) - \rho\sigma^2\sin(2\theta) \\
&= \sigma^2(1-\rho\sin(2\theta)) \\
\mathrm{Var}[Y_Q] &= \mathrm{Var}[X_Q\cos(\theta) + X_I\sin(\theta)] \\
&= \mathbb{E}[(X_Q\cos(\theta) + X_I\sin(\theta))^2] \\
&= \mathbb{E}[X_Q^2\cos^2(\theta)] + 2\mathbb{E}[X_IX_Q]\cos(\theta)\sin(\theta) + \mathbb{E}[X_I^2\sin^2(\theta)] \\
&= \mathbb{E}[X_Q^2]\cos^2(\theta) + \mathbb{E}[X_I^2]\sin^2(\theta) + \rho\sigma^2\sin(2\theta) \\
&= \sigma^2\cos^2(\theta) + \sigma^2\sin^2(\theta) + \rho\sigma^2\sin(2\theta) \\
&= \sigma^2(1+\rho\sin(2\theta)) \\
\mathbb{E}[Y_IY_Q] &= E[(X_I\cos(\theta) - X_Q\sin(\theta))(X_Q\cos(\theta) + X_I\sin(\theta))] \\
&= \mathbb{E}[X_IX_Q]\cos^2(\theta) + \mathbb{E}[X_I^2]\cos(\theta)\sin(\theta) - \mathbb{E}[X_Q^2]\sin(\theta)\cos(\theta) - \mathbb{E}[X_IX_Q]\sin^2(\theta) \\
&= \rho\sigma^2\cos^2(\theta) + \sigma^2\cos(\theta)\sin(\theta) - \sigma^2\sin(\theta)\cos(\theta) - \rho\sigma^2\sin^2(\theta) \\
&= \rho\sigma^2\cos^2(\theta) - \rho\sigma^2\sin^2(\theta) \\
&= \rho\sigma^2(\cos^2(\theta) - \sin^2(\theta)) \\
&= \rho\sigma^2\cos(2\theta).
\end{aligned}
$$

For example if $\rho = 1$ and the rotation is 90 degrees then the correlation coefficient of Y_I and Y_Q is -1. If X_I and X_Q are uncorrelated ($\rho = 0$) then Y_I and Y_Q are also uncorrelated for any value of θ.

Sometimes we represent a signal vector of length two using a complex numbers. The noise can also be represented with complex numbers. Suppose we have two independent Gaussian random variables X_I and X_Q and represent it as a single complex random variable $X = X_I + jX_Q$. We say the density of X is just the joint density of X_I and X_q.

$$
\begin{aligned}
f_X(x) &= \frac{1}{2\pi\sigma^2}\exp\{-[(x_I-\mu_I)^2 + (x_Q-\mu_Q)^2]/(2\sigma^2)\} \\
&= \frac{1}{2\pi\sigma^2}\exp\{-|x-\mu|^2/(2\sigma^2)\}
\end{aligned}
$$

where $\mu = \mu_I + j\mu_Q$ is the mean of the complex random variable and σ^2 is the common variance of X_I and X_Q. Note that the density depends on the magnitude square of the different between x and μ.

3.4 Random Processes

3.4.1 Basic Definitions

A random processes is a collection of random variables indexed in some way, often by time. At time t the random processes $X(t)$ is a random variable. The random variable $X(t)$ will have some distribution and for a continuous random process $X(t)$ will have a density function that might vary with time. The mean function of a random process is $\mu_X(t) = \mathbb{E}[X(t)]$ which also may be a function of time. Typically a noise random process will have mean of 0. The variance of a random process is $\sigma_X^2(t) = \mathbb{E}[(X(t) - \mu_X(t))^2]$ which also might be a function of time. The pair of random variables $X(t_1)$ and $X(t_2)$ will have a joint distribution function for any time variables t_1 and t_2. The autocorrelation of a random process $X(t)$ is defined as [1]

$$R_X(t,\tau) = \mathbb{E}[X(t)X^*(t-\tau)].$$

The autocorrelation measures the correlation between the noise at different points in time. For noise like signals the autocorrelation does not depend on the time but just the time difference between two samples. The (auto) covariance function is defined as

$$\begin{aligned} C_X(t_1,t_2) &= \mathbb{E}[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))^*] = \mathbb{E}[X(t_1)X^*(t_2)] - \mu_X(t_1)\mu_X^*(t_2) \\ &= R_X(t_1,t_2) - \mu_X(t_1)\mu_X^*(t_2). \end{aligned}$$

If the mean of the random process is zero, as is typical of noise, then the covariance function and the autocorrelation function are identical. A random process for which $\mu_X(t) = \mu_X(0)$ and $C_X(t_1,t_2) = C_X(0,t_2 - t_1)$ is called a wide-sense stationary (WSS) random process. For a WSS random process the mean is a constant (not a function of time) and the covariance function (and autocorrelation function) depends only on the time difference between the samples of the random process. For a wide-sense stationary random process the mean is $\mu_x(t) = \mu_X(0)$ and the autocorrelation is written as $R_X(\tau) = \mathbb{E}[X(t)X^*(t-\tau)] = \mathbb{E}[X^*(t)X(t+\tau)]$.

A random process $X(t)$ that is Gaussian distributed for each t with mean 0 and constant variance σ^2 and for which $X(t)$ and $X(s)$ are independent for $t \neq s$ is called a white Gaussian noise random processes. This process is also wide-sense stationary. For white Gaussian noise random process we have that $\mathbb{E}[X(t)X(s)] = 0$ for $t \neq s$.

The power of a (wide-sense stationary) random process $X(t)$ is

$$P_X = \mathbb{E}[|X(t)|^2] = R_X(0).$$

To see the meaning of this definition we start with the time average power of a particular realization of a random process over a finite interval. Then take the (probabilistic) average to get the average power.

$$\begin{aligned} P_X &= \mathbb{E}\left[\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |X(t)|^2 dt\right] \\ &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} E[|X(t)|^2]dt \\ &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} R_X(0)dt \\ &= R_X(0). \end{aligned}$$

[1]There are two different definitions of autocorrelation function for complex random processes that are complex conjugates of each other. The second definition of autocorrelation function is $R_X(t,\tau) = \mathbb{E}[X(t)X^*(t+\tau)]$. Mostly we are interested in the real part of the autocorrelation function so either definition is fine

The power spectral density of a wide sense stationary random process is defined as the Fourier transform of the autocorrelation function.

$$\begin{aligned} S_X(f) &= \int_{\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau}d\tau \\ R_X(\tau) &= \int_{\infty}^{\infty} S_X(f)e^{j2\pi f\tau}df. \end{aligned}$$

Then the average power of a random process is

$$P_X = R_X(0) = \int_{-\infty}^{\infty} S_X(f)df.$$

Thus the power spectral density is a measure of how the power is distributed over frequency. Another way to see that the Fourier transform of autocorrelation is power spectral density is to put a random process through a linear time-invariant filter which has a very narrow bandpass transfer function. That is, suppose a random process with power spectral density $S_X(f)$ is the input to a filter with transfer function $H(f) = 0$ except for $f \in [f_0, f_0 + \Delta f]$ for some very small Δf where the transfer function is 1. Then the power at the output of the filter is the power spectral density of the input evaluated at the frequency that the filter passes (f_0) times Δf.

3.4.2 White Gaussian Noise

Thermal motion of electrons in electronic circuits generates noise, called thermal noise or Johnson noise. Suppose that $S_n(f)$ is the power spectral density of a thermal noise random process. Since for a real random process the power at negative frequencies is the same as positive frequencies, i.e. $S_n(f) = S_n(-f)$, the "one-sided" power spectral density $\tilde{S}_n(f)$can be expressed in terms of the two sided power spectral density.

$$\begin{aligned} \tilde{S}_n(f) &= S_n(f) + S_n(-f) \\ &= 2S_n(f), \qquad f \geq 0. \end{aligned}$$

For thermal or Johnson noise the one-sided power spectral density is

$$\tilde{S}_n(f) = \frac{hf}{e^{hf/kT_0} - 1}, \quad f \geq 0 \tag{3.4}$$

$$S_n(f) = \frac{hf}{2(e^{hf/kT_0} - 1)}, \quad -\infty < f < \infty \tag{3.5}$$

where $h = 6.626 \times 10^{-34}$ Joules-seconds is Planck's constant, $k = 1.380 \times 10^{-23}$ Joules/Kelvin is Boltzmann's constant and T_0 is the temperature in Kelvin. The "one-sided" power spectral density considers only positive frequencies. That is, the power at negative frequencies is considered part of the positive frequency axis. This one-sided power spectral density is shown in Figure 3.8. As can be seen the power spectral density is essentially flat up till frequencies around 10^{13} Hz (10 Terra Hz). Since most radios operate at frequencies well below 1 THz, a good model for the noise is a flat two-sided power spectral density (same noise power at all frequencies).

$$S_n(f) \approx \frac{kT_0}{2}, \ |f| \ll \frac{kT_0}{h}$$

which is obtained by approximating $e^x \approx 1 + x$ for x small in (3.4). For temperature of 290 Kelvin, $kT_0/h = 6 \times 10^{12}$. So for frequencies less than 10^{12}=1THz a model for the noise that has flat power

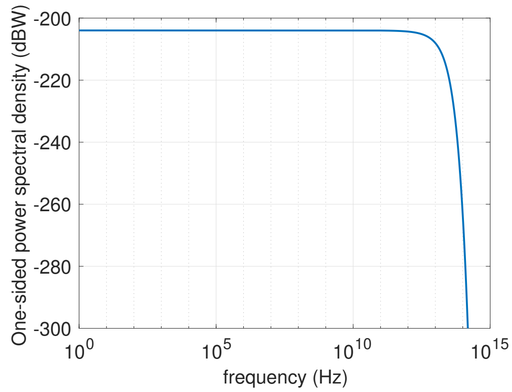

Figure 3.8: One-sided power spectral density of thermal (Johnson) noise

spectral density is appropriate. There is an error of about 8% for $f = 10^{12}$ and an error of 0.83% for $f = 100GHz$ Since the noise power spectral density is a constant (same power at all frequencies), this is called white noise. The two-sided power spectral density of white noise is $N_0/2$ where $N_0 = kT_0$. For room temperatures $T_0 = 290$K which makes $N_0 = 4 \times 10^{-21}$ Watts/Hz. In dBs this is $N_0 = -174$ dBm/Hz or $N_0 = -204$ dBW/Hz. Since the autocorrelation is the inverse Fourier transform of the power spectral density, the power spectral density and autocorrelation of white noise are given by

$$\begin{aligned} S_N(f) &= \frac{N_0}{2} \\ R_N(\tau) &= \frac{N_0}{2}\delta(\tau). \end{aligned}$$

The distribution of the noise (at any time) is zero mean Gaussian. Notice that the power, which is the integral of the power spectral density, of white noise is infinite. Since there is the same power at all frequencies the total power must be infinite. As mentioned above, white noise is only a model for what happens in practice. When operating at frequencies below frequencies of 1THz this is a very accurate model. Since the receiver always filters the signal and noise, the noise power at the output of the filter becomes finite. The ideal autocorrelation and power spectral density of white Gaussian noise are shown in Figure 3.9.

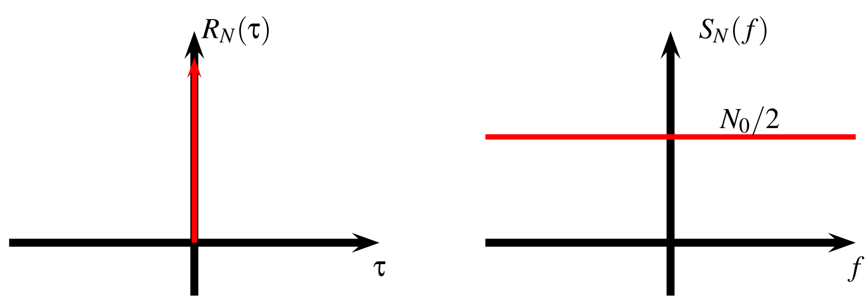

Figure 3.9: Autocorrelation and two-sided power spectral density of white noise

Simulation of white noise is not possible since, in theory, it has unlimited bandwidth. However, if we limit the bandwidth we can generate noise with flat (constant) power spectral density up to the simulation bandwidth. Figure 3.10 shows a realization in time of white Gaussian noise and the corresponding magnitude squared of the frequency content. In this example the simulation bandwidth is 5MHz. If

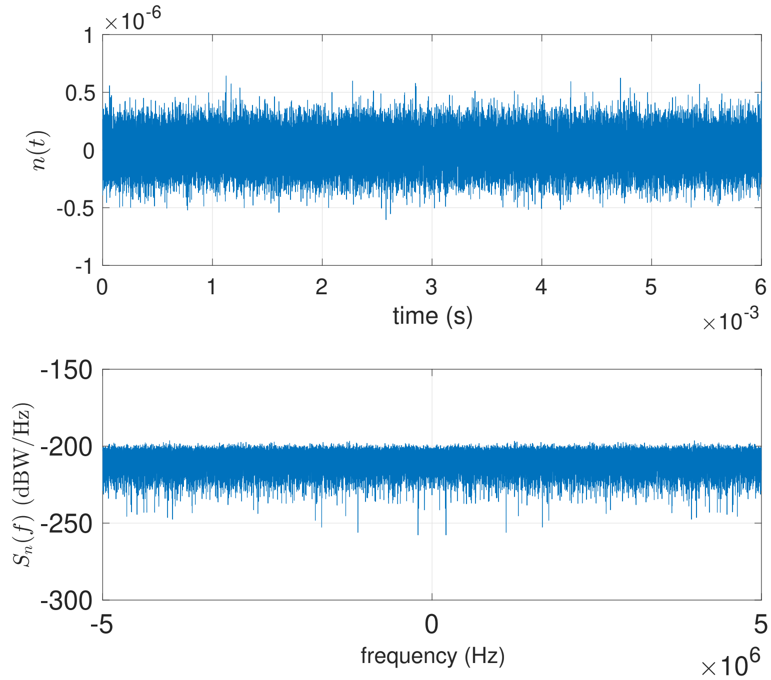

Figure 3.10: Noise in time domain and frequency domain

many realizations were generated and the frequency content averaged, it would become a constant power spectral density.

Example: What is the power of white Gaussian noise in the WiFi frequency band from 2400 MHz to 2483 MHz at room temperature? We need to consider the power at both the positive frequencies and negative frequencies.

$$
\begin{aligned}
P_N &= \int_{2.4\times 10^6}^{2.483\times 10^6} \frac{N_0}{2} df + \int_{-2.483\times 10^6}^{-2.4\times 10^6} \frac{N_0}{2} df \\
&= \int_{2.4\times 10^6}^{2.483\times 10^6} N_0 df = 83 \times 10^6 \times 4 \times 10^{-21} \\
&= 3.32 \times 10^{-13} \text{Watts} \\
&= -124 \text{dBW} \\
&= -94 \text{dBm}
\end{aligned}
$$

3.4.3 Noise through a filter

The receiver in a communication system typically has a filter that is able to select a frequency band where the signal is located and filter out signals outside that band. Noise at the input to a filter will be affected by the filtering process. We wish to characterize the statistics at the output of a filter when noise is the input to a filter. Of importance is the how the variance of the noise at the output of the filter depends on the filter characteristics when the input is white Gaussian noise. Noise will generally have zero mean at the input to the filter. The noise at the output of the filter will have zero mean if the input has zero mean. If the input to the filter is a Gaussian random process then the output will also be a Gaussian random process. To determine the complete distribution then we need to know the variance of the noise at the output of the filter. This will depend on the filter characteristics, i.e. the impulse response or transfer function. We can find the variance by finding the autocorrelation function since $\sigma_Y^2 = R_Y(0)$. Note that since noise has zero mean when we calculate the autocorrelation function, that is also the autocovariance function.

Consider noise at the input to a filter as shown in Figure 3.11.. The power spectral density of the

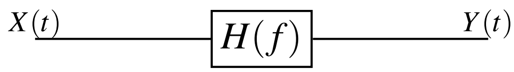

Figure 3.11: Noise at input and output of a filter

output of the filter is determined from the power spectral density at the input to the filter and the transfer function of the filter.

$$S_Y(f) = |H(f)|^2 S_X(f).$$

The autocorrelation of $Y(t)$ is given by

$$\begin{aligned}
R_Y(\tau) &= E[Y^*(t)Y(t+\tau)] = E[Y(t)Y^*(t-\tau)] \\
&= E[\int_{-\infty}^{\infty} X^*(t-\alpha)h^*(\alpha)d\alpha \int_{-\infty}^{\infty} X(t+\tau-\beta)h(\beta)d\beta] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[X^*(t-\alpha)X(t+\tau-\beta)]h^*(\alpha)h(\beta)d\alpha d\beta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_X(\tau+\alpha-\beta)]h^*(\alpha)h(\beta)d\alpha d\beta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_X(\tau-\gamma-\beta)\tilde{h}^*(\gamma)h(\beta)d\gamma d\beta \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} R_X((\tau-\beta)-\gamma)\tilde{h}^*(\gamma)d\gamma\right] h(\beta)d\beta \\
&= R_X(\tau) * \tilde{h}^* * h,
\end{aligned}$$

where $\tilde{h}(t) = h(-t)$ and the change of variables $\gamma = -\alpha$ was used. The power spectral density of the output is determined as

$$\begin{aligned}
S_Y(f) &= \int_\tau R_X(\tau)e^{-j2\pi f\tau}d\tau \\
&= \int_{\tau=-\infty}^{\infty}\int_{\beta=-\infty}^{\infty}\left[\int_{\gamma=-\infty}^{\infty} R_X((\tau-\beta)-\gamma)\tilde{h}^*(\gamma)d\gamma\right] h(\beta)d\beta e^{-j2\pi f\tau}d\tau \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} R_X(u)\tilde{h}^*(\gamma)d\gamma\right] h(\beta)e^{-j2\pi f(u+\beta+\gamma)}d\beta du \\
&= \int_{-\infty}^{\infty} R_X(u)e^{-j2\pi fu}du \int_{-\infty}^{\infty}\tilde{h}^*(\gamma)e^{-j2\pi f\gamma}d\gamma \int_{-\infty}^{\infty} h(\beta)e^{-j2\pi f\beta}d\beta \\
&= S_X(f)H^*(f)H(f) \\
&= S_X(f)|H(f)|^2.
\end{aligned}$$

At any particular time the output due to noise alone is a random variable with a certain density function. The mean of the output is the convolution of the mean of the input signal with the impulse response of the system. The variance of the output is

$$\begin{aligned}
\sigma_Y^2 = \mathrm{Var}[Y(t)] = R_Y(0) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_X(\alpha-\beta)h^*(\alpha)h(\beta)d\alpha d\beta \\
&= \int_{-\infty}^{\infty} |H(f)|^2 S_X(f)df
\end{aligned}$$

Example: Consider the case when the noise is white with power spectral density $N_0/2$ the variance of the output is

$$\begin{aligned}
\sigma_Y^2 &= \mathrm{Var}[Y(t)] \\
&= \frac{N_0}{2}\int_{-\infty}^{\infty} |h(t)|^2 dt \\
&= \frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2 df.
\end{aligned}$$

Figure 3.12 shows the power spectral density of white noise and the filter transfer function. Note that if the energy of the impulse response of the filter is 1, then the variance of the output of the filter is just $N_0/2$. If the filter has an ideal bandpass transfer function as shown in Figure 3.12 with bandwidth W then

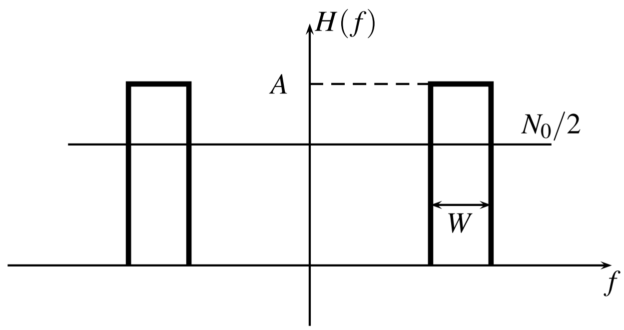

Figure 3.12: Ideal bandpass filtering of white noise

the variance of the noise at the output of the filter is

$$\sigma_Y^2 = A^2 W N_0.$$

A filter for which the noise variance is σ_Y^2 but does not have the brickwall shape is said to have noise bandwidth $\sigma_Y^2/(A^2N_0)$ where A is the peak output. As another example consider noise through a filter with impulse response

$$h(t) = p_T(t).$$

If the input is white Gaussian noise then the noise variance at the output of the receiver is

$$\sigma_Y^2 = \frac{N_0}{2}\int h^2(t)dt = \frac{N_0 T}{2}.$$

A simulation of the input noise and output noise is shown in Figure 3.13. In this figure the noise bandwidth is 10Hz. The filter is a low pass filter that integrates over a one second time period. The high frequency content of the noise is filtered out. As can be seen the noise variance at the output is significantly smaller than the noise variance of the input. In simulation the noise variance of the input is limited by the simulation bandwidth or sampling rate.

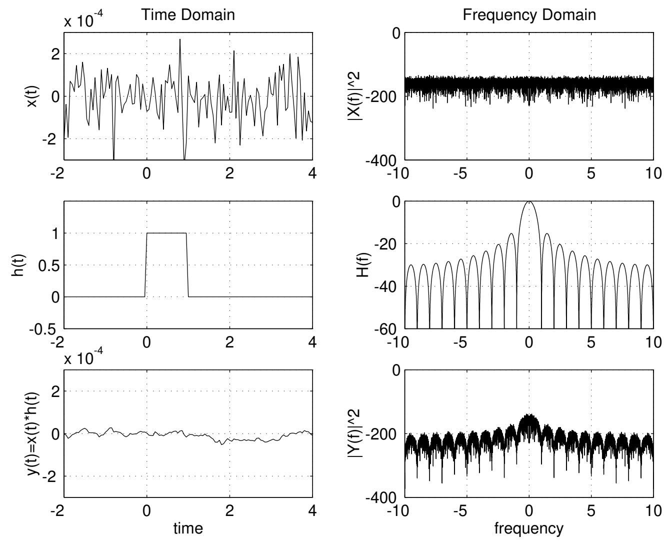

Figure 3.13: Example of noise through a filter

One other type of processing of noise (besides regular filtering) is correlating with a sinusoid. So consider (real) noise multiplied by a sinusoid into an integrator. That is,

$$\eta = \int_0^T n(t)A\cos(2\pi f_c t)dt$$

where $n(t)$ is white Gaussian noise with two sided power spectral density $N_0/2$. Assume that $f_c T \gg 1$. Because we are performing a linear operation on Gaussian noise, the resulting random variable η is Gaussian. Since the mean on $n(t)$ is zero, the mean of η is zero. The variance is determined as follows.

$$\begin{aligned}
\text{Var}[\eta] &= E[\eta^2] \\
&= E[\int_0^T n(t)A\cos(2\pi f_c t)dt \int_0^T n(s)A\cos(2\pi f_c s)ds] \\
&= E[\int_0^T \int_0^T n(t)n(s)A^2\cos(2\pi f_c t)\cos(2\pi f_c s)dtds] \\
&= \int_0^T \int_0^T E[n(t)n(s)]A^2\cos(2\pi f_c t)\cos(2\pi f_c s)dtds \\
&= \int_0^T \int_0^T \frac{N_0}{2}\delta(t-s)A^2\cos(2\pi f_c t)\cos(2\pi f_c s)dtds \\
&= \int_0^T \frac{N_0}{2}A^2\cos^2(2\pi f_c t)dt \\
&= \frac{A^2 N_0}{4}\int_0^T (1+\cos(2\pi 2 f_c t))dt = \frac{A^2 T N_0}{4}.
\end{aligned}$$

More generally, consider zero mean white Gaussian noise with power spectral density $S_x(f) = N_0/2$ or autocorrelation $R_x(\tau) = \frac{N_0}{2}\delta(\tau)$. If the noise is multiplied by a real waveform $x(t)$ and integrated then the variance can be calculated as follows.

$$
\begin{aligned}
\eta &= \int n(t)x(t)dt \\
\mathrm{Var}[\eta] &= \mathbb{E}[\eta^2] \\
&= \mathbb{E}[\int n(t)x(t)dt \int n(s)x(s)ds] \\
&= \int\int \mathbb{E}[n(t)n(s)]x(t)x(s)dtds \\
&= \int\int \frac{N_0}{2}\delta(t-s)x(t)x(s)dtds \\
&= \int \frac{N_0}{2}x^2(t)dt = \frac{N_0}{2}\int x^2(t)dt.
\end{aligned}
$$

Consider again zero mean white Gaussian noise with power spectral density $S_N(f) = N_0/2$. Let $\varphi_0(t)$ and $\varphi_1(t)$ be real orthonormal waveforms. If we correlate the noise with $\varphi_0(t)$ and $\varphi_1(t)$ the expected value of the product of the correlator outputs will be 0.

$$
\begin{aligned}
\eta_0 &= \int n(t)\varphi_0(t)dt \\
\eta_1 &= \int n(t)\varphi_1(t)dt \\
\mathbb{E}[\eta_0\eta_1] &= \mathbb{E}\left[\int n(t)\varphi_0(t)dt \int n(s)\varphi_1(s)ds\right] \\
&= \int \mathbb{E}[n(t)n(s)]\varphi_0(t)\varphi_1(s)dtds \\
&= \int\int \frac{N_0}{2}\delta(t-s)\varphi_0(t)\varphi_1(s)dtds \\
&= \int \frac{N_0}{2}\varphi_0(t)\varphi_1(t)dt \\
&= \frac{N_0}{2}\int \varphi_0(t)\varphi_1(t)dt \\
&= 0.
\end{aligned}
$$

Since for Gaussian noise at the input, the noise at the output is also Gaussian. Since η_0 and η_1 are Gaussian and uncorrelated, they are also independent.

The noise we consider is white Gaussian noise with power spectral density $N_0/2$. The noise is a random process. However, we can separate the randomness part from the time variation part using a complete orthonormal set of functions. A complete orthonormal set of waveforms is one such that any finite energy waveform can be expressed as a linear combination of the set of orthonormal waveforms with no error. One example is the waveforms over the interval from 0 to T with finite energy. These signals can be expressed as linear combination of orthonormal signals. In this example, the orthonormal signals are sinusoids of frequencies that are multiples of $f_0 = 1/T$. This is just the Fourier series.

Now suppose that $\psi_0, \psi_1, \ldots$ is a complete set of orthonormal waveforms over some interval of time. Then we can represent the noise $n(t)$ over that time interval as

$$n(t) = \sum_{m=0} \eta_m \psi_m(t)$$

where η_i is a Gaussian random variable with mean 0 and variance $N_0/2$. In addition η_i and η_j are uncorrelated and Gaussian and thus independent. The random variables η_m are determined as

$$\eta_m = \int n(t)\psi_m(t)dt.$$

So we can start with a random process $n(t)$ and find a representation of that random process in terms of random variables $\eta_m, m = 0, 1, \ldots$ and deterministic functions of time $\psi_m(t), m = 0, 1, \ldots$. Alternatively starting with the random variables and functions of time we can find a random process. This decomposition of a random process into a collection of random variables and deterministic functions of time is called the **Karhunen-Loève** transform.

Summary of Noise through a filter: Let $n(t)$ be white Gaussian noise (mean 0, power spectral density $N_0/2$). Then

$$\eta_1 = \int n(t)x(t)dt \quad \Rightarrow \quad \mathrm{Var}[\eta_1] = \tfrac{N_0}{2}\int x^2(t)dt$$

$$\eta_2 = \int_{-\infty}^{\infty} n(t)h(T-t)dt \quad \Rightarrow \quad \mathrm{Var}[\eta_2] = \tfrac{N_0}{2}\int_{-\infty}^{\infty} h^2(t)dt = \tfrac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2 df$$

$$\eta_3 = \int_0^T n(t)\cos(2\pi f_c t)dt \quad \Rightarrow \quad \mathrm{Var}[\eta_3] = \tfrac{N_0 T}{4}.$$

3.5 Bandwidth of Digital Signals

The goals of this section is to determine the bandwidth of digitally modulated signals. Because the signals are random (due to the random nature of the data) we need to determine the power spectral density of the random process corresponding to the transmitted signal.

In many digital communication systems the transmitted signal $W(t)$ is an infinite sequence of amplitude modulated pulses or waveforms i.e.

$$W(t) = \sum_{\ell=-\infty}^{\infty} b_\ell x(t-\ell T).$$

This signal is a random process because the data sequence b_ℓ is a random sequence. This corresponds to using time-shifted pulses to transmit data. The random process $W(t)$, however, is not wide sense stationary. That is, the autocorrelation $R_W(t,\tau) = E[W(t)W(t+\tau)]$ typically depends on both t and τ. This means we can not take the Fourier transform of the autocorrelation function to get the power spectral density (because the autocorrelation is a function of both t and τ). In this case a random delay is introduced in order to make the signal wide sense stationary so that the power spectral density can be determined.

A digital data signal is modeled as a random process $Y(t)$ which is a wide sense stationary version of the random process $W(t)$;

$$Y(t) = W(t-U)$$

where U is a random variable needed in order to make $Y(t)$ wide sense stationary. If U is uniformly distributed between 0 and T then $Y(t)$ is a wide sense stationary random process as we will see below. First note that if the expected value of the data symbols is 0 then the mean of $W(t)$ and $Y(t)$ is zero. Next we desire then to compute the auto correlation of $Y(t)$ and also the power spectral density (psd) of $Y(t)$. Assume that $\{b_\ell\}_{\ell=-\infty}^{\infty}$ is a sequences of i.i.d. random variables with zero mean and variance σ^2 For example if $P\{b_\ell = +1\} = 1/2$ and $P\{b_\ell = -1\} = 1/2$ then $\sigma^2 = 1$. Also we need to assume U and b_ℓ are independent. Then the power spectral density of the random process Y can be calculated as follows.

Claim:

$$\begin{aligned} R_Y(\tau) &= \frac{\sigma^2}{T}\int_{-\infty}^{\infty} x(t)x^*(t-\tau)dt \\ S_Y(f) &= \frac{\sigma^2}{T}|X(f)|^2 \text{ where } X(f) = \mathcal{F}\{x(t)\} \end{aligned}$$

Proof:

For any t and τ

$$\begin{aligned}
\mathbb{E}[Y(t)Y^*(t-\tau)] &= \mathbb{E}\left[\sum_{\ell=-\infty}^{\infty} b_\ell x(t-\ell T-U)\sum_{m=-\infty}^{\infty} b_m^* x^*(t-\tau-mT-U)\right] \\
&= \sum_{\ell=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} \underbrace{\mathbb{E}\{b_\ell b_m^*\}}_{\delta_{\ell m}=\begin{cases}\sigma^2, & \ell=m\\ 0, & \ell\neq m\end{cases}} \mathbb{E}\left[x(t-\ell T-U)x^*(t-\tau-mT-U)\right] \\
&= \sum_{\ell=-\infty}^{\infty} \sigma^2 E\left[x(t-\ell T-U)x^*(t-\tau-\ell T-U)\right] \\
&= \sum_{\ell=-\infty}^{\infty} \sigma^2 \frac{1}{T}\int_{u=0}^{T} x(t-\ell T-u)x^*(t-\tau-\ell T-u)du \\
&= \frac{1}{T}\sum_{\ell=-\infty}^{\infty} \sigma^2 \int_{\ell T}^{(\ell+1)T} x(t-v)x^*(t-\tau-v)dv \quad (v=\ell T+u, \quad dv=du) \\
&= \frac{\sigma^2}{T}\int_{-\infty}^{\infty} x(t-v)x^*(t-\tau-v)dv \qquad (s=t-v) \\
&= \frac{\sigma^2}{T}\int_{-\infty}^{\infty} x(s)x^*(s-\tau)ds = \frac{\sigma^2}{T}\int_{-\infty}^{\infty} x(t)x^*(t-\tau)dt.
\end{aligned}$$

Because the $\mathbb{E}[Y(t)Y^*(t-\tau)]$, the autocorrelation of $Y(t)$, depends only on τ and not on t, $Y(t)$ is wide sense stationary with autocorrelation

$$R_Y(\tau) = \frac{\sigma^2}{T}\int_{-\infty}^{\infty} x(t)x^*(t-\tau)dt.$$

Now let $g_1(t)=x(t)$ and $g_2(t)=x^*(-t)$ then

$$g_1 * g_2(\tau) = \int_{-\infty}^{\infty} g_1(t)g_2(\tau-t)dt = \int_{-\infty}^{\infty} x(t)x^*(t-\tau)du.$$

Thus the correlation of $x(t)$ with $x^*(t-\tau)$ can be written as a convolution of two functions. So

$$R_Y(\tau) = \frac{\sigma^2}{T}(g_1 * g_2)(\tau).$$

Now the power spectral density is the Fourier transform of the autocorrelation function.

$$S_Y(f) = \mathcal{F}\left\{\frac{\sigma^2}{T}(g_1 * g_2)(\tau)\right\} = \frac{\sigma^2}{T}G_1(f)G_2(f).$$

Now we relate the spectral content of $g_1(t)$ and $g_2(t)$ to the spectral content of $x(t)$ as

$$\begin{aligned}
G_1(f) &= \mathcal{F}\{x(t)\} = X(f) \\
G_2(f) &= \mathcal{F}\{x^*(-t)\} = X^*(f).
\end{aligned}$$

Thus

$$S_Y(f) = \frac{\sigma^2}{T}|X(f)|^2.$$

3.5.1 Example 1: Rectangular Pulses

The first example is that of a rectangular pulse shape of duration T seconds. That is,

$$x(t) = \sqrt{P}p_T(t) = \begin{cases} \sqrt{P} & 0 \le t \le T \\ 0 & \text{elsewhere} \end{cases}$$

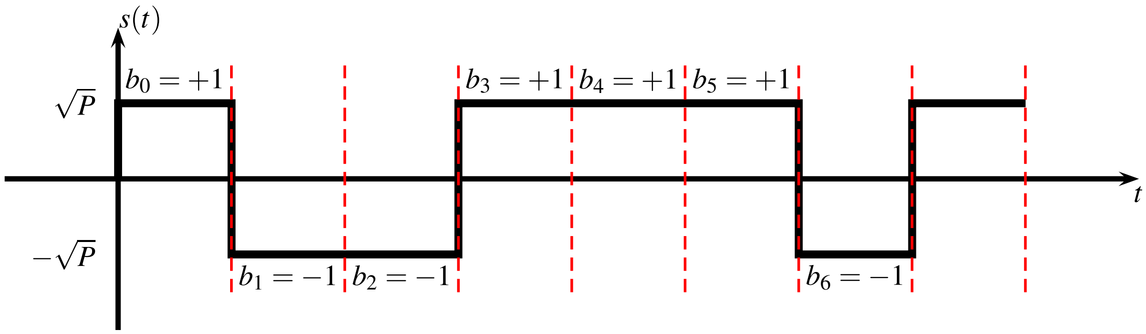

Figure 3.14: Digitally modulated signal using rectangular pulses

where $b_\ell \in \{\pm 1\}$ which means that the variance of the data $\sigma^2 = 1$. The signal $s(t)$ based on rectangular pulses is shown in Figure 3.14 To determine the power spectral density we first determine the autocorrelation of the pulse shape. First note that

$$
\begin{aligned}
p_T(t+\tau) &= \begin{cases} 1 & 0 \le t+\tau \le T \\ 0 & \text{elsewhere} \end{cases} \\
&= \begin{cases} 1 & -\tau \le t \le T-\tau \\ 0 & \text{elsewhere.} \end{cases}
\end{aligned}
$$

The integration can now be done with the help of Figure 3.15.

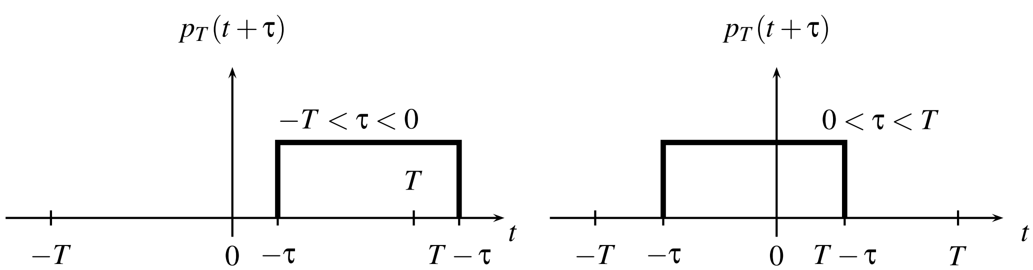

Figure 3.15: Time shifted pulses

$$
\int_{-\infty}^{\infty} x(t)x(t+\tau)dt = P\int_0^T p_T(t+\tau)dt \quad = \quad \begin{cases} P(T-\tau) & 0 \le \tau \le T \\ P(T+\tau) & -T \le \tau \le 0 \\ 0 & \text{elsewhere} \end{cases}
$$

The autocorrelation and the power spectral density of the random signal $Y(t)$ is

$$\begin{aligned}
R_Y(\tau) &= \begin{cases} \frac{P}{T}(T-|\tau|) & |\tau| \le T \\ 0 & \text{elsewhere} \end{cases} \\
S_Y(f) &= \mathcal{F}\{R_Y(\tau)\} = P[\mathcal{F}\{\Lambda_T(t)\}] \\
X(f) &= \mathcal{F}\{\sqrt{P}p_T(t)\} = \sqrt{P}T\left[\frac{\sin(\pi f T)}{\pi f T}\right]e^{-j\pi f T} \\
|X(f)|^2 &= PT^2\frac{\sin^2(\pi f T)}{(\pi f T)^2} \\
S_Y(f) &= \frac{1}{T}|X(f)|^2 = PT\frac{\sin^2(\pi f T)}{(\pi f T)^2} \\
S_Y(f) &= PT\frac{\sin^2(\pi f T)}{(\pi f T)^2} = PT\mathrm{sinc}^2(fT) \\
\mathrm{sinc}(x) &= \frac{\sin(\pi x)}{\pi x}.
\end{aligned}$$

The power spectral density $S_Y(f)$ is shown in Figures 3.16, 3.17, and 3.18 for $P=1$ and $T=1$. The nulls in the spectrum occur at frequencies $f=\pm n/T$ for $n=1,2,3,\ldots$ Note that for large f, the power spectral density decays with frequency as $1/(\pi^2 f^2 T^2)$ as seen in Figure 3.18. This rather slow decay in the spectrum at high frequencies is due to the high frequency components associated with instantaneous transitions in the time domain as seen in Figure 3.14. The next example eliminates the discontinuities in time.

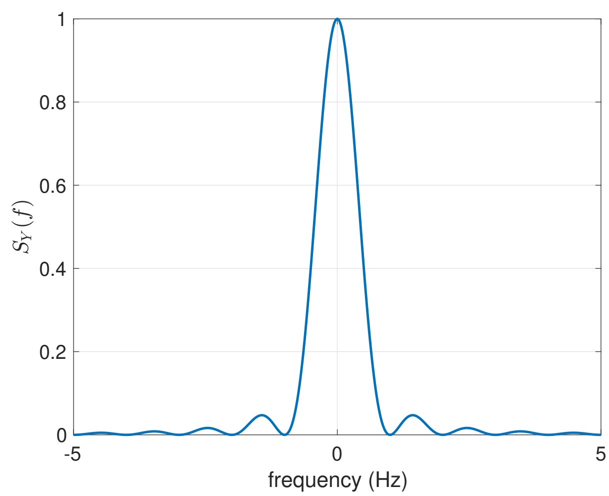

Figure 3.16: Spectrum for rectangular pulses, $P=1$, $T=1$.

3.5.2 Example 2: Half Cosine/Sine Pulses

Consider a pulse that consists of a half sinusoid. This pulse shape, when used to generate a signal is continuous, unlike the rectangular pulse shape.

$$x(t) = \sqrt{2P}\sin(\pi t/T)p_T(t)$$

The waveform $W(t)$ for this pulse shape is shown in Figure 3.19. The Fourier transform is given by

$$X(f) = \int_{-\infty}^{\infty}\sqrt{2P}\sin(\pi t/T)p_T(t)e^{-j2\pi f t}\,dt$$

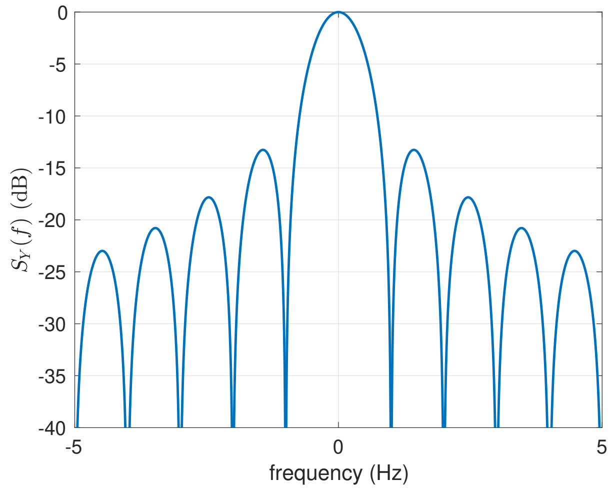

Figure 3.17: Spectrum in dB for rectangular pulses, $P = 1$, $T = 1$.

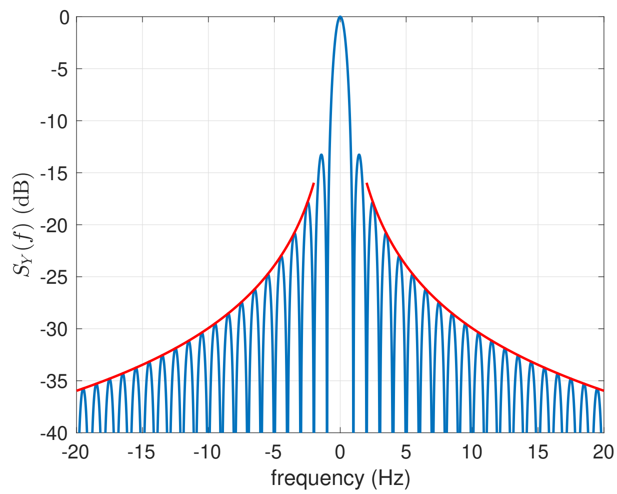

Figure 3.18: Spectrum in dB for rectangular pulses, $P = 1$, $T = 1$.

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{\sqrt{2P}}{2j}(e^{j\pi t/T} - e^{-j\pi t/T})p_T(t)e^{-j2\pi ft}dt \\
&= \frac{\sqrt{2P}}{2j}\int_0^T (e^{j\pi t/T} - e^{-j\pi t/T})e^{-j2\pi ft}dt \\
&= \frac{\sqrt{2P}}{2j}\int_0^T (e^{j\pi t/T(1-2fT)} - e^{-j\pi t/T(1+2fT)})dt \\
&= \frac{\sqrt{2P}}{2j}\left[\frac{e^{j\pi t/T(1-2fT)}}{j\pi(1-2fT)/T} + \frac{e^{-j\pi t/T(1+2fT)}}{j\pi(1+2fT)/T}\right]_0^T \\
&= \frac{\sqrt{2P}}{2j}\left[\frac{e^{j\pi(1-2fT)}-1}{j\pi(1-2fT)/T} + \frac{e^{-j\pi(1+2fT)}-1}{j\pi(1+2fT)/T}\right] \\
&= \frac{\sqrt{2P}T}{2j}\left[\frac{-e^{-j2\pi fT}-1}{j\pi(1-2fT)} + \frac{-e^{-j2\pi fT}-1}{j\pi(1+2fT)}\right]
\end{aligned}
$$

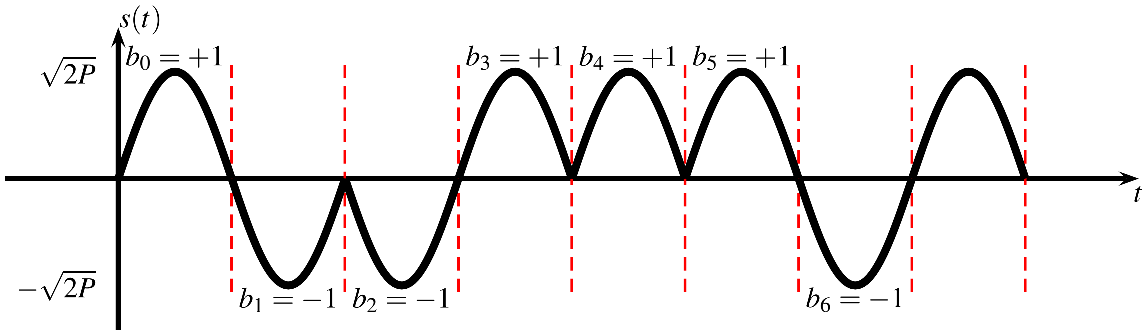

Figure 3.19: Signal using sine pulses

$$
\begin{aligned}
&= \frac{\sqrt{2P}T}{2\pi}(1+e^{-j2\pi fT})\left[\frac{1}{(1+2fT)}+\frac{1}{(1-2fT)}\right]\\
&= \frac{\sqrt{2P}T}{2\pi}(1+e^{-j2\pi fT})\left[\frac{2}{(1-4f^2T^2)}\right]\\
&= \frac{\sqrt{2P}T}{2\pi}e^{-j2\pi fT/2}(e^{j2\pi fT/2}+e^{-j2\pi fT/2})\left[\frac{2}{(1-4f^2T^2)}\right]\\
&= \frac{\sqrt{2P}T}{2\pi}e^{-j2\pi fT/2}(2\cos(2\pi fT/2))\left[\frac{2}{(1-4f^2T^2)}\right]\\
&= \frac{2\sqrt{2P}T}{\pi}e^{-j2\pi fT/2}(\cos(2\pi fT/2))\left[\frac{1}{(1-4f^2T^2)}\right]\\
|X(f)|^2 &= \frac{8PT^2}{\pi^2}\left[\frac{\cos^2(\pi fT)}{(1-4f^2T^2)^2}\right]\\
S_Y(f) &= \frac{1}{T}|X(f)|^2\\
&= \frac{8PT}{\pi^2}\left[\frac{\cos^2(\pi fT)}{(1-4f^2T^2)^2}\right].
\end{aligned}
$$

Figures 3.20, 3.21 and 3.22 show the power spectrum for sinusoidal pulses. The nulls of the spectrum occur at $f = \pm n/2T$ for $n = 3,5,7,\ldots$. The main lobe of the spectrum has width $3/T$. Notice that the spectrum falls of as $1/(2\pi^2 f^4 T^4)$ rather than $1/(\pi fT)^2$ for rectangular pulses. This is due to the fact that there are no discontinuities in the signal in the time domain. There are discontinuities in the derivative of the signal. By eliminating those we cause the spectrum to decay even faster. In the next example the discontinuity of derivative of the signal is eliminated. Figure 3.23 compares the spectrum of rectangular pulses and sinusoidal pulses.

Summary for Half Cosine Pulses

- $x(t) = \sqrt{2P}\sin(\pi t/T)p_T(t)$
- $X(f) = \frac{8PT^2\cos^2(\pi fT)}{\pi^2(1-4f^2T^2)^2}$
- The energy of the pulse is $E = \int x^2(t)dt = \int |X(f)|^2 df = PT$
- The power spectral density of the pulse is $S_Y(f) = \frac{8PT\cos^2(\pi fT)}{\pi^2(1-4f^2T^2)^2}$.
- The power $\int S_Y(f)df = P$

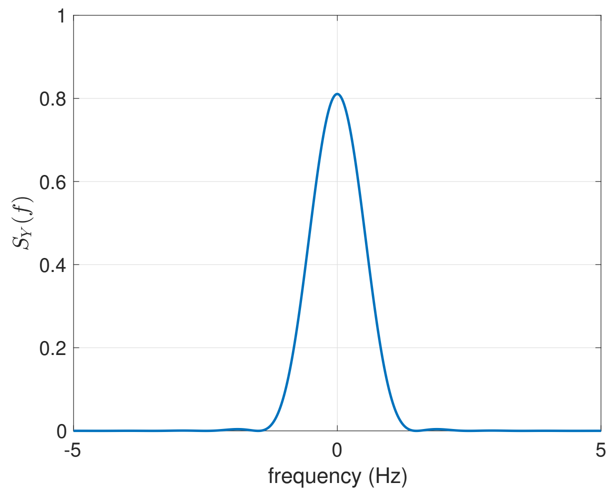

Figure 3.20: Spectrum for sine pulses, $P = 1$, $T = 1$.

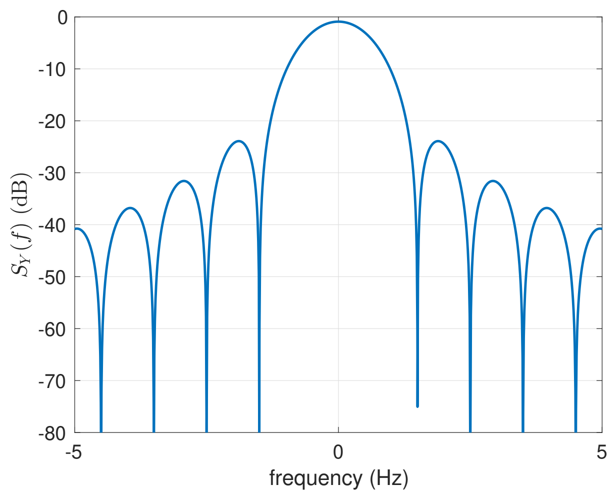

Figure 3.21: Spectrum in dB for sine pulses, $P = 1$, $T = 1$.

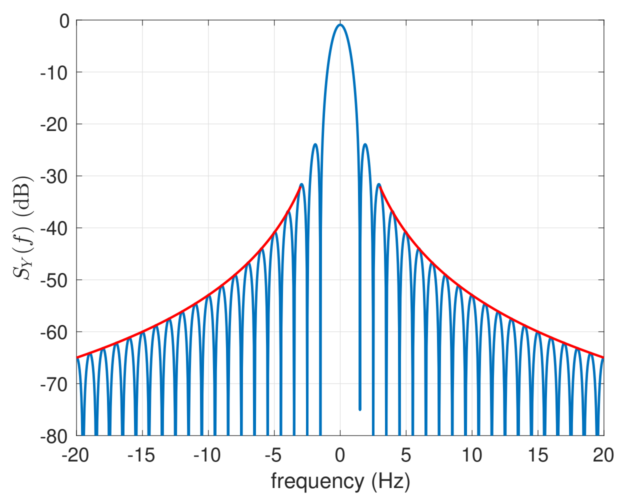

Figure 3.22: Spectrum in dB for sine pulses, $P = 1$, $T = 1$.

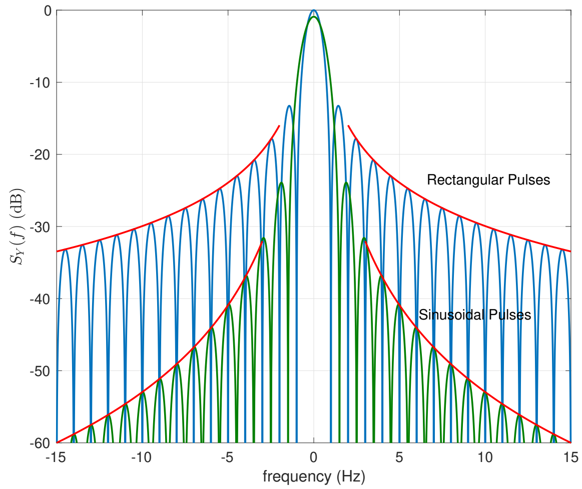

Figure 3.23: Spectrum in dB for rectangular and sine pulses, $P = 1$, $T = 1$.

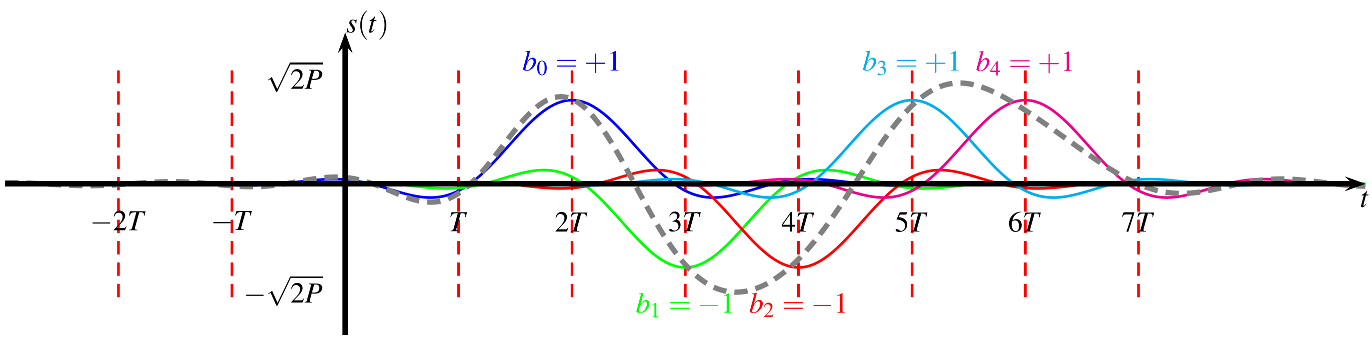

Figure 3.24: Signal using square-root raised cosine pulses

3.5.3 Example 3: Square-Root Raised-Cosine Pulses

The square-root raised cosine pulse we described in Chapter 2 is

$$x(t) = \sqrt{PT}\left[\frac{1}{\sqrt{T}}\left(\frac{\sin(\pi(1-\alpha)t/T) + 4\alpha t/T\cos(\pi(1+\alpha)t/T)}{\pi[1-(4\alpha t/T)^2]t/T}\right)\right].$$

This pulse shape is not time limited and thus extends over more than one symbol duration T. A sequence of delayed pulses is shown in Figure 3.24. The dotted line is the sum of the different pulses for different data bits. When this signal is filtered with a matched filter, the output does not contain any intersymbol interference. That is, the first pulse is orthogonal to a time shifted version of the pulse as was seen in Chapter 2. From Appendix 2B and 2C we see that

$$X(f) = \begin{cases} \sqrt{PT}, & 0 \le |f| \le \frac{1-\alpha}{2T} \\ T\sqrt{\frac{P}{2}[1-\sin(\pi T(|f|-\frac{1}{2T})/\alpha)]}, & \frac{1-\alpha}{2T} \le |f| \le \frac{1+\alpha}{2T} \\ 0, & \text{otherwise.} \end{cases}$$

$$|X(f)|^2 = \begin{cases} PT^2, & 0 \le |f| \le \frac{1-\alpha}{2T} \\ \frac{PT^2}{2}[1-\sin(\pi T(|f|-\frac{1}{2T})/\alpha)], & \frac{1-\alpha}{2T} \le |f| \le \frac{1+\alpha}{2T} \\ 0, & \text{otherwise.} \end{cases}$$

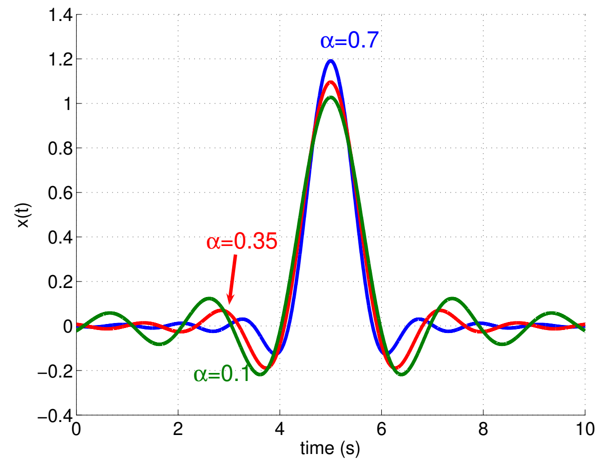

Figure 3.25: Spectrum for square-root raised-cosine pulses

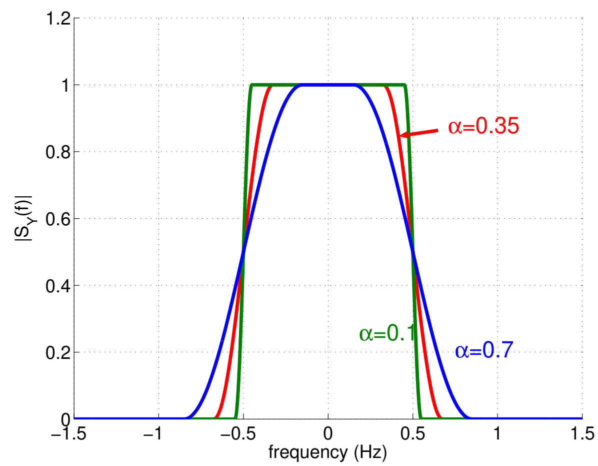

Figure 3.26: Spectrum for square-root raised-cosine pulses

$$S_Y(f) = \begin{cases} PT, & 0 \le |f| \le \frac{1-\alpha}{2T} \\ \frac{PT}{2}[1-\sin(\pi T(|f|-\frac{1}{2T})/\alpha)], & \frac{1-\alpha}{2T} \le |f| \le \frac{1+\alpha}{2T} \\ 0, & \text{otherwise.} \end{cases}$$

Note that the smaller α is the longer the time for the pulse to die out. However, the smaller α the narrower the spectrum. When $\alpha \to 0$ the spectrum becomes flat and concentrated over the interval $[-\frac{1}{2T}, \frac{1}{2T}]$. As mentioned in Chapter 2, the time duration is infinite and as such the pulse is typically truncated in time. If we truncate the pulse in the time domain then the frequency content will be altered. In particular there will be frequency content outside the band $[-(1+\alpha)/(2T), (1+\alpha)/(2T)]$. Consider truncating the pulse so that in the time domain the pulse amplitude is smaller than -40dB lower than the peak and truncating at a zero in the pulse time domain signal. For $\alpha = 0.25$ this corresponds to a time of $5.18T$ symbols on either side of the peak. For this truncation the power spectral density is shown in Figure 3.28. The sidelobes generated by truncation are about 34 dB below the peak of the power spectral density. There also is a very small (about .05dB) passband ripple in the power spectral density. Increasing the pulse duration in the time domain reduces the sidelobes in the frequency domain. Finally, truncation also will induce some level of intersymbol interference.

Ideal square-root raised cosine pulses have no intersymbol interference at the output of a matched filter. The main disadvantage of this pulse shape is the nonconstant envelope of the signal. In order to avoid any distortion, the amplifier generally needs to operate in the linear range and as a result the efficiency of the amplifier in converting

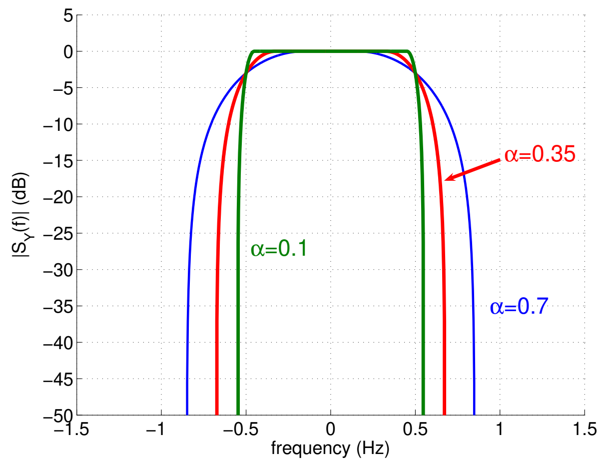

Figure 3.27: Spectrum for square-root raised-cosine pulses

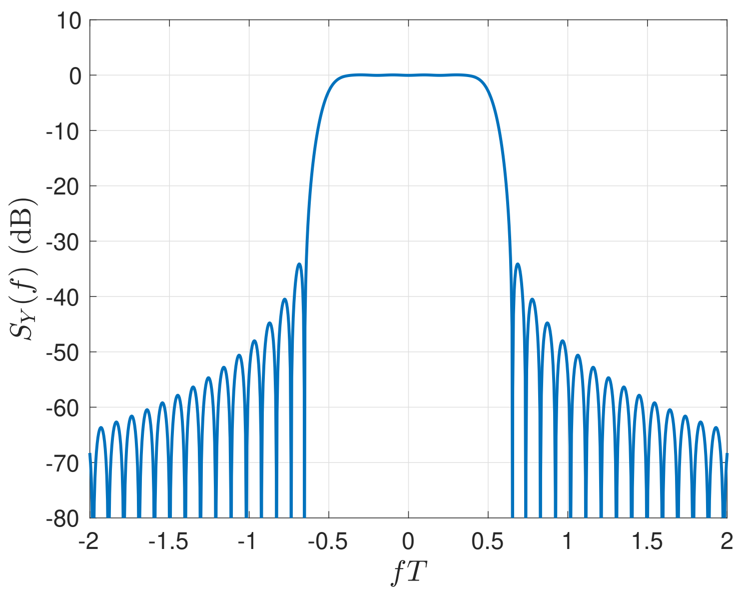

Figure 3.28: Spectrum for square-root raised-cosine pulses truncated with $\alpha = 0.25$.

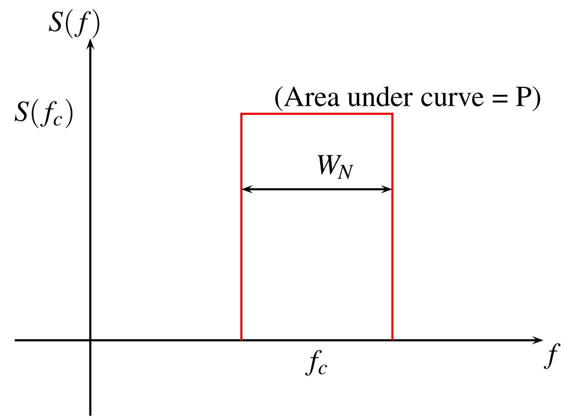

Figure 3.29: Defining noise bandwidth

DC energy to RF energy is reduced.

3.6 Definitions of Bandwidth for Digital Signals

The bandwidth of a signal or pulse is important for various reasons. One reason is that the spectral occupancy of a signal is typically constrained by regulation (for wireless applications) or by the physical media (for wired applications). There are various definitions of bandwidth that can be applied to signals. These definitions are the following (see [5]):

1. Null-to-Null bandwidth: This is defined as the bandwidth of main lobe of power spectral density

2. 99% power bandwidth containment: This is defined as the bandwidth such that 1/2% of power lies above upper band limit and 1/2% lies below lower band limit

3. x dB bandwidth: This is defined as the bandwidth such that spectrum is x dB below spectrum at center of band (e.g. 3dB bandwidth)

4. Noise bandwidth: $\triangleq W_N = P/S(f_c)$ where P is total power and $S(f_c)$ is value of spectrum at $f = f_c$. See Figure 3.29. This is useful in determining how much noise gets through a filter matched to the pulse shape.

5. Absolute bandwidth: $\triangleq W_A = \min\{W : S(f) = 0\ \forall |f| > U\}$. Many pulse shapes, especially those with constant envelope, have infinite absolute bandwidth.

6. Spectral Mask: This is not a definition of bandwidth but a mask that the power spectral density must fall below. By adjusting the pulse shape or the symbol rate the spectrum can be adjusted to fit in the mask. The mask appropriate for WiFi is shown in Figure 3.30. In Figure 3.31 the spectrum of a half sinusoidal pulse is shown with the mask.

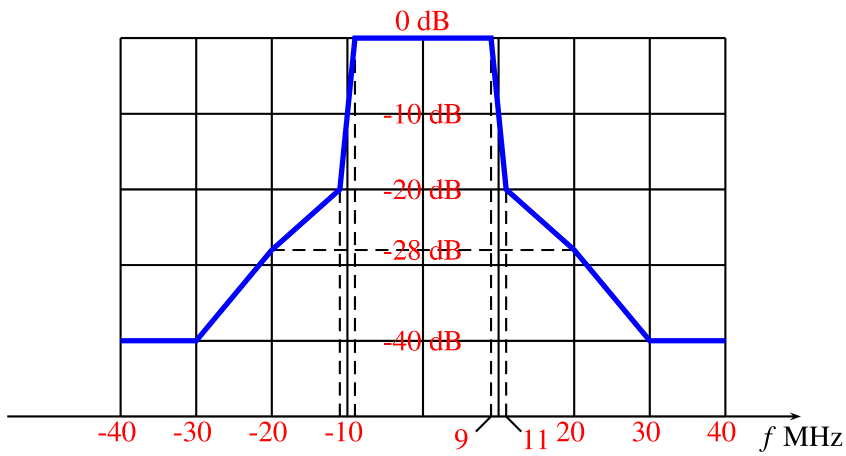

Figure 3.30: Spectral mask for WiFi

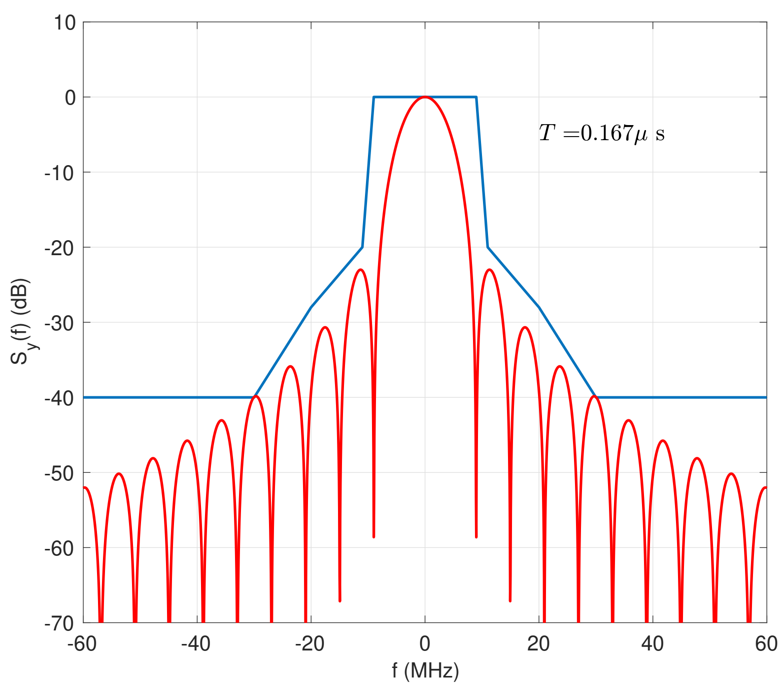

Figure 3.31: WiFi spectral mask and half sinusoidal pulse spectrum

In Table 3.3 we show the values of the time-bandwidth product for various definitions of bandwidth. For example, for the null-to-null bandwidth W_n we show W_nT.

These three modulation schemes all have same error probability as will be seen in the next Chapter. Half sinusoidal pulses have the minimum possible Gabor bandwidth over all modulation schemes whose basic pulse is limited to T seconds.All of these have infinite absolute bandwidth. It should be noted that for each of these pulse shapes there can be both an in-phase signal and a quadrature-phase signal occupying the same bandwidth. So If

	Bandwidth Definition				
Pulse Shape	Null-to-null	99% Power	35dB	Noise	3dB
Rectangular	2.0	20.56	35.12	1.00	0.88
Sinusoidal	3.00	1.18	3.24	1.23	1.18
Sinc pulse (square-root raised cosine with $\alpha =$ 0	1.00	1.00	1.00	1.00	1.00
Square-Root Raised Cosine ($\alpha = 0.05$)	1.05	1.03	1.05	1.0	1.0
Square-Root Raised Cosine ($\alpha = 0.25$)	1.25	1.10	1.24	1.0	1.0
Square-Root Raised Cosine ($\alpha = 0.50$)	1.50	1.27	1.49	1.0	1.0

Table 3.3: Different bandwidths for different pulses normalized to $1/T$.

we take the noise bandwidth as a measure of the bandwidth of a signal there are two dimensions in a bandwidth $1/T$ in T seconds (for rectangular, sinc and raised cosine pulses). This satisfies the $N = 2WT$ relation between dimensions, bandwidth, and time.

Summary of Chapter 3 concepts:

- For two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- The conditional probability of event A given event B is $P(A|B) = P(A \cap B)/P(B)$.
- For events A and B, $P(A) = P(A \cap B) + P(A \cap \bar{B})$. This is called the law of total probability.
- For events A and B, $P(A) = P(A|B)P(B) + P(A \cap \bar{B})P(\bar{B})$.
- Events A and B are independent if $P(A \cap B) = P(A)P(B)$.
- For a continuous random variable X with density function $f_X(x)$ the probability of $X \in I$ is $P(X \in I) = \int_I f_X(x)dx$.
- The expected value or mean of a continuous random variable X with density function $f_X(x)$ is

$$\mu_X = E[X] = \int x f_X(x)dx$$

- The variance of a continuous random variable X with density function $f_X(x)$ and mean μ is

$$\sigma_X^2 = E(X-\mu)^2] = \int (x-\mu)^2 f_X(x)dx$$

- A Gaussian random variable X with mean μ and variance σ^2 has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu/(2\sigma^2))}.$$

- If X is a Gaussian random variable with mean μ and variance σ^2, then

$$P(X \leq b) = \Phi(\frac{b-\mu}{\sigma}) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu/(2\sigma^2))} dx$$
$$P(X > c) = Q(\frac{c-\mu}{\sigma}) = \int_{c}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu/(2\sigma^2))} dx$$

- $Q(x) = 1 - \Phi(x) = \Phi(-x)$.
- Noise in a receiver is typically modeled as white Gaussian noise.
- White Gaussian noise has a flat power spectral density.
- White Gaussian noise into a filter yields Gaussian noise at the output of the filter.
- A randomly delayed linearly modulated signal with pulse shape $x(t)$ is

$$Y(t) = \sum_m b_m x(t - mT - U)$$

  where b_m is a i.i.d. sequence of random data symbols with $E = b_m^2 = \sigma_b^2$ and U is uniformly distributed over the interval $[0, T]$.
- The power spectral density of $Y(t)$ is

$$S_Y(f) = \frac{\sigma^2}{T} |X(f)|^2$$

  where $X(f)$ is the Fourier transform of $x(t)$.
- For rectangular pulses the null-to-null bandwidth is $2/T$ where T is the symbol duration.

Key Relations for Chapter 3:

- The variance of the noise at the output of a filter due to white Gaussian noise is $\eta = \frac{N_0}{2}\int h^2(t)dt = \frac{N_0}{2}\int |H(f)|^2 df$
- If $\eta = \int n(t)x(t)dt$ where $n(t)$ is white Gaussian noise with power spectral density $N_0/2$ then the variance of η is $N_0/2\int x^2(t)dt$.
- The power spectral density of a random data signal is

$$S_Y(f) = \frac{\sigma^2}{T}|X(f)|^2.$$

3.7 Problems

Add problem determing probability that jointly Gaussian random variable is in a region described by two radii and two phases.

1. Let X be a random variable with density function

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

   (a) Let $Y = +\sqrt{X}$ (positive square root). Find the density function, $f_Y(y)$, of Y for all values of y.

   (b) Let $Z = aX + b$ where a is a positive constant. Find the density function, $f_Z(z)$, for all values of Z.

2. Let $X(t)$ be a zero mean wide sense stationary Gaussian random process with autocorrelation function $R_X(\tau) = e^{-|\tau|}$.

   (a) What is the variance of $X(t)$?

   (b) Let $h(t) = \delta(t) - \delta(t-1)$ be the impulse response of a linear system with input $X(t)$ and output $Y(t)$. Find the mean and variance of $Y(t)$.

   (c) Is $Y(t)$ wide sense stationary?

   (d) Find $P\{Y(t) > 2\}$ (Express your answer in terms of the Q function defined on the first page.

3. Let $X(t)$ be a zero mean wide sense stationary Gaussian random process with power spectral density

$$S_X(f) = \begin{cases} 1, & |2\pi f| \leq 5 \\ 0, & |2\pi f| > 5. \end{cases}$$

   Let $H(f) = 5$ for all f be the transfer function of a linear system with input $X(t)$ and output $Y(t)$. Find the mean and variance of the output $Y(t)$.

4. Let $X(t)$ be a zero mean wide sense stationary Gaussian random process with autocorrelation function $R_X(\tau)$ given below. Let $Y(t)$ be the output of a linear time invariant filter with impulse response $h(t)$ given below. Find the variance of $Y(t)$.

$$R_X(\tau) = 5\delta(\tau)$$

   where $h(t) = p_T(t)$ ($p_T(t) = 1$, $0 < t < T$, and 0 elsewhere).

5. Let $X(t)$ be a wide sense stationary random process with autocorrelation function $R_X(\tau)$. Let $Y(t)$ be the random process defined by $Y(t) = X(t) - X(t-s)$ where s is a fixed number.

   (a) Find the autocorrelation of $Y(t)$ in terms of the autocorrelation of $X(t)$.

   (b) What is the mean of $Y(t)$?

   (c) Is $Y(t)$ wide sense stationary?

   (d) What is the power spectral density of $Y(t)$ in terms of the power spectral density of $X(t)$?

6. The wide sense stationary random process $X(t)$ has autocorrelation function $R_X(t) = 2\exp(-t)$.

   (a) What is the value of $E[X(t+1) - X(t-1)]^2$?

   (b) If $X(t)$ is also Gaussian with zero mean find $P(X^2(t) > 1)$ in terms of the standard Gaussian distribution function Φ defined by

$$\Phi(x) = \int_{-\infty}^{x} \exp(-u^2/2)du$$

7. Let X and Y are independent discrete-valued random variables with

$$P(X = j) = \begin{pmatrix} n \\ j \end{pmatrix} p^j (1-p)^n - j \qquad 0 \leq j \leq n, \quad 0 \leq p \leq 1$$

$$P(Y = k) = \begin{pmatrix} m \\ k \end{pmatrix} q^k (1-q)^m - k \qquad 0 \leq k \leq m, \quad 0 \leq q \leq 1.$$

   Find $P(X+Y \geq i)$ for $0 \leq i \leq n+m$.

8. Given X is a zero mean unit variance Gaussian (Normal) random variable. Find the density function of $Y = X^2$. (You must specify the density function for all possible values of the argument).

9. Given N is a Poisson random variable with parameter λ (i.e. $P(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}, \quad n = 0, 1, 2, \ldots$). Find $E(e^{sN})$.

10. A zero mean white Guassian random process $X(t)$ with power spectral density $N_0/2$ is the input to a linear system.

    (a) The impulse response of the linear system is

$$h(t) = A\cos(\omega_c t) p_T(t)$$

    where $p_T(t)$ is 1 for $0 \leq t \leq T$ and is zero elsewhere. Also $\omega_c T = 2\pi n$ (an integer number of cycles in T seconds). Find the mean and variance of the output of the filter.

    (b) If the transfer function of (a different) linear system is triangular,

$$H(f) = \begin{cases} T(fT+1), & -\frac{1}{T} < f \leq 0 \\ T(1-fT), & 0 \leq f < \frac{1}{T} \\ 0, & \text{elsewhere}, \end{cases}$$

    find the mean and variance of the output of the filter.

11. A signal $s(t)$ of duation T consists of 15 consectuive pusles (of duration $T/15$) of amplitude ± 1. The sequence of amplitudes is (-1 -1 -1 -1 +1 -1 +1 -1 -1 +1 +1 -1 +1 +1 +1). Assume that this signal is input to a linear time-invariant system (filter) with impulse response $h(t) = s(T-t)$. Find (plot) the output of the filter.

12. A signal $s(t)$ of duration T consists of 15 consecutive pulses (of duration $T/15$) of amplitude ± 1. The sequence of amplitudes is (-1 -1 -1 -1 +1 -1 +1 -1 -1 +1 +1 -1 +1 +1 +1). Data is transmitted using this signal. The data sequence 011010 is transmitted in 6T seconds by transmitting the signal $s(t)$ for the first T seconds, $-s(t-T)$ during the interval $[T, 2T]$, $-s(t-2T)$ during the interval $[2T, 3T]$ and so on. The total signal transmitted is thus

$$s(t) - s(t-T) - s(t-2T) + s(t-3T) - s(t-4T) + s(t-5T)$$

    Assume that this signal is input to a linear time-invariant system (filter) with impulse response $h(t) = s(T-t)$. Find (plot) the output of the filter.

13. Let $X(t)$ be a wide sense stationary random process with autocorrelation function $R_X(\tau)$. Let $Y(t)$ be the random process defined by $Y(t) = X(t) - X(t-s)$ where s is a fixed number. Find the autocorrelation of $Y(t)$ in terms of the autocorrelation of $X(t)$. Is Y(t) wide sense stationary? What is the mean of Y(t)? What is the power spectral density of $Y(t)$ in terms of the power spectral density of $X(t)$?

14. Consider a bandpass signals $x(t)$ with frequency content $X(f)$ generated from two signals $x_I(t)$ and $x_Q(t)$ with frequency content $X_I(f)$ and $X_Q(f)$ as in the notes. The point of this homework is to show you can recover $x_I(t)$ from $x(t)$.

    (i) Find the frequency content of the signal $y_I(t) = x(t)\sqrt{2}\cos(2\pi f_c t)$. in terms of $X_I(f)$ and $X_Q(f)$.

    (ii) Find the frequency content of $y_I(t) * g_{LP}(t)$ where $g_{LP}(t)$ is an ideal low pass filter.

15. Suppose $n(t)$ is white Gaussian noise with two-sided power spectral density $N_0/2$. That is,

$$\begin{aligned} R_n(\tau) &= E[n(t)n(t+\tau)] = \frac{N_0}{2}\delta(\tau) \\ S_n(f) &= \frac{N_0}{2}. \end{aligned}$$

    Suppose that

$$\begin{aligned} y_I(t) &= \int h(t-\tau)n(\tau)\cos(2\pi f_c \tau)d\tau \\ y_Q(t) &= \int h(t-\tau)n(\tau)\sin(2\pi f_c \tau)d\tau \end{aligned}$$

    where $h(t)$ is an ideal low pass filter that passes frequencies from $-W/2$ to $W/2$ and rejects all other frequencies.

    Determine the autocorrelation and power spectral density of $y_I(t)$ and $y_Q(t)$.

16. Let $X(t)$ be a zero mean wide sense stationary Gaussian random process with power spectral density

$$S_X(f) = \begin{cases} 1, & |f| \le 5 \\ 0, & |f| > 5. \end{cases}$$

    Let $H(f) = 5$ for all f be the transfer function of a linear system with input $X(t)$ and output $Y(t)$. Find the mean and variance of the output $Y(t)$.

Bibliography

[1] J. Soni and R. Goodman, *A mind at play: How Claude Shannon invented the information age*, Simon and Schuster, 2017.

[2] A. Leon-Garcia, *Probability, statistics, and random processes for electrical engineering*, Pearson Education, 2017.

[3] B. Hajek, *Random processes for engineers*, Cambridge University Press, 2015.

[4] M. B. Pursley, *Random processes in linear systems*, Prentice Hall, 2002.

[5] F. Amoroso, "The bandwidth of digital data signals," *IEEE Communications magazine*, vol. 18, no. 6, pp. 13–24, 1980.

[6] E. Agrell, J. Lassing, E. G. Strom, and T. Ottosson, "On the optimality of the binary reflected gray code," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3170–3182, 2004.

[7] J. Lassing, E. G. Strom, E. Agrell, and T. Ottosson, "Computation of the exact bit-error rate of coherent m-ary psk with gray code bit mapping," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1758–1760, 2003.

[8] P. K. Sinha, P. Biswas, and A. Sinha, "Design and implementation of high speed digital multi-programme system based on itu-t j. 83 annexures a and c," in *2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, pp. 1–5. IEEE, 2015.

[9] Y. Okumura, "Field strength and its variability in vhf and uhf land-mobile radio service," *Rev. Electr. Commun. Lab.*, vol. 16, pp. 825–873, 1968.

[10] M. Hata, "Empirical formula for propagation loss in land mobile radio services, ieee transa-ctions on vehicular technology," *Vol. VT-29*, no. 3, , 1980.

[11] H. Hashemi, "The indoor radio propagation channel," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, 1993.

[12] N. Zhang and S. W. Golomb, "Sixty-phase generalized Barker sequences," *IEEE Transactions on Information theory*, vol. 35, no. 4, pp. 911–912, 1989.

[13] M. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication-part i: System analysis," *IEEE Transactions on communications*, vol. 25, no. 8, pp. 795–799, 1977.

[14] *Global Positioning System Standard Positioning Service Signal Specification*, 1995.

[15] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Transactions on information theory*, vol. 18, no. 4, pp. 531–532, 1972.

[16] S. W. Golomb et al., *Shift register sequences*, Aegean Park Press, 1967.

[17] R. J. McEliece, *Finite fields for computer scientists and engineers*, volume 23, Springer Science & Business Media, 2012.

[18] P. Fan and M. Darnell, *Sequence design for communications applications*, volume 1, Research Studies Press, 1996.

[19] D. V. Sarwate and M. B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," *Proceedings of the IEEE*, vol. 68, no. 5, pp. 593–619, 1980.