Background and Communication Channels

Suayb S. Arslan

August 31, 2014

1 INTRODUCTION

In digital communication, the information being transferred is represented in digital form, most commonly as binary digits, or bits. This is in contrast to analog information, which takes on a continuum of values. Information is transferred digitally, but we, humans, consume information or produce it in analog forms. Why digital?

- A desirable property of a typical storage or communication system is to be be independent of the source characteristics, so that a variety of information sources can share the same communication medium and use the same communication system. Such an approach will alleviate the need for reinterpretation of the information every time it is to be transferred from one point in time or space to another.
- Another striking advantage is the prevention of noise accumulation due to regeneration of transferred bits at each recevier in a multi-hop communication scenario.

One of the important concepts of communication theory is the idea of modulation. Modulation refers to the representation of digital information in terms of analog waveforms that can be transmitted over physical channels. These bits are translated into symbols using a bit-to-symbol map, which in this case could be as simple as mapping the bit 0 to the symbol +1, and the bit 1 to the symbol -1. After the bit-to-symbol mapping operation, symbols go through a shaping function, most of the time an LTI system (filter) to output an analog waveform to be communicated over the channel.

Analog impulse response of the filtering employed for modulation are often constrained in the frequency domain. Such constraints arise either from the physical characteristics of the communication medium, or from external factors such as government regulation of spectrum usage. Thus, it is typical to classify channels, and the signals transmitted over them, in terms of the frequency bands they occupy.

Apart from necessary fundamental principles of signals and systems, probability and stochastic processes for understanding the material presented here, modulation is the main introductory concept to the communication theory. However in order to excel into this concept, we need lay out some background about frequency characterizations of signals and certain probability principles.

As a final note, I would like to acknowledge Laurence B. Milstein of University of California, San Diego whose graduate class notes have been quite instrumental in writing this short note. I am sure it is his passion and love for communication theory which inspired many other researchers around the world as well as myself to read, write and research about this engineering path.

1.1 Preliminaries

Let us start with Euler's identity which decomposes a complex exponential into real valued sinusoids,

$$e^{\pm j\theta} = \cos\theta \pm j\sin\theta \tag{1.1}$$

This quantity is a special form of a general complex number expressed as $z = x + jy$ where $x = Re\{z\}$ is the real part and $y = Im\{z\}$ is the imaginary part. The phase of z is defined to be $\tan^{-1}\frac{y}{x}$ and the magnitude of z is $\sqrt{x^2+y^2}$. The conjugate of z is denoted by $z^* = x - jy$. We can see that $e^{j\theta}$ has a unit magnitude and phase θ. Equation (1.1) also leads to following relationships,

$$\cos(\theta) = \frac{e^{j\theta} + e^{-j\theta}}{2}, \quad \sin(\theta) = \frac{e^{j\theta} - e^{-j\theta}}{2j} \tag{1.2}$$

We can also show the following properties to hold for any pair of complex numbers z_1 and z_2.

$$(z_1 \pm z_2)^* = z_1^* \pm z_2^* \tag{1.3}$$
$$(z_1 z_2)^* = z_1^* z_2^* \tag{1.4}$$
$$(z_1/z_2)^* = z_1^*/z_2^* \text{ assuming } z_2 \text{ is nonzero.} \tag{1.5}$$
$$e^{z^*} = (e^z)^* \tag{1.6}$$

Communication theory is all about relative geometries of the signals used, which are predominantly governed by the inner products of signals. For two complex valued $k \times 1$ vectors $\mathbf{s} = (s_1, s_2, \ldots, s_k)$ and $\mathbf{r} = (r_1, r_2, \ldots, r_k)$, the inner product is given by

$$< \mathbf{s}, \mathbf{r} >= \sum_{i=1}^{k} = s_i r_i^* = \mathbf{r}^H \mathbf{s} \tag{1.7}$$

which can also be defined for time dependent functions $s(t)$ and $r(t)$ as follows

$$< s, r >= \int_{-\infty}^{\infty} = s(t) r^*(t) dt. \tag{1.8}$$

Due to linearity of summing and integration for example, we can easily see relationships like $< a_1 s_1 + a_2 s_2, b_1 r_1 + b_2 r_2 >= a_1 b_1^* < s_1, r_1 > +a_2 b_1^* < s_2, r_1 > +a_1 b_2^* < s_1, r_2 > +a_1 b_2^* < s_1, r_2 >$. Another closely related definition is the energy of a signal which is defined to be the inner product of the signal with itself, given by

$$E_s = ||s||^2 =< s, s >= \int_{-\infty}^{\infty} = |s(t)|^2 dt. \tag{1.9}$$

where $||s||$ is the norm of the signal $s(t)$ and is the square root of the energy of the signal. The inner product obeys the following relationship

$$| < s, r > | \leq ||s|| \, ||r||, \tag{1.10}$$

which is also known as a form of Cauchy-Schwartz inequality. To prove this inequality, let us define the vector for some scalar c, $z = cr + s$. Consider

$$||z||^2 = ||cr - s||^2 = |c|^2 ||r||^2 - 2|c| < s, r > +||s||^2 \geq 0 \tag{1.11}$$

Let us reexpress this as a quadratic equation $f(|c|) = |c|^2||r||^2 - 2|c| < s, r > +||s||^2$, and simple quadratic algebra yields if $f(|c|) \geq 0 \Rightarrow f(\frac{<s,r>}{||r||^2}) \geq 0$ since $f(\frac{<s,r>}{||r||^2})$ is the minimum of the function. If we rewrite equation (1.11), we have

$$||z||^2 = -\frac{(< s, r >)^2}{||r||^2} + ||s||^2 \geq 0 \Rightarrow ||s||^2 \; ||r||^2 \geq (< s, r >)^2 \tag{1.12}$$

from which the result follows. It is also easy to see that the equality holds if and only if $s(t)$ is some scalar multiple of $r(t)$ i.e., $s(t) = cr(t)$.

One of the major operations of linear time invariant systems (LTI) is the convolution. The convolution of the two signals $s(t)$ and $r(t)$ is defined to be of the form,

$$(s * r)(t) = s(t) * r(t) = \int_{-\infty}^{+\infty} s(u) r(t - u) du \tag{1.13}$$

which is usually referred as "convolution of $s(t)$ and $r(t)$ at time t". Due to LTI nature, delay or scalar multiplier operations can be performed before or after convolution operation without effecting the eventual outcome. The result of convolution gets more interesting when one of the signals is a delta dirac function i.e.,

$$(s * \delta)(u_0) = \int_{-\infty}^{+\infty} \delta(u - u_0) s(u) du = s(u_0) \tag{1.14}$$

where we used the fact that $\delta(u - u_0) = \delta(u_0 - u)$. This result is due to the "sifting" property of the delta dirac function. Since this results implies that $\delta(t_0) * s(t) = s(t_0)$, it also implies $\delta(t - t_0) * s(t) = s(t - t_0)$ i.e., the convolution of a signal with a shifted version of the delta function results in a shifted version of the signal $s(t)$. The role of the convolution operation is pretty significant in communication system modeling and implementation.

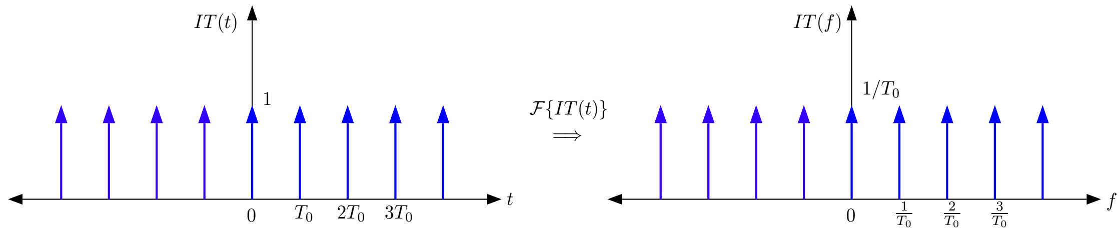

Figure 1.1: Impulse train in time and frequency domains.

INDICATOR AND SINC FUNCTIONS There are few functions that are extremely useful in developing fundamentals of communication theory. One of them is the indicator function which is defined for a set S in the following way,

$$\mathbf{I}_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases} \tag{1.15}$$

Similarly, a sinc function is defined as a function of a sinusoidal function as follows,

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \tag{1.16}$$

where $\text{sinc}(0) = 1$ due to the limiting behavior of the function as x tends to 0.

FOURIER SERIES The frequency content of periodic signals can easily be captured using Fourier series. In other words, every periodic function has a Fourier series representation. A periodic function $p(t)$ with period T_0 can be represented by an infinite series of exponential time functions as follows,

$$p(t) = \sum_{n=-\infty}^{\infty} P[n] e^{j2\pi nt/T_0} \tag{1.17}$$

where the Fourier series coefficients are given by

$$P[n] = \frac{1}{T} \int_{-T/2}^{T/2} p(t) e^{-j2\pi nt/T_0} \tag{1.18}$$

and $f_0 = 1/T_0$ is the fundamental frequency of the periodic signal. All frequencies in the series representation are harmonically related; the ratio of two distinct frequencies is a rational number.

One of the well known periodic functions of interest is the *impulse train* given by

$$IT(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0) \tag{1.19}$$

The Fourier series coefficients for this periodic function can be found as

$$\begin{aligned} IT[n] &= \frac{1}{T} \int_{-T/2}^{T/2} \sum_{n=-\infty}^{\infty} \delta(t - nT_0) e^{-j2\pi nt/T_0} && (1.20) \\ &= \frac{1}{T} \int_{-T/2}^{T/2} \delta(t) e^{-j2\pi nt/T_0} dt = 1/T && (1.21) \end{aligned}$$

Thus, we have two different expressions to represent an impulse train,

$$IT(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_0) = \frac{1}{T} \sum_{n=-\infty}^{\infty} e^{j2\pi nt/T} \tag{1.22}$$

If we take the Fourier transform of both expressions we obtain,

$$IT(f) = \sum_{n=-\infty}^{\infty} e^{-j2\pi nT_0 f} = \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T_0}) \tag{1.23}$$

which implies that Fourier transform of a periodic impulse train in time domain is a periodic impulse train in frequency domain.

Let $s_T(t)$ be a time limited function whose support is symmetric around the origin. The periodic function $s(t)$ can be obtained by convolving $s_T(t)$ by the impulse train $IT(t)$ i.e., $s(t) = s_T(t) * IT(t) \Leftrightarrow S(f) = S_T(f)IT(f)$. Thus, we have

$$S(f) = S_T(f) \times \frac{1}{T_0} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T_0}) = \sum_{n=-\infty}^{\infty} \frac{S_T(n/T_0)}{T_0} \delta(f - \frac{n}{T_0}) \tag{1.24}$$

which implies that the Fourier transform of a periodic signal is bound to be discrete. Note also that as $T_0 \to \infty$, the gap between discrete pulses shrinks and disappears i.e., non-periodic time limited $s(t)$ has a continuous band unlimited frequency response. We will also see later that for an arbitrary periodic continuous time function, the Fourier transform consists of impulses (located at the harmonic frequencies) whose areas are the Fourier series coefficients. This discussion also shows us a way to calculate Fourier series coefficients from the Fourier transform of the periodic function as $T_0 \to \infty$.

FOURIER TRANSFORM Fourier transform of a signal (possible complex valued) $s(t)$ is an invertible transform technique, employed to transform time signal $s(t)$ to the frequency domain representation $S(f) = \mathfrak{F}(s(t))$, given by

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} dt = < s(t), e^{j2\pi ft} > . \tag{1.25}$$

Similarly, the inverse Fourier transform is defined to be of the form,

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft} dt = < S(f), e^{-j2\pi ft} > . \tag{1.26}$$

Based on these definitions, let us calculate the Fourier transform of the time domain square function $I_{[-T/2,T/2]}(t)$,

$$S(f) = \int_{-T/2}^{T/2} e^{-j2\pi ft} dt = \frac{e^{-j\pi fT} - e^{j\pi fT}}{-j2\pi f} = \frac{\sin(\pi fT)}{\pi f} = T\mathrm{sinc}(fT). \tag{1.27}$$

Similarly, $s(t) = \delta(t)$ has the Fourier transform $S(f) = 1$ due to delta function's sifting property. We usually use $\delta(t) \Leftrightarrow 1$ to indicate the Fourier pair for short hand notation for the rest of our discussion. Some of the basic properties of Fourier transform can be itemized as follows,

- if $s(t) \Leftrightarrow S(f)$, then $r(t) = S(t) \Leftrightarrow R(f) = s(-f)$. This is known as the time-frequency duality of the Fourier transform operation.

- if $s(t) \Leftrightarrow S(f)$, then $s^*(t) \Leftrightarrow S^*(-f)$ and $s^*(-t) \Leftrightarrow S^*(f)$. This also implies if $s(t) = s^*(t)$ i.e., $s(t)$ is a real valued signal, then $S(f) = S^*(-f)$ i.e., their Fourier transforms are conjugate symmetric of eachother.

- if $s(t) \Leftrightarrow S(f)$, then $s(t-t_0) \Leftrightarrow S(f)e^{-j2\pi f t_0}$ and $s(t)e^{j2\pi f_0 t} \Leftrightarrow S(f-f_0)$.

- if $s_1(t) \Leftrightarrow S_1(f)$ and $s_2(t) \Leftrightarrow S_2(f)$, then $(s_1 * s_2)(t) \Leftrightarrow S_1(f)S_2(f)$ and $s_1(t)s_2(t) \Leftrightarrow (S_1 * S_2)(f)$.

- if $s(t) \Leftrightarrow S(f)$, then $s(at) \Leftrightarrow \frac{1}{|a|}S(\frac{f}{a})$

Next, Let us present Parseval's theorem which plays a crucial role when one talks about signal energy or power. From above we have $s^*(t) \Leftrightarrow S^*(-f)$, first we observe through change of variables that

$$s^*(t) = \int_{-\infty}^{\infty} S^*(-f)e^{j2\pi f t} df = \int_{-\infty}^{\infty} S^*(f)e^{-j2\pi f t} df \tag{1.28}$$

Let us consider the inner product

$$\begin{aligned}
< s_1, s_2 > &= \int_{-\infty}^{\infty} s_1(t)s_2^*(t)dt && (1.29)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_1(f)e^{j2\pi f t}df \int_{-\infty}^{\infty} S_2^*(g)e^{-j2\pi g t}dgdt && (1.30)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_1(f)S_2^*(g)\int_{-\infty}^{\infty} e^{j2\pi(f-g)t}dtdfdg && (1.31)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_1(f)S_2^*(g)\delta(f-g)dfdg && (1.32)\\
&= \int_{-\infty}^{\infty} S_1(f)S_2^*(f)df && (1.33)\\
&= < S_1, S_2 > . && (1.34)
\end{aligned}$$

If we set $s_1(t) = s_2(t) = s(t)$, we obtain the energy of the signal $s(t)$ (summing of all the instantaneous powers over time) as follows

$$E_s = ||s(t)||^2 = \int_{-\infty}^{\infty} |s(t)|^2 dt = \int_{-\infty}^{\infty} |S(f)|^2 df \tag{1.35}$$

where we assume E_s is finite. This particularly shows the energy conservation principle, after and before the transform operation.

Exercise 1: If we let $s(t) = \sum_{n=1}^{\infty} a_n \sin(\frac{n\pi t}{L})$ for all $0 < t < L$, then we have

$$\frac{2}{L}\int_0^L s^2(t)dt = \sum_{n=1}^{\infty} a_n^2. \tag{1.36}$$

Hint: First show that

$$\int_0^L \sin\left(\frac{m\pi t}{L}\right)\sin\left(\frac{n\pi t}{L}\right)dt = \begin{cases} L/2 & \text{if } n=m \\ 0 & \text{Otherwise.}\end{cases} \quad (1.37)$$

Average power of the signal $s(t)$ is defined to be of the form

$$P_s = \langle |s(t)|^2\rangle = \lim_{L\to\infty}\frac{1}{2L}\int_{-A}^{A}|s(t)|^2 dt \quad (1.38)$$

A bounded signal $s(t)$ can either be an energy signal in which E_s is finite and $P_s = 0$ or a power signal in which E_s is infinite and P_s is finite. Suppose that the signal $s(t)$ passes through a bandpass filter $H(f)$ defined as follows,

$$H(f) = \begin{cases} 1 & \text{if } f_0 - \Delta f/2 < f < f_0 + \Delta f/2 \\ 0 & \text{Otherwise.}\end{cases} \quad (1.39)$$

The instantaneous power or the energy spectral density $E_s(f_0)$ of $s(t)$ at frequency f_0 is defined to be the energy of the signal at the output of the filter divided by the width of the filter Δf as Δf tends to zero. It is easy to see that the signal energy at the output is approximately $|S(f)|^2\Delta f$ and therefore $E_s(f_0) = |S(f)|^2$. Therefore the integral of the energy spectral density gives us the signal energy. Let $s_L(t)$ denote the signal $s(t)$ for $-L \le t \le L$ and zero otherwise, and $S_L(f)$ the corresponding Fourier transform, then

$$S_L(f) = \int_{-\infty}^{\infty} s_L(t)e^{-j2\pi f t}dt = \int_{-L}^{L} s(t)e^{-j2\pi f t}dt \quad (1.40)$$

Power spectral density for the signal $s_L(t)$ is therefore given by

$$\Phi_{S_L S_L}(f) = \lim_{L\to\infty}\frac{1}{2L}|S_L(f)|^2 \quad (1.41)$$

Another popular quantity related to the analysis of signals is the autocorrelation function which measure how closely the signal $s(t)$ approximates delayed versions of itself. For a given delay τ, autocorrelation function of $s(t)$ is given by

$$R_s(\tau) = \int_{-\infty}^{\infty} s(t)s^*(t-\tau)dt \quad (1.42)$$

where we note that at $\tau = 0$, i.e., convolution with zero shifted version of the signal itself, we obtain the inner product of the signal with itself. This is the energy of the signal i.e., $R_s(0) = E_s$. If we look at the expression (1.41) carefully, we can realize that (using change of variables and integral algebra) that

$$\Phi_{S_L S_L}(f) = \int_{-\infty}^{\infty} R_s(\tau)e^{-j2\pi f\tau}d\tau \Longrightarrow R_s(\tau) = \int_{-\infty}^{\infty}\Phi_{S_L S_L}(f)e^{j2\pi f\tau}df \quad (1.43)$$

This implies that we can relate the average energy (power in a unit time) to the power spectral density as follows,

$$E_s = R_s(0) = \int_{-\infty}^{\infty}\Phi_{S_L S_L}(f)df. \quad (1.44)$$

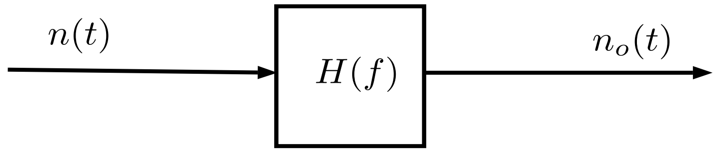

Figure 1.2: Power spectral density of a random process $n(t)$ changes after a linear filter with frequency response $H(f)$.

Finally, we consider a stochastic process going through a linear filter. In a communication context, that random process is usually the noise. This is illustrated in Fig. 1.2. Without loss of generality, we assume the filter system is bounded response and the input random process is stationary and has a finite mean. Thus, we have

$$\begin{aligned}\mathbb{E}[n_0(t)] &= \int_{-\infty}^{-\infty} h(\tau)\mathbb{E}[n(t-\tau)]d\tau &(1.45)\\ &= \mathbb{E}[n]\int_{-\infty}^{-\infty} h(\tau)d\tau = \mathbb{E}[n]H(0). &(1.46)\end{aligned}$$

Next, let us look at the autocorrelation function of the output process,

$$\begin{aligned}R_{n_0}(t_1,t_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)\mathbb{E}[n(t_1-\tau_1)n(t_2-\tau_2)]d\tau_1 d\tau_2 &(1.47)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_n(t_1-t_2-\tau_1+\tau_2)d\tau_1 d\tau_2 &(1.48)\end{aligned}$$

which shows that the output process is wide-sense stationary. Let $\tau = t_1 - t_2$, we have

$$\begin{aligned}R_{n_0}(t_1,t_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_n(\tau-\tau_1+\tau_2)d\tau_1 d\tau_2 &(1.49)\\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} h(\tau_1)R_n(\tau+\tau_2-\tau_1)d\tau_1\right)h(\tau_2)d\tau_2 &(1.50)\end{aligned}$$

which shows that

$$R_{n_0}(\tau) = R_n(\tau) * h(t) * h(-t). \tag{1.51}$$

Taking the Fourier transform of both sides yield

$$\Phi_{N_0N_0}(f) = \Phi_{NN}(f)H(f)H^*(f) = \Phi_{NN}(f)|H(f)|^2. \tag{1.52}$$

NYQUIST'S SAMPLING THEOREM In a nutshell, Nyquist's sampling theorem provides a prescription of an upper bound for the sampling interval required to avoid aliasing. Consider a signal $s(t)$ band limited to $[-B/2, B/2]$. Sampling with period T_0 is nothing but the multiplication of $s(t)$ with $IT(t)$. In frequency domain this corresponds to

$$S(f) * IT(f) = \frac{1}{T_0}\sum_{-\infty}^{\infty} S(f - \frac{n}{T_0}) \tag{1.53}$$

Since $s(t)$ is band limited to $[-B/2, B/2]$, in order not to cause aliasing, $1/B \geq T_0$. This implies the maximum sampling interval is $1/B$ i.e., we sample at a rate of B, which is also known as critical frequency. In more formal terms, we have the following theorem if we assume the band limited signal is critically sampled.

Theorem 1: For any signal $s(t)$ band limited to $[-B/2, B/2]$ can be described completely by its samples $\{s(n/B)\}$ at rate B. Furthermore, $s(t)$ can be recovered from its samples using the following interpolation formula:

$$s(t) = \sum_{n=-\infty}^{\infty} s\left(\frac{n}{B}\right) \frac{\sin(2\pi \frac{B}{2}(t - \frac{n}{B}))}{2\pi \frac{B}{2}(t - \frac{n}{B})} = \sum_{n=-\infty}^{\infty} s\left(\frac{n}{B}\right) \mathrm{sinc}(Bt - n) \tag{1.54}$$

BASEBAND AND PASSBAND SIGNALS A signal $s(t)$ is considered to be *baseband* if

$$S(f) = 0, \quad |f| > W \tag{1.55}$$

for some $W > 0$. Similarly, a signal $s(t)$ is said to be *passband* if

$$S(f) = 0, \quad |f \pm f_c| > W \tag{1.56}$$

for some $f_c > W > 0$. These definitions for signals is also applicable to Linear Time Invariant (LTI) system transfer functions as well. For some class of signals, although they have infinite bandwidth their most of the energy is contained in a finite frequency band. Using a slight abuse of notation, previous descriptions can still be used with $S(f) \approx 0$. The following exercise demonstrates a case in which the bandwidth is defined as the size of an appropriately chosen interval where most of the energy of the signal lies.

Exercise 1: Let $s(t) = I_{0,T}(t)$ be a rectangle shape signal for $0 < t < T$, zero otherwise. Let $S(f)$ be the Fourier transform of $s(t)$ and $0 < a \leq 1$ be the fraction of energy contained in the band $[-B, B]$. Show that a satisfies the following relationship

$$\int_0^B T\mathrm{sinc}^2(fT)df = a/2 \tag{1.57}$$

and for $T = 1$, B turns out to be 10.2 for $a = 0.99$ i.e., 99% and 0.85 for $a = 0.9$ i.e., 90% energy containment.

2 COMPLEX BASEBAND REPRESENTATION

2.1 BASEBAND SIGNALS

Communication channels as well as signals used to transmit information are almost always of passband nature. In otherwords, user narrow band signals are often transmitted using some type of carrier modulation. However, passband signal processing is a challenging task because it requires much higher sampling rates to discritize/digitize the analog information for processing. Therefore, it maybe profoundly practical to express real-valued passband waveforms as a complex-valued baseband signal. This transformation allows many modern communication systems to implement sophisticated signal processing algorithms digitally on complex baseband representations of real-valued pass band signals. Such a transformed representation enables a modular transceiver design as well. In otherwords, all the processing and algorithms are developed for baseband signals, independent of the physical frequency band used for the actual communication link.

Let us assume a real-valued bandpass signal $s(t)$ has a Fourier transform $S(f)$ which consists of a positive spectrum component $S_+(f)$ and a negative spectrum component $S_-(f)$. In other words, $S_+(f)$ the spectrum of the resulting analytic signal $s_+(t)$ is defined as

$$S_+(f) = \mathfrak{F}\{s_+(t)\} = 2u(f)S(f) \tag{2.1}$$

where $u(x)$[1] is the unit step function expressed as

$$u(x) = \begin{cases} 1 & \text{if } x > 0 \\ 1/2 & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.2}$$

Before finding a closed for expression for $s_+(t)$, let us observe the following equivalence between the unit step function and sign function,

$$u(x) = \frac{1}{2} + \frac{1}{2}\mathrm{sgn}(t) \tag{2.3}$$

which implies that the inverse Fourier transform of $2u(f)$ can be found using linearity and duality as

$$\mathfrak{F}\{2u(f)\} = \delta(f) + \left(\frac{j}{\pi f}\right) \tag{2.4}$$

where $\mathfrak{F}\{\mathrm{sgn}(t)\} = 1/j\pi f$ and $\delta(f) = \delta(-f)$.

Using inverse Fourier transform, the analytic signal can be found as follows,

$$\begin{aligned} s_+(t) = F^{-1}\{S_+(f)\} &= F^{-1}\{2u(f)S(f)\} && (2.5)\\ &= F^{-1}\{2u(f)\} * F^{-1}\{S(f)\} && (2.6)\\ &= \left(\delta(f) + \left(\frac{j}{\pi f}\right)\right) * s(t) && (2.7)\\ &= s(t) + j\frac{1}{\pi t} * s(t) && (2.8)\\ &= s(t) + js^H(t) && (2.9) \end{aligned}$$

where $s^H(t) = s(t) * 1/\pi t$ is called the Hilbert transform of $s(t)$. The complex baseband representation of $s(t)$, denoted as $s_b(t)$, can be obtained by left shifting and scaling the analytic signal $s_+(t)$. In otherwords, the spectrum $S_b(f) = \mathfrak{F}\{s_b(f)\}$ can be derived from $S_+(f)$ using

$$S_b(f) = \frac{1}{\sqrt{2}}S_+(f + f_c) \tag{2.10}$$

[1] Here we use a dummy variable x so that the same function can be used for both in frequency and in time.

where f_c is the carrier frequency assuming that $s(t)$ is obtained through modulating a baseband signal using a carrier. The complex baseband representation $s_b(t)$ can be found by taking inverse Fourier transform of $S_b(f)$,

$$s_b(t) = \mathfrak{F}^{-1}\{S_b(f)\} = \frac{1}{\sqrt{2}}\mathfrak{F}^{-1}\{S_+(f+f_c)\} = \frac{1}{\sqrt{2}}s_+(t)e^{-j2\pi f_c t} = \frac{1}{\sqrt{2}}\left(s(t)+js^H(t)\right)e^{-j2\pi f_c} \tag{2.11}$$

This equation can be rewritten as,

$$s(t)+js^H(t) = \sqrt{2}s_b(t)e^{2j\pi f_c t} \tag{2.12}$$

to see that $s(t) = \sqrt{2}\text{Re}\{s_b(t)e^{2j\pi f_c t}\}$. Since $s_b(t)$ is complex valued in general, it can be expressed as in terms of a real and an imaginary components,

$$s_b(t) = r(t) + ji(t) \tag{2.13}$$

where $r(t) = \text{Re}\{s_b(t)\}$ and $i(t) = \text{Im}\{s_b(t)\}$. This leads to a different definition of the passband signal $s(t)$. Thus, any passband signal $s(t)$ can be rewritten as in the following form,

$$s(t) = \sqrt{2}r(t)\cos(2\pi f_c t) - \sqrt{2}i(t)\sin(2\pi f_c t) \tag{2.14}$$

where $r(t)$ and $i(t)$ are also sometimes referred as inphase and quadrature signal components. The complex baseband representation $s_b(t)$ can also be expressed in polar form by defining the envelope signal $e(t)$ and the phase signal $\theta(t)$ as follows,

$$e(t) = |s_b(t)| = \sqrt{r^2(t)+i^2(t)}, \quad \theta(t) = \arctan\frac{i(t)}{r(t)}. \tag{2.15}$$

Using $s_b(t) = e(t)e^{j\theta(t)}$ and $s(t) = \sqrt{2}\text{Re}\{s_b(t)e^{2j\pi f_c t}\}$, we obtain yet another expression for the passband signal $s(t)$ as follows,

$$s(t) = \sqrt{2}e(t)\cos(2\pi f_c t + \theta(t)) \tag{2.16}$$

Exercise 3: Suppose that $r(t)$ and $i(t)$ components are real valued low pass signals. The corresponding passband signal $s(t)$ is obtained through an operation called modulation as follows,

$$s(t) = \sqrt{2}r(t)\cos(2\pi f_c t) - \sqrt{2}i(t)\sin(2\pi f_c t) \tag{2.17}$$

The objective of the demodulator is to extract $r(t)$ and $i(t)$. Show that $r(t)$ and $i(t)$ has the following relations to $s(t)$.

$$r(t) = \frac{1}{\sqrt{2}}\left[s(t)\cos(2\pi f_c t) + s^H(t)\sin(2\pi f_c t)\right] \tag{2.18}$$

$$i(t) = \frac{1}{\sqrt{2}}\left[s^H(t)\cos(2\pi f_c t) - s(t)\sin(2\pi f_c t)\right] \tag{2.19}$$

Let us also calculate the energy of the baseband representation of $s(t)$. We start with the definition of the energy of $s(t)$,

$$E_s = \int_{-\infty}^{\infty} s^2(t)dt = \int_{-\infty}^{\infty} |S(f)|^2 df \tag{2.20}$$

and recognizing that

$$
\begin{aligned}
S(f) &= \sqrt{2}\int_{-\infty}^{\infty} \mathrm{Re}\{s_b(t)e^{2j\pi f_c t}\}e^{-j2\pi f t}dt &(2.21)\\
&= \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}(s_b(t)e^{2\pi f t} + s_b^*(t)e^{-j2\pi f t})dt &(2.22)\\
&= \frac{1}{\sqrt{2}}\left[S_b(f-f_c) + S_b^*(-f-f_c)\right] &(2.23)
\end{aligned}
$$

Plugging equation 2.23 in place of $s(t)$, the energy expression becomes

$$
E_s = \frac{1}{2}\int_{-\infty}^{\infty}\left|S_b(f-f_c) + S_b^*(-f-f_c)\right|^2 df = \int_{-\infty}^{\infty}|S_b(f)|^2 df = \int_{-\infty}^{\infty}|s_b(t)|^2 dt \qquad (2.24)
$$

where frequency shift and conjugation operation does not change the energy content of the signal $s_b(t)$ and cross terms disappear because we assume that bandwidth of $S_b(f)$ is usually much smaller than carrier frequency f_c i.e., $S_b(f-f_c)S_b(-f-f_c) = 0$ and $S_b^*(f-f_c)S_b^*(-f-f_c) = 0$. This shows that the energy of the complex baseband representation is identical to the corresponding passband signal $s(t)$. This derivation also explains the $\sqrt{2}$ term we have been using in this section.

2.1.1 Baseband Channels

The complex baseband representation of channels or linear systems are pretty similar to that of signals except few minor but important changes. A bandpass channel with the impulse response $h(t)$ and corresponding frequency response $H(f)$ has one sided frequency response $H_+(f)$ defined as,

$$
H_+(f) = 2u(f)H(f) \qquad (2.25)
$$

where the time domain representation $h_+(t)$ is simply the inverse Fourier transform of $H_+(f)$. However, the definition of the complex baseband representation is slightly different and given by

$$
H_b(f) = \frac{1}{2}H_+(f+f_c) \qquad (2.26)
$$

where the constant term is $1/2$ instead of $1/\sqrt{2}$. Using these definitions and assuming that $h(t)$ is real valued impulse response i.e., $H(f) = H^*(-f)$, we can obtain

$$
H(f) = H_b(f-f_c) + H_b^*(-f-f_c) \qquad (2.27)
$$

from which we find by taking the inverse Fourier transform

$$
h(t) = h_b(t)e^{j2\pi f_c t} + h_b^*(t)e^{-j2\pi f_c t} = 2\mathrm{Re}\{h_b(t)e^{2j\pi f_c t}\} \qquad (2.28)
$$

2.1.2 Transmission of a passband signal through a passband channel

One of the fundamental questions of this section is whether linear filtering in the passband has any equivalence in linear filtering in the complex baseband. It turns out there is such an equivalence and it is perfectly fine to do the linear filtering operation in the complex baseband. To show this, let us assume $r(t)$ to be the output of linear filter $s(t)$ with $h(t)$. We have,

$$\begin{aligned} R(f) &= S(f)H(f) && (2.29)\\ &= \frac{1}{\sqrt{2}}\left(S_b(f-f_c)+S_b^*(-f-f_c)\right)\left(H_b(f-f_c)+H_b^*(-f-f_c)\right) && (2.30)\\ &= \frac{1}{\sqrt{2}}\left(S_b(f-f_c)H_b(f-f_c)+S_b^*(-f-f_c)H_b^*(-f-f_c)\right) && (2.31)\\ &= \frac{1}{\sqrt{2}}\left(R_b(f-f_c)+R_b^*(-f-f_c)\right) && (2.32)\end{aligned}$$

which means that all the linear filtering operations can be done in the complex baseband and yet the actual result for the passband can be obtained through simple transformation.

2.2 Random Processes and Noise

Random processes usually serve very useful for characterization of noise, interference, and the input-output relationship of certain class of communication channels. A random process, also known as stochastic process, is a family of random variables,, indexed by a parameter t from an indexing set $\mathscr{T}$. For each experiment outcome $\omega \in \Omega$, we assign a measure X that depends on t

$$X(t,\omega) \quad t \in \mathscr{T}, \omega \in \Omega \qquad (2.33)$$

where t usually represents time that can be either discrete or continuous. For example, if t is continuous we often denote the random process $X(t)$ and at each time t,$X(t)$ becomes a random variable. In otherwords, it is clear that for a fixed t, $X(t,\omega)$ is a random variable and for a fixed ω, $X(t,\omega)$ is a realization. The statistical properties of a random process is completely characterized by the collection of joint cumulative distribution function of the set of random variables

$$\{X(t_1), X(t_2), \ldots, X(t_n)\} \qquad (2.34)$$

for any set of time samples $\{t_1, t_2, \ldots, t_n\}$ and any n. Sometimes, a complete specification of a random process may not possible. Instead moments of time samples are used to partially specify it. One of the important measures is autocorrelation function, defined as the correlation between the two time samples $X(t_1)$ and $X(t_2)$ i.e., $R_X(t_1,t_2) = \mathbb{E}[X(t_1)X^*(t_2)]$. For real random processes, $R_X(t_1,t_2)$ is symmetric. Similarly, autocovariance function can be defined as follows,

$$\begin{aligned} C_x(t_1,t_2) &= \mathbb{E}[(X(t_1)-\mathbb{E}[X(t_1)])(X(t_2)-\mathbb{E}[X(t_2)])^*] && (2.35)\\ &= R_X(t_1,t_2)-\mathbb{E}[X(t_1)]\mathbb{E}[X^*(t_2)] && (2.36)\end{aligned}$$

In many random processes that we will consider in communication systems, the statistics do not change over time. A random process is called mth order stationary if joint cdf of any m time samples is independent of the time origin i.e.,

$$[X(t_1), X(t_2), \dots, X(t_m)] \sim [X(t_1+\tau), X(t_2+\tau), \dots, X(t_m+\tau)] \tag{2.37}$$

If we impose this condition for any $m > 0$, the process is called strictly stationary, which is usually very strong requirement. A random process $X(t)$ is called wide sense stationary (WSS) if the following properties hold,

$$\mathbb{E}[X(t)] = m \;\; \forall t \tag{2.38}$$

$$R_X(t_1, t_2) = R_X(t_2 - t_1) \;\; \forall t_1, t_2 \tag{2.39}$$

from which we note that the conditions of WSS is more relaxed than that of strict stationarity. Note that $R_X(0) = \mathbb{E}[X(t)^2]$ is the power of the process and is positive. Furthermore, $R_X(\tau)$ is an even function and attains its maximum at $\tau = 0$. Finally, if a Gaussian process is WSS, it is also strictly stationary. Thus, a Gaussian process can be specified by only knowing the common mean m and the covariance $C_X(\tau)$. The random process $X(t)$ is cyclostationary with repsect to time interval T if it is statistically indistinguishable

Our final note about random processes is the concept of ergodicity. A quantity that can be obtained through measurements is the ensemble average. The estimate of the mean value of $X(t)$ is given by

$$\widehat{m}_X(t) = \frac{1}{N}\sum_{i=1}^{N} X(t, w_i) \tag{2.40}$$

where ω_i is the outcome of the ith random experiment. If the process is strictly stationary, the mean value should not change over time. Therefore, the natural question would be to ask whether the mean can be estimated based on realizations over time. In other words, if we define the time average as follows,

$$\mathbb{E}_T[X(t)] = \frac{1}{2T}\int_{-T}^{T} X(t,\omega)dt \tag{2.41}$$

our question would be "when does the time average should converge to the ensemble average?". If a random process is ergodic, its time and ensemble averages converge.

Exercise 4: Suppose that $X_n = X(t,\omega)$ is a discrete stationary random process where X_ns can be interpreted as a sequence of i.i.d. random variables with mean $\mathbb{E}[X_n] = m$. Show that $X(t,\omega)$ is ergodic.

Example 1: Let us consider a random modulated information signal

$$s(t) = \sum_{-\infty}^{\infty} a[n]p(t - nT) \tag{2.42}$$

where $a[n]$ is the random information sequence (which may be complex valued) and $p(t)$ is the signal shaping pulse and nonzero only for $[0, T]$. Let us show how to compute the PSD of $s(t)$. The first operation is to window the signal for an interval of $[0, NT]$ to obtain,

$$s_N(t) = \sum_{n=0}^{N-1} a[n]p(t - nT) \tag{2.43}$$

Next, let us take the Fourier transform of the windowed signal,

$$S_N(f) = \mathfrak{F}\{s_N(t)\} = \sum_{n=0}^{N-1} a[n]P(f)e^{-j2\pi fnT} = P(f)\sum_{n=0}^{N-1} a[n]e^{-j2\pi fnT} \tag{2.44}$$

The estimate of the power spectral density is given by

$$\frac{|S_N(f)|^2}{NT} = \frac{|P(f)|^2|\sum_{n=0}^{N-1} a[n]e^{-j2\pi fnT}|^2}{NT} \tag{2.45}$$

$$= \frac{|P(f)|^2\sum_{n=0}^{N-1}\sum_{n=0}^{N-1} a[n]a^*[m]e^{-j2\pi f(n-m)T}}{NT} \tag{2.46}$$

If we take the limit $N \to \infty$, we obtain

$$\lim_{N\to\infty}\frac{|S_N(f)|^2}{NT} = \frac{|P(f)|^2}{T}\lim_{N\to\infty}\frac{\sum_{n=0}^{N-1}\sum_{n=0}^{N-1} a[n]a^*[m]e^{-j2\pi f(n-m)T}}{N} \tag{2.47}$$

Furthermore, if we assume the time average of $a[n]a^*[m]$ is zero i.e., source information samples are completely uncorrelated, we shall simplify the equation and rewrite

$$\lim_{N\to\infty}\frac{|S_N(f)|^2}{NT} = \frac{|P(f)|^2}{T}\lim_{N\to\infty}\frac{\sum_{n=0}^{N-1}|a[n]|^2}{N} = \frac{|P(f)|^2}{T}\sigma_a^2 \tag{2.48}$$

where σ_a^2 is the time average of $|a[n]|^2$. We observe that the PSD of the modulated source symbols scales as the magnitude squared of the spectrum of the signal shaping (modulating) function. If we delay $s(t)$ by an integer multiple of the pulse length kT, we also observe that

$$s(t-kT) = \sum_{-\infty}^{\infty} a[n]p(t-(n+k)T) = \sum_{-\infty}^{\infty} a[n-k]p(t-nT) \tag{2.49}$$

from which we can deduce that if the source information sequence is stationary, $s(t)$ and its delay version are statistically indistinguishable. However, this is only true if the delay happens an integer multiple of T. Thus, a modulated waveform with a stationary input sequence $\{a[n]\}$ is cyclostationary process.

A random process is baseband/passband if its PSD is baseband/passband. Similar to deterministic signals, complex envelope can be defined for passband random processes. Let $X(t)$ be a passband random process with the complex envelope $X_b(t)$. Their spectral relationship can be shown to be of the form,

$$\Phi_X(f) = \frac{1}{2}(\Phi_{X_b}(f-f_c) + \Phi^*_{X_b}(-f-f_c)) \tag{2.50}$$

Let us explore more about passband and baseband relationship through one of the most popular random processes taking place in communication systems: noise process. We primarily assume a narrow band WSS noise process $n(t)$ with zero mean and PSD $\Phi_{NN}(f) \neq 0$ only if $f_c - B/2 \leq |f| \leq f_c + B/2$. Equivalent complex baseband signal $n_b(t) = x(t) + jy(t)$ where

$x(t)$ and $y(t)$ are real valued baseband noise components. Passband noise can be expressed as

$$n(t) = \sqrt{2}\text{Re}\{n_b(t)e^{j2\pi f_c t}\} \tag{2.51}$$
$$= \sqrt{2}x(t)\cos\{2\pi f_c t\} - \sqrt{2}y(t)\sin\{2\pi f_c t\} \tag{2.52}$$

Let us look at the autocorrelation function of $n(t)$. We define $\phi_{XX}(\tau) = \mathbb{E}[x(t)x(t+\tau)]$, $\phi_{YY}(\tau) = \mathbb{E}[y(t)y(t+\tau)]$ and $\phi_{XY}(\tau) = \mathbb{E}[x(t)y(t+\tau)]$. Using these definitions, we have

$$\phi_{NN}(t,t+\tau) = \mathbb{E}[n(t)n^*(t+\tau)] \tag{2.53}$$
$$= (\phi_{XX}(\tau)+\phi_{YY}(\tau))\cos(2\pi f_c\tau) \tag{2.54}$$
$$+(\phi_{XX}(\tau)-\phi_{YY}(\tau))\cos(2\pi f_c(2t+\tau)) \tag{2.55}$$
$$-(\phi_{YX}(\tau)-\phi_{XY}(\tau))\sin(2\pi f_c\tau) \tag{2.56}$$
$$-(\phi_{YX}(\tau)+\phi_{XY}(\tau))\sin(2\pi f_c(2t+\tau)) \tag{2.57}$$

where we used the appropriate trigonometric identities. Since $n(t)$ is narrowband WSS process, its autocorrelation function should only depend on τ. Thus, this implies $\phi_{XX}(\tau) = \phi_{YY}(\tau)$ and $\phi_{YX}(\tau) = -\phi_{XY}(\tau)$. Using these results, we finally arrive at

$$\phi_{NN}(\tau) = 2\phi_{XX}(\tau)\cos(2\pi f_c\tau) - 2\phi_{YX}(\tau)\sin(2\pi f_c\tau) \tag{2.58}$$

If we also compute the autocorrelation function of $n_b(t)$, we obtain

$$\phi_{N_bN_b}(\tau) = \mathbb{E}[n_b(t)n_b^*(t+\tau)] = \phi_{XX}(\tau)+\phi_{YY}(\tau) - j\phi_{XY}(\tau) + j\phi_{YX}(\tau) \tag{2.59}$$
$$= 2\phi_{XX}(\tau)+2j\phi_{YX}(\tau) \tag{2.60}$$

implying that we have $\phi_{NN}(\tau) = \text{Re}\{\phi_{N_bN_b}(\tau)e^{j2\pi f_c\tau}\}$. Note here that if $x(t)$ and $y(t)$ are uncorrelated, $\phi_{XY}(\tau) = 0$ for all τ, i.e., $\phi_{N_bN_b}(\tau) = 2\phi_{XX}(\tau)$ is real valued. The PSD of $n(t)$ can therefore be computed in terms of PSD of $n_b(t)$ as follows

$$\Phi_{NN}(f) = \int_{-\infty}^{+\infty}\text{Re}\{\phi_{N_bN_b}(\tau)e^{j2\pi f_c\tau}\}e^{-j2\pi f\tau}d\tau \tag{2.61}$$
$$= \frac{1}{2}\int_{-\infty}^{+\infty}(\phi_{N_bN_b}(\tau)e^{j2\pi f_c\tau}+\phi^*_{N_bN_b}(\tau)e^{-j2\pi f_c\tau})e^{-j2\pi f\tau}d\tau \tag{2.62}$$
$$= \frac{1}{2}\left(\Phi_{N_bN_b}(f-f_c)+\Phi_{N_bN_b}(-f-f_c)\right) \tag{2.63}$$

In addition to the relationship $\phi_{YX}(\tau) = -\phi_{XY}(\tau)$ due to stationarity condition, we also observe that

$$\phi_{YX}(\tau) = \mathbb{E}[y(t)x(t+\tau)] = \mathbb{E}[y(t'-\tau)x(t')] = \phi_{XY}(-\tau) \tag{2.64}$$

together which implies

$$\phi_{XY}(\tau) = -\phi_{XY}(-\tau) \tag{2.65}$$

and that $\phi_{XY}(0) = 0$, $\phi_{XY}(\tau)$ is an odd function of τ. Now using the equation (2.63) at $f_c = 0$, we have

$$\mathfrak{F}\{\text{Re}\{\phi_{N_bN_b}(\tau)\}\} = \frac{1}{2}\left(\Phi_{N_bN_b}(f)+\Phi_{N_bN_b}(-f)\right) \tag{2.66}$$

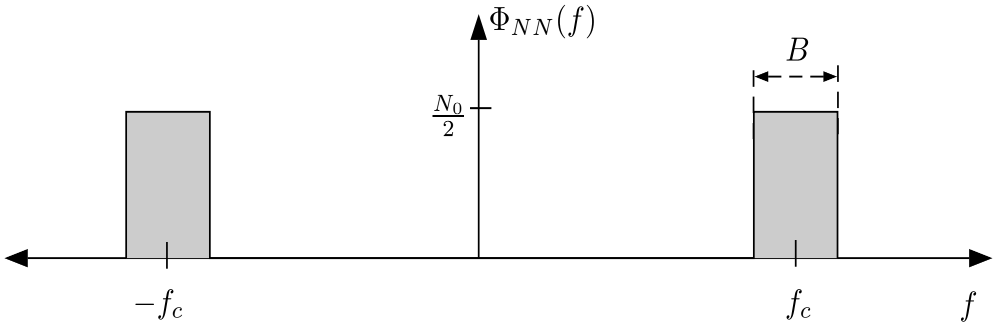

Figure 2.1: Power spectral density of a narrow band WSS white noise process $n(t)$.

Furthermore if we assume $x(t)$ and $y(t)$ are uncorrelated, $\phi_{N_bN_b}(\tau)$ is real and hence

$$\Phi_{N_bN_b}(f) = \frac{1}{2}\left(\Phi_{N_bN_b}(f) + \Phi_{N_bN_b}(-f)\right) \Rightarrow \Phi_{N_bN_b}(f) = \Phi_{N_bN_b}(-f) \tag{2.67}$$

that the complex baseband representation of $n(t)$ is symmetric around $f = 0$. The WSS noise process is called "white" if in the frequency of interest i.e., $f_c - B/2 \leq |f| \leq f_c + B/2$ the PSD is relatively flat and equals $N_0/2$. The PSD of white noise $n(t)$ is illustrated in Fig. 2.1.

Exercise 5: Show that the complex baseband representation of the passband white noise has PSD $\Phi_{N_bN_b}(f) = N_0$ for $|f| \leq B/2$ and zero otherwise and an autocorrelation function $\phi_{N_bN_b}(\tau) = N_0\frac{\sin(\pi B\tau)}{\pi\tau}$.

If we let $B \to \infty$, we can show that $\Phi_{N_bN_b}(\tau) \to N_0$ and $\phi_{N_bN_b}(\tau) \to N_0\delta(\tau)$. The noise process with a flat spectrum for all frequencies is called *white noise*. In short, we have the following sequence of relations,

$$\Phi_{N_bN_b}(f) \text{ is symmetric around } f = 0 \Rightarrow \phi_{XY}(\tau) = 0, \forall\tau \Rightarrow \phi_{N_bN_b} = 2\phi_{XX} = 2\phi_{YY} \tag{2.68}$$

which implies that autocorrelation function of $n_b(t)$ is real valued, $x(t)$ and $y(t)$ are uncorrelated and therefore $\phi_{XX} = \phi_{XX} = \frac{N_0}{2}\delta(\tau)$.

As we have already discussed, the complex baseband representation of an overall communication system is quite useful for both simulation and analysis of passband communication systems. This typical scenario is illustrated in Fig. 2.2, now including the WSS white noise process.

3 Signal Space Representations

Signals can be represented as arrays of coefficients or vectors over a set of appropriate basis. In fact, it is usually necessary and convenient to represent signals as sums of orthogonal/orthonormal signals. Let us assume we have N orthonormal functions $\{f_i(t), i = 1, 2, \ldots, N\}$ i.e.,

$$< f_n(t), f_m(T) >= \int_{-\infty}^{\infty} f_n(t)f_m^*(t) = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases} \tag{3.1}$$

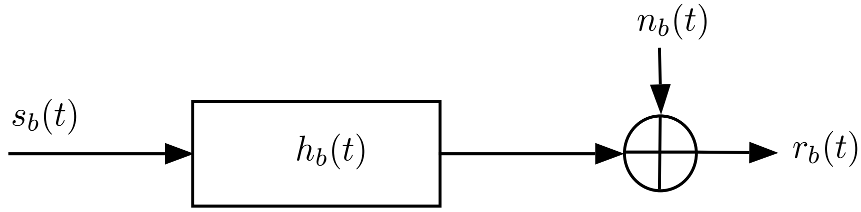

Figure 2.2: Equivalent baseband form of a typical communication channel.

The signal $s(t)$ can be represented by a linear combination of these functions expressed as

$$s(t) = \sum_{i=1}^{N} s_i f_i(t) \tag{3.2}$$

with coefficients $\{s_i, i = 1, 2, \dots, N\}$. Of course this formation assumes the signal $s(t)$ to be contained in the signal space spanned by $f_i(t)$s. Otherwise, we can only find an approximation by choosing the appropriate set $\{s_i\}_{i=1}^{N}$. Through straighforward algebra, it can be shown that the optimal s_is (optimal in mean square sense) turn out to be of the form

$$s_i =< s(t), f_i(t) >= \int_{-\infty}^{\infty} s(t) f_i^*(t) dt, \;\; i = 1, 2, \dots, N \tag{3.3}$$

which is nothing but the projection of $s(t)$ onto each and every orthonormal basis function. In general, we are provided with the set of signal waveforms instead of the orthonormal basis functions. Thus, we should be able to go from given signal waveforms to basis functions. In turns out that there is a nice workaround to that called *Gram-Schmidt* (GM) orthogonalization procedure of Linear algebra. Let us assume we are given M signal waveforms $\{s_i(t), i = 1, 2, \dots, M\}$ with the corresponding energies $\{E_i, i = 1, 2, \dots, M\}$. Outline of the orthogonalization procedure is as follows.

- $f_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$
- Let $f_2' = s_2(t) - < s_2(t), f_1(t) > f_1(t)$, $f_2 = \frac{f_2'}{\sqrt{E_2}}$
- ...
- Let $f_k' = s_k(t) - \sum_{i=1}^{k-1} < s_k(t), f_i(t) > f_i(t)$, $f_k = \frac{f_k'}{\sqrt{E_k}}$
- The procedure ceases after all $s_i(t)$s are processed.

We note that number of non-zero orthonormal basis functions is $\leq M$ i.e., some of the $f_i(t)$ could simply be zero if they depend on the $f_1(t), \dots, f_{i-1}(t)$. Also, these basis functions are not unique and can change depending on the order of processing signals. Suppose that GM process results in N orthonormal basis functions, then each signal $s_k(t)$ can be expressed in terms of these basis functions as follows,

$$s_k(t) = \sum_{i=1}^{N} s_{ki} f_i(t), \qquad k = 1, 2, \dots, M \tag{3.4}$$

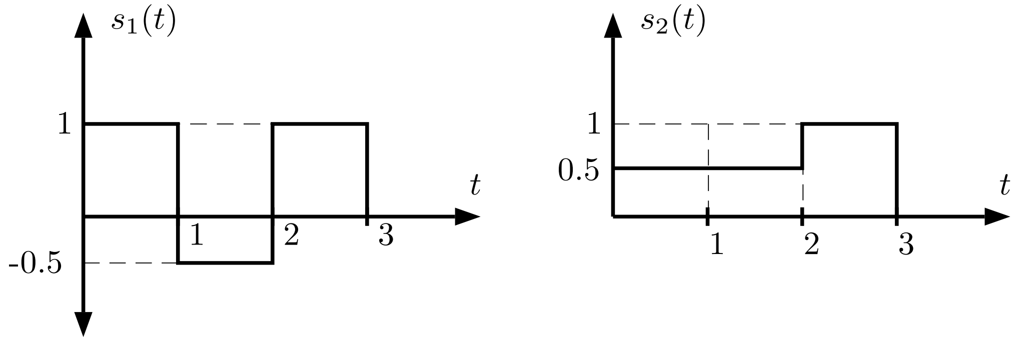

Figure 3.1: Given two signals to show GM procedure.

with coefficients

$$s_{ki} = < s_k(t), f_i(t) > = \int_{-\infty}^{\infty} s_k(t) f_i^*(t) dt, \quad i = 1, 2, \ldots, N \tag{3.5}$$

Since orthonormal basis functions have unit energy, we have two alternative ways to compute signal energies given by

$$E_{s_k} = \int_{-\infty}^{\infty} |s_k(t)|^2 dt = \sum_{i=1}^{N} |s_{ki}|^2 \tag{3.6}$$

Example 2: Let us consider two signals as shown in Fig. 3.1. Note that the energy of s_1 is

$$E_{s_1} = \int_{-\infty}^{\infty} |s_1(t)|^2 dt = 2.25 \tag{3.7}$$

The first step of GM gives us the first orthonormal basis function as follows

$$f_1(t) = s_1(t)/\sqrt{2.25} \tag{3.8}$$

Next, the inner product of $s_2(t)$ with $f_1(t)$ can be calculated,

$$< s_2(t), f_1(t) > = \int_{-\infty}^{\infty} s_2(t) f_1^*(t) dt = \frac{1}{\sqrt{2.25}} \left(\frac{1}{2} - \frac{1}{4} + 1 \right) = \frac{5}{4\sqrt{2.25}} \tag{3.9}$$

Thus,

$$f_2'(t) = s_2(t) - \frac{5}{4\sqrt{2.25}} \frac{s_1(t)}{\sqrt{2.25}} = s_2(t) - \frac{5}{9} s_1(t) \tag{3.10}$$

and we obtain the following piecewise function

$$f_2'(t) = \begin{cases} -1/18 & \text{for } 0 \leq t < 1 \\ 7/9 & \text{for } 1 \leq t < 2 \\ 4/9 & \text{for } 2 \leq t < 3 \end{cases} \tag{3.11}$$

whose energy is ≈ 0.8056 and finally $f_2(t) \approx f_2'(t)/0.8056$.

Exercise 6: Let us consider two different signals $s_k(t)$ and $s_m(t)$. Moreover let $s_{bk}(t)$ and $s_{bm}(t)$ be their complex baseband representations, respectively. Show that

$$\int_{-\infty}^{\infty} s_m(t)s_k^*(t)dt = \int_{-\infty}^{\infty} \mathrm{Re}\{s_{bm}(t)s_{bk}^*(t)\}dt \tag{3.12}$$

Exercise 7: Now consider signal space representation of $s_k(t)$, $s_m(t)$, $s_{bk}(t)$ and $s_{bm}(t)$ and denote them as $\boldsymbol{s}_k$, $\boldsymbol{s}_m$, $\boldsymbol{s}_{bk}$ and $\boldsymbol{s}_{bm}$. Show that

$$\boldsymbol{s}_k\boldsymbol{s}_m^T = \mathrm{Re}\{\boldsymbol{s}_{bk}\boldsymbol{s}_{bm}^T\} \tag{3.13}$$

where ^{T} represents the transpose operation.

Let also define the Euclidian distance between two passband signals $s_k(t)$ and $s_m(t)$ as follows

$$\begin{aligned}
d(s_k(t), s_m(t)) &= \sqrt{\int_{-\infty}^{\infty} |s_k(t) - s_m(t)|^2 dt} && (3.14)\\
&= ||\boldsymbol{s}_k - \boldsymbol{s}_m|| = \sqrt{||\boldsymbol{s}_k||^2 + ||\boldsymbol{s}_m||^2 - 2\boldsymbol{s}_k\boldsymbol{s}_m^T} && (3.15)\\
&= \sqrt{||\boldsymbol{s}_{bk}||^2 + ||\boldsymbol{s}_{bm}||^2 - 2\mathrm{Re}\{\boldsymbol{s}_{bk}\boldsymbol{s}_{bm}\}^T} && (3.16)\\
&= ||\boldsymbol{s}_{bk} - \boldsymbol{s}_{bm}|| && (3.17)\\
&= \sqrt{\int_{-\infty}^{\infty} |s_{bk}(t) - s_{bm}(t)|^2 dt} && (3.18)
\end{aligned}$$

which means that that the Euclidean distance of the passband signals is identical to that of the corresponding baseband signals.

4 Digital Modulation

Suppose a sequence of binary digits a_n is to be transmitted over the physical channel. The main purpose of the modulator is to map every k bits to one of the $M = 2^k$ symbols/waveforms $\{s_i(t), i = 1, 2, \ldots, M\}$. If the transmitted waveforms $s_i(t)$ depends only on the current k bits, the modulation is called memoryless. Furthermore, if the underlying mapping is linear, the modulation is named linear, otherwise non-linear.

4.1 M-ary Pulse Amplitude Modulation (MPAM)

In this type of modulation, the amplitude of the signal is used to carry the information. Thus, it is also known as “amplitude shift keying" in literature. The MPAM passband waveform is represented by

$$s_m(t) = \sqrt{2}A_m g(t)\cos(2\pi f_c t) \tag{4.1}$$

where $A_m = (2m-1-M)d, m = 1, 2, \ldots, M$ are the M possible amplitudes and d is half the minimum symbol distance. We will later see that the minimum distance between two symbols

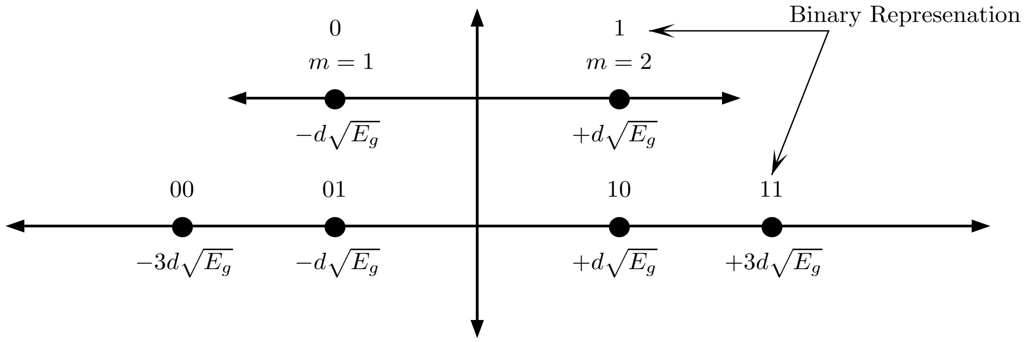

Figure 4.1: Signal space represenations of MPAM signals for $M = 1,2$. Also shown are the binary representation of symbols. This representation is usually known as mapping. We used gray mapping in this example where the adjacent symbols' binary representations differ at most 1 bit location.

of a given modulation has a major impact on the error probability of information delivery. The pulse shaping function $g(t)$ is a real-valued signal of duration T. From its definition, the equivalent baseband representation of a MPAM signal is $s_{bm} = A_m g(t)$.

Let us consider a spacial case when $M = 2$. The set of all possible signals contain,

$$s_1(t) = -\sqrt{2} d g(t) \cos(2\pi f_c t) \tag{4.2}$$

$$s_2(t) = \sqrt{2} d g(t) \cos(2\pi f_c t) \tag{4.3}$$

from which we immediately realize that $s_1(t) = -s_2(t)$. Due this property, this modulation is called antipodal signalling. Signal energy can be calculated to be

$$E_m = A_m^2 \int_0^T |g(t)|^2 dt = A_m^2 E_g = d^2 E_g \tag{4.4}$$

In order to find a basis function, we define $s_m(t)$ by its energy to obtain

$$f(t) = \sqrt{\frac{2}{E_g}} g(t) \cos(2\pi f_c t) \tag{4.5}$$

If we express $s_1(t)$ and $s_2(t)$ in terms of $f(t)$ we get

$$s_1(t) = -d\sqrt{E_g} \quad s_2(t) = d\sqrt{E_g} \tag{4.6}$$

4.2 M-ary Phase Shift Keying (MPSK)

If the information is transmitted using the phase of the modulating signal, it is called phase shift keying. The MPSK passband waveform is represented by

$$\begin{aligned} s_m(t) &= \sqrt{2} g(t) \cos(2\pi f_c t + \Theta_m) && (4.7) \\ &= \sqrt{2} g(t) \cos(\Theta_m) \cos(2\pi f_c t) - \sqrt{2} g(t) \sin(\Theta_m) \sin(2\pi f_c t) && (4.8) \end{aligned}$$

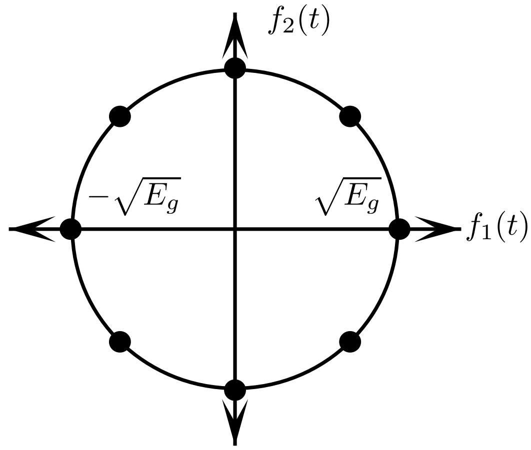

Figure 4.2: Signal space represenation of 8PSK signal.

where

$$\Theta_m = 2\pi(m-1)/M, \quad m = 1,2,\ldots,M \tag{4.9}$$

denotes the information conveying phase of the carrier. Note that all the signals of the PSK modulation has the same energy E_g. Then, the passband PSK signal can be represented by

$$s_m(t) = s_{m1}(t) f_1(t) + s_{m2}(t) f_2(t) \tag{4.10}$$

with the orthonormal basis functions given by

$$f_1(t) = \sqrt{\frac{2}{E_g}} g(t) \cos(2\pi f_c t) \tag{4.11}$$

$$f_2(t) = -\sqrt{\frac{2}{E_g}} g(t) \sin(2\pi f_c t) \tag{4.12}$$

The signal space representation is given by

$$\boldsymbol{s}_m = \left[\sqrt{E_g}\cos(\Theta_m) \quad \sqrt{E_g}\sin(\Theta_m) \right]^T \tag{4.13}$$

Finally, the euclidian distance between two signals of MPSK is given by

$$\begin{aligned} ||s_m - s_n|| &= \sqrt{\left|\sqrt{E_g}e^{j\Theta_m} - \sqrt{E_g}e^{j\Theta_n}\right|^2} && (4.14)\\ &= \sqrt{E_g(e^{j\Theta_m} - e^{j\Theta_n})(e^{-j\Theta_m} - e^{-j\Theta_n})} && (4.15)\\ &= \sqrt{2E_g(1 - \mathrm{Re}\{e^{j(\Theta_m - \Theta_n)}\})} && (4.16)\\ &= \sqrt{2E_g(1 - \cos(2\pi(m-n)/M)} && (4.17) \end{aligned}$$

from which we can deduce the minimum Euclidian distance between two MPSK signals to be

$$\sqrt{2E_g(1 - \cos(2\pi/M)} = 2\sqrt{E_g}\sin(\pi/M) \tag{4.18}$$

4.3 M-ARY QUADRATURE AMPLITUDE MODULATION (MQAM)

Note that with MPSK signalling, inphase and quadrature components are interrelated to eachother such that the amplitude of the signals are the same. Relaxing this constaint gives us MQAM signalling given by,

$$s_m(t) = \sqrt{2}a_m g(t)\cos(2\pi f_c t) - \sqrt{2}b_m g(t)\sin(2\pi f_c t) \tag{4.19}$$

where $m = 1,\ldots,M$ and $(\cos(2\pi f_c t), \sin(2\pi f_c t))$ pair are referred as quadrature carriers. Note that half of $\log_2 M$ bits are mapped to a_m and the other half of bits are mapped to b_m. By recognizing the complex equivalent envelope being $(a_m + jb_m)g(t)$, the energy of $s_m(t)$ can be calculated as

$$\int_0^T |s_m|^2 dt = |a_m + jb_m|^2 \int_0^T |g(t)|^2 dt = (a_m^2 + b_m^2)E_g \tag{4.20}$$

Similar to previous arguments, the signal space representation is given by

$$\boldsymbol{s}_m = \left[\sqrt{E_g}a_m \quad \sqrt{E_g}b_m\right]^T \tag{4.21}$$

using the same basis function derived for MPSK signalling. Finally, the euclidian distance between two signals of MQAM is given by

$$\begin{aligned} ||s_m - s_n|| &= \sqrt{E_g\left|a_m + jb_m - a_n - jb_n\right|^2} && (4.22)\\ &= \sqrt{E_g}\left|(a_m - a_n) + j(b_m - b_n)\right| && (4.23)\\ &= \sqrt{E_g}\sqrt{(a_m - a_n)^2 + (b_m - b_n)^2} && (4.24) \end{aligned}$$

As you might have noticed, we have made no attempt so far pertaining to the particular choice of (a_m, b_m) pair for the specification of s_m. It is general practice to assume equidistant signals i.e., $a_m, b_m \in \{\pm d, \pm 3d, \ldots, \pm(M-1)d\}$ in which case the minimum distance between constituent signals is going to be $2d\sqrt{E_g}$.

4.4 M-ARY FREQUENCY SHIFT KEYING (MFSK)

MFSK is an orthogonal (multi-dimensional) modulation scheme whose passband expression can be given by

$$s_m(t) = \sqrt{\frac{2E_s}{T_s}}\cos(2\pi(f_c + m\Delta f)t), \qquad m = 1,2,\ldots,M \tag{4.25}$$

where $s_m(t) \in [0, T_s]$ and E_s is the energy of the signal $s_m(t)$. For two MFSK signals $s_n(t)$ and $s_m(t)$, the correlation of the two shall give us

$$\begin{aligned} &= \frac{2E_s}{T_s}\int_0^{T_s}\cos(2\pi(f_c + m\Delta f)t)\cos(2\pi(f_c + n\Delta f)t)dt && (4.26)\\ &= \frac{E_s}{T_s}\int_0^{T_s}\cos(2\pi(m-n)\Delta f t)dt && (4.27)\\ &= \frac{E_s}{T_s}\frac{\sin(2\pi(m-n)\Delta f T_s)}{2\pi(m-n)\Delta f} && (4.28) \end{aligned}$$

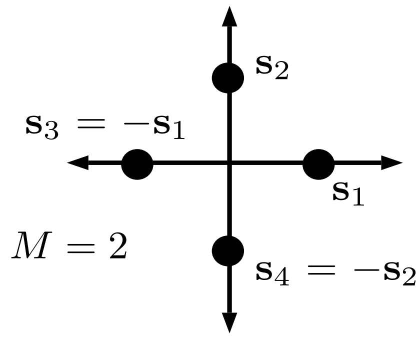

Figure 4.3: Biorthogonal signal representation using $M = 2$.

which is zero only if $\Delta f T_s = k/2$ for $k = \{\pm 1, \pm 2, \ldots\}$. Thus, the smallest separation that results in M orthogonal signals is $\Delta f = 1/2T$. From bandwidth efficiency perspective that is what is usually chosen.

If we select $\Delta f = 1/2T$, all M signals become orthogonal. Therefore they are indeed basis function with a proper normalization (to make it orthonormal) i.e.,

$$f_m(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t + \pi m t/T) \tag{4.29}$$

which results in a signal representation for $1 \le i \le M$

$$\mathbf{s}_i = [0\,0\ldots 0\,\sqrt{E_s}\,0\ldots 0\,0]; \tag{4.30}$$

whose ith entry is $\sqrt{E_s}$. We can realize that the minimum Euclidian distance between any signal points is $\sqrt{2E_s}$.

Additionally, we can add the set $\{-s_m(t)\}$ to the set of M orthogonal signals $\{s_m(t)\}$ to create $2M$ biorthogonal signal set. In this case the distance between signals is either $2\sqrt{E}$ or $\sqrt{2E}$. An example for $M = 2$ is shown in Fig. 4.3.

Exercise 8: Consider a set of M orthogonal signals $\{s_m(t)\}$ and define the mean of the signal set to be

$$\bar{\mathbf{s}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{s}_m \tag{4.31}$$

We can realize that this set of signals has non-zero mean. To remove it, we generate another set of signals given by

$$\mathbf{s}'_m = \mathbf{s}_m - \bar{\mathbf{s}} \tag{4.32}$$

First show that the set of signals $\{\mathbf{s}'_m\}$ has zero mean. Next, show that each signal has energy $E_s(1-1/M)$. This is interesting because These waveforms require less energy than orthogonal waveforms to achieve the same minimum Euclidean distance (you can simply verify that). Finally show that the correlation of the two distinct signals from this set is nonzero and equals to $-1/(M-1)$.

4.5 M-ary Differential Phase Shift Keying (DPSK)

In a differential phase shift keying, bits are mapped to the difference of two consecutive phases instead of phases themselves. In other words at time k, a phase difference level is selected depending on the transmitted symbol, i.e., $\Delta\Theta[k]$. The absolute value of the transmitted value can be found to be,

$$\Theta[k] = \Theta[k-1] + \Delta\Theta[k] \tag{4.33}$$

Note that the information is carried through the phase differences. This leads to a simpler receiver architecture i.e., the detection can be performed based on the phase difference of two waveforms of the consecutive symbol intervals. This modulation type is a non-coherent version of the MPSK modulation.

4.6 Minimum Shift Keying (MSK)

MSK is in fact a special case of the continuous phase frequency shift keying (CPFSK) in which the parameter, defined as the modulation index, equals a constant value. Since MSK has all the desirable properties of CPFSK and is easier to analyze (less sophisticated receiver architecture), we shall start by describing it and its interesting relationship with respect to FSK and QPSK modulation formats. Main motivation behind MSK (and hence CPFSK) is to come up with a more spectrally efficient modulation format by observing that frequency and phase changes between two consecutive different symbol transmissions are abrupt.

Let us start with a minor variation of 4PSK or also known as $QPSK$ called, Offset $QPSK$ or $OQPSK$. The signal representation of $OQPSK$ is exactly same as the one given for MPSK with $M = 4$ earlier. The difference between $QPSK$ and $OQPSK$ is in the alignment of inphase and quadrature bit streams. If the symbolling period is T, each components will have bitstreams sampled at $1/2T$ to be modulated. Thus, $QPSK$ may introduce $\pm\pi/2$ or π phase changes at every $2T$. This sharp jumps in the waveform leads to side lobe elevation and hence increasing the out-of-band radiation. On the other hand $OQPSK$ has one of the bit streams shifted by T to eliminate the phase change of π. Two important features of $OQPSK$ is its constant envelope signaling and smaller side lobes of its frequency response. Introduction of $OQPSK$ suggests that further suppression of out-of-band interference might be possible if the phase changes can completely be eliminated i.e., constant envelope modulation schemes with continuous phase might be needed.

MSK can be thought as a special case of $OQPSK$ where $g(t)$ is a sinusoidal pulse instead of a rectangular shape. We have the MSK signal expressed as

$$s(t) = a_I(t)\cos(\frac{\pi t}{2T}) + a_Q(t)\sin(\frac{\pi t}{2T}) \tag{4.34}$$

where $a_I(t) = +1, -1, -1, +1$ and $a_Q(t) = +1, -1, +1, +1$ are $2T$ long stream of bits as shown as a example in Fig. 4.4 (a) and (c). Fig. 4.4 (b) and (d) show how the waveform changes when modulated with carries. We can observe that the quadrature component $a_Q(t)$ is shifted right by T relative to the inphase component $a_I(t)$. The composite signal $s(t)$, the addition of Fig. 4.4 (b) and (d), is shown in Fig. 4.4 (e).

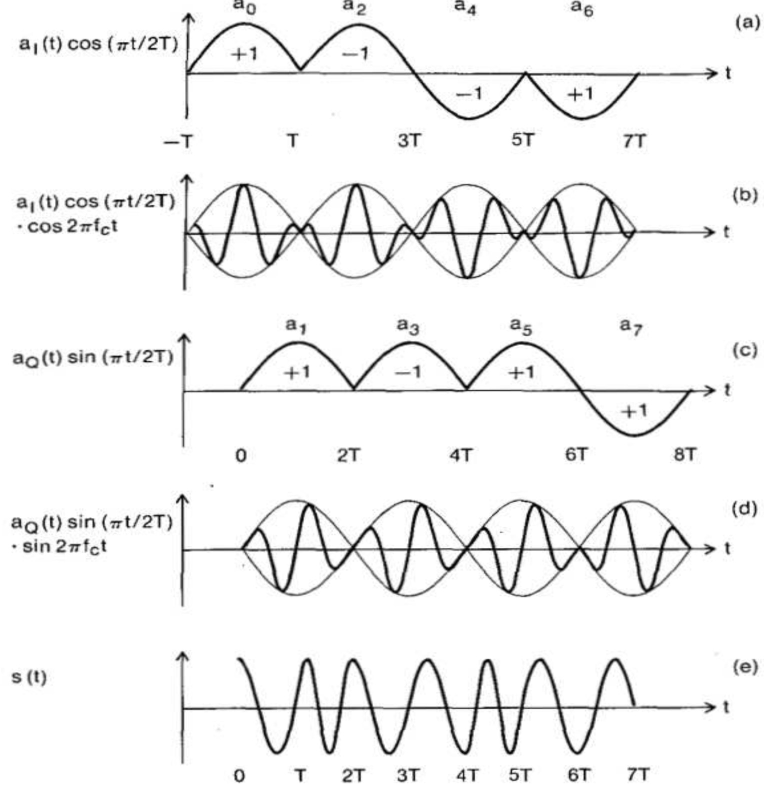

Figure 4.4: Representation of MSK symbols [1].

Note that using trigonometric identities, the equation (4.34), can be written as

$$s(t) = \cos(2\pi f_c t + b_k(t)\frac{\pi t}{2T} + \beta_k(t)) \tag{4.35}$$

where $b_k(t) = -a_I(t)a_Q(t)$ and $\beta_k(t) = (1 - a_I(t))\pi/2$. Observing Fig. 4.4, we notice that MSK waveform is a constant envelope as desired. Additionally, the phase is continuous at bit transitions. Equation 4.35 can be interpreted as an FSK signal whose frequency can be varied from $f_c - 1/4T$ to $f_c + 1/4T$ with phase changes of 0 or π. This leads to a frequency variation of $1/2T$. From our previous discussion of FSK, we can notice that $1/2T$ is the minimum frequency separation that allows FSK signals to be orthogonal. This is indeed why this modulation scheme is named “minimum shift keying". Overall, MSK can be both viewed as $OQPSK$ with sinusoidal pulse weighting or as $CPFSK$ with a frequency separation of $1/2T$.

The spectrum of MSK has lower sidelobes than that of $QPSK$ and $OQPSK$. This makes MSK extremely spectrum efficient modulation scheme. However MSK has a bigger mainlobe, suggesting that MSK may not be a good modulation scheme for narrow band communication links. MSK has also other special properties which make them quite industry standard. For example, MSK has simple demodulation and synchronization circuits [1]. Additionally, spectal properties of MSK can be improved by shaping the data pulses further. For example input symbols can be applied a Gaussian filtering before MSK modulation. This scheme is adopted by GSM operations.

5 Demodulation (Signal Recovery) for Digital Transmission

5.1 Preliminaries

In a reception scenario, the amplification or the attenuation of the received waveform leads to amplification as well as the attenuation of the transmitted waveform and the noise together. The parameter of interest is therefore the ratio of the signal power to the power of the undesired noise in the received waveform. This ratio is well known by the name signal-to-noise ratio, abbreviated as S/N or SNR.

In a filtering process, the input SNR can be improved at the output of the filter. This ratio can be expressed as

$$\Delta SNR = \frac{SNR_{out}}{SNR_{in}} \tag{5.1}$$

5.2 Matched Filter

There are two major operations conducted upon the received signal. Analog receivers (such as ones that are used with AM, FM, etc) typically aimed at reconstructing the transmitted waveform as closely as possible, while the digital receivers attempt to "pull" the signal out of the background noise without respecting the original waveform shape. In other words, they try to maximize the output SNR without regard to preserving the shape of the original signal.

The matched filter is a linear system that maximizes the output SNR. Here, the SNR is defined as the power ratio between a signal (meaningful information) and the background noise (undesired signal). A block diagram of the matched filter is roughly shown in Fig. 5.1. The output SNR is given by

$$\gamma_{SNR} = \frac{s_0^2(T)}{n_0^2(T)} = \frac{\left|\int_{-\infty}^{\infty} S(f)H(f)e^{j2\pi fT}df\right|^2}{\int_{-\infty}^{\infty}|H(f)|^2\Phi_{NN}(f)df} \tag{5.2}$$

where the denominator is the variance of the output noise process. We wish to choose $H(f)$ so as to maximize equation (5.2). However, this is about finding an optimal function rather than finding an optimal value. A solution to this problem can be given using the Cauchy-Schwartz's inequality which can be given for all functions $f(x)$ and $g(x)$ as follows,

$$\left|\int_{-\infty}^{\infty} f(x)g^*(x)dx\right|^2 \leq \int_{-\infty}^{\infty}|f(x)|^2dx\int_{-\infty}^{\infty}|g(x)|^2dx \tag{5.3}$$

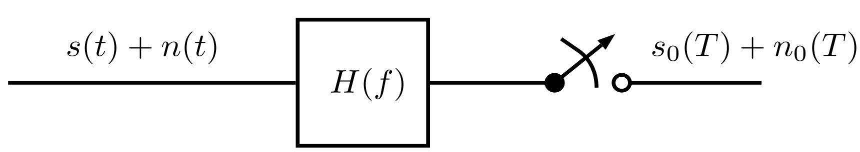

Figure 5.1: System diagram of a matched filter.

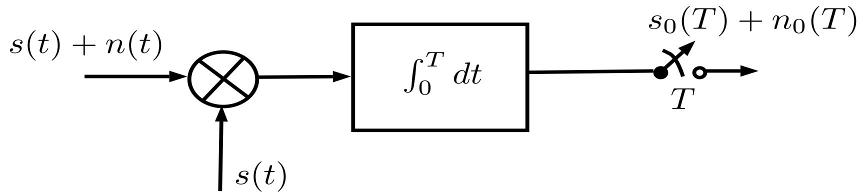

Figure 5.2: System diagram of a Correlator.

where the inequality can be achieved if $f(x) = cg(x)$ for some constant c. Let $g(x) = \frac{S^*(f)}{\sqrt{\Phi_{NN}(f)}}$ and $f(x) = H(f)\sqrt{\Phi_{NN}(f)}e^{j2\pi fT}$ using this inequality, we rewrite the numerator as follows,

$$\begin{aligned} \left|\int_{-\infty}^{\infty} S(f)H(f)e^{j2\pi fT}df\right|^2 &= \left|\int_{-\infty}^{\infty} \frac{S(f)}{\sqrt{\Phi_{NN}(f)}}H(f)\sqrt{\Phi_{NN}(f)}e^{j2\pi fT}df\right|^2 \\ &\leq \int_{-\infty}^{\infty}|H(f)|^2\Phi_{NN}(f)df\int_{-\infty}^{\infty}\frac{|S(f)|^2}{\Phi_{NN}(f)}df \end{aligned} \tag{5.4}$$

Combining this, equation (5.2) can be rewritten as,

$$\gamma_{SNR} \leq \int_{-\infty}^{\infty}\frac{|S(f)|^2}{\Phi_{NN}(f)}df \tag{5.5}$$

Therefore, an $H(f)$ can be seeked to achieve the upper bound. Yet, we know that upper bound is achievable if

$$H(f)\sqrt{\Phi_{NN}(f)}e^{j2\pi fT} = c\frac{S^*(f)}{\sqrt{\Phi_{NN}(f)}} \tag{5.6}$$

which implies for some constant c,

$$H(f) = ce^{-j2\pi fT}\frac{S^*(f)}{\Phi_{NN}(f)} \tag{5.7}$$

If we assume a constant noise spectrum ρ (like a white noise), using inverse Fourier transform, we have the time domain expressions

$$h(t) = c's^*(T-t) \tag{5.8}$$

where $c' = c \times \rho$. Since c' is a dummy variable, it can be ignored for simplicity. Thus, we have the following relationship between the input and output,

$$s_0(t) + n_0(t) = [s(t) + n(t)] * h(t) \tag{5.9}$$

in other words,

$$s_0(t) + n_0(t) = \int_0^T [s(\tau) + n(\tau)]s(T-(t-\tau))d\tau \tag{5.10}$$

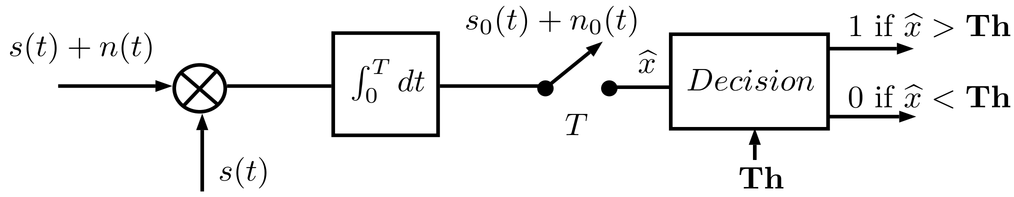

Figure 5.3: System diagram of a Baseband Communication System.

and at time $t = T$ (sampling at time T), we have

$$s_0(t) + n_0(t) = \int_0^T [s(\tau) + n(\tau)]\, s(\tau)\, d\tau \tag{5.11}$$

This shows the equivalence of a matched filter to the receiver system shown in Fig. 5.2. What this system does is nothing but correlating the original signal with the received signal and sampling the output signal. In other words, the receiver is finding the projection of the received signal in the direction of the transmitted signal. Due to inherent alignment of the system towards receiving $s(t)$, the output SNR is maximized.

5.3 Performance of Baseband Transmission

There are two distinct demodulation paradigms. First and the most popular one is “Coherent (synchronous) demodulation/detection". In this modulation paradigm, perfect carrier and phase synchronization is assumed at the receiver. The second one is “Noncoherent demodulation" which is usually achieved simple methods like envelope detection. Noncoherent systems usually does not require a carrier and phase regeneration in order to accurately demodulate the incoming signal.

In a typical synchronous demodulation scenario, the received waveform that passes though the analog front end goes through a band pass filter to reject out-of-band noise. Next, the incoming waveform is multiplied by a cosine of the carrier frequency. The resultant waveform goes though a low pass filter in order to obtain the original transmitted waveform. After the carrier multiplication, the final operations resembles to base band operations. Let us explore the performance of a simple baseband transmission using synchronous optimal receiver architectures. We can change Fig. 5.2 slightly to obtain Fig. 5.3 for our performance calculations. We assume the channel noise to be a bandpass white Gaussian noise with a two sided power spectral density. Note that the integrator is nothing but a low pass filter. The signal is assumed to pass through the lowpass filter without any clear distortion.

After low pass filtering and sampling received signal is $s_0(T) + n_0(T)$ where $s_0(t)$ is a binary valued function (either $-A_1$ representing “0" bit or A_2 representing “1" bit value) and $n_0(t)$ is white Gaussian noise with variance σ^2. The general problem of communication is to transmit binary information to the receiver side ideally without error. However, due to noise the reception process is never perfect. One of the problems regarding this process is to set the optimal threshold to minimize the the error probability. There are two possibilities for error to occur:

- A symbol 0 is transmitted, but the decision turns out to be 1.
- A symbol 1 is transmitted, but the decision turns out to be 0.

Apparently, the conditional probabilities of error, given that symbol 0 or 1 was transmitted can be expressed as

$$P_{e|0} = \frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbf{Th}}^{\infty} exp\left\{-\frac{(n+A_1)^2}{2\sigma^2}\right\}dn \tag{5.12}$$

$$P_{e|1} = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mathbf{Th}} exp\left\{-\frac{(n-A_2)^2}{2\sigma^2}\right\}dn \tag{5.13}$$

The average probability of error can be found by averaging over the a priori probabilities p_0 and p_1 of transmitting symbols 0 and 1, respectively.

$$\begin{aligned} P_e(\mathbf{Th}) &= p_0 P_{e|0} + p_1 P_{e|1} \\ &= \frac{p_0}{\sigma\sqrt{2\pi}}\int_{\mathbf{Th}}^{\infty} exp\left\{-\frac{(n+A_1)^2}{2\sigma^2}\right\}dn + \frac{1-p_0}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mathbf{Th}} exp\left\{-\frac{(n-A_2)^2}{2\sigma^2}\right\}dn \end{aligned} \tag{5.14}$$

So we choose $\mathbf{Th}^*$ such that average error probability is minimized. We set

$$\frac{dP_e(\mathbf{Th})}{d\mathbf{Th}} = 0 \tag{5.15}$$

Using the Leibnitz rule of differentiation[2]and equation (5.13), we can obtain

$$\frac{dP_e(\mathbf{Th})}{d\mathbf{Th}} = -\frac{p_0}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(\mathbf{Th}+A_1)^2}{2\sigma^2}\right\} + \frac{1-p_0}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(\mathbf{Th}-A_2)^2}{2\sigma^2}\right\} = 0 \tag{5.17}$$

This implies that

$$\frac{1-p_0}{p_0} = exp\left\{-\frac{(\mathbf{Th}^*+A_1)^2-(\mathbf{Th}^*-A_2)^2}{2\sigma^2}\right\} \tag{5.18}$$

$$= exp\left\{-\frac{2\mathbf{Th}^*(A_1+A_2)+A_1^2-A_2^2}{2\sigma^2}\right\} \tag{5.19}$$

from which we deduce

$$\mathbf{Th}^* = -\frac{\sigma^2}{A_1+A_2}\ln\left(\frac{1-p_0}{p_0}\right) + \frac{A_2^2-A_1^2}{2(A_1+A_2)} \tag{5.20}$$

Note that if $A_1 = A_2 = A$ and $p_0 = p_1 = 0.5$, the optimal threshold $\mathbf{Th}^* = 0$ as expected. If we further assume $z = \frac{n+B}{\sigma}$, we shall have

$$dz = \frac{dn}{\sigma} \Longrightarrow dn = \sigma dz. \tag{5.21}$$

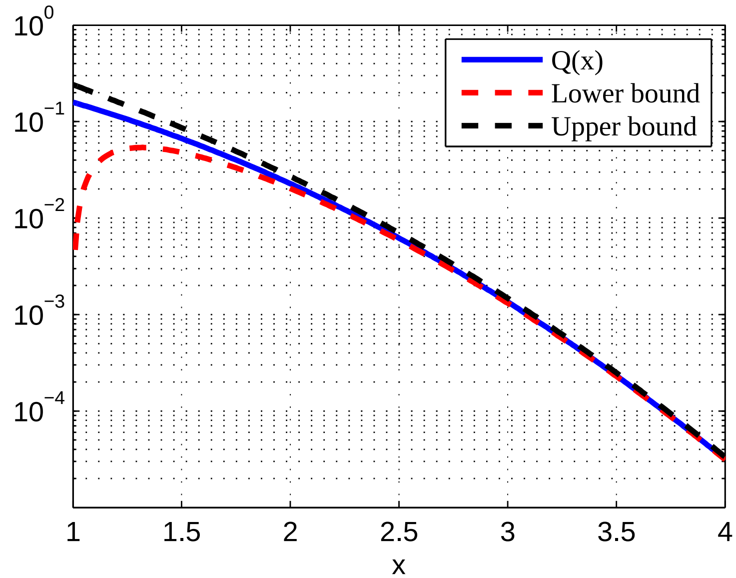

Figure 5.4: System diagram of a Baseband Communication System.

Then, if we let $B = A_1$ we can simply write down

$$P_{e|0} = \frac{1}{\sigma\sqrt{2\pi}} \int_{(\mathbf{Th}^*+A_1)/\sigma}^{\infty} exp\left\{-\frac{z^2}{2}\right\} \sigma dz = \frac{1}{\sqrt{2\pi}} \int_{(\mathbf{Th}^*+A_1)/\sigma}^{\infty} exp\left\{-\frac{z^2}{2}\right\} dz \quad (5.22)$$

$$= Q((\mathbf{Th}^* + A_1)/\sigma) \quad (5.23)$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} exp\left\{-\frac{z^2}{2}\right\} dz \quad (5.24)$$

is the well known Q error function used frequently for evaluating error performance of various modulation schemes. Similarly, we have $P_{e|1} = Q(-(\mathbf{Th}^* - A_2)/\sigma)$, Therefore,

$$P_e(\mathbf{Th}^*) = p_0 Q((\mathbf{Th}^* + A_1)/\sigma) + p_1 Q(-(\mathbf{Th}^* - A_2)/\sigma) \quad (5.25)$$

$$= p_0 Q((\mathbf{Th}^* + A_1)/\sigma) + p_1 - p_1 Q((\mathbf{Th}^* - A_2)/\sigma) \quad (5.26)$$

For example BPSK signaling with $p_0 = p_1 = 0.5$ and $\mathbf{Th}^* = 0$, the average probability of symbol/bit error is given by $Q(A/\sigma)$. Finally, we make note of the important bounds for $Q(x)$ function as follows,

$$\frac{1}{\sqrt{2\pi}x}\left(1 - \frac{1}{x^2}\right) e^{-x^2/2} < Q(x) < \frac{1}{\sqrt{2\pi}x} e^{-x^2/2} \quad (5.27)$$

[2]Let $I(\lambda) = \int_{a(\lambda)}^{b(\lambda)} f(x;\lambda)dx$ be the continuous function parameterized by λ, then

$$\frac{dI(\lambda)}{d\lambda} = \frac{db(\lambda)}{d\lambda} f(b(\lambda);\lambda) - \frac{da(\lambda)}{d\lambda} f(a(\lambda);\lambda) + \int_{a(\lambda)}^{b(\lambda)} \frac{\partial f(x;\lambda)}{\partial \lambda} dx \quad (5.16)$$

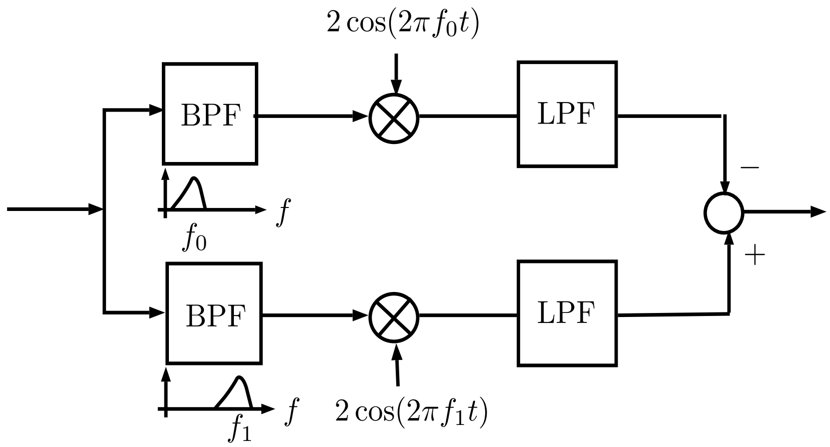

Figure 5.5: Coherent FSK demodulation. The two BPFs are non-overlapping in frequency spectrum.

where the upper bound is usually is not tight for large x. Instead, the following provides much better upper bound on $Q(x)$ function

$$Q(x) \leq \frac{1}{2}e^{-x^2/2} \tag{5.28}$$

Let us consider the coherent detection of binary FSK signalling. The receiver architecture for $M = 2$ case is shown in Fig. 5.5. Let us assume the signal is sent on the carrier $f_i, 0 \leq i \leq M-1$ i.e., $s_i(t) = \sqrt{2E_s/T_s}\cos(2\pi f_i t)$. At the output of each LPF, there will remain a noise term and only the ith branch will have the signal term. Thus, the output test statistics would be

$$y = \sqrt{2E_s/T_s} + n_1(t) - n_2(t) \tag{5.29}$$

Note that noises in the two channels are independent, if their spectra are non-overlapping. We also note that noise variance add, and hence the overall noise process is Gaussian with variance $2\sigma^2$. Therefore the symbol error rate can simply be found as

$$Q\left(\sqrt{\frac{E_s}{T_s\sigma^2}}\right) \tag{5.30}$$

To be continued...

References

[1] S. Pasupathy, “Minimum Shift Keying: A Spectrally Efficient Modulation," *IEEE Commun. Mag.,* vol. 17, no. 7, pp. 14âĂŞ22, July 1979.