Lecture 13: QAM and Noise

1 Implementation of QAM

Last time we described QAM and the principles of its implementation. The incoming bits are encoded into complex valued symbols in a given constellation $\mathcal{A}$. The sequence of symbols are then mapped into a complex baseband waveform $u(t) = \sum_k a_k p(t-kT)$. The baseband waveform is then modulated up to passband to get the transmitted waveform, given by

$$\begin{aligned} x(t) &= 2\Re\left\{e^{2\pi i f_c t}\, u(t)\right\} \\ &= 2\cos(2\pi f_c t)\ \Re\{u(t)\} - 2\sin(2\pi f_c t)\ \Im\{u(t)\}. \end{aligned} \tag{1}$$

For implementation purposes, each complex multiplication above corresponds to 4 real multiplications. This becomes somewhat simpler both to implement and to visualize if the modulating pulse $p(t)$ is also real. In discussing the Nyquist criterion in the previous lecture, we saw that this is not a fundamental limitation. We also let a'_k and a''_k be the real and imaginary parts of $a_k = a'_k + ia''_k$ and assume that the symbols are generated as real and imaginary parts (as opposed to magnitude and phase, for example). We then have $\Re\{u(t)\} = \sum_k a'_k p(t-kT)$ and $\Im\{u(t)\} = \sum_k a''_k(t)p(t-kT)$. From (1), $x(t)$ becomes

$$x(t) = 2\left(\cos 2\pi f_c t\right)\left(\sum_k a'_k p(t-kT)\right) - 2\left(\sin 2\pi f_c t\right)\left(\sum_k a''_k p(t-kT)\right).$$

This can be realized as two parallel PAM systems, followed by "double-sideband" modulation by "quadrature carriers" $\cos 2\pi f_c t$ and $-\sin 2\pi f_c t$. These are then summed (with the usual factor of 2), as shown in Figure 1. This realization of QAM is called *double-sideband quadrature-carrier* (DSB-QC) modulation[1].

A QAM receiver must first demodulate the received waveform $y(t)$. Assuming the scaling and receiver time reference discussed before, this received waveform is assumed to be simply $y(t) = x(t) + n(t)$. Here we also assume that there is no noise, so that $y(t)$ is simply the transmitted waveform $x(t)$. The first task of the receiver is to demodulate $x(t)$ back to baseband. Here, as explained shortly, this is done by multiplying the received waveform by both $\cos(2\pi f_c t)$ and $-\sin(2\pi f_c t)$. The two resulting waveforms are each filtered by a filter with impulse response $q(t)$ and then sampled at T spaced intervals. The resulting DSB-QC receiver is shown in Figure 2.

[1] The terminology comes from analog modulation where two real analog waveforms are modulated respectively onto cosine and sine carriers. For analog modulation, it is customary to transmit an additional component of carrier from which timing and phase can be recovered. As we see shortly, no such additional carrier is necessary here.

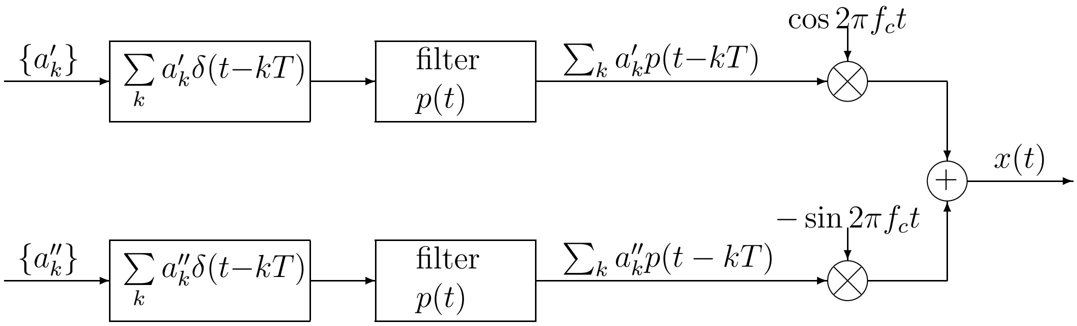

Figure 1. DSB-QC modulation.

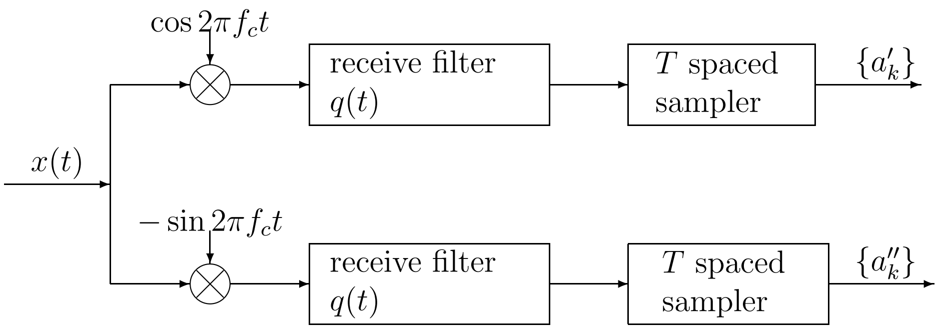

Figure 2. DSB-QC demodulation.

The multiplication by $\cos(2\pi f_c t)$ at the receiver moves the positive frequency part of $x(t)$ both up and down in frequency by f_c, and does the same with the negative frequency part. We are assuming throughout that both the transmit pulse $p(t)$ and the receive pulse $q(t)$ are in fact baseband waveforms relative to the carrier frequency (specifically, that $\hat{p}(f) = 0$ and $\hat{q}(f) = 0$ for $f \geq f_c$). Thus the result of multiplying the modulated waveform $x(t)$ by $\cos(2\pi f_c t)$ yields a response at baseband and also yields responses around $2f_c$ and $-2f_c$. The receive filter $q(t)$ then eliminates the double frequency terms. We can see the effect of the multiplication by $\cos(2\pi f_c t)$ both at transmitter and receiver from the following trigonometric identity:

$$2\cos(2\pi f_c t)\ \cos(2\pi f_c t) = 1 + \cos(4\pi f_c t)$$

Thus the receive filter $q(t)$ in the upper (cosine) part of the demodulator filters the real part of the original baseband waveform, resulting in the output $\sum_k a'_k (p * q)(t - kT)$. Assuming that the cascade $g(t)$ of the filters $p(t)$ and $q(t)$ is ideal Nyquist, the sampled output retrieves the real part of the original symbols without intersymbol interference. The filter $q(t)$ also rejects the double frequency terms.

The multiplication by $-\sin(2\pi f_c t)$ similarly moves the received waveform to a baseband component plus double carrier frequency terms. The effect of multiplying by $-\sin(2\pi f_c t$

at both transmitter and receiver is given by

$$2\sin(2\pi f_c t)\sin(2\pi f_c t) = 1 - \cos(4\pi f_c t)$$

Again, (assuming that $p(t) * q(t)$ is ideal Nyquist) the filter $q(t)$ in the lower (sine) part of the receiver retrieves the imaginary components of the orignal symbols without inter-symbol interference. Finally, from the identity

$$2\cos(2\pi f_c t)\sin(2\pi f_c t) = \sin(4\pi f_c t),$$

we see that there is no crosstalk at baseband between the real and imaginary parts of the original symbols.

It is important to go through the above argument to realize that our earlier approach of multiplying $u(t)$ by $e^{2\pi i f_c t}$ for modulation and then by $e^{-2\pi i f_c t}$ for demodulation is just a notationally more convenient way of doing the same thing. Working with sines and cosines is much more concrete, but is messier and makes it harder to see the whole picture.

2 Carrier recovery in QAM systems

Consider a QAM receiver and visualize the passband-to-baseband conversion as multiplying the positive frequency passband by the complex sinusoid $e^{-2\pi i f_c t}$. If the receiver has a phase error $\phi(t)$ in its estimate of the phase of the transmitted carrier, then it will instead multiply the incoming waveform by $e^{-2\pi i f_c t + i\phi(t)}$. We assume in this analysis that the time reference at the receiver is perfectly known, so that the sampling of the filtered output is done at the correct time. Thus the assumption is that the oscillator at the receiver is not quite in phase with the oscillator at the transmitter. Another point of view here is that the carrier frequency is usually orders of magnitude higher than the baseband bandwidth, and thus a small error in timing is significant in terms of carrier phase but not in terms of sampling. The carrier phase error will rotate the correct complex baseband signal $u(t)$ by $\phi(t)$; *i.e.*, the actual received baseband signal $v(t)$ will be

$$v(t) = e^{i\phi(t)}u(t).$$

If $\phi(t)$ is slowly time-varying relative to the response $q(t)$ of the receiver filter, then the samples $\{v_k\}$ of the filter output will be

$$v_k \approx e^{i\phi(kT)}a_k.$$

as illustrated in Figure 3. The phase error $\phi(t)$ is said to come through *coherently*. This phase coherence makes carrier recovery easy in QAM systems.

As can be seen from the figure, if the phase error is small enough, and the set of points in the constellation are well enough separated, then the phase error can be simply corrected by moving to the closest signal point and adjusting the phase of the demodulating carrier accordingly.

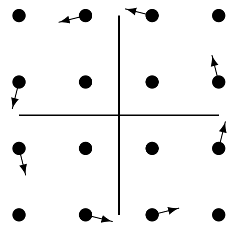

Figure 3: Rotation of constellation points by phase error

There are two complicating factors here. The first is that we have not taken noise into account yet. When the received signal $y(t)$ is $x(t) + n(t)$, then the output of the T spaced sampler is not the original symbols $\{a_k\}$, but rather a noise corrupted version of them. The second problem is that if a large phase error ever occurs, it can not be corrected. For example, in Figure 3, if $\phi(t) = \pi/2$, then even in the absence of noise, the received samples always line up with symbols from the constellation (but of course, not the transmitted symbols).

2.1 Tracking phase in the presence of noise

The problem of *deciding on* or *detecting* the symbols $\{a_k\}$ from the received samples $\{v_k\}$ in the presence of noise is a major topic that we will take up shortly. Here, however, we have the added complication of both detecting the transmitted symbols and also tracking and eliminating the phase error.

Fortunately, the problem of decision making and that of phase tracking are largely separable. The reason is that the oscillators used to generate the modulating and demodulating carriers are relatively stable and have phases which change quite slowly relative to each other. Thus the phase error with any kind of reasonable tracking will be quite small, and thus the data symbols can be detected from the received samples almost as if the phase error were zero. The difference between the received sample and the detected data symbol will still be non-zero, mostly due to noise but partly due to phase error. However, the noise has zero mean (as we understand later) and thus tends to average out over many sample times. Thus the general approach is to make decisions on the data symbols as if the phase error is zero, and then to make slow changes to the phase based on averaging over many sample times. This approach is called *decision directed carrier recovery.* Note that if we track the phase as phase errors occur, we are also tracking the carrier, in both frequency and phase.

In a decision directed scheme, assume that, for the received sample v_k, a decision d_k is made on the transmitted signal point a_k, and assume that $d_k = a_k$ with very high probability. The apparent phase error for the kth sample is then the difference between the phase of v_k and the phase of d_k. Any method for feeding back the apparent phase error to the generator of the sinusoid $e^{-2\pi i f_c t + i\phi(t)}$ in such a way as to slowly reduce the apparent phase error will tend to produce a robust carrier recovery system.

In one popular method, the feedback signal is taken as the imaginary part of $v_k d_k^*$. If the phase angle from d_k to v_k is ϕ_k, then

$$v_k d_k^* = |v_k||d_k|\ e^{i\phi_k},$$

so the imaginary part is $|v_k||d_k|\sin\phi_k \approx |v_k||d_k|\phi_k$, when ϕ_k is small. Decision-directed carrier recovery based on such a feedback signal can be extremely robust even in the presence of substantial distortion and large initial phase errors. With a second-order phase-locked carrier recovery loop, it turns out that the carrier frequency f_c can be recovered as well.

2.2 Large phase errors

A problem with decision-directed carrier recovery and with many other approaches is that the recovered phase may settle into any value for which the received eye pattern (*i.e.*, the pattern of a long string of received samples as viewed on a scope) "looks OK." With $(M \times M)$-QAM signal sets, as in Figure 3, the signal set has four-fold symmetry, and phase errors of $90°, 180°$, or $270°$ are not detectable. Simple differential coding methods that transmit the "phase" (quadrantal) part of the symbol information as a change of phase from the previous symbol rather than as an absolute phase can easily overcome this problem. Another approach is to resynchronize the system frequently by sending some known pattern of symbols. This latter approach is frequently used in wireless systems where fading sometimes causes a loss of phase synchronization.

3 Noise

It is now time to bring noise into the picture. We have managed to avoid it up until now by looking at the mechanics of modulation and intersymbol interference. However, noise is usually the fundamental limitation for communication over physical channels. It is the reason why the constellations we have discussed cannot have points that are too close together. However, to understand this at a more precise level, we need to model the noise. We will see that noise is right up with death and taxes as things that cannot be avoided but can be ameliorated somewhat by understanding.

We continue to take the view here that the channel output waveform is $y(t) = x(t) + n(t)$ where $x(t)$ is the channel input and $n(t)$ is the noise waveform. Naturally if the noise waveform were known, it could be removed and there would be little point talking about it. Thus it makes sense to assume that the noise waveform is a sample function of a random waveform. Random waveforms are called *stochastic processes*, so we must pause to introduce the notion of a stochastic process.

Recall that when we discussed source coding for analog waveforms, we were also faced with the problem of random waveforms. We handled the problem there by simply expanding any given source waveform in an orthonormal expansion and viewing the resulting samples as random variables.

Here we need to be a little more careful, since the behavior of the noise is critical to everything we do subsequently. Fortunately, our primary concern is a very simple model of noise called a *Gaussian process* where a few very basic ideas of stochastic processes will suffice.

4 Stochastic Processes

A *stochastic process* $\{N(t); -\infty < t < \infty\}$ is a collection of random variables, one for each value of the parameter t. We usually visualize t as denoting time, so there is one random variable for each instant of time. There are two major classes of stochastic processes, first, *discrete time stochastic processes*, where t takes on only integer values, and second, *continuous time stochastic processes*, where t takes on arbitrary real values. Here we are interested almost exclusively in the continuous time case, and assume that in what follows. We often restrict our attention to some finite interval of time or to non-negative times, but if we simply refer to a stochastic process as $\{N(t)\}$, we are implicitly considering a continuous time process over $-\infty < t < \infty$.

We will refer to specific points of time within the given range as epochs. There is an underlying sample space Ω over which these random variables are defined. That is, for each epoch t, the random variable $N(t)$ is a function $N(t, \omega)$ mapping sample points $\omega \in \Omega$ to real numbers.

In more advanced treatements, the random variable $N(t)$ at each epoch t is often generalized to be a complex random variable or a vector random variable (thus leading to complex or vector stochastic processes), but all of the random variables for now will be real (recall that random variables are defined as mappings from the sample space to the reals).

A given sample point ω within the underlying sample space determines the sample values of $N(t)$ for each epoch t. The collection of all these sample values for a given sample point ω, *i.e.*, $\{N(t, \omega)\}$ is called a *sample function* $n(t)$ of the process.

We could view a stochastic process as a completely abstract mathematical entity, but usually we want to view it as a model for some experiment where t represents time and where the sample function above represents the result of the experiment as a function of time, usually from minus to plus infinity. At first, this seems slightly inconsistent with the intuitive view of a probabilistic experiment in which, before the experiment is performed (at $t = -\infty$), any sample point $\omega \in \Omega$ might occur, and, after the experiment is performed (at $t = +\infty$), one and only one of these sample points has actually occurred.

There are two ways out of this perceived inconsistency. One is to recognize that the notion of 'before and after' in our intuitive view is inessential; the only important thing is the view that a multiplicity of sample points might occur, but only one actually occurs. This point of view is appropriate in designing a cellular telephone for manufacture. Each individual phone that is sold then experiences its own noise waveform, but the device is manufactured to work over a sample space of possible such waveforms.

The other way is to recognize that whether we view a function of time as going from $-\infty$ to $+\infty$ or going from some large negative to large positive time is primarily a matter of mathematical convenience. We often model signals as persisting from $-\infty$ to $+\infty$, but we know very well that they don't start before the device creating them is built, and don't continue after it is destroyed. More to the point, the time scales of interest in many communication problems are on the order of a second or less. Modeling assumptions over larger time intervals are irrelevant to the problem being modeled and are made for mathematical convenience.

In order to specify a stochastic process $\{N(t)\}$, some kind of rule is required from which joint distribution functions can, at least in principle, be calculated. That is, for all positive integers k, and all choices of k epochs $t_1, t_2, \dots, t_k$, it must be possible to find the joint distribution function,

$$F_{N(t_1),\dots,N(t_k)}(n_1, \dots, n_k) = \Pr\{N(t_1) \le n_1, \dots, N(t_k) \le n_k\} \tag{2}$$

for all choices of the real numbers $n_1, \dots, n_k$. Equivalently, if densities exist, it must be possible, in principle, to find the joint density,

$$f_{N(t_1),\dots,N(t_k)}(n_1, \dots, n_k) = \frac{\partial^k F_{N(t_1),\dots,N(t_k)}(n_1, \dots, n_k)}{\partial n_1 \cdots \partial n_k} \tag{3}$$

for all real $n_1, \dots, n_k$. Since k can be arbitrarily large in (2) and (3), it might seem difficult for a simple rule to specify all these quantities, but we shall see a number of simple rules in the examples to follow that allow all these quantities to be calculated.

Often the first thing we want to know about a process is the mean at each epoch t and the covariance between any two epochs t, τ. The mean, $\mathsf{E}[N(t)] = \overline{N}(t)$ is simply a function of t and can be found directly from $F_{N(t)}(n)$ or $f_{N(t)}(n)$. The covariance[2], as a function of the two epochs t, τ is denoted by $\mathsf{K}_N(t, \tau)$ and is given by

$$\mathsf{K}_N(t, \tau) = \mathsf{E}\left[[N(t) - \overline{N}(t)][N(\tau) - \overline{N}(\tau)]\right] \tag{4}$$

This can be calculated from the joint distribution function $F_{N(t),N(\tau)}(x_1, x_2)$ or from the density $f_{N(t),N(\tau)}(x_1, x_2)$. To make the covariance function look a little simpler, we usually split each random variable $N(t)$ into its mean $\overline{N}(t)$ and its fluctuation, $\widetilde{N}(t) = N(t) - \overline{N}(t)$. The covariance function is then

$$\mathsf{K}_N(t, \tau) = \mathsf{E}\left[\widetilde{N}(t)\widetilde{N}(\tau)\right] \tag{5}$$

The stochastic processes of most interest to us are used to model noise waveforms and usually have zero mean, in which case $N(t) = \widetilde{N}(t)$. In other cases, it usually aids intuition to separate the process into its mean (which is simply an ordinary function) and its fluctuation, which is then zero mean.

We now look at Gaussian processes. These are the canonic types of processes used to model noise, and many other stochastic processes are defined in terms of Gaussian processes. Many rules of thumb in engineering and statistics are stated without any mention of Gaussian processes, but are in fact valid only for Gaussian processes.

[2]This is often called the *autocovariance* to distinguish it from the covariance between two processes; we will not need to refer to this latter idea

5 Gaussian process

We will first define Gaussian random variables, then jointly Gaussian random variables, and finally Gaussian processes. A random variable (rv) N is a *normalized Gaussian* rv if it has the probability density

$$f_N(n) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-n^2}{2}\right].$$

It can be seen by symmetry that the mean of N is zero. The variance is 1, as you probably remember from elementary probability or can look up in any table of integrals. A random variable Z is a *Gaussian* rv if it is a scaled and shifted version of a normalized Gaussian rv, *i.e.*, if $Z = \sigma(N - \bar{Z})$ for a normalized Gaussian rv N. It can be seen that $\bar{Z}$ is the mean of Z and σ^2 is the variance. The density of Z is

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(z-\bar{Z})^2}{(2\sigma^2)}\right]. \tag{6}$$

We denote a Gaussian rv Z of mean $\bar{Z}$ and variance σ^2 as $Z \sim \mathcal{N}(\bar{Z}, \sigma^2)$.

Gaussian rv's are very important in modeling noise and other random phenemena for the following reasons:

- The central limit theorem roughly says that if we add a very large number of independent random variables, the sum tends to a Gaussian rv.
- Gaussian rv's have a number of extremal properties; as discussed later, they are, in some sense, the most random and the most noisy random variables for a given variance.
- After getting used to them, they are very easy to manipulate analytically.
- They are very commonly used as noise models for channels. Results in the literature often assume that noise random variables are Gaussian even without explicitly stating it.

A set of k random variables, $Z_1, \ldots, Z_k$ is said to be *zero mean jointly Gaussian* if there is a set of iid normalized Gaussian rv's $N_1, \ldots, N_m$ such that Z_l for each $l, 1 \le l \le k$, can be expressed as

$$Z_l = \sum_{j=1}^{m} a_{lj} N_j \tag{7}$$

where $a_{l1}, a_{l2}, \ldots, a_{lm}$ is a set of real numbers. It is convenient notationally to refer to a set of k random variables, $Z_1, \ldots, Z_k$ as a random vector (r**v**) $\boldsymbol{Z}$. If we then let A be a k by m matrix with the elements $\{a_{lj}\}$, we can express (7) more compactly as

$$\boldsymbol{Z} = A\,\boldsymbol{N} \tag{8}$$

Finally, $\boldsymbol{Z}' = (Z'_1, \dots, Z'_k)$ is *jointly Gaussian* if $\boldsymbol{Z}' = \boldsymbol{Z} + \bar{\boldsymbol{Z}}$ where $\boldsymbol{Z}$ is a zero mean jointly Gaussian r**v** and $\bar{\boldsymbol{Z}}$ is a k tuple of real numbers.

A sum of zero mean independent Gaussian rv's is a zero mean Gaussian rv, so each Z_l in (7) is zero mean Gaussian. Jointly Gaussian means much more than this, however. As can be seen from (8), any set of linear combinations of zero mean jointly Gaussian rv's is also a zero mean jointly Gaussian random vector.

If we interpret Gaussian rv's in terms of the central limit theorem, then it is natural to expect that statistically dependent Gaussian rv's are different linear combinations of the same set of small underlying rv's, and thus that they can be expressed as in (8).

Including a mean in the general definition of jointly Gaussian rv's is almost an afterthought. Noise is almost always viewed as zero mean, and if a rv Z has a mean $\bar{Z}$, it is best considered separately as some deterministic quantity which is added to the zero mean random variable $Z - \bar{Z}$. This zero mean variable, $\widetilde{Z} = Z - \bar{Z}$ is known as the *fluctuation* of Z.

We will come back later to describe the properties of zero mean jointly Gaussian rv's, and in particular to derive their joint probability density. For the moment, the only property that we need is given by the following theorem, which will be proved in the following lecture.

Theorem 5.1 (Jointly Gaussian density) *For any set $\{N_1, N_2, \dots, N_k\}$ of jointly Gaussian rv's, the joint density $f_{N_1,\dots,N_k}(n_1, \dots, n_k)$ is a function only of $\mathsf{E}[N_j]$ for each j, $1 \le j \le k$ and of $\mathsf{E}[\tilde{N}_j \tilde{N}_l]$ for each $1 \le l, j \le k$.*

We now have the machinery to define Gaussian processes.

Definition $\{N(t)\}$ is a *Gaussian process* if, for all positive integers k and all choices of epochs $t_1, \dots, t_k$, the set of random variables $N(t_1), \dots, N(t_k)$ is a jointly Gaussian set of random variables.

If we know the mean, $\overline{N}(t) = \mathsf{E}[N(t)]$ for each epoch t, and the covariance, $\mathsf{K}_N(t,\tau) = \mathsf{E}[\widetilde{N}(t)\widetilde{N}(\tau)]$ for each pair of epochs t, τ, then we know $\mathsf{E}\left[\widetilde{N}(t_i)\widetilde{N}(t_j)\right]$ for all i, j, and thus the joint density is specified for $t_1, \dots, t_k$. Thus a Gaussian process is completely specified by its mean and covariance function.

Example 1 Let $\{N(t);\, t \in \Re\}$ be defined by $N(t) = tZ$ for all $t \in \mathbb{R}$ where $Z \sim \mathcal{N}(0,1)$. This is an example of a zero mean Gaussian process. Its covariance function is $\mathsf{K}_N(t,\tau) = t\tau$. Our main purpose in discussing it is to show that a stochastic process can be very degenerate; a sample function of this process is fully specified by the sample value of the single random variable Z.

Example 2 As another zero mean Gaussian process, let $\{Z_k\}$ for integer k be a sequence of zero mean iid Gaussian rv's $\mathcal{N}(0, \sigma^2)$. Let

$$N(t) = \sum_k Z_k \operatorname{sinc}\left(\frac{t - kT}{T}\right)$$

It can be seen that the sample functions for this process are simply functions limited to the baseband $(-1/(2T),\, 1/(2T))$. This example brings out a mathematical issue that we will have to deal with later. The expected energy in a waveform is $\sum_k T\mathsf{E}[Z_k^2]$, which is infinite. The simple way to deal with this is to let $\{Z_k\}$ be a finite set of variables, say for $-10^8 \le k \le 10^8$.

Example 3 As an example of a non-Gaussian process, consider $\{N^2(t)\}$ where $\{N(t)\}$ is a Gaussian process. We could similarly consider any function $g(N(t))$. Such processes are usually best analyzed by focusing on the underlying Gaussian process.

6 Stationarity and Related Concepts

Many of the most useful stochastic processes have the property that the location of the time origin is irrelevant, *i.e.*, they "behave" the same way at one time as at any other time. This property is called *stationarity* and such a process is called a *stationary process*. We now define stationarity precisely.

Since the location of the time origin must be irrelevant for stationarity, stochastic processes that are defined over any interval other than $(-\infty, \infty)$ cannot be stationary. Thus we restrict attention to processes defined over the entire time interval.

The next requirement for a stochastic process $\{N(t)\}$ to be stationary is that $N(t)$ must be identically distributed for all epochs $t \in \mathbb{R}$. This means that, for any epochs t and $t+\tau$, and for any real number x, $\Pr\{N(t) \le x\} = \Pr\{N(t+\tau) \le x\}$. This does not mean that $N(t)$ and $N(t+\tau)$ are the same random variables; for a given sample outcome ω of the experiment, $N(t,\omega)$ is typically unequal to $N(t+\tau,\omega)$. It simply means that $N(t)$ and $N(t+\tau)$ have the same distribution function, *i.e.*,

$$F_{N(t)}(x) = F_{N(t+\tau)}(x) \qquad \text{for all } x \tag{9}$$

This is still not enough for stationarity, however. We also require that joint distributions over any set of epochs remain the same if all those epochs are shifted to new epochs by an arbitrary shift τ. This includes the previous requirement as a special case, so we have the definition:

Definition: A stochastic process $\{N(t);\, t\in\mathbb{R}\}$ is *stationary* if, for all positive integers k, for all sets of epochs $t_1 \in \mathbb{R}, \dots, t_k \in \mathbb{R}$, and for all shifts $\tau \in \mathbb{R}$,

$$F_{N(t_1),\dots,N(t_k)}(n_1 \dots, n_k) = F_{N(t_1+\tau),\dots,N(t_k+\tau)}(n_1 \dots, n_k) \tag{10}$$

for all $n_1, \dots, n_k \in \mathbb{R}$. For the typical case where densities exist, this can be rewritten as

$$f_{N(t_1),\dots,N(t_k)}(n_1 \dots, n_k) = f_{N(t_1+\tau),\dots,N(t_k+\tau)}(n_1 \dots, n_k) \tag{11}$$

for all $n_1, \dots, n_k \in \mathbb{R}$.

For any Gaussian process, the joint distribution of $N(t_1), \dots, N(t_k)$ depends only on the mean and covariance of those variables. Thus, for this distribution to be the same as that of $N(t_1+\tau), \dots, N(t_k+\tau)$, it is necessary and sufficient, first, that $\mathsf{E}[N(t_i)] = \mathsf{E}[N(t_i+\tau)]$ for $1 \leq i \leq n$, and, second, that $\mathsf{K}_N(t_i, t_j) = \mathsf{K}_N(t_i+\tau, t_j+\tau)$ for $1 \leq i, j \leq k$. For a Gaussian process $\{N(t)\}$ to be stationary, the above conditions must be satisfied for all τ, all k, and all $t_1, \dots, t_k$. Necessary and sufficient conditions for this, are, first, that $\mathsf{E}[N(t)] = \mathsf{E}[N(t+\tau)]$ for all t, τ, and, second, that $\mathsf{K}_N(t_1, t_2) = \mathsf{K}_N(t_1+\tau, t_2+\tau)$ for all t_1, t_2, τ. The following theorem states this in an even simpler form.

Theorem 6.1 *A Gaussian process $\{N(t);\, t \in \Re\}$ is stationary iff $\mathsf{E}[N(t)] = \mathsf{E}[N(0)]$ and $\mathsf{K}_N(t, t+\tau) = \mathsf{K}_N(0, \tau)$ for all $t, \tau \in \mathbb{R}$.*

With this theorem, we see that Example 2 is a stationary process, whereas Example 1 is non-stationary since $\mathsf{K}_N(t,t) = t^2$. Also, Example 3 is stationary if the underlying process $\{N(t)\}$ is stationary.

Example 2 brings out an interesting theoretical non-sequitur. Stochastic processes are simplified by assuming stationarity. However, non-trivial stationary stochastic processes have sample functions with infinite energy, and therefore the sample functions don't fit into the $\mathcal{L}_2$ theory of waveforms that we have been developing.

The practical solution to this is simple, and is the same solution that we used in Example 2. We simply truncate the stochastic process in any way that is convenient. Thus, when we say that noise is stationary, we mean that it is stationary over a much longer time interval than the the interval over which communication is to take place. This is not very nice theoretically, but it is better at this level to understand the problem than to try to develop some very awkward mathematics to hide the problem. We will soon develop the concept of white Gaussian noise, and will find that the problem there exists in both time and frequency.

There are many results in probability theory that depend only on the means and covariances of the random variables of interest. For stochastic processes, a number of these classical results become simpler when the process is stationary. For many of these results, only the part of the definition of stationarity that refers to mean and covariance is required. This leads to the following definition:

Definition: A stochastic process $\{N(t);\, t \in \Re\}$ is *wide sense stationary (WSS)* if $\mathsf{E}[N(t)] = \mathsf{E}[N(0)]$ and $\mathsf{K}_N(t, t+\tau) = \mathsf{K}_N(0, \tau)$ for all $t, \tau \in \mathbb{R}$. respectively.

Since the covariance function $\mathsf{K}_N(t, t+\tau)$ of a WSS process is a function of only one variable τ, we will often write the covariance function as a function of one variable, namely $\mathsf{K}_N(\tau)$ in place of $\mathsf{K}_N(t, t+\tau)$. In other words, the single variable in the single argument form represents the difference between the two arguments in two argument form. Thus, the covariance function $\mathsf{K}_N(t, \tau)$ of a WSS process must be a function only of $t - \tau$ and is expressed in single argument form as $\mathsf{K}_N(t - \tau)$.

The reader should not conclude from the frequent usage of the term WSS that there are many important processes that are WSS but not stationary. Rather we use the notion of wide sense stationarity to make clear, for some results, that they depend only on the mean and covariance, thus perhaps making it easier to understand them.