Chapter 2

Discrete-time and continuous-time AWGN channels

In this chapter we begin our technical discussion of coding for the AWGN channel. Our purpose is to show how the continuous-time AWGN channel model $Y(t) = X(t) + N(t)$ may be reduced to an equivalent discrete-time AWGN channel model $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, without loss of generality or optimality. This development relies on the sampling theorem and the theorem of irrelevance. More practical methods of obtaining such a discrete-time model are orthonormal PAM or QAM modulation, which use an arbitrarily small amount of excess bandwidth. Important parameters of the continuous-time channel such as SNR, spectral efficiency and capacity carry over to discrete time, provided that the bandwidth is taken to be the nominal (Nyquist) bandwidth. Readers who are prepared to take these assertions on faith may skip this chapter.

2.1 Continuous-time AWGN channel model

The continuous-time AWGN channel is a random channel whose output is a real random process

$$Y(t) = X(t) + N(t),$$

where $X(t)$ is the input waveform, regarded as a real random process, and $N(t)$ is a real white Gaussian noise process with single-sided noise power density N_0 which is independent of $X(t)$.

Moreover, the input $X(t)$ is assumed to be both power-limited and band-limited. The average input power of the input waveform $X(t)$ is limited to some constant P. The channel band B is a positive-frequency interval with *bandwidth* W Hz. The channel is said to be baseband if $B = [0, W]$, and passband otherwise. The (positive-frequency) support of the Fourier transform of any sample function $x(t)$ of the input process $X(t)$ is limited to B.

The *signal-to-noise ratio* SNR of this channel model is then

$$\mathrm{SNR} = \frac{P}{N_0 W},$$

where $N_0 W$ is the total noise power in the band B. The parameter N_0 is defined by convention to make this relationship true; *i.e.,* N_0 is the noise power per positive-frequency Hz. Therefore the double-sided power spectral density of $N(t)$ must be $S_{nn}(f) = N_0/2$, at least over the bands $\pm B$.

The two parameters W and SNR turn out to characterize the channel completely for digital communications purposes; the absolute scale of P and N_0 and the location of the band B do not affect the model in any essential way. In particular, as we will show in Chapter 3, the capacity of any such channel in bits per second is

$$C_{[\text{b/s}]} = W \log_2(1 + \text{SNR}) \quad \text{b/s}.$$

If a particular digital communication scheme transmits a continuous bit stream over such a channel at rate R b/s, then the *spectral efficiency* of the scheme is said to be $\rho = R/W$ (b/s)/Hz (read as "bits per second per Hertz"). The Shannon limit on spectral efficiency is therefore

$$C_{[(\text{b/s})/\text{Hz}]} = \log_2(1 + \text{SNR}) \quad (\text{b/s})/\text{Hz};$$

i.e., reliable transmission is possible when $\rho < C_{[(\text{b/s})/\text{Hz}]}$, but not when $\rho > C_{[(\text{b/s})/\text{Hz}]}$.

2.2 Signal spaces

In the next few sections we will briefly review how this continuous-time model may be reduced to an equivalent discrete-time model via the sampling theorem and the theorem of irrelevance. We assume that the reader has seen such a derivation previously, so our review will be rather succinct.

The set of all real finite-energy signals $x(t)$, denoted by $\mathcal{L}_2$, is a real vector space; *i.e.,* it is closed under addition and under multiplication by real scalars. The inner product of two signals $x(t), y(t) \in \mathcal{L}_2$ is defined by

$$\langle x(t), y(t) \rangle = \int x(t)y(t)\ dt.$$

The squared Euclidean norm (energy) of $x(t) \in \mathcal{L}_2$ is defined as $||x(t)||^2 = \langle x(t), x(t) \rangle < \infty$, and the squared Euclidean distance between $x(t), y(t) \in \mathcal{L}_2$ is $d^2(x(t), y(t)) = ||x(t) - y(t)||^2$. Two signals in $\mathcal{L}_2$ are regarded as the same ($\mathcal{L}_2$-equivalent) if their distance is 0. This allows the following strict positivity property to hold, as it must for a proper distance metric:

$$||x(t)||^2 \geq 0, \quad \text{with strict inequality unless } x(t) \text{ is } \mathcal{L}_2\text{-equivalent to } 0.$$

Every signal $x(t) \in \mathcal{L}_2$ has an $\mathcal{L}_2$ Fourier transform

$$\hat{x}(f) = \int x(t)e^{-2\pi i f t} dt,$$

such that, up to $\mathcal{L}_2$-equivalence, $x(t)$ can be recovered by the inverse Fourier transform:

$$x(t) = \int \hat{x}(f)e^{2\pi i f t} df.$$

We write $\hat{x}(f) = \mathcal{F}(x(t)), x(t) = \mathcal{F}^{-1}(\hat{x}(f))$, and $x(t) \leftrightarrow \hat{x}(f)$.

It can be shown that an $\mathcal{L}_2$ signal $x(t)$ is $\mathcal{L}_2$-equivalent to a signal which is continuous except at a discrete set of points of discontinuity ("almost everywhere"); therefore so is $\hat{x}(f)$. The values of an $\mathcal{L}_2$ signal or its transform at points of discontinuity are immaterial.

By Parseval's theorem, the Fourier transform preserves inner products:

$$\langle x(t), y(t)\rangle = \langle \hat{x}(f), \hat{y}(f)\rangle = \int \hat{x}(f)\hat{y}^*(f)\ df.$$

In particular, $||x(t)||^2 = ||\hat{x}(f)||^2$.

A signal space is any subspace $\mathcal{S} \subseteq \mathcal{L}_2$. For example, the set of $\mathcal{L}_2$ signals that are time-limited to an interval $[0, T]$ ("have support $[0, T]$") is a signal space, as is the set of $\mathcal{L}_2$ signals whose Fourier transforms are nonzero only in $\pm B$ ("have frequency support $\pm B$").

Every signal space $\mathcal{S} \subseteq \mathcal{L}_2$ has an orthogonal basis $\{\phi_k(t), k \in \mathcal{I}\}$, where $\mathcal{I}$ is some discrete index set, such that every $x(t) \in \mathcal{S}$ may be expressed as

$$x(t) = \sum_{k\in\mathcal{I}} \frac{\langle x(t), \phi_k(t)\rangle}{||\phi_k(t)||^2}\phi_k(t),$$

up to $\mathcal{L}_2$ equivalence. This is called an orthogonal expansion of $x(t)$.

Of course this expression becomes particularly simple if $\{\phi_k(t)\}$ is an orthonormal basis with $||\phi_k(t)||^2 = 1$ for all $k \in \mathcal{I}$. Then we have the orthonormal expansion

$$x(t) = \sum_{k\in\mathcal{I}} x_k\phi_k(t),$$

where $\mathbf{x} = \{x_k = \langle x(t), \phi_k(t)\rangle, k \in \mathcal{I}\}$ is the corresponding set of orthonormal coefficients. From this expression, we see that inner products are preserved in an orthonormal expansion; *i.e.*,

$$\langle x(t), y(t)\rangle = \langle \mathbf{x}, \mathbf{y}\rangle = \sum_{k\in\mathcal{I}} x_k y_k.$$

In particular, $||x(t)||^2 = ||\mathbf{x}||^2$.

2.3 The sampling theorem

The sampling theorem allows us to convert a continuous signal $x(t)$ with frequency support $[-W, W]$ (*i.e.*, a baseband signal with bandwidth W) to a discrete-time sequence of samples $\{x(kT), k \in \mathbb{Z}\}$ at a rate of $2W$ samples per second, with no loss of information.

The sampling theorem is basically an orthogonal expansion for the space $\mathcal{L}_2[0, W]$ of signals that have frequency support $[-W, W]$. If $T = 1/2W$, then the complex exponentials $\{\exp(2\pi i f kT), k \in \mathbb{Z}\}$ form an orthogonal basis for the space of Fourier transforms with support $[-W, W]$. Therefore their scaled inverse Fourier transforms $\{\phi_k(t) = \mathrm{sinc}_T(t - kT), k \in \mathbb{Z}\}$ form an orthogonal basis for $\mathcal{L}_2[0, W]$, where $\mathrm{sinc}_T(t) = (\sin \pi t/T)/(\pi t/T)$. Since $||\mathrm{sinc}_T(t)||^2 = T$, every $x(t) \in \mathcal{L}_2[0, W]$ may therefore be expressed up to $\mathcal{L}_2$ equivalence as

$$x(t) = \frac{1}{T}\sum_{k\in\mathbb{Z}} \langle x(t), \mathrm{sinc}_T(t - kT)\rangle \mathrm{sinc}_T(t - kT).$$

Moreover, evaluating this equation at $t = jT$ gives $x(jT) = \frac{1}{T}\langle x(t), \mathrm{sinc}_T(t - jT)\rangle$ for all $j \in \mathbb{Z}$ (provided that $x(t)$ is continuous at $t = jT$), since $\mathrm{sinc}_T((j - k)T) = 1$ for $k = j$ and $\mathrm{sinc}_T((j - k)T) = 0$ for $k \neq j$. Thus if $x(t) \in \mathcal{L}_2[0, W]$ is continuous, then

$$x(t) = \sum_{k\in\mathbb{Z}} x(kT)\mathrm{sinc}_T(t - kT).$$

This is called the sampling theorem.

Since inner products are preserved in an orthonormal expansion, and here the orthonormal coefficients are $x_k = \frac{1}{\sqrt{T}}\langle x(t), \mathrm{sinc}_T(t - kT)\rangle = \sqrt{T}x(kT)$, we have

$$\langle x(t), y(t)\rangle = \langle \mathbf{x}, \mathbf{y}\rangle = T\sum_{k\in\mathbb{Z}} x(kT)y(kT).$$

The following exercise shows similarly how to convert a continuous passband signal $x(t)$ with bandwidth W (*i.e.,* with frequency support $\pm[f_c - W/2, f_c + W/2]$ for some center frequency $f_c > W/2$) to a discrete-time sequence of sample pairs $\{(x_{c,k}, x_{s,k}), k \in \mathbb{Z}\}$ at a rate of W pairs per second, with no loss of information.

Exercise 2.1 (Orthogonal bases for passband signal spaces)

(a) Show that if $\{\phi_k(t)\}$ is an orthogonal set of signals in $\mathcal{L}_2[0, W]$, then $\{\phi_k(t)\cos 2\pi f_c t, \phi_k(t)\sin 2\pi f_c t\}$ is an orthogonal set of signals in $\mathcal{L}_2[f_c - W, f_c + W]$, the set of signals in $\mathcal{L}_2$ that have frequency support $\pm[f_c - W, f_c + W]$, provided that $f_c \ge W$.

[Hint: use the facts that $\mathcal{F}(\phi_k(t)\cos 2\pi f_c t) = (\hat{\phi}_k(f - f_c) + \hat{\phi}_k(f + f_c))/2$ and $\mathcal{F}(\phi_k(t)\sin 2\pi f_c t) = (\hat{\phi}_k(f - f_c) - \hat{\phi}_k(f + f_c))/2i$, plus Parseval's theorem.]

(b) Show that if the set $\{\phi_k(t)\}$ is an orthogonal basis for $\mathcal{L}_2[0, W]$, then the set $\{\phi_k(t)\cos 2\pi f_c t, \phi_k(t)\sin 2\pi f_c t\}$ is an orthogonal basis for $\mathcal{L}_2[f_c - W, f_c + W]$, provided that $f_c \ge W$.

[Hint: show that every $x(t) \in \mathcal{L}_2[f_c - W, f_c + W]$ may be written as $x(t) = x_c(t)\cos 2\pi f_c t + x_s(t)\sin 2\pi f_c t$ for some $x_c(t), x_s(t) \in \mathcal{L}_2[0, W]$.]

(c) Conclude that every $x(t) \in \mathcal{L}_2[f_c - W, f_c + W]$ may be expressed up to $\mathcal{L}_2$ equivalence as

$$x(t) = \sum_{k\in\mathbb{Z}} (x_{c,k}\cos 2\pi f_c t + x_{s,k}\sin 2\pi f_c t)\,\mathrm{sinc}_T(t - kT), \quad T = \frac{1}{2W},$$

for some sequence of pairs $\{(x_{c,k}, x_{s,k}), k \in \mathbb{Z}\}$, and give expressions for $x_{c,k}$ and $x_{s,k}$. □

2.4 White Gaussian noise

The question of how to define a white Gaussian noise (WGN) process $N(t)$ in general terms is plagued with mathematical difficulties. However, when we are given a signal space $\mathcal{S} \subseteq \mathcal{L}_2$ with an orthonormal basis as here, then defining WGN with respect to $\mathcal{S}$ is not so problematic. The following definition captures the essential properties that hold in this case:

Definition 2.1 (White Gaussian noise with respect to a signal space $\mathcal{S}$) *Let $\mathcal{S} \subseteq \mathcal{L}_2$ be a signal space with an orthonormal basis $\{\phi_k(t), k \in \mathcal{I}\}$. A Gaussian process $N(t)$ is defined as white Gaussian noise with respect to $\mathcal{S}$ with single-sided power spectral density N_0 if*

(a) The sequence $\{N_k = \langle N(t), \phi_k(t)\rangle, k \in \mathcal{I}\}$ is a sequence of iid Gaussian noise variables with mean zero and variance $N_0/2$;

(b) Define the "in-band noise" as the projection $N_{|\mathcal{S}}(t) = \sum_{k\in\mathcal{I}} N_k\phi_k(t)$ of $N(t)$ onto the signal space $\mathcal{S}$, and the "out-of-band noise" as $N_{|\mathcal{S}^\perp}(t) = N(t) - N_{|\mathcal{S}}(t)$. Then $N_{|\mathcal{S}^\perp}(t)$ is a process which is jointly Gaussian with $N_{|\mathcal{S}}(t)$, has sample functions which are orthogonal to $\mathcal{S}$, is uncorrelated with $N_{|\mathcal{S}}(t)$, and thus is statistically independent of $N_{|\mathcal{S}}(t)$.

For example, any stationary Gaussian process whose single-sided power spectral density is equal to N_0 within a band B and arbitrary elsewhere is white with respect to the signal space $\mathcal{L}_2(B)$ of signals with frequency support $\pm B$.

Exercise 2.2 (Preservation of inner products) Show that a Gaussian process $N(t)$ is white with respect to a signal space $\mathcal{S} \subseteq \mathcal{L}_2$ with psd N_0 if and only if for any signals $x(t), y(t) \in \mathcal{S}$,

$$\mathsf{E}[\langle N(t), x(t)\rangle\langle N(t), y(t)\rangle] = \frac{N_0}{2}\langle x(t), y(t)\rangle. \qquad \Box$$

Here we are concerned with the detection of signals that lie in some signal space $\mathcal{S}$ in the presence of additive white Gaussian noise. In this situation the following theorem is fundamental:

Theorem 2.1 (Theorem of irrelevance) *Let $X(t)$ be a random signal process whose sample functions $x(t)$ lie in some signal space $\mathcal{S} \subseteq \mathcal{L}_2$ with an orthonormal basis $\{\phi_k(t), k \in \mathcal{I}\}$, let $N(t)$ be a Gaussian noise process which is independent of $X(t)$ and white with respect to $\mathcal{S}$, and let $Y(t) = X(t) + N(t)$. Then the set of samples*

$$Y_k = \langle Y(t), \phi_k(t)\rangle, \quad k \in \mathcal{I},$$

is a set of sufficient statistics for detection of $X(t)$ from $Y(t)$.

Sketch of proof. We may write

$$Y(t) = Y_{|\mathcal{S}}(t) + Y_{|\mathcal{S}^\perp}(t),$$

where $Y_{|\mathcal{S}}(t) = \sum_k Y_k\phi_k(t)$ and $Y_{|\mathcal{S}^\perp}(t) = Y(t) - Y_{|\mathcal{S}}(t)$. Since $Y(t) = X(t) + N(t)$ and

$$X(t) = \sum_k \langle X(t), \phi_k(t)\rangle\phi_k(t),$$

since all sample functions of $X(t)$ lie in $\mathcal{S}$, we have

$$Y(t) = \sum_k Y_k\phi_k(t) + N_{|\mathcal{S}^\perp}(t),$$

where $N_{|\mathcal{S}^\perp}(t) = N(t) - \sum_k\langle N(t), \phi_k(t)\rangle\phi_k(t)$. By Definition 2.1, $N_{|\mathcal{S}^\perp}(t)$ is independent of $N_{|\mathcal{S}}(t) = \sum_k\langle N(t), \phi_k(t)\rangle\phi_k(t)$, and by hypothesis it is independent of $X(t)$. Thus the probability distribution of $X(t)$ given $Y_{|\mathcal{S}}(t) = \sum_k Y_k\phi_k(t)$ and $Y_{|\mathcal{S}^\perp}(t) = N_{|\mathcal{S}^\perp}(t)$ depends only on $Y_{|\mathcal{S}}(t)$, so without loss of optimality in detection of $X(t)$ from $Y(t)$ we can disregard $Y_{|\mathcal{S}^\perp}(t)$; *i.e.,* $Y_{|\mathcal{S}}(t)$ is a sufficient statistic. Moreover, since $Y_{|\mathcal{S}}(t)$ is specified by the samples $\{Y_k\}$, these samples equally form a set of sufficient statistics for detection of $X(t)$ from $Y(t)$. $\Box$

The sufficient statistic $Y_{|\mathcal{S}}(t)$ may alternatively be generated by filtering out the out-of-band noise $N_{|\mathcal{S}^\perp}(t)$. For example, for the signal space $\mathcal{L}_2(B)$ of signals with frequency support $\pm B$, we may obtain $Y_{|\mathcal{S}}(t)$ by passing $Y(t)$ through a brick-wall filter which passes all frequency components in B and rejects all components not in B.[1]

[1] Theorem 2.1 may be extended to any model $Y(t) = X(t) + N(t)$ in which the out-of-band noise $N_{|\mathcal{S}^\perp}(t) = N(t) - N_{|\mathcal{S}}(t)$ is independent of both the signal $X(t)$ and the in-band noise $N_{|\mathcal{S}}(t) = \sum_k N_k\phi_k(t)$; *e.g.,* to models in which the out-of-band noise contains signals from other independent users. In the Gaussian case, independence of the out-of-band noise is automatic; in more general cases, independence is an additional assumption.

Combining Definition 2.1 and Theorem 2.1, we conclude that for any AWGN channel in which the signals are confined to a sample space $\mathcal{S}$ with orthonormal basis $\{\phi_k(t), k \in \mathcal{I}\}$, we may without loss of optimality reduce the output $Y(t)$ to the set of samples

$$Y_k = \langle Y(t), \phi_k(t)\rangle = \langle X(t), \phi_k(t)\rangle + \langle N(t), \phi_k(t)\rangle = X_k + N_k, \quad k \in \mathcal{I},$$

where $\{N_k, k \in \mathcal{I}\}$ is a set of iid Gaussian variables with mean zero and variance $N_0/2$. Moreover, if $x_1(t), x_2(t) \in \mathcal{S}$ are two sample functions of $X(t)$, then this orthonormal expansion preserves their inner product:

$$\langle x_1(t), x_2(t)\rangle = \langle \mathbf{x}_1, \mathbf{x}_2\rangle,$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the orthonormal coefficient sequences of $x_1(t)$ and $x_2(t)$, respectively.

2.5 Continuous time to discrete time

We now specialize these results to our original AWGN channel model $Y(t) = X(t) + N(t)$, where the average power of $X(t)$ is limited to P and the sample functions of $X(t)$ are required to have positive frequency support in a band B of width W. For the time being we consider the baseband case in which $B = [0, W]$.

The signal space is then the set $\mathcal{S} = \mathcal{L}_2[0, W]$ of all finite-energy signals $x(t)$ whose Fourier transform has support $\pm B$. The sampling theorem shows that $\{\phi_k(t) = \frac{1}{\sqrt{T}}\text{sinc}_T(t-kT), k \in \mathbb{Z}\}$ is an orthonormal basis for this signal space, where $T = 1/2W$, and that therefore without loss of generality we may write any $x(t) \in \mathcal{S}$ as

$$x(t) = \sum_{k \in \mathbb{Z}} x_k \phi_k(t),$$

where x_k is the orthonormal coefficient $x_k = \langle x(t), \phi_k(t)\rangle$, and equality is in the sense of $\mathcal{L}_2$ equivalence.

Consequently, if $X(t)$ is a random process whose sample functions $x(t)$ are all in $\mathcal{S}$, then we can write

$$X(t) = \sum_{k \in \mathbb{Z}} X_k \phi_k(t),$$

where $X_k = \langle X(t), \phi_k(t)\rangle = \int X(t)\phi_k(t)\ dt$, a random variable that is a linear functional of $X(t)$. In this way we can identify any random band-limited process $X(t)$ of bandwidth W with a discrete-time random sequence $\mathbf{X} = \{X_k\}$ at a rate of $2W$ real variables per second. Hereafter the input will be regarded as the sequence $\mathbf{X}$ rather than $X(t)$.

Thus $X(t)$ may be regarded as a sum of amplitude-modulated orthonormal pulses $X_k\phi_k(t)$. By the Pythagorean theorem,

$$||X(t)||^2 = \sum_{k \in \mathbb{Z}} ||X_k\phi_k(t)||^2 = \sum_{k \in \mathbb{Z}} X_k^2,$$

where we use the orthonormality of the $\phi_k(t)$. Therefore the requirement that the average power (energy per second) of $X(t)$ be less than P translates to a requirement that the average energy of the sequence $\mathbf{X}$ be less than P per $2W$ symbols, or equivalently less than $P/2W$ per symbol.[2]

[2]The requirement that the sample functions of $X(t)$ must be in $\mathcal{L}_2$ translates to the requirement that the sample sequences $\mathbf{x}$ of $\mathbf{X}$ must have finite energy. This requirement can be met by requiring that only finitely many elements of $\mathbf{x}$ be nonzero. However, we do not pursue such finiteness issues.

Similarly, the random Gaussian noise process $N(t)$ may be written as

$$N(t) = \sum_{k \in \mathbb{Z}} N_k \phi_k(t) + N_{|\mathcal{S}^\perp}(t)$$

where $\mathbf{N} = \{N_k = \langle N(t), \phi_k(t) \rangle\}$ is the sequence of orthonormal coefficients of $N(t)$ in $\mathcal{S}$, and $N_{|\mathcal{S}^\perp}(t) = N(t) - \sum_k N_k \phi_k(t)$ is out-of-band noise. The theorem of irrelevance shows that $N_{|\mathcal{S}^\perp}(t)$ may be disregarded without loss of optimality, and therefore that the sequence $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ is a set of sufficient statistics for detection of $X(t)$ from $Y(t)$.

In summary, we conclude that the characteristics of the discrete-time model $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ mirror those of the continuous-time model $Y(t) = X(t) + N(t)$ from which it was derived:

- The symbol interval is $T = 1/2W$; equivalently, the symbol rate is $2W$ symbols/s;
- The average signal energy per symbol is limited to $P/2W$;
- The noise sequence $\mathbf{N}$ is iid zero-mean (white) Gaussian, with variance $N_0/2$ per symbol;
- The signal-to-noise ratio is thus SNR $= (P/2W)/(N_0/2) = P/N_0 W$, the same as for the continuous-time model;
- A data rate of ρ bits per two dimensions (b/2D) translates to a data rate of $R = W\rho$ b/s, or equivalently to a spectral efficiency of ρ (b/s)/Hz.

This important conclusion is the fundamental result of this chapter.

2.5.1 Passband case

Suppose now that the channel is instead a passband channel with positive-frequency support band $B = [f_c - W/2, f_c + W/2]$ for some center frequency $f_c > W/2$.

The signal space is then the set $\mathcal{S} = \mathcal{L}_2[f_c - W/2, f_c + W/2]$ of all finite-energy signals $x(t)$ whose Fourier transform has support $\pm B$.

In this case Exercise 2.1 shows that an orthogonal basis for the signal space is a set of signals of the form $\phi_{k,c}(t) = \mathrm{sinc}_T(t - kT) \cos 2\pi f_c t$ and $\phi_{k,s}(t) = \mathrm{sinc}_T(t - kT) \sin 2\pi f_c t$, where the symbol interval is now $T = 1/W$. Since the support of the Fourier transform of $\mathrm{sinc}_T(t - kT)$ is $[-W/2, W/2]$, the support of the transform of each of these signals is $\pm B$.

The derivation of a discrete-time model then goes as in the baseband case. The result is that the sequence of real pairs

$$(Y_{k,c}, Y_{k,s}) = (X_{k,c}, X_{k,s}) + (N_{k,c}, N_{k,s})$$

is a set of sufficient statistics for detection of $X(t)$ from $Y(t)$. If we compute scale factors correctly, we find that the characteristics of this discrete-time model are as follows:

- The symbol interval is $T = 1/W$, or the symbol rate is W symbols/s. In each symbol interval a pair of two real symbols is sent and received. We may therefore say that the rate is $2W = 2/T$ real dimensions per second, the same as in the baseband model.
- The average signal energy per dimension is limited to $P/2W$;

- The noise sequences $\mathbf{N}_c$ and $\mathbf{N}_s$ are independent real iid zero-mean (white) Gaussian sequences, with variance $N_0/2$ per dimension;
- The signal-to-noise ratio is again SNR $= (P/2W)/(N_0/2) = P/N_0W$;
- A data rate of ρ b/2D again translates to a spectral efficiency of ρ (b/s)/Hz.

Thus the passband discrete-time model is effectively the same as the baseband model.

In the passband case, it is often convenient to identify real pairs with single complex variables via the standard correspondence between $\mathbb{R}^2$ and $\mathbb{C}$ given by $(x, y) \leftrightarrow x + iy$, where $i = \sqrt{-1}$. This is possible because a complex iid zero-mean Gaussian sequence $\mathbf{N}$ with variance N_0 per complex dimension may be defined as $\mathbf{N} = \mathbf{N}_c + i\mathbf{N}_s$, where $\mathbf{N}_c$ and $\mathbf{N}_s$ are independent real iid zero-mean Gaussian sequences with variance $N_0/2$ per real dimension. Then we obtain a complex discrete-time model $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ with the following characteristics:

- The symbol interval is $T = 1/W$, or the rate is W complex dimensions/s.
- The average signal energy per complex dimension is limited to P/W;
- The noise sequence $\mathbf{N}$ is a complex iid zero-mean Gaussian sequence, with variance N_0 per complex dimension;
- The signal-to-noise ratio is again SNR $= (P/W)/N_0 = P/N_0W$;
- A data rate of ρ bits per complex dimension translates to a spectral efficiency of ρ (b/s)/Hz.

This is still the same as before, if we regard one complex dimension as two real dimensions.

Note that even the baseband real discrete-time model may be converted to a complex discrete-time model simply by taking real variables two at a time and using the same map $\mathbb{R}^2 \to \mathbb{C}$.

The reader is cautioned that the correspondence between $\mathbb{R}^2$ and $\mathbb{C}$ given by $(x, y) \leftrightarrow x + iy$ preserves some algebraic, geometric and probabilistic properties, but not all.

Exercise 2.3 (Properties of the correspondence $\mathbb{R}^2 \leftrightarrow \mathbb{C}$) Verify the following assertions:

(a) Under the correspondence $\mathbb{R}^2 \leftrightarrow \mathbb{C}$, addition is preserved.

(b) However, multiplication is not preserved. (Indeed, the product of two elements of $\mathbb{R}^2$ is not even defined.)

(c) Inner products are not preserved. Indeed, two orthogonal elements of $\mathbb{R}^2$ can map to two collinear elements of $\mathbb{C}$.

(d) However, (squared) Euclidean norms and Euclidean distances are preserved.

(e) In general, if $\mathbf{N}_c$ and $\mathbf{N}_s$ are real jointly Gaussian sequences, then $\mathbf{N}_c + i\mathbf{N}_s$ is not a proper complex Gaussian sequence, even if $\mathbf{N}_c$ and $\mathbf{N}_s$ are independent iid sequences.

(f) However, if $\mathbf{N}_c$ and $\mathbf{N}_s$ are independent real iid zero-mean Gaussian sequences with variance $N_0/2$ per real dimension, then $\mathbf{N}_c + i\mathbf{N}_s$ is a complex zero-mean Gaussian sequence with variance N_0 per complex dimension. □

2.6 Orthonormal PAM and QAM modulation

More generally, suppose that $X(t) = \sum_k X_k \phi_k(t)$, where $\mathbf{X} = \{X_k\}$ is a random sequence and $\{\phi_k(t) = p(t - kT)\}$ is an orthonormal sequence of time shifts $p(t - kT)$ of a basic modulation pulse $p(t) \in \mathcal{L}_2$ by integer multiples of a symbol interval T. This is called *orthonormal pulse-amplitude modulation (PAM).*

The signal space $\mathcal{S}$ is then the subspace of $\mathcal{L}_2$ spanned by the orthonormal sequence $\{p(t-kT)\}$; *i.e.,* $\mathcal{S}$ consists of all signals in $\mathcal{L}_2$ that can be written as linear combinations $\sum_k x_k p(t - kT)$.

Again, the average power of $X(t) = \sum_k X_k p(t - kT)$ will be limited to P if the average energy of the sequence $\mathbf{X}$ is limited to PT per symbol, since the symbol rate is $1/T$ symbol/s.

The theorem of irrelevance again shows that the set of inner products

$$Y_k = \langle Y(t), \phi_k(t)\rangle = \langle X(t), \phi_k(t)\rangle + \langle N(t), \phi_k(t)\rangle = X_k + N_k$$

is a set of sufficient statistics for detection of $X(t)$ from $Y(t)$. These inner products may be obtained by filtering $Y(t)$ with a *matched filter* with impulse response $p(-t)$ and sampling at integer multiples of T as shown in Figure 1 to obtain

$$Z(kT) = \int Y(\tau)p(\tau - kT)\, d\tau = Y_k,$$

Thus again we obtain a discrete-time model $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where by the orthonormality of the $p(t - kT)$ the noise sequence $\mathbf{N}$ is iid zero-mean Gaussian with variance $N_0/2$ per symbol.

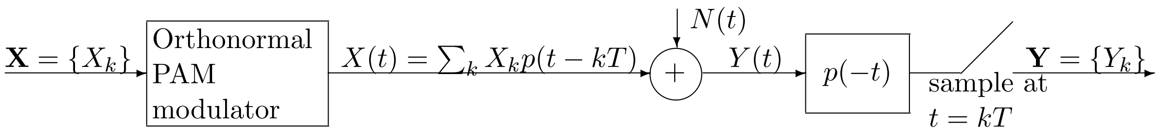

Figure 1. Orthonormal PAM system.

The conditions that ensure that the time shifts $\{p(t-kT)\}$ are orthonormal are determined by Nyquist theory as follows. Define the composite response in Figure 1 as $g(t) = p(t) * p(-t)$, with Fourier transform $\hat{g}(f) = |\hat{p}(f)|^2$. (The composite response $g(t)$ is also called the autocorrelation function of $p(t)$, and $\hat{g}(f)$ is also called its power spectrum.) Then:

Theorem 2.2 (Orthonormality conditions) *For a signal $p(t) \in \mathcal{L}_2$ and a time interval T, the following are equivalent:*

(a) The time shifts $\{p(t - kT), k \in \mathbb{Z}\}$ are orthonormal;

(b) The composite response $g(t) = p(t) * p(-t)$ satisfies $g(0) = 1$ and $g(kT) = 0$ for $k \neq 0$;

(c) The Fourier transform $\hat{g}(f) = |\hat{p}(f)|^2$ satisfies the Nyquist criterion for zero intersymbol interference, namely

$$\frac{1}{T} \sum_{m \in \mathbb{Z}} \hat{g}(f - m/T) = 1 \quad \text{for all } f.$$

Sketch of proof. The fact that (a) $\Leftrightarrow$ (b) follows from $\langle p(t - kT), p(t - k'T)\rangle = g((k - k')T)$. The fact that (b) $\Leftrightarrow$ (c) follows from the aliasing theorem, which says that the discrete-time Fourier transform of the sample sequence $\{g(kT)\}$ is the aliased response $\frac{1}{T}\sum_m \hat{g}(f - m/T)$. $\square$

It is clear from the Nyquist criterion (c) that if $p(t)$ is a baseband signal of bandwidth W, then

(i) The bandwidth W cannot be less than $1/2T$;

(ii) If $W = 1/2T$, then $\hat{g}(f) = T, -W \leq f \leq W$, else $\hat{g}(f) = 0$; *i.e.,* $g(t) = \mathrm{sinc}_T(t)$;

(iii) If $1/2T < W \leq 1/T$, then any real non-negative power spectrum $\hat{g}(f)$ that satisfies $\hat{g}(1/2T + f) + \hat{g}(1/2T - f) = T$ for $0 \leq f \leq 1/2T$ will satisfy (c).

For this reason $W = 1/2T$ is called the *nominal* or *Nyquist bandwidth* of a PAM system with symbol interval T. No orthonormal PAM system can have bandwidth less than the Nyquist bandwidth, and only a system in which the modulation pulse has autocorrelation function $g(t) = p(t) * p(-t) = \mathrm{sinc}_T(t)$ can have exactly the Nyquist bandwidth. However, by (iii), which is called the *Nyquist band-edge symmetry condition*, the Fourier transform $|\hat{p}(f)|^2$ may be designed to roll off arbitrarily rapidly for $f > W$, while being continuous and having a continuous derivative.

Figure 2 illustrates a raised-cosine frequency response that satisfies the Nyquist band-edge symmetry condition while being continuous and having a continuous derivative. Nowadays it is no great feat to implement such responses with excess bandwidths of 5–10% or less.

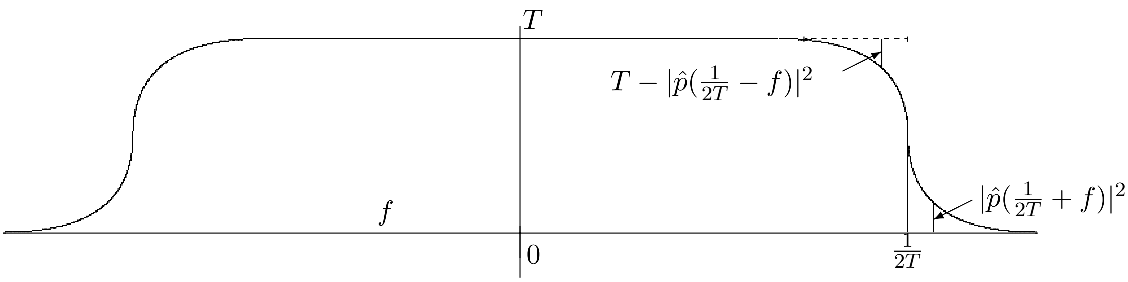

Figure 2. Raised-cosine spectrum $\hat{g}(f) = |\hat{p}(f)|^2$ with Nyquist band-edge symmetry.

We conclude that an orthonormal PAM system may use arbitrarily small excess bandwidth beyond the Nyquist bandwidth $W = 1/2T$, or alternatively that the power in the out-of-band frequency components may be made to be arbitrarily small, without violating the practical constraint that the Fourier transform $\hat{p}(f)$ of the modulation pulse $p(t)$ should be continuous and have a continuous derivative.

In summary, if we let W denote the Nyquist bandwidth $1/2T$ rather than the actual bandwidth, then we again obtain a discrete-time channel model $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ for any orthonormal PAM system, not just a system with the modulation pulse $p(t) = \frac{1}{\sqrt{T}}\mathrm{sinc}_T(t)$, in which:

- The symbol interval is $T = 1/2W$; equivalently, the symbol rate is $2W$ symbols/s;
- The average signal energy per symbol is limited to $P/2W$;
- The noise sequence $\mathbf{N}$ is iid zero-mean (white) Gaussian, with variance $N_0/2$ per symbol;
- The signal-to-noise ratio is $\mathrm{SNR} = (P/2W)/(N_0/2) = P/N_0W$;
- A data rate of ρ bits per two dimensions (b/2D) translates to a data rate of $R = \rho/W$ b/s, or equivalently to a spectral efficiency of ρ (b/s)/Hz.

Exercise 2.4 (Orthonormal QAM modulation)

Figure 3 illustrates an orthonormal quadrature amplitude modulation (QAM) system with symbol interval T in which the input and output variables X_k and Y_k are complex, $p(t)$ is a complex finite-energy modulation pulse whose time shifts $\{p(t-kT)\}$ are orthonormal (the inner product of two complex signals is $\langle x(t), y(t)\rangle = \int x(t)y^*(t)\ dt$), the matched filter response is $p^*(-t)$, and $f_c > 1/2T$ is a carrier frequency. The box marked $2\Re\{\cdot\}$ takes twice the real part of its input— *i.e.,* it maps a complex signal $f(t)$ to $f(t) + f^*(t)$— and the Hilbert filter is a complex filter whose frequency response is 1 for $f > 0$ and 0 for $f < 0$.

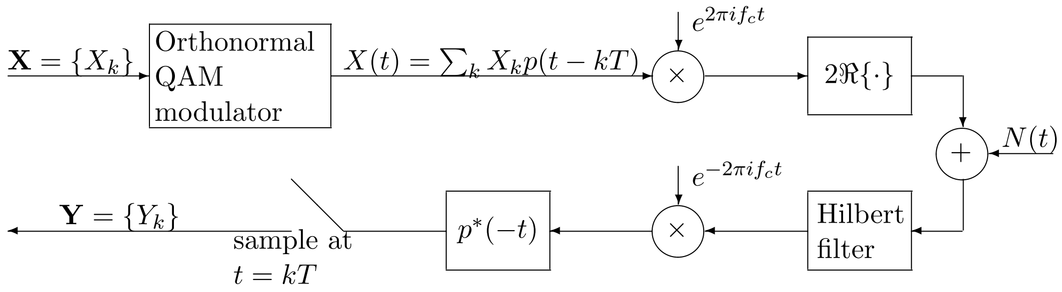

Figure 3. Orthonormal QAM system.

(a) Assume that $\hat{p}(f) = 0$ for $|f| \geq f_c$. Show that the Hilbert filter is superfluous.

(b) Show that Theorem 2.2 holds for a complex response $p(t)$ if we define the composite response (autocorrelation function) as $g(t) = p(t) * p^*(-t)$. Conclude that the bandwidth of an orthonormal QAM system is lowerbounded by its Nyquist bandwidth $W = 1/T$.

(c) Show that $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{N}$ is an iid complex Gaussian noise sequence. Show that the signal-to-noise ratio in this complex discrete-time model is equal to the channel signal-to-noise ratio $\mathrm{SNR} = P/N_0W$, if we define $W = 1/T$. [Hint: use Exercise 2.1.]

(d) Show that a mismatch in the receive filter— *i.e.,* an impulse response $h(t)$ other than $p^*(-t)$— results in linear intersymbol interference— *i.e.,* in the absence of noise $Y_k = \sum_j X_j h_{k-j}$ for some discrete-time response $\{h_k\}$ other than the ideal response δ_{k0} (Kronecker delta).

(e) Show that a phase error of θ in the receive carrier— *i.e.,* demodulation by $e^{-2\pi i f_c t + i\theta}$ rather than by $e^{-2\pi i f_c t}$— results (in the absence of noise) in a phase rotation by θ of all outputs Y_k.

(f) Show that a sample timing error of δ— *i.e.,* sampling at times $t = kT + \delta$— results in linear intersymbol interference. □

2.7 Summary

To summarize, the key parameters of a band-limited continuous-time AWGN channel are its bandwidth W in Hz and its signal-to-noise ratio SNR, regardless of other details like where the bandwidth is located (in particular whether it is at baseband or passband), the scaling of the signal, etc. The key parameters of a discrete-time AWGN channel are its symbol rate W in two-dimensional real or one-dimensional complex symbols per second and its SNR, regardless of other details like whether it is real or complex, the scaling of the symbols, etc. With orthonormal PAM or QAM, these key parameters are preserved, regardless of whether PAM or QAM is used, the precise modulation pulse, etc. The (nominal) spectral efficiency ρ (in (b/s)/Hz or in b/2D) is also preserved, and (as we will see in the next chapter) so is the channel capacity (in b/s).