# ECE 461: Digital Communication
# Lecture 8b: Pulse Shaping and Sampling

## Introduction

Information is digital in today's world but the physical world is still analog. Digital communication entails mapping digital information into electromagnetic energy (voltage waveforms) and transmitting over an appropriate physical medium (over a wire or wireless). At the receiver, we record the electromagnetic energy (voltage waveform again) and based on this knowledge, try to recover the original information bits. In the first lecture, we pointed out that for engineering convenience, the mapping between digital information and analog voltage waveforms is divided into two separate parts. At the transmitter:

- we first map digital information into a *discrete sequence* of voltage levels; this is the modulation or coding step.
- next, we interpolate between these voltage levels to produce an analog voltage waveform that is then transmitted; this is the DAC (digital to analog conversion) step.

At the receiver:

- we *sample* the received analog voltage waveform to produce a discrete sequence of voltage levels; this is the ADC (analog to digital conversion) step.
- next, we map the discrete sequence of sampled voltage levels to the information bits; this is the demodulation or decoding step.

These operations are depicted in Figure 1, in the context of transmission over the AWGN channel.

We have seen in the previous lectures, in substantial detail, the steps of modulation (coding) and demodulation (decoding). In this lecture, we will delve deeper into the DAC and ADC steps. At the end of this lecture, we will be able to derive a relationship between the sampling and interpolation rates (of the ADC and DAC, respectively) with an important physical parameter of an analog voltage waveform: *bandwidth*.

## Digital to Analog Conversion (DAC)

How do we map a sequence of voltages, $\{x[m]\}$, to waveforms, $x(t)$? This mapping is known as DAC. There are a few natural conditions we would like such a mapping to meet:

1. Since the digital information is contained in the discrete sequence of voltages $\{x[m]\}$, we would like these voltages to be readily recovered from the voltage waveform $x(t)$. One way of achieving this is to set

$$x(mT) = x[m] \tag{1}$$

where $T$ is the time period between voltage samples. This way, all the information we want to communicate is present in the transmitted waveform and readily extractable too.

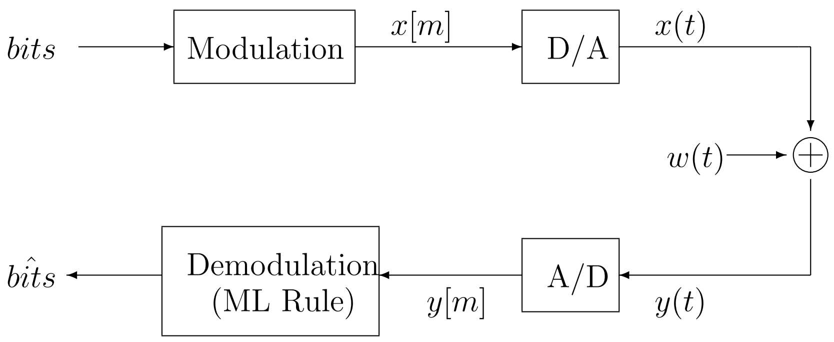


Figure 1: The basic transmit and receive operations in the context of communicating over the AWGN channel.

2. We could potentially pick any waveform that satisfies Equation (1). In other words, we seek to *interpolate* between the uniformly spaced voltage sequence. Of course, this interpolation should be universal, i.e., it should work for *any* sequence of discrete voltage levels (this is because the voltage sequence varies based on the coding method (sequential vs block) and also the information bits themselves). While such interpolation could be done in any arbitrary fashion (as long as Equation (1) is obeyed), there are some natural considerations to keep in mind:

   (a) We would like the resulting waveform $x(t)$ to have the smallest possible bandwidth. As we will see shortly, physical media, both wireless and wireline, impose spectral restrictions on the waveform transmitted through them; the most common type of these restrictions is that the bandwidth be as small as possible.

   (b) With an eye towards ease of implementation, we would like to have a *systematic* way of interpolating the voltage sequences. It would also be useful from an implementation stand point if the waveform value at any time can be generated using the discrete voltage values in its *immediate neighborhood*.

A convenient way of taking care of these conditions while interpolating, is to use the *pulse shaping* filter:

$$x(t) = \sum_{m>0} x[m]g\left(t - mT\right) \tag{2}$$

where $T$ is once again the time period between samples, and $g(t)$ is our "pulse", also known as the "interpolation filter". Essentially, a DAC that uses this pulse shaping equation will overlay (or convolve) the pulse over the voltage impulse defined by $x[m]$, and add all the convolutions together. Thus, the DAC in a transmitter is also called a "pulse shaper". This ensures a systematic way of generating the voltage waveform, but how well the other two conditions enumerated above are met depends on the choice of the pulse $g(t)$:

1. From Equation (2), we see that the bandwidth of $x(t)$ is exactly the *same* as the bandwidth of $g(t)$. So, controlling the bandwidth of $x(t)$ is the same as appropriate design of the pulse $g(t)$.

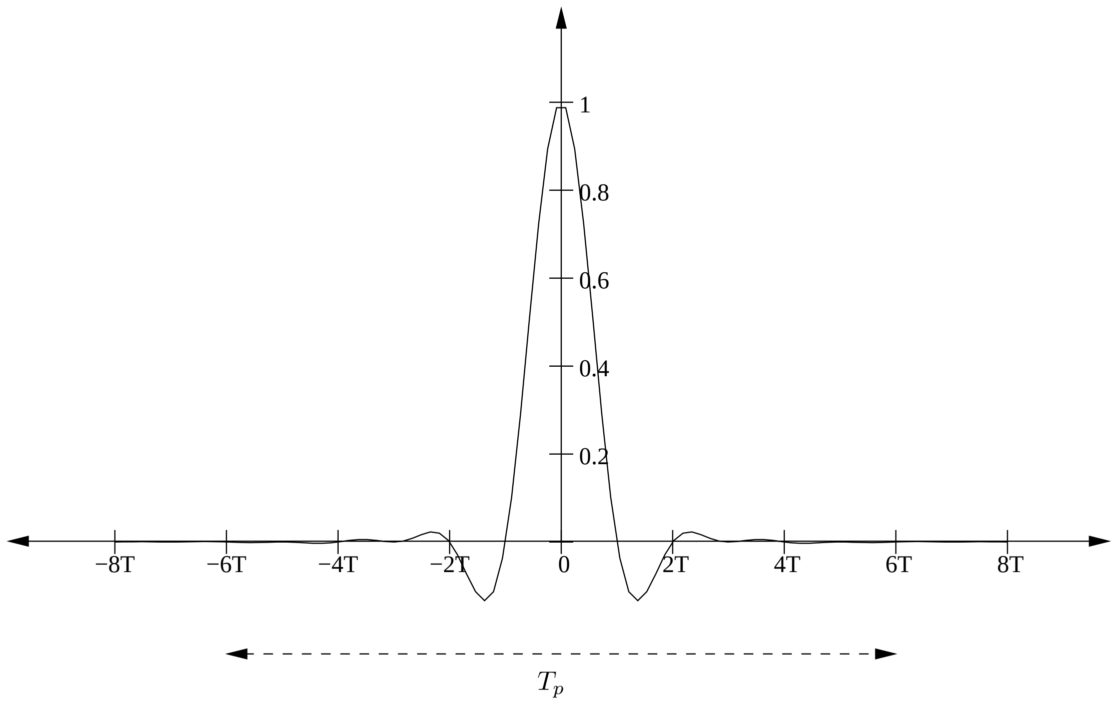


Figure 2: A pulse and its spread $T_p$.

2. How many neighboring discrete voltage values $x[m]$ affect the actual value of the voltage waveform $x(t)$ at any time $t$ depends on the *spread* of the pulse: the larger the spread, the more the impact of the number of neighboring discrete voltage values in deciding the waveform voltage value. Figure 2 illustrates this point: the number of neighboring voltages that make an impact is approximately the ratio of the spread of the pulse $T_p$ to the time period between the discrete voltage values $T$.

These two aspects are better appreciated in the concrete context of the three example pulses discussed below.

## Exemplar Pulses

- Rectangular Pulse: The rectangular pulse (`rect` pulse for short), or Zero Order Hold (ZOH) would overlay a voltage sample with a waveform that looks like that in Figure 3. As is, the `rect` pulse is nonzero for negative values of $t$, making it depend on a future discrete voltage level. An alternative shifted version of the ZOH simply holds each voltage until the next time instant: thus exactly one discrete voltage level is all that is needed (the immediately previous one) to decide the voltage waveform value at any time $t$.

  The greatest advantage of the ZOH is that it is very simple to implement with minimal memory requirements. On the flip side however, we know from our previous knowledge on waveforms and the Fourier transform that sharp edges in the time domain means large bandwidth in the frequency domain. Thus, this is the main disadvantage of

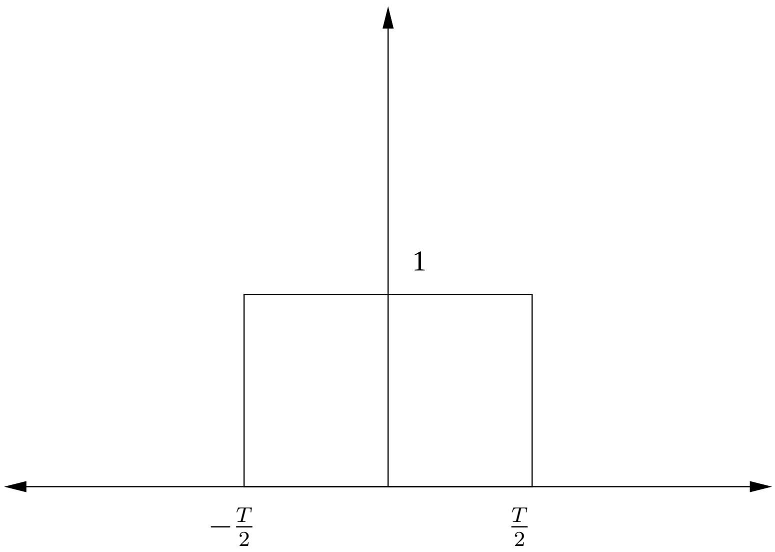


Figure 3: The rectangular pulse waveform $g(t) = \text{rect}\left(\frac{t}{T}\right)$

the ZOH, since its Fourier transform is actually a `sinc` function, which has infinite bandwidth. As a result, the rectangular pulse is not a very practical interpolation filter to use, since it is not possible to keep the spectral content of the waveform within any bandwidth constraint (that a channel would impose).

- Sinc Pulse: From the rectangular pulse, we learn that we want to constrain the bandwidth of the interpolation filter that we choose. If that is the case, then a natural choice for our pulse should have a power spectrum with a rectangular shape similar to that in Figure 3. A pulse shape with this spectrum is the "ideal interpolation filter" in signal processing, and corresponds to $g(t) = \text{sinc}\left(\frac{t}{T}\right) \stackrel{\text{def}}{=} \frac{\sin\left(\frac{\pi t}{T}\right)}{\frac{\pi t}{T}}$ in the time domain (see Figure 4).

  The advantages of the `sinc` pulse are similar to the advantages of the rectangular pulse, except that they hold in the frequency domain rather than in the time domain. It completely restricts the frequency content within a compact box so that we can tightly meet any bandwidth constraint. But such a perfect cut-off in the frequency domain implies that the time domain sequence has to be of infinite length, which means that the pulse spread is very large.

- Raised Cosine Pulse: A solution to the problems of the rectangular and `sinc` filters is to choose a pulse that lies in between. This way, the pulse would not be infinitely long in either the time or frequency domain. An example of such a pulse is the *raised cosine* pulse, whose waveform is shown in Figure 5. Mathematically, the raised cosine pulse is given by

$$g_{rc}(t) = \text{sinc}\left(\frac{t}{T}\right) \frac{\cos\left(\frac{\pi\beta t}{T}\right)}{1 - \frac{4\beta^2 t^2}{T^2}}. \tag{3}$$

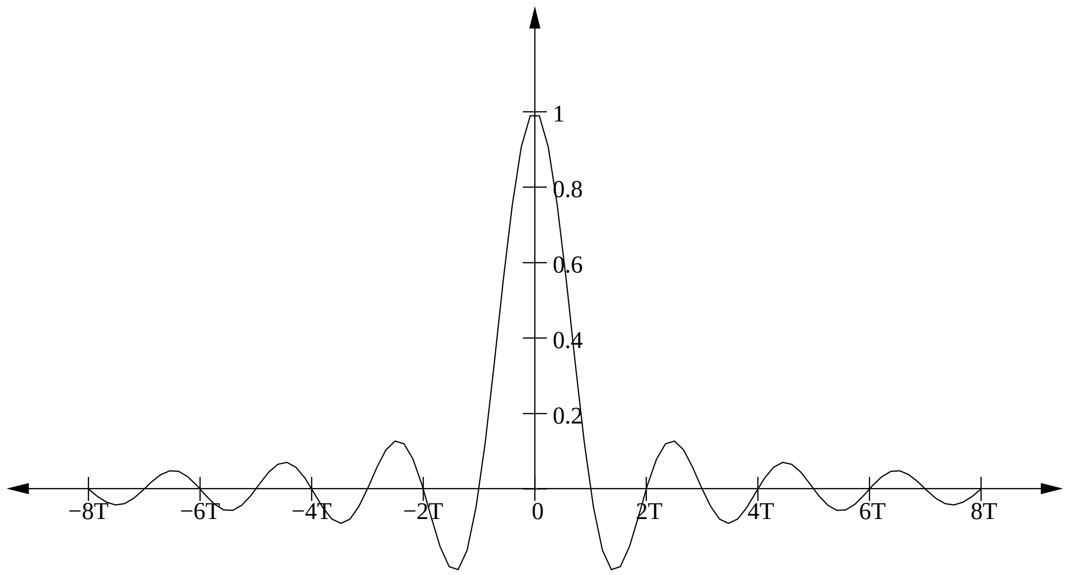


Figure 4: The sinc pulse waveform $g(t) = \text{sinc}\left(\frac{t}{T}\right)$.

By varying the parameter $\beta$ between 0 and 1, we can get from a `sinc` pulse ($\beta = 0$) to a much dampened version ($\beta = 1$).

As the illustration in Figure 5 demonstrates, the raised cosine has a smooth roll-off in the time domain, restricting the amount of bandwidth it uses. It also dampens to zero more quickly than the `sinc`, meaning that it is also more practical to use.

## Bandwidth and Narrowest Spacing

What is the bandwidth of the transmit signal with the exemplar pulses studied above? There are several definitions of bandwidth, but let us agree on an approximate definition that is good for our purposes: bandwidth of a voltage waveform $x(t)$ is the measure of the smallest set of frequencies where most of the energy of the *Fourier* transform $X(f)$ is contained in. For instance: the Fourier transform of the pulse sinc $\left(\frac{t}{T}\right)$ is $T\text{rect}\,(fT)$ (cf. Homework exercise). We see that all the energy of the Fourier transform is contained entirely within the finite spread of frequencies $\left[-\frac{1}{2T}, \frac{1}{2T}\right]$. So, we say that the bandwidth is $1/T$. More generally, the bandwidth of the raised cosine pulse depends on the parameter $\beta$: it increases as $\beta$ increases. Specifically, the bandwidth is equal to $1/T$ when $\beta = 0$, about $\frac{1.5}{T}$ when $\beta = 0.5$ and about $\frac{2}{T}$ when $\beta = 1$ (cf. Homework exercise).

In each of the cases studied above the bandwidth of the transmit signal is directly proportional to $1/T$, exactly the rate of discrete transmit voltage sequences. But our study of the special cases above was motivated by engineering purposes (simple methods of interpolation, etc). It is useful to know how much we are losing by taking the engineering considerations very seriously: this entails asking the following fundamental question:

> Supposing any interpolation strategy, what is the smallest bandwidth one can get as a function of the discrete transmit voltage rate $\frac{1}{T}$?

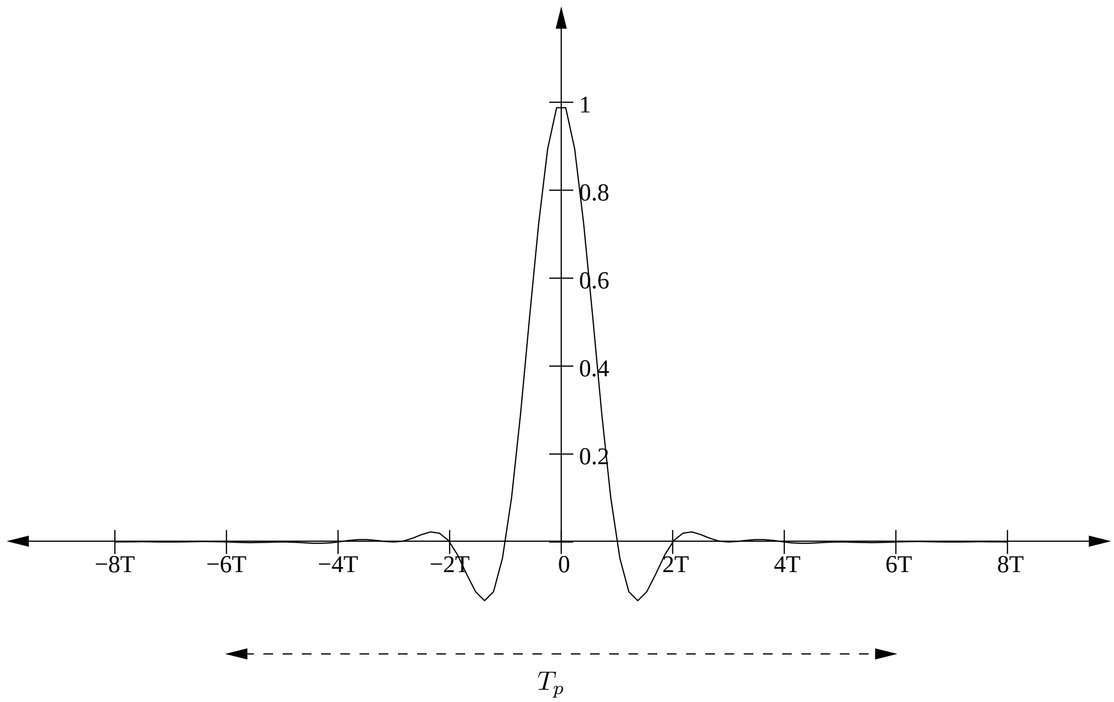


Figure 5: The raised cosine pulse waveform.

A fundamental result in communication theory says that the answer is approximately equal to $\frac{1}{T}$.[1] Concretely, this is exactly achieved by using the *sinc* pulse to shape the transmit waveform. In engineering practice, a raised cosine pulse (with an appropriate choice of $\beta$) is used. While that will entail a "loss" in bandwidth usage, we will simply use the term $1/T$ for the bandwidth of the transmit signal. This will not seriously affect any of the conceptual developments to follow (the numerical values of data rate might change).

## Analog to Digital Conversion (ADC)

The received voltage waveform $y(t)$ has potentially larger bandwidth than $x(t)$ due to the addition of the noise waveform $w(t)$. However, the information content is available only within the bandwidth of the transmit waveform $x(t)$. So we can, without loss of any information, filter the received waveform and restrict its bandwidth to be the same as that of the transmit waveform $x(t)$. We would like to convert this waveform into a discrete sequence of voltage levels that can be further processed to decode the information bits. The natural thing to do is to *sample* the received waveform at the same rate $1/T$: this creates a discrete

[1] We say approximately because our definition of bandwidth involved the mathematically ambiguous word "most".

sequence of voltage levels

$$\begin{aligned} y[m] &\stackrel{\text{def}}{=} y(mT) && (4)\\ &= x(mT) + w(mT) && (5)\\ &= x[m] + w[m], && (6) \end{aligned}$$

that are the transmit voltage sequence corrupted by additive noise. This is the basic AWGN channel we have studied reliable communication over. In engineering practice, it is quite common to over sample at the receiver: for instance at the rate of $2/T$ rather than $1/T$ samples per second. This leads to a somewhat more involved model where the intermediate received samples are noisy versions of a weighted sum of transmit voltages. We will take a careful look at the analytical forms of this channel model in the next lecture when we begin studying wireline communication. Further, we will see the engineering rationale for over sampling at the receiver during our upcoming study of reliable communication over wireline channels.

## Bandwidth and Capacity of the AWGN Channel

Consider the AWGN channel from Equation (6):

$$y[m] = x[m] + w[m], \quad m \geq 1. \tag{7}$$

We have seen that there are $W$ channel uses per second if we constrain the bandwidth of the analog transmit voltage waveform to $W$ Hz. How does the variance of the additive noise $w[m]$ depend on the bandwidth $W$? To understand this, we could look at the (random) noise waveform $w(t)$ and it's (random) Fourier transform. Since noise waveforms are *power* signals, i.e.,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (w(t))^2 \, dt \quad = \quad \sigma^2 > 0 \tag{8}$$

the Fourier transform is not well defined. To avoid this problem we could consider the time-restricted noise waveform

$$w_T(t) \stackrel{\text{def}}{=} \begin{cases} w(t) & -T \leq t \leq T \\ 0 & \text{else} \end{cases} \tag{9}$$

which is an *energy* signal. We denote the Fourier transform of $w_T(t)$ by $W_T(f)$. The *average* variance of the Fourier transform of the time-restricted noise in the limit of no restriction is called the *power spectral density*:

$$\text{PSD}_w(f) \stackrel{\text{def}}{=} \lim_{T\to\infty} \frac{1}{2T} \text{Var}\left(W_T(f)\right). \tag{10}$$

Based on measurements of additive noise, a common model for the power spectral density is that it is constant, denoted by $\frac{N_0}{2}$, measured in Watts/Hz. Furthermore this model holds over a very wide range of frequencies of interest to communication engineers: practical

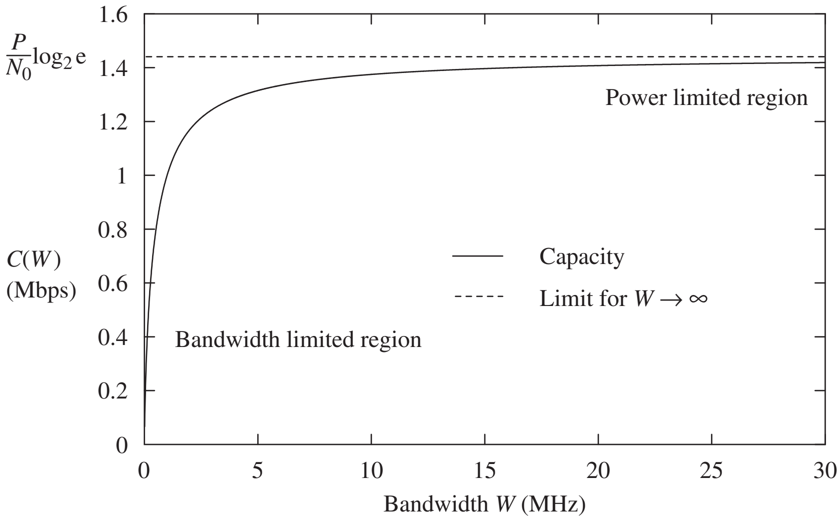


Figure 6: Capacity as a function of bandwidth for SNR per Hz $2P/N_0 = 10^6$.

measurement data suggests a value of about $10^{-14}$ Watts/Hz for $N_0$. With a (double sided) bandwidth of $W$, the total area under the power spectral density is

$$\int_{-\frac{W}{2}}^{\frac{W}{2}} \mathrm{PSD}(f)\, df = \frac{N_0 W}{2}. \tag{11}$$

It turns out that the variance of the noise sample $w[m]$, at time sample $m$, is *exactly equal* to the expression in Equation (11)! This remarkable calculation is explored in a homework exercise. So, we conclude that the variance of the noise sample increases in direct proportion to the bandwidth $W$.

We can now put together these observations into our earlier discussion of the capacity of the discrete time AWGN channel to arrive at the capacity of the analog (or continuous time) AWGN channel:

$$C = \frac{W}{2} \log_2 \left( 1 + \frac{2P}{N_0 W} \right) \quad \text{bits/s}. \tag{12}$$

Here the transmit power constraint is denoted by $P$ and is measured in Watts. Now we can see how the capacity depends on the bandwidth $W$. Surely the capacity can only increase as $W$ increases (one can always ignore the extra bandwidth). One can directly show that the capacity is a *concave* function of the bandwidth $W$ (this is explored in an exercise). Figure 6 plots the variation of the capacity as a function of bandwidth for an exemplar value of SNR per Hz.

Two important implications follow:

- When the bandwidth is small, the capacity is very sensitive to changes in bandwidth: this is because the SNR per Hz is quite large and then the capacity is pretty much linearly related to the bandwidth. This is called the *bandwidth limited* regime.

- When the bandwidth is large, the SNR per Hz is small and

$$\frac{W}{2}\log_2\left(1+\frac{2P}{N_0W}\right) \approx \frac{W}{2}\left(\frac{2P}{N_0W}\right)\log_2 e \tag{13}$$

$$= \frac{P}{N_0}\log_2 e. \tag{14}$$

  In this regime, the capacity is proportional to the total power $P$ received over the whole band. It is insensitive to the bandwidth and increasing the bandwidth has only a small impact on capacity. On the other hand, the capacity is now linear in the received power and increasing power does have a significant effect. This is called the *power limited* regime.

As $W$ increases, the capacity increases monotonically and reaches the asymptotic limit

$$C_\infty = \frac{P}{N_0}\log_2 e \quad \text{bits/s}. \tag{15}$$

This is the capacity of the AWGN channel with only a power constraint and no bandwidth constraint. It is important to see that the capacity is finite even though the bandwidth is not. The connection of this expression to energy efficient communication is explored in a homework exercise.

## Looking Ahead

Starting next lecture, we shift gears and start looking at communication over wires. We start by looking at how a wireline channel affects voltage waveforms passing through it. We will be able to arrive at a discrete time wireline channel model by combining the effect of the wire on the voltage waveforms passing through it along with the DAC and ADC operations.