ECE 361: Lecture 14: Capacity of the Continuous-Time AWGN Channel

14.1. Back from the Discrete to the Continuous

In Lecture 12, we noted that the capacity of the discrete-time Gaussian channel with an average power constraint is

$$C = \frac{1}{2}\log_2(1 + \mathsf{SNR}) \text{ bits per channel use} \tag{14.1}$$

where $\mathsf{SNR} = \mathcal{E}/\sigma^2$ with $\mathcal{E}$ being the average amount of energy received during each channel use and σ^2 being the noise variance. The underlying continuous-time communication system (operating over an additive white Gaussian noise channel with two-sided power spectral density $\frac{\mathsf{N}_0}{2}$) from which we extracted this model had a single unit-energy signal pulse $\psi(t)$ of finite duration T. More specifically, $\psi(t) = 0$ for $t < 0$ or $t > T$, and also

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \, dt = \int_0^T |\psi(t)|^2 \, dt = 1.$$

The transmitter signal was a sequence of pulses of different amplitudes, with each pulse delayed by T seconds from the previous pulse. Thus, the transmitted signal was of the form

$$s(t) = \sum_{i=0}^{n-1} \mathbb{X}_i \sqrt{\mathcal{E}}\psi(t - iT).$$

so that (ignoring channel attenuation) the total energy received over nT seconds is

$$\int_0^{nT} |s(t)|^2 \, dt = \mathcal{E}\sum_{i=0}^{n-1} \mathbb{X}_i^2.$$

The receiver consisted of a matched filter (with sampling time $T_0 > T$) for $\psi(t)$. Thus, the impulse response of the receiver filter was $h(t) = \psi(T_0 - t)$, and the sampled value of the filter response to $s(t)$ at time $t = T_0 + iT$ was a Gaussian random variable with mean $\mathbb{X}_i\sqrt{\mathcal{E}}$ and variance $\sigma^2 = \frac{\mathsf{N}_0}{2}\int_0^T |\psi(t)|^2 \, dt = \frac{\mathsf{N}_0}{2}$. These random variables were independent Gaussian random variables. The average power constraint requires that *on average*, a fixed amount of energy $n\mathcal{E}$ be received over the nT second interval, that is,

$$n\mathcal{E} = \mathsf{E}\left[\mathcal{E}\sum_{i=0}^{n-1} \mathbb{X}_i^2\right] = \mathcal{E}\sum_{i=0}^{n-1} \mathsf{E}[\mathbb{X}_i^2] \Rightarrow \sum_{i=0}^{n-1} \mathsf{E}[\mathbb{X}_i^2] = n.$$

With energy $n\mathcal{E}$ being received on average over a nT second interval, the average received power P is given by $\mathsf{P} = n\mathcal{E}/(nT) = \mathcal{E}/T$. Thus, we can re-write (14.1) as

$$C = \frac{1}{2T}\log_2\left(1 + \frac{\mathsf{P}T}{\mathsf{N}_0/2}\right) \text{ bits/second} \tag{14.2}$$

where the units of measurement have also been changed from bits per channel use to bits per second: each channel use takes T seconds.

14.2. Multiplexed Signaling Schemes

Suppose that there exists another unit-energy signal $\hat{\psi}(t)$ with the property $\hat{\psi}(t) = 0$ for $t < 0$ or $t > T$. Then, we could use $\hat{\psi}(t)$ in place of $\psi(t)$ as described in Section 14.1, with a signal

$$\hat{s}(t) = \sum_{i=0}^{n-1} \hat{\mathbb{X}}_i \sqrt{\mathcal{E}} \hat{\psi}(t - iT)$$

transmitting energy $n\mathcal{E}$ on average, etc. and achieving capacity given by (14.1) and so on and so forth. Now suppose further that $\psi(t)$ and $\hat{\psi}(t)$ are *orthogonal* signals, that is,

$$\int_{-\infty}^{\infty} \psi(t)\hat{\psi}(t)\,dt = \int_{0}^{T} \psi(t)\hat{\psi}(t)\,dt = 0.$$

If the signal $s(t) + \hat{s}(t)$ is transmitted and demodulated by the receivers with matched filters $\psi(T_0 - t)$ and $\hat{\psi}(T_0 - t)$, then each receiver demodulates the signal intended for it without any interference from the other signal! To see this, note that the response of the filter with impulse response $\psi(T_0 - t)$ to the signal $\hat{\psi}(t)$ at $\tau = T_0$ is

$$\psi(T_0 - t) \star \hat{\psi}(t)\Big|_{\tau=T_0} = \int_{-\infty}^{\infty} \psi(T_0 - t)\hat{\psi}(T_0 - t)\,dt = \int_{-\infty}^{\infty} \psi(t)\hat{\psi}(t)\,dt = 0$$

and similarly for other time instants $T_0 + iT$ and the other receiver.[1] That is, we can *multiplex* signals $s(t)$ and $\hat{s}(t)$ over the same physical medium (in other words, transmit the sum) and recover the data in each signal, with each receiver being oblivious to the existence of the other signal in its input. Generally, if we have M orthogonal signals, we can multiplex all of them and thus increase the capacity to as large a value as we desire. There is a minor glitch, though. With the average power being constrained to be P, this power must be shared among the M transmissions, and thus with M multiplexed signals, we can get a capacity of

$$C = \frac{M}{2T} \log_2\left(1 + \frac{(\mathsf{P}/M)T}{\mathsf{N}_0/2}\right) \text{ bits/second.} \tag{14.3}$$

So, it would seem that there is a free lunch after all; we can increase capacity by just multiplexing more and more signals. Unfortunately, the lunch is not free. As we use more and more orthogonal signals, we use more and more of a very precious resource – *bandwidth.* This notion is discussed next.

14.3. Orthogonal Signals and Signal Bandwidth

The late U.S. Supreme Court Justice Potter Stewart once rendered a decision in which he memorably said

"I may not know how to define it legally, but I know it when I see it."

He was, of course, speaking of pornography, but the notion of *bandwidth* of a signal is very similar. Every engineer understands the notion, perhaps only at an intuitive level, but a definition of bandwidth that is satisfactory for all purposes remains as elusive as ever.

A signal $s(t)$ is said to be *strictly time-limited* if $s(t)$ is 0 outside a time interval of finite length, e.g. $s(t) = 0$ if $t < 0$ or $t > T$. It is said to be *strictly band-limited* if its Fourier transform $S(f)$ is 0 outside a frequency interval of finite length, e.g. $S(f) = 0$ if $|f| > W$, or $S(f) = 0$ if $|f| > f_c + W/2$, or $|f| < f_c - W/2$ corresponding respectively to lowpass or bandpass signals of bandwidth W. Unfortunately, a signal cannot be both strictly time-limited and strictly band-limited. Now, let $s(t)$ be a *unit-energy* strictly time-limited function. In particular, suppose that $s(t) = 0$ if $t < 0$ or if $t > T$. Let η denote a *small* positive number, that is, $0 < \eta \ll 1$. Let W be a number such that

$$\int_{-W}^{W} |S(f)|^2\,df > 1 - \eta,$$

[1] The orthogonality of $\psi(T_0 - t)$ and $\hat{\psi}(T_0 - t)$ also ensures that the Gaussian random variables in the two receivers are (jointly Gaussian) uncorrelated random variables and thus independent Gaussian random variables.

that is, *almost all* of the energy in $s(t)$ is in the frequency band $[-W, W]$. We say that $s(t)$ is *essentially band-limited* to W Hz or an essentially low-pass signal of bandwidth W Hz. Now, suppose that $\{s(t)\}$ denotes a collection of unit-energy signals that are all strictly time-limited to $[0, T]$ and all essentially band-limited to $[-W, W]$. How many mutually orthogonal signals does the set $\{s(t)\}$ contain? The answer is given by a result called the Landau-Pollak Theorem. The number of orthogonal signals is less than $2WT/(1-\eta)$. Since η is very close to 0, the denominator is just slightly less than 1, and so we conclude that the number of orthogonal signals strictly time-limited to $[0, T]$ and essentially band-limited to W Hz is approximately $M = 2WT$.

14.4. Capacity of the Band-Limited AWGN Channel

Since approximately $M = 2WT$ orthogonal signals time-limited to T seconds and essentially band-limited to W Hz can be found, these can be multiplexed to get reliable communication over the AWGN channel. Substituting into (14.3), we get Shannon's celebrated result.

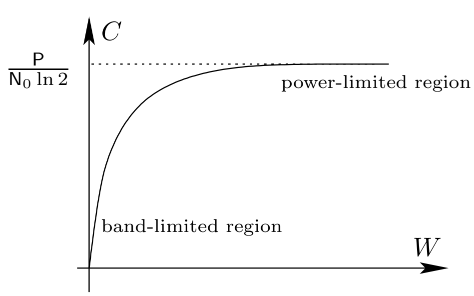
> If signals of average power P watts are used over an additive white Gaussian noise channel with two-sided power spectral density $\mathsf{N}_0/2$ watts/Hz and bandwidth W Hz, then the channel capacity is
> $$C = W \log_2\left(1 + \frac{\mathsf{P}}{\mathsf{N}_0 W}\right) \text{ bits/second.}$$

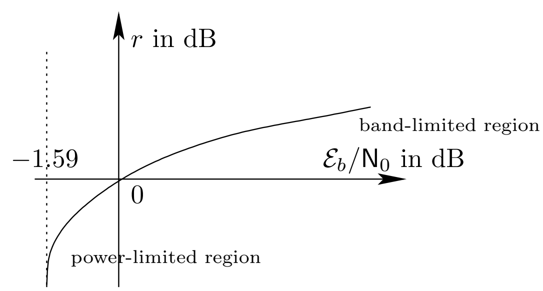
Reliable communication is possible over the band-limited additive white Gaussian noise channels at all rates less than C. Conversely, any communication scheme that attempts to communicate at rates above capacity is doomed to suffer large error probability.

When W is small, the capacity is approximately $W \log_2(\mathsf{P}/\mathsf{N}_0 W)$ and increases very rapidly and almost linearly with W. On the other hand, when W is large, the capacity increases very slowly with W, and approaches the limiting value $\dfrac{\mathsf{P}}{\mathsf{N}_0 \ln 2} = \dfrac{\mathsf{P}}{\mathsf{N}_0} \log_2 e$ as $W \to \infty$. See the left-hand figure below. Since reliable communication is possible at rates close to capacity, that is, nearly C bits are being transmitted each second with energy $\mathsf{P} \times 1 = \mathsf{P}$, the *energy per bit* $\mathcal{E}_b$ satisfies the inequality $\mathcal{E}_b > \mathsf{P}/C = \mathsf{N}_0 \ln 2$. We saw previously in Lecture 8 that for M-ary orthogonal signaling, $\mathcal{E}_b$ must be larger than $2\sigma^2 \ln 2 = \mathsf{N}_0 \ln 2$ for reliable communication. The result here shows that this is not just a peculiarity of M-ary orthogonal signaling. *Any* communication scheme must have a *bit-energy-to-noise-density* ratio $\mathcal{E}_b/\mathsf{N}_0 > \ln 2 \approx 0.69$ or ≈ -1.59 dB in order to communicate data reliably over an AWGN channel. Let $r = R/W$ denote the *spectral bit rate* of a communication scheme transmitting R bits per second over an AWGN channel of bandwidth W. Then, since $\mathcal{E}_b = \mathsf{P}/R$, we get that r must satisfy

$$r < \log_2\left(1 + r\frac{\mathcal{E}_b}{\mathsf{N}_0}\right).$$

The graph of the function $r = \log_2(1 + r\mathcal{E}_b/\mathsf{N}_0)$ as a function of $\mathcal{E}_b/\mathsf{N}_0$ is as shown in the right-hand figure below. Note that $r \to 0$ ($-\infty$ dB) as $\mathcal{E}_b/\mathsf{N}_0 \to \ln 2$ and that $r = 1$ (0 dB) at $\mathcal{E}_b/\mathsf{N}_0 = 1$ (0 dB).





The rate R or spectral rate r must lie below the curves shown in the above two figures.