So,

$$h(X) \leq \frac{1}{2} \log (2\pi e)^n |K|$$

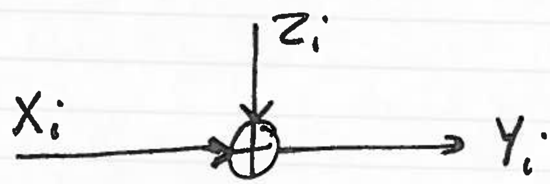
X Lecture 8, Oct. 21, 2003

Gaussian Channel

The channel we consider here is a discrete time channel with inputs $X_1, X_2, \ldots$ and output $Y_1, Y_2, \ldots$ where

$$Y_i = X_i + Z_i$$

where $Z_i \sim N(0, N)$



if the noise power (variance) N is zero or the transmission power is limitless, then it is possible to transmit an infinite number of bits per use. In such a case (unconstrained) capacity of the channel is infinite. In practice, however, there is a limit on the transmi power, e.g., for any Codeword $(x_1, x_2, \ldots, x_n)$ transmitted over the channel, we may require

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$$

Then, there is a limit to how close the codeword can be to each other and, therefore, a limit on the number of Codewords and consequently, the rate. For example, we can take a system using two symbols (Codewords) with two magnitudes $\pm\sqrt{P}$ for 0 and 1, respectively. This turns the Gaussian channel to a BSC with crossover probability,

$$\varepsilon = P_e = \int_{\sqrt{P}}^{\infty} \frac{1}{\sqrt{2\pi N}} e^{-\frac{x^2}{2N}} dx = Q\left(\sqrt{\frac{P}{N}}\right)$$

Capacity of Gaussian Channel:

Definition: The information capacity of the Gaussian Channel with power constraint P is:

$$C = \max_{p(x):\, E[X^2] \leq P} I(X;Y)$$

Calculating the information Capacity:

$$I(X;Y) = h(Y) - h(Y|X)$$
$$= h(Y) - h(X+Z|X)$$
$$= h(Y) - h(Z|X)$$
$$= h(Y) - h(Z)$$

Since $Z \sim N(0, N)$, $h(Z) = \frac{1}{2}\log(2\pi e N)$

Also:

$$E[Y^2] = E[(X+Z)^2] = E[X^2] + 2\cancelto{0}{E[X]}E[Z] + E[Z^2]$$
$$= P + N$$

Y ishas Variance $P+N$ we have

$$h(Y) \leq \frac{1}{2}\log 2\pi e(P+N)$$

and, therefore,

$$I(X;Y) \leq \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi e N$$
$$= \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$$

with equality if Y is normal, this occurs when X is Gaussian with $E[X^2] = P$. So,

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right),$$

Now, we will ~~present~~ present and prove the channel coding and its converse. Before doing this, we ~~give~~ ~~present~~ present the definition of a (M, n) code and achievability of a rate.

Definition: A (M, n) code for a Gaussian channel with power constraint P consists of the following:

1) An index set $\{1, 2, \dots, M\}$.

2) An encoding function $x: \{1, 2, \dots, M\} \to \mathcal{X}^n$ yielding codewords $x^n(1), x^n(2), \dots, x^n(M)$, satisfying the power constraint P, i.e., for every codeword

$$\sum_{i=1}^{n} x_i^2(w) \leq nP \qquad w = 1, 2, \dots, M.$$

3) A decoding function

$$g: \mathcal{Y}^n \to \{1, 2, \dots, M\}$$

The rate of the code is $R = \frac{\log_2 M}{n}$

Definition: A rate R is achievable if there exists a sequence of $(2^{nR}, n)$ codes with codewords satisfying the power constraint such that the maximal probability of error $\lambda^{(n)} \to 0$. The capacity of the channel is the supremum of the achievable rates.

We now present the channel coding theorem for the Gaussian channel indicating that the capacity of this channel (the supremum achievable rate) is the <u>information capacity</u> found before.

<u>Theorem</u> : The capacity of a Gaussian channel with power constraint P is

$$C = \frac{1}{2} \log\left(1 + \frac{P}{N}\right) \text{ bits per transmission.}$$

To prove this theorem we need to prove two things:

1) That for every $R < C$, there is a code sequence of $(n, 2^{nR})$ codes with $\lambda^{(n)} \to 0$. This is usually called the channel coding theorem.

2) That for any code with $\lambda^{(n)} \to 0$ we need to have $R \leq C$. This is usually called converse to the channel Coding Theorem.

Proof of the achievability:

The sequence of events involved in the coding/decoding is:

1) Codebook generation: We generate 2^{nR} codewords each of length n by generating $X_i(w)$, $i=1,2,\dots,n$, $w=1,2,\dots,2^{nR}$ i.i.d. according to a Gaussian distribution with zero mean and variance $P-\varepsilon$.

Since $\frac{1}{n}\sum_{i=1}^{n} X_i^2(w) \longrightarrow P-\varepsilon$, the probability that a codeword ~~has~~ violates the power constraint is arbitrarily small. The codebook $\underline{X}^n(1), \dots X^n(2^{nR})$ is revealed to both transmitter and receiver

2) Encoding: To send the message index w, the transmitter sends $X^n(w)$.

3) The receiver, upon receiving Y^n checks the codebook. If there is a unique $X^n(\hat{w})$ that is jointly typical with Y^n, (and satisfying the power constraint) the decoder ~~decdes~~ decides in favour of $\hat{w}$. Else, i.e., 1) if there is no codeword jointly typical with Y^n, or 2) there is more than one 3) the ~~co~~ codeword found violates the power constraint, the decoder declares failure.

4) Analysis of probability of error: Without loss of generality assume that $W=1$ was sent.

Thus $y^n = x^n(1) + z^n$.

Define the events

$$E_0 = \{\frac{1}{n}\sum_{i=1}^{n} x_i^2(1) > P\}$$

and

$$E_i = \{(x^n(i), y^n) \in A_\varepsilon^{(n)}\}$$

Then,

$$P(\mathcal{E}|W=1) = P(\mathcal{E}) = P(E_0 \cup E_1^c \cup E_2 \cup \cdots \cup E_{2^{nR}})$$
$$\leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$

By the AEP, $P(E_1^c) \to 0$, i.e., for large n

$$P(E_1^c) \leq \varepsilon$$

So,

$$P_e^{(n)} = P(\mathcal{E}) = P_r(\mathcal{E}|W=1) \leq P(E_0) + P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i)$$
$$\leq \varepsilon + \varepsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\varepsilon)}$$
$$= 2\varepsilon + (2^{nR}-1)\, 2^{-n(I(X;Y)-3\varepsilon)}$$
$$\leq 3\varepsilon \quad \text{provideded that } R < I(X;Y) - 3\varepsilon.$$

Now, picking the best codebook among all random codebooks and deleting the worst half of codewords, we obtain a codebook with low maximal probability of error. This code achieves a rate arbitrarily close to capacity and has an arbitrarily small maximal probability of error.

<u>Proof</u> of the converse to channel coding theorem:

We want to show that if $P\,\lambda^{(n)} \to 0$ then R has to be less than or equal to C, i.e.,

$$R \leq \frac{1}{2}\log\left(1+\frac{P}{N}\right).$$

<u>Proof</u>: if $\lambda^{(n)} \to 0$ then $P_e^{(n)} \to 0$.

$$nR = H(W) = I(W;\hat{W}) + H(W|\hat{W})$$

but $H(W|\hat{W}) \leq 1 + nRP_e^{(n)} = n\varepsilon_n$ where $\varepsilon_n \to 0$

so,

$$\begin{aligned} nR &\leq I(W;\hat{W}) + n\varepsilon_n \\ &\leq I(X^n;Y^n) + n\varepsilon_n \\ &= h(Y^n) - h(Y^n|X^n) + n\varepsilon_n \\ &= h(Y^n) - h(Z^n) + n\varepsilon_n \\ &\leq \sum_{i=1}^{n} h(Y_i) - h(Z^n) + n\varepsilon_n \\ &= \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Z_i) + n\varepsilon_n \\ &= \sum_{i=1}^{n} I(X_i;Y_i) + n\varepsilon_n \end{aligned}$$

$$nR \leq \sum_{i=1}^{n} [h(Y_i) - h(Z_i)] + n\varepsilon_n$$

Since $Y_i = X_i + Z_i$, the variance Y_i is $P_i + N$

where

$$P_i = \frac{1}{2^{nR}} \sum_{w} x_i^2(w)$$

so,

$$h(Y_i) \leq \frac{1}{2} \log 2\pi e (P_i + N)$$

and $$h(Z_i) = \frac{1}{2} \log 2\pi e N$$

so,

$$nR \leq \sum \left(\frac{1}{2} \log 2\pi e (P_i + N) - \frac{1}{2} \log 2\pi e N\right) + n\varepsilon_n$$

$$= \sum_{i=1}^{N} \frac{1}{2} \log \left(1 + \frac{P_i}{N}\right)$$

or

$$R \leq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} \log \left(1 + \frac{P_i}{N}\right) + \varepsilon_n \leq \frac{1}{2} \log \left(1 + \frac{1}{n} \sum_{i=1}^{n} \frac{P_i}{N}\right) + \varepsilon_n$$

$$\leq \frac{1}{2} \log \left(1 + \frac{P}{N}\right) + \varepsilon_n$$

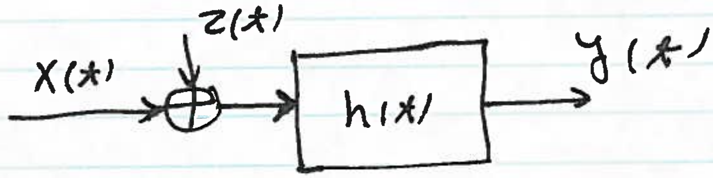
where we have used Jensen's inequality due to the fact that $f(x) = \frac{1}{2} \log(1 + x)$ is a concave function.

Jensen's inequality: for a convex function:

$$E[f(X)] \geq f(E(X)).$$

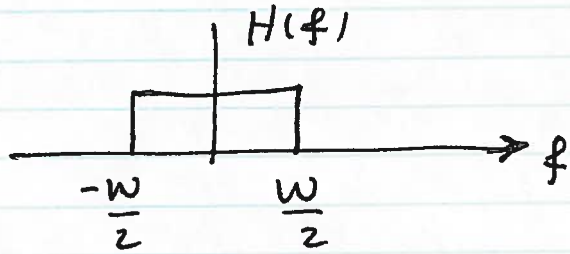
Band limited Gaussian Channel:

$$Y(t) = (X(t) + Z(t)) * h(t)$$

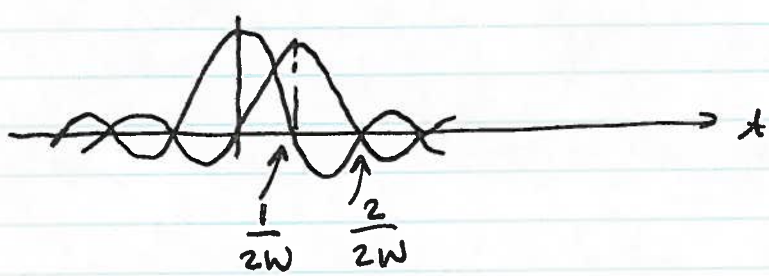


if the bandwidth of the channel is W, i.e.,

$$H(f) = 0 \qquad |f| > \frac{W}{2}$$



then Signalling rate $R_s \leq 2W$. That is, we can send a signal every $T_s = \frac{1}{R_s} \geq \frac{1}{2W}$.



So,

$$C = 2W\left[\frac{1}{2}\log\left(1+\frac{P}{N}\right)\right] = W\log\left(1+\frac{P}{N}\right)$$

and $N = N_0 W$. Also, $P = E_b R$

So

$$R \leq W \log\left(1 + \frac{E_b}{N_0} \times \frac{R}{W}\right)$$

When $R = C$, we have

$$\frac{R}{W} = \log\left(1 + \frac{E_b}{N_0} \times \frac{R}{W}\right)$$

or

$$\frac{R}{W} \cdot \frac{E_b}{N_0} = 2^{R/W} - 1$$

$$\frac{E_b}{N_0} = \frac{2^{R/W} - 1}{R/W}$$

$\eta = \frac{R}{W}$ in bits/sec./Hz. is the signalling (modulation) efficiency.

Take BPSK where $\frac{R}{W} = 1$ bps/Hz. at most

$\frac{R}{W} = 1 \Rightarrow \frac{E_b}{N_0} = 1 = 0$ dB

But with $\frac{E_b}{N_0} = 0$ dB one gets $BER \geq 0.08$.

To get $BER = 10^{-5}$, we need $\frac{E_b}{N_0} = 9.6$ dB, i.e., we are ≈ 9.6 dB off from the theoretical value (even if we accept 10^{-5} as zero). Using a rate $\frac{1}{2}$ code (convolution code) and QPSK modulation, we get 10^{-5} at $\frac{E_b}{N_0} = 4.4$ dB (an improvement of 5.2 dB at the expense of doubling the Bandwidth

Just using QPSK results in being off by 9.6-1.7:

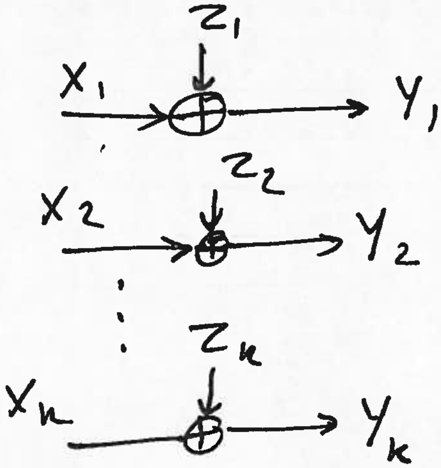
at $\frac{R}{W} = 2$ bps/Hz. Using a Turbo Codes with blocklength of 65000 bits gives $BER = 10^{-5}$ at $\frac{E_b}{N_0} = 0.7\,dB$ only $0.7\,dB$ off the Shannon limit.

In the limit where $\frac{R}{W} \to 0$, we have

$$\lim_{\frac{R}{W}\to 0} \frac{E_b}{N_0} = \lim \frac{2^{R/W} \ln 2}{1} = \ln 2 = -1.6\,dB$$

Lecture 9, Oct. 28, 2003

Parallel Gaussian Channels:



$$Y_i = X_i + Z_i \qquad i = 1, 2, \cdots k$$

$$Z_i \sim N(0, N_i)$$

w.r.t. constraint

$$E\left[\sum_{i=1}^{k} X_i^2\right] \le P$$

$$C = \max_{f(x_1, \cdots x_k) : \sum_i E X_i^2 \le P} I(X_1, \cdots X_k ; Y_1, \cdots Y_k)$$