Advanced Wireless Communications (Lecture 9: Capacity of Time-Invariant AWGN Channels)

Dr. Laxminarayana S Pillutla  
DA-IICT, Gandhinagar, Gujarat, India.  
Email: laxminarayana_pillutla@daiict.ac.in  

October 30, 2009
Acknowledgment

- The material in this lecture is based on the book “Fundamentals of Wireless Communications” by David Tse (University of California, Berkeley) and Pramod Viswanath (University of Illinois, Urbana Champaign).
- Another good reference is “Elements of Information Theory” by Thomas Cover and Joy Thomas.
Capacity of Wireless Channels

- The notion of *channel capacity* was introduced by *Claude Shannon* in 1948.
  - In his pioneering work Shannon showed a surprising result that, it is possible to communicate *reliably* at a rate less than or equal to *channel capacity*.
  - This can be done by coding intelligently.
  - Until then it was believed that the only way for reliable communication is to reduce the transmission rate (by say repetition coding).

- **Formal definition of capacity:**

\[
C = \max_{p(x): \mathbb{E}(c(x)) \leq A} I(x; y),
\]

where \(I(x; y)\) denotes the mutual information between \(x\) and \(y\).

- **Different notions of capacity:**
  - Ergodic capacity
  - Outage capacity.

- The above two notions of capacity are applicable for fading channels.
Capacity of Additive White Gaussian Noise (AWGN) channel

- Here we aim to derive the capacity of a power-constrained AWGN channel modelled as:

\[
y = x + n
\]

(2)

- Now to compute the capacity \( C \) in (1), we need to compute the mutual information as follows:

\[
I(x; y) = h(y) - h(y|x) \text{ (by definition)}
\]

\[
= h(y) - h(w) \tag{3}
\]

where \( h(.) \) denote the corresponding differential entropies. The (3) follows since given \( x, y \) has the same distribution as \( w \) with mean shifted by \( x \).

- Recall from the lecture on infotheory basics that the differential entropy of a Gaussian rv is equal to \( (1/2) \ln(2\pi e\sigma^2) \), \( \sigma^2_x \) denotes the variance of rv \( x \). Therefore

\[
h(w) = \frac{1}{2} \ln(2\pi e\sigma^2) \tag{4}
\]

where \( \sigma^2 \) is the variance of the noise \( w \).
Capacity of Additive White Gaussian Noise (AWGN) channel (ctd...)

- Therefore $I(x; y)$ in (3) can be expressed as
  \[ I(x; y) = h(y) - \frac{1}{2} \ln(2\pi e\sigma^2) . \]  
  (5)

- Hence the capacity of an AWGN channel $C_{\text{AWGN}}$ in (1) can be expressed as
  \[
  C_{\text{AWGN}} = \max_{p(x): \mathbb{E}(c(x)) \leq A} h(y) - \frac{1}{2} \ln(2\pi e\sigma^2)
  \]
  \[
  = \max_{p(x): \mathbb{E}(x^2) \leq P} h(y) - \frac{1}{2} \ln(2\pi e\sigma^2) \quad \text{(taking} \ c(x) = x^2) \]
  \[
  = \max_{p(y): \mathbb{E}(y^2) \leq P + \sigma^2} h(y) - \frac{1}{2} \ln(2\pi e\sigma^2) \quad \text{(taking} \ c(x) = x^2) \]
  (6)

  where (6) follows from the fact that the variance of $y \leq P + \sigma^2$.

- Thus the problem of finding the input distribution $p(x)$ has been transformed to that of finding the output distribution $p(y)$ in (6).
Capacity of Additive White Gaussian Channel (ctd...)

- Again recall from the information theory basics that a Gaussian distribution maximizes the differential entropy.
- Therefore $p(y) \sim \mathcal{N}(0, p + \sigma^2)$.
- Using the above result, in (6), we obtain $C_{\text{AWGN}}$ as

$$C_{\text{AWGN}} = \frac{1}{2} \ln \left( 2\pi e (P + \sigma^2) \right) - \frac{1}{2} \ln(2\pi e \sigma^2)$$

$$= \frac{1}{2} \ln(1 + \frac{P}{\sigma^2}) . \quad (7)$$

- Therefore capacity of an AWGN channel is $\frac{1}{2} \ln(1 + (P/\sigma^2))$, where $P$ denotes the transmit power and $\sigma^2$ denotes the noise variance.
  - The quantity $P/\sigma^2$ is nothing but the received signal-to-noise ratio (SNR).
Sphere Packing Interpretation of Channel Capacity of an AWGN Channel

- Consider transmissions with a codeword $x_i$ of length $N$. Suppose the code words come from a set $\mathcal{E}$ and all the code words are equally likely i.e.,

$$\mathcal{E} = \{x_1, x_2, \cdots, x_{|\mathcal{E}|}\} \tag{8}$$

where the operator $|.|$ denotes the cardinality of a given set.
- The energy of the $N$-dimensional received vector cannot be greater than $N(P + \sigma^2)$, therefore it lies in a sphere of radius $\sqrt{N(P + \sigma^2)}$ with high probability.
- If we try to partition the received vector space into $|\mathcal{E}|$ decision regions corresponding to each code word $x_i$, such that the regions do not intersect (so that the codewords can be decoded reliably).
  - Intuitively the above mentioned decision regions have to be spheres of equal volume.
  - For example, imagine that a circle can be divided into $N$ equal area circles only.
• Thus from the above explanation $|\epsilon| = V_y/V_N$, where $V_y$ corresponds to the volume of the received vector space and $V_N$ corresponds to the volume of the noise sphere.

• In general, the volume of a sphere is given by the formula $A_N r^N$, where $A_N$ is a constant and $r$ is the radius of the sphere.
  - For example for $N = 2$, i.e., the sphere corresponds to a circle, then $A_N = \pi$.
  - If $N = 3$, $A_N = 3$, i.e., the sphere corresponds to a three dimensional sphere, then $A_N = 4\pi/3$. 
Sphere Packing Interpretation of Channel Capacity of an AWGN Channel (ctd...)

- Thus, \( V_y = A_N \left( \sqrt{N(P + \sigma^2)} \right)^N \) and \( V_N = A_N \left( \sqrt{N\sigma^2} \right)^N \). Therefore

\[
|\varepsilon| = \frac{A_N \left( \sqrt{N(P + \sigma^2)} \right)^N}{A_N \left( \sqrt{N\sigma^2} \right)^N} \tag{9}
\]

- Therefore the number of bits per code word that can be communicated reliably is at most

\[
\frac{1}{N} \log(|\varepsilon|) = \frac{1}{N} \log \left( \frac{\left[ \sqrt{N(P + \sigma^2)} \right]^N}{\left[ \sqrt{N\sigma^2} \right]^N} \right) = \frac{1}{2} \log(1 + \frac{P}{\sigma^2}) \tag{10}
\]
Capacity of a Time-Invariant Single Input Multiple Output (SIMO) Channel

- Consider a SIMO channel i.e., with one transmit antenna and $L$ receive antennae. The received signal on the $l^{th}$ antenna at time instant $m$ denoted as $y_l[m]$ can be written as

$$y_l[m] = h_l x[m] + w_l[m], \quad l = 1, \ldots, L, \quad (11)$$

where $h_l$ denotes the fixed channel gain from the transmitter to the $l^{th}$ receive antenna, $x[m]$ denotes the transmitted symbol at time instant $m$ and $w_l[m]$ denotes the background noise at the $l^{th}$ receive antenna at time instant $m$. For our purposes we assume that $w$ is independent both spatially and temporally.

- The (11) can be written compactly in vector form as

$$y[m] = hx[m] + w[m], \quad (12)$$

where $y[m] = [y_1[m], \ldots, y_L[m]]^T$, $h = [h_1, \ldots, h_L]^T$ and $w[m] = [w_1[m], \ldots, w_L[m]]^T$. 
Capacity of a Time-Invariant Single Input Multiple Output (SIMO) Channel (ctd...)

- Note that the temporal and spatial independence of \( \{w_i[m]\}_{i=1}^L \) implies that \( E[w[m]w[m]^H] = N_0 I \), where \( N_0 \) denotes the noise variance of \( w_i[m], i = 1, \cdots, L \).

- Now, suppose we pre-multiply (12) by \( U^H \), where \( U \) is a unitary matrix (its columns are orthogonal to each other and have unit norm) and \( H \) denotes the Hermitian operator. The matrix \( U \) has the following form:

\[
U = [v, v_2, \cdots, v_L], \text{ where } v = \frac{h}{||h||}. \tag{13}
\]

- Therefore

\[
U^H y[m] = \begin{bmatrix}
    v^H h \\
    v_2^H h \\
    \vdots \\
    v_L^H h
\end{bmatrix} x[m] + \begin{bmatrix}
    v^H w[m] \\
    v_2^H w[m] \\
    \vdots \\
    v_L^H w[m]
\end{bmatrix}. \tag{14}
\]
Capacity of a Time-Invariant Single Input Multiple Output (SIMO) Channel (ctd...)

- Now notice that $\mathbf{v}^H \mathbf{h} = \frac{\mathbf{h}^H \mathbf{h}}{||\mathbf{h}||} = ||\mathbf{h}||$. Also $\mathbf{v}_i^H \mathbf{h} = \mathbf{v}_i^H \mathbf{v} ||\mathbf{h}|| = 0$ (follows by definition of $\mathbf{v}$ and since $\mathbf{v}_i$ is orthogonal to $\mathbf{v}$).
- Therefore (14) becomes

$$
\mathbf{U}^H \mathbf{y}[m] = \begin{bmatrix}
||\mathbf{h}|| x[m] \\
0 \\
\vdots \\
0
\end{bmatrix} + \begin{bmatrix}
\mathbf{v}_1^H \mathbf{w}[m] \\
\mathbf{v}_2^H \mathbf{w}[m] \\
\vdots \\
\mathbf{v}_L^H \mathbf{w}[m]
\end{bmatrix} \quad (15)
$$

- From (15), we note that the signal $x[m]$ is contained only in the first component of the transformed signal $\mathbf{U}^H \mathbf{y}$. The other components contain only the noise.
- Therefore it suffices to consider only the first component in (15) and accordingly is a **sufficient statistic to estimate** $x[m]$. 

Capacity of a Time-Invariant Single Input Multiple Output (SIMO) Channel (ctd...)

- Thus

\[ v^H y[m] = ||h|| x[m] + v^H w[m] \]
\[ = h^H y[m] = ||h||^2 x[m] + h^H w[m] \]  

(16)

where (16) follows upon substituting \( v = \frac{h}{||h||} \).

- From these analysis we transformed the SIMO channel into a scalar AWGN channel with received SNR \( P||h||^2/N_0 \), where \( P \) is the average energy per transmit symbol. Therefore the capacity of the SIMO channel can be computed as per (7) to be

\[ C_{SIMO} = 2 \times \frac{1}{2} \left[ \ln \left( 1 + \frac{P||h||^2}{N_0} \right) \right] . \]  

(17)

- The factor of 2 in (17) accounts for the complex channel.
- Multiple antennae therefore increase the capacity by increasing the received SNR, thereby providing power gain. For example, if \( L = 2 \), then power gain is 3 dB.
Consider a MISO channel i.e., with $L$ transmit antennae and one receive antennae. The received signal $y[m]$ for the time instant $m$ is equal to

$$y[m] = h^H x[m] + w[m]$$  \hspace{1cm} (18)

Recall from the SIMO experience

- Projections in orthogonal directions contain noise and is not helpful in signal detection.

Therefore a natural strategy would be to steer the signal in the direction of $h$. Therefore we set

$$x[m] = \frac{h}{||h||} \tilde{x}[m]$$  \hspace{1cm} (19)

With the transformation in (19), the MISO channel is reduced to the AWGN channel

$$y[m] = ||h|| \tilde{x}[m] + w[m]$$  \hspace{1cm} (20)
Therefore the capacity of a MISO channel is equal to the capacity of SIMO channel and is given as

\[ C_{\text{MISO}} = \ln \left( 1 + \frac{P||h||^2}{N_0} \right) . \] (21)

The strategy in (19) of aligning the transmit symbol in the direction of the channel is known as \textit{transmit beamforming}.

An obvious question that one can get is “Can we do better than in (19)”. The answer is NO because:

- It can be shown that \( P||h||^2 N_0 \) is the maximum received SNR under the transmit power constraint (can be shown by using Cauchy-Schwarz inequality on the expression for received SNR).
- Therefore the expression in (21) is indeed the capacity of the MISO channel.
Capacity of Frequency Selective Channels

• Consider a time-invariant $L$-tap frequency-selective AWGN channel:

$$y[m] = \sum_{l=0}^{L-1} h_l x[m - l] + w[m].$$  \hspace{1cm} (22)

• We assume that the input symbols $x[m]$ are subject to an average power constraint.

• In the last lecture on OFDM, we have seen that a frequency-selective channel can be converted into $N_c$ independent sub-carriers by adding a cyclic-prefix of length $L - 1$, where $L$ denotes the number of multi-paths.

• Now suppose that the operation is repeated over blocks of data symbols as shown in the figure of next slide:
The communication over the $i^{th}$ OFDM block can be written as

$$\tilde{y}_n[i] = \tilde{h}_n \tilde{d}_n[i] + \tilde{w}_n[i], \quad n = 0, 1, \ldots, N_c - 1$$  \hspace{1cm} (23)

Let

$$\tilde{d}[i] = [\tilde{d}_0[i], \ldots, \tilde{d}_{N_c-1}[i]]$$  \hspace{1cm} (24)

$$\tilde{w}[i] = [\tilde{w}_0[i], \ldots, \tilde{w}_{N_c-1}[i]]$$  \hspace{1cm} (25)

$$\tilde{y}[i] = [\tilde{y}_0[i], \ldots, \tilde{y}_{N_c-1}[i]]$$  \hspace{1cm} (26)

$$\tilde{h} \triangleq \sqrt{N_c} \text{DFT} [h]$$  \hspace{1cm} (27)
Capacity of Frequency Selective Channels (ctd...)

- If $N_c$ is large then the loss due to cyclic prefix can be made arbitrarily small.
- Therefore the capacity of the original frequency-selective channel is same as the capacity of the transformed channel as $N_c \to \infty$.
- The average power constraint on the input symbols for the $i^{th}$ block can be written as
  \[
  \frac{1}{N_c} \sum_{n=0}^{N_c-1} |d_n[i]|^2 \leq P \tag{28}
  \]
  \[
  \implies \mathbb{E} \left[ ||d_i||^2 \right] \leq P \quad \text{(since $N_c$ is large)} \tag{29}
  \]
- The power constraint in (28) (and hence in (29)) can be re-written in terms of the transformed vector $\tilde{d}[i]$ using the Parseval’s theorem for DFT.

- **Parseval’s theorem:**
  \[
  \int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(e^{j\omega})|^2 d\omega \quad \text{(for continuous FT)} \tag{30}
  \]
  \[
  \sum_{n=\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{j\omega})|^2 d\omega \quad \text{(for DTFT)} \tag{31}
  \]
  \[
  \sum_{n=0}^{N_c-1} |x(n)|^2 = \frac{1}{N_c} \sum_{n=0}^{N_c-1} |X(n)|^2 \quad \text{(for DFT)} \tag{32}
  \]
Capacity of Frequency Selective Channels (ctd...)

- From our definition of DFT

\[
\tilde{d}_n[i] = \frac{1}{\sqrt{N_c}} \sum_{i=0}^{N_c-1} d_n e^{-j\frac{2\pi in}{N_c}}
\]  

(33)

- Now by Parseval’s theorem for DFT

\[
\sum_{n=0}^{N_c-1} |d_n[i]|^2 = \frac{1}{N_c} \sum_{n=0}^{N_c-1} |\tilde{d}_n[i]|^2
\]

\[
\Rightarrow \sum_{n=0}^{N_c-1} |d_n[i]|^2 = \mathbb{E} \left[ ||\tilde{d}_i||^2 \right] \leq N_c P \text{ (since } N_c \text{ is large)}
\]  

(34)

- Since the transformed channels defined in (23) are non-interfering, therefore they constitute \( N_c \) parallel channels. These parallel channels are subject to a total power constraint given by (34)
A natural strategy for reliable communication over a parallel AWGN channel is to allocate the power $P_n$ to each $n^{th}$ sub-channel so that the total power constraint is met.

Upon allocating power a separate capacity achieving AWGN code can be used to communicate over each of the sub-channels. This is depicted in the following figure:
Capacity of Frequency Selective Channels (ctd...)

- With the independent coding strategy shown in the Fig. of the previous slide, the maximum reliable rate of communication is given by

\[
\sum_{n=0}^{N_c-1} \ln \left( 1 + \frac{P_n|\tilde{h}_n|^2}{N_0} \right)
\]

(35)

- The power allocation (i.e., determining the \( P_n \) values) can be chosen separately to maximize (35). Therefore the optimal power allocation, thus is the solution to the optimization problem below:

\[
C_{N_c} = \max_{[P_0, \ldots, P_{N_c-1}]} \sum_{n=0}^{N_c-1} \ln \left( 1 + \frac{P_n|\tilde{h}_n|^2}{N_0} \right)
\]

(36)

subject to

\[
\sum_{n=0}^{N_c-1} P_n = N_c P
\]

(37)

\[
P_n \geq 0 \quad , \quad n = 0, \ldots, N_c - 1
\]

(38)
The optimization problem defined in (36)- (38) is convex. Therefore, KKT conditions are necessary and sufficient.

KKT conditions for the optimization problem in (36) - (38):

\[ P_n^* \geq 0, \ n = 0, \cdots, N_c - 1 \]  
\[ \sum_{n=0}^{N_c-1} P_n^* - N_c P = 0 \]  
\[ \lambda_n^* \geq 0, \ n = 0, \cdots, N_c - 1 \]  
\[ \lambda_n^* P_n^* = 0, \ n = 0, \cdots, N_c - 1 \]  
\[ -\left( \frac{1}{1 + \frac{P_n^* |\tilde{h}_n|^2}{N_0}} \right) \frac{|\tilde{h}_n|^2}{N_0} + v^* - \lambda_n^* = 0, \ n = 0, \cdots, N_c - 1 \]  

where \( P_n^* \ (n = 0, \cdots, N_c - 1) \) and \( \lambda_n^* \ (n = 0, \cdots, N_c - 1) \), \( v^* \) are primal and dual optimal respectively.
Capacity of Frequency-Selective Channels (ctd...)

- The (43) implies

\[
\lambda_n^* = v^* - \frac{|\tilde{h}_n|^2}{N_0} \left( \frac{1}{1 + \frac{P_n^*|\tilde{h}_n|^2}{N_0}} \right), n = 0, \cdots, N_c - 1 \tag{44}
\]

\[
\implies v^* \geq \frac{|\tilde{h}_n|^2}{N_0} \left( \frac{1}{1 + \frac{P_n^*|\tilde{h}_n|^2}{N_0}} \right), n = 0, \cdots, N_c - 1. \tag{45}
\]

where (45) follows since \( \lambda_n^* \geq 0, n = 0, \cdots, N_c - 1. \)

- Now consider the case \( v^* < \frac{|\tilde{h}_n|^2}{N_0} \). This assumption together with (45) implies

\[
\frac{1}{P_n^* + \frac{N_0}{|\tilde{h}_n|^2}} \leq v^* < \frac{N_0}{|\tilde{h}_n|^2} \tag{46}
\]

- The inequality in (46) can only hold if \( P_n^* > 0 \). Therefore from the complementary slackness condition in (42), we get \( \lambda_n^* = 0. \)
Capacity of Frequency Selective Channels (ctd...)

- Using $\lambda_n^* = 0$ in (44), we obtain

$$v^* = \frac{1}{P_n^* + \frac{N_0}{|h_n|^2}}$$

$$\implies P_n^* = \frac{1}{v^*} - \frac{N_0}{|\tilde{h}_n|^2} \text{ if } v^* < \frac{|\tilde{h}_n|^2}{N_0} \quad (47)$$

- Now consider the case when $v^* \geq \frac{1}{\frac{N_0}{|h_n|^2}}$, then

$$v^* \geq \frac{1}{\frac{N_0}{|h_n|^2}} > \frac{1}{P_n^* + \frac{N_0}{|h_n|^2}} \quad (48)$$

$$\implies v^* - \frac{1}{P_n^* + \frac{N_0}{|\tilde{h}_n|^2}} > 0 \quad (49)$$
Thus (49) implies $\lambda_n^* > 0$ (see (44)), therefore from the complementary slackness condition $P_n^* = 0$. Hence

$$P_n^* = 0 \text{ if } v^* \geq \frac{1}{\frac{N_0}{|h_n|^2}} \quad (50)$$

Combining (47) and (50), we can write

$$P_n^* = \left( \frac{1}{v^*} - \frac{N_0}{|\tilde{h}_n|^2} \right)^+ \quad (51)$$

where $x^+ = \max(x, 0)$.

The Lagrange multiplier in (51) can be determined such that sum power constraint in (37) is met:

$$\frac{1}{N_c} \sum_{n=0}^{N_c-1} \left( \frac{1}{v^*} - \frac{N_0}{|\tilde{h}_n|^2} \right)^+ = P \quad (52)$$
Capacity of Frequency Selective Channels (ctd...)

- The optimal power allocation strategy shown in (51) is termed as the **water-filling or water-pouring** power allocation strategy.

- As per the waterfilling strategy more power is allocated to the sub-carrier with better channel conditions and less power is allocated to the sub-carrier with poor channel conditions *(which intuitively makes sense.)*
Capacity of Frequency Selective Channels (ctd...)

- Note that $\tilde{h}_n$ is $\sqrt{N_c}$ time the DFT of $\{h_l\}_{l=0}^{N_c-1}$ i.e.,

$$
\tilde{h}_n = \sum_{l=0}^{L-1} h_l \exp \left( -j\frac{2\pi ln}{N_c} \right).
$$

- In general the DFT is obtained by sampling the DTFT $H(f)$ at $f = nW/N_c$. Therefore

$$
H(f) = \sum_{l=0}^{L-1} h_l \exp \left( -j\frac{2\pi lf}{W} \right), \quad f \in [0, W].
$$

- Note that as $N_c$ grows, the frequency width $W/N_c$ of the sub-carriers goes to zero and thus they represent a finer sampling of the continuous spectrum. Therefore, the optimal power allocation in (51) converges to

$$
P^*(f) = \left( \frac{1}{v^*} - \frac{N_0}{|H(f)|^2} \right) +
$$
The $\nu^*$ in (55) is chosen such that

$$\int_0^W P^*(f)df = P$$  \hspace{1cm} (56)$$

The power allocation strategy in (55) can be interpreted as water-filling over frequency and is shown in the figure below:
Capacity of Frequency Selective Channels (ctd...)

- With $N_c$ sub-carriers, the largest reliable communication rate with independent coding is $C_{N_c}$ bits per OFDM symbol (or) bits/s/Hz, where $C_{N_c}$ is defined as in (35). Now as $N_c \to \infty$

\[
\frac{W C_{N_c}}{N_c} = \lim_{N_c \to \infty} \sum_{n=0}^{N_c-1} \frac{W}{N_c} \ln \left( 1 + \frac{P_n |\tilde{h}_n|^2}{N_0} \right)
\]

\[
C \xrightarrow{N_c \to \infty} \int_0^W \ln \left( 1 + \frac{P^*(f) |H(f)|^2}{N_0} \right)
\]

(57)

where (57) follows because as $N_c \to \infty$, $P_n^* \to P^*(f)$ and $\tilde{h}_n \to H(f)$.

- For large $N_c$, coding jointly across the sub-carriers cannot give any benefit in terms of probability of error. Therefore the expression in (57) corresponds to the capacity of the time invariant frequency selective AWGN channel.

- A more rigorous information theoretic derivation can be found in the Appendix B.6 of the book by Tse and Viswanath.