Basics of Communication - Information Theory

Monica Navarro, Stephan Pfletschinger

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)

Wireless Networks: From Energy Harvesting to Information Processing

European School of Antennas/WIPE-COST ACTION IC1301

9 – 13 Nov. 2015, Castelldefels, Spain

Centre
Tecnològic
de Telecomunicacions
de Catalunya

Outline

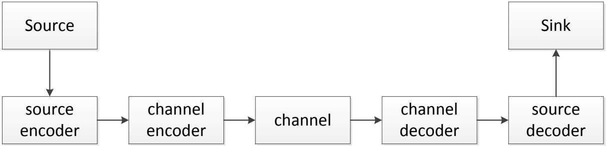
- Introduction
- Basic Concepts of Information Theory
    - Entropy and mutual information
    - Fundamental channel models
- Capacities of Single-User Channels
    - BSC, BEC
    - AWGN, AWGN with discrete input
    - Fading channels

Communication System

- Lets consider the simplified representation of a communication system
- Goal: achieve reliable transmission; recover (at the sink) the transmitted information from the source with as little distortion as possible.

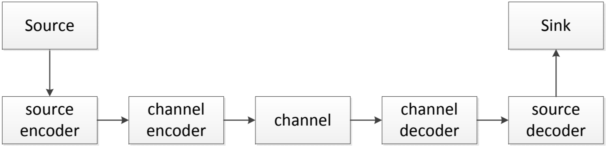
Communication System

- Lets consider the simplified representation of a communication system
- Goal: achieve reliable transmission; recover (at the sink) the transmitted information from the source with as little distortion as possible.
- source-channel separation theorem



Communication System

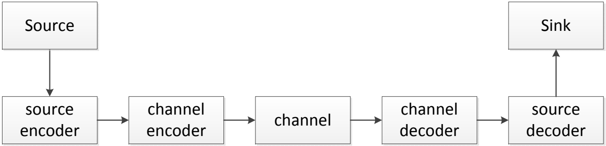
- Lets consider the simplified representation of a communication system
- Goal: achieve reliable transmission; recover (at the sink) the transmitted information from the source with as little distortion as possible.
- source-channel separation theorem



source coding theorem: for a given source and distortion measure, there exists a minimum rate $R(d)$ necessary (and sufficient) to describe this source with distortion $\leq d$.

Communication System

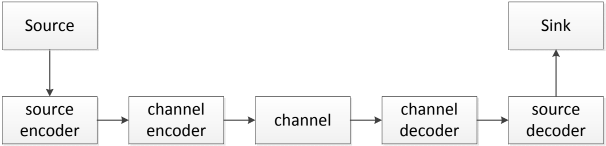
- Lets consider the simplified representation of a communication system
- Goal: achieve reliable transmission; recover (at the sink) the transmitted information from the source with as little distortion as possible.
- source-channel separation theorem



source coding theorem: for a given source and distortion measure, there exists a minimum rate $R(d)$ necessary (and sufficient) to describe this source with distortion $\leq d$.
channel coding theorem: there exist a maximum rate (bits per channel use) at which information can be transmitted reliably (probability of error $\rightarrow 0$) over a given channel. $\Rightarrow$ maximum rate: capacity of the channel C.

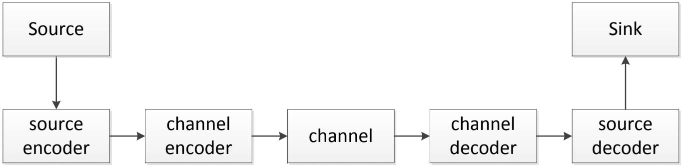
Communication System

- Lets consider the simplified representation of a communication system
- Goal: achieve reliable transmission; recover (at the sink) the transmitted information from the source with as little distortion as possible.
- source-channel separation theorem



Communication System

- Lets consider the simplified representation of a communication system
- Goal: achieve reliable transmission; recover (at the sink) the transmitted information from the source with as little distortion as possible.
- source-channel separation theorem



the source can be reconstructed at the receiver with a distortion of at most d if $R(d) < C$

Limits for channel coding: channel capacity

- The performance limits for channel coding are given by the channel capacity
- Results from information theory are mathematically exact, but have to be interpreted with caution:
  - channel capacity holds for infinite block length
  - channel models are highly simplified
  - decoding complexity is not considered
- Nevertheless useful to obtain guidelines for practical design

Basic Concepts of Information Theory

Entropy and Mutual Information

We consider a discrete random variable $X \in \mathbb{X} = \{x_1, x_2, \ldots, x_N\}$, where $\mathbb{X}$ denotes an alphabet of cardinality N. The probability mass function (pmf) is denoted by

$$p_i = p(x_i) = P[X = x_i], \quad \text{with } \sum_{i=1}^{N} p_i = 1$$

The entropy of the X is defined as

$$H(X) = \sum_{x \in \mathbb{X}} p(x) \mathrm{ld} \frac{1}{p(x)} = \mathop{\mathbb{E}}_{X} [-\mathrm{ld}\, p(X)] \quad (1)$$

- logarithms are base two ("logarithmus dualis"), the entropy is measured in bits
- $H(X)$ depends only on the distribution of X, i.e. the p_i, and not on the values of X itself, i.e. the x_i

Entropy and Mutual Information

- We can think of the entropy $H(X)$ as a measure of
  - the amount of "information" provided by an observation of X
  - our "uncertainty" of X
  - the "randomness" of X
- Properties of $H(X)$
  1. $0 \leq H(X) \leq \mathrm{ld} N$
  2. $H(X) = 0$ iff $p_i = 1$ for some i
  3. $H(X) = \mathrm{ld} N$ iff $p_i = 1/N$ for all i

  - $H(X)$ vanishes only if X is deterministic and it is maximum if X is uniformly distributed

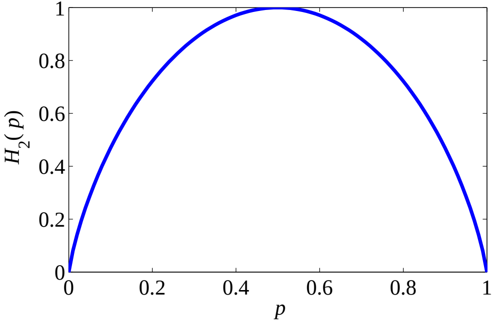
Entropy and Mutual Information

Binary entropy function

For $\mathbb{X} = \{0, 1\}$, $P[X = 0] = p$, we define

$$H_2(p) \triangleq -p\,\mathrm{ld}\,p - (1-p)\,\mathrm{ld}(1-p) \quad (2)$$

- $H_2(p)$ is a concave function in p
- attains its maximum at $p = 1/2$
- our uncertainty about a binary random variable is maximum if both outcomes are equiprobable



Now we consider two discrete RV, $X \in \mathbb{X}$ and $Y \in \mathbb{Y}$ and define the joint entropy and the conditional entropy

$$H(X,Y) = -\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\,\mathrm{ld}\,p(x,y) = -\mathbb{E}\left[\mathrm{ld}\,p(X,Y)\right] \quad (3)$$

$$H(X|Y) = \sum_{y\in\mathbb{Y}} p(y) H(X|Y=y) = -\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\,\mathrm{ld}\,p(x|y) \quad (4)$$

where $H(X|Y=y) \triangleq -\sum_{x\in\mathbb{X}} p(x|y)\,\mathrm{ld}\,p(x|y)$

The conditional entropy $H(X|Y)$ is our uncertainty about X after having observed Y.

- Conditioning reduces entropy:

$$H(X|Y) \leq H(X) \tag{5}$$

  - equality holds if and only if X and Y are independent
  - $H(X|Y=y)$ can be greater than $H(X)$, but on the average the knowledge of Y reduces our uncertainty of X

- Chain rule

$$H(X,Y) = H(X) + H(Y|X) \leq H(X) + H(Y) \tag{6}$$

$$H(\mathbf{X}) \leq \sum_{i=1}^{n} H(X_i) \tag{7}$$

where $\mathbf{X} = (X_1, X_2, \dots, X_n)$

- The mutual information between X and Y is

$$I(X;Y) = \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\,\mathrm{ld}\frac{p(x,y)}{p(x)p(y)} \quad (8)$$

$$= H(X) - H(X|Y) \quad (9)$$

$$= H(X) + H(Y) - H(X,Y) \quad (10)$$

  - $I(X;Y) \geq 0$ with equality iff X and Y are independent
  - $I(X;Y)$ is the reduction in uncertainty of X due to the knowledge of Y
  - If X is transmitted over a channel and received as Y, $I(X;Y)$ is the transmitted amount of information

Entropy and Mutual Information

- Mutual information for random vectors

$$
\begin{aligned}
I(X_1, X_2; Y) &= H(X_1, X_2) - H(X_1, X_2|Y) \\
I(\mathbf{X}; Y) &= H(\mathbf{X}) - H(\mathbf{X}|Y)
\end{aligned}
\tag{11}
$$

- Conditional mutual information

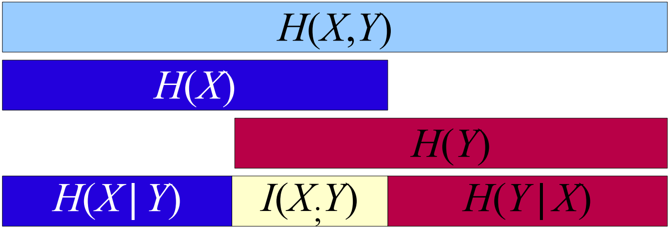
$$
I(X; Y|Z) = H(X|Z) - H(X|Y, Z) = \sum_{z \in \mathbb{Z}} p(z) I(X; Y|Z = z) \tag{12}
$$

- Chain rule

$$
\begin{aligned}
I(X_1, X_2; Y) &= I(X_1; Y) + I(X_2; Y|X_1) \\
&= I(X_2; Y) + I(X_1; Y|X_2)
\end{aligned}
\tag{13}
$$

Entropy and Mutual Information

Relationship between entropy, conditional entropy, joint entropy and mutual information.



Entropy and Mutual Information

For a continuous random variable, we can define the differential entropy

$$h(X) \triangleq -\int p(x) \mathrm{ld} p(x) \, \mathrm{d}x \tag{14}$$

The conditional differential entropy and the mutual information are defined accordingly

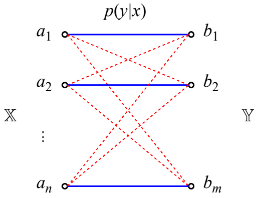
$$\begin{aligned} h(X|Y) &= -\iint p(x,y) \mathrm{ld} p(x|y) \, \mathrm{d}x \, \mathrm{d}y & (15) \\ I(X;Y) &= h(X) - h(X|Y) & (16) \end{aligned}$$

- $h(X)$ might be negative
- $h(X)$ for a discrete random variable is $-\infty$
- $I(X;Y) \geq 0$ like in the discrete case
- chain rules hold like in discrete case

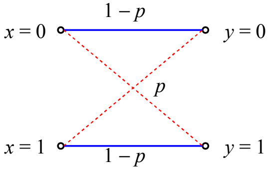
Basic Channel Models

Discrete memoryless channel

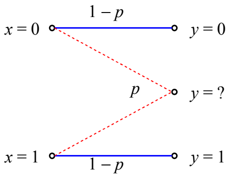
- The discrete memoryless channel (DMC)
  - is defined by an input alphabet $\mathbb{X}$, an output alphabet $\mathbb{Y}$ and the transition probabilities $p(y|x)$
  - memoryless means $p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i)$
  - The simplest examples are the binary symmetric channel (BSC) and the binary erasure channel (BEC)



Binary symmetric channel



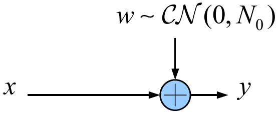
Binary erasure channel



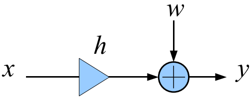
Basic Channel Models

Single-user channels

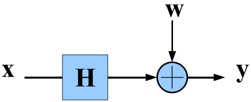
The following channels have continuous inputs and outputs.



- AWGN channel: $y = x + w$



- Rayleigh fading channel: $y = h \cdot x + w$

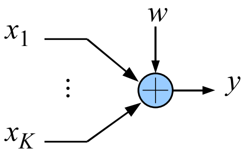


- Vector Gaussian channel: $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$

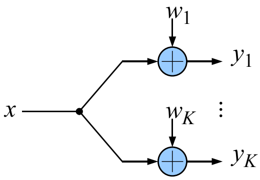
Basic Channel Models

Multi-user channels

- Multiple-access channel (MAC)
  - uplink
  - K users
  - K different messages
  - K power constraints

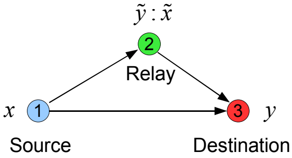


- Broadcast channel (BC)
  - downlink
  - K users
  - K different messages, one common message
  - one power constraint

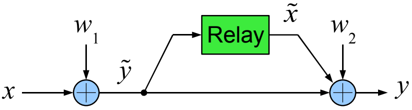


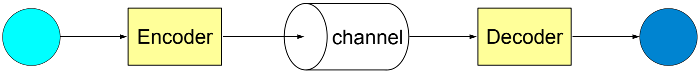
Basic Channel Models
Relay channel

- Relay channel, general model
  - Relay helps source to transmit its message
  - Capacity is unknown



- Degraded Gaussian relay channel
  - Capacity is known





An (n, M) codebook consists of

- an encoding function $u \rightarrow \mathbf{x} = (x_1, \ldots, x_n)$
- a decoding function $\mathbf{y} \rightarrow \hat{u}$

- Probability of error: $P_\mathrm{e}^n = P[u \neq \hat{u}]$
- The rate $R = \frac{\mathrm{ld} M}{n}$ is achievable if there exists a sequence of $(n, 2^{nR})$ codebooks such that $P_\mathrm{e}^n \rightarrow 0$ for $n \rightarrow \infty$.

$\Leftrightarrow$ for every $\epsilon > 0$ we can find an $(n, 2^{nR})$ codebook such that $P_\mathrm{e}^n < \epsilon$.

Random coding with ML decoding

- Generate (n, M) codebook: $\mathcal{C} = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(M)\}$
  - codebook is known to transmitter and receiver

$$\begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & & & \vdots \\ x_1(M) & x_2(M) & \cdots & x_n(M) \end{pmatrix}$$

- Coding procedure
  1. Select message $u \in \{1, 2, \dots, M\}$
  2. Transmit $\mathbf{x}(u) = (x_1(u), x_2(u), \dots, x_n(u))$
  3. Receive $\mathbf{y} = (y_1, \dots, y_n)$
  4. Maximum likelihood decoding: $\hat{u} = \arg\max_i p\left(\mathbf{y}|\mathbf{x}(i)\right)$

Capacities of Single-User Channels

- The channel capacity is the supremum of all achievable rates:

$$C = \sup R$$

Definition (Channel capacity)

The channel capacity of a channel with input X and output Y is given by

$$C = \max_{p(x)} I(X;Y)$$

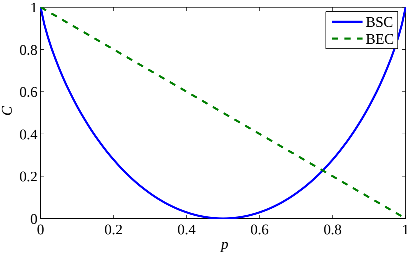
Channel Capacity of BSC and BEC

- BSC (binary symmetric channel)

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_{x \in \mathbb{X}} p(x) H(Y|X=x)$$
$$= H(Y) - \sum_{x \in \{0,1\}} p(x) H_2(p) = H(Y) - H_2(p)$$

  - $H(Y)$ is maximum for $p_x = 1/2$, then $C_{\mathrm{BSC}} = 1 - H_2(p)$

- BEC (binary erasure channel): $C_{\mathrm{BEC}} = 1 - p$
  - on average, p bits get lost



Channel Capacity of the AWGN

- AWGN channel with continuous input:

$$\text{real-valued: } y = x + w, \quad x \in \mathbb{R},\ w \sim \mathcal{N}(0, N_0/2) \quad (17)$$

$$\text{complex-valued: } y = x + w, \quad x \in \mathbb{C},\ w \sim \mathcal{CN}(0, N_0) \quad (18)$$

- the capacities are

$$C_{\mathrm{real}} = \frac{1}{2}\mathrm{ld}\left(1 + \frac{2E_{\mathrm{S}}}{N_0}\right), \quad C_{\mathrm{complex}} = \mathrm{ld}\left(1 + \frac{E_{\mathrm{S}}}{N_0}\right) \quad (19)$$

  - To achieve capacity, the channel input must be normal distributed: $x \sim \mathcal{N}(0, E_{\mathrm{S}})$ or $x \sim \mathcal{CN}(0, E_{\mathrm{S}})$.
  - this holds for the discrete-time channel, capacity is measured in bits per channel use
  - for the continuous-time channel, this corresponds to the spectral efficiency, measured in bps/Hz
  - Note: $N_0 = 2\sigma^2$ by convention

Channel Capacity of the AWGN

Outline of derivation for real-valued AWGN

- Power constraint: $\mathbb{E}\left[x^2\right] \leq E_\mathrm{S}$, mutual information: $I(X;Y) = h(Y) - h(Y|X)$

$$C = \max_{p(x):\mathbb{E}[x^2]\leq E_\mathrm{S}} \{h(Y) - h(Y|X)\}$$

- the differential entropy of a $\mathcal{N}(\mu, \sigma^2)$ distributed random variable is $\frac{1}{2}\mathrm{ld}(2\pi e\sigma^2)$, independent of μ. Then

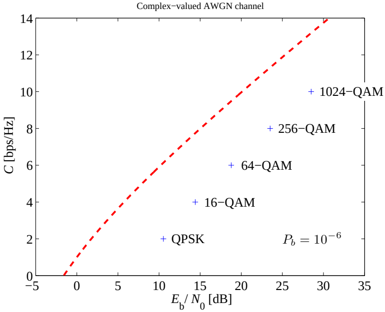
$$h(Y|X) = \int_{-\infty}^{\infty} p(x) h(\underbrace{Y|X=x}_{\sim\mathcal{N}(x,N_0/2)})\,\mathrm{d}x = h(W) = \frac{1}{2}\mathrm{ld}(\pi e N_0)$$

- The normal distribution maximizes the differential entropy for a given second moment $\Rightarrow$ Y is normal distributed $\Rightarrow$ $X \sim \mathcal{N}(0, E_\mathrm{S})$, $Y \sim \mathcal{N}(0, E_\mathrm{S} + \frac{N_0}{2})$, $h(Y) = \frac{1}{2}\mathrm{ld}\left(\pi e(2E_\mathrm{S} + N_0)\right)$ and finally

$$C = \frac{1}{2}\mathrm{ld}\left(1 + \frac{2E_\mathrm{S}}{N_0}\right)$$

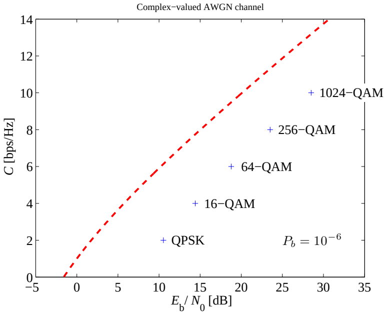
Channel Capacity of the AWGN

- Consider the real-valued AWGN with continuous input
  - E_S is the energy per transmitted symbol
  - in the context of channel coding, we often use the energy per bit $E_\mathrm{b} = \frac{E_\mathrm{S}}{R}$
  - the maximum rate is $R = \frac{1}{2}\mathrm{ld}\left(1 + \frac{2RE_\mathrm{b}}{N_0}\right)$, thus $\frac{E_\mathrm{b}}{N_0} = \frac{2^{2R}-1}{2R}$
  - for $R \to 0$, $\frac{E_\mathrm{b}}{N_0} \to \ln 2$, i.e. -1.59 dB



Channel Capacity of the AWGN

- Consider the real-valued AWGN with continuous input
  - E_{S} is the energy per transmitted symbol
  - in the context of channel coding, we often use the energy per bit $E_{\mathrm{b}} = \frac{E_{\mathrm{S}}}{R}$
  - the maximum rate is $R = \frac{1}{2}\mathrm{ld}\left(1 + \frac{2RE_{\mathrm{b}}}{N_0}\right)$, thus $\frac{E_{\mathrm{b}}}{N_0} = \frac{2^{2R}-1}{2R}$
  - for $R \to 0$, $\frac{E_{\mathrm{b}}}{N_0} \to \ln 2$, i.e. -1.59 dB
  - For $E_{\mathrm{b}}/N_0 < -1.59$ dB, no reliable transmission!



Channel Capacity of Parallel AWGN Channels

- Consider N inputs $X_1, X_2, \ldots, X_N$
- N outputs $Y_1, Y_2, \ldots, Y_N$
- each with $Y_i = X_i + W_i$, $W_i \sim \mathcal{N}(0, \sigma_i^2)$
- subject to a total power constraint

$$\mathbb{E} = \left[\sum_{i=1}^{N} X_i^2 \right] \leq P$$

- The capacity is given by

$$C = \sum_{i=1}^{N} \frac{1}{2} \mathrm{ld} \left(1 + \frac{P_i}{\sigma_i^2} \right)$$

- is achieved for independent Gaussian $X_i \sim (0, P_i)$
- with $P_i = \max\{\nu - \sigma_i^2, 0\}$,$\nu$ a constant and $\sum_{i=1}^{N} P_i = P$

Channel Capacity of Parallel AWGN Channels

- Consider N inputs $X_1, X_2, \ldots, X_N$
- N outputs $Y_1, Y_2, \ldots, Y_N$
- each with $Y_i = X_i + W_i$, $W_i \sim \mathcal{N}(0, \sigma_i^2)$
- subject to a total power constraint

$$\mathbb{E} = \left[\sum_{i=1}^{N} X_i^2 \right] \leq P$$

- The capacity is given by

$$C = \sum_{i=1}^{N} \frac{1}{2} \mathrm{ld} \left(1 + \frac{P_i}{\sigma_i^2} \right)$$

- is achieved for independent Gaussian $X_i \sim (0, P_i)$
- with $P_i = \max\{\nu - \sigma_i^2, 0\}$,$\nu$ a constant and $\sum_{i=1}^{N} P_i = P$
  $\rightarrow$ Waterfilling

Channel Capacity of the AWGN with Discrete Input

- AWGN channel with discrete input
  - Transmit symbol (channel input) is taken out of a discrete alphabet, e.g. a PAM constellation: $x \in \mathbb{X} = \{a_1, a_2, \ldots, a_M\} \subset \mathbb{R}$. We additionally assume that all constellation points are equiprobable.

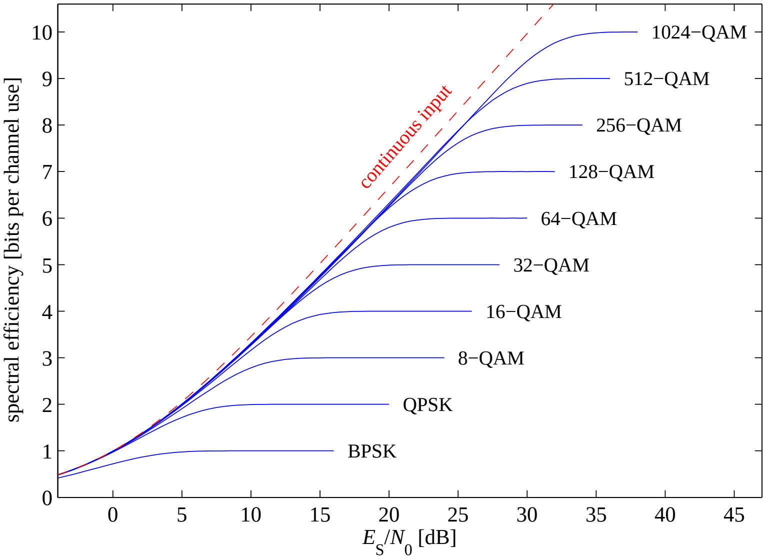
$$h(Y|X) = \sum_{x \in \mathbb{X}} p(x) h(Y|X = x) = h(W) = \frac{1}{2} \mathrm{ld}\left(\pi e N_0\right)$$

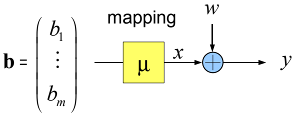
$$h(Y) = -\mathbb{E}\left[\mathrm{ld} p(y)\right] = -\mathbb{E}\left[\mathrm{ld}\left(\frac{1}{M\sqrt{\pi N_0}} \sum_{i=1}^{M} \exp\left(-\frac{(y - a_i)^2}{N_0}\right)\right)\right]$$

$$C = h(Y) - h(Y|X)$$

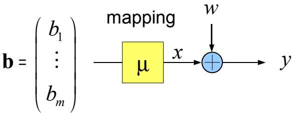
  - Note: most QAM constellations can be separated into two PAM constellations

Channel Capacity of the AWGN with Discrete Input





- With little loss of generality, we consider real-valued constellations
- Definitions and assumptions for real-valued AWGN
  - channel: $y = x + w$, noise: $w \sim \mathcal{N}(0, N_0/2)$, $p(y|x) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(y-x)^2}{N_0}\right)$
  - input bit vector: $\mathbf{b} = (b_1, \ldots, b_m)^{\mathrm{T}} \in \{0,1\}^m$
  - mapping function: $\mu : \{0,1\}^m \to \mathbb{X} = \{a_1, a_2, \ldots, a_M\} \subset \mathbb{R}$, with $M = 2^m$
  - uniform input distribution: $P[b_i = 0] = 0.5 \; \forall i$, hence $P[x = a_i] = 2^{-m} = \frac{1}{M}$



- With little loss of generality, we consider real-valued constellations
- Definitions and assumptions for real-valued AWGN
  - channel: $y = x + w$, noise: $w \sim \mathcal{N}(0, N_0/2)$, $p(y|x) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(y-x)^2}{N_0}\right)$
  - input bit vector: $\mathbf{b} = (b_1, \ldots, b_m)^{\mathrm{T}} \in \{0,1\}^m$
  - mapping function: $\mu : \{0,1\}^m \to \mathbb{X} = \{a_1, a_2, \ldots, a_M\} \subset \mathbb{R}$, with $M = 2^m$
  - uniform input distribution: $P[b_i = 0] = 0.5 \; \forall i$, hence $P[x = a_i] = 2^{-m} = \frac{1}{M}$
  - No optimization over input distribution

Capacities of AWGN channel with QAM

Definitions

- We can define the following capacities:
  1. $C^{\text{CM}} = I(X;Y) = I(\mathbf{B};Y)$, "coded modulation" capacity
  2. $C_q^{\text{CM}} = I(B_q;Y|B_1 \cdots B_{q-1})$, CM subchannel capacity
  3. $C_q^{\text{BICM}} = I(B_q;Y)$, BICM subchannel capacity
  4. $C^{\text{BICM}} = \sum_{q=1}^{m} I(B_q;Y)$, BICM capaciy
- Note: only C^{CM} is independent of the mapping μ
- Derivation of capacities

  From the chain rule,

$$C^{\text{CM}} = I(B_1 \cdots B_m;Y) = \sum_{q=1}^{m} I(B_q;Y|B_1 \cdots B_{q-1}) = \sum_{q=1}^{m} C_q^{\text{CM}}$$

again with chain rule,

$$C_q^{\text{CM}} = \underbrace{I(B_q \cdots B_m; Y | B_1 \cdots B_{q-1})}_{\triangleq R_{q-1}} - \underbrace{I(B_{q+1} \cdots B_m; Y | B_1 \cdots B_q)}_{=R_q}$$

$$R_q = \sum_{(b_1 \cdots b_q) \in \{0,1\}^q} P[B_1 = b_1, \ldots, B_q = b_q] \cdot I(B_{q+1} \cdots B_m; Y | b_1 \cdots b_q)$$

$$= 2^{-q} \sum_{j=0}^{2^q-1} \underbrace{I\left(B_{q+1} \cdots B_m; Y | (b_1 \cdots b_q) = \text{bin}(j)\right)}_{\triangleq R_{q,j}}$$

$R_{q,j} = C\left(\mathcal{A}\left((b_1 \cdots b_q) = \text{bin}(j)\right)\right)$, where $\mathcal{A}\left((b_1 \cdots b_q)\right)$ denotes the subconstellation with bits $b_1, \ldots, b_q$ fixed and $C(\mathcal{A})$ is its capacity

$\Rightarrow$ we require the capacity $C(\mathcal{A})$

Capacities of AWGN channel with QAM

Derivation

For a discrete set $\mathcal{A} = \{a_1, \ldots, a_M\}$ and equally probable constellation points, we have

$$C(\mathcal{A}) = \frac{1}{M}\sum_{i=1}^{M}\int_{-\infty}^{\infty} p(y|a_i)\,\mathrm{ld}\frac{p(y|a_i)}{\frac{1}{M}\sum_{j=1}^{M} p(y|a_j)}\,\mathrm{d}y$$

$$= \mathrm{ld} M - \frac{1}{M}\sum_{i=1}^{M}\int_{-\infty}^{\infty} p(y|a_i)\,\mathrm{ld}\sum_{j=1}^{M}\frac{p(y|a_j)}{p(y|a_i)}\,\mathrm{d}y$$

$$= \mathrm{ld} M - \frac{1}{M}\sum_{i=1}^{M}\mathop{\mathbb{E}}_{y|a_i}\left[\mathrm{ld}\sum_{j=1}^{M}\exp\left(-\frac{(y-a_i)^2-(y-a_j)^2}{N_0}\right)\right]$$

since the expectation is over $y = a_i + \sqrt{\frac{N_0}{2}}w$, with $w \sim \mathcal{N}(0,1)$, we can write

$$C(\mathcal{A}) = \mathrm{ld} M - \frac{1}{M}\sum_{i=1}^{M}\mathop{\mathbb{E}}_{w}\left[\mathrm{ld}\sum_{j=1}^{M}\exp\left(-\frac{(a_i - a_j + \sqrt{\frac{N_0}{2}}w)^2 - \frac{N_0}{2}w^2}{N_0}\right)\right]$$

- Finally, for numerical computation, we can approximate the expectation with a normally distributed sequence $w_1, w_2, \ldots, w_N$ for $N \to \infty$ and obtain

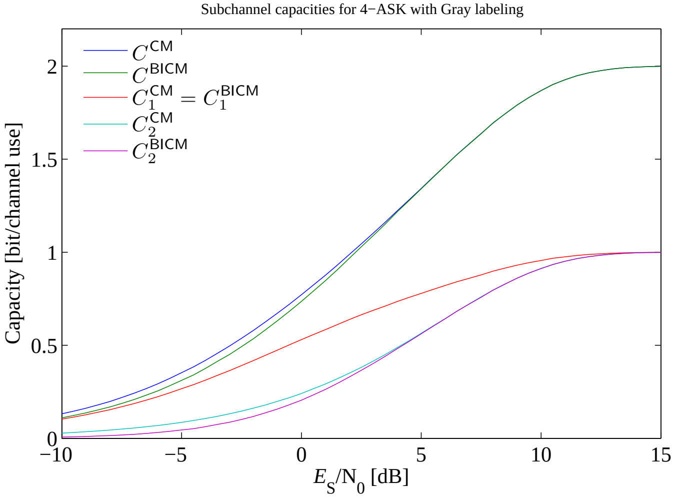
$$C(\mathcal{A}) \approx \mathrm{ld} M - \frac{1}{NM} \sum_{n=1}^{N} \sum_{i=1}^{M} \mathrm{ld} \sum_{j=1}^{M} \exp\left(-\frac{(a_i - a_j)^2}{N_0} + \sqrt{\frac{2}{N_0}}(a_i - a_j) w_n \right)$$

- Hence, we obtain $R_{q,j} \to R_q \to C_q^{\mathrm{CM}}$. Note that we can compute C^{CM} directly.

- For BICM, we have the subchannel capacities

$$\begin{aligned} C_q^{\mathrm{BICM}} &= I(\mathbf{B}; Y) - I(B_1 \cdots B_{q-1}, B_{q+1} \cdots B_m; Y | B_q) \\ &= C^{\mathrm{CM}} - \frac{C(\mathcal{A}(b_q = 0)) + C(\mathcal{A}(b_q = 1))}{2} \end{aligned}$$

Capacities for 4-ASK with Gray labeling

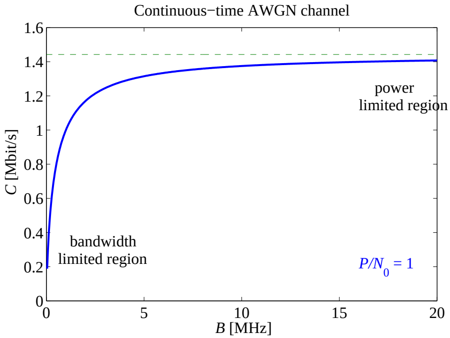


AWGN Capacity – Continuous Time

- We consider a passband channel with bandwidth B and noise power spectral density $N_0/2$ (note the redefinition of N_0!). Its capacity in bit/s is

$$C = B\,\mathrm{ld}\left(1 + \frac{P}{N_0 B}\right) \tag{20}$$

- The capacity for infinite bandwidth is $\lim_{B\to\infty} C = \frac{1}{\ln 2}\frac{P}{N_0}$



Fading Channels and Outage Capacity

- Flat fading channel: $y_i = h_i \cdot x_i + w_i, \quad w_i \sim \mathcal{CN}(0, N_0)$
  - Power gain is unity, i.e. $\mathbb{E}\left[|h_i|^2\right] = 1$, e.g. Rayleigh fading: $h_i \sim \mathcal{CN}(0, 1)$
  - Average SNR is $\bar{\gamma} = \frac{E_\mathrm{S}}{N_0}$

- Slow fading, no CSI at transmitter
  - channel coefficient is constant during one codeword: $h_i = h \, \forall i$, the "capacity" is hence $\mathrm{ld}(1 + |h|^2\bar{\gamma})$. The transmitter sends at rate R. The channel is in outage if the rate is too high,

$$P_\mathrm{out}(R) = P\left[\mathrm{ld}(1 + |h|^2\bar{\gamma}) < R\right] = 1 - \exp\left(-\frac{2^R - 1}{\bar{\gamma}}\right)$$

Fading Channels and Outage Capacity

- Flat fading channel: $y_i = h_i \cdot x_i + w_i, \quad w_i \sim \mathcal{CN}(0, N_0)$
  - Power gain is unity, i.e. $\mathbb{E}\left[|h_i|^2\right] = 1$, e.g. Rayleigh fading: $h_i \sim \mathcal{CN}(0, 1)$
  - Average SNR is $\bar{\gamma} = \frac{E_\mathrm{S}}{N_0}$

- Slow fading, no CSI at transmitter
  - channel coefficient is constant during one codeword: $h_i = h \; \forall i$, the "capacity" is hence $\mathrm{ld}(1 + |h|^2\bar{\gamma})$. The transmitter sends at rate R. The channel is in outage if the rate is too high,

$$P_\mathrm{out}(R) = P\left[\mathrm{ld}(1 + |h|^2\bar{\gamma}) < R\right] = 1 - \exp\left(-\frac{2^R - 1}{\bar{\gamma}}\right)$$

  - Channel capacity is zero!
  - We need another metric to describe this channel.

Fading Channels and Outage Capacity

- We define the outage capacity C_ϵ as the largest rate such that the outage probability is less than ϵ:

$$C_\epsilon = \mathrm{ld}\left(1 + \bar{\gamma} \cdot F_g^{-1}(\epsilon)\right) \tag{21}$$

  where $F_g(x) = P[g \leq x]$ is the cdf (cumulative distribution function) of the channel power gain $g = |h|^2$.
- Fast fading, no CSI at transmitter
  - h_i is i.i.d., while the codeword length $n \to \infty$. In this case, we apply the ergodic capacity

$$C = \mathbb{E}\left[\mathrm{ld}\left(1 + |h|^2 \bar{\gamma}\right)\right] \tag{22}$$

Remarks on practical Channel Coding schemes

Channel capacity and limits for infinite blocklength

- Shannon limit: reliable communication is only possible if the data rate is below the channel capacity

- For system design we may want to calculate the minimum SNR $\gamma_{\min}$ required to achieve a target perfomance BER $= p$ with rate R
- Recall that for the Gaussian channel

$$y = x + w$$

  with $w \sim \mathcal{N}(0, N_0/2)$
- capacity is achieved with Gaussian input distribution $x \sim \mathcal{N}(0, E_s)$

$$C_{\mathrm{G}}(\gamma) = \frac{1}{2}\mathrm{ld}(1 + 2\gamma)$$

Channel capacity and limits for infinite blocklength

- and the maximum achievable rate R for a given probability of bit error $p > 0$ is related to the channel capacity by

$$R(\gamma, p) = \frac{C_G(\gamma)}{1 - H_2(p)}$$

- where $H_2(p) = p\mathrm{ld}(p) - (1-p)\mathrm{ld}(1-p)$

$$\gamma_{\mathrm{min,G}} = \frac{1}{2}\left(2^{2R(1-H_2(p))} - 1\right)$$

- Equivalently, for binary antipodal signalling $x \in \{+1, -1\}$ the capacity is

$$C_{\mathrm{B}}(\gamma) = J(\sqrt{8\gamma})$$

with function $J(x)$ defined

$$J(x) = 1 - \frac{1}{\sqrt{2\pi}x} \int_{-\infty}^{\infty} \exp\left(-\frac{-(t - x^2/2)^2}{2x^2}\right) \mathrm{ld}(1 + \exp(-t))\, \mathrm{d}t$$

- $J(x)$ and its inverse $J^{-1}(y)$ are typically calculated through numerical approximations [tenBrink01].
- The minimum SNR γ_{min} required to achieve a target perfomance BER $= p$ with rate R with binary antipodal signalling is

$$\gamma_{\mathrm{min,B}} = \frac{1}{8}\left(J^{-1}\left(R(1 - H_2(p))\right)\right)^2$$

Channel capacity and limits for infinite blocklength

- Curve fitting approximations of the $J(\cdot)$-function

$$J(\sigma) = \begin{cases} -0.0421061\sigma^3 + 0.209252\sigma^2 - 0.00640081\sigma & 0 \leq \sigma \leq \sigma_{th} \\ 1 - \exp\left(0.00181491\sigma^3 - 0.142675\sigma^2 - 0.0822054\sigma + 0.0549608\right) & \sigma_{th} < \sigma < \infty \end{cases}$$

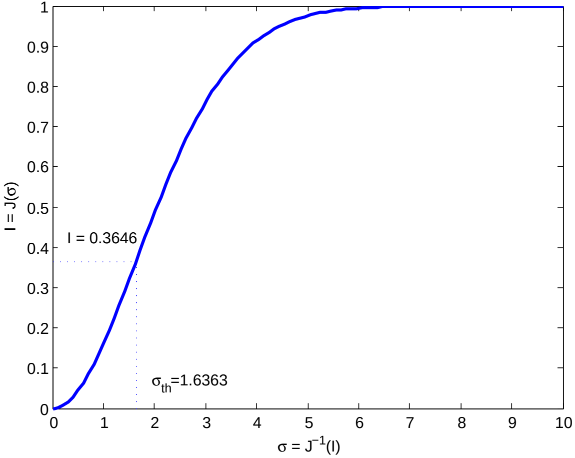
- Curve fitting approximations of the $J^{-1}(\cdot)$-function

$$J^{-1}(I) = \begin{cases} 1.09542I^2 + 0.214217I + 2.33727\sqrt{I} & 0 \leq I \leq I_{th} \\ -0.706692\ln\left(-0.386013(I-1)\right) + 1.75017I & I_{th} < I < 1 \end{cases}$$

[tenBrink01] S.ten Brink,"Design of Concatenated Coding Schemes based on Iterative Decoding Convergence", PhD thesis, Apr. 2001.

Channel capacity and limits for infinite blocklength

- Approximation of the $J(\cdot)$-function



Channel capacity and limits for finite blocklength

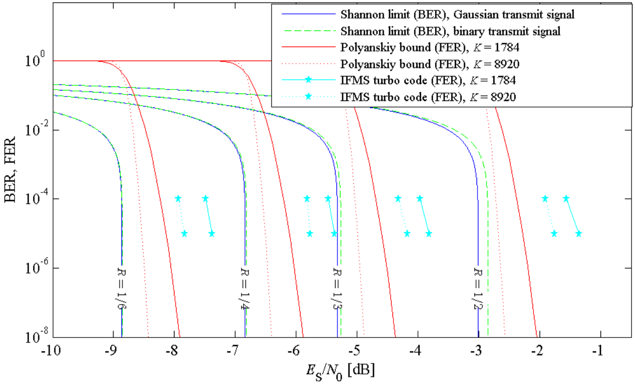
- However, practical communication systems use finite blocklength coding
- Polyanksi, Poor and Verdú [Pol10] derived frame error probability bounds for finite length N schemes (Gaussian input distribution)
- The following expression provides a FER lower bound for any practical coding/decoding scheme

$$p_{\mathrm{F}} \geq Q\left((1+2\gamma)\sqrt{\frac{N}{8\gamma(1+\gamma)}}\left(\ln(1+\gamma)+\frac{\ln N}{N}-2R\ln N\right)\right)$$

[Pol10]Y. Polyanskiy, H. V. Poor, and S. Verd, Channel coding rate in the finite blocklength regime, IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307-2359, May 2010.

Channel capacity and limits for finite blocklength

An example



Shannon limits for Gaussian and binary transmit signals, with Polyanskiy bound for codeword lengths $N = K/R_c$ and code rates $R_c = 1/2, 1/3, 1/4$ and $1/6$ and the performance of IFMS turbo codes.

Research trends in Channel Coding

- Today's channel coding schemes: LPDC, turbo-codes are capacity achieving for point to point BIAWGN channels
- There is room for improvement for multiuser networks
- But also for point to point links
  - short block lengths (e.g. Massive Machine Type Communications, IoT framework)
  - multi-terminal coding/decoding
  - improve computational power efficiency $\rightarrow$ transmission power efficiency to be expected
  - joint source-channel coding
  - security (network coding $\rightarrow$ physical layer security)