Digital Communications
SSY125, Lectures 5 and 6

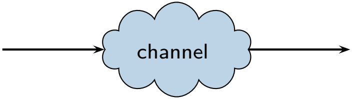
Communication over a Noisy Channel
(Chapters 4 and 5)

Christian Häger
Slides prepared by Alexandre Graell i Amat

November 8, 2023

CHALMERS

digital data, voice
movie, audio
100110101



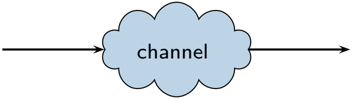
satellite link,
fiber, telephone line,
hard-disk drive, DVD

100011101

In These Lectures...

- How much information can we transmit reliably over an unreliable channel?
- How do we achieve this in practice?

digital data, voice
movie, audio
100110101



satellite link,
fiber, telephone line,
hard-disk drive, DVD

100011101

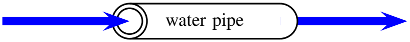
In These Lectures...

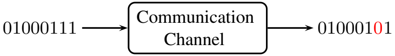
- How much information can we transmit reliably over an unreliable channel?
- How do we achieve this in practice?

Shannon's Seminal Contribution

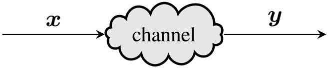
There is a fundamental limit, i.e., a highest rate, at which information can be transmitted reliably over the channel: channel capacity.

Model of a Noisy Communication Channel

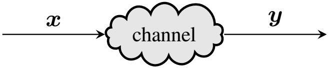




Model of a Noisy Communication Channel

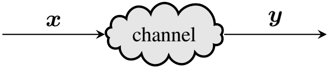


Model of a Noisy Communication Channel



Discrete Memoryless Channel

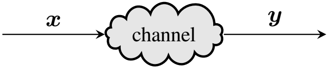
Model of a Noisy Communication Channel



Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.

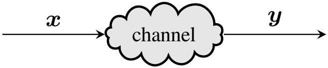
Model of a Noisy Communication Channel



Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$

Model of a Noisy Communication Channel

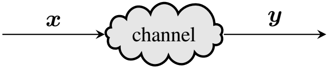


Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ with (memoryless)

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) =$$

Model of a Noisy Communication Channel

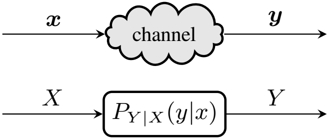


Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ with (memoryless)

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i P_{Y|X}(y_i|x_i).$$

Model of a Noisy Communication Channel

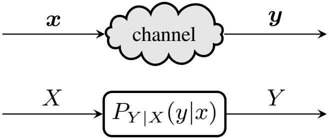


Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ with (memoryless)

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i P_{Y|X}(y_i|x_i).$$

Model of a Noisy Communication Channel

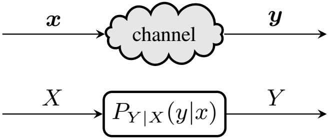


Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \dots,)$ and $\boldsymbol{y} = (y_1, y_2, \dots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ with (memoryless)

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i P_{Y|X}(y_i|x_i).$$

Model of a Noisy Communication Channel



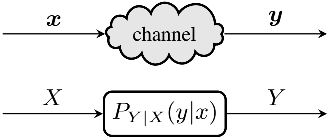
Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ with (memoryless)

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i P_{Y|X}(y_i|x_i).$$

- Channel input: $X \in \mathcal{X}$.

Model of a Noisy Communication Channel



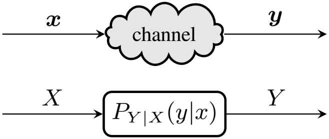
Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ with (memoryless)

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i P_{Y|X}(y_i|x_i).$$

- Channel input: $X \in \mathcal{X}$.
- Channel output: $Y \in \mathcal{Y}$.

Model of a Noisy Communication Channel



Discrete Memoryless Channel

- The input and output of the channel, $\boldsymbol{x} = (x_1, x_2, \ldots,)$ and $\boldsymbol{y} = (y_1, y_2, \ldots,)$, are discrete, i.e., $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$.
- Channel is defined by $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ with (memoryless)

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_i P_{Y|X}(y_i|x_i).$$

- Channel input: $X \in \mathcal{X}$.
- Channel output: $Y \in \mathcal{Y}$.
- Entirely specified by the conditional PMF $P_{Y|X}(y|x)$.

System Entropies

- Both the channel input and the channel output are random variables

System Entropies

- Both the channel input and the channel output are random variables $\longrightarrow$ The information content of a random variable, its uncertainty, is given by the entropy!

System Entropies

- Both the channel input and the channel output are random variables $\longrightarrow$ The information content of a random variable, its uncertainty, is given by the entropy!

Average amount of uncertainty on the input of the channel X:

$$\mathsf{H}(X) = \sum_{x \in \mathcal{X}} P(x) \log \frac{1}{P(x)}.$$

System Entropies

- Both the channel input and the channel output are random variables $\longrightarrow$ The information content of a random variable, its uncertainty, is given by the entropy!

Average amount of uncertainty on the input of the channel X:

$$\mathsf{H}(X) = \sum_{x \in \mathcal{X}} P(x) \log \frac{1}{P(x)}.$$

We have shown...

$$0 \leq \mathsf{H}(X) \leq \log |\mathcal{X}|.$$

System Entropies

- Both the channel input and the channel output are random variables $\longrightarrow$ The information content of a random variable, its uncertainty, is given by the entropy!

Average amount of uncertainty on the input of the channel X:

$$\mathsf{H}(X) = \sum_{x \in \mathcal{X}} P(x) \log \frac{1}{P(x)}.$$

We have shown...

$$0 \leq \mathsf{H}(X) \leq \log |\mathcal{X}|.$$

Average amount of uncertainty on the channel output Y:

$$\mathsf{H}(Y) = \sum_{y \in \mathcal{Y}} P(y) \log \frac{1}{P(y)}.$$

System Entropies

Definition (Conditional Entropy of X given $Y = y$)

The conditional entropy of X given the event $Y = y$ is

$$\mathsf{H}(X|Y = y) \triangleq \sum_{x \in X} P(x|y) \log \frac{1}{P(x|y)}$$

System Entropies

Definition (Conditional Entropy of X given $Y = y$)

The conditional entropy of X given the event $Y = y$ is

$$\mathsf{H}(X|Y = y) \triangleq \sum_{x \in X} P(x|y) \log \frac{1}{P(x|y)}$$

Definition (Conditional Entropy)

The conditional entropy of X given Y is

$$\mathsf{H}(X|Y) =$$

System Entropies

Definition (Conditional Entropy of X given $Y = y$)

The conditional entropy of X given the event $Y = y$ is

$$\mathsf{H}(X|Y = y) \triangleq \sum_{x \in X} P(x|y) \log \frac{1}{P(x|y)}$$

Definition (Conditional Entropy)

The conditional entropy of X given Y is

$$\mathsf{H}(X|Y) = \sum_{y \in \mathcal{Y}} P(y) \mathsf{H}(X|Y = y)$$

System Entropies

Definition (Conditional Entropy of X given $Y = y$)

The conditional entropy of X given the event $Y = y$ is

$$\mathsf{H}(X|Y = y) \triangleq \sum_{x \in X} P(x|y) \log \frac{1}{P(x|y)}$$

Definition (Conditional Entropy)

The conditional entropy of X given Y is

$$\begin{aligned} \mathsf{H}(X|Y) &= \sum_{y \in \mathcal{Y}} P(y) \mathsf{H}(X|Y = y) \\ &= \sum_{y \in \mathcal{Y}} P(y) \sum_{x \in \mathcal{X}} P(x|y) \log \frac{1}{P(x|y)} \end{aligned}$$

System Entropies

Definition (Conditional Entropy of X given $Y = y$)

The conditional entropy of X given the event $Y = y$ is

$$\mathsf{H}(X|Y = y) \triangleq \sum_{x \in X} P(x|y) \log \frac{1}{P(x|y)}$$

Definition (Conditional Entropy)

The conditional entropy of X given Y is

$$\begin{aligned}
\mathsf{H}(X|Y) &= \sum_{y \in \mathcal{Y}} P(y) \mathsf{H}(X|Y = y) \\
&= \sum_{y \in \mathcal{Y}} P(y) \sum_{x \in \mathcal{X}} P(x|y) \log \frac{1}{P(x|y)} \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(y) P(x|y) \log \frac{1}{P(x|y)}
\end{aligned}$$

System Entropies

Definition (Conditional Entropy of X given $Y = y$)

The conditional entropy of X given the event $Y = y$ is

$$\mathsf{H}(X|Y = y) \triangleq \sum_{x \in X} P(x|y) \log \frac{1}{P(x|y)}$$

Definition (Conditional Entropy)

The conditional entropy of X given Y is

$$\begin{aligned}
\mathsf{H}(X|Y) &= \sum_{y \in \mathcal{Y}} P(y) \mathsf{H}(X|Y = y) \\
&= \sum_{y \in \mathcal{Y}} P(y) \sum_{x \in \mathcal{X}} P(x|y) \log \frac{1}{P(x|y)} \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(y) P(x|y) \log \frac{1}{P(x|y)} \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x, y) \log \frac{1}{P(x|y)}.
\end{aligned}$$

System Entropies

Theorem (Conditioning reduces entropy)

For any two random variables X and Y,

$$\mathsf{H}(X|Y) \leq \mathsf{H}(X).$$

with equality if and only if X and Y are statistically independent.

System Entropies

Definition (Joint Entropy)

The joint entropy of X and Y is defined as

$$\mathsf{H}(X,Y) =$$

System Entropies

Definition (Joint Entropy)

The joint entropy of X and Y is defined as

$$\mathsf{H}(X,Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log\frac{1}{P(x,y)}.$$

System Entropies

Definition (Joint Entropy)

The joint entropy of X and Y is defined as

$$\mathsf{H}(X,Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x,y) \log \frac{1}{P(x,y)}.$$

It gives the uncertainty of the overall system.

System Entropies

Definition (Joint Entropy)

The joint entropy of X and Y is defined as

$$\mathsf{H}(X,Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log\frac{1}{P(x,y)}.$$

It gives the uncertainty of the overall system.

Lemma

The joint entropy of two statistically independent random variables X and Y is

System Entropies

Definition (Joint Entropy)

The joint entropy of X and Y is defined as

$$\mathsf{H}(X,Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log\frac{1}{P(x,y)}.$$

It gives the uncertainty of the overall system.

Lemma

The joint entropy of two statistically independent random variables X and Y is

$$\mathsf{H}(X,Y) = \mathsf{H}(X) + \mathsf{H}(Y).$$

System Entropies

For any *useful* communication channel, the output Y and the input X are not independent.

System Entropies

For any *useful* communication channel, the output Y and the input X are not independent.

Lemma (Chain Rule)

The joint entropy of two random variables X and Y is given by

$$\mathsf{H}(X,Y) =$$

System Entropies

For any *useful* communication channel, the output Y and the input X are not independent.

Lemma (Chain Rule)

The joint entropy of two random variables X and Y is given by

$$\mathsf{H}(X, Y) = \mathsf{H}(X) + \mathsf{H}(Y|X).$$

System Entropies

For any *useful* communication channel, the output Y and the input X are not independent.

Lemma (Chain Rule)

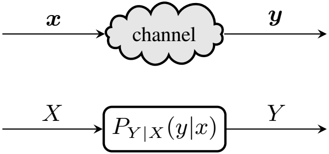
The joint entropy of two random variables X and Y is given by

$$\mathsf{H}(X, Y) = \mathsf{H}(X) + \mathsf{H}(Y|X).$$

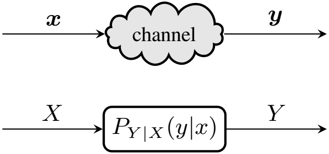
Since $P(x, y) = P(y, x)$, we can also write

$$\mathsf{H}(X, Y) = \mathsf{H}(Y) + \mathsf{H}(X|Y).$$

Information Conveyed by the Channel

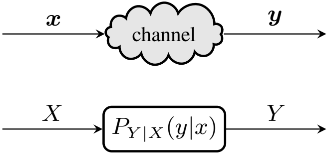


Information Conveyed by the Channel



- How much information about the transmitted random variable X we gain from Y?

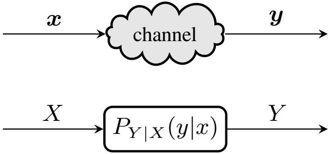
Information Conveyed by the Channel



- How much information about the transmitted random variable X we gain from Y?

Equivalently...

Information Conveyed by the Channel

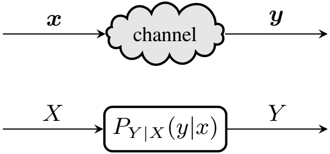


- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y

Information Conveyed by the Channel

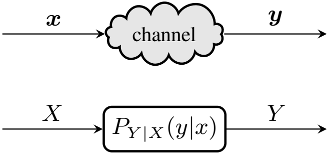


- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y $\longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

Information Conveyed by the Channel



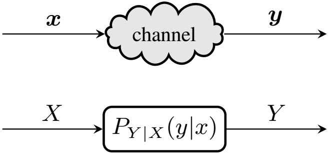
- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y $\longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

- How much information have we conveyed over the channel?

Information Conveyed by the Channel



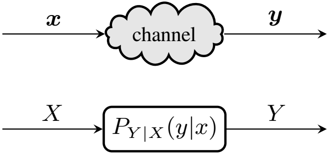
- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y $\longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

- How much information have we conveyed over the channel?
  - *Before* receiving Y: Uncertainty on X is

Information Conveyed by the Channel



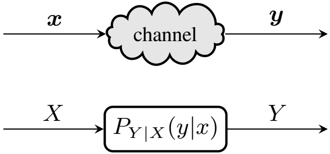
- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y $\longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

- How much information have we conveyed over the channel?
  - *Before* receiving Y: Uncertainty on X is $\mathsf{H}(X)$

Information Conveyed by the Channel



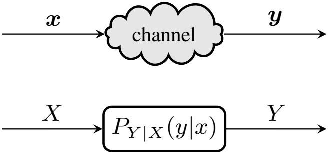
- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y $\longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

- How much information have we conveyed over the channel?
  - *Before* receiving Y: Uncertainty on X is $\mathsf{H}(X)$
  - *After* receiving Y: Uncertainty on X is

Information Conveyed by the Channel



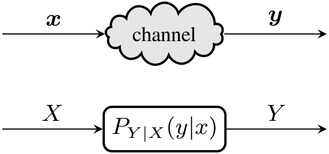
- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y $\longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

- How much information have we conveyed over the channel?
  - *Before* receiving Y: Uncertainty on X is $\mathsf{H}(X)$
  - *After* receiving Y: Uncertainty on X is $\mathsf{H}(X|Y)$

Information Conveyed by the Channel



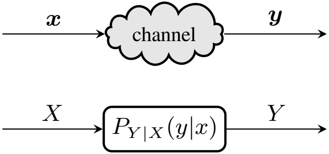
- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving Y $\longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

- How much information have we conveyed over the channel?
  - *Before* receiving Y: Uncertainty on X is $\mathsf{H}(X)$
  - *After* receiving Y: Uncertainty on X is $\mathsf{H}(X|Y)$
  - Thus, the information conveyed is:

Information Conveyed by the Channel



- How much information about the transmitted random variable X we gain from Y?

Equivalently...

How much uncertainty remains on the knowledge of X after receiving $Y \longrightarrow$ The conditional entropy of X given Y, $\mathsf{H}(X|Y)$!

- How much information have we conveyed over the channel?
  - *Before* receiving Y: Uncertainty on X is $\mathsf{H}(X)$
  - *After* receiving Y: Uncertainty on X is $\mathsf{H}(X|Y)$
  - Thus, the information conveyed is: $\mathsf{H}(X) - \mathsf{H}(X|Y)$!

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

- The mutual information measures the average reduction in uncertainty about X that results from learning the value of Y.

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

- The mutual information measures the average reduction in uncertainty about X that results from learning the value of Y.

By simple manipulation of the entropies,

$$\mathsf{I}(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}.$$

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

- The mutual information measures the average reduction in uncertainty about X that results from learning the value of Y.

By simple manipulation of the entropies,

$$\mathsf{I}(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}.$$

Properties

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

- The mutual information measures the average reduction in uncertainty about X that results from learning the value of Y.

By simple manipulation of the entropies,

$$\mathsf{I}(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}.$$

Properties

P1 $\mathsf{I}(X;Y) \geq 0$,

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

- The mutual information measures the average reduction in uncertainty about X that results from learning the value of Y.

By simple manipulation of the entropies,

$$\mathsf{I}(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}.$$

Properties

P1 $\mathsf{I}(X;Y) \geq 0$,

P2 $\mathsf{I}(X;Y) = 0$ if and only if X and Y are independent,

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

- The mutual information measures the average reduction in uncertainty about X that results from learning the value of Y.

By simple manipulation of the entropies,

$$\mathsf{I}(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}.$$

Properties

P1 $\mathsf{I}(X;Y) \geq 0$,

P2 $\mathsf{I}(X;Y) = 0$ if and only if X and Y are independent,

P3 $\mathsf{I}(X;Y) = \mathsf{I}(Y;X)$,

Information Conveyed by the Channel

Definition (Mutual Information)

The mutual information between X and Y is

$$\mathsf{I}(X;Y) \triangleq \mathsf{H}(X) - \mathsf{H}(X|Y).$$

- The mutual information measures the average reduction in uncertainty about X that results from learning the value of Y.

By simple manipulation of the entropies,

$$\mathsf{I}(X;Y) = \sum_{x\in X}\sum_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}.$$

Properties

P1 $\mathsf{I}(X;Y) \geq 0$,

P2 $\mathsf{I}(X;Y) = 0$ if and only if X and Y are independent,

P3 $\mathsf{I}(X;Y) = \mathsf{I}(Y;X)$,

P4 $\mathsf{I}(X;Y) \leq \min(\mathsf{H}(X), \mathsf{H}(Y))$.

The Channel Capacity

- How much information can we transmit over the cannel?

The Channel Capacity

- How much information can we transmit over the cannel?
  - No control over the channel.

The Channel Capacity

- How much information can we transmit over the cannel?
  - No control over the channel.
  - However, control over $P_X(x)$.

The Channel Capacity

- How much information can we transmit over the cannel?
  - No control over the channel.
  - However, control over $P_X(x)$.
- The mutual information $\mathsf{I}(X;Y)$ depends on $P_X(x)$!.

The Channel Capacity

- How much information can we transmit over the cannel?
  - No control over the channel.
  - However, control over $P_X(x)$.
- The mutual information $\mathsf{I}(X;Y)$ depends on $P_X(x)$!.

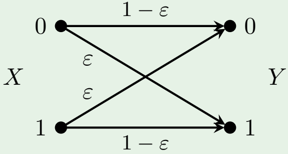
Definition (Channel Capacity)

For a given channel, the channel capacity is defined to be the maximum of the mutual information, maximized over all possible input distributions $P_X(x)$,

$$\mathsf{C} \triangleq \max_{P_X} \mathsf{I}(X;Y).$$

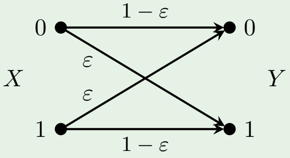
Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel



Mutual Information and Channel Capacity

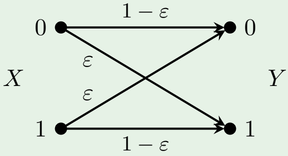
Running Example: The Binary Symmetric Channel



- **Input:** $X \in \mathcal{X} = \{0, 1\}$.

Mutual Information and Channel Capacity

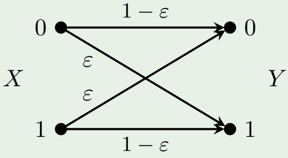
Running Example: The Binary Symmetric Channel



- Input: $X \in \mathcal{X} = \{0, 1\}$.
- Output: $Y \in \mathcal{Y} = \{0, 1\}$.

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel



- Input: $X \in \mathcal{X} = \{0, 1\}$.
- Output: $Y \in \mathcal{Y} = \{0, 1\}$.
- Channel defined by the transition probabilities

$$P(0|0) = P(1|1) = 1 - \varepsilon$$
$$P(1|0) = P(0|1) = \varepsilon.$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$\mathsf{I}(X;Y)$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$\mathsf{I}(X;Y) = \mathsf{H}(Y) - \mathsf{H}(Y|X)$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$
\begin{aligned}
\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_{x\in\mathcal{X}} P(x)\mathsf{H}(Y|X=x)
\end{aligned}
$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$
\begin{aligned}
\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_{x \in \mathcal{X}} P(x)\mathsf{H}(Y|X=x) \\
&= \mathsf{H}(Y) - (P(0)\mathsf{H}(Y|X=0) + P(1)\mathsf{H}(Y|X=1))
\end{aligned}
$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$\begin{aligned}
\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_{x\in\mathcal{X}} P(x)\mathsf{H}(Y|X=x) \\
&= \mathsf{H}(Y) - (P(0)\mathsf{H}(Y|X=0) + P(1)\mathsf{H}(Y|X=1)) \\
&= \mathsf{H}(Y) - \underbrace{(P(0)+P(1))}_{=1}\mathsf{H}_\mathsf{b}(\varepsilon)
\end{aligned}$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$
\begin{aligned}
\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_{x \in \mathcal{X}} P(x)\mathsf{H}(Y|X=x) \\
&= \mathsf{H}(Y) - \left(P(0)\mathsf{H}(Y|X=0) + P(1)\mathsf{H}(Y|X=1)\right) \\
&= \mathsf{H}(Y) - \underbrace{(P(0)+P(1))}_{=1}\mathsf{H}_{\mathsf{b}}(\varepsilon) \\
&= \mathsf{H}(Y) - \mathsf{H}_{\mathsf{b}}(\varepsilon)
\end{aligned}
$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$
\begin{aligned}
\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_{x\in\mathcal{X}} P(x)\mathsf{H}(Y|X=x) \\
&= \mathsf{H}(Y) - \left(P(0)\mathsf{H}(Y|X=0) + P(1)\mathsf{H}(Y|X=1)\right) \\
&= \mathsf{H}(Y) - \underbrace{(P(0)+P(1))}_{=1}\mathsf{H}_\mathsf{b}(\varepsilon) \\
&= \mathsf{H}(Y) - \mathsf{H}_\mathsf{b}(\varepsilon) \\
&\leq
\end{aligned}
$$

Mutual Information and Channel Capacity

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$
\begin{aligned}
\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_{x\in\mathcal{X}} P(x)\mathsf{H}(Y|X=x) \\
&= \mathsf{H}(Y) - (P(0)\mathsf{H}(Y|X=0) + P(1)\mathsf{H}(Y|X=1)) \\
&= \mathsf{H}(Y) - \underbrace{(P(0)+P(1))}_{=1}\mathsf{H}_{\mathsf{b}}(\varepsilon) \\
&= \mathsf{H}(Y) - \mathsf{H}_{\mathsf{b}}(\varepsilon) \\
&\leq 1 - \mathsf{H}_{\mathsf{b}}(\varepsilon).
\end{aligned}
$$

Running Example: The Binary Symmetric Channel

- We compute first the mutual information,

$$\begin{aligned}
\mathsf{I}(X;Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_{x\in\mathcal{X}} P(x)\mathsf{H}(Y|X=x) \\
&= \mathsf{H}(Y) - (P(0)\mathsf{H}(Y|X=0) + P(1)\mathsf{H}(Y|X=1)) \\
&= \mathsf{H}(Y) - \underbrace{(P(0)+P(1))}_{=1}\mathsf{H}_\mathsf{b}(\varepsilon) \\
&= \mathsf{H}(Y) - \mathsf{H}_\mathsf{b}(\varepsilon) \\
&\leq 1 - \mathsf{H}_\mathsf{b}(\varepsilon).
\end{aligned}$$

$\mathsf{I}(X;Y)$ depends on P_Y, thus on P_X!

The Channel Capacity

Running Example: The Binary Symmetric Channel

$$\mathsf{I}(X;Y) = \mathsf{H}(Y) - \mathsf{H}_\mathsf{b}(\varepsilon)$$

- Let $P(X=0)=\delta$ and $P(X=1)=1-\delta$.

The Channel Capacity

Running Example: The Binary Symmetric Channel

$$\mathsf{I}(X;Y) = \mathsf{H}(Y) - \mathsf{H_b}(\varepsilon)$$

- Let $P(X=0) = \delta$ and $P(X=1) = 1-\delta$.
- The mutual information is maximized for

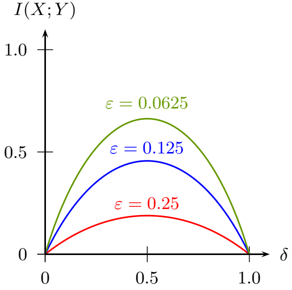
The Channel Capacity

Running Example: The Binary Symmetric Channel

$$\mathsf{I}(X;Y) = \mathsf{H}(Y) - \mathsf{H_b}(\varepsilon) \leq 1 - \mathsf{H_b}(\varepsilon).$$

- Let $P(X=0)=\delta$ and $P(X=1)=1-\delta$.
- The mutual information is maximized for $\delta = \frac{1}{2}$ (equally likely symbols)

The Channel Capacity

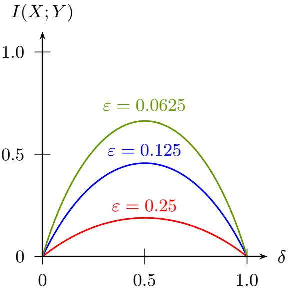


Running Example: The Binary Symmetric Channel

$$\mathsf{I}(X;Y) = \mathsf{H}(Y) - \mathsf{H_b}(\varepsilon) \leq 1 - \mathsf{H_b}(\varepsilon).$$

- Let $P(X=0) = \delta$ and $P(X=1) = 1-\delta$.
- The mutual information is maximized for $\delta = \frac{1}{2}$ (equally likely symbols)

The Channel Capacity

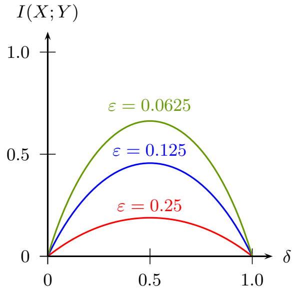


Running Example: The Binary Symmetric Channel

$$\mathsf{I}(X;Y) = \mathsf{H}(Y) - \mathsf{H_b}(\varepsilon) \leq 1 - \mathsf{H_b}(\varepsilon).$$

- Let $P(X=0) = \delta$ and $P(X=1) = 1-\delta$.
- The mutual information is maximized for $\delta = \frac{1}{2}$ (equally likely symbols)
- The maximum of the mutual information is the channel capacity,

The Channel Capacity



Running Example: The Binary Symmetric Channel

$$\mathsf{I}(X;Y) = \mathsf{H}(Y) - \mathsf{H_b}(\varepsilon) \leq 1 - \mathsf{H_b}(\varepsilon).$$

- Let $P(X=0) = \delta$ and $P(X=1) = 1-\delta$.
- The mutual information is maximized for $\delta = \frac{1}{2}$ (equally likely symbols)
- The maximum of the mutual information is the channel capacity,

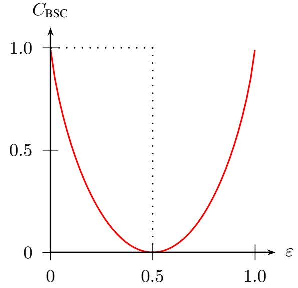
$$\mathsf{C_{BSC}} = 1 - \mathsf{H_b}(\varepsilon) \text{ bits.}$$

The Channel Capacity

Running Example: The Binary Symmetric Channel

$$\mathsf{C}_{\mathsf{BSC}} = 1 - \mathsf{H}_{\mathsf{b}}(\varepsilon) \quad \text{bits.}$$

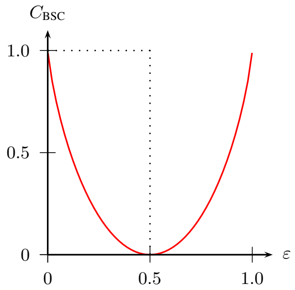
The Channel Capacity



Running Example: The Binary Symmetric Channel

$$\mathsf{C}_{\mathsf{BSC}} = 1 - \mathsf{H}_{\mathsf{b}}(\varepsilon) \quad \text{bits.}$$

The Channel Capacity

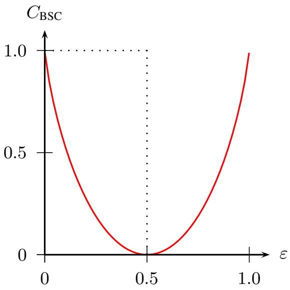


Running Example: The Binary Symmetric Channel

$$\mathsf{C}_{\mathsf{BSC}} = 1 - \mathsf{H}_{\mathsf{b}}(\varepsilon) \quad \text{bits.}$$

- It is maximum $(1$ bit$)$ for $\varepsilon = 0$ and $\varepsilon = 1$.

The Channel Capacity



Running Example: The Binary Symmetric Channel

$$\mathsf{C}_{\mathsf{BSC}} = 1 - \mathsf{H}_{\mathsf{b}}(\varepsilon) \text{ bits.}$$

- It is maximum (1 bit) for $\varepsilon = 0$ and $\varepsilon = 1$.
- It is zero for $\varepsilon = \frac{1}{2} \longrightarrow$ The channel is useless!

The Channel Coding Theorem

Definition (Channel Capacity)

For a given channel, the channel capacity is defined to be the maximum of mutual information, maximized over all possible input distributions $P(x)$,

$$\mathsf{C} \triangleq \max_{P_X(x)} \mathsf{I}(X;Y).$$

The Channel Coding Theorem

Definition (Channel Capacity)

For a given channel, the channel capacity is defined to be the maximum of mutual information, maximized over all possible input distributions $P(x)$,

$$\mathsf{C} \triangleq \max_{P_X(x)} \mathsf{I}(X;Y).$$

- The channel capacity is a measure of the information conveyed by a channel, but...

The Channel Coding Theorem

Definition (Channel Capacity)

For a given channel, the channel capacity is defined to be the maximum of mutual information, maximized over all possible input distributions $P(x)$,

$$\mathsf{C} \triangleq \max_{P_X(x)} \mathsf{I}(X;Y).$$

- The channel capacity is a measure of the information conveyed by a channel, but... What is its operational meaning?

The Channel Coding Theorem

The channel capacity is the maximum transmission rate at which we can communicate reliably over the channel!

The Channel Coding Theorem

The channel capacity is the maximum transmission rate at which we can communicate reliably over the channel!

Theorem (Shannon's Channel Coding Theorem)

For a discrete-time channel, it is possible to transmit information with an arbitrarily small probability of error if the communication rate R is below the channel capacity, i.e., $R < \mathsf{C}$. More precisely, for any $R \leq \mathsf{C}$, there exist a sequence of coding schemes of length N with average error probability $P_{\mathsf{e}}^{(N)}$ that tends to zero as $N \to \infty$, i.e., $P_{\mathsf{e}}^{(N)} \to 0$ as $N \to \infty$.
Conversely, any sequence of coding schemes with vanishing error probability must have $R \leq C$. Hence, the probability of error for transmission above capacity is bounded above zero.

The Channel Coding Theorem

The channel capacity is the maximum transmission rate at which we can communicate reliably over the channel!

Theorem (Shannon's Channel Coding Theorem)

For a discrete-time channel, it is possible to transmit information with an arbitrarily small probability of error if the communication rate R is below the channel capacity, i.e., $R < \mathsf{C}$. More precisely, for any $R \leq \mathsf{C}$, there exist a sequence of coding schemes of length N with average error probability $P_{\mathsf{e}}^{(N)}$ that tends to zero as $N \to \infty$, i.e., $P_{\mathsf{e}}^{(N)} \to 0$ as $N \to \infty$.
Conversely, any sequence of coding schemes with vanishing error probability must have $R \leq C$. Hence, the probability of error for transmission above capacity is bounded above zero.

Running Example: The Binary Symmetric Channel

For $\varepsilon = 0.25$, $\mathsf{C}_{\mathsf{BSC}} = 1 - \mathsf{H}_{\mathsf{b}}(\varepsilon) = 1 - \mathsf{H}_{\mathsf{b}}(0.25) = 0.1887$

The Channel Coding Theorem

The channel capacity is the maximum transmission rate at which we can communicate reliably over the channel!

Theorem (Shannon's Channel Coding Theorem)

For a discrete-time channel, it is possible to transmit information with an arbitrarily small probability of error if the communication rate R is below the channel capacity, i.e., $R < \mathsf{C}$. More precisely, for any $R \leq \mathsf{C}$, there exist a sequence of coding schemes of length N with average error probability $P_{\mathrm{e}}^{(N)}$ that tends to zero as $N \to \infty$, i.e., $P_{\mathrm{e}}^{(N)} \to 0$ as $N \to \infty$.
Conversely, any sequence of coding schemes with vanishing error probability must have $R \leq C$. Hence, the probability of error for transmission above capacity is bounded above zero.

Running Example: The Binary Symmetric Channel

For $\varepsilon = 0.25$, $\mathsf{C}_{\mathsf{BSC}} = 1 - \mathsf{H}_{\mathsf{b}}(\varepsilon) = 1 - \mathsf{H}_{\mathsf{b}}(0.25) = 0.1887 \longrightarrow$ Reliable communication is possible as long as we transmit at a rate < 0.1887 bits per channel use.

Communication Over the AWGN Channel

The AWGN Channel

Additive White Gaussian Noise (AWGN) Channel

Continuous-time, complex AWGN channel,

$$y(t) = x(t) + n(t),$$

where $x(t)$ is bandlimited with bandwidth W and has signal power P, the symbol interval is $T = 1/W$, and $n(t)$ is complex AWGN with PSD N_0 and

$$\mathsf{SNR} \triangleq \frac{P}{\mathsf{N}_0 W}.$$

The AWGN Channel

Additive White Gaussian Noise (AWGN) Channel

Continuous-time, complex AWGN channel,

$$y(t) = x(t) + n(t),$$

where $x(t)$ is bandlimited with bandwidth W and has signal power P, the symbol interval is $T = 1/W$, and $n(t)$ is complex AWGN with PSD N_0 and

$$\mathsf{SNR} \triangleq \frac{P}{\mathsf{N}_0 W}.$$

Discrete-time AWGN channel,

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n},$$

where $\boldsymbol{x} = (x_1, x_2, \ldots)$ is the transmitted sequence of constellation symbols with average energy per symbol $\mathsf{E_s}$, and $\boldsymbol{n}$ is a sequence of i.i.d. Gaussian noise random variables with zero mean and variance $\sigma^2 = \mathsf{N}_0/2$ per real dimension.

The AWGN Channel

The AWGN Channel

- Input of the channel X is distributed according to to $p(x) = p_X(x)$.

The AWGN Channel

- Input of the channel X is distributed according to to $p(x) = p_X(x)$.
- Channel law (conditional PDF):

$$p(y|x) = p_{Y|X}(y|x) = \frac{1}{2\pi\sigma^2} e^{-\frac{|y-x|^2}{2\sigma^2}}.$$

The AWGN Channel

- Input of the channel X is distributed according to to $p(x) = p_X(x)$.
- Channel law (conditional PDF):

$$p(y|x) = p_{Y|X}(y|x) = \frac{1}{2\pi\sigma^2} e^{-\frac{|y-x|^2}{2\sigma^2}}.$$

- Signal-to-noise ratio:

$$\mathsf{SNR} \triangleq \frac{\mathsf{E_s}}{\mathsf{N_0}} = \frac{\mathsf{E_s}}{2\sigma^2}.$$

Mutual Information and Channel Capacity

Mutual Information and Channel Capacity

- For a given p_X, the amount of information that can be conveyed over the channel is given by the mutual information

$$\mathsf{I}(X;Y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy,$$

Mutual Information and Channel Capacity

- For a given p_X, the amount of information that can be conveyed over the channel is given by the mutual information

$$\mathsf{I}(X;Y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy,$$

- If X is a discrete RV,

$$\mathsf{I}(X;Y) = \sum_{x \in X} \int P(x)p(y|x) \log \frac{p(x,y)}{P(x)p(y)} dy,$$

Mutual Information and Channel Capacity

- For a given p_X, the amount of information that can be conveyed over the channel is given by the mutual information

$$\mathsf{I}(X;Y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy,$$

- If X is a discrete RV,

$$\mathsf{I}(X;Y) = \sum_{x \in X} \int P(x)p(y|x) \log \frac{p(x,y)}{P(x)p(y)} dy,$$

- The ultimate limit at which we can transmit reliably is given by the channel capacity,

$$\mathsf{C} \triangleq \max_{p_X} \mathsf{I}(X;Y).$$

Capacity of the AWGN channel

Capacity of the Discrete-Time AWGN Channel

The channel capacity of the discrete-time complex AWGN channel with average energy per symbol $\mathsf{E_s}$, noise variance $\sigma^2 = \mathsf{N_0}/2$ per dimension, and $\mathsf{SNR} = \frac{\mathsf{E_s}}{2\sigma^2}$ is

$$\mathsf{C_{AWGN-D}} = \log\left(1 + \mathsf{SNR}\right) \text{ [bits/channel use] or [bits/symbol]},$$

and is achieved by a Gaussian input distribution, i.e., $X \sim \mathcal{CN}(0, \mathsf{E_s})$.

Capacity of the AWGN channel

Capacity of the Discrete-Time AWGN Channel

The channel capacity of the discrete-time complex AWGN channel with average energy per symbol $\mathsf{E_s}$, noise variance $\sigma^2 = \mathsf{N_0}/2$ per dimension, and $\mathsf{SNR} = \frac{\mathsf{E_s}}{2\sigma^2}$ is

$$\mathsf{C_{AWGN-D}} = \log\left(1 + \mathsf{SNR}\right) \; [\text{bits/channel use}] \text{ or } [\text{bits/symbol}],$$

and is achieved by a Gaussian input distribution, i.e., $X \sim \mathcal{CN}(0, \mathsf{E_s})$.

Capacity of the Continuous-Time AWGN Channel

The channel capacity of the continuous-time AWGN channel of bandwidth W is

$$\mathsf{C_{AWGN-C}} = W \log\left(1 + \mathsf{SNR}\right) \quad [\text{bits/s}],$$

where the $\mathsf{SNR} = \frac{P}{\mathsf{N_0} W}$.

Capacity of the AWGN channel

Capacity of the Discrete-Time AWGN Channel

The channel capacity of the discrete-time complex AWGN channel with average energy per symbol $\mathsf{E_s}$, noise variance $\sigma^2 = \mathsf{N_0}/2$ per dimension, and $\mathsf{SNR} = \frac{\mathsf{E_s}}{2\sigma^2}$ is

$$\mathsf{C_{AWGN-D}} = \log\left(1 + \mathsf{SNR}\right) \text{ [bits/channel use] or [bits/symbol]},$$

and is achieved by a Gaussian input distribution, i.e., $X \sim \mathcal{CN}(0, \mathsf{E_s})$.

Capacity of the Continuous-Time AWGN Channel

The channel capacity of the continuous-time AWGN channel of bandwidth W is

$$\mathsf{C_{AWGN-C}} = W \log\left(1 + \mathsf{SNR}\right) \quad \text{[bits/s]},$$

where the $\mathsf{SNR} = \frac{P}{\mathsf{N_0} W}$.

The capacity $\mathsf{C_{AWGN-C}}$ depends on only two parameters, the channel bandwidth W and the SNR.

Capacity of the AWGN channel

Capacity of the Discrete-Time AWGN Channel

The channel capacity of the discrete-time real AWGN channel with average energy per symbol $\mathsf{E_s}$, noise variance $\sigma^2 = \mathsf{N_0}/2$, and $\mathsf{SNR} = \frac{\mathsf{E_s}}{\sigma^2}$ is

$$\mathsf{C_{AWGN-D}} = \frac{1}{2}\log\left(1+\mathsf{SNR}\right)\ [\text{bits/channel use}] \text{ or } [\text{bits/symbol}],$$

and is achieved by a Gaussian input distribution, i.e., $X \sim \mathcal{N}(0, \mathsf{E_s})$.

The Channel Coding Theorem for the AWGN Channel

Theorem (Channel Coding Theorem, Discrete-Time Channel)

All rates R below $\mathsf{C}_{\mathsf{AWGN-D}}$ are achievable, i.e., for every $R < \mathsf{C}_{\mathsf{AWGN-D}}$ there exists a sequence of coding schemes with vanishing error probability $P_{\mathsf{e}}^{(N)} \to 0$ as the block length $N \to \infty$. Conversely, any sequence of coding schemes of rate R and block length N with error probability $P_e^N \to 0$ must have a rate $R < \mathsf{C}_{\mathsf{AWGN-D}}$.

The Channel Coding Theorem for the AWGN Channel

Theorem (Channel Coding Theorem, Discrete-Time Channel)

All rates R below $\mathsf{C}_{\mathsf{AWGN-D}}$ are achievable, i.e., for every $R < \mathsf{C}_{\mathsf{AWGN-D}}$ there exists a sequence of coding schemes with vanishing error probability $P_{\mathsf{e}}^{(N)} \to 0$ as the block length $N \to \infty$. Conversely, any sequence of coding schemes of rate R and block length N with error probability $P_e^N \to 0$ must have a rate $R < \mathsf{C}_{\mathsf{AWGN-D}}$.

What the Theorem Tells Us: Achievable Rates

Arbitrarily reliable transmission can be achieved on the discrete-time channel at any rate $R < \mathsf{C}_{\mathsf{AWGN-D}}$ if codes of large length are used.

The Channel Coding Theorem for the AWGN Channel

Theorem (Channel Coding Theorem, Discrete-Time Channel)

All rates R below $\mathsf{C}_{\mathsf{AWGN-D}}$ are achievable, i.e., for every $R < \mathsf{C}_{\mathsf{AWGN-D}}$ there exists a sequence of coding schemes with vanishing error probability $P_{\mathsf{e}}^{(N)} \to 0$ as the block length $N \to \infty$. Conversely, any sequence of coding schemes of rate R and block length N with error probability $P_e^N \to 0$ must have a rate $R < \mathsf{C}_{\mathsf{AWGN-D}}$.

What the Theorem Tells Us: Achievable Rates

Arbitrarily reliable transmission can be achieved on the discrete-time channel at any rate $R < \mathsf{C}_{\mathsf{AWGN-D}}$ if codes of large length are used. $\longrightarrow$ Rate R is achievable.

The Channel Coding Theorem for the AWGN Channel

Theorem (Channel Coding Theorem, Discrete-Time Channel)

All rates R below $\mathsf{C}_{\mathsf{AWGN-D}}$ are achievable, i.e., for every $R < \mathsf{C}_{\mathsf{AWGN-D}}$ there exists a sequence of coding schemes with vanishing error probability $P_{\mathsf{e}}^{(N)} \to 0$ as the block length $N \to \infty$. Conversely, any sequence of coding schemes of rate R and block length N with error probability $P_e^N \to 0$ must have a rate $R < \mathsf{C}_{\mathsf{AWGN-D}}$.

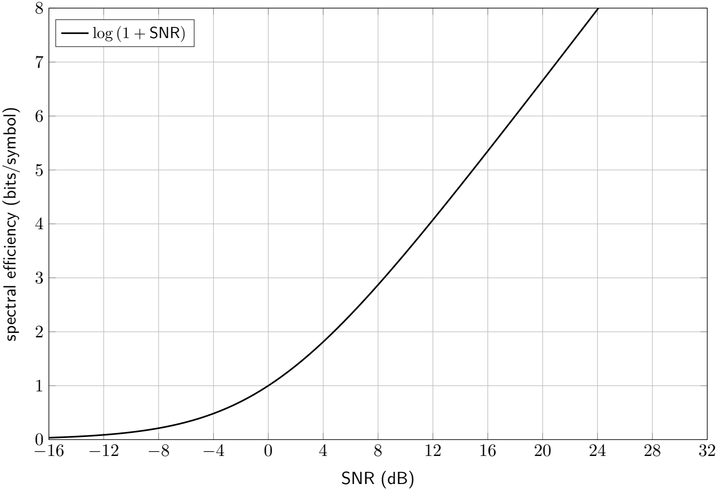
What the Theorem Tells Us: Achievable Rates

Arbitrarily reliable transmission can be achieved on the discrete-time channel at any rate $R < \mathsf{C}_{\mathsf{AWGN-D}}$ if codes of large length are used. $\longrightarrow$ Rate R is achievable.

For the Continuous-Time Channel...

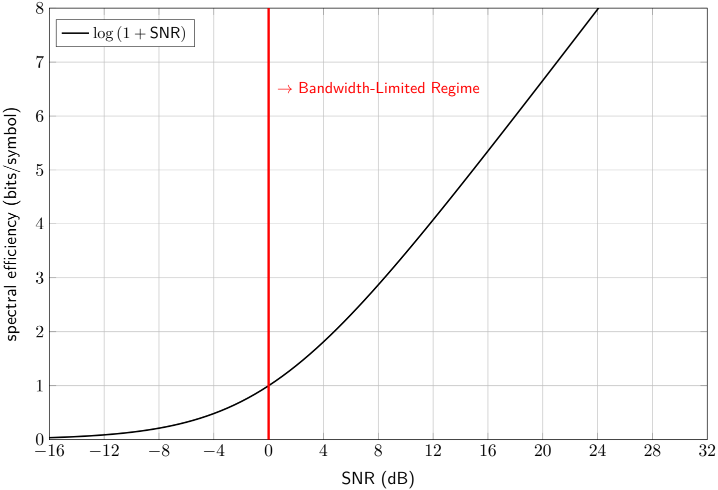
Reliable transmission can be achieved on the continuous-time channel at any bit rate R_{b} [bits/second] ($R_{\mathsf{b}} = R/T$) such that $R_{\mathsf{b}} < \mathsf{C}_{\mathsf{AWGN-C}}$.

AWGN Channel Capacity



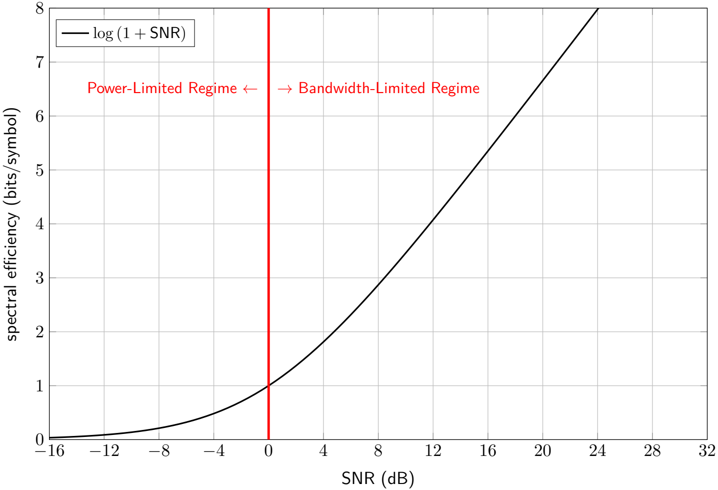
- To support spectral efficiency R, we must have SNR $> 2^R - 1$

AWGN Channel Capacity



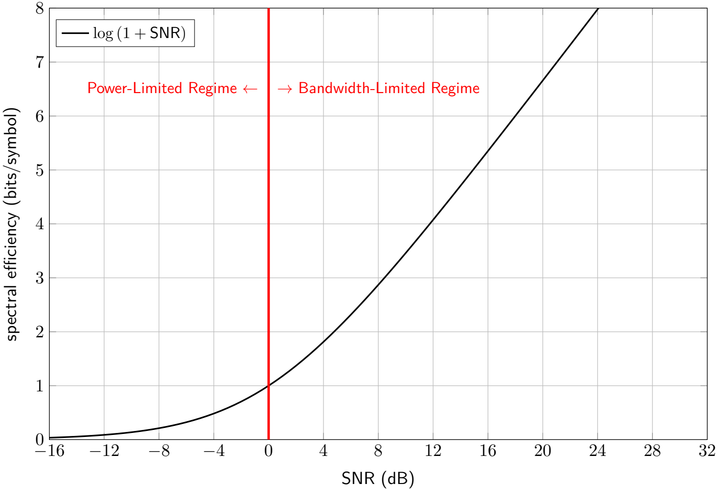
- To support spectral efficiency R, we must have SNR $> 2^R - 1$

AWGN Channel Capacity



- To support spectral efficiency R, we must have SNR $> 2^R - 1$

AWGN Channel Capacity



- To support spectral efficiency R, we must have SNR $> 2^R - 1$
- The capacity behaves quite differently in the two regimes

Power-Limited and Bandwidth-Limited Regimes

Power-Limited and Bandwidth-Limited Channels

Ideal band-limited AWGN channels may be classified as bandwidth-limited ($\mathsf{SNR} \gg 1$) or power-limited ($\mathsf{SNR} \ll 1$) according to whether they permit transmission at high spectral efficiencies or not.

Power-Limited Regime

Power-Limited Regime, SNR$\ll 1$ (While W Can Grow Very Large)

We can approximate channel capacity as

$$\mathsf{C}_{\mathsf{AWGN-D}} = \log(1 + \mathsf{SNR})$$

Power-Limited Regime

Power-Limited Regime, SNR$\ll 1$ (While W Can Grow Very Large)

We can approximate channel capacity as

$$\mathsf{C}_{\mathsf{AWGN-D}} = \log(1 + \mathsf{SNR}) \approx$$

Power-Limited Regime

Power-Limited Regime, SNR$\ll 1$ (While W Can Grow Very Large)

We can approximate channel capacity as

$$\mathsf{C}_{\mathsf{AWGN-D}} = \log(1 + \mathsf{SNR}) \approx \frac{1}{\ln 2}\mathsf{SNR},$$

Power-Limited Regime

Power-Limited Regime, SNR$\ll 1$ (While W Can Grow Very Large)

We can approximate channel capacity as

$$\mathsf{C}_{\mathsf{AWGN-D}} = \log(1+\mathsf{SNR}) \approx \frac{1}{\ln 2}\mathsf{SNR},$$

i.e., $\mathsf{C}_{\mathsf{AWGN-D}}$ increases linearly with SNR.

Power-Limited Regime

Power-Limited Regime, SNR$\ll 1$ (While W Can Grow Very Large)

We can approximate channel capacity as

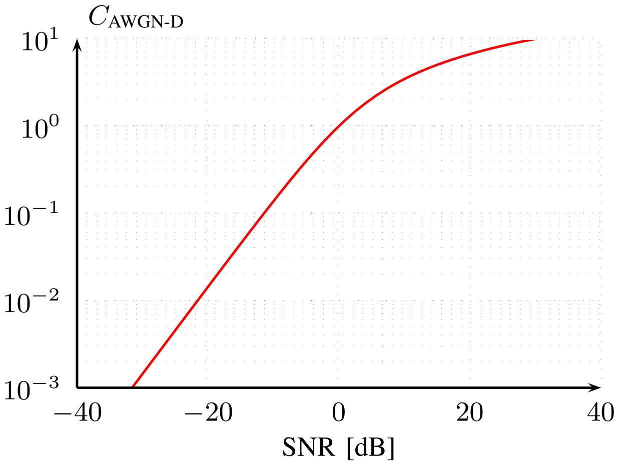
$$\mathsf{C}_{\mathsf{AWGN-D}} = \log(1+\mathsf{SNR}) \approx \frac{1}{\ln 2}\mathsf{SNR},$$

i.e., $\mathsf{C}_{\mathsf{AWGN-D}}$ increases linearly with SNR.

In the Power-Limited Regime

- Doubling the SNR doubles the capacity.

Capacity Curve



- For low SNR the gain is linear.

Bandwidth-Limited Regime

Bandwidth-Limited Regime, SNR$\gg 1$

When the SNR is large, we have

$$\mathsf{C}_{\mathsf{AWGN-D}} = \log(1 + \mathsf{SNR})$$

Bandwidth-Limited Regime

Bandwidth-Limited Regime, SNR$\gg 1$

When the SNR is large, we have

$$\begin{aligned}\mathsf{C}_{\mathsf{AWGN-D}} &= \log(1+\mathsf{SNR})\\ &\approx \log \mathsf{SNR},\end{aligned}$$

Bandwidth-Limited Regime

Bandwidth-Limited Regime, SNR$\gg 1$

When the SNR is large, we have

$$\begin{aligned} \mathsf{C}_{\mathsf{AWGN-D}} &= \log(1+\mathsf{SNR}) \\ &\approx \log \mathsf{SNR}, \end{aligned}$$

i.e., $\mathsf{C}_{\mathsf{AWGN-D}}$ increases logarithmically with SNR.

Bandwidth-Limited Regime

Bandwidth-Limited Regime, SNR$\gg 1$

When the SNR is large, we have

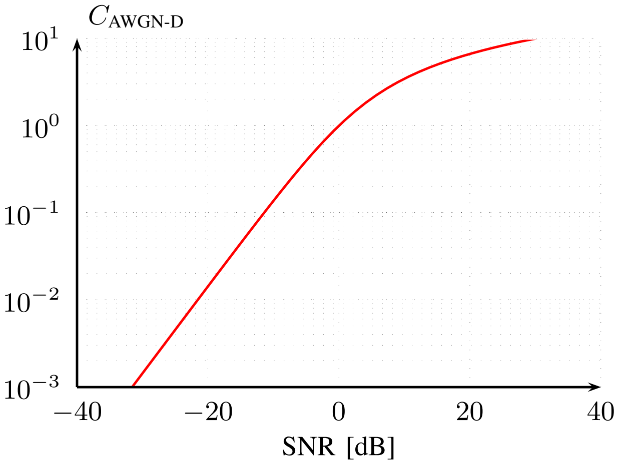
$$\begin{aligned} \mathsf{C}_{\mathsf{AWGN-D}} &= \log(1+\mathsf{SNR}) \\ &\approx \log \mathsf{SNR}, \end{aligned}$$

i.e., $\mathsf{C}_{\mathsf{AWGN-D}}$ increases logarithmically with SNR.

In the Bandwidth-Limited Regime

- Doubling the SNR (every additional 3 dB in SNR) yields an increase in achievable spectral efficiency of only 1 (bit/s)/Hz.

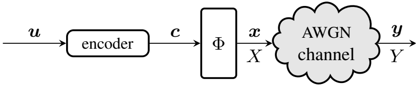
Capacity Curve



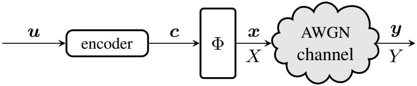
- For low SNR the gain is linear.
- For high SNR the gain is only logarithmic.

How Do We Design Good Coding Schemes in Practice?

How Do We Design Good Coding Schemes in Practice?

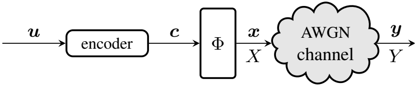


How Do We Design Good Coding Schemes in Practice?



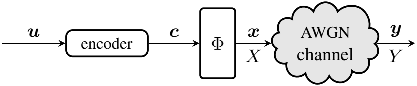
- A binary channel encoder maps information bits $\boldsymbol{u}$ to binary codewords $\boldsymbol{c}$

How Do We Design Good Coding Schemes in Practice?



- A binary channel encoder maps information bits $\boldsymbol{u}$ to binary codewords $\boldsymbol{c}$
- A symbol mapper Φ maps the binary codewords to a sequence of complex symbols $\boldsymbol{x} = (x_1, x_2, \dots, x_N)$

How Do We Design Good Coding Schemes in Practice?



- A binary channel encoder maps information bits $\boldsymbol{u}$ to binary codewords $\boldsymbol{c}$
- A symbol mapper Φ maps the binary codewords to a sequence of complex symbols $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$
- Each symbol x_i takes on values in $\mathcal{X} = \{\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_M\} \subset \mathbb{C}$ according to $P_X(x_i)$
- Example: M-QAM constellations with $P_X(x) = 1/M$ for all $x \in \mathcal{X}$

How Do We Design Good Coding Schemes in Practice?

- A binary channel encoder maps information bits $\boldsymbol{u}$ to binary codewords $\boldsymbol{c}$
- A symbol mapper Φ maps the binary codewords to a sequence of complex symbols $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$
- Each symbol x_i takes on values in $\mathcal{X} = \{\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_M\} \subset \mathbb{C}$ according to $P_X(x_i)$
- Example: M-QAM constellations with $P_X(x) = 1/M$ for all $x \in \mathcal{X}$

How Do We Design Good Coding Schemes in Practice?

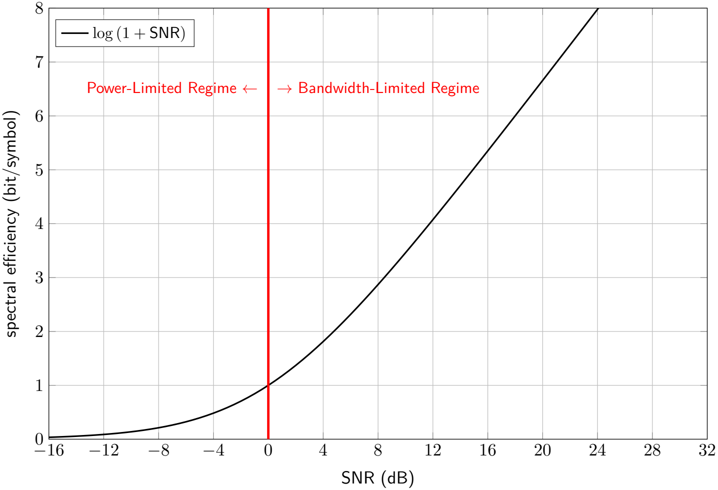
- A binary channel encoder maps information bits $\boldsymbol{u}$ to binary codewords $\boldsymbol{c}$
- A symbol mapper Φ maps the binary codewords to a sequence of complex symbols $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$
- Each symbol x_i takes on values in $\mathcal{X} = \{\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_M\} \subset \mathbb{C}$ according to $P_X(x_i)$
- Example: M-QAM constellations with $P_X(x) = 1/M$ for all $x \in \mathcal{X}$
- At the receiver, apply appropriate "inverse" operations: demapper Φ^{-1} and channel decoder

How Do We Design Good Coding Schemes in Practice?

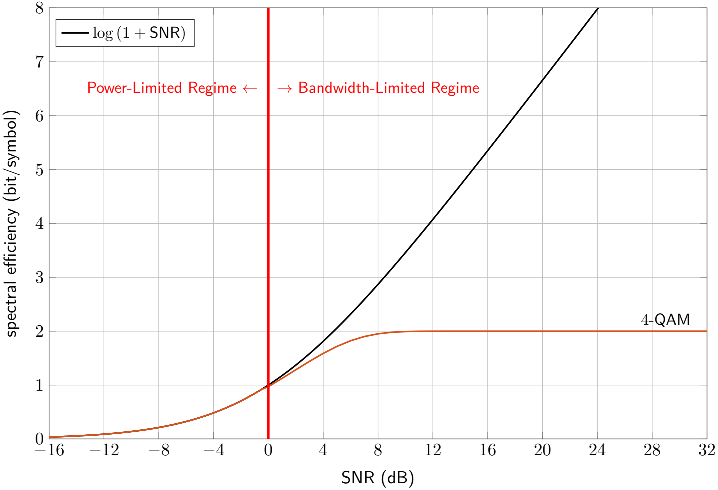
- A binary channel encoder maps information bits $\boldsymbol{u}$ to binary codewords $\boldsymbol{c}$
- A symbol mapper Φ maps the binary codewords to a sequence of complex symbols $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$
- Each symbol x_i takes on values in $\mathcal{X} = \{\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_M\} \subset \mathbb{C}$ according to $P_X(x_i)$
- Example: M-QAM constellations with $P_X(x) = 1/M$ for all $x \in \mathcal{X}$
- At the receiver, apply appropriate "inverse" operations: demapper Φ^{-1} and channel decoder

How much do we loose by restricting ourselves to M-QAM?

Mutual Information of M-QAM

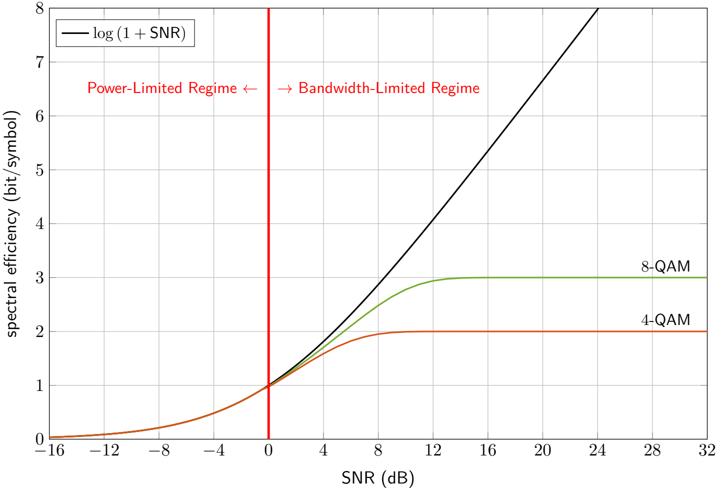


Mutual Information of M-QAM



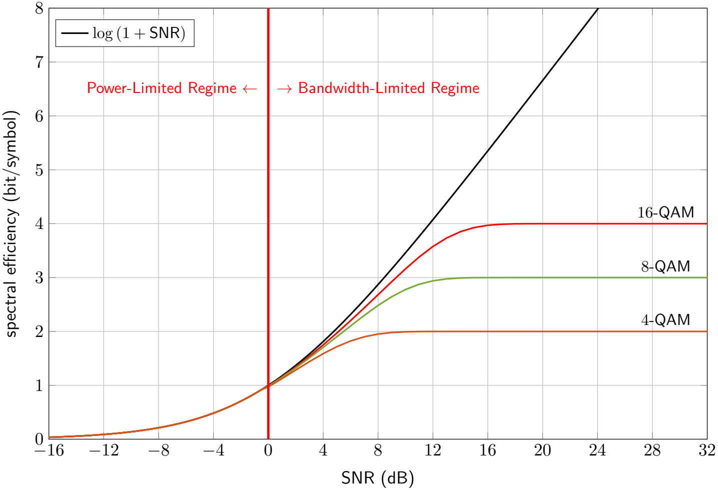
- Power-Limited: We can restrict ourselves to 4-QAM (i.e., binary signalling in each dimension) with almost no loss

Mutual Information of M-QAM



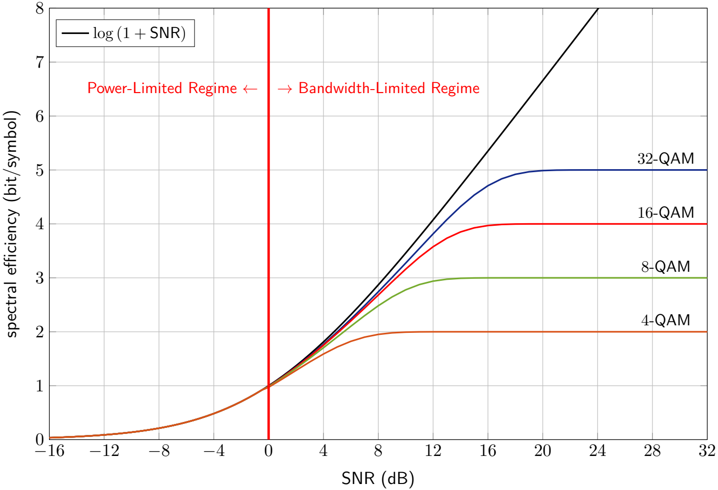
- Power-Limited: We can restrict ourselves to 4-QAM (i.e., binary signalling in each dimension) with almost no loss

Mutual Information of M-QAM



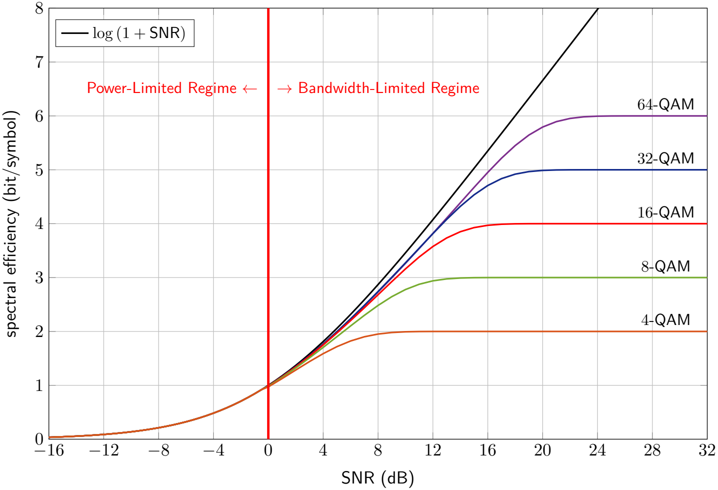
- Power-Limited: We can restrict ourselves to 4-QAM (i.e., binary signalling in each dimension) with almost no loss

Mutual Information of M-QAM



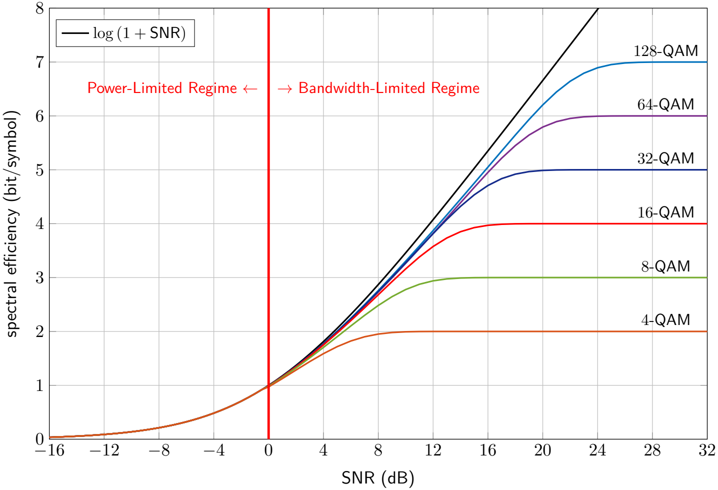
- Power-Limited: We can restrict ourselves to 4-QAM (i.e., binary signalling in each dimension) with almost no loss

Mutual Information of M-QAM



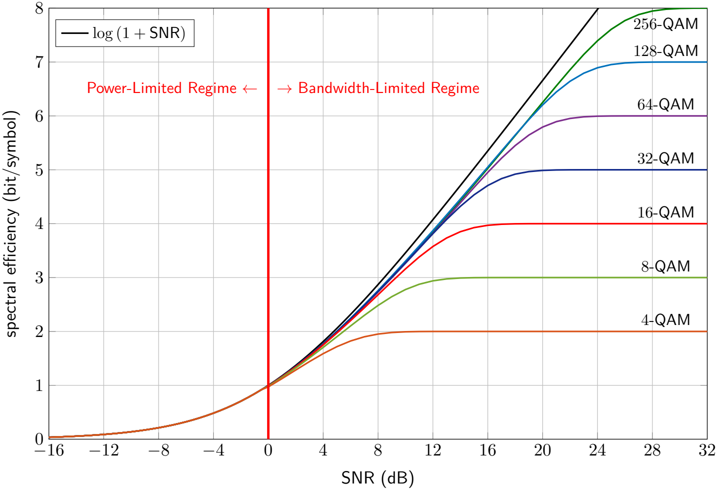
- Power-Limited: We can restrict ourselves to 4-QAM (i.e., binary signalling in each dimension) with almost no loss

Mutual Information of M-QAM



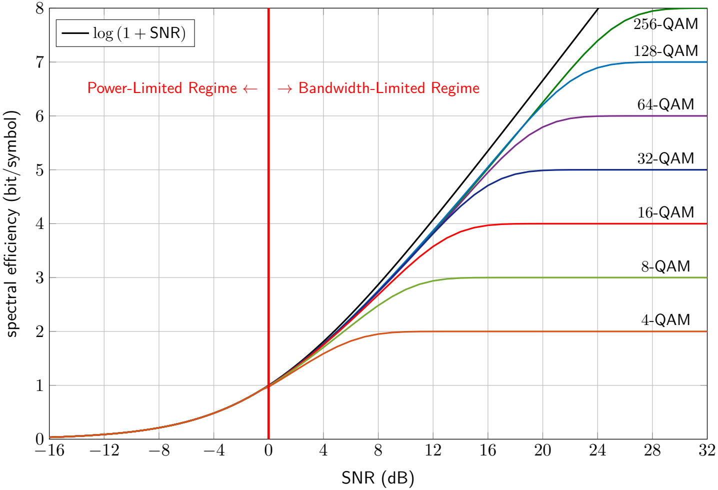
- Power-Limited: We can restrict ourselves to 4-QAM (i.e., binary signalling in each dimension) with almost no loss

Mutual Information of M-QAM



- Power-Limited: We can restrict ourselves to 4-QAM (i.e., binary signalling in each dimension) with almost no loss

Mutual Information of M-QAM



- Bandwidth-Limited: We need to use higher-order QAM constellations
- There is a gap between the mutual information curves and the capacity.

Power Efficiency and Energy per Information Bit

- To compare coded communication systems we consider the ratio

$$\frac{\mathsf{E_b}}{\mathsf{N_0}},$$

where $\mathsf{E_b}$ is the energy per information bit. $\mathsf{E_b}/\mathsf{N_0}$ is referred to as the power efficiency.

Power Efficiency and Energy per Information Bit

- To compare coded communication systems we consider the ratio

$$\frac{\mathsf{E_b}}{\mathsf{N_0}},$$

  where $\mathsf{E_b}$ is the energy per information bit. $\mathsf{E_b}/\mathsf{N_0}$ is referred to as the power efficiency.

- For a coded system that encodes sequences $\boldsymbol{u}$ of K bits onto sequences $\boldsymbol{x}$ of N symbols with average energy per symbol $\mathsf{E_s}$. Then,

$$\mathsf{E_b} =$$

Power Efficiency and Energy per Information Bit

- To compare coded communication systems we consider the ratio

$$\frac{\mathsf{E_b}}{\mathsf{N_0}},$$

where $\mathsf{E_b}$ is the energy per information bit. $\mathsf{E_b}/\mathsf{N_0}$ is referred to as the power efficiency.

- For a coded system that encodes sequences $\boldsymbol{u}$ of K bits onto sequences $\boldsymbol{x}$ of N symbols with average energy per symbol $\mathsf{E_s}$. Then,

$$\mathsf{E_b} = \frac{\mathsf{E_s}}{R}.$$

Power Efficiency and Energy per Information Bit

- To compare coded communication systems we consider the ratio

$$\frac{\mathsf{E_b}}{\mathsf{N_0}},$$

where $\mathsf{E_b}$ is the energy per information bit. $\mathsf{E_b}/\mathsf{N_0}$ is referred to as the power efficiency.

- For a coded system that encodes sequences $\boldsymbol{u}$ of K bits onto sequences $\boldsymbol{x}$ of N symbols with average energy per symbol $\mathsf{E_s}$. Then,

$$\mathsf{E_b} = \frac{\mathsf{E_s}}{R}.$$

- $\mathsf{SNR} = \frac{\mathsf{E_s}}{\mathsf{N_0}} = R\frac{\mathsf{E_b}}{\mathsf{N_0}}$.

Power Efficiency and Bandwidth Fundamental Tradeoff

Communication is reliable if

$$R < \mathsf{C}_{\mathsf{AWGN-D}}$$

Power Efficiency and Bandwidth Fundamental Tradeoff

Communication is reliable if

$$\begin{aligned} R &< \mathsf{C}_{\mathsf{AWGN-D}} \\ &= \log\left(1 + \frac{\mathsf{E_s}}{2\sigma^2}\right) \end{aligned}$$

Power Efficiency and Bandwidth Fundamental Tradeoff

Communication is reliable if

$$\begin{aligned} R &< \mathsf{C}_{\mathsf{AWGN-D}} \\ &= \log\left(1 + \frac{\mathsf{E_s}}{2\sigma^2}\right) \\ &= \log\left(1 + R\frac{\mathsf{E_b}}{2\sigma^2}\right) \end{aligned}$$

Power Efficiency and Bandwidth Fundamental Tradeoff

Communication is reliable if

$$
\begin{aligned}
R &< \mathsf{C}_{\mathsf{AWGN-D}} \\
&= \log\left(1 + \frac{\mathsf{E_s}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{\mathsf{N_0}}\right).
\end{aligned}
$$

Power Efficiency and Bandwidth Fundamental Tradeoff

Communication is reliable if

$$
\begin{aligned}
R &< \mathsf{C}_{\mathsf{AWGN-D}} \\
&= \log\left(1 + \frac{\mathsf{E_s}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{\mathsf{N_0}}\right).
\end{aligned}
$$

For reliable transmission, the minimum $\mathsf{E_b}/\mathsf{N_0}$ to support a rate R is

$$
\frac{\mathsf{E_b}}{\mathsf{N_0}} > \frac{2^R - 1}{R}.
$$

Power Efficiency and Bandwidth Fundamental Tradeoff

Communication is reliable if

$$
\begin{aligned}
R &< \mathsf{C}_{\mathsf{AWGN-D}} \\
&= \log\left(1 + \frac{\mathsf{E_s}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{\mathsf{N_0}}\right).
\end{aligned}
$$

For reliable transmission, the minimum $\mathsf{E_b}/\mathsf{N_0}$ to support a rate R is

$$
\frac{\mathsf{E_b}}{\mathsf{N_0}} > \frac{2^R - 1}{R}.
$$

Using, $\lim_{R\to 0} \frac{2^R-1}{R} = \ln 2$,

$$
\mathsf{E_b}/\mathsf{N_0} > \ln 2 = -1.59 \text{ dB},
$$

Power Efficiency and Bandwidth Fundamental Tradeoff

Communication is reliable if

$$
\begin{aligned}
R &< \mathsf{C}_{\mathsf{AWGN-D}} \\
&= \log\left(1 + \frac{\mathsf{E_s}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{2\sigma^2}\right) \\
&= \log\left(1 + R\frac{\mathsf{E_b}}{\mathsf{N_0}}\right).
\end{aligned}
$$

For reliable transmission, the minimum $\mathsf{E_b}/\mathsf{N_0}$ to support a rate R is

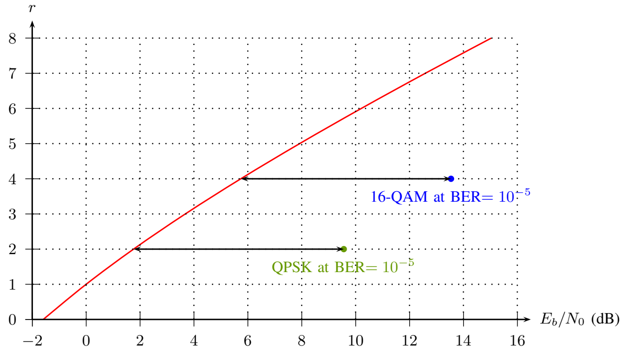
$$
\frac{\mathsf{E_b}}{\mathsf{N_0}} > \frac{2^R - 1}{R}.
$$

Using, $\lim_{R\to 0} \frac{2^R-1}{R} = \ln 2$,

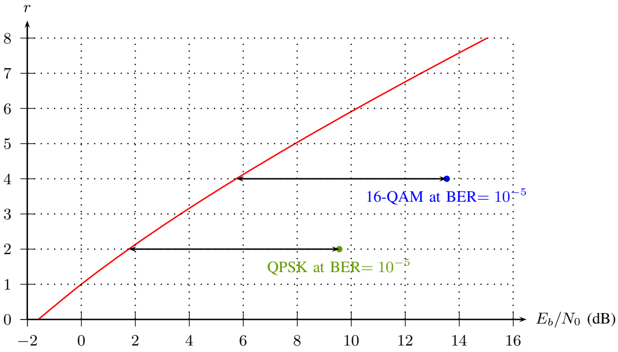
$$
\mathsf{E_b}/\mathsf{N_0} > \ln 2 = -1.59 \text{ dB},
$$

i.e., it is not possible to transmit reliably over the AWGN channel at $\mathsf{E_b}/\mathsf{N_0}$ smaller than -1.59 dB, even when we let $R \to 0!$.

Power Efficiency and Bandwidth Fundamental Tradeoff

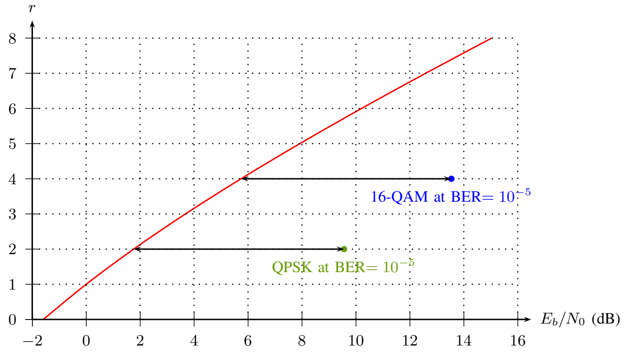


Power Efficiency and Bandwidth Fundamental Tradeoff



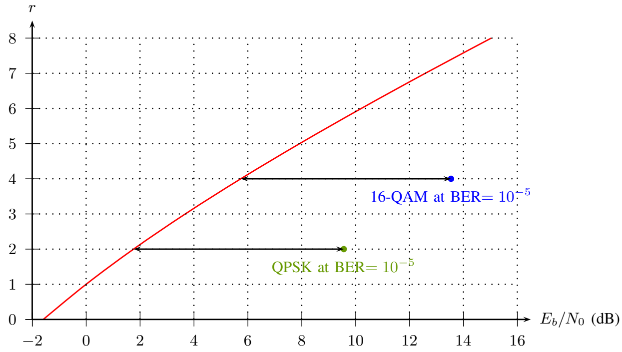
- All rates R below the red curve are achievable.

Power Efficiency and Bandwidth Fundamental Tradeoff



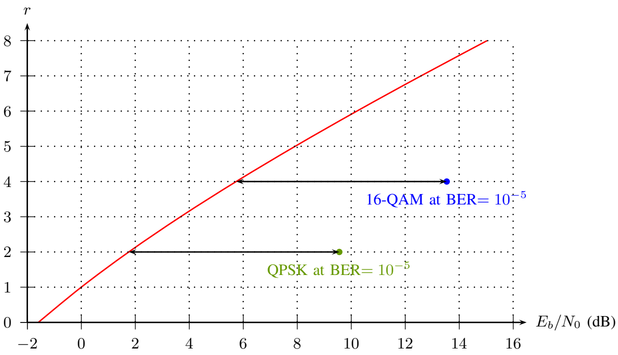
- All rates R below the red curve are achievable.
- For a given R, all values of $\mathsf{E_b/N_0}$ in the right of the red curve are achievable.

Power Efficiency and Bandwidth Fundamental Tradeoff



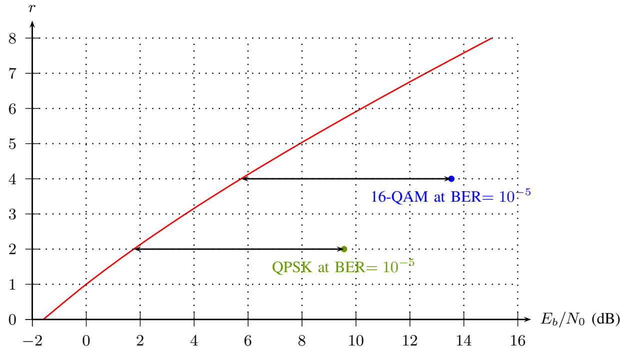
- All rates R below the red curve are achievable.
- For a given R, all values of $\mathsf{E_b/N_0}$ in the right of the red curve are achievable.
- R is usually referred to as spectral efficiency, since determines how efficiently we use the spectrum ($R = R_\mathsf{b}T = R_\mathsf{b}/W$).

Power Efficiency and Bandwidth Fundamental Tradeoff

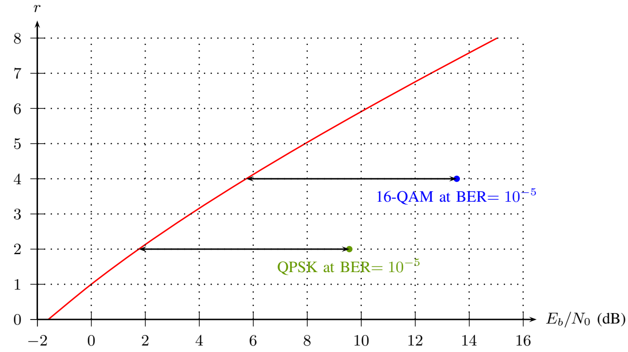


- All rates R below the red curve are achievable.
- For a given R, all values of $\mathsf{E_b}/\mathsf{N_0}$ in the right of the red curve are achievable.
- R is usually referred to as spectral efficiency, since determines how efficiently we use the spectrum $(R = R_\mathsf{b} T = R_\mathsf{b}/W)$.
- Uncoded transmission performs far away from the theoretical limit.

Power Efficiency and Bandwidth Fundamental Tradeoff

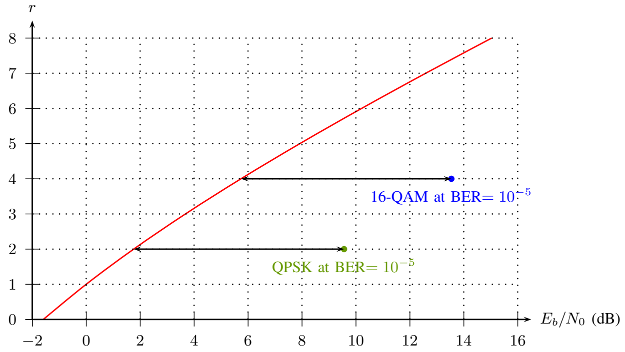


Power Efficiency and Bandwidth Fundamental Tradeoff



A fundamental tradeoff between power and bandwidth

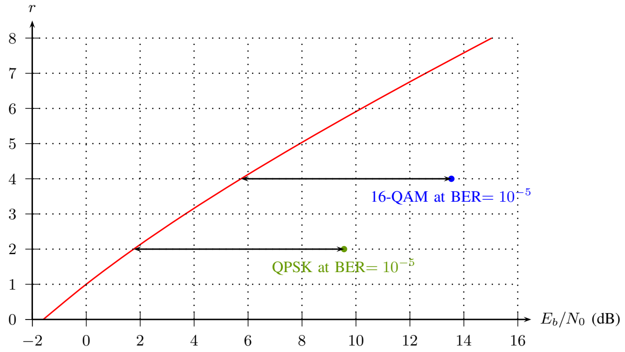
Power Efficiency and Bandwidth Fundamental Tradeoff



A fundamental tradeoff between power and bandwidth

- The required $\mathsf{E_b/N_0}$ increases with increasing R, while increasing R decreases the required bandwidth to support the same information rate R_b.

Power Efficiency and Bandwidth Fundamental Tradeoff



A fundamental tradeoff between power and bandwidth

- The required $\mathsf{E_b/N_0}$ increases with increasing R, while increasing R decreases the required bandwidth to support the same information rate R_b.
- Decreasing R requires less $\mathsf{E_b/N_0}$, but higher bandwidth to support the same R_b.