Information Theory Problems

• How to transmit or store information as efficiently as possible.
• What is the maximum amount of information that can be transmitted or stored reliably?
• How can information be kept secure?
Digital Communications System

\[ \hat{W} \] is an estimate of \( W \)

Data source

Source encoder → Channel encoder → Digital modulator

Channel

Data sink

Source decoder ← Channel decoder ← Digital demodulator

\( W \) \hspace{1cm} \( X \) \hspace{1cm} \( P_X \) \hspace{1cm} \( P_{Y|X} \) \hspace{1cm} \( Y \) \hspace{1cm} \( P_Y \)
Communication Channel

Encoder -> Channel $p(y|x)$ -> Decoder

$W$ Message $\rightarrow X^n \rightarrow Y^n \rightarrow \hat{W}$ Estimate of Message
Discrete Memoryless Channel

$X \xrightarrow{X_1} Y \xrightarrow{y_1}
X_2 \xrightarrow{\cdot \cdot \cdot} Y_2
X_k \xrightarrow{\cdot \cdot \cdot} Y_k
X_{J} \xrightarrow{\cdot \cdot \cdot} Y_{J}$

$J \geq K \geq 2$
Discrete Memoryless Channel

Channel Transition Matrix

\[
P = \begin{bmatrix}
p(y_1 | x_1) & \cdots & p(y_1 | x_k) & \cdots & p(y_1 | x_K) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
p(y_j | x_1) & \cdots & p(y_j | x_k) & \cdots & p(y_j | x_K) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
p(y_j | x_1) & \cdots & p(y_j | x_k) & \cdots & p(y_j | x_K)
\end{bmatrix}
\]
Binary Symmetric Channel

Transition probabilities

\[
\begin{pmatrix}
0 & 1 - p \\
p & 1 - p \\
p & p \\
p & 0
\end{pmatrix}
\]

channel matrix
Binary Errors with Erasure Channel

\[ x_1 \quad 0 \]
\[ x_2 \quad 1 \]

\[ \begin{array}{c}
0 \\
1-p-q \\
p \\
q
\end{array} \quad \begin{array}{c}
0 \\
q \\
p \\
1-p-q
\end{array} \quad \begin{array}{c}
e \\
y_2
\end{array} \quad \begin{array}{c}
y_1
\end{array} \quad \begin{array}{c}
y_3
\end{array} \quad \begin{array}{c}
1
\end{array} \]
BPSK Modulation $K=2$

\[ x_1(t) = +\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \quad 0 \text{ bit} \]

\[ x_2(t) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t) \quad 1 \text{ bit} \]

$E_b$ is the energy per bit

$T_b$ is the bit duration

$f_c$ is the carrier frequency
BPSK Demodulation in AWGN

\[
p(y|x_2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - x_2)^2}{2\sigma^2}\right]
\]

\[
p(y|x_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - x_1)^2}{2\sigma^2}\right]
\]
BPSK Demodulation $K=2$

For $J=3$:

- $\sqrt{E_b}$
- $0$
- $+\sqrt{E_b}$

$y_3 \quad y_2 \quad y_1$

For $J=4$:

- $\sqrt{E_b}$
- $0$
- $+\sqrt{E_b}$

$y_4 \quad y_3 \quad y_2 \quad y_1$

For $J=8$:

- $\sqrt{E_b}$
- $0$
- $+\sqrt{E_b}$

$y_8 \quad y_1$
Mutual Information for a BSC

channel matrix

$$P_{Y|X} = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$

crossover probability $p$

$\bar{p} = 1 - p$

$p(x = 0) = w$

$p(x = 1) = 1 - w = \bar{w}$
Mutual information $I(X; Y)$

$1 - h(p)$

Probability of a "0" at input, $\omega$
**Convex Functions**

**Definition (Convex function):**

A real function $f(x)$, defined on a convex set $S$ (e.g., input symbol distributions), is **concave** (convex down, convex “cap” or convex ∩) if, for any point $x$ on the straight line between the pair of points $x_1$ and $x_2$, i.e., $x = \lambda x_1 + (1 - \lambda)x_2$ ($\lambda \in [0, 1]$), in the convex set $S$:

\[
f(x) \geq \lambda f(x_1) + (1 - \lambda)f(x_2)
\]

otherwise, if:

\[
f(x) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)
\]

then the function is said to be simply **convex** (convex up, convex “cup” or convex ∪).
Concave Function

Figure 3.3: Convex \( \cap \) (convex down or convex “cap”) function.
Convex Function

Figure 3.4: Convex $\cup$ (convex up or convex “cup”) function.
Mutual Information

\[ I(X; Y) = \sum_{k=1}^{K} \sum_{j=1}^{J} p(x_k)p(y_j|x_k) \log_b \left[ \frac{p(y_j|x_k)}{\sum_{l=1}^{K} p(x_l)p(y_j|x_l)} \right] \]

\[ = f [p(x_k), p(y_j|x_k)] \]

\[ I(X; Y) = f (p, P) \]
Theorem (Convexity of the mutual information function):

The (average) mutual information $I(X; Y)$ is a concave (or convex “cap”, or convex $\cap$) function over the convex set $S_p$ of all possible input distributions $\{p\}$. 
\[ I(X;Y) = I(p;P) \]
Mutual information $I(X;Y)$

$1 - h(p)$

Probability of a "0" at input, $\omega$
Theorem (Convexity of the mutual information function):

The (average) mutual information $I(X;Y)$ is a convex (or convex “cup”, or convex $\cup$) function over the convex set $S_P$ of all possible transition probability matrices $\{P\}$. 
\[ I(X; Y) = I(p; P) \]
BSC $I(X;Y)$
Channel Capacity

The *maximum* value of $I(X; Y)$ as the input probabilities $p(x_i)$ are varied is called the Channel Capacity

$$C = \max_{p(x_i)} I(X : Y)$$
Binary Symmetric Channel

\[ C = 1 - h(p) \]
Properties of the Channel Capacity

- $C \geq 0$ since $I(X;Y) \geq 0$
- $C \leq \log|X| = \log(K)$ since $C = \max I(X;Y) \leq \max H(X) = \log(K)$
- $C \leq \log|Y| = \log(J)$ for the same reason
- $I(X;Y)$ is a concave function of $p(X)$, so a local maximum is a global maximum
Symmetric Channels

A discrete memoryless channel is said to be **symmetric** if the set of output symbols \(\{y_j\}, j = 1, 2, ..., J,\) can be partitioned into subsets such that for each subset of the matrix of transition probabilities, each column is a permutation of the other columns, and each row is also a permutation of the other rows.
Binary Channels

Symmetric channel matrix

\[ P = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \]

Non-symmetric channel matrix

\[ P = \begin{bmatrix} 1-p_1 & p_2 \\ p_1 & 1-p_2 \end{bmatrix} \quad p_1 \neq p_2 \]
Binary Errors with Erasure Channel

\[ x_1 \quad 0 \quad 1-p-q \quad 0 \quad y_1 \]
\[ x_2 \quad 1 \quad p \quad q \quad e \quad y_2 \]
\[ \quad p \quad q \quad 1-p-q \quad 1-p-q \quad 1 \quad y_3 \]
Symmetric Channels

- No partition required $\rightarrow$ strongly symmetric
- Partition required $\rightarrow$ weakly symmetric
Capacity of a **Strongly Symmetric** Channel

**Theorem**

For a discrete symmetric channel, the channel capacity $C$ is achieved with an equiprobable input distribution, i.e., $p(x_k) = \frac{1}{K}$, $\forall k$, and is given by:

$$C = \left[ \sum_{j=1}^{J} p(y_j | x_k) \log_b p(y_j | x_k) \right] + \log_b J$$

$$I(X;Y) = \sum_{k=1}^{K} p(x_k) \sum_{j=1}^{J} p(y_j | x_k) \log p(y_j | x_k) + H(Y)$$

$$= \sum_{j=1}^{J} p(y_j | x_k) \log p(y_j | x_k) + H(Y)$$
Example $J = K = 3$
Example

\[
P_{Y|X} = \begin{bmatrix}
0.7 & 0.2 & 0.1 \\
0.1 & 0.7 & 0.2 \\
0.2 & 0.1 & 0.7 \\
\end{bmatrix}
\]

\[
\sum_{k=1}^{K} p(x_k) \sum_{j=1}^{J} p(y_j | x_k) \log p(y_j | x_k)
\]

\[
= p(x_1)[0.7 \log 0.7 + 0.1 \log 0.1 + 0.2 \log 0.2] \\
+ p(x_2)[0.2 \log 0.2 + 0.7 \log 0.7 + 0.1 \log 0.1] \\
+ p(x_3)[0.1 \log 0.1 + 0.2 \log 0.2 + 0.7 \log 0.7]
\]

\[
= 0.7 \log 0.7 + 0.2 \log 0.2 + 0.1 \log 0.1
\]

\[
= \sum_{j=1}^{J} p(y_j | x_k) \log p(y_j | x_k)
\]
Example

\[ H(Y) = - \sum_{j=1}^{J} p(y_j) \log p(y_j) \]

\[ p(y_1) = \sum_{k=1}^{K} p(y_1 | x_k) p(x_k) \]

\[ p(y_2) = \sum_{k=1}^{K} p(y_2 | x_k) p(x_k) \]

\[ \vdots \]

\[ p(y_J) = \sum_{k=1}^{K} p(y_J | x_k) p(x_k) \]

\[ p(y_1) = .7p(x_1) + .2p(x_2) + .1p(x_3) \]

\[ p(y_2) = .1p(x_1) + .7p(x_2) + .2p(x_3) \]

\[ p(y_3) = .2p(x_1) + .1p(x_2) + .7p(x_3) \]
r-ary Symmetric Channel

\[
P = \begin{bmatrix}
1-p & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\
\frac{p}{r-1} & 1-p & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\
\frac{p}{r-1} & \frac{p}{r-1} & 1-p & \cdots & \frac{p}{r-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & 1-p 
\end{bmatrix}
\]
r-ary Symmetric Channel

\[
C = (1 - p)\log(1 - p) + (r - 1)\frac{p}{r - 1}\log\left(\frac{p}{r - 1}\right) + \log r
\]

\[
= \log r + (1 - p)\log(1 - p) + p\log\left(\frac{p}{r - 1}\right)
\]

\[
= \log r + (1 - p)\log(1 - p) + p\log(p) - p\log(r - 1)
\]

\[
= \log r - h(p) - p\log(r - 1)
\]

• If \( r = 2 \)   \( C = 1 - h(p) \)
• If \( r = 3 \)   \( C = \log_2 3 - h(p) - p \)
• If \( r = 4 \)   \( C = 2 - h(p) - p\log_2 3 \)
Binary Errors with Erasure Channel
Binary Errors with Erasure Channel

\[ P_{Y|X} = \begin{bmatrix} .8 & .05 \\ .15 & .15 \\ .05 & .8 \end{bmatrix} \]

\[ P_X = \begin{bmatrix} .5 \\ .5 \end{bmatrix} \]

\[ P_Y = P_{Y|X} \times P_X = \begin{bmatrix} .425 \\ .15 \\ .425 \end{bmatrix} \]
Capacity of a Weakly Symmetric Channel

\[ C = \sum_{i=1}^{L} q_i C_i \]

- \( q_i \) – probability of channel \( i \)
- \( C_i \) – capacity of channel \( i \)
Binary Errors with Erasure Channel

\[ P_1 = \begin{bmatrix} .9412 & .0588 \\ .0588 & .9412 \end{bmatrix} \]

\[ P_2 = \begin{bmatrix} 1.0 & 1.0 \end{bmatrix} \]
Binary Errors with Erasure Channel

\[ C = \sum_{i=1}^{L} q_i C_i \]

\[ C_1 = 0.9412 \log_{10} 0.9412 + 0.0588 \log_{10} 0.0588 + \log_{10} 2 = 0.6773 \]

\[ C_2 = 1 \log_{10} 1 + \log_{10} 1 = 0.0 \]

\[ C = 0.85(0.6773) + 0.15(0.0) = 0.5757 \]
Binary Erasure Channel

\[ C = 1 - p \]
Z Channel (Optical)

0 \rightarrow 1 \rightarrow 0 \text{ (light on)}

1 \rightarrow p \rightarrow 1 \text{ (light off)}

1-p
Z Channel (Optical)

\[
I(X;Y) = \sum_{k=1}^{2} \sum_{j=1}^{2} p(x_k)p(y_j | x_k) \log \left( \frac{p(y_j | x_k)}{p(y_j)} \right)
\]

\[
l(x_1;Y) = w \log \left( \frac{1}{w + pw} \right)
\]

\[
l(x_2;Y) = \bar{w}p \log \left( \frac{p}{w + pw} \right) + \bar{w}(1-p) \log \left( \frac{1-p}{(1-p)w} \right)
\]

\[
w = 1 - \frac{1}{(1-p)(1+2^{h(p)/(1-p)})}
\]

\[
C = \log_2 \left( 1 + (1-p)p^{p/(1-p)} \right)
\]
Channel Capacity for the Z, BSC and BEC
Blahut-Arimoto Algorithm

\[ I(X;Y) = \sum_{k=1}^{K} p(x_k) \sum_{j=1}^{J} p(y_j \mid x_k) \log \left( \frac{p(y_j \mid x_k)}{\sum_{l=1}^{K} p(x_l) p(y_j \mid x_l)} \right) \]

- An analytic solution for the capacity can be very difficult to obtain
- The alternative is a numerical solution
  - Arimoto Jan. 1972
  - Blahut Jul. 1972
- Exploits the fact that \( I(X;Y) \) is a concave function of \( p(x_k) \)
Blahut-Arimoto Algorithm

\[ c_k = \exp \left[ \sum_{j=1}^{J} p(y_j|x_k) \ln \left( \frac{p(y_j|x_k)}{\sum_{l=1}^{K} p(x_l) p(y_j|x_l)} \right) \right] \quad \text{for } k = 1, \ldots, K \]

\[ I_L = \ln \sum_{k=1}^{K} p(x_k) c_k \]

\[ I_U = \ln \left( \max_{k=1,\ldots,K} c_k \right) \]
Blahut-Arimoto Algorithm

• Update the probabilities

\[ p^{(n+1)}(x_k) = \frac{p^{(n)}(x_k) \ c_k}{\sum_{l=1}^{K} p(x_l)^{(n)} \ c_l} \]
\[ p^{(n)} = p^{(0)} \]

\[ c_k(p^{(n)}) = \exp \left[ \sum_{j=1}^{J} p(y_j|x_k) \ln \left( \frac{p(y_j|x_k)}{\sum_{l=1}^{K} p(x_l) \ p(y_j|x_l)} \right) \right] \]

\[ I_L = \ln \sum_{k=1}^{K} p(x_k) \ c_k(p^{(n)}) \]

\[ I_U = \ln \left[ \max_{k=1,\ldots,K} c_k(p^{(n)}) \right] \]

\[ I_U - I_L < \epsilon \]

\[ C = I_L \]

\[ p^{(n+1)}(x_k) = p^{(n)}(x_k) \frac{c_k(p^{(n)})}{\sum_{l=1}^{K} p(x_l)^{(n)} \ c_l(p^{(n)})} \]

stop
Symmetric Channel Example

\[ P_1 = \begin{pmatrix} 0.4000 & 0.3000 & 0.2000 & 0.1000 \\ 0.1000 & 0.4000 & 0.3000 & 0.2000 \\ 0.3000 & 0.2000 & 0.1000 & 0.4000 \\ 0.2000 & 0.1000 & 0.4000 & 0.3000 \end{pmatrix} \]

<table>
<thead>
<tr>
<th>n</th>
<th>p(x_1)</th>
<th>p(x_2)</th>
<th>p(x_3)</th>
<th>p(x_4)</th>
<th>( I_U )</th>
<th>( I_L )</th>
<th>( \epsilon )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.1064</td>
<td>0.1064</td>
<td>0.0000</td>
</tr>
<tr>
<td>n</td>
<td>$p(x_1)$</td>
<td>$p(x_2)$</td>
<td>$p(x_3)$</td>
<td>$p(x_4)$</td>
<td>$I_U$</td>
<td>$I_L$</td>
<td>$\epsilon$</td>
</tr>
<tr>
<td>----</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>1</td>
<td>0.1000</td>
<td>0.6000</td>
<td>0.2000</td>
<td>0.1000</td>
<td>0.1953</td>
<td>0.0847</td>
<td>0.1106</td>
</tr>
<tr>
<td>2</td>
<td>0.1073</td>
<td>0.5663</td>
<td>0.2155</td>
<td>0.1126</td>
<td>0.1834</td>
<td>0.0885</td>
<td>0.0949</td>
</tr>
<tr>
<td>3</td>
<td>0.1141</td>
<td>0.5348</td>
<td>0.2287</td>
<td>0.1249</td>
<td>0.1735</td>
<td>0.0916</td>
<td>0.0819</td>
</tr>
<tr>
<td>4</td>
<td>0.1204</td>
<td>0.5061</td>
<td>0.2394</td>
<td>0.1369</td>
<td>0.1650</td>
<td>0.0942</td>
<td>0.0708</td>
</tr>
<tr>
<td>5</td>
<td>0.1264</td>
<td>0.4802</td>
<td>0.2480</td>
<td>0.1484</td>
<td>0.1576</td>
<td>0.0963</td>
<td>0.0613</td>
</tr>
<tr>
<td>10</td>
<td>0.1524</td>
<td>0.3867</td>
<td>0.2668</td>
<td>0.1963</td>
<td>0.1326</td>
<td>0.1024</td>
<td>0.0302</td>
</tr>
<tr>
<td>20</td>
<td>0.1912</td>
<td>0.3037</td>
<td>0.2604</td>
<td>0.2448</td>
<td>0.1219</td>
<td>0.1057</td>
<td>0.0162</td>
</tr>
<tr>
<td>40</td>
<td>0.2320</td>
<td>0.2622</td>
<td>0.2476</td>
<td>0.2578</td>
<td>0.1110</td>
<td>0.1064</td>
<td>0.0046</td>
</tr>
<tr>
<td>80</td>
<td>0.2482</td>
<td>0.2515</td>
<td>0.2486</td>
<td>0.2516</td>
<td>0.1068</td>
<td>0.1064</td>
<td>0.0004</td>
</tr>
<tr>
<td>202</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.1064</td>
<td>0.1064</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Non-Symmetric Channel Example

\[
P_2 = \begin{pmatrix} 
0.1000 & 0.2500 & 0.2000 & 0.1000 \\
0.1000 & 0.2500 & 0.6000 & 0.2000 \\
0.7000 & 0.2500 & 0.1000 & 0.2000 \\
0.1000 & 0.2500 & 0.1000 & 0.5000 
\end{pmatrix}
\]
<table>
<thead>
<tr>
<th>$n$</th>
<th>$p(x_1)$</th>
<th>$p(x_2)$</th>
<th>$p(x_3)$</th>
<th>$p(x_4)$</th>
<th>$I_U$</th>
<th>$I_L$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.4498</td>
<td>0.2336</td>
<td>0.2162</td>
</tr>
<tr>
<td>2</td>
<td>0.3103</td>
<td>0.1861</td>
<td>0.2545</td>
<td>0.2228</td>
<td>0.4098</td>
<td>0.2504</td>
<td>0.1595</td>
</tr>
<tr>
<td>3</td>
<td>0.3640</td>
<td>0.1428</td>
<td>0.2653</td>
<td>0.2036</td>
<td>0.3592</td>
<td>0.2594</td>
<td>0.0998</td>
</tr>
<tr>
<td>4</td>
<td>0.4022</td>
<td>0.1118</td>
<td>0.2804</td>
<td>0.1899</td>
<td>0.3289</td>
<td>0.2647</td>
<td>0.0642</td>
</tr>
<tr>
<td>8</td>
<td>0.4522</td>
<td>0.0450</td>
<td>0.3389</td>
<td>0.1639</td>
<td>0.2988</td>
<td>0.2763</td>
<td>0.0225</td>
</tr>
<tr>
<td>16</td>
<td>0.4629</td>
<td>0.0076</td>
<td>0.3732</td>
<td>0.1565</td>
<td>0.2848</td>
<td>0.2830</td>
<td>0.0018</td>
</tr>
<tr>
<td>32</td>
<td>0.4641</td>
<td>0.0002</td>
<td>0.3769</td>
<td>0.1588</td>
<td>0.2846</td>
<td>0.2844</td>
<td>0.0003</td>
</tr>
<tr>
<td>64</td>
<td>0.4640</td>
<td>0.0000</td>
<td>0.3768</td>
<td>0.1592</td>
<td>0.2844</td>
<td>0.2844</td>
<td>0.0000</td>
</tr>
<tr>
<td>65</td>
<td>0.4640</td>
<td>0.0000</td>
<td>0.3768</td>
<td>0.1592</td>
<td>0.2844</td>
<td>0.2844</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
\[ P_1 = \begin{bmatrix} .98 & .05 \\ .02 & .95 \end{bmatrix} \quad P_2 = \begin{bmatrix} .80 & .05 \\ .20 & .95 \end{bmatrix} \]

\[ P_3 = \begin{bmatrix} .80 & .10 \\ .20 & .90 \end{bmatrix} \quad P_4 = \begin{bmatrix} .60 & .01 \\ .40 & .99 \end{bmatrix} \]

\[ P_5 = \begin{bmatrix} .80 & .30 \\ .20 & .70 \end{bmatrix} \]
<table>
<thead>
<tr>
<th></th>
<th>( C )</th>
<th>( p^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7859</td>
<td>(0.5129, 0.4871)</td>
</tr>
<tr>
<td>2</td>
<td>0.4813</td>
<td>(0.4676, 0.5324)</td>
</tr>
<tr>
<td>3</td>
<td>0.3976</td>
<td>(0.4824, 0.5176)</td>
</tr>
<tr>
<td>4</td>
<td>0.3688</td>
<td>(0.4238, 0.5762)</td>
</tr>
<tr>
<td>5</td>
<td>0.1912</td>
<td>(0.5100, 0.4900)</td>
</tr>
<tr>
<td>$P_i$</td>
<td>$C$</td>
<td>$p^*$</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>$P_1$</td>
<td>.7859</td>
<td>(.5129 .4871)</td>
</tr>
<tr>
<td>$P_2$</td>
<td>.4813</td>
<td>(.4676 .5324)</td>
</tr>
<tr>
<td>$P_3$</td>
<td>.3976</td>
<td>(.4824 .5176)</td>
</tr>
<tr>
<td>$P_4$</td>
<td>.3688</td>
<td>(.4238 .5762)</td>
</tr>
<tr>
<td>$P_5$</td>
<td>.1912</td>
<td>(.5100 .4900)</td>
</tr>
</tbody>
</table>
Communication over Noisy Channels

- **W**: original message
- **X**: original codeword
- **Y**: corrupted codeword
- **\( \tilde{\mathbf{W}} \)**: decoded message

Diagram:
- Encoder: \( \mathbf{x} \in \mathcal{C} \)
- Noisy channel: \( P = p(y|x) \)
- Decoder: \( y = x \)
Binary Symmetric Channel

channel matrix

$$\begin{bmatrix}
p & \bar{p} \\
p & p \\
\bar{p} & p \\
p & \bar{p}
\end{bmatrix}$$
Binary Symmetric Channel

• Consider a block of $N = 1000$ bits
  – if $p = 0$, 1000 bits are received correctly
  – if $p = 0.01$, 990 bits are received correctly
  – if $p = 0.5$, 500 bits are received correctly

• When $p > 0$, we do not know which bits are in error
  – if $p = 0.01$, $C = .919$ bit
  – if $p = 0.5$, $C = 0$ bit
Triple Repetition Code

- $N = 3$

<table>
<thead>
<tr>
<th>message $w$</th>
<th>codeword $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>000</td>
</tr>
<tr>
<td>1</td>
<td>111</td>
</tr>
</tbody>
</table>
Binary Symmetric Channel Errors

- If $N$ bits are transmitted, the probability of an $m$ bit error pattern is
  \[ p^m (1-p)^{N-m} \]
- The probability of exactly $m$ errors is
  \[ p^m (1-p)^{N-m} \binom{N}{m} \]
- The probability of $m$ or more errors is
  \[ \sum_{i=m}^{N} p^i (1-p)^{N-i} \binom{N}{i} \]
Triple Repetition Code

• $N = 3$
• The probability of 0 errors is $(1 - p)^3$
• The probability of 1 error is $3p(1 - p)^2$
• The probability of 2 errors is $3p^2(1 - p)$
• The probability of 3 errors is $p^3$
## Triple Repetition Code – Decoding

<table>
<thead>
<tr>
<th>Received Word</th>
<th>Codeword</th>
<th>Error Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0</td>
<td>0 0 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>0 0 1</td>
<td>0 0 0</td>
<td>0 0 1</td>
</tr>
<tr>
<td>0 1 0</td>
<td>0 0 0</td>
<td>0 1 0</td>
</tr>
<tr>
<td>1 0 0</td>
<td>0 0 0</td>
<td>1 0 0</td>
</tr>
<tr>
<td>1 1 1</td>
<td>1 1 1</td>
<td>0 0 0</td>
</tr>
<tr>
<td>1 1 0</td>
<td>1 1 1</td>
<td>0 0 1</td>
</tr>
<tr>
<td>1 0 1</td>
<td>1 1 1</td>
<td>0 1 0</td>
</tr>
<tr>
<td>0 1 1</td>
<td>1 1 1</td>
<td>1 0 0</td>
</tr>
</tbody>
</table>
Triple Repetition Code

• The BSC bit error probability is \( p < \frac{1}{2} \)

• majority vote or nearest neighbor decoding
  
  000, 001, 010, 100 → 000
  111, 110, 101, 011 → 111

• the probability of a decoding error is
  
  \[ P_e = 3p^2(1-p) + p^3 = 3p^2 - 2p^3 < p \]

• If \( p = 0.01 \), then \( P_e = 0.000298 \) and only one word in 3356 will be in error after decoding.

• A reduction by a factor of 33.
Code Rate

• After compression, the data is (almost) memoryless and uniformly distributed (equiprobable)

• Thus the entropy of the messages (codewords) is

\[ H(W) = \log_b M \]

• The blocklength of a codeword is \( N \)
Code Rate

• The code rate is given by

\[ R = \frac{\log_b M}{N} \] bits per channel use

• \( M \) is the number of codewords
• \( N \) is the block length
• For the triple repetition code

\[ R = \frac{\log_2 2}{3} = \frac{1}{3} \]
Repetition

Probability of message error

Code rate R
Shannon’s Noisy Coding Theorem

For any $\varepsilon > 0$ and for any rate $R$ less than the channel capacity $C$, there is an encoding and decoding scheme that can be used to ensure that the probability of decoding error is less than $\varepsilon$ for a sufficiently large block length $N$. 
Error Correction Coding $N = 3$

- $R = 1/3 \ M = 2$
  - $0 \rightarrow 000$
  - $1 \rightarrow 111$

- $R = 1 \ M = 8$
  - $000 \rightarrow 000 \ 001 \rightarrow 001 \ 010 \rightarrow 010 \ 011 \rightarrow 011$
  - $111 \rightarrow 111 \ 110 \rightarrow 110 \ 101 \rightarrow 101 \ 100 \rightarrow 100$

- Another choice $R = 2/3 \ M = 4$
  - $00 \rightarrow 000 \ 01 \rightarrow 011$
  - $10 \rightarrow 101 \ 11 \rightarrow 110$
Error Correction Coding $N = 3$

- BSC $p = 0.01$
- $M$ is the number of codewords

<table>
<thead>
<tr>
<th>Code Rate R</th>
<th>$P_e$</th>
<th>$M = 2^{NR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0297</td>
<td>8</td>
</tr>
<tr>
<td>2/3</td>
<td>0.0199</td>
<td>4</td>
</tr>
<tr>
<td>1/3</td>
<td>$2.98 \times 10^{-4}$</td>
<td>2</td>
</tr>
</tbody>
</table>

- Tradeoff between code rate and error rate
Codes for $N=3$

$R=1$

$R=2/3$

$R=1/3$
Error Correction Coding $N = 5$

- BSC $p = 0.01$

<table>
<thead>
<tr>
<th>Code Rate R</th>
<th>$P_e$</th>
<th>$M=2^{NR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0490</td>
<td>32</td>
</tr>
<tr>
<td>4/5</td>
<td>0.0394</td>
<td>16</td>
</tr>
<tr>
<td>3/5</td>
<td>0.0297</td>
<td>8</td>
</tr>
<tr>
<td>2/5</td>
<td>$9.80 \times 10^{-4}$</td>
<td>4</td>
</tr>
<tr>
<td>1/5</td>
<td>$9.85 \times 10^{-6}$</td>
<td>2</td>
</tr>
</tbody>
</table>

- Tradeoff between code rate and error rate
Error Correction Coding $N = 7$

- BSC $p = 0.01$ $N = 7$

<table>
<thead>
<tr>
<th>Code Rate R</th>
<th>$P_e$</th>
<th>$M=2^{NR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0679</td>
<td>128</td>
</tr>
<tr>
<td>6/7</td>
<td>0.0585</td>
<td>64</td>
</tr>
<tr>
<td>5/7</td>
<td>0.0490</td>
<td>32</td>
</tr>
<tr>
<td>4/7</td>
<td>2.03×10^{-3}</td>
<td>16</td>
</tr>
<tr>
<td>3/7</td>
<td>1.46×10^{-3}</td>
<td>8</td>
</tr>
<tr>
<td>2/7</td>
<td>9.80×10^{-4}</td>
<td>4</td>
</tr>
<tr>
<td>1/7</td>
<td>3.40×10^{-7}</td>
<td>2</td>
</tr>
</tbody>
</table>

- Tradeoff between code rate and error rate
Best Code Comparison

- \( p = .01 \)
- \( N = 3 \quad R = 2/3 \quad P_e = 2.0 \times 10^{-2} \)
- \( N = 12 \quad R = 2/3 \quad P_e = 6.2 \times 10^{-3} \)
- \( N = 30 \quad R = 2/3 \quad P_e = 3.3 \times 10^{-3} \)

- The code rates are the same but the error rate is decreasing
- Thus for fixed \( R \), \( P_e \) can be decreased by increasing \( N \)
\[ C = \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_m \\ \vdots \\ \mathbf{c}_M \end{bmatrix} = \begin{bmatrix} c_{1,1} & \cdots & c_{1,n} & \cdots & c_{1,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{m,1} & \cdots & c_{m,n} & \cdots & c_{m,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{M,1} & \cdots & c_{M,n} & \cdots & c_{M,N} \end{bmatrix} \]
Binary Codes

• For given values of $M$ and $N$, there are $2^{MN}$ possible codes.
• Of these, some will be bad, some will be best (optimal), and some will be good, in terms of $P_e$
• An average code will be good.
**Theorem (Shannon’s channel coding theorem):**

Let $C$ be the information transfer capacity of a memoryless channel defined by its transition probabilities matrix $\mathbf{P} = \{p(y|x)\}$. If the code rate $R < C$, then there exists a channel code $\mathcal{C}$ of size $M$ and blocklength $N$, such that the probability of decoding error $P_e$ is *upperbounded* by an arbitrarily small number $\epsilon$;

\[
P_e \leq \epsilon
\]

provided that the blocklength $N$ is sufficiently large (i.e., $N \geq N_0$).
Channel Capacity

• To prove that information can be transmitted reliably over a noisy channel at rates up to the capacity, Shannon used a number of new concepts
  – Allowing an arbitrarily small but nonzero probability of error
  – Using long codewords
  – Calculating the average probability of error over a random choice of codes to show that at least one good code exists
Channel Coding Theorem

• Random coding used in the proof
• Joint typicality used as the decoding rule
• Shows that good codes exist which provide an arbitrarily small probability of error
• Does not provide an explicit way of constructing good codes
• If a long code (large $N$) is generated randomly, the code is likely to be good but is difficult to decode
Noisy Communication System

\[ W \xrightarrow{ \text{channel encoder} } X \xrightarrow{ \text{noisy channel} } Y \xrightarrow{ \text{channel decoder} } \tilde{W} \]

- **Original message**: \( W \)
- **Channel encoder**: \( x \in \mathcal{C} \)
- **Noisy channel**: \( P = p(y|x) \)
- **Corrupted codeword**: \( y \)
- **Channel decoder**: \( y = x \)
- **Decoded message**: \( \tilde{W} \)
The Data Processing Inequality
Cascaded Channels

The mutual information $I(W;Y)$ for the cascade cannot be larger than $I(W;X)$ or $I(X;Y)$, so that

$$I(W;Y) \leq I(W;X) \quad I(W;Y) \leq I(X;Y)$$
**Theorem** *(Converse of the channel coding theorem)*:

Let a memoryless channel with capacity $C$ be used to transmit codewords of blocklength $N$ and input information $R$. Then the error decoding probability $P_e$ satisfies the following inequality:

$$P_e(N, R) \geq 1 - \frac{C}{R} - \frac{1}{NR}$$

If the rate $R > C$, then the error decoding probability $P_e$ is bounded away from zero.
Channel Capacity: Weak Converse

\[ P_e(N, R) \geq 1 - \frac{C}{R} - \frac{1}{NR} \]

For \( R > C \), the decoding error probability is bounded away from 0.
Channel Capacity: Weak Converse

- $C = 0.3$
Fano’s Inequality

\[ h(P_e) + P_e \log_b (r-1) \]

\( \log_b r \)

\( \log_b (r-1) \)

\( (r-1)/r \)

\( 1 \)
Fano’s Inequality

\[ H(X|Y) \leq h(P_e) + P_e \log_b (r-1) \]

\[ H(X|Y) \leq 1 + NRP_e(N,R) \]
Channel Capacity: Strong Converse

• For rates above capacity (R > C)

\[ P_e(N,R) \geq 1 - 2^{-N E_A(R)} \]

• where \( E_A(R) \) is Arimoto’s error exponent
Arimoto’s Error Exponent $E_A(R)$
$E_A(R)$ for a BSC with $p=0.1$
• The capacity is a very clear dividing point
• At rates below capacity, $P_e \to 0$ exponentially as $N \to \infty$
• At rates above capacity, $P_e \to 1$ exponentially as $N \to \infty$