Seperation of source and channel coding

The source coding depiction can be considered as in fig 1,

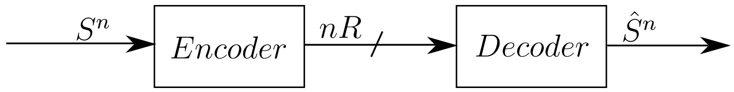

Fig. 1. Source Coding

In the source coding problem we showed that if $H(S) \leq R$ then R is an achievable rate, meaning there exists a sequence of codes such that $Pr(S^n \neq \hat{S^n}) \overset{n\to\infty}{\longrightarrow} 0$. Since R is the compression ration, measured in bits per symbol, we would like it to be as small as possible. R cannot be smaller than $H(S)$ (See lecture 5).
For a channel coding depiction consider the following figure,

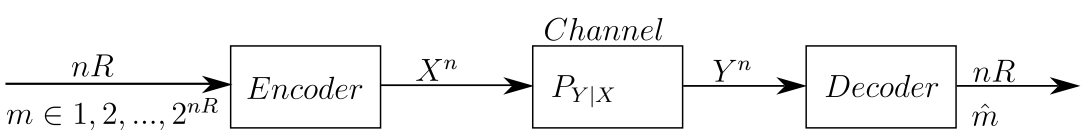

Fig. 2. Channel Coding

Here, R represents the number of bits each symbol represents, and therefore we would like it to be as big as possible. For the channel, a rate R is achievable if there exists a sequence of codes $(2^{nR}, n)$ such that $Pr(M \neq \hat{M}) \overset{n\to\infty}{\longrightarrow} 0$.
Now let us consider the problem of a joint source-channel problem as seen in fig 3. What would be the optimal source code and channel code given a specific channel? In other words, what would be the sequence of codes which achieves an arbitrarily small probability of error, using the least amount of symbols?

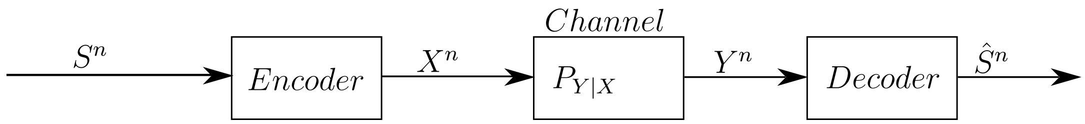

Fig. 3. Joint source-channel Coding

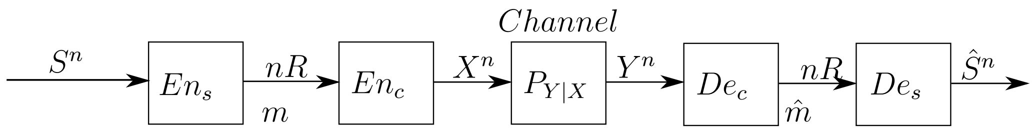

Fig. 4. Source-channel seperation scheme

A possible solution is a source-channel seperation scheme as follows,

In this scheme, the channel coding does not take into account the source coding.

The solution scheme must maintain both the channel- and source-coding constraints, i.e

- $Pr(M \neq \hat{M}) \longrightarrow 0$ for $R < C$
- $Pr(S^n \neq \hat{S^n}) \longrightarrow 0$ for $H(S) < R$

where R is the bit rate and

$$C = \max_{P_X} I(X;Y) \tag{1}$$

is the Capacity of the channel.

Therefore any valid source-channel coding scheme must maintain,

$$H(S) < C \Rightarrow Pr(S^n \neq \hat{S^n}) \longrightarrow 0 \tag{2}$$

Notice that the probability of error for each individual message goes to zero since the maximum probablity of error over all messages goes to zero.

Theorem 1 (A source-channel seperation is optimal) If $H(S) > C$ then one cannot transmit S^n lossless through the channel, and if $H(S) \leq C$ one can use the source-channel seperation scheme.

Proof:

For our purpose, let us consider only discrete memoryless channels (A more general proof exists).

For a memoryless channel,

$$P(Y_i|Y^{i-1}, X^i) = P(Y_i|X^i) \tag{3}$$

We will prove that if $Pr(S^n \neq \hat{S^n}) \overset{n\to\infty}{\longrightarrow} 0$ then $H(S) < C$,

$$\begin{aligned} nH(S) &\overset{(a)}{=} H(S^n) \\ &= H(S^n) - H(S^n|Y^n) + H(S^n|Y^n) \\ &= I(S^n; Y^n) + H(S^n|Y^n) \end{aligned} \tag{4}$$

Where

(a) follows from $S^n \sim P_S$ i.i.d

Assume a code that achieves $Pr(S^n \neq \hat{S^n}) = P_\epsilon \longrightarrow 0$ exists.

Using Fano's inequality:

$$nH(S) \leq I(S^n; Y^n) + (1 + nP_\epsilon \log |\mathcal{S}|) \tag{5}$$

$$H(S) \leq \frac{1}{n} I(S^n; Y^n) + \epsilon_n \tag{6}$$

where

$$\epsilon_n = \frac{1}{n} + P_\epsilon \log |\mathcal{S}| \overset{P_\epsilon \to 0}{\longrightarrow} 0 \tag{7}$$

Now, since $S^n - X^n - Y^n$ form a markov chain, by the Data Processing inequality we get

$$\begin{aligned} I(S^n; Y^n) &\leq I(X^n; Y^n) \\ &= H(Y^n) - H(Y^n|X^n) \\ &\overset{(a)}{=} \sum_{i=1}^{n} [H(Y_i|Y^{i-1}) - H(Y_i|X_i)] \\ &\overset{(b)}{\leq} \sum_{i=1}^{n} [H(Y_i) - H(Y_i|X_i)] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} I(X_i;Y_i) \\
&\stackrel{(c)}{\leq} nC \qquad (8)
\end{aligned}
$$

Where

(a) follows from the definition of a memoryless channel,

(b) conditioning reduces entropy

(c) follows from the definition of Capacity (1) and X_i, Y_i being i.i.d

So if there exists a code for which the probability for an error goes to zero then

$$
\begin{aligned}
H(S) &\leq \frac{1}{n} I(S^n;Y^n) + \epsilon_n \\
&\leq C + \epsilon_n \qquad (9)
\end{aligned}
$$

With ϵ_n arbitrarily small ■