Simultaneous Transmission of Data and State with Common Knowledge

Yossef Steinberg

Technion—Israel Institute of Technology

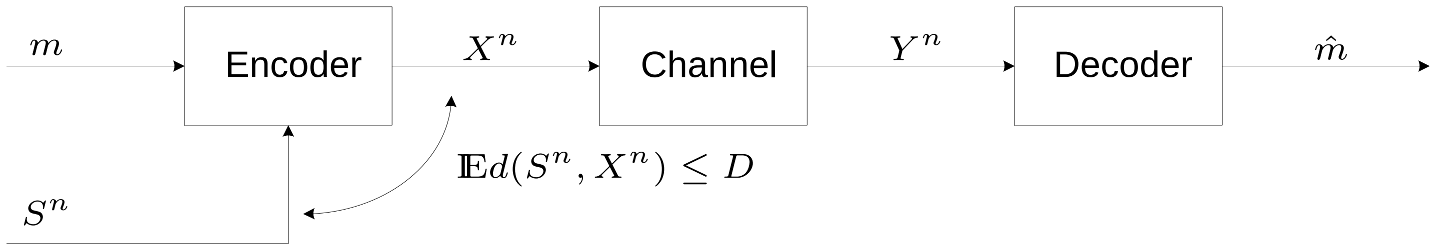
ysteinbe@ee.technion.ac.il

Introduction

Outline

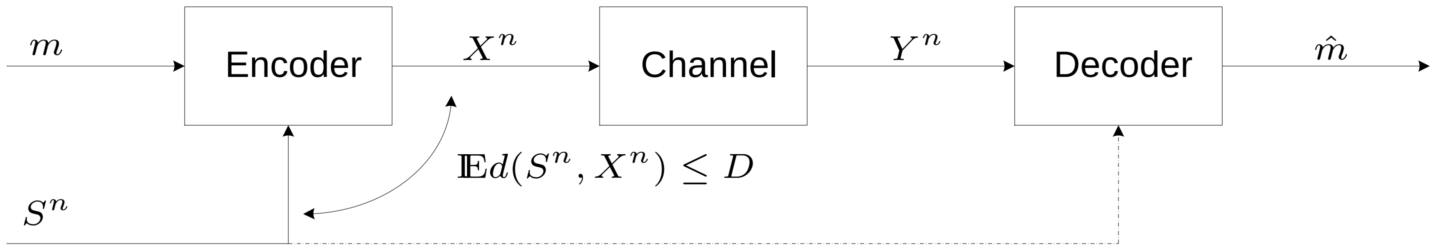
- The general Information Embedding problem
- Reversible Information Embedding, with and without distortion
- The common knowledge constraint
- Main result
- Example

The Information Embedding (IE) Problem



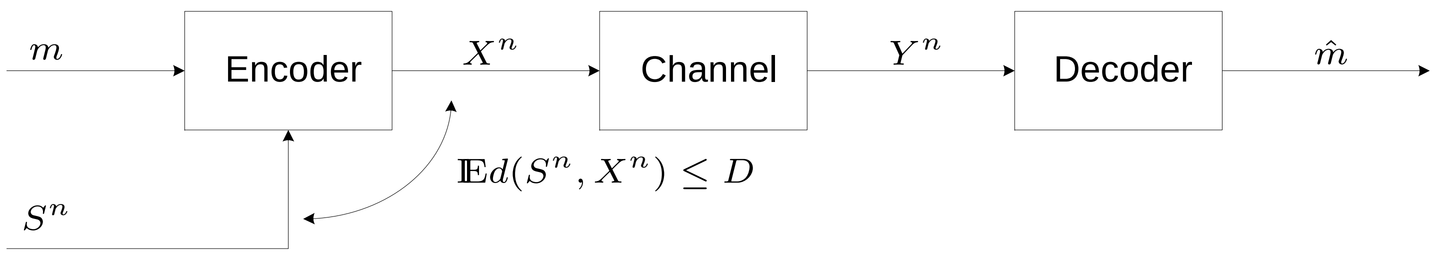
- A message m is embedded into host signal S^n, producing data set X^n
- X^n is transmitted via $P_{Y|X}$ (attack channel) to its destination
- At the destination, a noisy version Y^n of the data set is received, from which m is decoded.
- In IE, m is embedded into S^n in a manner that is transparent to the unintended observer $\Rightarrow$ a distortion constraint between S^n and X^n
- Public IE – The host S^n is available only at the encoder

The Information Embedding (IE) Problem



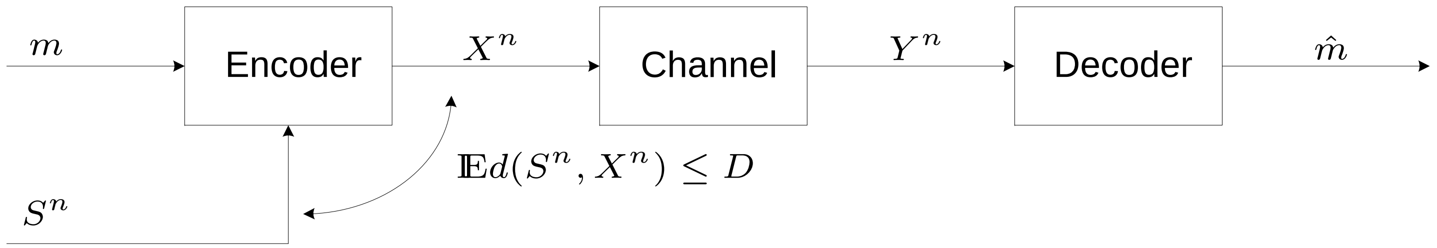
- A message m is embedded into host signal S^n, producing data set X^n
- X^n is transmitted via $P_{Y|X}$ (attack channel) to its destination
- At the destination, a noisy version Y^n of the data set is received, from which m is decoded.
- In IE, m is embedded into S^n in a manner that is transparent to the unintended observer $\Rightarrow$ a distortion constraint between S^n and X^n
- Public IE – The host S^n is available only at the encoder
- Private IE – The host S^n is available at both, encoder and decoder

The IE Problem (cont'd)



- The distortion constraint is imposed in order to hide the fact that communication (beyond that of S^n) is taking place
- Classical IE puts emphasis on embedding rate vs. input distortion D.

The IE Problem (cont'd)



- The distortion constraint is imposed in order to hide the fact that communication (beyond that of S^n) is taking place
- Classical IE puts emphasis on embedding rate vs. input distortion D.

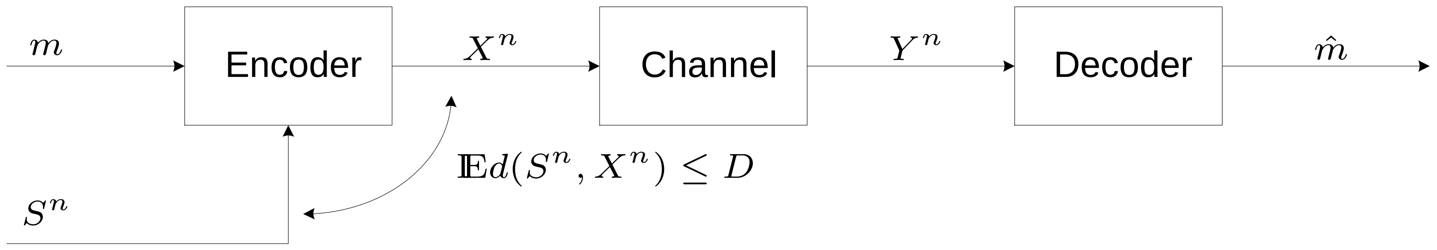
Closely related to Gel'fand & Pinsker channel [Moulin & O'Sullivan, 2003], due to the distortion constraint. Thus

$$C = \max\left[I(U;Y) - I(U;S)\right]$$

where the max is over all $P_{UX|S}$ satisfying the input distortion constraint

$$\mathbb{E}d(S,X) \leq D$$

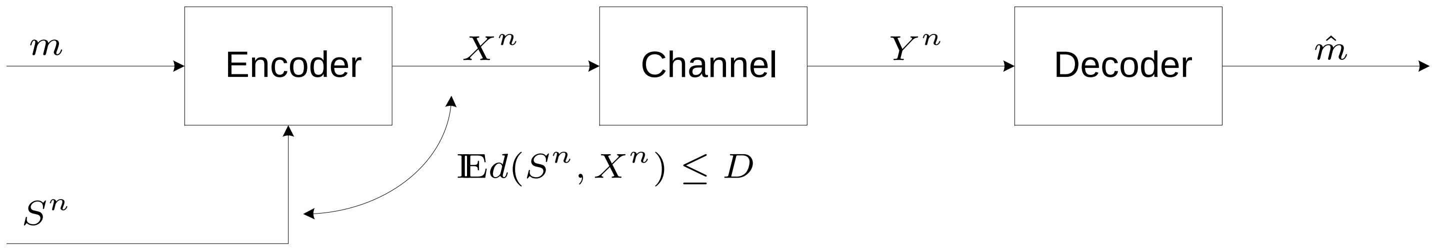
The IE Problem (cont'd)



Classical IE – embedding rate vs. input distortion:

$$C = \max_{\mathbb{E}d(S,X) \le D} [I(U;Y) - I(U;S)]$$

The IE Problem (cont'd)

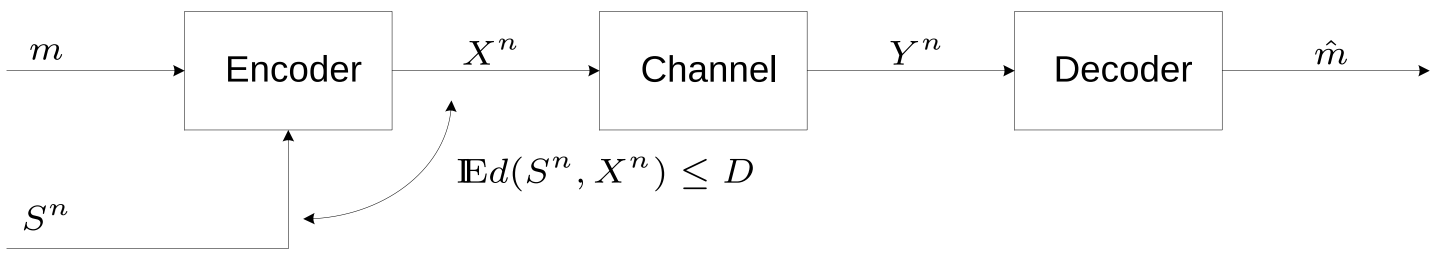


Classical IE – embedding rate vs. input distortion:

$$C = \max_{\mathbb{E}d(S,X) \leq D} [I(U;Y) - I(U;S)]$$

- The host S^n is of value at the destination (the reason for communicating from the first place)
- The destination obtains a noisy version of the data set X^n

The IE Problem (cont'd)



Classical IE – embedding rate vs. input distortion:

$$C = \max_{\mathbb{E}d(S,X)\le D} [I(U;Y) - I(U;S)]$$

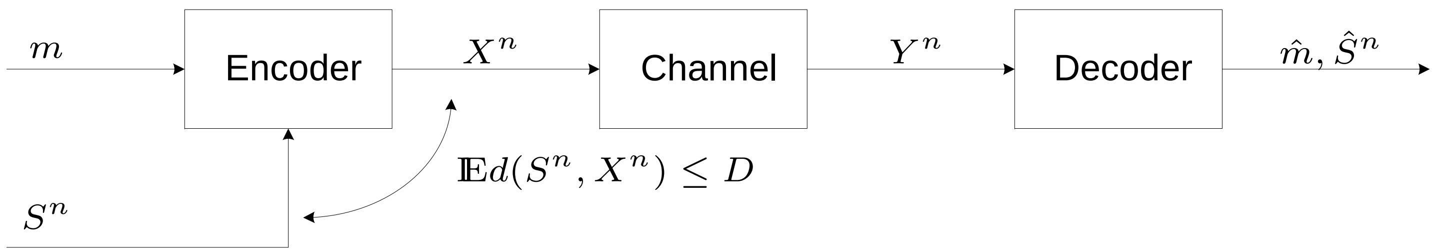
- The host S^n is of value at the destination (the reason for communicating from the first place)
- The destination obtains a noisy version of the data set X^n

Some applications cannot tolerate high distortion at the destination (e.g., medical imagery).

$\Rightarrow$ Reversible Information Embedding

Reversible IE (with and without distortion)

[Fridrich, Goljan, Du *SPIE* 2002], [Kalker & Willems, 2002]



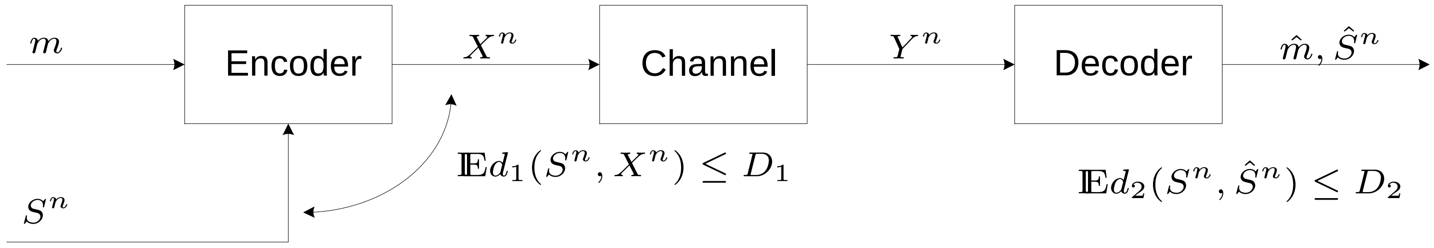
In reversible IE (RIE), an additional constraint is imposed, that S^n can be faithfully restored from Y^n. The constraint $\mathbb{E}d(S, X) \leq D$ is still relevant

$C = \max H(X) - H(S)$ (no attack channel, Kalker & Willems)

$C = \max I(X;Y) - H(S)$ (with channel, Kalker & Willems, Kotagiri & Laneman '05)

Reversible IE *(cont'd)*

With distortion (Sutivong *et. al.* 2002, 2005)



We end up with two distortion constraints. An achievable rate

$$R = \max[I(U;Y) - I(U;S)]$$

where the max is over all $P_{X,U|S}$ such that

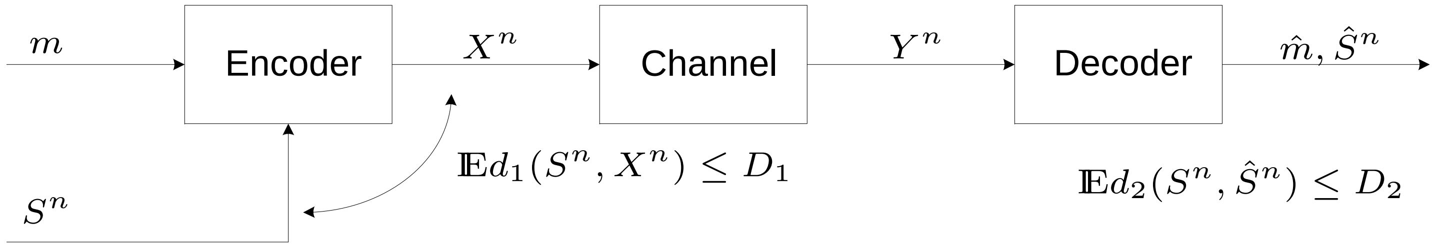
$$\mathbb{E}d_1(S,X) \leq D_1$$

$$\mathbb{E}d_2(S,\phi(U,Y)) \leq D_2$$

for some function $\phi: \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{S}}$.

Reversible IE (cont'd)

With distortion (Sutivong *et. al.* 2002, 2005)



We end up with two distortion constraints. An achievable rate

$$R = \max[I(U;Y) - I(U;S)] \geq 0$$

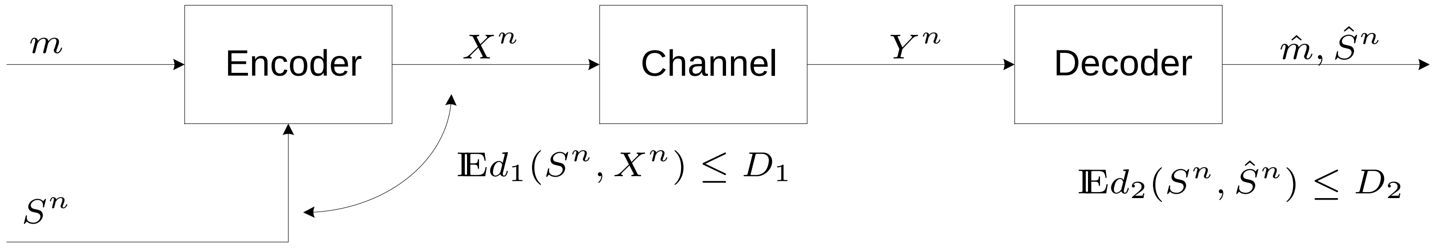
where the max is over all $P_{X,U|S}$ such that

$$\mathbb{E}d_1(S, X) \leq D_1$$
$$\mathbb{E}d_2(S, \phi(U, Y)) \leq D_2$$

for some function $\phi : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{S}}$.

Reversible IE *(cont'd)*

With distortion (Sutivong *et. al.* 2002, 2005)



We end up with two distortion constraints. An achievable rate

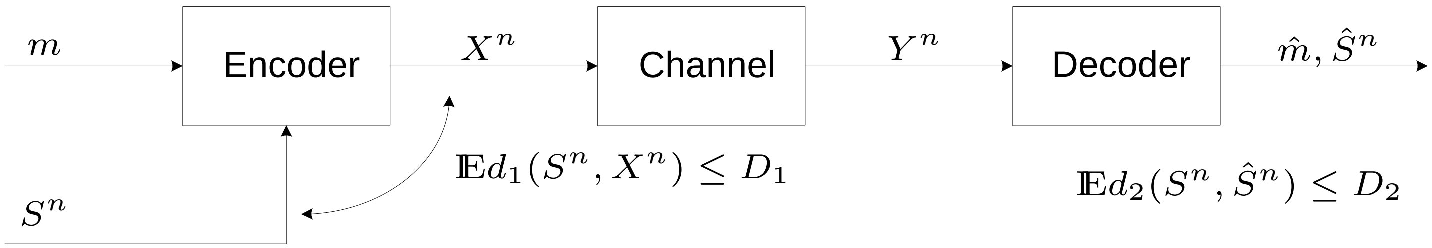
$$R = \max[I(U;Y) - I(U;S)] \ge 0$$

where the max is over all $P_{X,U|S}$ such that

$$\mathbb{E}d_1(S,X) \le D_1$$
$$\mathbb{E}d_2(S,\phi(U,Y)) \le D_2$$

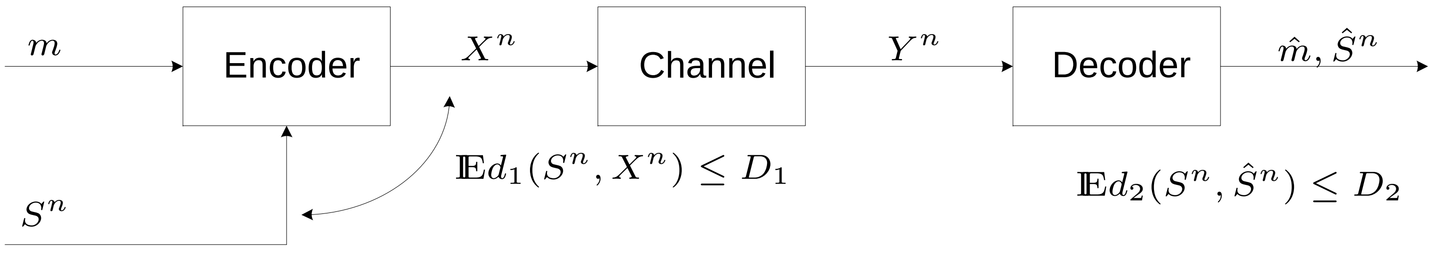
for some function $\phi : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{S}}$. Solved for the Gaussian case (Sutivong *et. al.* 2005). Contributions by Merhav & Shamai 2007, Cover, Kim, and Sutivong 2007.

Reversible IE with distortion (cont'd)



$$R = \max[I(U;Y) - I(U;S)]$$
$$\mathbb{E}d_1(S, X) \le D_1$$
$$\mathbb{E}d_2(S, \phi(U,Y)) \le D_2$$

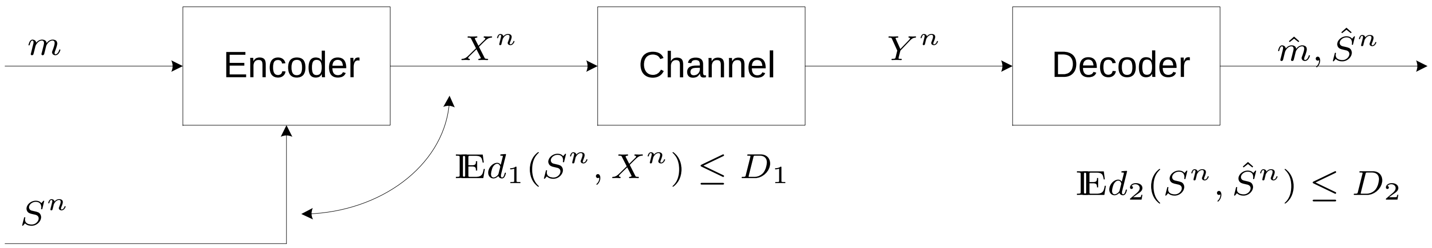
Reversible IE with distortion *(cont'd)*



$$R = \max[I(U;Y) - I(U;S)]$$
$$\mathbb{E}d_1(S, X) \le D_1$$
$$\mathbb{E}d_2(S, \phi(U, Y)) \le D_2$$

- A separation based strategy is suboptimal.
- Part of the distortion is introduced by the channel noise. Thus the estimated host $\hat{S}^n$ cannot be reproduced at the sender side

Reversible IE with distortion *(cont'd)*

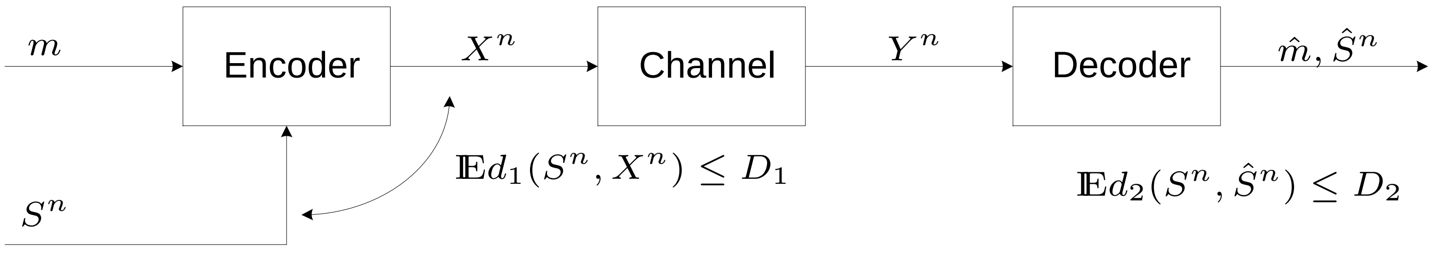


$$R = \max[I(U;Y) - I(U;S)]$$
$$\mathbb{E}d_1(S, X) \leq D_1$$
$$\mathbb{E}d_2(S, \phi(U,Y)) \leq D_2$$

- A separation based strategy is suboptimal.
- Part of the distortion is introduced by the channel noise. Thus the estimated host $\hat{S}^n$ cannot be reproduced at the sender side

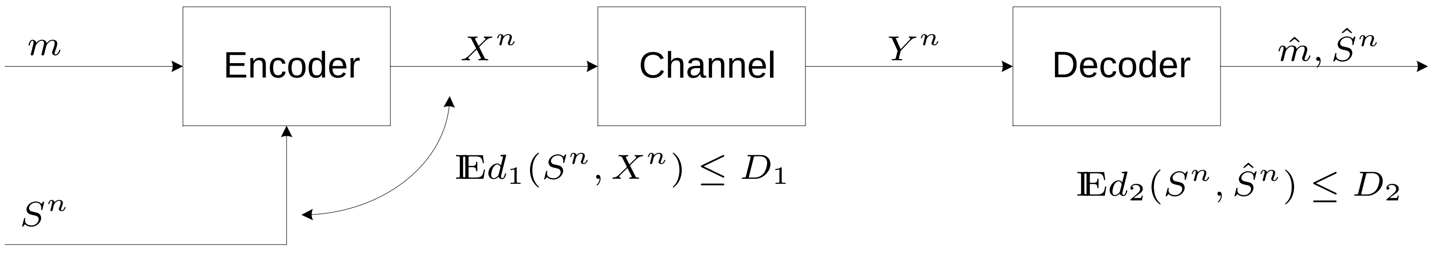
In some applications, this is a drawback.

Reversible IE with distortion *(cont'd)*



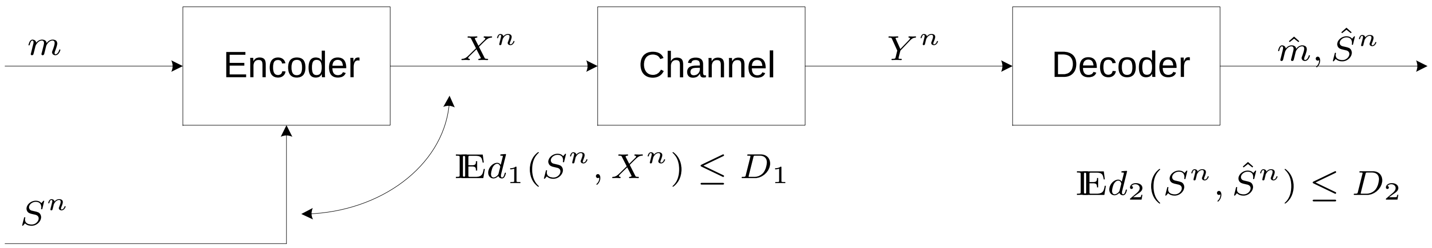
- Medical data S^n is watermarked (ID, authentication...) and sent to an expert, for consultation.

Reversible IE with distortion *(cont'd)*



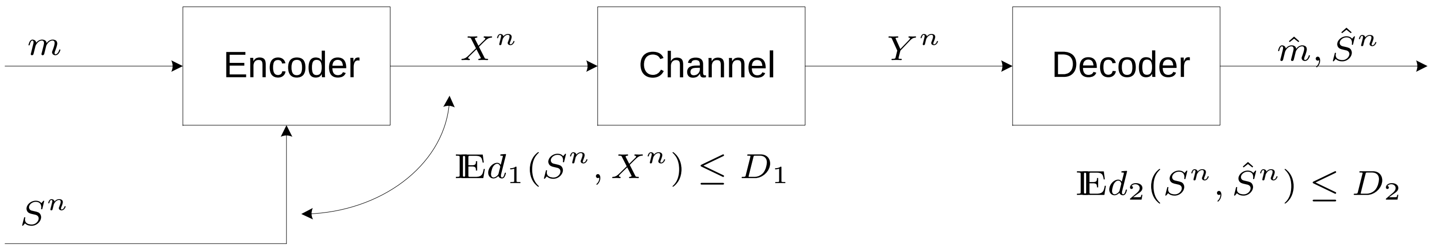
- Medical data S^n is watermarked (ID, authentication...) and sent to an expert, for consultation.
- Lossy transmission, due to limitations of the channel.

Reversible IE with distortion (cont'd)



- Medical data S^n is watermarked (ID, authentication...) and sent to an expert, for consultation.
- Lossy transmission, due to limitations of the channel.
- The coding scheme guarantees *average* distortion.
  The distortion pattern is not known at the sender side.

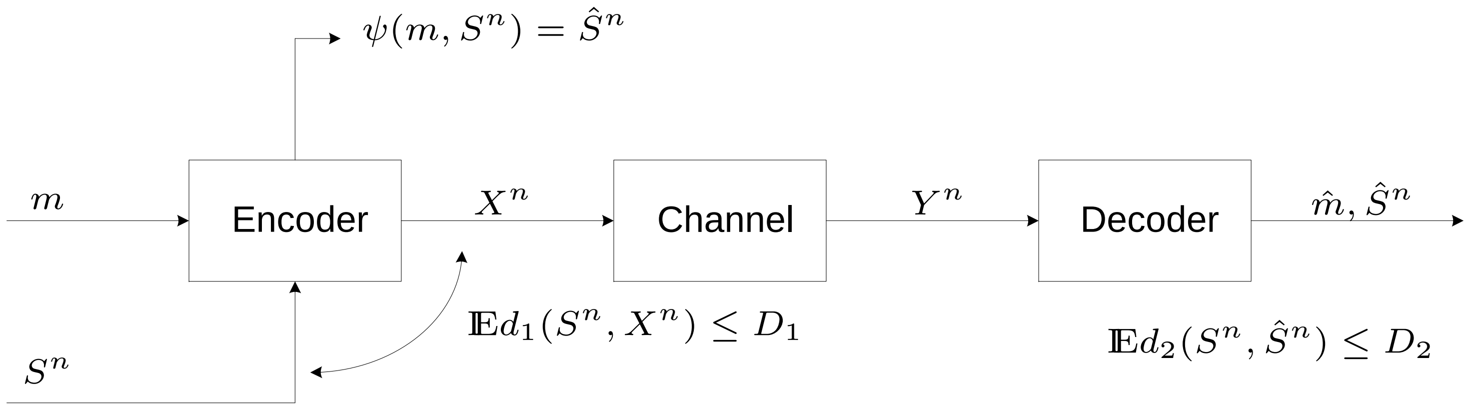
Reversible IE with distortion (cont'd)



- Medical data S^n is watermarked (ID, authentication...) and sent to an expert, for consultation.
- Lossy transmission, due to limitations of the channel.
- The coding scheme guarantees *average* distortion.
  The distortion pattern is not known at the sender side.

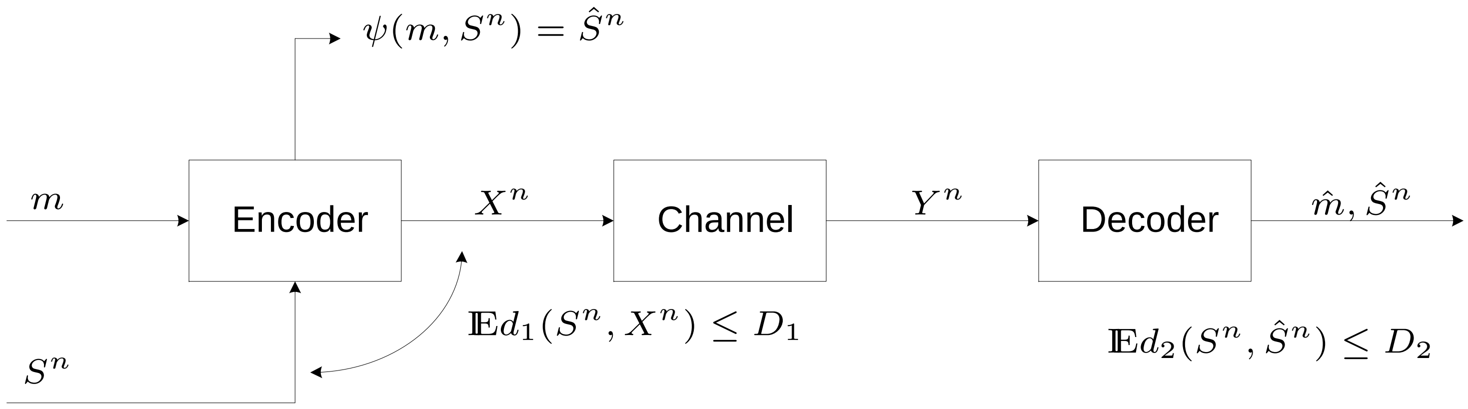
Important details can be blurred during transmission. Sender is unaware.

Reversible IE with distortion (cont'd)



Devise a coding scheme that enables the sender to produce locally $\hat{S}^n$.

Reversible IE with distortion *(cont'd)*

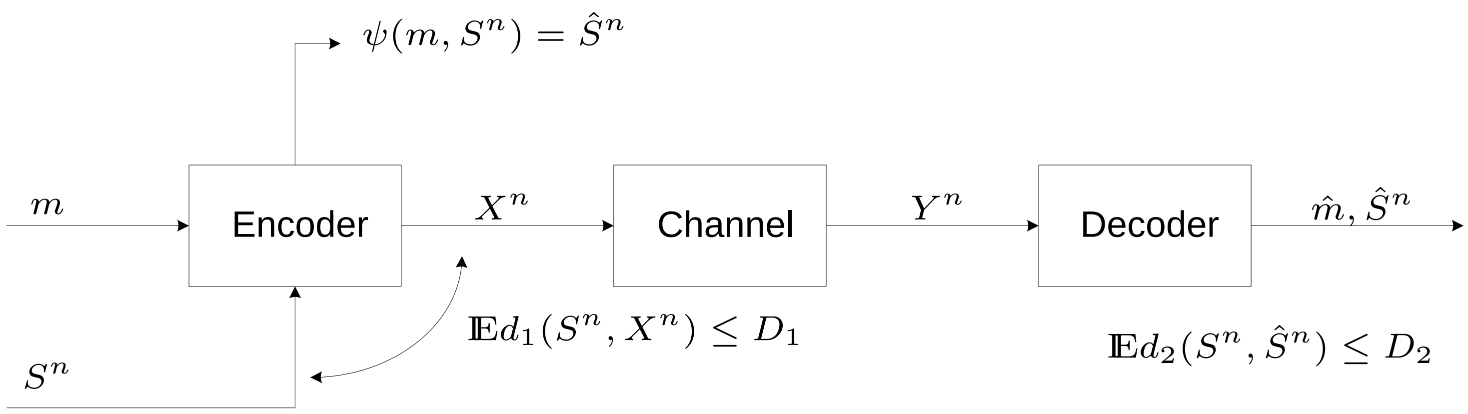


Devise a coding scheme that enables the sender to produce locally $\hat{S}^n$.

Role of the common reproduction:

- Re-transmit in case that the distortion pattern is "bad."
- Common reference for the consultation, where the sender and expert (destination) know what the data at the other side look like.

Reversible IE with distortion (cont'd)



Devise a coding scheme that enables the sender to produce locally $\hat{S}^n$.

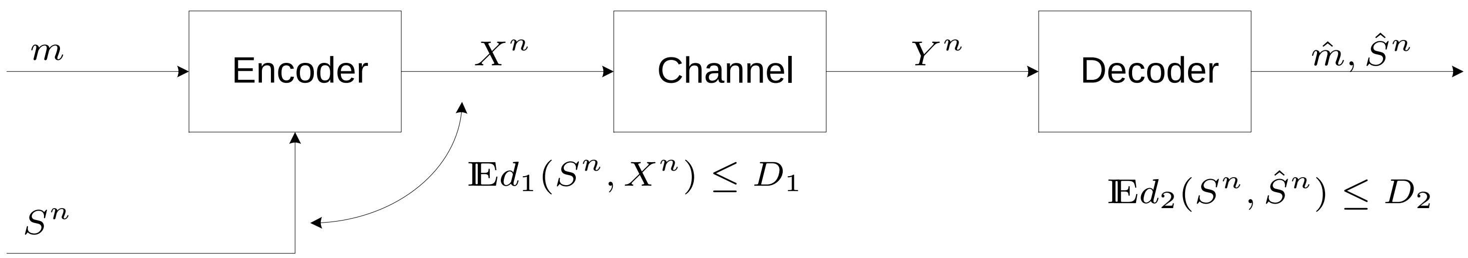
Role of the common reproduction:

- Re-transmit in case that the distortion pattern is "bad."
- Common reference for the consultation, where the sender and expert (destination) know what the data at the other side look like.

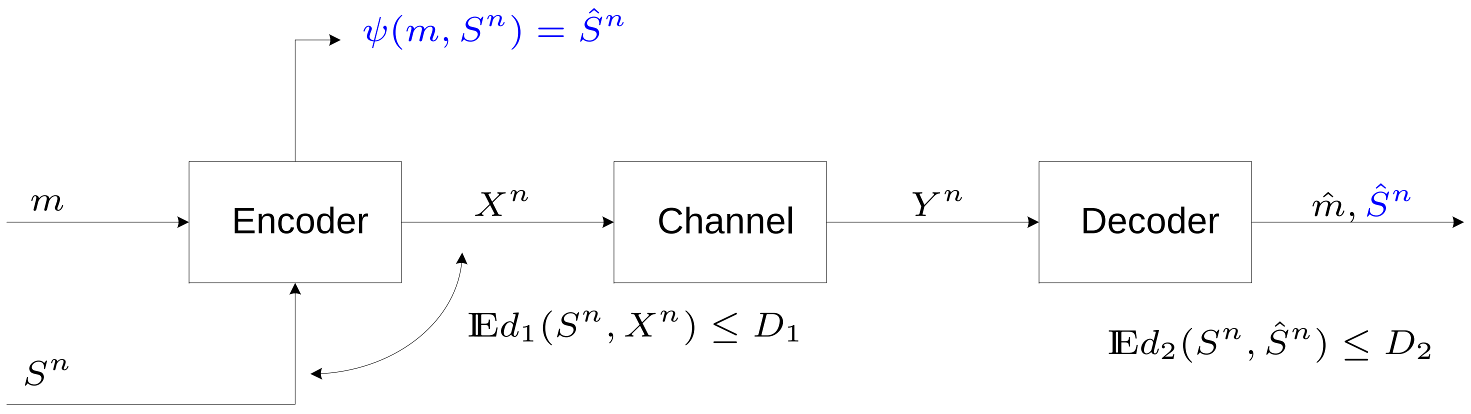
$\Longrightarrow$ *Common Knowledge (CK)* constraint

The Common Knowledge Constraint

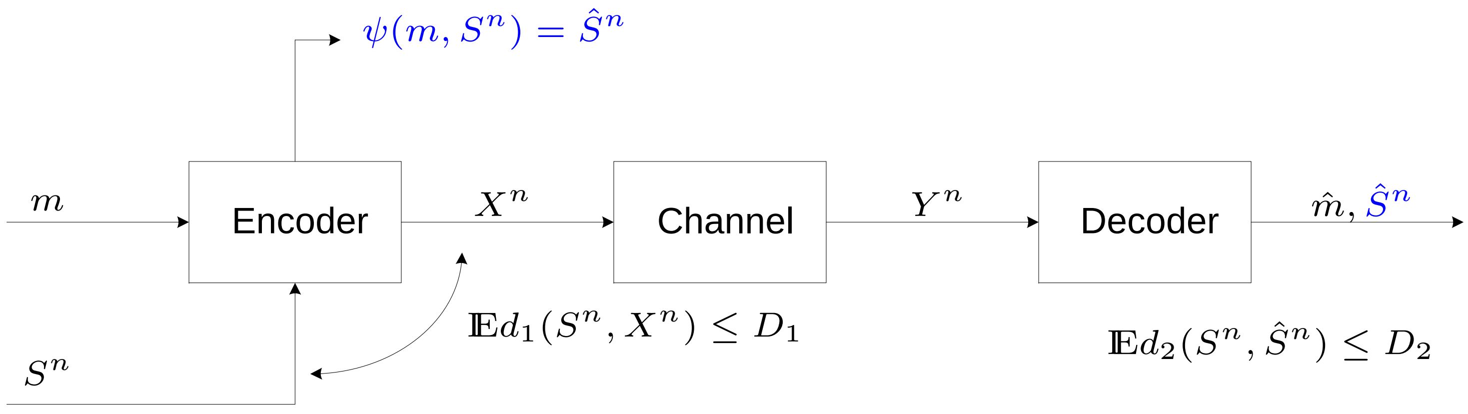
Problem Formulation



Problem Formulation



Problem Formulation



Definition: Let $\mathcal{M} = \{1, 2, \ldots, 2^{nR}\}$. An $(n, 2^{nR}, D_1, D_2, \epsilon)$ common knowledge (CK) code consists of

$$f : \mathcal{M} \times \mathcal{S}^n \to \mathcal{X}^n, \quad \text{encoder map}$$

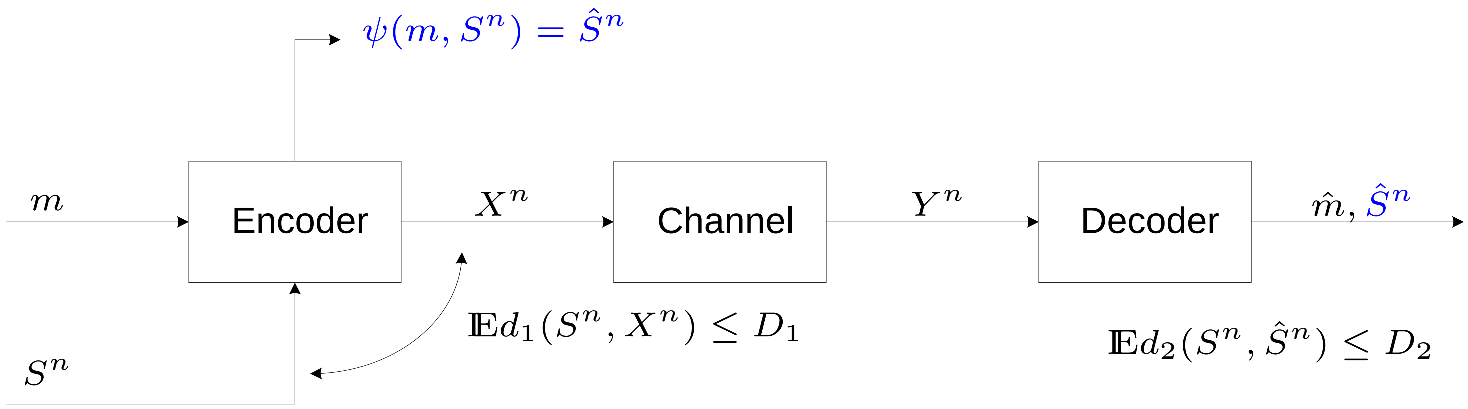
$$g : \mathcal{Y}^n \to \mathcal{M}, \quad g_s : \mathcal{Y}^n \to \hat{\mathcal{S}}^n \quad \text{decoding maps}$$

$$\psi : \mathcal{M} \times \mathcal{S}^n \to \hat{\mathcal{S}}^n, \quad \text{sender reconstruction map}$$

such that the probability of error is bounded by ϵ, the distortion constraints are satisfied, and

$$P(\psi(m, S^n) \neq \hat{S}^n) \leq \epsilon.$$

Problem Formulation



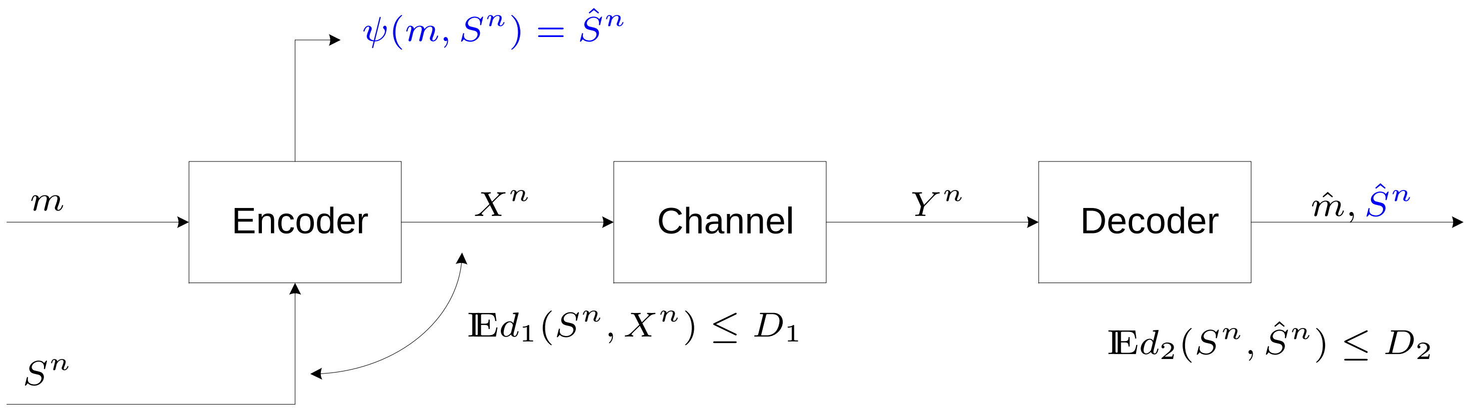
Definition: Let $\mathcal{M} = \{1, 2, \ldots, 2^{nR}\}$. An $(n, 2^{nR}, D_1, D_2, \epsilon)$ common knowledge (CK) code consists of

$f : \mathcal{M} \times \mathcal{S}^n \to \mathcal{X}^n,$	encoder map
$g : \mathcal{Y}^n \to \mathcal{M}, \quad g_s : \mathcal{Y}^n \to \hat{\mathcal{S}}^n$	decoding maps
$\psi : \mathcal{M} \times \mathcal{S}^n \to \hat{\mathcal{S}}^n,$	sender reconstruction map

such that the probability of error is bounded by ϵ, the distortion constraints are satisfied, and

$$P(\psi(m, S^n) \neq \hat{S}^n) \leq \epsilon.$$

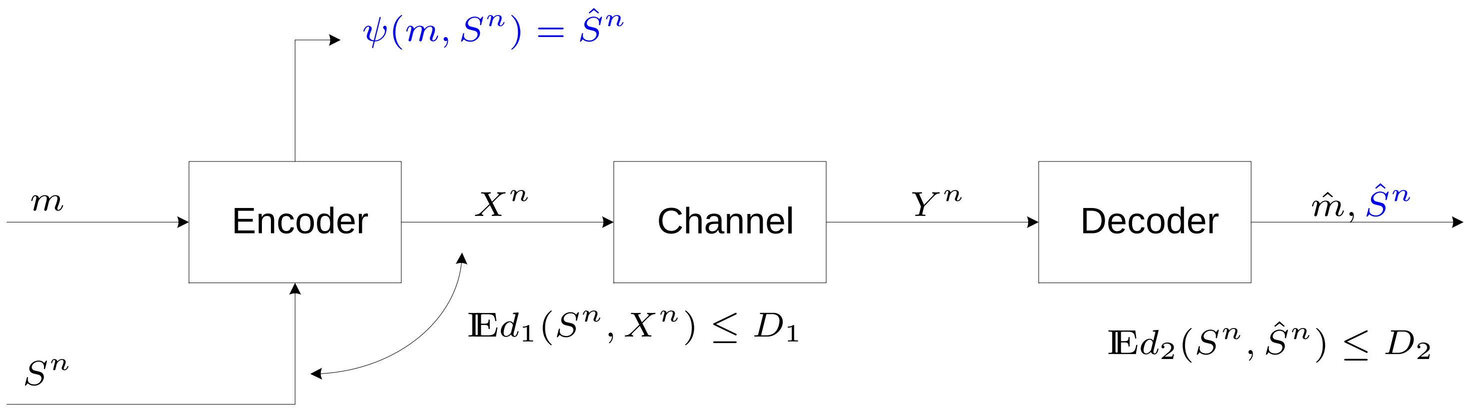
Problem Formulation *(cont'd)*



$$\hat{S}^n = g_s(Y^n)$$

$$P(\psi(m, S^n) \neq \hat{S}^n) \leq \epsilon$$

Problem Formulation (cont'd)

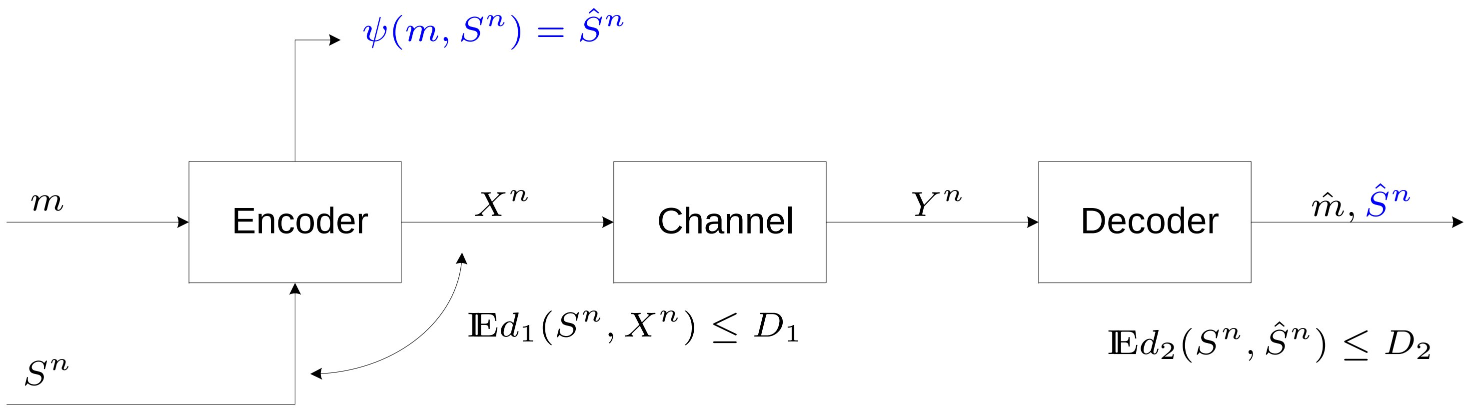


$$\hat{S}^n = g_s(Y^n)$$

$$P(\psi(m, S^n) \neq \hat{S}^n) \leq \epsilon$$

- The CK embedding capacity, $C_{ck}(D_1, D_2)$, is the maximal achievable embedding rate with input and output distortions (D_1, D_2), and arbitrarily small ϵ.

Main Result



Theorem 1

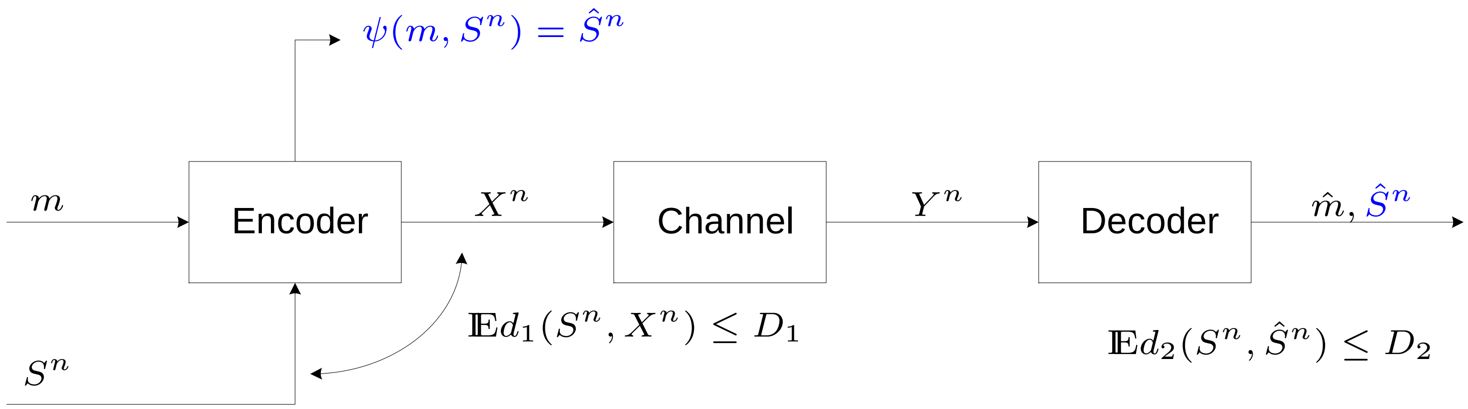
$$C_{ck}(D_1, D_2) = \max[I(W;Y) - I(W;S)] \geq 0$$

where the maximum is over all (W, X) $W \circ\!\!- X \circ\!\!- Y$, *and* $\varphi : \mathcal{W} \to \hat{\mathcal{S}}$ *such that*

$$\mathbb{E}d_1(S, X) \leq D_1,$$

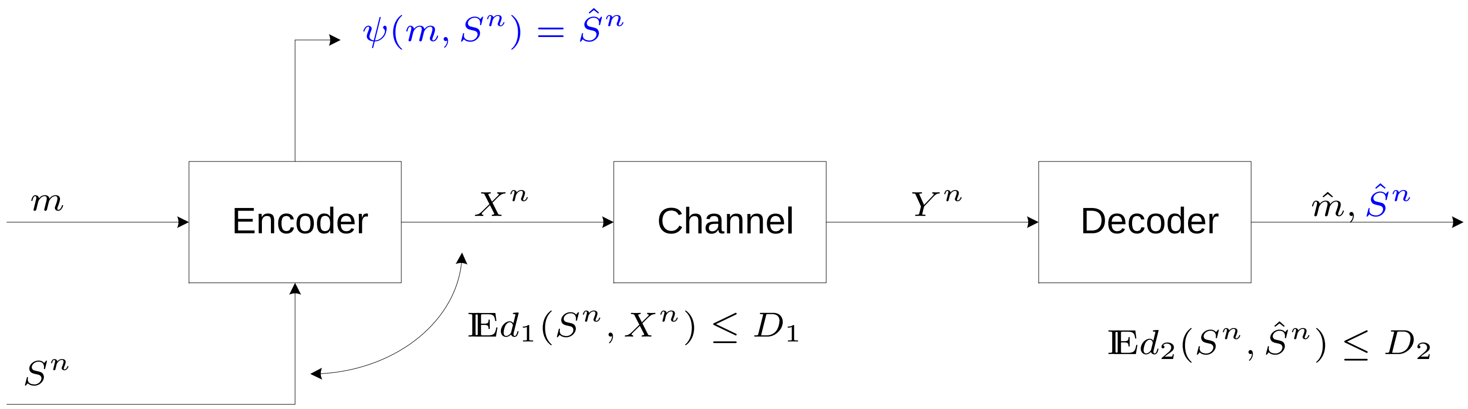
$$\mathbb{E}d_2(S, \varphi(W)) \leq D_2.$$

Main Result (cont'd)



$$
\begin{aligned}
C_{ck}(D) &= \max[I(W;Y) - I(W;S)], \qquad W \multimap X \multimap Y \\
\mathbb{E}d_1(S, X) &\leq D_1, \\
\mathbb{E}d_2(S, \varphi(W)) &\leq D_2.
\end{aligned}
$$

Main Result (cont'd)

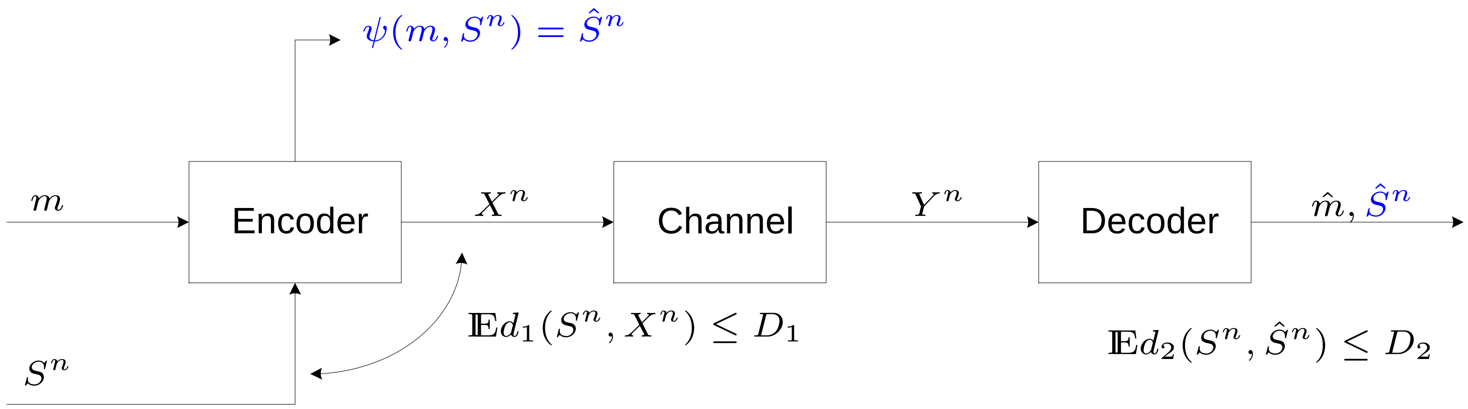


$$
\begin{aligned}
C_{ck}(D) &= \max[I(W;Y) - I(W;S)], \quad W \multimap X \multimap Y \\
\mathbb{E}d_1(S,X) &\leq D_1, \\
\mathbb{E}d_2(S,\varphi(W)) &\leq D_2.
\end{aligned}
$$

Achievable rate without the CK constraint

$$
\begin{aligned}
R &= \max[I(U;Y) - I(U;S)], \quad U \multimap X \multimap Y \\
\mathbb{E}d_1(S,X) &\leq D_1, \\
\mathbb{E}d_2(S,\phi(U,Y)) &\leq D_2.
\end{aligned}
$$

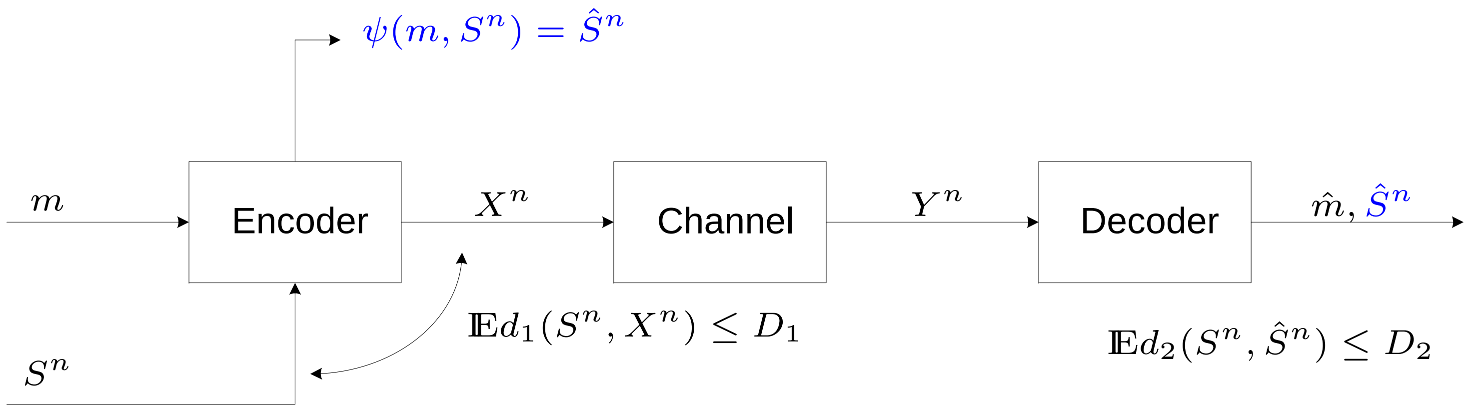
Main Result (cont'd)



$$
\begin{aligned}
C_{ck}(D) &= \max[I(W;Y) - I(W;S)], \qquad W \multimap X \multimap Y \\
\mathbb{E}d_1(S,X) &\leq D_1, \\
\mathbb{E}d_2(S,\varphi(W)) &\leq D_2.
\end{aligned}
$$

- Construct codewords W^n. Binning operation is as usual
- Y^n is used only to resolve the binning. It cannot be used in the estimation phase, to further reduce the distortion.

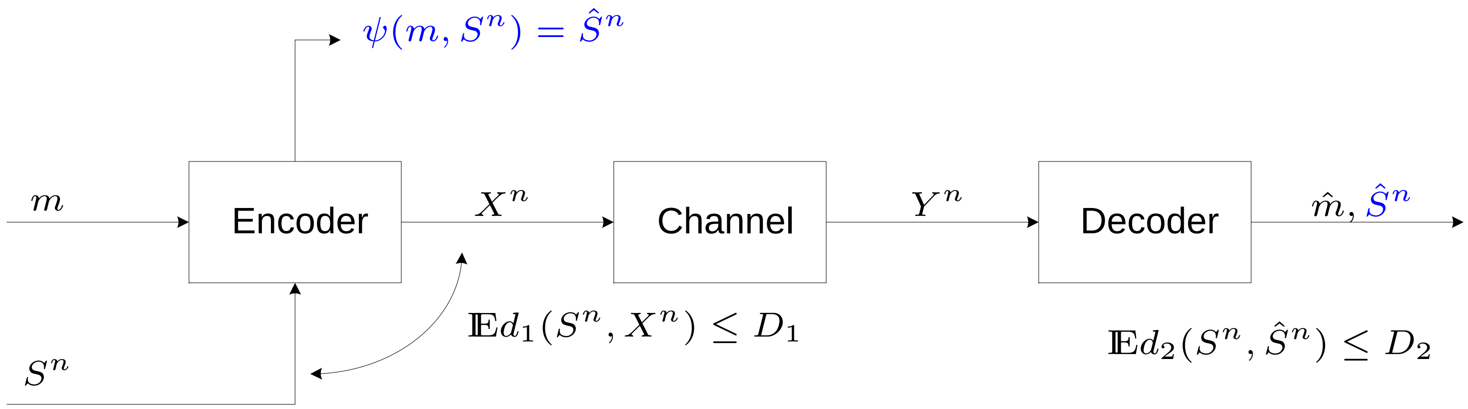
Main Result (cont'd)



$$
\begin{aligned}
C_{ck}(D) &= \max[I(W;Y) - I(W;S)], \qquad W \multimap X \multimap Y \\
\mathbb{E}d_1(S, X) &\le D_1, \\
\mathbb{E}d_2(S, \varphi(W)) &\le D_2.
\end{aligned}
$$

- $\max_{\hat{S}}[I(\hat{S};Y) - I(\hat{S};S)]$ is achievable, but sub-optimal.

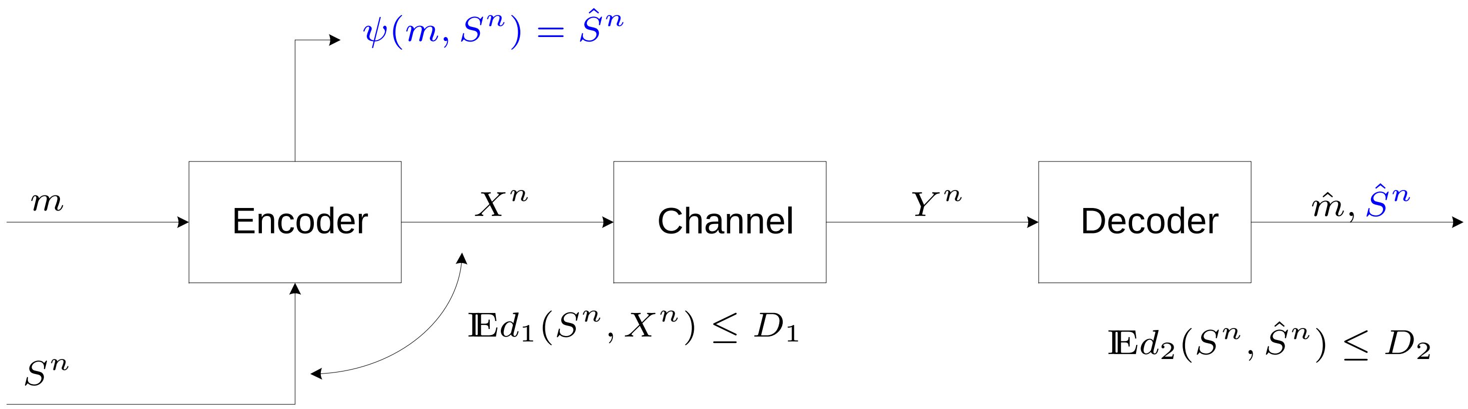
Main Result (cont'd)



$$
\begin{aligned}
C_{ck}(D) &= \max[I(W;Y) - I(W;S)], \quad W \multimap X \multimap Y \\
\mathbb{E}d_1(S, X) &\le D_1, \\
\mathbb{E}d_2(S, \varphi(W)) &\le D_2.
\end{aligned}
$$

- $\max_{\hat{S}}[I(\hat{S};Y) - I(\hat{S};S)]$ is achievable, but sub-optimal.
- Although W (or $\varphi(W)$) is a good source code generated by the encoder, separation is sub-optimal.

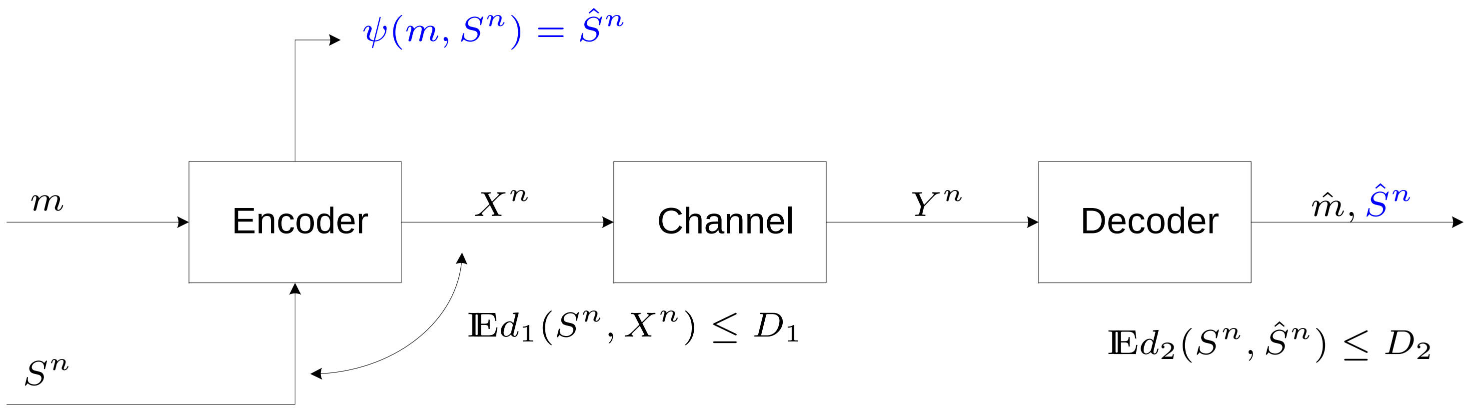
Main Result (cont'd)



$$\begin{aligned} C_{ck}(D) &= \max[I(W;Y) - I(W;S)], \qquad W -\!\circ\!- X -\!\circ\!- Y \\ \mathbb{E}d_1(S,X) &\leq D_1, \\ \mathbb{E}d_2(S,\varphi(W)) &\leq D_2. \end{aligned}$$

- $\max_{\hat{S}}[I(\hat{S};Y) - I(\hat{S};S)]$ is achievable, but sub-optimal.
- Although W (or $\varphi(W)$) is a good source code generated by the encoder, separation is sub-optimal.
- The channel can depend explicitly on S ($P_{Y|X,S}$), in which case $W -\!\circ\!- (X,S) -\!\circ\!- Y$.
  $\Rightarrow$ Solves the problem of joint transmission of data and state (Sutivong *et. al.*) under CK constraint.

Main Result *(cont'd)*



$$\begin{aligned} C_{ck}(D) &= \max[I(W;Y) - I(W;S)], \qquad W \multimap X \multimap Y \\ \mathbb{E}d_1(S,X) &\le D_1, \\ \mathbb{E}d_2(S,\varphi(W)) &\le D_2. \end{aligned}$$

- $\max_{\hat{S}}[I(\hat{S};Y) - I(\hat{S};S)]$ is achievable, but sub-optimal.
- Although W (or $\varphi(W)$) is a good source code generated by the encoder, separation is sub-optimal.
- The channel can depend explicitly on S ($P_{Y|X,S}$), in which case $W \multimap (X,S) \multimap Y$.
  $\Rightarrow$ Solves the problem of joint transmission of data and state (Sutivong *et. al.*) under CK constraint.
- Can bound $|\mathcal{W}| \le |\mathcal{X}||\mathcal{S}| + 3$.

Example

Example 1 *Binary symmetric channel & host, Hamming distortion measures*

$$Y = X \oplus Z, \quad Z \sim \textit{Bernoulli}(p_z), \quad S \sim \textit{Bernoulli}(1/2),$$

Example

Example 1 *Binary symmetric channel & host, Hamming distortion measures*

$$Y = X \oplus Z, \quad Z \sim \text{Bernoulli}(p_z), \quad S \sim \text{Bernoulli}(1/2),$$

$$\mathbb{E} d_H(S^n, X^n) \leq D_1, \quad \mathbb{E} d_H(S^n, \hat{S}^n) \leq D_2, \quad P(\hat{S}^n \neq \psi(m, S^n)) \leq \epsilon.$$

Example

Example 1 *Binary symmetric channel & host, Hamming distortion measures*

$$Y = X \oplus Z, \quad Z \sim \textit{Bernoulli}(p_z), \quad S \sim \textit{Bernoulli}(1/2),$$

$$\mathbb{E} d_H(S^n, X^n) \leq D_1, \quad \mathbb{E} d_H(S^n, \hat{S}^n) \leq D_2, \quad P(\hat{S}^n \neq \psi(m, S^n)) \leq \epsilon.$$

1. *A separation approach:*

$$R_S(D_2) = 1 - h(D_2), \quad C = \text{u.c.e}\{g(D_1)\}$$

where

$$g(D_1) = \begin{cases} 0 & 0 \leq D_1 \leq p_z \\ h(D_1) - h(p_z) & p_z \leq D_1 \leq 1/2. \end{cases}$$

Example

Example 1 *Binary symmetric channel & host, Hamming distortion measures*

$$Y = X \oplus Z, \quad Z \sim \textit{Bernoulli}(p_z), \quad S \sim \textit{Bernoulli}(1/2),$$

$$\mathbb{E} d_H(S^n, X^n) \leq D_1, \quad \mathbb{E} d_H(S^n, \hat{S}^n) \leq D_2, \quad P(\hat{S}^n \neq \psi(m, S^n)) \leq \epsilon.$$

1. A separation approach:

$$R_S(D_2) = 1 - h(D_2), \quad C = \text{u.c.e}\{g(D_1)\}$$

where

$$g(D_1) = \begin{cases} 0 & 0 \leq D_1 \leq p_z \\ h(D_1) - h(p_z) & p_z \leq D_1 \leq 1/2. \end{cases}$$

The achievable distortions with zero embedding rate

$$1 - h(D_2) \leq \text{u.c.e}\{g(D_1)\}.$$

Example

Example 1 *Binary symmetric channel & host, Hamming distortion measures*

$$Y = X \oplus Z, \quad Z \sim \text{Bernoulli}(p_z), \quad S \sim \text{Bernoulli}(1/2),$$

$$\mathbb{E} d_H(S^n, X^n) \leq D_1, \quad \mathbb{E} d_H(S^n, \hat{S}^n) \leq D_2, \quad P(\hat{S}^n \neq \psi(m, S^n)) \leq \epsilon.$$

1. *A separation approach:*

$$R_S(D_2) = 1 - h(D_2), \quad C = \text{u.c.e}\{g(D_1)\}$$

where

$$g(D_1) = \begin{cases} 0 & 0 \leq D_1 \leq p_z \\ h(D_1) - h(p_z) & p_z \leq D_1 \leq 1/2. \end{cases}$$

The achievable distortions with zero embedding rate

$$1 - h(D_2) \leq \text{u.c.e}\{g(D_1)\}.$$

$\Rightarrow$ *For* $p_z = 0$

$$1 - h(D_2) \leq h(D_1).$$

Example (cont'd)

2. *But for this system, can achieve* $D_1 = D_2 = 0$, *if we set*

$$X_i = S_i \quad \forall i.$$

Example *(cont'd)*

2. *But for this system, can achieve* $D_1 = D_2 = 0$, *if we set*

$$X_i = S_i \quad \forall i.$$

In the capacity formula, choose

$$W = X = S, \quad \varphi(W) = \varphi(S) = S.$$

Then $D_1 = D_2 = 0$ *are achievable. Moreover*

$$I(W;Y) - I(W;S) = I(S;S) - I(S;S) = 0,$$

hence this substitution is valid.

Thank You!