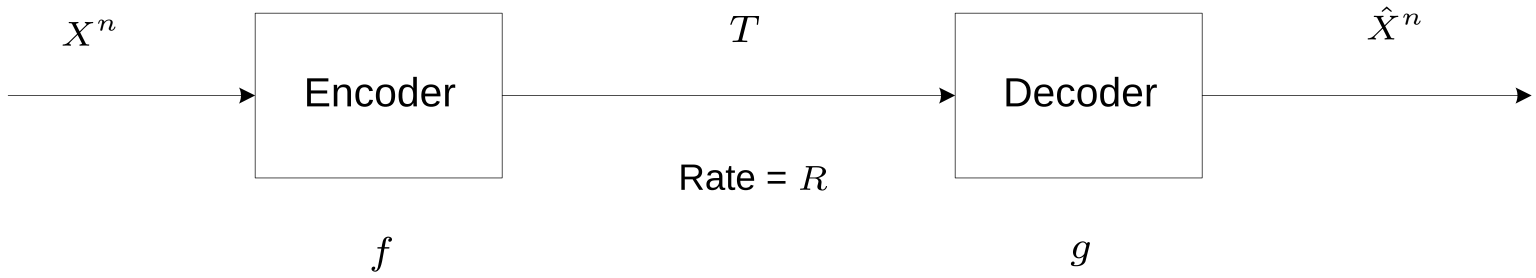
Coding and Common Knowledge

Yossef Steinberg
Technion—Israel Institute of Technology
ysteinbe@ee.technion.ac.il

Introduction

Motivation

Classical Rate-Distortion Theory:



- The code designer is concerned with reducing the rate R under a constraint on the distortion.

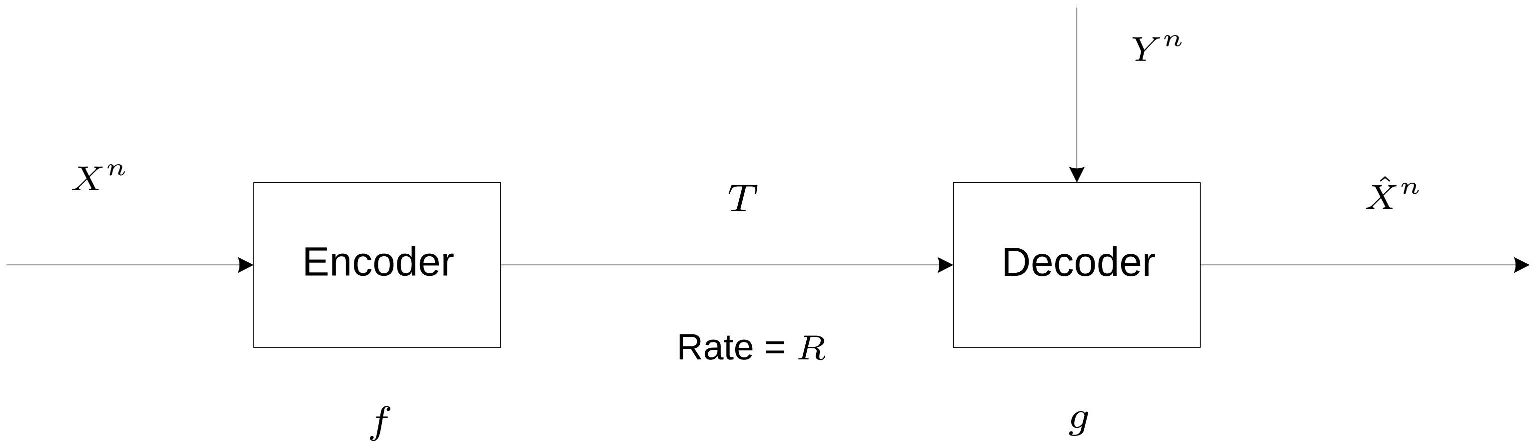
  The question of whether the sender knows $\hat{X}^n$ is not raised.

- $\hat{X}^n = g(f(X^n)), \quad \mathbb{E}d(X^n, \hat{X}^n) \leq D$

- $\hat{X}^n$ is known at the encoder.

Motivation *(cont'd)*

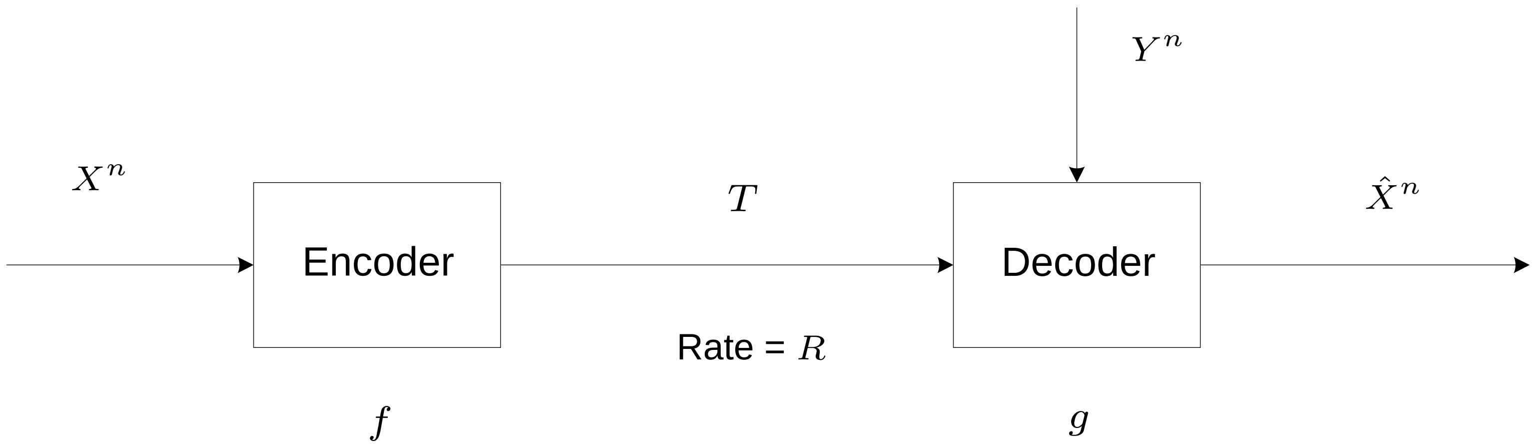
The Wyner-Ziv Problem:



- The realization of Y^n is unknown to the encoder.
- $\hat{X}^n = g(f(X^n), Y^n), \quad \mathbb{E}d(X^n, \hat{X}^n) \leq D$.
- The reconstruction $\hat{X}^n$ cannot be reproduced at the sender side.

Motivation (cont'd)

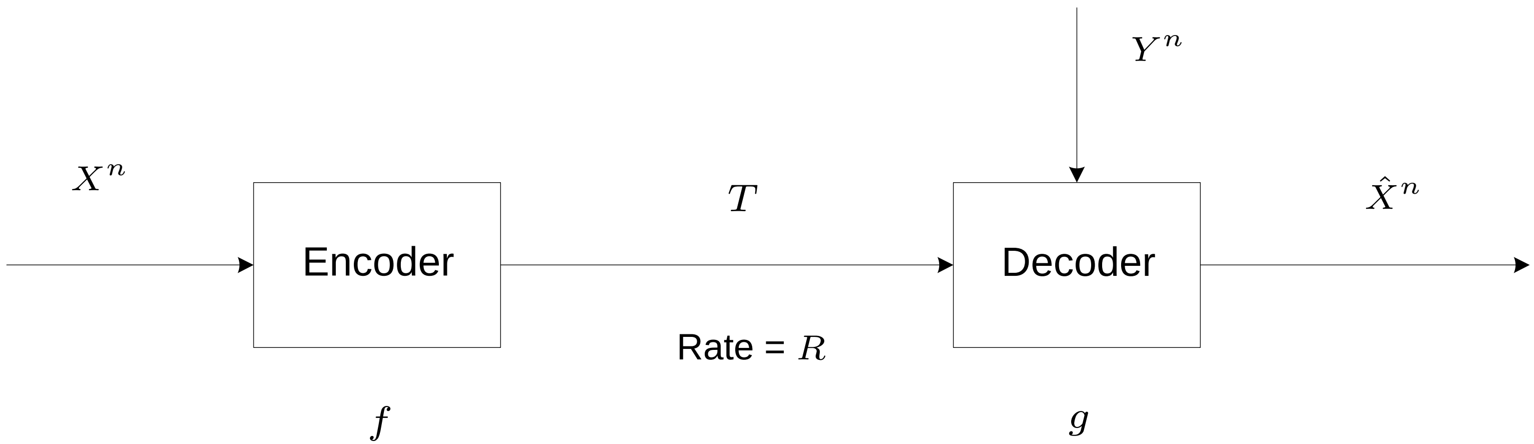
The Wyner-Ziv Problem:



- The realization of Y^n is unknown to the encoder.
- $\hat{X}^n = g(f(X^n), Y^n), \quad \mathbb{E}d(X^n, \hat{X}^n) \le D$.
- The reconstruction $\hat{X}^n$ cannot be reproduced at the sender side. The sender does not know the distortion pattern.

Motivation (cont'd)

The Wyner-Ziv Problem:

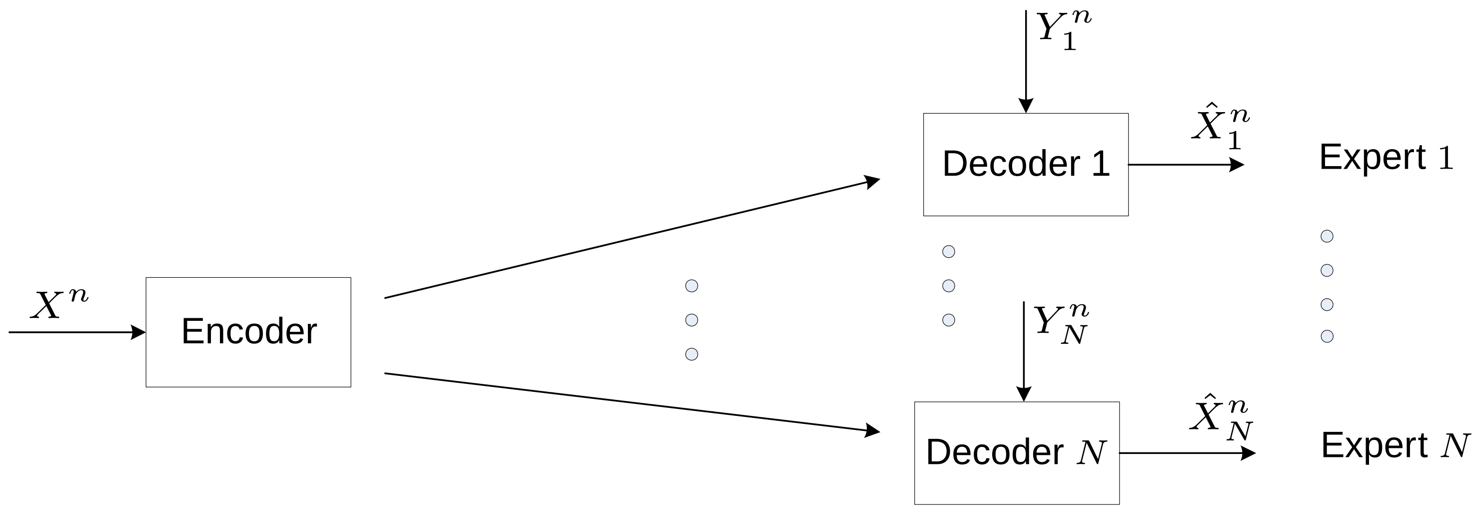


- The realization of Y^n is unknown to the encoder.
- $\hat{X}^n = g(f(X^n), Y^n), \quad \mathbb{E}d(X^n, \hat{X}^n) \leq D.$
- The reconstruction $\hat{X}^n$ cannot be reproduced at the sender side. The sender does not know the distortion pattern.

In some applications, this is a drawback.

Motivation (cont'd)

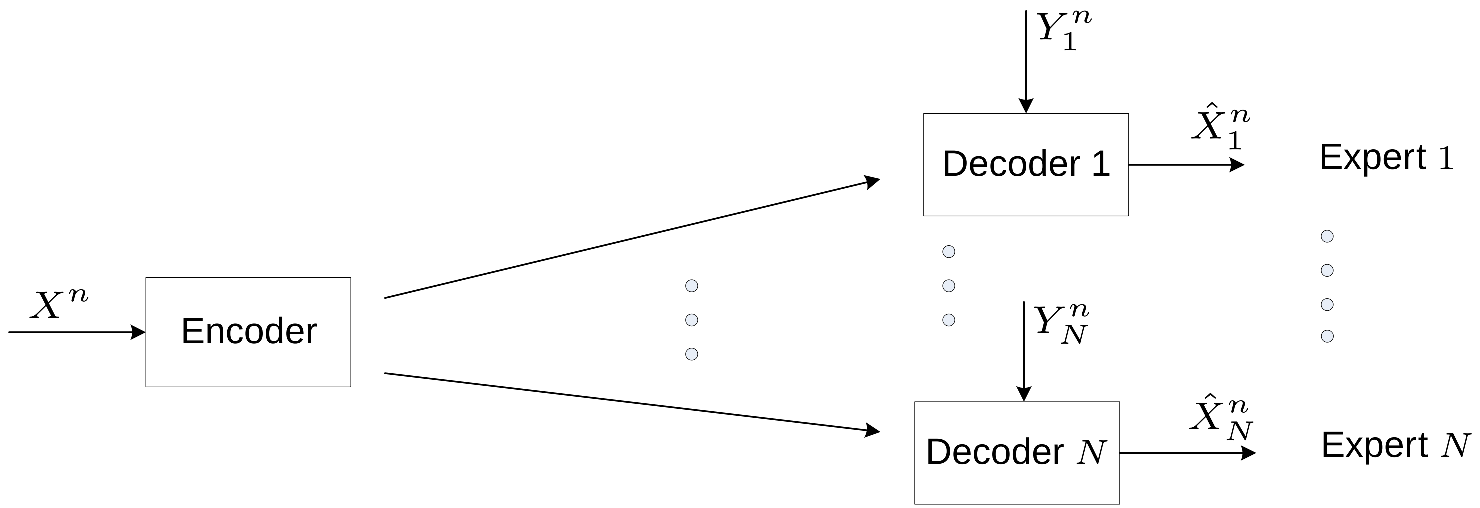
Consider the transmission of medical information over noisy BC



- Medical data is sent to a number of experts, for consultation.
  Each has side information Y_j^n, $\quad j = 1, 2, \ldots N$ about the patient.

Motivation (cont'd)

Consider the transmission of medical information over noisy BC



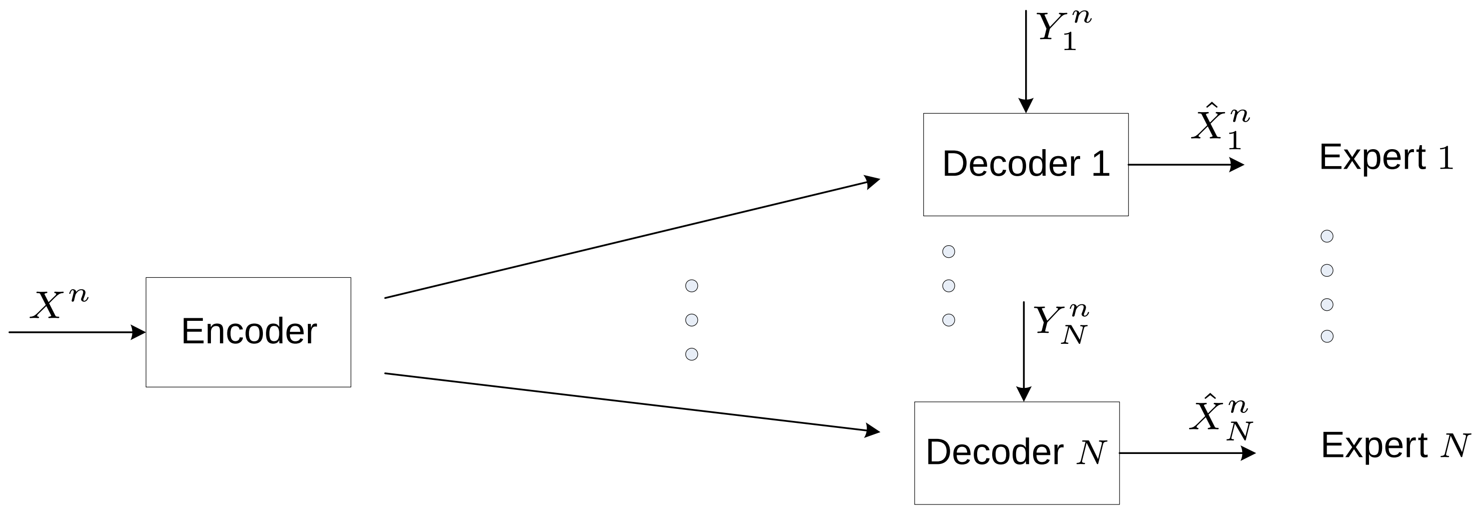
- Medical data is sent to a number of experts, for consultation.

  Each has side information $Y_j^n, \quad j = 1, 2, \ldots N$ about the patient.

- Lossy transmission, due to limitations of the channel.

Motivation *(cont'd)*

Consider the transmission of medical information over noisy BC



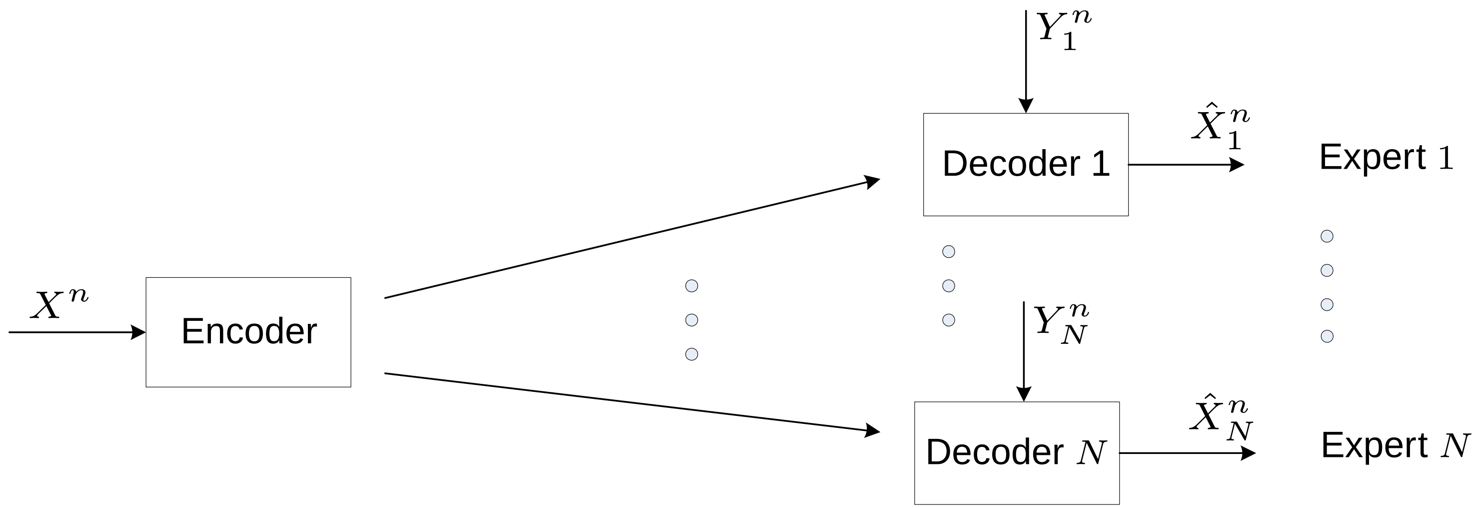
- Medical data is sent to a number of experts, for consultation.

  Each has side information $Y_j^n, \quad j = 1, 2, \ldots N$ about the patient.
- Lossy transmission, due to limitations of the channel.
- The coding scheme guarantees *average* distortion.

  The distortion pattern is not known at the sender side.

Motivation (cont'd)

Consider the transmission of medical information over noisy BC



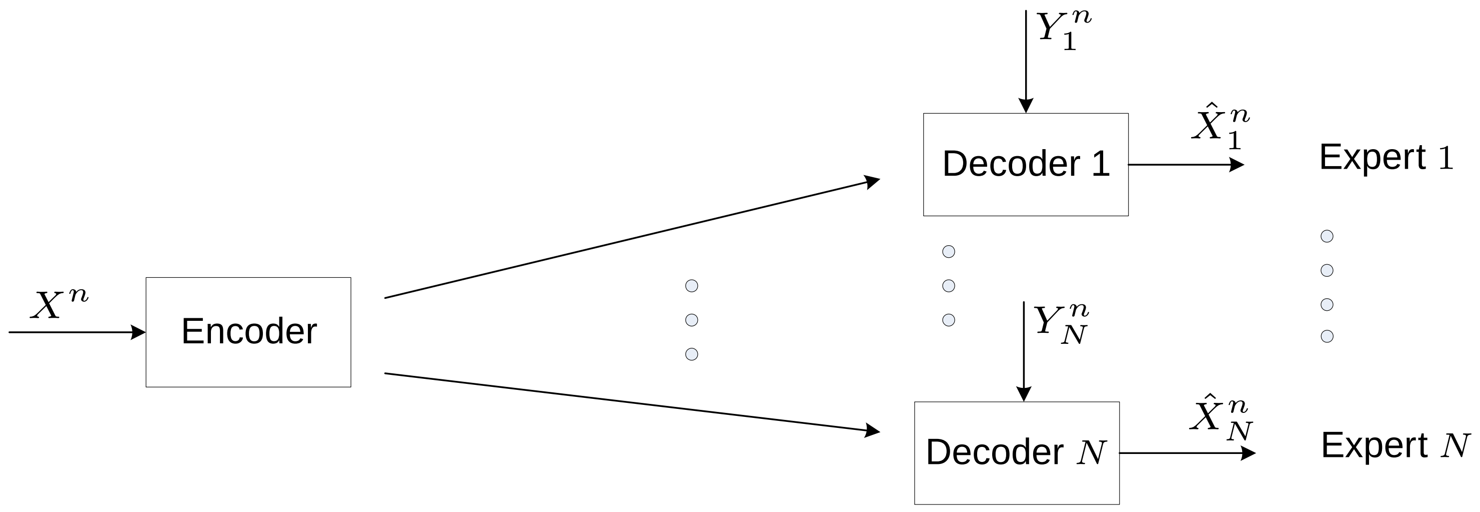
- Medical data is sent to a number of experts, for consultation.

  Each has side information $Y_j^n, \quad j = 1, 2, \ldots N$ about the patient.
- Lossy transmission, due to limitations of the channel.
- The coding scheme guarantees *average* distortion.

  The distortion pattern is not known at the sender side. (BC,

Motivation (cont'd)

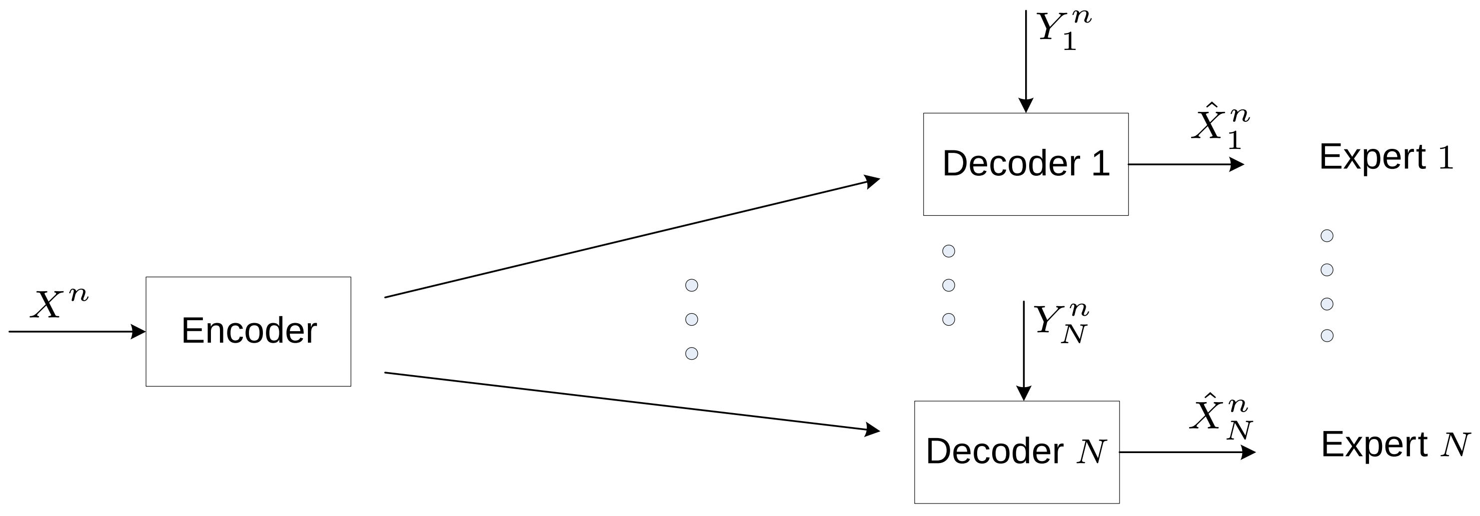
Consider the transmission of medical information over noisy BC



- Medical data is sent to a number of experts, for consultation.
  Each has side information $Y_j^n, \quad j = 1, 2, \ldots N$ about the patient.
- Lossy transmission, due to limitations of the channel.
- The coding scheme guarantees *average* distortion.
  The distortion pattern is not known at the sender side. (BC, SI)

Motivation (cont'd)

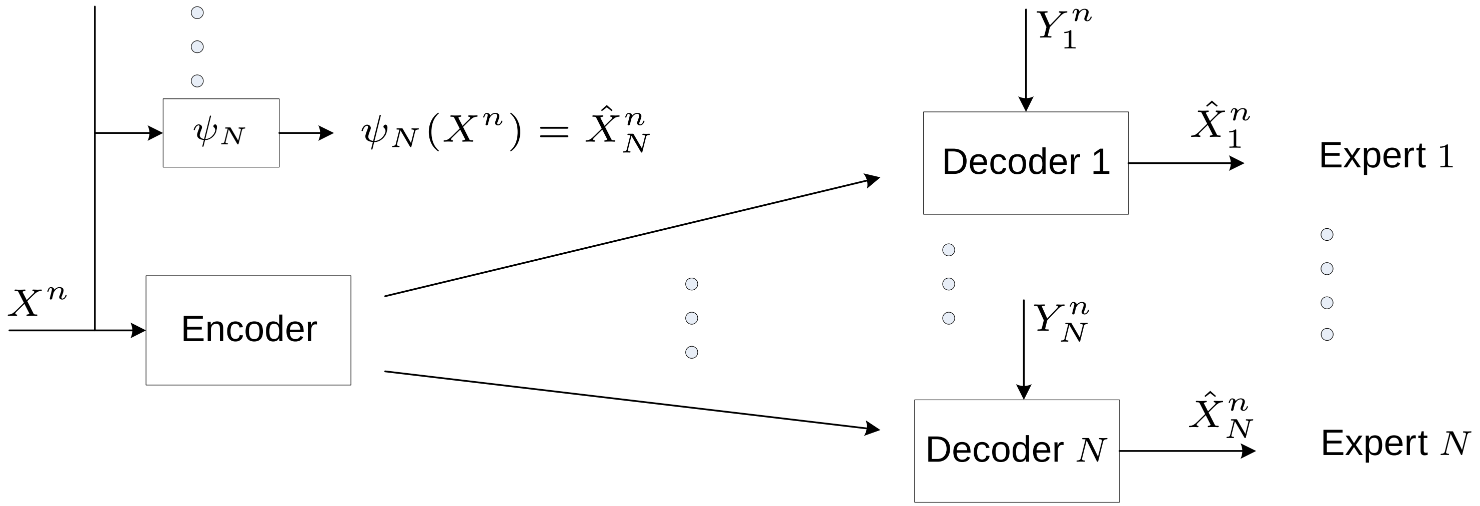
Consider the transmission of medical information over noisy BC



- Medical data is sent to a number of experts, for consultation.
  Each has side information $Y_j^n, \quad j = 1, 2, \ldots N$ about the patient.
- Lossy transmission, due to limitations of the channel.
- The coding scheme guarantees *average* distortion.
  The distortion pattern is not known at the sender side. (BC, SI)

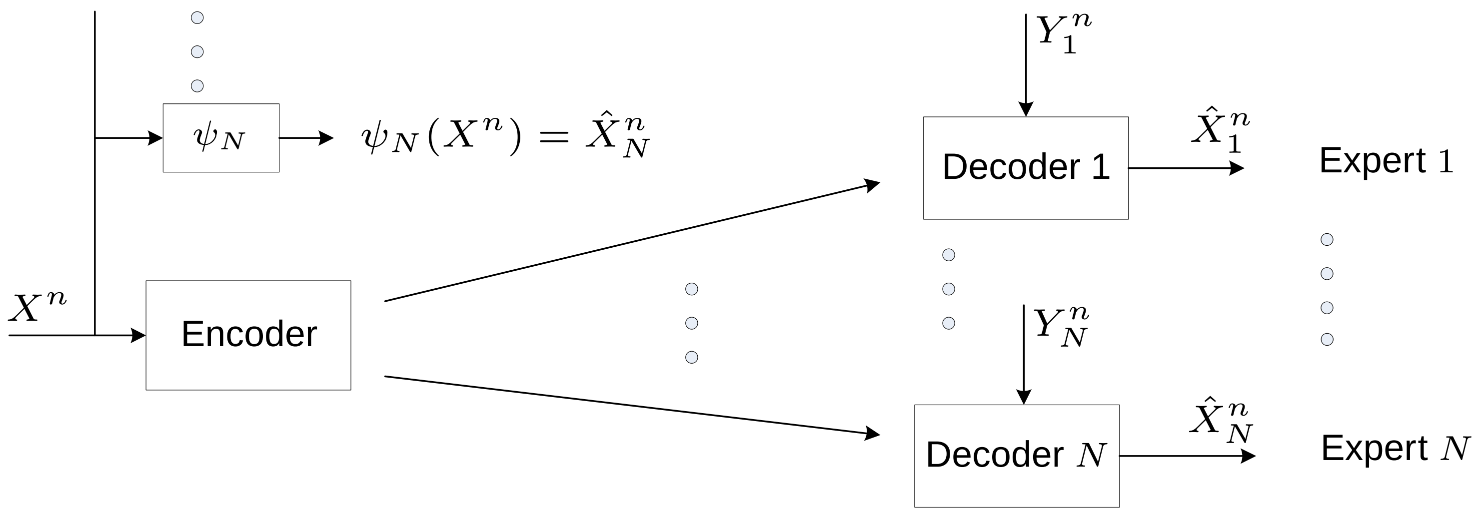
Important details can be blurred during transmission. Sender is unaware.

Motivation (cont'd)



Devise a coding scheme that enables the sender to produce locally $\hat{X}_1^n \ldots, \hat{X}_N^n$.

Motivation (cont'd)

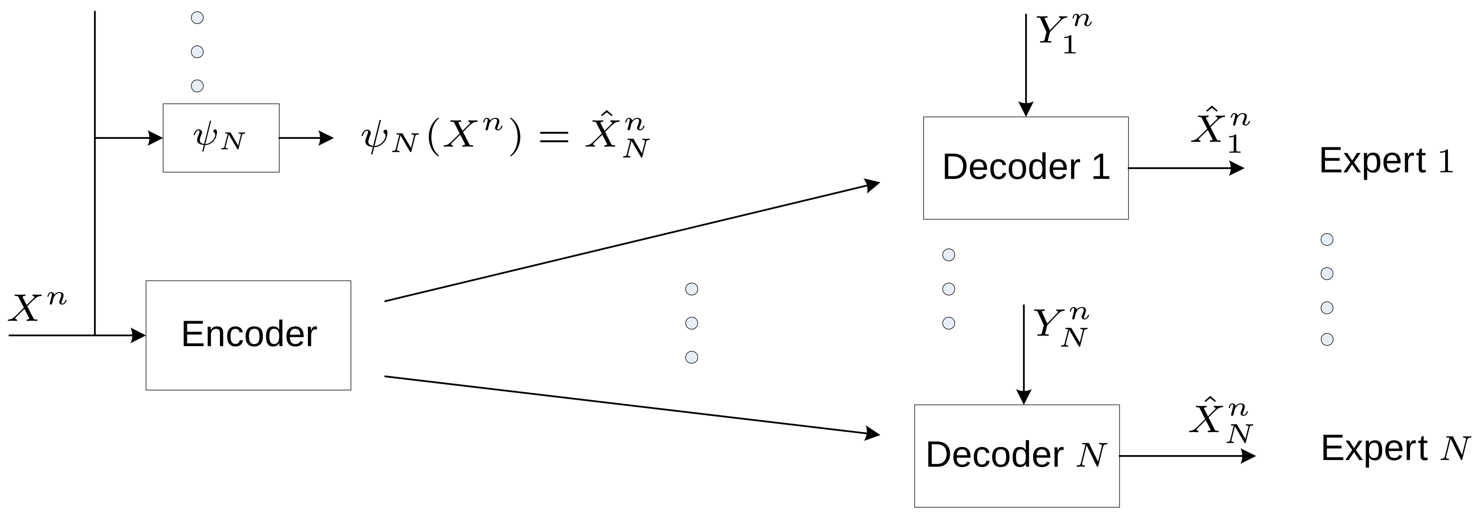


Devise a coding scheme that enables the sender to produce locally $\hat{X}_1^n \ldots, \hat{X}_N^n$.

Role of the common reproduction:

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

Motivation *(cont'd)*



Devise a coding scheme that enables the sender to produce locally $\hat{X}_1^n \ldots, \hat{X}_N^n$.

Role of the common reproduction:

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

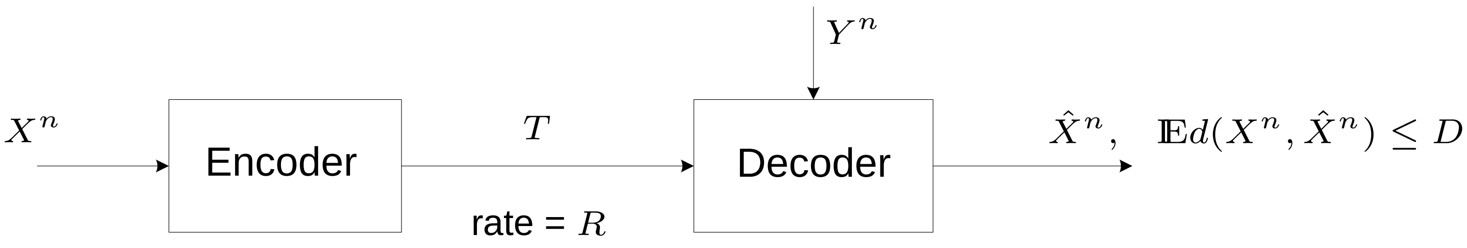
$\Longrightarrow$ *Common Knowledge (CK)* constraint

Outline

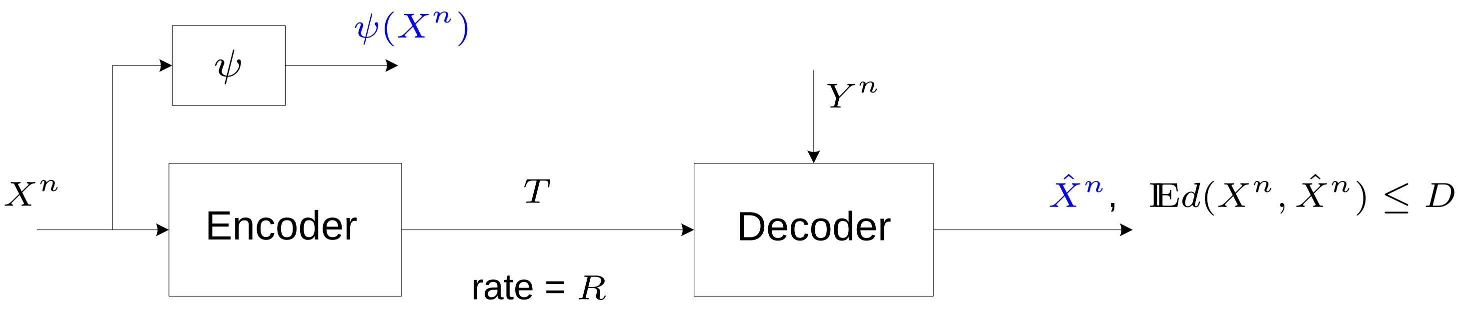
- Source coding with side information at the decoder
- Examples
- Joint source-channel coding for the broadcast channel
- Does common knowledge constraint imply the optimality of a separation-based scheme?

Source coding with side information

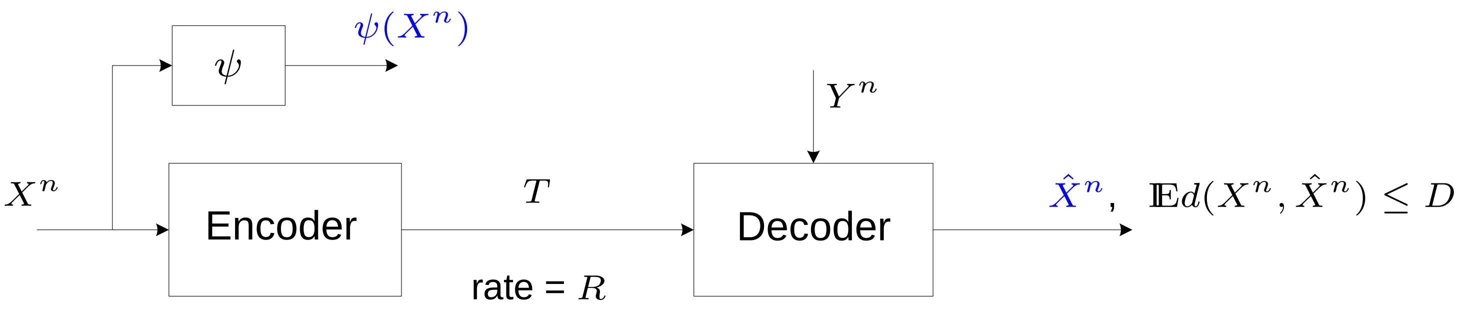
Problem formulation



Problem formulation *(cont'd)*



Problem formulation (cont'd)



Definition: Let $\mathcal{T} = \{1, 2, \dots, 2^{nR}\}$. An $(n, 2^{nR}, D, \epsilon)$ common knowledge (CK) code for the source X with decoder side information Y consists of an encoder-decoder pair

$$f : \mathcal{X}^n \to \mathcal{T}, \qquad g : \mathcal{T} \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n,$$

and a sender reconstruction map

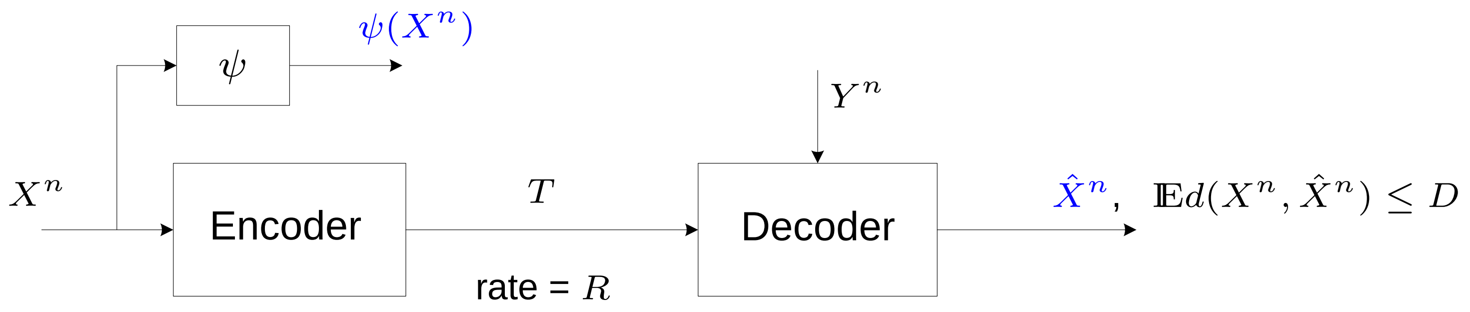
$$\psi : \mathcal{X}^n \to \hat{\mathcal{X}}^n,$$

such that

$$\mathbb{E}d(X^n, g(f(X^n), Y^n)) \le D,$$

$$P_{XY}\left(\psi(X^n) \ne g(f(X^n), Y^n)\right) \le \epsilon.$$

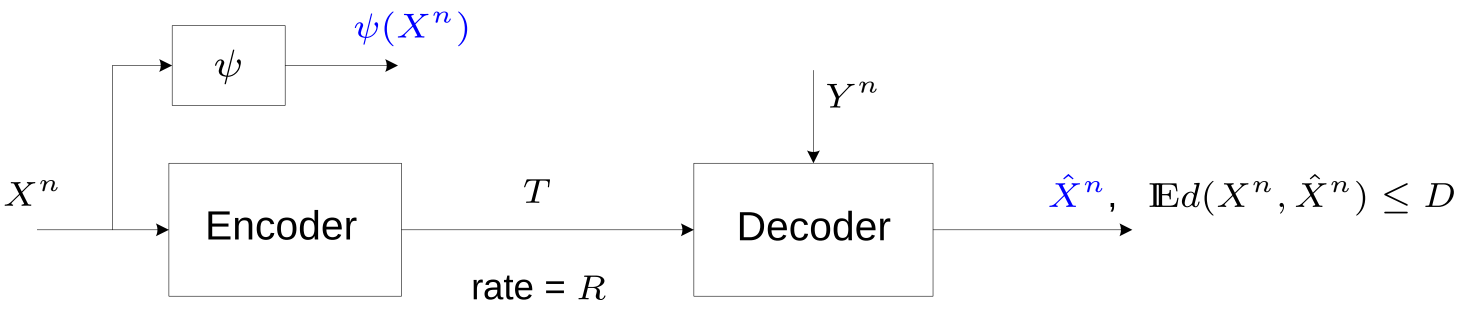
Problem formulation (cont'd)



$$\hat{X}^n = g(f(X^n), Y^n)$$

$$P_{XY}\left(\psi(X^n) \neq \hat{X}^n\right) \leq \epsilon \quad \text{(CK constraint)}.$$

Problem formulation (cont'd)

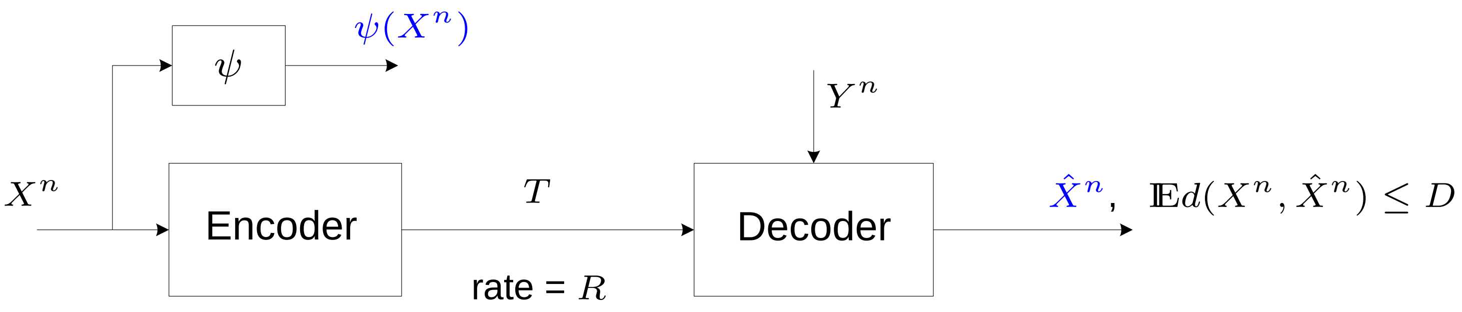


$$\hat{X}^n = g(f(X^n), Y^n)$$

$$P_{XY}\left(\psi(X^n) \neq \hat{X}^n\right) \leq \epsilon \quad \text{(CK constraint)}.$$

- The CK rate-distortion function, $R_{ck}(D)$, is the minimal achievable CK coding rate, under average distortion D and arbitrarily small ϵ.

Main result



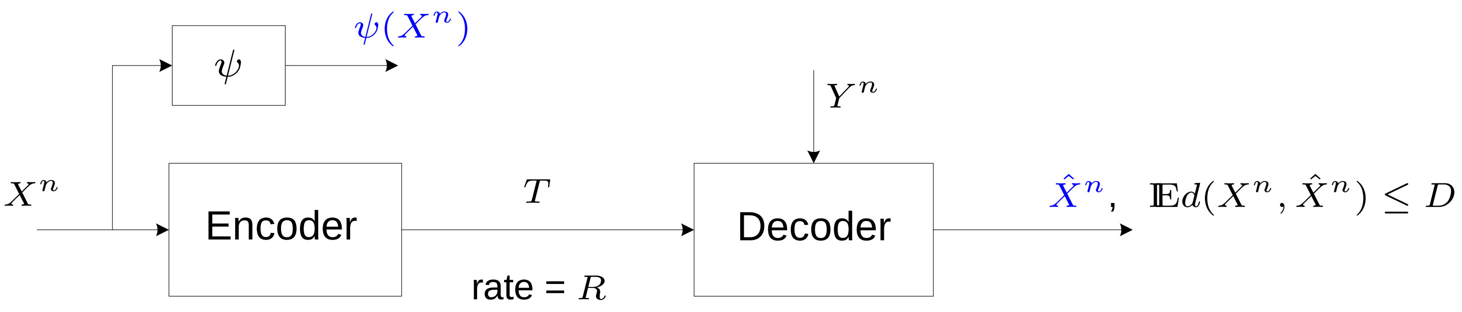
Theorem 1

$$R_{ck}(D) = \min[I(\hat{X}; X) - I(\hat{X}; Y)]$$

where the minimum is over all $\hat{X}$ *such that* $\hat{X} \multimap X \multimap Y$ *and*

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

Main result



Theorem 1

$$R_{ck}(D) = \min[I(\hat{X}; X) - I(\hat{X}; Y)]$$

where the minimum is over all $\hat{X}$ *such that* $\hat{X} \multimap X \multimap Y$ *and*

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

Due to the Markov conditions, $R_{ck}(D) = \min I(\hat{X}; X|Y)$.

Comparison with the Wyner-Ziv rate-distortion function

Let U and $\hat{X}$ satisfy $U \multimap X \multimap Y$, $\hat{X} \multimap X \multimap Y$. Then

$$R_{WZ}(D) = \min[I(U;X) - I(U;Y)] \quad \mathbb{E}d(X, \phi(U,Y)) \leq D$$

$$R_{ck}(D) = \min[I(\hat{X};X) - I(\hat{X};Y)] \quad \mathbb{E}d(X, \hat{X}) \leq D$$

Comparison with the Wyner-Ziv rate-distortion function

Let U and $\hat{X}$ satisfy $U \multimap X \multimap Y$, $\hat{X} \multimap X \multimap Y$. Then

$$R_{WZ}(D) = \min[I(U;X) - I(U;Y)] \quad \mathbb{E}d(X, \phi(U,Y)) \le D$$

$$R_{ck}(D) = \min[I(\hat{X};X) - I(\hat{X};Y)] \quad \mathbb{E}d(X, \hat{X}) \le D$$

- In the WZ problem, the codewords U^n need not satisfy the distortion constraint by themselves. The side information Y is used for *binning* and *estimation.*

Comparison with the Wyner-Ziv rate-distortion function

Let U and $\hat{X}$ satisfy $U \multimap X \multimap Y$, $\hat{X} \multimap X \multimap Y$. Then

$$R_{WZ}(D) = \min[I(U;X) - I(U;Y)] \quad \mathbb{E}d(X, \phi(U,Y)) \le D$$

$$R_{ck}(D) = \min[I(\hat{X};X) - I(\hat{X};Y)] \quad \mathbb{E}d(X, \hat{X}) \le D$$

- In the WZ problem, the codewords U^n need not satisfy the distortion constraint by themselves. The side information Y is used for *binning* and *estimation.*
- In the CK problem, the codewords $\hat{X}^n$ satisfy the distortion constraint by themselves. The decoder uses Y to resolve the binning, but cannot use it to further improve the estimation.

Comparison with the Wyner-Ziv rate-distortion function

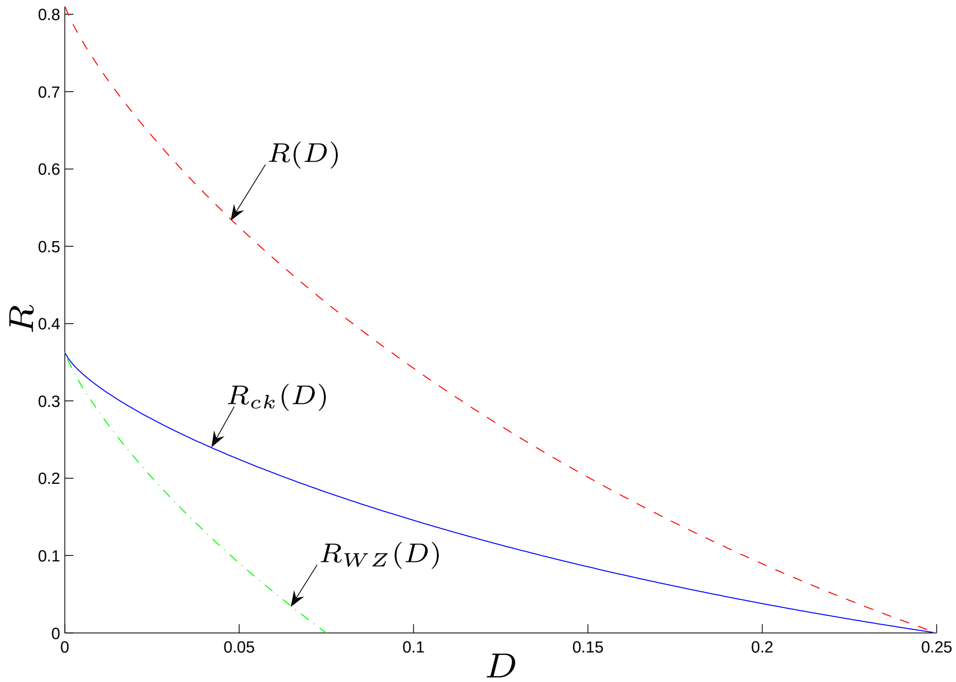
Let U and $\hat{X}$ satisfy $U \multimap X \multimap Y$, $\hat{X} \multimap X \multimap Y$. Then

$$R_{WZ}(D) = \min[I(U;X) - I(U;Y)] \quad \mathbb{E}d(X, \phi(U,Y)) \leq D$$

$$R_{ck}(D) = \min[I(\hat{X};X) - I(\hat{X};Y)] \quad \mathbb{E}d(X, \hat{X}) \leq D$$

- In the WZ problem, the codewords U^n need not satisfy the distortion constraint by themselves. The side information Y is used for *binning* and *estimation.*
- In the CK problem, the codewords $\hat{X}^n$ satisfy the distortion constraint by themselves. The decoder uses Y to resolve the binning, but cannot use it to further improve the estimation.
- There is no external random variable in the CK problem.

Typical curves

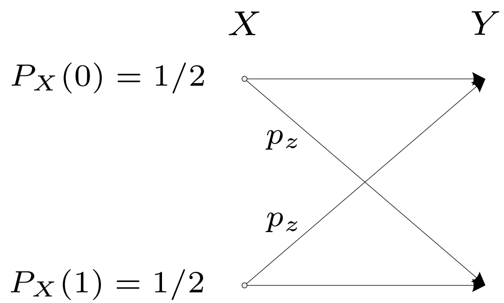


$$R_{ck}(D) = \min[I(\hat{X};X) - I(\hat{X};Y)] \qquad \mathbb{E}d(X,\hat{X}) \leq D$$

$$\hat{X} \multimap X \multimap Y$$

Examples

Example 1 *The doubly symmetric binary source, Hamming distortion measure*

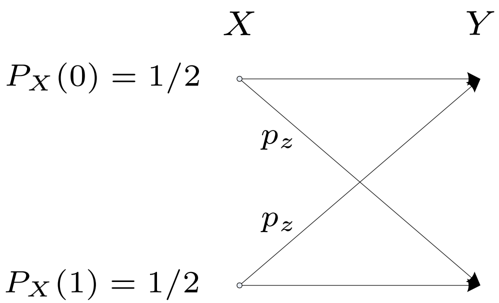


$$Y = X \oplus Z, \qquad Z \sim \text{Bernoulli}(p_z)$$

$$R_{ck}(D) = h(p_z \star D) - h(D), \qquad 0 \leq D \leq 1/2.$$

Examples

Example 1 *The doubly symmetric binary source, Hamming distortion measure*



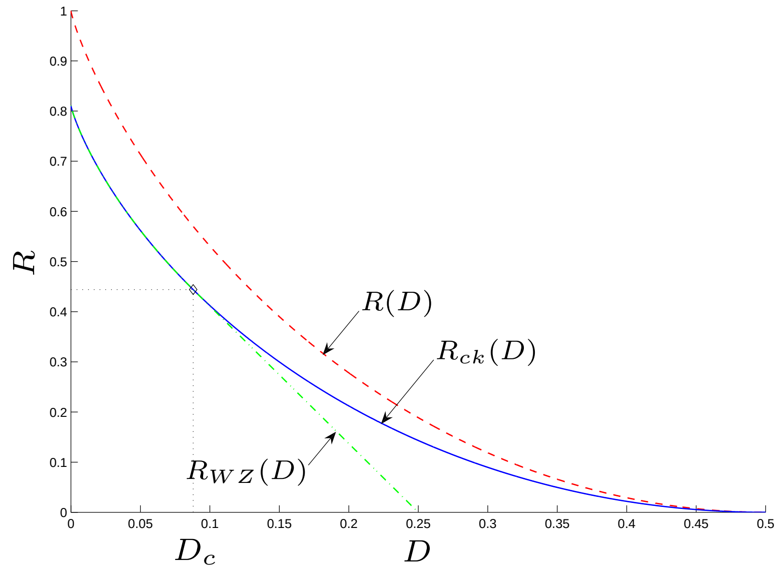
$$Y = X \oplus Z, \qquad Z \sim \text{Bernoulli}(p_z)$$

$$R_{ck}(D) = h(p_z \star D) - h(D), \qquad 0 \leq D \leq 1/2.$$

Closely related to the Wyner-Ziv rate-distortion function

$$R_{WZ}(D) = \text{l.c.e}\left\{h(p_z \star D) - h(D), (p_z, 0)\right\}.$$

Examples – doubly symmetric source (cont'd)



Corollary 1 *For the binary doubly symmetric source with Hamming distortion measure, no penalty is incurred due to the common knowledge constraint in the region* $0 \leq D \leq D_c$.

Examples *(cont'd)*

Example 2 *Gaussian source and square error distortion measure.*

$$Y = X + Z, \qquad X \sim \mathcal{N}(0, \sigma_X^2), \qquad Z \sim \mathcal{N}(0, \sigma_Z^2), \qquad X \perp V$$

Examples *(cont'd)*

Example 2 *Gaussian source and square error distortion measure.*

$$Y = X + Z, \qquad X \sim \mathcal{N}(0, \sigma_X^2), \qquad Z \sim \mathcal{N}(0, \sigma_Z^2), \qquad X \perp V$$

$$R_{ck}(D) = \frac{1}{2} \log \left(\frac{\sigma_X^2 \sigma_Z^2}{(\sigma_X^2 + \sigma_Z^2) D} \cdot \frac{D + \sigma_Z^2}{\sigma_Z^2} \right).$$

Examples (cont'd)

Example 2 *Gaussian source and square error distortion measure.*

$$Y = X + Z, \qquad X \sim \mathcal{N}(0, \sigma_X^2), \qquad Z \sim \mathcal{N}(0, \sigma_Z^2), \qquad X \perp V$$

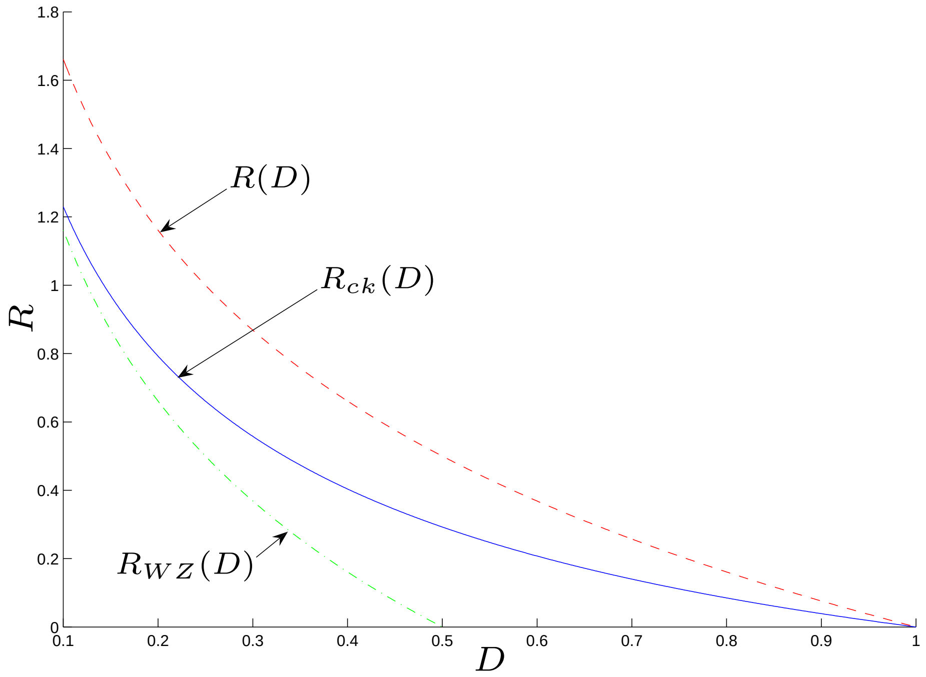
$$R_{ck}(D) = \frac{1}{2}\log\left(\frac{\sigma_X^2\sigma_Z^2}{(\sigma_X^2 + \sigma_Z^2)D} \cdot \frac{D + \sigma_Z^2}{\sigma_Z^2}\right).$$

Note that

$$R_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2}\log\left(\frac{\sigma_X^2\sigma_Z^2}{(\sigma_X^2 + \sigma_Z^2)D}\right),$$

therefore $\frac{1}{2}\log\left(\frac{D+\sigma_Z^2}{\sigma_Z^2}\right)$ is the penalty due to the CK constraint.

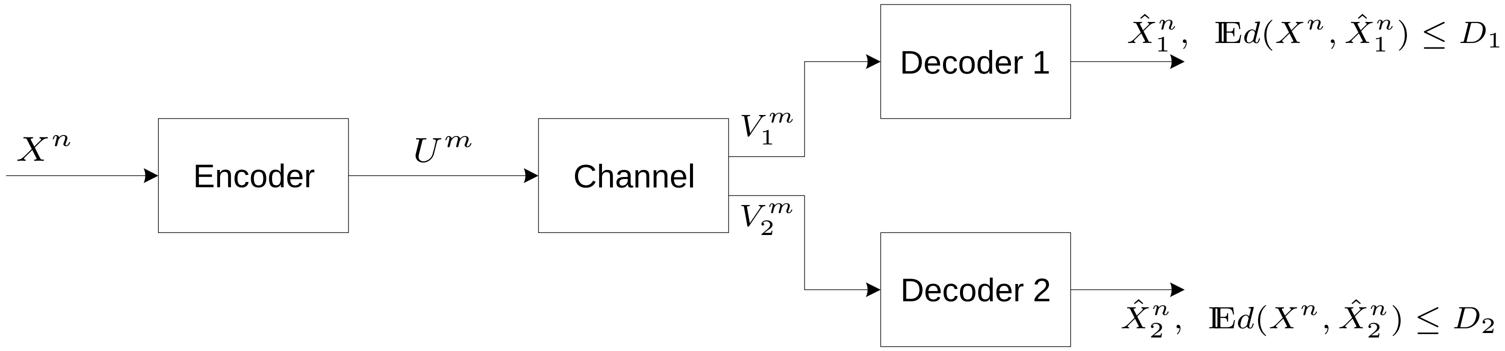
Examples – Gaussian source (cont'd)



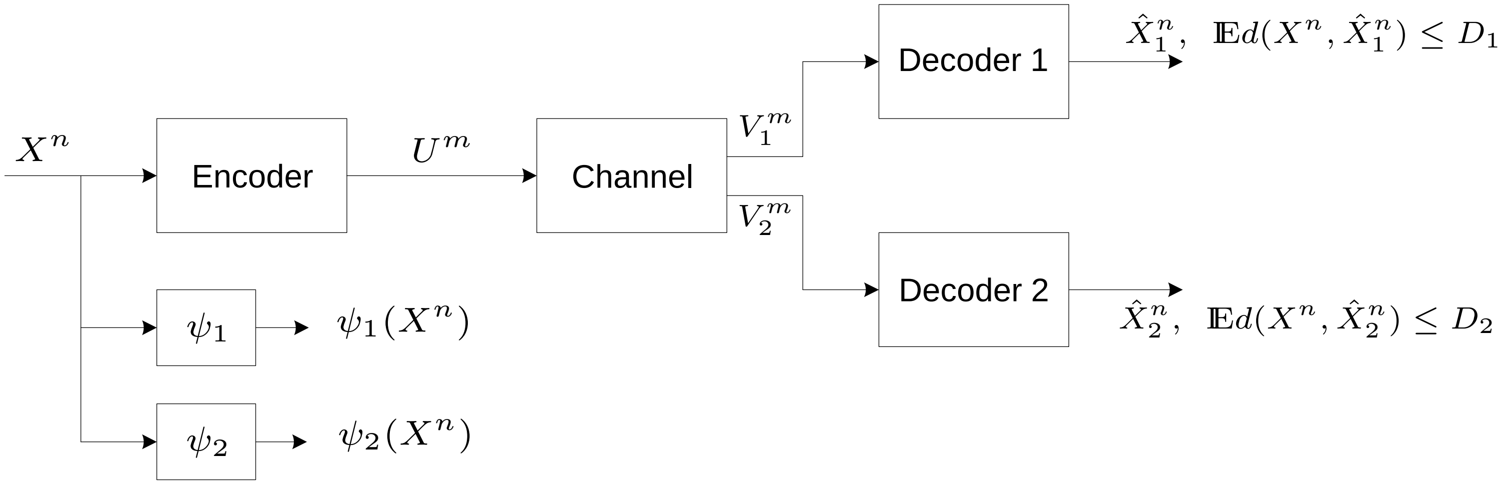
Here $\sigma_X^2 = \sigma_Z^2 = 1$.

Joint source-channel coding for the BC

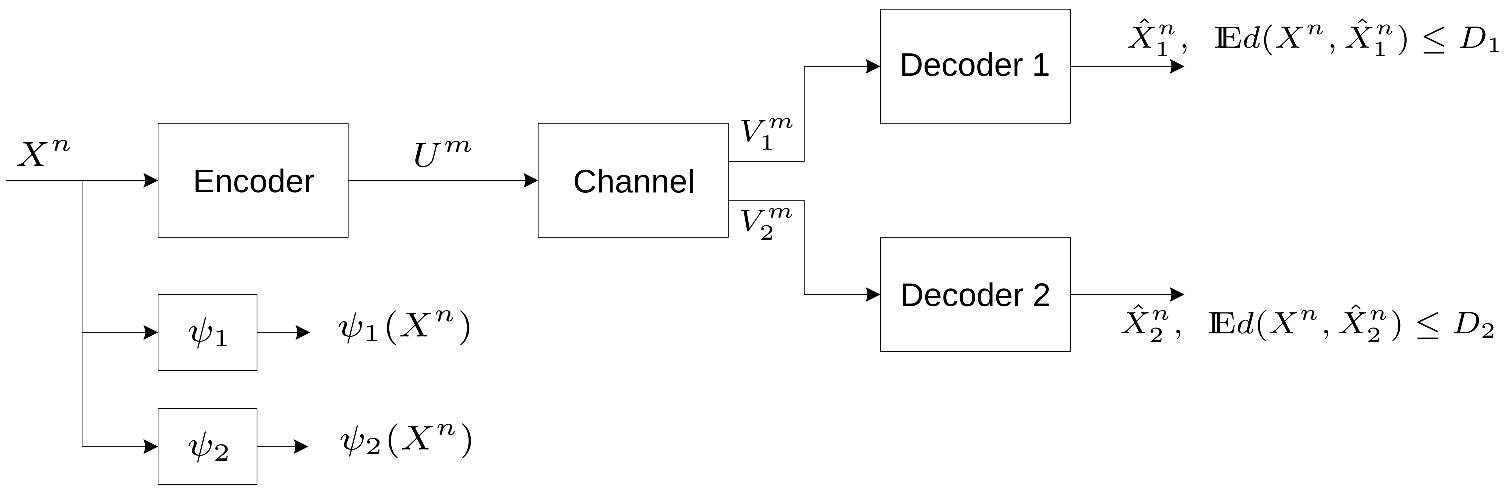
Problem formulation



Problem formulation

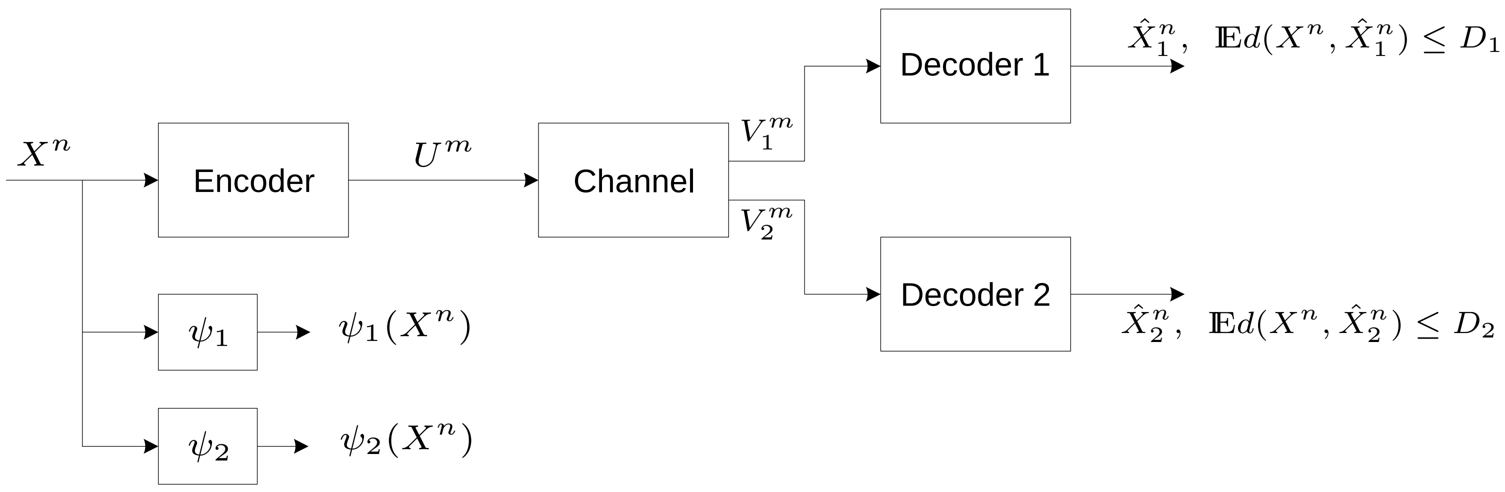


Problem formulation



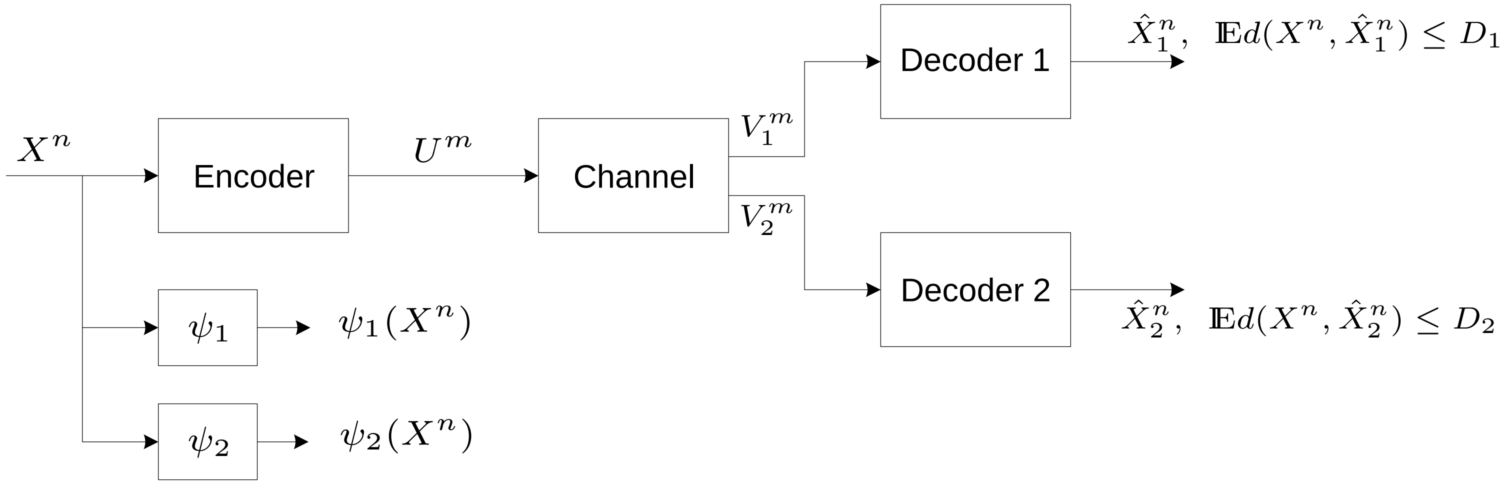
- Memoryless, degraded broadcast channel $U \multimap V_2 \multimap V_1$.

Problem formulation



- Memoryless, degraded broadcast channel $U \multimap V_2 \multimap V_1$.
- Bandwidth expansion ratio $\rho = m/n$.

Problem formulation



- Memoryless, degraded broadcast channel $U \multimap V_2 \multimap V_1$.
- Bandwidth expansion ratio $\rho = m/n$.
- CK constraints:

$$P\left(\psi_j(X^n) \neq \hat{X}_j^n\right) \leq \epsilon, \qquad j = 1, 2.$$

Main result

- $\mathcal{C}$ – the capacity region of the degraded broadcast channel $P_{V_1,V_2|U}$

Main result

- $\mathcal{C}$ – the capacity region of the degraded broadcast channel $P_{V_1,V_2|U}$
- $\mathcal{R}_X(D_1, D_2)$ – the successive refinement rate region of the source X at distortions (D_1, D_2)

Main result

- $\mathcal{C}$ – the capacity region of the degraded broadcast channel $P_{V_1, V_2|U}$
- $\mathcal{R}_X(D_1, D_2)$ – the successive refinement rate region of the source X at distortions (D_1, D_2)

Theorem 2 *Under the CK constraint, the distortion pair (D_1, D_2) is achievable with bandwidth expansion ratio ρ if and only if*

$$\mathcal{R}_X(D_1, D_2) \cap \rho\mathcal{C} \neq \varnothing$$

Main result

- $\mathcal{C}$ – the capacity region of the degraded broadcast channel $P_{V_1, V_2|U}$
- $\mathcal{R}_X(D_1, D_2)$ – the successive refinement rate region of the source X at distortions (D_1, D_2)

Theorem 2 *Under the CK constraint, the distortion pair (D_1, D_2) is achievable with bandwidth expansion ratio ρ if and only if*

$$\mathcal{R}_X(D_1, D_2) \cap \rho\mathcal{C} \neq \varnothing$$

$\Longrightarrow$ Separation yields optimal distortion pairs.

CK and separation

Does CK imply separation?

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

- Without CK: Single letter codes yield optimal performance.

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

- Without CK: Single letter codes yield optimal performance.
  - The allowed distortions are incurred by the channel noise.

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

- Without CK: Single letter codes yield optimal performance.
  - The allowed distortions are incurred by the channel noise.
  - Separation is suboptimal.

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

- Without CK: Single letter codes yield optimal performance.
  - The allowed distortions are incurred by the channel noise.
  - Separation is suboptimal.
- With CK: Cannot rely on channel noise “to do the work,” since this cannot be reproduced at the sender side.

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

- Without CK: Single letter codes yield optimal performance.
  - The allowed distortions are incurred by the channel noise.
  - Separation is suboptimal.
- With CK: Cannot rely on channel noise “to do the work,” since this cannot be reproduced at the sender side.
  - Distortion must be introduced by the sender.

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

- Without CK: Single letter codes yield optimal performance.
  - The allowed distortions are incurred by the channel noise.
  - Separation is suboptimal.
- With CK: Cannot rely on channel noise “to do the work,” since this cannot be reproduced at the sender side.
  - Distortion must be introduced by the sender.
  - Separation is optimal.

Does CK imply separation?

Lossy transmission of a Gaussian source over the Gaussian BC:

- Without CK: Single letter codes yield optimal performance.
  - The allowed distortions are incurred by the channel noise.
  - Separation is suboptimal.
- With CK: Cannot rely on channel noise “to do the work,” since this cannot be reproduced at the sender side.
  - Distortion must be introduced by the sender.
  - Separation is optimal.

If CK requires the distortion to be introduced by the sender, is separation always optimal under CK?

Does CK imply separation?

Unfortunately, there are situations where separation is suboptimal even under the CK constraint

Does CK imply separation?

Unfortunately, there are situations where separation is suboptimal even under the CK constraint

- Separation is suboptimal in *lossless* transmission of a joint source over the MAC (Cover, El Gamal, and Salehi 1980)

Does CK imply separation?

Unfortunately, there are situations where separation is suboptimal even under the CK constraint

- Separation is suboptimal in *lossless* transmission of a joint source over the MAC (Cover, El Gamal, and Salehi 1980)
- Separation is suboptimal in lossy transmission of state over the state dependent channel, even under CK constraint.

Thank You