Coding for Single- and Multi- User Systems with Constrained and Unconstrained Side Information

Yossef Steinberg

Technion—Israel Institute of Technology

ysteinbe@ee.technion.ac.il

Outline

- Introduction – state-dependent channels with side information
  - Motivation
  - Basic single-user models
- The single-user case
  - Causal vs. non-causal
  - Examples
  - Algorithms
- The Wyner-Ziv problem
- Constrained side information
  - General achievability result
  - Special cases
- Multi User models
  - The state dependent broadcast channel
  - The state dependent multiple access channel

Introduction

Motivation

The most common example of a channel that depends on random parameters (state), is the Gaussian fading channel:

$$Y_i = S_i X_i + V_i \quad i = 1, 2, \dots n$$

where:

- V – Complex, circularly symmetric additive white Gaussian noise (AWGN)
- S – Fading coefficients, $S \sim P_S(\cdot)$
- X – Channel input
- Y – Channel output.

The fading (state) S is independent of the channel noise V. Notation:

$$S = Re^{j\Theta}, \quad \Theta \sim \mathrm{U}[0, 2\pi).$$

Motivation *(cont'd)*

$$Y_i = S_i X_i + V_i \quad i = 1, 2, \ldots n, \quad S = Re^{j\Theta}, \quad \Theta \sim \mathsf{U}[0, 2\pi)$$

Faithfully describes practical channels in wireless communications. Most common assumptions on P_S:

1. Rayleigh:

$$f_R(r) = \frac{2r}{\Omega} e^{-r^2/\Omega}, \quad r \geq 0, \quad \Omega = \mathbb{E}(R^2).$$

   Suitable to describe a channel with large number of scatterers (ionospheric or tropospheric propagation)

2. Nakagami-m:

$$f_R(r) = \frac{2}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m r^{2m-1} e^{-mr^2/\Omega}$$

   Urban radio channels

3. Rice distribution:

$$f_R(r) = \frac{r}{\sigma^2} e^{-(r^2+\bar{r}^2)/2\sigma^2} I_0\left(\frac{r\bar{r}}{\sigma^2}\right)$$

   Line-of-sight communication link, where $\bar{r}$ is average of main link, and σ^2 its variance.

Motivation (cont'd)

A model which has gained much attention in recent years:

$$Y_i = X_i + S_i + V_i$$

Possible applications of this model:

1. Wireless communications – Here S can describe
   (a) interference from an adjacent channel
   (b) a message we send in the current channel to a second user.
2. Watermarking.

Basic single-user models

$$Y_i = S_i X_i + V_i, \quad (1)$$

$$Y_i = X_i + S_i + V_i \quad (2)$$

Various assumptions can be made on who knows what:

- S unknown at the encoder and decoder $C = \max_{P_X} I(X;Y)$
- S known at the decoder (CSIR) $C = \max_{P_X} I(X;YS) = \max_{P_X} I(X;Y|S)$
- S known at the encoder (CSIT).

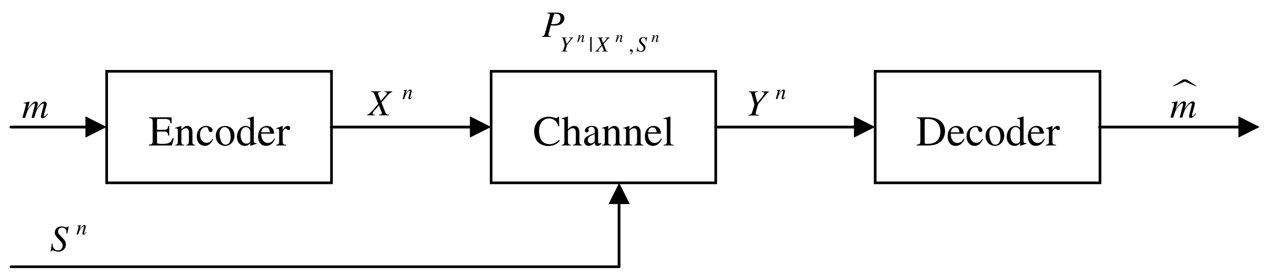
  Here have to specify whether known in causal or non causal manner
- S known at both ends $C = \max_{P_{X|S}} I(X;Y|S)$

  Capacity can be achieved by time-multiplexing of codes, each optimal for a specific realization of S. Time multiplexing according to the probabilities of s. Therefore, capacity is invariant to whether S is known causally or non-causally.

The single-user model

Basic single-user models

General channel model



Assumptions and notation:

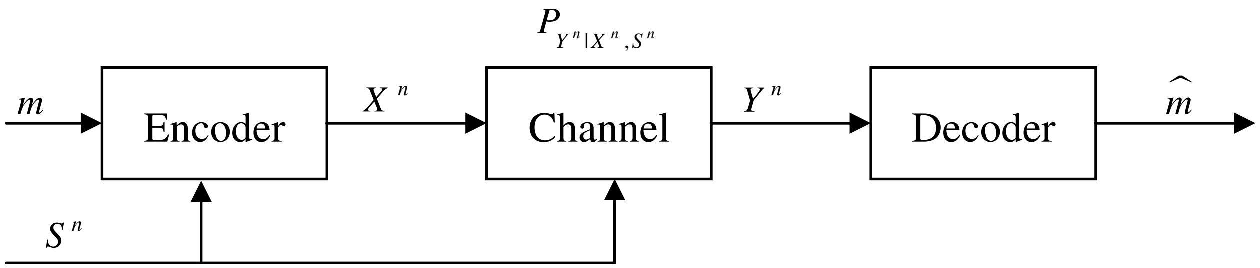
- Finite input, state, and output alphabets: $\mathcal{X}$, $\mathcal{S}$, $\mathcal{Y}$
- Memoryless, time-invariant channel and state

$$P_{Y^n|X^n,S^n}(y^n|x^n,s^n) = \prod_{i=1}^{n} P_{Y|X,S}(y_i|x_i,s_i)$$

$$P_{S^n}(s^n) = \prod_{i=1}^{n} P_S(s_i)$$

The channel is defined by the pair $\{P_S, P_{Y|X,S}\}$.

Causal SI at the encoder



Definition: An (n, M, ϵ) code for state-dependent channel with causal side information at the encoder consists of a causal encoder map

$$f_i : \{1, 2, \ldots, M\} \times \mathcal{S}^i \rightarrow \mathcal{X}, \ \ i = 1, 2, \ldots, n$$

and a decoder map

$$g : \mathcal{Y}^n \rightarrow \{1, 2, \ldots, M\}$$

such that

$$P_e \triangleq \frac{1}{M} \sum_{m=1}^{M} \sum_{s^n} P_{S^n}(s^n) P_{Y^n|X^n,S^n}([g^{-1}(m)]^c | \boldsymbol{f}(m, s^n), s^n) \leq \epsilon$$

$\boldsymbol{f}(m, s^n) = (f_1(m, s_1), f_2(m, s^2), \ldots, f_n(m, s^n))$ is the channel input. The *rate* of the code is

$$R = \frac{1}{n} \log M.$$

Causal SI at the encoder (cont'd)

A rate R is said to be *achievable* if for any $\epsilon > 0$ and sufficiently large n, there exists an $(n, 2^{nR}, \epsilon)$ code for the channel $\{P_S, P_{Y|X,S}\}$, with causal side information.

The *capacity* C of the channel is the supremum of all achievable rates.

Causal SI at the encoder (cont'd)

The causal case was solved by Shannon in 1958. Introduced the concept of *strategies*, i.e., deterministic mappings from $\mathcal{S}$ to $\mathcal{X}$.

Causal SI at the encoder (cont'd)

The causal case was solved by Shannon in 1958. Introduced the concept of *strategies*, i.e., deterministic mappings from $\mathcal{S}$ to $\mathcal{X}$.

- Any deterministic map $t : \mathcal{S} \to \mathcal{X}$ induces a distribution on $\mathcal{Y}$ via the simple relation:

$$P_{Y|T}(y|t) = \sum_s P_S(s) P_{Y|X,S}(y|t(s), s).$$

Causal SI at the encoder (cont'd)

The causal case was solved by Shannon in 1958. Introduced the concept of *strategies*, i.e., deterministic mappings from $\mathcal{S}$ to $\mathcal{X}$.

- Any deterministic map $t : \mathcal{S} \to \mathcal{X}$ induces a distribution on $\mathcal{Y}$ via the simple relation:

$$P_{Y|T}(y|t) = \sum_s P_S(s) P_{Y|X,S}(y|t(s), s).$$

- Thus we can define a memoryless channel $P_{Y|T}$, whose input alphabet is the space $\mathcal{T}$ of all mappings $t : \mathcal{S} \to \mathcal{X}$, and output alphabet is $\mathcal{Y}$.

Causal SI at the encoder *(cont'd)*

The causal case was solved by Shannon in 1958. Introduced the concept of *strategies*, i.e., deterministic mappings from $\mathcal{S}$ to $\mathcal{X}$.

- Any deterministic map $t : \mathcal{S} \to \mathcal{X}$ induces a distribution on $\mathcal{Y}$ via the simple relation:

$$P_{Y|T}(y|t) = \sum_s P_S(s) P_{Y|X,S}(y|t(s), s).$$

- Thus we can define a memoryless channel $P_{Y|T}$, whose input alphabet is the space $\mathcal{T}$ of all mappings $t : \mathcal{S} \to \mathcal{X}$, and output alphabet is $\mathcal{Y}$.

- The channel $P_{Y|T}$ is "regular", i.e., no SI.

Causal SI at the encoder (cont'd)

The causal case was solved by Shannon in 1958. Introduced the concept of *strategies*, i.e., deterministic mappings from $\mathcal{S}$ to $\mathcal{X}$.

- Any deterministic map $t : \mathcal{S} \to \mathcal{X}$ induces a distribution on $\mathcal{Y}$ via the simple relation:
$$P_{Y|T}(y|t) = \sum_{s} P_S(s) P_{Y|X,S}(y|t(s), s).$$
- Thus we can define a memoryless channel $P_{Y|T}$, whose input alphabet is the space $\mathcal{T}$ of all mappings $t : \mathcal{S} \to \mathcal{X}$, and output alphabet is $\mathcal{Y}$.
- The channel $P_{Y|T}$ is "regular", i.e., no SI.

Shannon showed that $P_{Y|X,S}$ with causal SI is *equivalent* to $P_{Y|T}$.

Causal SI at the encoder *(cont'd)*

The causal case was solved by Shannon in 1958. Introduced the concept of *strategies*, i.e., deterministic mappings from $\mathcal{S}$ to $\mathcal{X}$.

- Any deterministic map $t : \mathcal{S} \to \mathcal{X}$ induces a distribution on $\mathcal{Y}$ via the simple relation:

$$P_{Y|T}(y|t) = \sum_s P_S(s) P_{Y|X,S}(y|t(s), s).$$

- Thus we can define a memoryless channel $P_{Y|T}$, whose input alphabet is the space $\mathcal{T}$ of all mappings $t : \mathcal{S} \to \mathcal{X}$, and output alphabet is $\mathcal{Y}$.

- The channel $P_{Y|T}$ is "regular", i.e., no SI.

Shannon showed that $P_{Y|X,S}$ with causal SI is *equivalent* to $P_{Y|T}$.

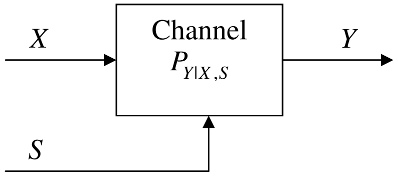
Theorem 1 *[Shannon 1958] The capacity of $P_{Y|X,S}$ with causal side information S at the encoder is given by*

$$C = \max_{P_T} I(T;Y)$$

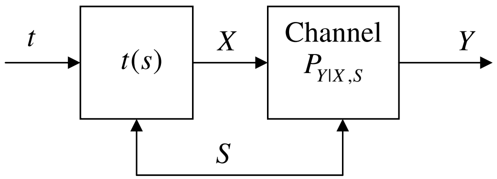
P_T is a distribution on $\mathcal{T}$. The random strategies T are drawn according to P_T, independently of S.

Causal SI at the encoder (cont'd)

Original channel:



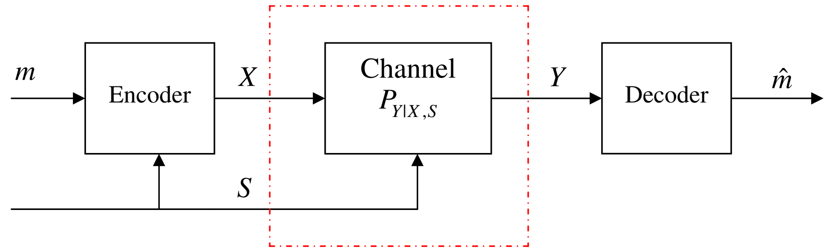
Under causal coding ($x_i = f_i(m, s^i)$), equivalent to:



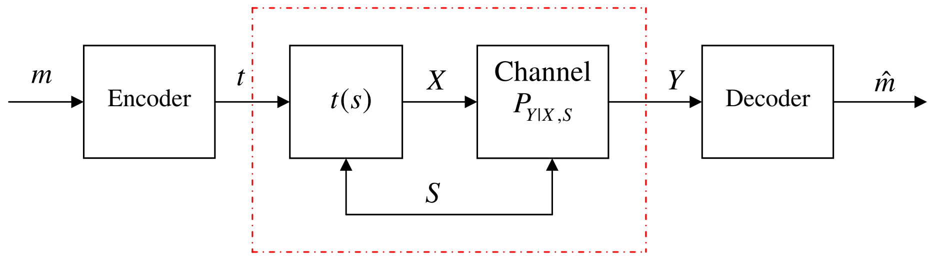
New input alphabet size - $|\mathcal{T}| = \mathcal{X}^{\mathcal{S}}$.

Causal SI at the encoder (cont'd)

Original channel:



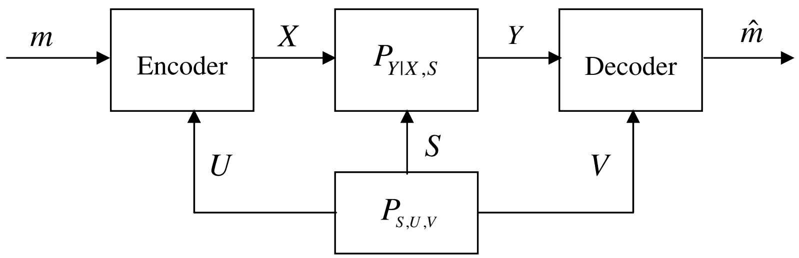
Under causal coding ($x_i = f_i(m, s^i)$), equivalent to:



New input alphabet size - $|\mathcal{T}| = \mathcal{X}^{\mathcal{S}}$.

Causal SI - a simple generalization

- Let the triple (S, U, V) be iid, $\sim P_{S,U,V}$.
- U causal SI at the encoder. V SI at the decoder.



Here the capacity is:

$$C = \max_{P_T} I(T; Y, V) = \max_{P_T} I(T; Y|V)$$

where $t : \mathcal{U} \to \mathcal{X}$.

Examples

Evaluation of capacity for specific models turns out to be a difficult problem, since it involves maximization w.r.t. distributions over the space of strategies $\mathcal{T}$.

Only examples known today:

1. The state sequence known at the encoder is a subset of the channel output [Caire and Shamai, 1999]:

   $$U_i = \psi(V_i), \quad \text{for some deterministic } \psi.$$

   In this case, the capacity is expressed as

   $$C = \max_{P_{X|U}} I(X;Y|U,V).$$

   I.e. - no need to work with strategies. A reminiscent of our remark at introduction.
2. Discrete memoryless state dependent modulo-additive channels [Erez and Zamir, 2000]

Examples *(cont'd)*

The discrete memoryless state-dependent modulo-additive channels [Erez and Zamir, 2000]

$$Y = X \oplus Z_S$$

where

- Additive noise Z_s is distributed according to $P_{Z|S}(\cdot|s)$, and $\oplus$ is the modulo-addition operation.
- Random state S is distributed according to P_S, and is known causally at the encoder.
- No input constraint.

The capacity of this channel with causal knowledge of S at the encoder is given by

$$C = \log|\mathcal{X}| - \min_{t:s\to\mathcal{X}} H(Z_S \ominus t(S))$$

where $\ominus$ stands for the modulo subtraction operation.

Examples (cont'd)

$$Y = X \oplus Z_S, \quad Z_s \sim P_{Z|S}(\cdot|s)$$

The capacity of this channel

$$C = \log|\mathcal{X}| - \min_{t:s\to\mathcal{X}} H(Z_S \ominus t(S)).$$

Interpretation:

- The strategies t serve as noise (Z) predictors
- The optimal (capacity achieving) strategy is the one that minimizes the *entropy* of the noise prediction error. Note that in general, this does not coincide with minimal probability of error predictor.
- Structure of optimal code: A code for *regular* (i.e., no state) modulo-additive channel with additive noise $\tilde{Z}$

$$P_{\tilde{Z}}(\tilde{z}) = \sum_s P_S(s) P_{Z|S}(\tilde{z} \oplus t(s))$$

followed by the noise predictor t.

Note: the code is state-independent.

Examples *(cont'd)*

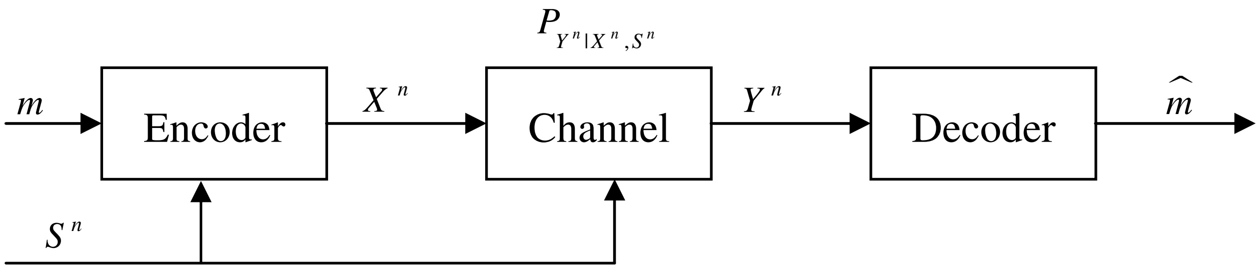
$$Y = X \oplus Z_S, \quad Z_s \sim P_{Z|S}(\cdot|s)$$

$$C = \log|\mathcal{X}| - \min_{t:s\to\mathcal{X}} H(Z_S \ominus t(S)).$$

- In spite of the simplicity of the capacity formula, still have to find optimal predictor t that minimizes error entropy. It can be computed for few special cases (e.g., BSC).

- Prediction with minimal error entropy was introduced by Elias in the context of predictive coding [“Predictive Coding," Elias, 1955].

Noncausal SI



Definition: An (n, M, ϵ) code for state-dependent channel with causal side information at the encoder consists of an encoder map

$$f : \{1, 2, \ldots, M\} \times \mathcal{S}^n \to \mathcal{X}^n$$

and a decoder map

$$g : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$$

such that

$$P_e \triangleq \frac{1}{M} \sum_{m=1}^{M} \sum_{s^n} P_{Y^n|X^n,S^n}([g^{-1}(m)]^c | f(m, s^n), s^n) \leq \epsilon$$

The *rate* of the code is

$$R = \frac{1}{n} \log M.$$

Noncausal SI (cont'd)

Previous work (a very partial list):

- First introduced by Kusnetsov & Tsybakov, "Coding in a memory with defective cells," *PPI* 1974. Coding when the locations of defective cells are known a priori.

  Initiated a series of works [Tsybakov, *PPI* 1975], [Kusnetsov, Kasami, & Tamamura, *IEEE IT* 1978], and more, that dealt with construction of codes for memories with defective cells.

- Problem solved in full generality by Gel'fand & Pinsker, "Coding for channel with random parameters," *Probl. Inform. & Control*, 1980.

- Further relevant contributions by Heegard & El Gamal, "On the capacity of computer memory with defects," *IEEE IT* 1983.

Noncausal SI (cont'd)

The main result by Gel'fand & Pinsker:

Theorem 2 *The capacity of discrete memoryless state-dependent channel, with states known non-causally at the encoder, is given by*

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

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

$$U \multimap (X,S) \multimap Y$$

Noncausal SI *(cont'd)*

The main result by Gel'fand & Pinsker:

Theorem 2 *The capacity of discrete memoryless state-dependent channel, with states known non-causally at the encoder, is given by*

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

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

$$U \multimap (X,S) \multimap Y$$

Define the rate $R(P_{U,X\,|\,S}) = I(U;Y) - I(U;S)$.

Noncausal SI *(cont'd)*

The main result by Gel'fand & Pinsker:

Theorem 2 *The capacity of discrete memoryless state-dependent channel, with states known non-causally at the encoder, is given by*

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

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

$$U \multimap (X,S) \multimap Y$$

Define the rate $R(P_{U,X\,|\,S}) = I(U;Y) - I(U;S)$.

Proposition 1 *[GP 1980]*

1. $R(P_{U,X\,|\,S})$ *is a convex* $\cup$ *function of* $P_{X\,|\,U,S}$*, for fixed* $P_{U\,|\,S}$
2. $R(P_{U,X\,|\,S})$ *is a concave* $\cap$ *function of* $P_{U\,|\,S}$*, for fixed* $P_{X\,|\,U,S}$.

Noncausal SI *(cont'd)*

The main result by Gel'fand & Pinsker:

Theorem 2 *The capacity of discrete memoryless state-dependent channel, with states known non-causally at the encoder, is given by*

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

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

$$U \multimap (X,S) \multimap Y$$

Define the rate $R(P_{U,X\,|\,S}) = I(U;Y) - I(U;S)$.

Proposition 1 *[GP 1980]*

1. $R(P_{U,X\,|\,S})$ is a convex $\cup$ function of $P_{X\,|\,U,S}$, for fixed $P_{U\,|\,S}$

2. $R(P_{U,X\,|\,S})$ is a concave $\cap$ function of $P_{U\,|\,S}$, for fixed $P_{X\,|\,U,S}$.

Thus, the optimal $P_{X\,|\,U,S}$ is a zero-one law.

$\Rightarrow$ To exhaust C, it is enough to restrict X to be a deterministic function of (U,S), $X = f(U,S)$.

Noncausal SI (cont'd)

Proof overview

Direct part is based on *binning*

Noncausal SI (cont'd)

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X\,|\,S}$ with $X = f(U, S)$, and set $R = I(U; Y) - I(U; S) - 2\delta$.

Noncausal SI *(cont'd)*

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.

Noncausal SI (cont'd)

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.

Noncausal SI *(cont'd)*

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.

Noncausal SI (cont'd)

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.

Noncausal SI *(cont'd)*

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.
- The encoder has at hand m and S^n.
  It looks in bin m for $\boldsymbol{u}(m, S^n)$, the first vector $\boldsymbol{u}$ that is jointly typical ($P_{U,S}$) with S^n.
  Since the size of each bin is $2^{n(I(U;S)+\delta)}$, such a vector is likely to exist. (See rate-distortion theory.)

Noncausal SI (cont'd)

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.
- The encoder has at hand m and S^n.
  It looks in bin m for $\boldsymbol{u}(m, S^n)$, the first vector $\boldsymbol{u}$ that is jointly typical ($P_{U,S}$) with S^n.
  Since the size of each bin is $2^{n(I(U;S)+\delta)}$, such a vector is likely to exist. (See rate-distortion theory.)

Noncausal SI *(cont'd)*

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X\,|\,S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.
- The encoder has at hand m and S^n.
  It looks in bin m for $\boldsymbol{u}(m, S^n)$, the first vector $\boldsymbol{u}$ that is jointly typical ($P_{U,S}$) with S^n. Here non-causality comes into account
  Since the size of each bin is $2^{n(I(U;S)+\delta)}$, such a vector is likely to exist. (See rate-distortion theory.)

Noncausal SI *(cont'd)*

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.
- The encoder has at hand m and S^n.
  It looks in bin m for $\boldsymbol{u}(m, S^n)$, the first vector $\boldsymbol{u}$ that is jointly typical ($P_{U,S}$) with S^n. Here non-causality comes into account
  Since the size of each bin is $2^{n(I(U;S)+\delta)}$, such a vector is likely to exist. (See rate-distortion theory.)

Noncausal SI (cont'd)

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.
- The encoder has at hand m and S^n.
  It looks in bin m for $\boldsymbol{u}(m, S^n)$, the first vector $\boldsymbol{u}$ that is jointly typical ($P_{U,S}$) with S^n. Here non-causality comes into account
  Since the size of each bin is $2^{n(I(U;S)+\delta)}$, such a vector is likely to exist. (See rate-distortion theory.)
- The word sent via the channel is $\boldsymbol{x}$, where

$$x_i = f((\boldsymbol{u}(m, s^n))_i, S_i).$$

Noncausal SI *(cont'd)*

Proof overview

Direct part is based on *binning*

- Fix $P_{U,X|S}$ with $X = f(U,S)$, and set $R = I(U;Y) - I(U;S) - 2\delta$.
- Generate $2^{n(I(U;Y)-\delta)}$ words, iid, according to P_U.
  Distribute them among 2^{nR} bins, each associated with one message.
  Thus each bin contains $2^{n(I(U;S)+\delta)}$ codewords $\boldsymbol{u}$.
  Reveal the codewords and bins to the decoder.
- The encoder has at hand m and S^n.
  It looks in bin m for $\boldsymbol{u}(m, S^n)$, the first vector $\boldsymbol{u}$ that is jointly typical ($P_{U,S}$) with S^n. Here non-causality comes into account
  Since the size of each bin is $2^{n(I(U;S)+\delta)}$, such a vector is likely to exist. (See rate-distortion theory.)
- The word sent via the channel is $\boldsymbol{x}$, where

$$x_i = f((\boldsymbol{u}(m, s^n))_i, S_i).$$

Note: with high probability, the resulting triplet $(\boldsymbol{u}(m, S^n), \boldsymbol{x}, S^n)$ is jointly typical.

Noncausal SI (cont'd)

Proof overview (cont'd)

- Since $(\boldsymbol{u}(m, S^n), \boldsymbol{x}, S^n)$ is jointly typical, and the total number of $\boldsymbol{u}$ vectors is $2^{n(I(U;Y)-\delta)}$, the decoder can decode $\boldsymbol{u}(m, S^n)$. Say $\hat{\boldsymbol{u}}$.
- The decoder declares $\hat{m}$ to be the bin number in which $\hat{\boldsymbol{u}}$ resides.

Noncausal SI *(cont'd)*

Proof overview (cont'd)

For the converse, start with Fano inequality, and proceed either as in [GP 1980], or [Csiszár & Körner, *IEEE IT* May 1978], utilizing a decomposition lemma by Csiszár.

$$\begin{aligned} nR - n\epsilon_n &\leq I(m;Y^n) = I(m;Y^n) - I(m;S^n) \\ &= \sum_{i=1}^{n} I(mY^{i-1}S_{i+1}^n;Y_i|Y^{i-1}) - I(mY^{i-1}S_{i+1}^n;S_i|S_{i+1}^n) \\ &\leq \sum_{i=1}^{n} I(mY^{i-1}S_{i+1}^n;Y_i) - I(mY^{i-1}S_{i+1}^n;S_i) \end{aligned}$$

Define

$$U_i = mY^{i-1}S_{i+1}^n$$

So we have

$$R - \epsilon_n \leq \frac{1}{n}\sum_{i=1}^{n} I(U_i;Y_i) - I(U_i;S_i)$$

Now apply standard time-sharing arguments, using the concavity of the functional in the sum.

Noncausal SI (cont'd)

Proof overview (cont'd)

$$R - \epsilon_n \leq \frac{1}{n}\sum_{i=1}^{n} I(U_i;Y_i) - I(U_i;S_i), \qquad U_i = mY^{i-1}S_{i+1}^{n} \tag{3}$$

Time-sharing arguments + concavity:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}[I(U_i;Y_i) - I(U_i;S_i)] \\
&= \quad [I(U_J;Y_J|J) - I(U_J;S_J|J)] \quad J \sim U\{1,2,\ldots,n\} \\
&\leq \quad I(U_J,J;Y) - I(U_J,J;S) \\
&= \quad I(\tilde{U};Y) - I(\tilde{U};S)
\end{aligned}$$

Note: we could arrive to the same result by taking the maximal term in (3). But this will not work in the presence of an input constraint.

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.
  - If U and S are independent of each other ($P_{U,S} = P_U P_S$), any choice of $\boldsymbol{u}$ will do

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.
  - If U and S are independent of each other ($P_{U,S} = P_U P_S$), any choice of $\boldsymbol{u}$ will do $\Rightarrow$ no need to know S^n a priori

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.
  - If U and S are independent of each other ($P_{U,S} = P_U P_S$), any choice of $\boldsymbol{u}$ will do $\Rightarrow$ no need to know S^n a priori
- Proof of converse, GP:

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.
  - If U and S are independent of each other ($P_{U,S} = P_U P_S$), any choice of $\boldsymbol{u}$ will do $\Rightarrow$ no need to know S^n a priori
- Proof of converse, GP:
  - The external random variable $U_i = mY^{i-1}S^n_{i+1}$ depends on S_i via Y^{i-1}.

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.
  - If U and S are independent of each other ($P_{U,S} = P_U P_S$), any choice of $\boldsymbol{u}$ will do $\Rightarrow$ no need to know S^n a priori
- Proof of converse, GP:
  - The external random variable $U_i = mY^{i-1}S_{i+1}^n$ depends on S_i via Y^{i-1}.
  - If the encoder is causal, Y^{i-1} is independent of S_i

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.
  - If U and S are independent of each other ($P_{U,S} = P_U P_S$), any choice of $\boldsymbol{u}$ will do $\Rightarrow$ no need to know S^n a priori
- Proof of converse, GP:
  - The external random variable $U_i = mY^{i-1}S_{i+1}^n$ depends on S_i via Y^{i-1}.
  - If the encoder is causal, Y^{i-1} is independent of S_i $\Rightarrow$ with a causal encoder, U_i is independent of S_i

Causal SI revisited

The GP technique can be used to derive capacity formula for causal SI (Shannon)

- Proof of direct, GP:
  - Non-causality is crucial in the binning step, when looking in bin m for a $\boldsymbol{u}$ that is jointly typical $P_{U,S}$ with S^n.
  - If U and S are independent of each other ($P_{U,S} = P_U P_S$), any choice of $\boldsymbol{u}$ will do $\Rightarrow$ no need to know S^n a priori
- Proof of converse, GP:
  - The external random variable $U_i = mY^{i-1}S_{i+1}^n$ depends on S_i via Y^{i-1}.
  - If the encoder is causal, Y^{i-1} is independent of S_i $\Rightarrow$ with a causal encoder, U_i is independent of S_i

With causal knowledge of S

$$C = \max I(U;Y) - I(U;S)$$

where the max is over $P_{U,X|S}$ such that U is independent of S, and X is a deterministic function of (U, S).

Noncausal vs. causal

A comparison of the different formulas for non-causal and causal SI:

Non causal

Noncausal vs. causal

A comparison of the different formulas for non-causal and causal SI:

Non causal

$$\max_{p(u|s)} I(U;Y) - I(U;S)$$

$$X = f(U,S)$$

Noncausal vs. causal

A comparison of the different formulas for non-causal and causal SI:

Non causal

$$\max_{p(u|s)} I(U;Y) - I(U;S)$$

$$X = f(U,S)$$

Noncausal vs. causal

A comparison of the different formulas for non-causal and causal SI:

Non causal

$$\max_{p(u|s)} I(U;Y) - I(U;S)$$

$$X = f(U,S)$$

Noncausal vs. causal

A comparison of the different formulas for non-causal and causal SI:

Non causal	Causal
$\max_{p(u\|s)} I(U;Y) - I(U;S)$	$\max_{p(u)} I(U;Y)$
$X = f(U,S)$	$X = f(U,S)$
$\max_{p(t\|s)} I(T;Y) - I(T;S)$	$\max_{p(t)} I(T;Y)$
$t : \mathcal{S} \to \mathcal{X}$	$t : \mathcal{S} \to \mathcal{X}$

Noncausal vs. causal

A comparison of the different formulas for non-causal and causal SI:

Non causal	Causal
$\max_{p(u\|s)} I(U;Y) - I(U;S)$	$\max_{p(u)} I(U;Y)$
$X = f(U,S)$	$X = f(U,S)$
$\max_{p(t\|s)} I(T;Y) - I(T;S)$	$\max_{p(t)} I(T;Y)$
$t : \mathcal{S} \to \mathcal{X}$	$t : \mathcal{S} \to \mathcal{X}$

We have shown the direction "$\Longrightarrow$" in the first line

Noncausal vs. causal

A comparison of the different formulas for non-causal and causal SI:

Non causal	Causal
$\max_{p(u\|s)} I(U;Y) - I(U;S)$ $X = f(U,S)$	$\max_{p(u)} I(U;Y)$ $X = f(U,S)$
$\max_{p(t\|s)} I(T;Y) - I(T;S)$ $t : \mathcal{S} \to \mathcal{X}$	$\max_{p(t)} I(T;Y)$ $t : \mathcal{S} \to \mathcal{X}$

We have shown the direction "$\Longrightarrow$" in the first line

- Can show "$\Longrightarrow$" in the second line, and "$\Updownarrow$" in both columns. (For $\Updownarrow$, have to use $X = f(U,S)$, a result of the convexity properties of $R(P_{U,X|S})$.)

Noncausal vs. causal

- An *achievability* result for causal setting can be obtained from an achievability result for non-causal setting, by taking all the external random variables to be independent of S.

  This will be expressed in terms of external random variables.

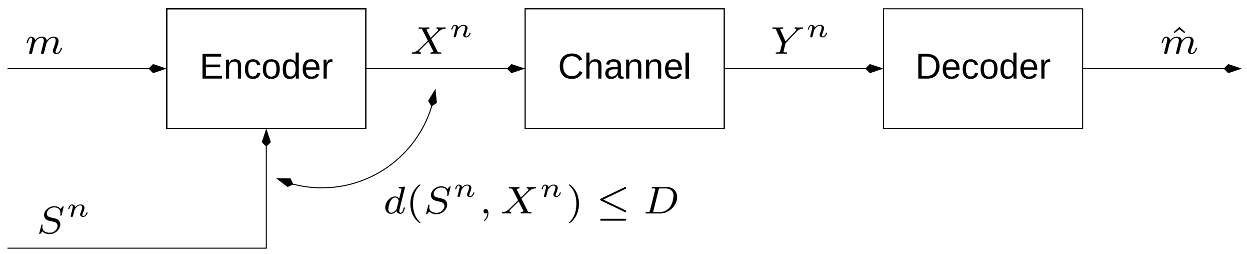
- In case the original achievable region satisfies some convexity properties, can use strategies instead of *part* of the external random variables.

Applications - the non-causal case

Possible applications of the non-causal model:

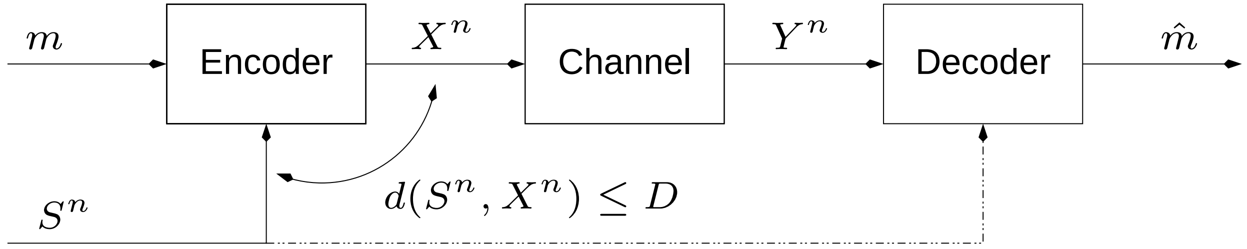
- Communication systems employing FDM, where coding is done across frequencies.
- Watermarking (WM), or Information Embedding (IE).

The Information Embedding (IE) Problem



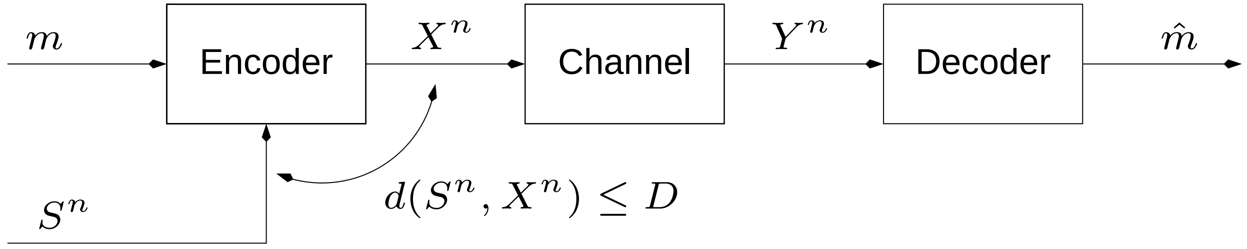
- A message m is embedded into host signal S^n, producing data set X^n
- X^n is transmitted via $P_{Y|X}$ to its destination
- At the destination, a noisy version Y^n of the data set is received, from which m is decoded.
- In IE, m is embedded into S^n in a manner that is transparent to the unintended 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 $P_{Y|X}$ 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.
  That is, hide the fact that messages are embedded into S^n.
  Thus, the host signal S^n is also termed as *covertext*.
- Reduce total distortion at the output

Classical IE puts emphasis on embedding rate (rate of messages m) vs. input distortion D.

Closely related to Gel'fand & Pinsker channel [Moulin & O'Sullivan, 2003], via the constraint. Thus, the *embedding capacity* is given by

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

Examples

Evaluation of the GP capacity is usually hard. There are two canonical models, however, for which capacity can be computed

1. The additive white Gaussian noise (AWGN) channel, with additive known interference, and input power constraint.

2. Binary symmetric channel, with Bernoulli($\frac{1}{2}$) modulo-additive known interference, and input Hamming constraint.

Examples *(cont'd)*

A key example – the AWGN channel with additive interference, studied in M.H.M. Costa, "Writing on dirty paper," *IEEE IT* 1983

$$Y_i = X_i + S_i + V_i$$

where

V_i – Additive white Gaussian noise (AWGN), $V_i \sim \mathcal{N}(0, \sigma_v^2)$

S_i – Additive interference, known non-causally at the encoder, independent of $\{V_i\}_i$, iid, $S_i \sim \mathcal{N}(0, \sigma_s^2)$

X_i – Channel input, subject to power constraint:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P$$

The GP formula applies. Thus

$$C = \max_{P_{X,U|S}} [I(U;Y) - I(U;S)]$$

where the maximization is subject to the constraint

$$\mathbb{E}X^2 \leq P, \quad U \multimap (X,S) \multimap Y$$

and X can be taken to be a deterministic function of (U,S) $(\star)$

Examples *(cont'd)*

$$Y_i = X_i + S_i + V_i$$

A naive approach would be to try to cancel the additive interference S_i (or part of it) at the encoder. With such a strategy, if $P > \sigma_s^2$, we get

$$R = \frac{1}{2}\log\left(1 + \frac{P - \sigma_s^2}{\sigma_v^2}\right)$$

Examples *(cont'd)*

Writing on dirty paper (WDP)

$$Y_i = X_i + S_i + V_i$$

$$C = \max I(U;Y) - I(U;S), \quad \mathbb{E}X^2 \leq P$$

Costa suggested the following substitutions in the GP formula:

$$\begin{aligned} U &= X + \alpha S \\ & X \sim \mathcal{N}(0, P) \text{ independent of } S \end{aligned}$$

Examples *(cont'd)*

Writing on dirty paper (WDP)

$$Y_i = X_i + S_i + V_i$$

$$C = \max I(U;Y) - I(U;S), \quad \mathbb{E}X^2 \le P$$

Costa suggested the following substitutions in the GP formula:

$$U = X + \alpha S$$

$X \sim \mathcal{N}(0, P)$ independent of S

Examples *(cont'd)*

Writing on dirty paper (WDP)

$$Y_i = X_i + S_i + V_i$$

$$C = \max I(U;Y) - I(U;S), \quad \mathbb{E}X^2 \le P$$

Costa suggested the following substitutions in the GP formula:

$$U = X + \alpha S$$

$X \sim \mathcal{N}(0, P)$ independent of S X does not "fight" S

$$\alpha = \frac{P}{P + \sigma_v^2}$$

from which he obtained

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_v^2}\right) \quad (!)$$

That is, in terms of capacity, no penalty is incurred due to the presence of additive interference, provided it is known a priori at the encoder.

(penalty – relative to the case of no interference)

No need to prove that this substitution is optimal

Examples *(cont'd)*

Binary symmetric channel with known interference

$$Y_i = X_i \oplus S_i \oplus V_i$$

where $\mathcal{X} = \mathcal{S} = \mathcal{V} = \{0,1\}$, $\{S_i\}_i$ and $\{V_i\}_i$ are iid and independent of each other, and

$$P_V(1) = p, \quad P_S(1) = \frac{1}{2}, \quad \mathbb{E}(X) = \delta.$$

The capacity of this channel with noncausal knowledge of S at the encoder is given by

$$C = \text{U.C.E}\,\{H(\delta) - H(p)\}$$

where U.C.E stands for upper convex envelope. Here the additive interference does incur a penalty relative to the case of no interference. (Without interference, the capacity is $H(p \star \delta) - H(p)$.)

Examples *(cont'd)*

These examples parallel similar ones in the context of watermarking.

Embedding messages in a Gaussian host S, with quadratic distortion D, over an AWGN channel:

$$Y_i = X_i + V_i$$

$$C = \max I(U;Y) - I(U;S), \quad \mathbb{E}[(X-S)^2] \leq D$$

Make the following substitutions:

$$\begin{aligned} U &= Z + \alpha S, \quad Z \sim \mathcal{N}(0, D) \text{ indep. of } X \\ X &= U + (1-\alpha)S \\ \alpha &= \frac{D}{D + \sigma_v^2} \end{aligned}$$

from which it follows

$$C \geq \frac{1}{2} \log\left(1 + \frac{D}{\sigma_v^2}\right)$$

In fact, equality holds, as otherwise we contradict the converse in Costa's result (note that here $X = Z + S$).

Examples (cont'd)

Embedding messages in a binary $(1/2, 1/2)$ host S, with Hamming distortion δ, over BSC(p)

$$Y_i = X_i \oplus V_i$$

where $\mathcal{X} = \mathcal{S} = \mathcal{V} = \{0, 1\}$, $\{S_i\}_i$ and $\{V_i\}_i$ are iid and independent of each other, and

$$P_V(1) = p, \quad P_S(1) = \frac{1}{2}, \quad \mathbb{E}(X \ominus S) = \delta.$$

The capacity of this IE system is given by

$$C = \text{U.C.E}\,\{H(\delta) - H(p)\}$$

Computational algorithms

Since evaluation of capacity directly via the GP formula is, in general, a prohibitively difficult problem, computational algorithms based on alternate maximization (Arimoto-Blahut like) have been developed.

Computational algorithms

Since evaluation of capacity directly via the GP formula is, in general, a prohibitively difficult problem, computational algorithms based on alternate maximization (Arimoto-Blahut like) have been developed.

Such an algorithm has, in general, the following ingredients

Computational algorithms

Since evaluation of capacity directly via the GP formula is, in general, a prohibitively difficult problem, computational algorithms based on alternate maximization (Arimoto-Blahut like) have been developed.

Such an algorithm has, in general, the following ingredients

1. An alternate maximization (AM) scheme, which converges to the capacity

Computational algorithms

Since evaluation of capacity directly via the GP formula is, in general, a prohibitively difficult problem, computational algorithms based on alternate maximization (Arimoto-Blahut like) have been developed.

Such an algorithm has, in general, the following ingredients

1. An alternate maximization (AM) scheme, which converges to the capacity
2. A "geometric" upper bound: a functional of the variables (distributions) of the AM scheme, which upper bounds the capacity, and coincides with the capacity when the capacity-achieving variables (distributions) are plugged in.

Computational algorithms

Since evaluation of capacity directly via the GP formula is, in general, a prohibitively difficult problem, computational algorithms based on alternate maximization (Arimoto-Blahut like) have been developed.

Such an algorithm has, in general, the following ingredients

1. An alternate maximization (AM) scheme, which converges to the capacity
2. A “geometric” upper bound: a functional of the variables (distributions) of the AM scheme, which upper bounds the capacity, and coincides with the capacity when the capacity-achieving variables (distributions) are plugged in.
3. A stopping rule based on comparing the output of the AM scheme with the geometric upper bound.

Computational algorithms *(cont'd)*

Two schemes were suggested for the GP capacity formula

1. C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE IT* Sep. 1983.

   Here an AM schemes is applied on

   $$I(U;Y) - I(U;S), \quad X = f(U,S).$$

   Each step of the scheme involves exhaustive search over a subset of the functions $f(u,s)$.

2. F. Dupuis, Wei Yu, and F. Willems, "Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side-information," *Proc. of 2004 Int. Symp. Inf. Theory*, Chicago, IL, 2004.

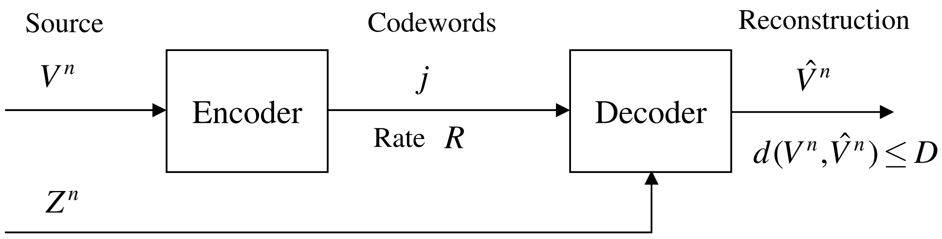
   Here the AM scheme is applied on

   $$I(T;Y) - I(T;S), \quad X = T(S),$$

   where P_T is one of the variables of the AM scheme. Thus exhaustive search is not needed.

The Wyner-Ziv problem

Source coding with side information



V^n – Memoryless source, alphabet $\mathcal{V}$. To be compressed to rate R, and reconstructed at distortion $\leq D$

$\hat{V}^n$ – Reconstruction at the decoder, alphabet $\hat{\mathcal{V}}$

d – distortion measure

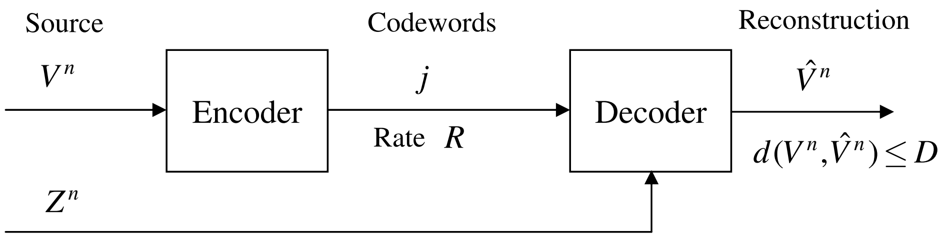
$$d(V^n, \hat{V}^n) = \frac{1}{n} \sum_{i=1}^{n} d(V_i, \hat{V}_i).$$

Z^n – Side information, available only at the decoder.

The source and side information are memoryless, with joint distribution $P_{V,Z}$.

<u>Q:</u> What is the minimal rate R such that it is possible to reconstruct V at distortion level D, with SI Z at the decoder?

Source coding with SI *(cont'd)*



Definition: An (n, M, D) code for V with side information Z is an encoder map f and a decoder map g

$$f : \mathcal{V}^n \to \{1, 2, \ldots, M\}$$
$$g : \{1, 2, \ldots, M\} \times \mathcal{Z}^n \to \hat{\mathcal{V}}^n$$

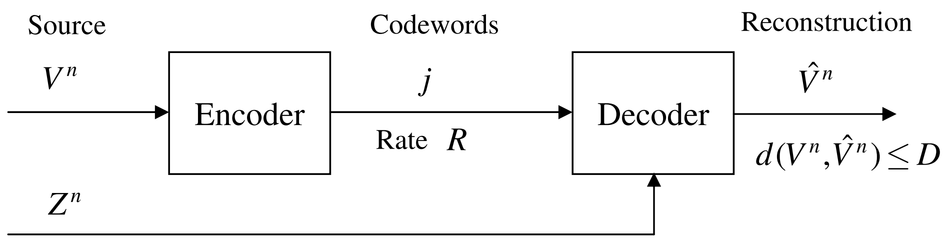
with average distortion not exceeding D

$$\mathbb{E}d(V^n, g(f(V^n), Z^n)) = \frac{1}{n}\sum_{i=1}^{n} d(V_i, g_i(f(V^n), Z^n)) \leq D.$$

The rate of the code is $R = \frac{1}{n}\log M$.

The minimal achievable rate with distortion D and decoder SI Z is denoted by $R(D|Z)$.

Source coding with SI *(cont'd)*



The main result on $R(D|Z)$ (A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoders," *IEEE IT*, Jan. 1976.)

Theorem 3 *For any discrete memoryless source V with decoder side information Z,*

$$R(D|Z) = \min[I(W;V) - I(W;Z)]$$

where the minimum is over all external random variables W such that there exists a deterministic mapping

$$\phi : \mathcal{W} \times \mathcal{Z} \to \hat{\mathcal{V}}$$

satisfying

$$\mathbb{E}d(V, \phi(W, Z)) \leq D,$$

and the Markov chain $W \multimap V \multimap Z$ *holds.*

Source coding with SI *(cont'd)*

$$R(D|Z) = \min_{W,\phi} [I(W;V) - I(W;Z)],$$

$$\mathbb{E}d(V,\phi(W,Z)) \leq D, \quad W \multimap V \multimap Z.$$

- The GP formula resembles the WZ formula.
- Both use binning, where in WZ(76) it is applied in the context of rate-distortion theory, whereas in GP(80) in the context of channel coding.
- Due to the Markov structure, the WZ formula can be written as

$$R(D|Z) = \min_{W,\phi} I(W;V|Z) \quad W \multimap V \multimap Z.$$

  Note that the only difference between this formula and the formula for RD with SI at both ends (encoder and decoder) is the Markov structure. When SI is present at both sides, no Markov structure is imposed.
- The two problems – WZ and GP – are considered dual. Duality is not full, since the WZ formula can be written as single conditional mutual information, but the GP cannot.

Constrained SI in single user systems

How SI is provided?

A typical user in a communication system (represented as “encoder” or “decoder”), seldom has the possibility of measuring the channel state S directly.

How SI is provided?

A typical user in a communication system (represented as “encoder” or “decoder”), seldom has the possibility of measuring the channel state S directly.

- Channel state information at the decoder

How SI is provided?

A typical user in a communication system (represented as “encoder” or “decoder”), seldom has the possibility of measuring the channel state S directly.

- Channel state information at the decoder
  - If estimated from channel output, then it is a system with no SI

How SI is provided?

A typical user in a communication system (represented as “encoder” or “decoder”), seldom has the possibility of measuring the channel state S directly.

- Channel state information at the decoder
  - If estimated from channel output, then it is a system with no SI
  - Otherwise, it is provided by a third party.

How SI is provided?

A typical user in a communication system (represented as “encoder” or “decoder”), seldom has the possibility of measuring the channel state S directly.

- Channel state information at the decoder
  - If estimated from channel output, then it is a system with no SI
  - Otherwise, it is provided by a third party.
- Channel state information at the encoder

How SI is provided?

A typical user in a communication system (represented as “encoder” or “decoder”), seldom has the possibility of measuring the channel state S directly.

- Channel state information at the decoder
  - If estimated from channel output, then it is a system with no SI
  - Otherwise, it is provided by a third party.

- Channel state information at the encoder
  - If provided from the receiver, then the right tool to deal with it is “feedback channels.”

How SI is provided?

A typical user in a communication system (represented as “encoder” or “decoder”), seldom has the possibility of measuring the channel state S directly.

- Channel state information at the decoder
  - If estimated from channel output, then it is a system with no SI
  - Otherwise, it is provided by a third party.
- Channel state information at the encoder
  - If provided from the receiver, then the right tool to deal with it is “feedback channels.”
  - Otherwise, it is provided by a third party.

How SI is provided? *(cont'd)*

Thus, many times, SI is provided by a third party, which is part of the system:

How SI is provided? *(cont'd)*

Thus, many times, SI is provided by a third party, which is part of the system:

- A central station that collects data from other active users around, estimates *on site* channel state, and transmits it to the active users via wayside links.

How SI is provided? (cont'd)

Thus, many times, SI is provided by a third party, which is part of the system:

- A central station that collects data from other active users around, estimates *on site* channel state, and transmits it to the active users via wayside links.
- A "regular" user in a network, who sends information about his messages to neighboring users, in order to reduce potential interference (e.g., transmitter 1 in Natasha's talk).

How SI is provided? *(cont'd)*

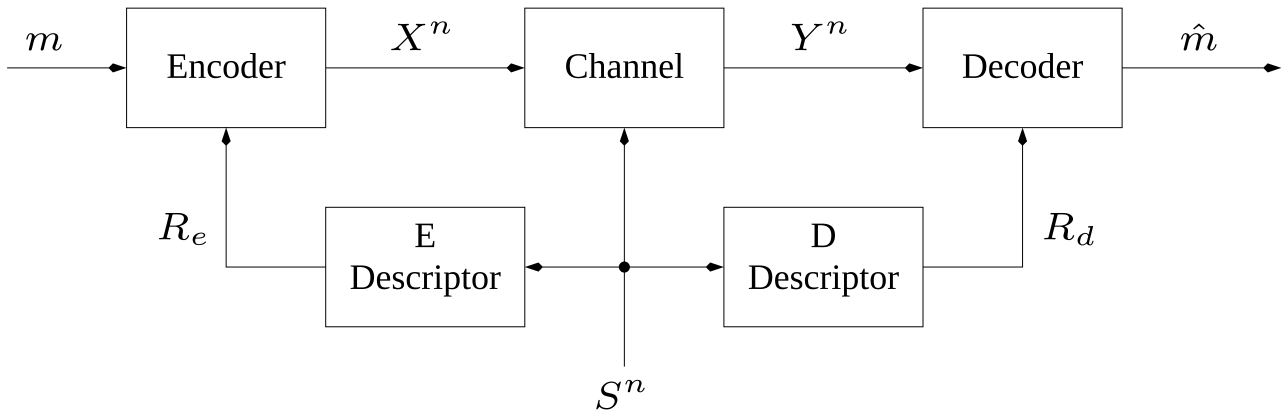
Thus, many times, SI is provided by a third party, which is part of the system:

- A central station that collects data from other active users around, estimates *on site* channel state, and transmits it to the active users via wayside links.
- A "regular" user in a network, who sends information about his messages to neighboring users, in order to reduce potential interference (e.g., transmitter 1 in Natasha's talk).

In such cases, system resources must be allocated in order to provide SI to the transmitters or receivers.

The general model

Heegard & El Gamal, 1983, "On the capacity of computer memory with defects."
Introduced coding for state dependent channels with rate limited side information at both ends. Devised an achievable region.

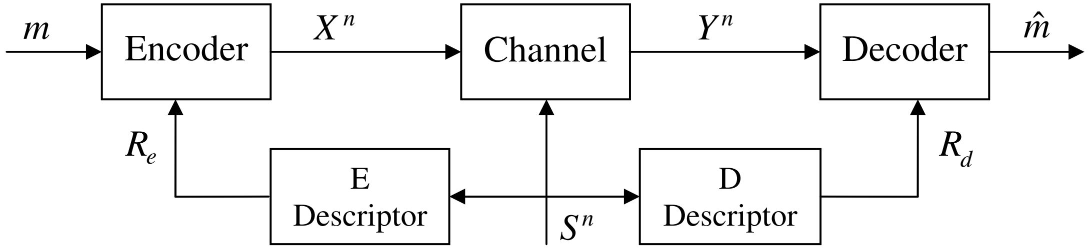


Theorem 4 (Heegard & El Gamal, 1983) *All triples* (R, R_e, R_d) *in the convex hull of the set*

$$\begin{aligned}
R_e &\geq I(S_0, S_e; S) \\
R_d &\geq I(S_0, S_d; S) - I(S_0, S_d; Y) \\
R_d &\geq I(S_d; S|S_0) - I(S_d; Y|S_0) \\
R_e + R_d &\geq I(S_0, S_e, S_d; S) - I(S_0, S_d; Y) + I(S_e; S_d|S_0) \\
R_e + R_d &\geq I(S_e, S_d; S|S_0) - I(S_d; Y|S_0) + I(S_e; S_d|S_0) \\
R &\leq I(U; Y, S_d|S_0) - I(U; S_e|S_0)
\end{aligned}$$

for some $P_{S,S_0,S_e,S_d,U,X} = P_S P_{S_0,S_e,S_d|S} P_{U,X|S_0,S_e}$, *are achievable.*

Few special cases

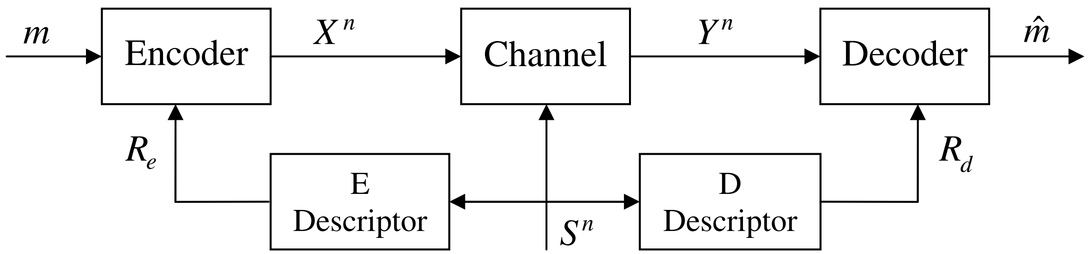


Heegard & El Gamal showed that the region is tight for the cases:

1. $R_e = 0$, $R_d = 0$
2. $R_e = H(S)$, $R_d = H(S|Y)$ (both sides fully informed)
3. $R_e = H(S)$, $R_d = 0$ (the GP model)
4. R_e arbitrary, $R_d = H(S|Y)$ (rate-limited SI @ encoder, fully informed decoder).

Case 4 was treated also by Rosenzweig *et al*, 2005.

Few special cases (cont'd)



Case 4. R_e arbitrary, $R_d = H(S|Y)$ (rate-limited SI @ encoder, fully informed decoder).

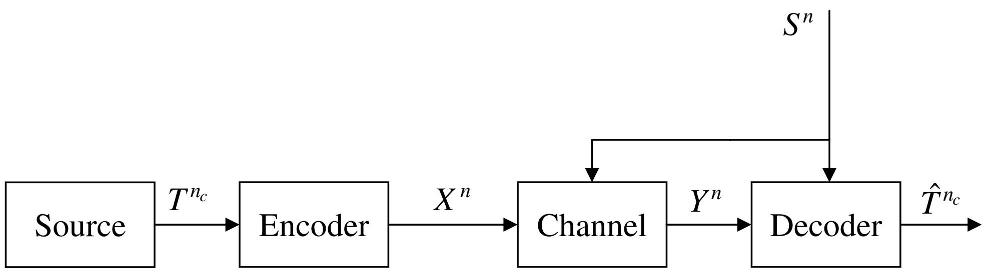
A complete single letter characterization of this region was given in [Rosezweig, Steinberg, Shamai, *IEEE IT*, May 2005]:

Theorem 5 *For any state dependent discrete memoryless channel with full SI at the decoder and rate limited (*R_e*) SI at the encoder,* (R, R_e) *is achievable if and only if*

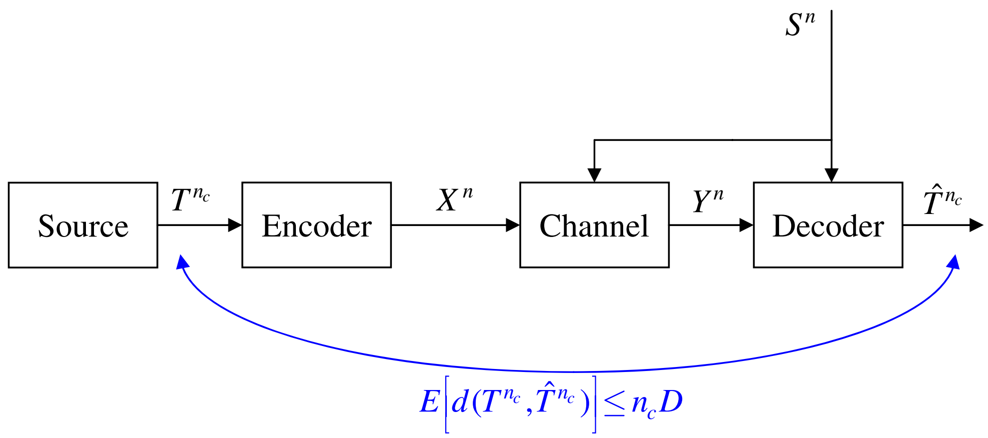
$$\begin{aligned} R &\leq I(X;Y|S,S_e) \\ R_e &\geq I(S;S_e) \end{aligned}$$

for some S_e *such that* $X \multimap S_e \multimap S \quad S_e \multimap (S,X) \multimap Y$

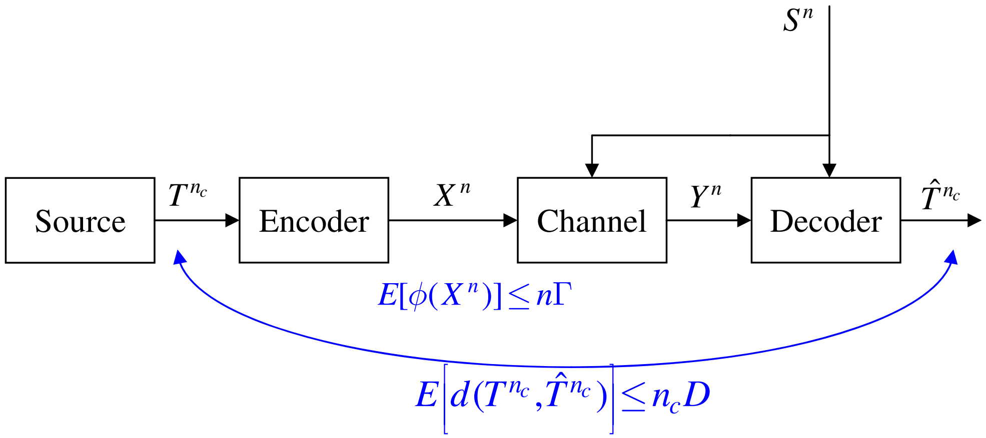
Joint state-source-channel coding



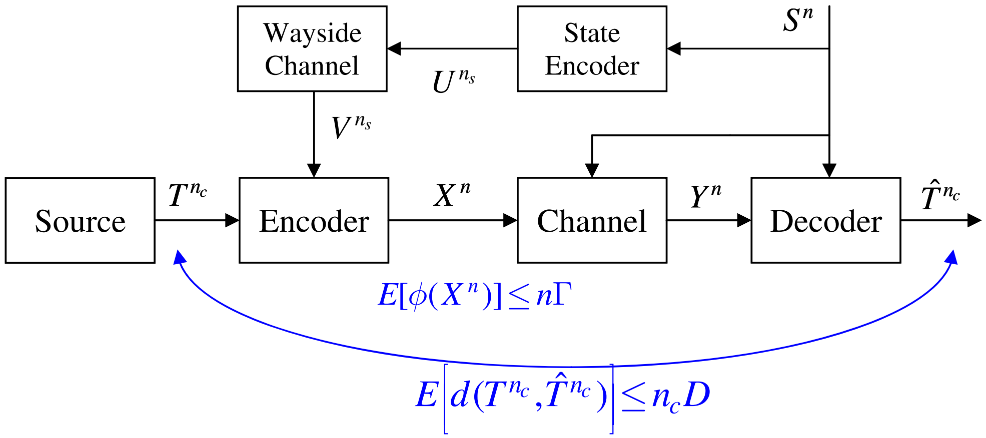
Joint state-source-channel coding



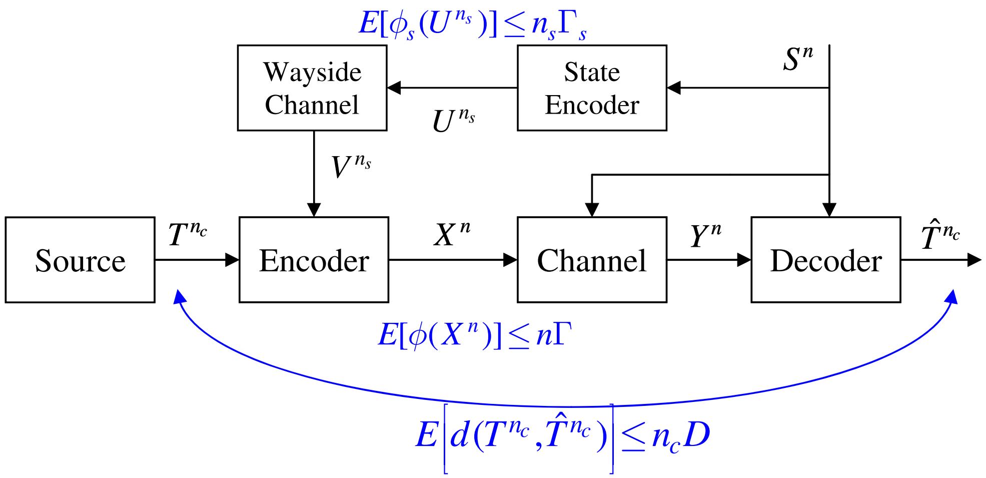
Joint state-source-channel coding



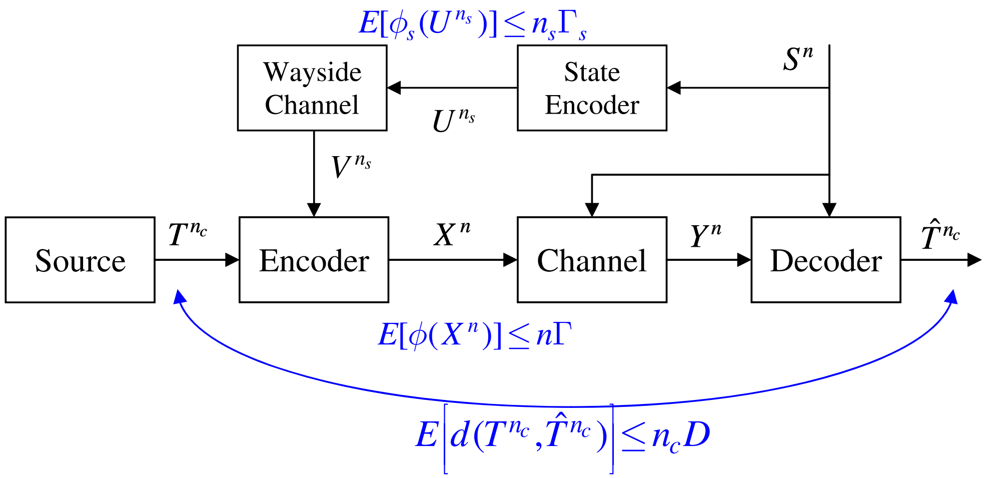
Joint state-source-channel coding



Joint state-source-channel coding

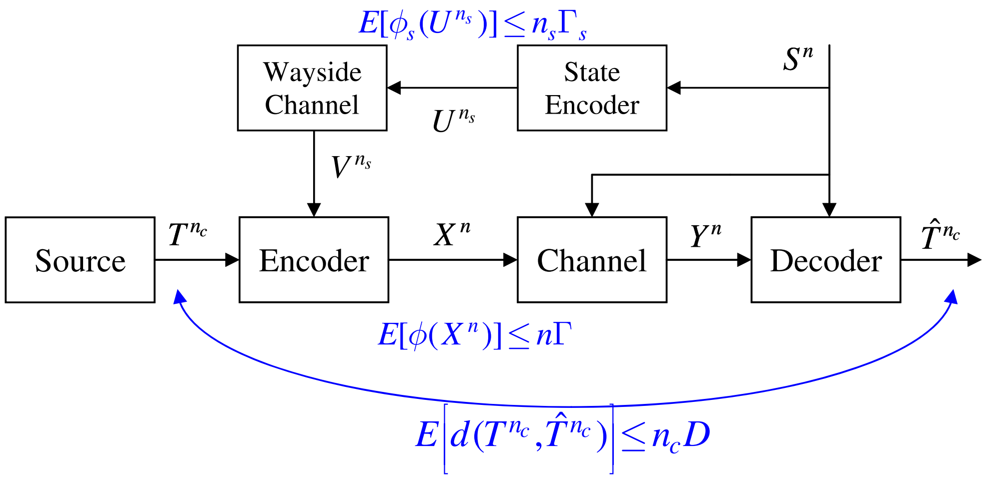


Joint state-source-channel coding



All channels and sources are discrete memoryless: Forward channel $P_{Y|X,S}$, wayside channel $P_{V|U}$, channel state S (P_S), and source T (P_T).

Joint state-source-channel coding



All channels and sources are discrete memoryless: Forward channel $P_{Y|X,S}$, wayside channel $P_{V|U}$, channel state S (P_S), and source T (P_T).

- ρ_c, ρ_s – bandwidth expansion factors: $\rho_c = n_c/n$, and $\rho_s = n_s/n$.
- There is no decoder for S, since we are not interested in reproducing it. We are interested in reducing D (or Γ, Γ_s).

We are interested in the region of all achievable (D, Γ, Γ_s), for given (ρ_c, ρ_s)

Joint state-source-channel coding (cont'd)

Relevant Works

CS available at encoder:

- Gel'fand & Pinsker, 1980 – Capacity of channel with random state, known non-causally at the encoder (GP Channel).
- Merhav & Shamai, 2003 – Joint source-channel coding for Wyner-Ziv source and GP channel. Separation holds.

Rate-limited CSI at encoder and/or decoder:

- Heegard & El Gamal, 1983 – Achievable region for channel with rate limited CSI at encoder and decoder, (R_e, R_d). Tight for some cases.
- Rosenzweig, Steinberg, Shamai, 2005 – Capacity of channel with rate limited CSI at the encoder, full CS at the decoder.

Joint state-source-channel coding (cont'd)

Definition: An $(n, n_c, n_s, D, \Gamma, \Gamma_s)$ code consists of three mappings:

$$\begin{aligned} f_s &: \mathcal{S}^n \to \mathcal{U}^{n_s} \\ f &: \mathcal{T}^{n_c} \times \mathcal{V}^{n_s} \to \mathcal{X}^n \\ g &: \mathcal{Y}^n \times \mathcal{S}^n \to \hat{\mathcal{T}}^{n_c} \end{aligned}$$

such that

$$\mathbb{E}[\phi(f(T^{n_c}, V^{n_s}))] \leq n\Gamma, \quad \mathbb{E}[\phi_s(f_s(S^n))] \leq n_s\Gamma_s$$

$$\mathbb{E}[d(T^{n_c}, g(Y^n, S^n))] \leq n_c D.$$

The distortion-cost triple (D, Γ, Γ_s) is *achievable* with (ρ_c, ρ_s) if $\forall \epsilon > 0$ and s.l.n there exists an $(n, \rho_c n, \rho_s n, D + \epsilon, \Gamma, \Gamma_s)$ code for $(P_T, P_S, P_{Y|XS})$.

Joint state-source-channel coding (cont'd)

Theorem 6 (Cemal & Steinberg, ISIT 2005) *The distortion-cost triplet* (D, Γ, Γ_s) *is achievable with bandwidth expansion factors* (ρ_c, ρ_s) *iff*

$$\exists S_0 : \quad S \multimap S_0 \multimap X, \quad S_0 \multimap (X, S) \multimap Y$$

such that

$$\begin{aligned} \rho_c R_T(D) &\leq I(X; Y|S, S_0) \\ I(S; S_0) &\leq \rho_s C_g(\Gamma_s) \\ \mathbb{E}[\phi(X)] &\leq \Gamma \end{aligned}$$

where

$R_T(D)$ – *rate-distortion function of the source* P_T

$C_g(\Gamma_s)$ – *capacity of the genie channel* $P_{V|U}$ *with input constraint* Γ_s.

Joint state-source-channel coding (cont'd)

(D, Γ, Γ_s) achievable with (ρ_c, ρ_s) iff

$$\exists S_0 : \quad S \multimap S_0 \multimap X, \quad S_0 \multimap (X,S) \multimap Y$$

such that

$$\begin{aligned} \rho_c R_T(D) &\leq I(X;Y|S,S_0) \\ I(S;S_0) &\leq \rho_s C_g(\Gamma_s) \\ \mathbb{E}[\phi(X)] &\leq \Gamma \end{aligned}$$

- Separation holds for:
  - Coding the source P_T independently of the channels $P_{Y|X,S}$, $P_{V|U}$, and state S.
  - Coding the state S independently of P_T, $P_{V|U}$
- The code of the state S (S_0) does depend on the forward channel $P_{Y|X,S}$.

Joint state-source-channel coding (cont'd)

The same holds for transmission of messages instead of source T: (R, Γ, Γ_s) achievable with ρ_s iff

$$\exists S_0 : \quad S \multimap S_0 \multimap X, \quad S_0 \multimap (X,S) \multimap Y$$

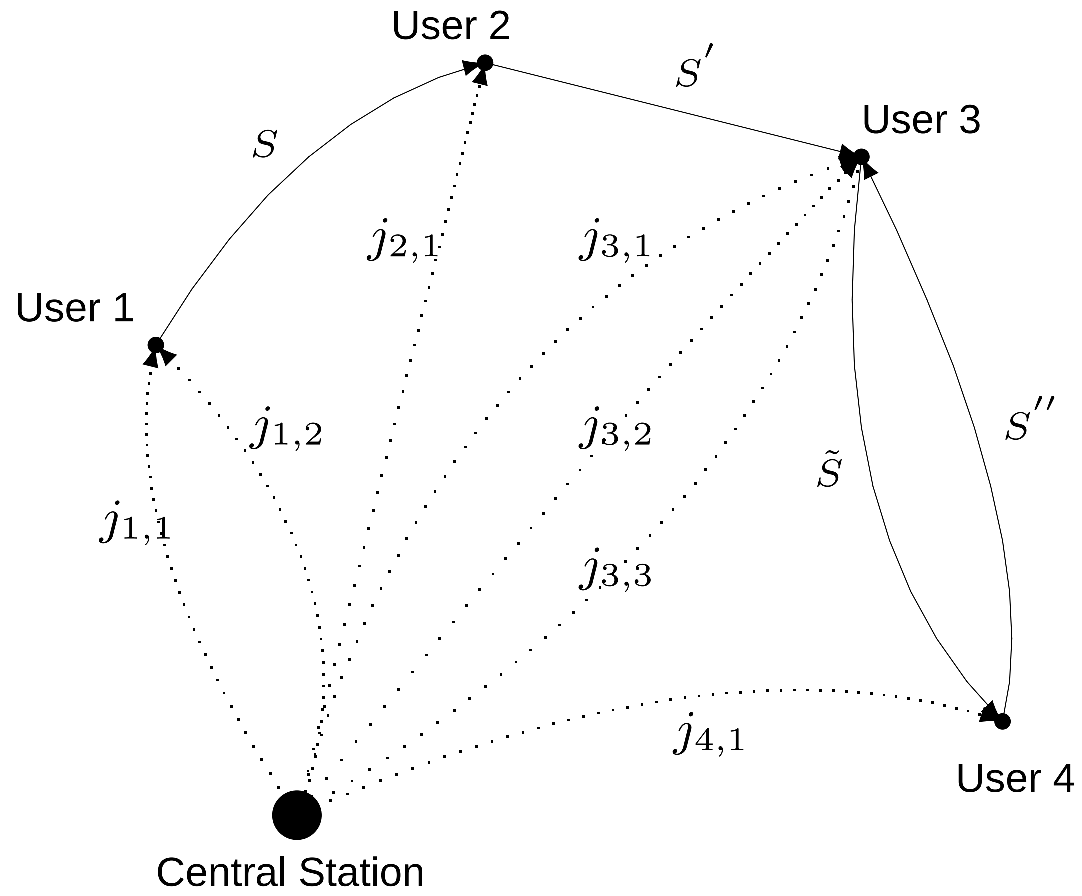
such that

$$\begin{aligned} R &\leq I(X;Y|S,S_0) \\ I(S;S_0) &\leq \rho_s C_g(\Gamma_s) \\ \mathbb{E}[\phi(X)] &\leq \Gamma \end{aligned}$$

- Separation holds for:
  - Coding the state S independently of the wayside channel $P_{V|U}$
- The code of the state S (S_0) does depend on the forward channel $P_{Y|X,S}$.

Coding with multiple descriptions of CS

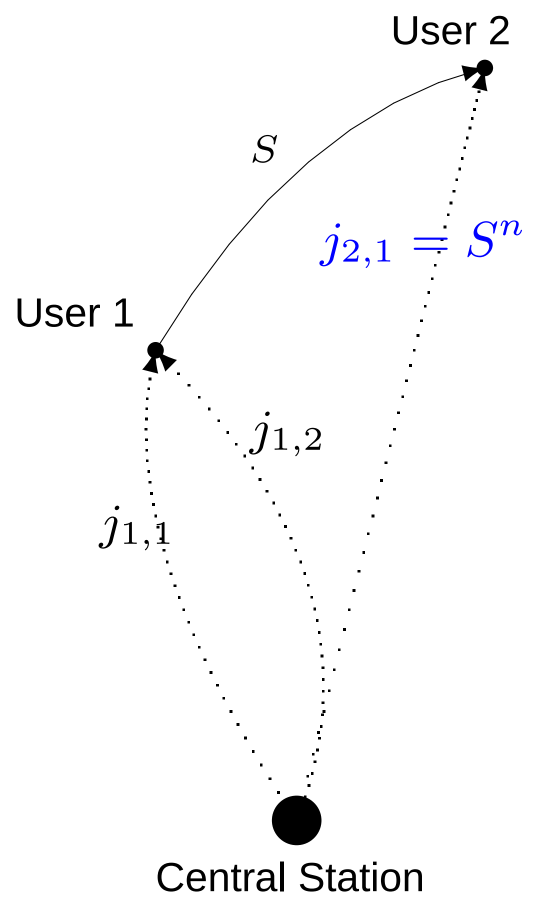
A Network Scenario:



The central station sends information on the forward channels states (S, S' etc) to the users.

Coding with multi-csi *(cont'd)*

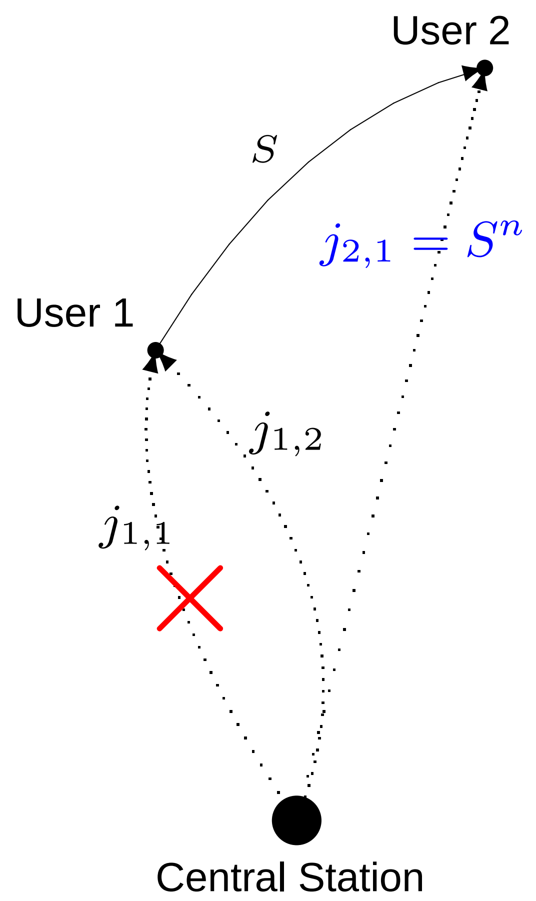
A Network Scenario:



The central station sends information on the forward channels states (S, S' etc) to the users.

Coding with multi-csi (cont'd)

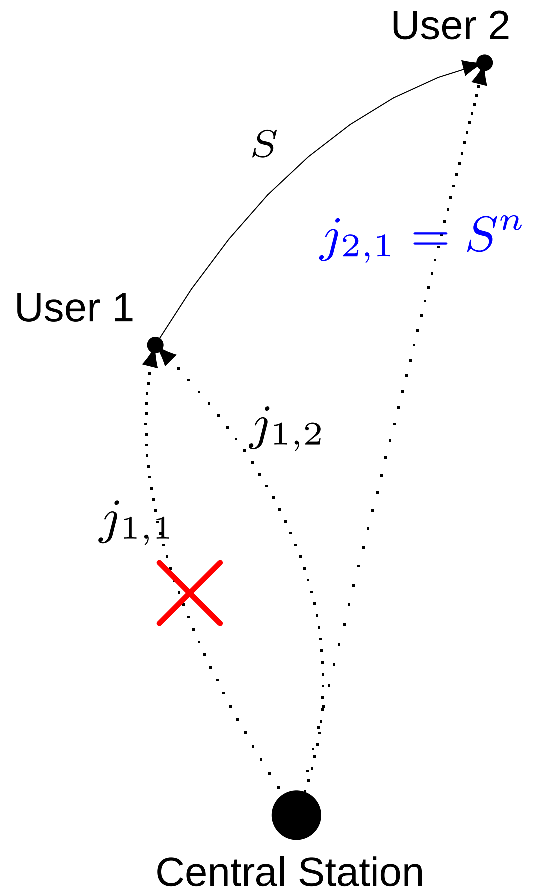
A Network Scenario:



The central station sends information on the forward channels states (S, S' etc) to the users.

Coding with multi-csi (cont'd)

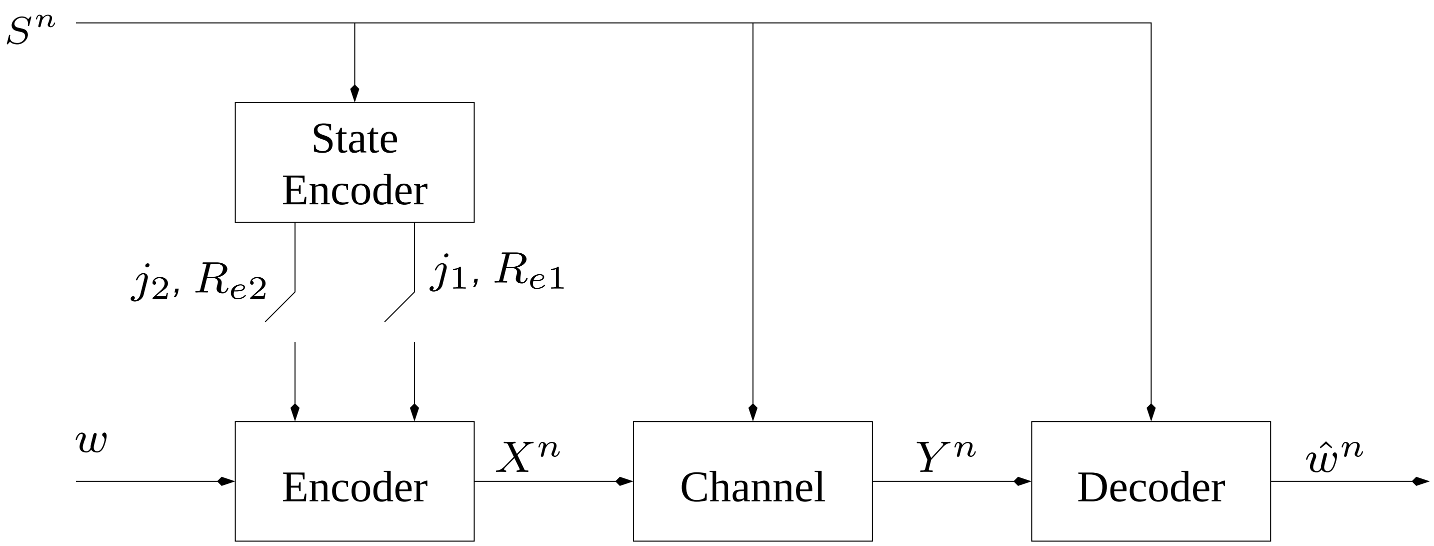
A Network Scenario:



Goal: Design a scheme as robust as possible to failure of wayside links $\Longrightarrow$ Multiple Descriptions.

Coding with multi-csi (cont'd)

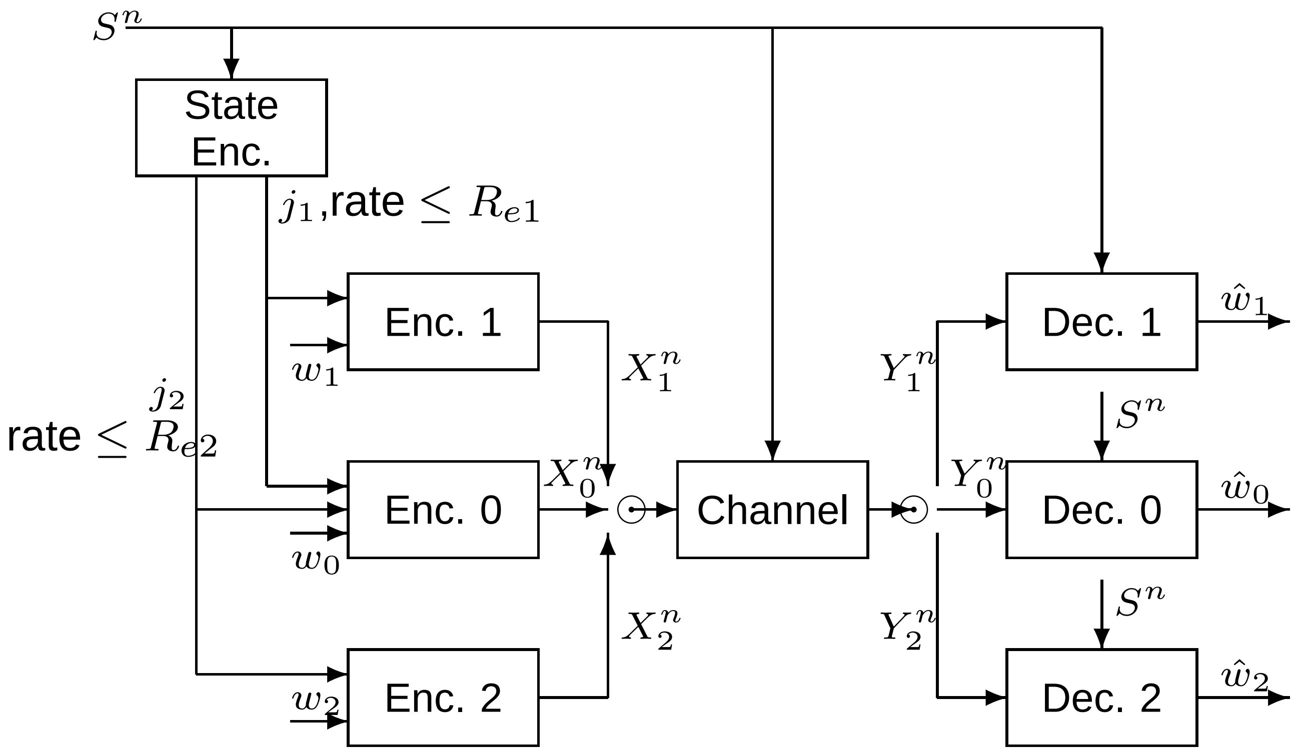
Multiple Descriptions (MD) of CS



- Encoder and decoder know the position of each of the switches
- The coding scheme in use depends on the switches (hence so does X^n)
- The wayside channels are represented by noiseless links (separation theorem...)

Coding with multi-csi *(cont'd)*

Multiple Descriptions (MD) of CS - an equivalent description:



Coding with multi-csi *(cont'd)*

R_{e1}, R_{e2} – rates of CSI streams j_1, j_2, sent to the main transmitter via noiseless wayside links.

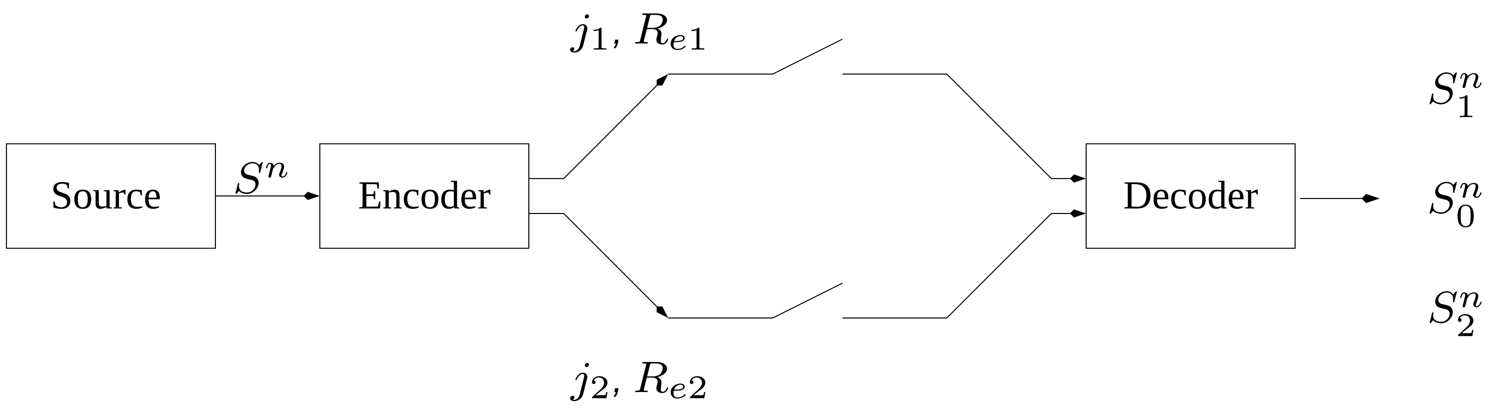
R_0 – Forward transmission rate when both streams, j_1 and j_2 arrive to the main transmitter.

R_i – Forward transmission rate when only stream j_i arrives to the main transmitter, $i = 1, 2$.

We are interested in $\mathcal{R}_{\text{MDCSI}}$, the region of all achievable quintuples $(R_{e1}, R_{e2}, R_0, R_1, R_2)$. [Cemal & Steinberg, ISIT 2005]

Coding with multi-csi *(cont'd)*

MD in Source Coding



The decoder produces S_0^n if both stream, j_1 and j_2, arrive, and S_i^n if only stream j_i arrive, $i = 1, 2$.

$$\mathbb{E}d(S^n, S_i^n) \leq D_i, \quad i = 0, 1, 2.$$

The MD problem: Characterize the set of all achievable $(R_{e1}, R_{e2}, D_0, D_1, D_2)$.

Coding with multi-csi (cont'd)

Relevant work MD has not been suggested before in CSI setting.

In the context of source coding:

- Posed by Gersho, Witsenhausen, Wolf, Wyner, Ziv, and Ozarow, at 1979 IT Workshop.
- Contributions by: Witsenhausen 1980, Wolf, Wyner,& Ziv 1980, Ozarow 1980, Witsenhausen & Wyner 1980,
- El Gamal & Cover 1982 – achievable region
- Berger & Zhang 83 – No excess rate for Bernoulli($1/2$) source, with $D_0 = 0$ (perfect reconstruction with the two streams)
- Ahlswede 1985 – The rate-distortion region for multiple descriptions without excess rate. General source.

Coding with multi-csi (cont'd)

Achievable Region [Cemal & Steinberg, 2005]

$\mathcal{R}_i$ – CH of all $(R_{e1}, R_{e2}, R_0, R_1, R_2)$ satisfying

$$\begin{aligned} R_{e1} &\geq I(S; S_1) \\ R_{e2} &\geq I(S; S_2) \\ R_{e1} + R_{e2} &\geq I(S; S_0, S_1, S_2) + I(S_1; S_2) \end{aligned}$$

$$\begin{aligned} R_0 &\leq I(X_0; Y_0 | S, S_0) \\ R_1 &\leq I(X_1; Y_1 | S, S_1) \\ R_2 &\leq I(X_2; Y_2 | S, S_2) \end{aligned}$$

for some (S_0, S_1, S_2)

$$\begin{aligned} S \multimap S_i \multimap X_i, &\quad i = 0, 1, 2 \\ S_i \multimap (X_i, S) \multimap Y_i &\quad i = 0, 1, 2. \end{aligned}$$

Coding with multi-csi (cont'd)

Outer Region [Cemal & Steinberg, 2005]

$\mathcal{R}_o$ – CH of all $(R_{e1}, R_{e2}, R_0, R_1, R_2)$ satisfying

$$\begin{aligned} R_{e1} &\geq I(S; S_1) \\ R_{e2} &\geq I(S; S_2) \\ R_{e1} + R_{e2} &\geq I(S; S_0, S_1, S_2) \end{aligned}$$

$$\begin{aligned} R_0 &\leq I(X_0; Y_0 | S, S_0) \\ R_1 &\leq I(X_1; Y_1 | S, S_1) \\ R_2 &\leq I(X_2; Y_2 | S, S_2) \end{aligned}$$

for some (S_0, S_1, S_2)

$$\begin{aligned} S \multimap S_i \multimap X_i, \quad & i = 0, 1, 2 \\ S_i \multimap (X_i, S) \multimap Y_i \quad & i = 0, 1, 2. \end{aligned}$$

Coding with multi-csi *(cont'd)*

Theorem 7 (Cemal & Steinberg, 2005) *For any discrete memoryless channel and state, and fully informed decoder,*

$$\mathcal{R}_i \subseteq \mathcal{R}_{\mathit{MDCSI}} \subseteq \mathcal{R}_o.$$

Coding with multi-csi (cont'd)

Achievable Region

$\mathcal{R}_i$ – CH of all $(R_{e1}, R_{e2}, R_0, R_1, R_2)$ satisfying

$$\begin{aligned} R_{e1} &\geq I(S; S_1) \\ R_{e2} &\geq I(S; S_2) \\ R_{e1} + R_{e2} &\geq I(S; S_0, S_1, S_2) + I(S_1; S_2) \end{aligned}$$

$$\begin{aligned} R_0 &\leq I(X_0; Y_0 | S, S_0) \\ R_1 &\leq I(X_1; Y_1 | S, S_1) \\ R_2 &\leq I(X_2; Y_2 | S, S_2) \end{aligned}$$

In MD for source coding, there are no Markov conditions, and the rate constraints (on R_0, R_1, R_2) are replaced by

$$D_i \geq \mathbb{E} d(S, S_i), \quad i = 0, 1, 2.$$

Coding with multi-csi (cont'd)

No Excess Rate

Some Notation:

$C(R_e)$ = capacity of forward channel with rate R_e CSI (optimal CSI coding in one stream). By HEG-83 and RSS-04,

$$\begin{aligned} C(R_e) &= \max I(X;Y|S,S_0) \\ \text{subject to} \quad & R_e \geq I(S;S_0), \quad S \multimap S_0 \multimap X, \quad S_0 \multimap (X,S) \multimap Y. \end{aligned}$$

Coding with multi-csi (cont'd)

No Excess Rate

Some Notation:

$C(R_e)$ = capacity of forward channel with rate R_e CSI (optimal CSI coding in one stream). By HEG-83 and RSS-04,

$$\begin{aligned} C(R_e) &= \max I(X;Y|S,S_0) \\ \text{subject to} \quad & R_e \geq I(S;S_0), \quad S \multimap S_0 \multimap X, \quad S_0 \multimap (X,S) \multimap Y. \end{aligned}$$

MD without excess rate: MD coding of CS with rates

$$(R_{e1}, R_{e2}, R_0, R_1, R_2)$$

such that

$$R_0 = C(R_{e1} + R_{e2}).$$

Coding with multi-csi *(cont'd)*

Notation:

$\mathcal{Q}$ – region of all achievable $(R_{e1}, R_{e2}, R_0, R_1, R_2)$ with $R_0 = C(R_{e1} + R_{e2})$.

$\mathcal{Q}_p$ – region of all achievable $(R_{e1}, R_{e2}, R_0, R_1, R_2)$ with $R_0 = C(R_{e1} + R_{e2})$, with probability of error decaying as $c/n^{1+\delta}$, for some $\delta > 0$.

We are interested in characterizing $\mathcal{Q}$. ($\mathcal{Q}_p$).

Coding with multi-csi *(cont'd)*

$\mathcal{R}$ – CH of all $(R_{e1}, R_{e2}, R_0, R_1, R_2)$ satisfying

$$\begin{aligned} R_{ei} &\geq I(S; S_i), \quad i = 1, 2 \\ R_{e1} + R_{e2} &\geq I(S; S_0, S_1, S_2) \\ R_i &\leq I(X_i; Y_i | S, S_i), \quad i = 0, 1, 2 \\ R_0 &= C(R_{e1} + R_{e2}) \end{aligned}$$

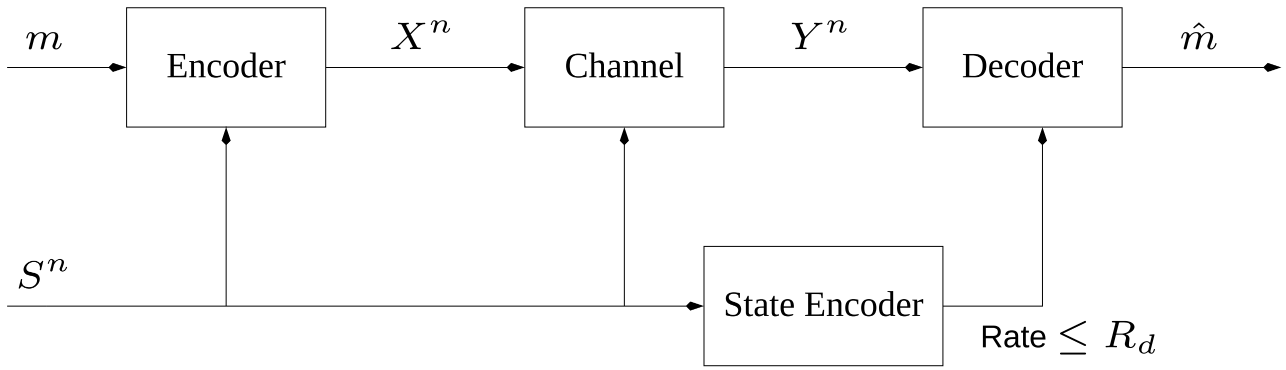
for some (S_0, S_1, S_2) such that $I(S_1; S_2) = 0$, and

$$S \multimap S_i \multimap X_i, \quad S_i \multimap (X_i, S) \multimap Y_i, \quad i = 0, 1, 2.$$

Theorem 8 *[Cemal & Steinberg 2005]*

$$\mathcal{Q}_p = \mathcal{R}.$$

Rate limited SI at the decoder



- Memoryless channel $P_{Y|X,S}(y|x,s)$ and state $P_S(s)$
- State sequence S^n known a priori at the encoder
- A compressed version of S^n, with rate$(S^n) \leq R_d$, is provided to the decoder.

We are interested in the region of all achievable rates and input costs:

$$R = \frac{\log|\mathcal{M}|}{n}, \quad R_d = \frac{\log|\mathcal{T}|}{n}, \quad \Gamma = E\phi(X^n).$$

Rate limited SI at the decoder *(cont'd)*

Possible applications:

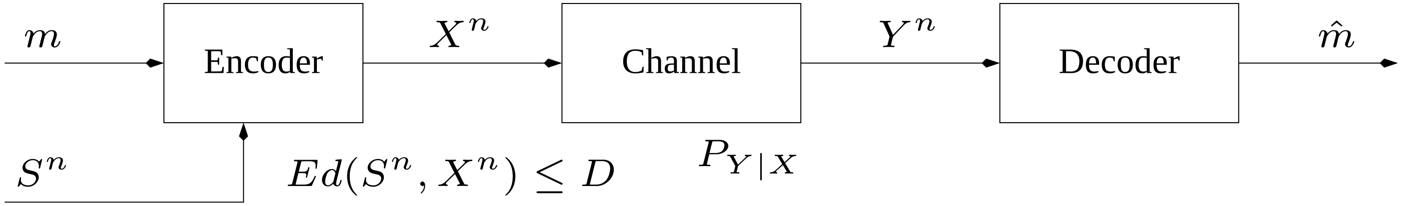
- Communication systems:

  OFDM + coding, where coding is done across frequencies. The sender knows channels states (fading), and sends it via a wayside channel to the receiver.

- Watermarking (WM) with compressed host at the decoder.

Rate limited SI at the decoder (cont'd)

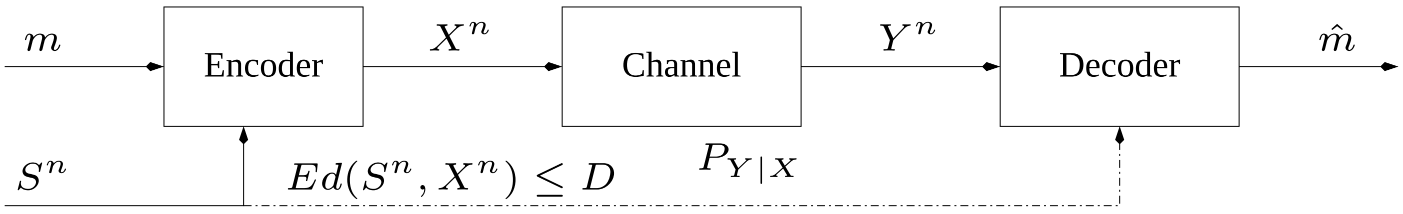
Watermarking applications:



- Public Watermarking – The host data S^n is available only at the encoder.

Rate limited SI at the decoder (cont'd)

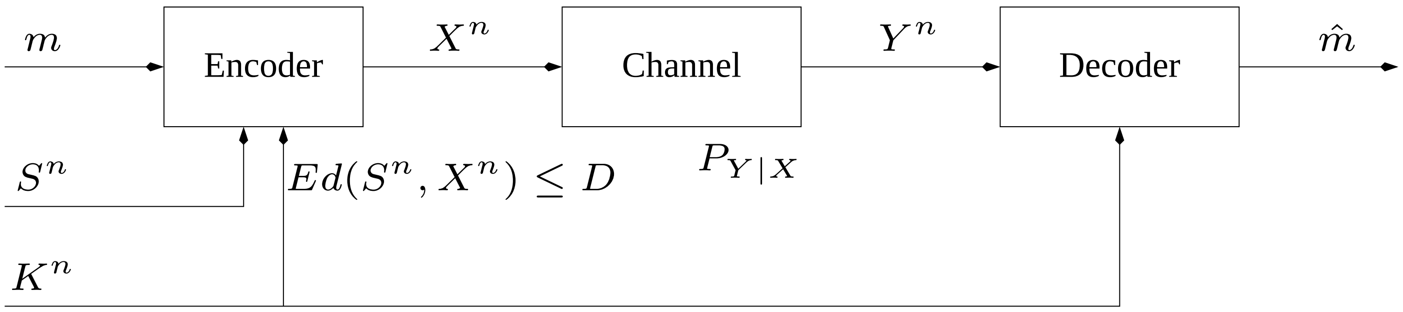
Watermarking applications:



- Public Watermarking – The host data S^n is available only at the encoder.
- Private Watermarking – The host data S^n is available at both, encoder and decoder.

Rate limited SI at the decoder (cont'd)

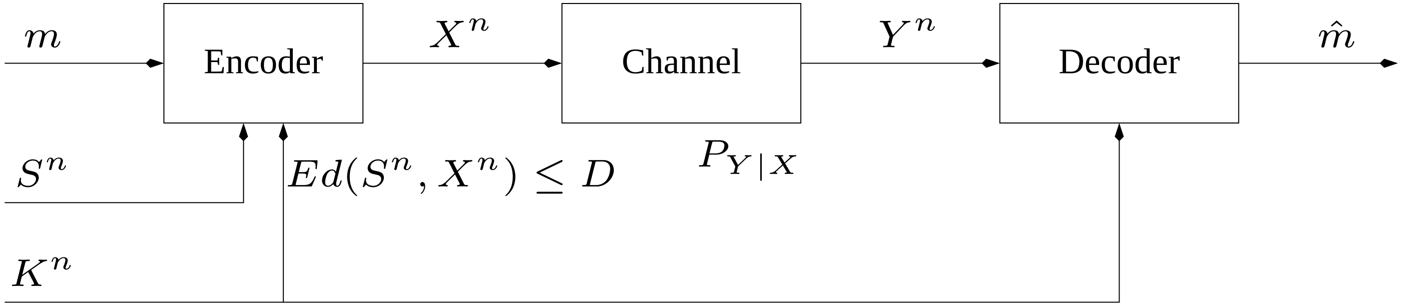
Watermarking applications:



- Public Watermarking – The host data S^n is available only at the encoder.
- Private Watermarking – The host data S^n is available at both, encoder and decoder.
- A bridge between the versions [Moulin & O'Sullivan] – A key K^n is present at the encoder and decoder, with a given $P_{S,K}$.

Rate limited SI at the decoder (cont'd)

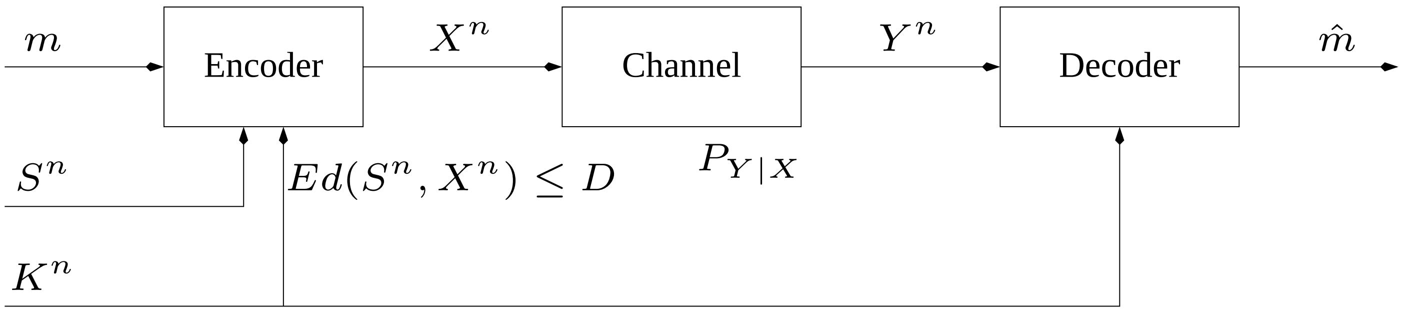
Watermarking applications:



- Public Watermarking – The host data S^n is available only at the encoder.
- Private Watermarking – The host data S^n is available at both, encoder and decoder.
- A bridge between the versions [Moulin & O'Sullivan] – A key K^n is present at the encoder and decoder, with a given $P_{S,K}$.
  - K^n is provided to the decoder at no cost

Rate limited SI at the decoder (cont'd)

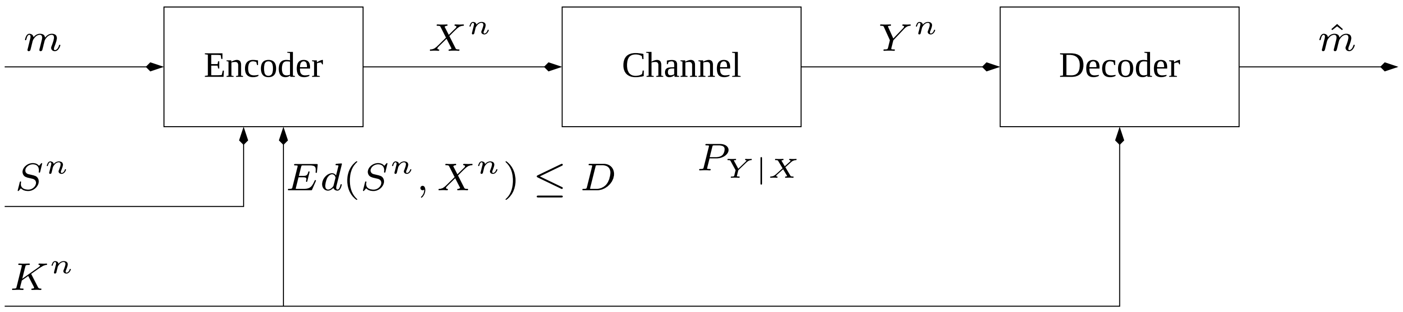
Watermarking applications:



- Public Watermarking – The host data S^n is available only at the encoder.
- Private Watermarking – The host data S^n is available at both, encoder and decoder.
- A bridge between the versions [Moulin & O'Sullivan] – A key K^n is present at the encoder and decoder, with a given $P_{S,K}$.
    - K^n is provided to the decoder at no cost
    - How to choose $P_{K|S}$?

Rate limited SI at the decoder (cont'd)

Watermarking applications:

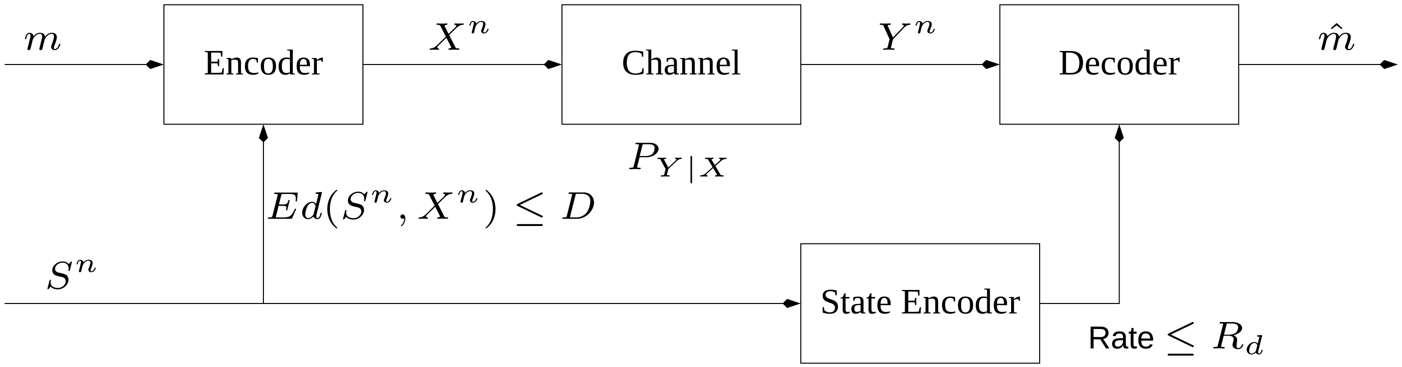


- Public Watermarking – The host data S^n is available only at the encoder.
- Private Watermarking – The host data S^n is available at both, encoder and decoder.
- A bridge between the versions [Moulin & O'Sullivan] – A key K^n is present at the encoder and decoder, with a given $P_{S,K}$.
  - K^n is provided to the decoder at no cost
  - How to choose $P_{K|S}$?

$\Rightarrow$ Quantify the deocder's a priori knowledge by a rate-limit

Rate limited SI at the decoder *(cont'd)*

Watermarking applications:



Problem:

Characterize the region of all achievable (R, R_d, D), where:

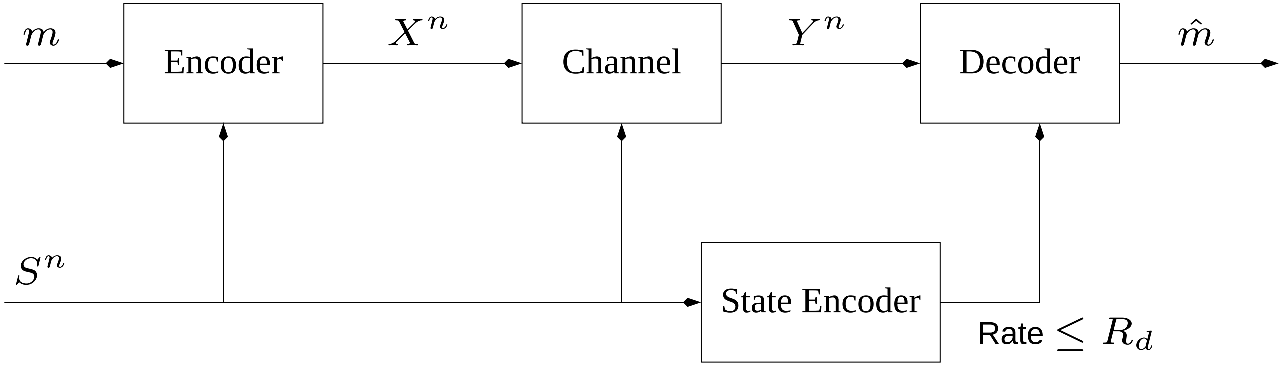
R – Embedding rate,

R_d – rate of compressed SI @ decoder

D – distortion between host and input.

Rate limited SI at the decoder (cont'd)

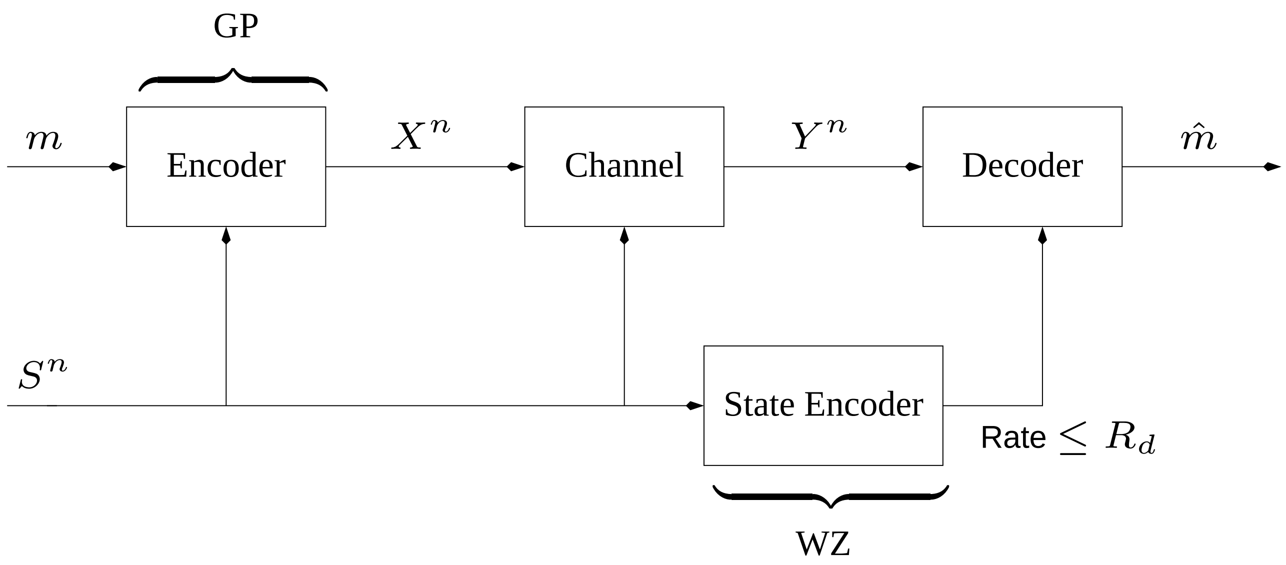
Related problems



- S^n is known noncausally at the encoder $\Rightarrow$ channel coding part is related to the Gel'fand-Pinsker (GP) problem.
- Y^n depends statistically on S^n and can serve as side information (SI) in retrieving the compressed state at the decoder $\Rightarrow$ coding of S^n is related to the Wyner-Ziv (WZ) problem.

Rate limited SI at the decoder (cont'd)

Related problems



- For the WZ problem, the SI Y^n is not memoryless
- There is no distortion constraint in retrieving S^n at the decoder (instead, maximize capacity of the main channel)

Rate limited SI at the decoder (cont'd)

Relevant work

- Wyner & Ziv, 1976
- Gel'fand & Pinsker, 1980
- Heegard & El Gamal, 1983, "On the capacity of computer memory with defects."

  Introduced coding for state dependent channels with rate limited side information at both ends. Devised an achievable region.

The current model is a special case of Heegard & El Gamal's model.

Rate limited SI at the decoder *(cont'd)*

Works related to WM: (very partial list)

- Moulin & O'Sullivan, 2003 – Introduced WM from IT viewpoint. Connection to GP. Bridging between public and private WM, via K^n.
- Willems & kalker, 2002 – WM system without attack channel. Two new ingerdients:
  - The host S^n is reconstructed within distortion D_2 at the decoder
  - *Composite rate limit*: a rate limit is put on the data set X^n. (Huffman code.)
- Maor & Merhav, 2005a, 2005b – Extended Willems & kalker work: (a) general lossless codes, (b) attack channel.

Rate limited SI at the decoder *(cont'd)*

Define: $\mathcal{R}^*$ – collection of all (R, R_d, Γ) satisfying

$$\begin{aligned} R &\leq I(U;Y|S_d) - I(U;S|S_d) \\ R_d &\geq I(S;S_d) - I(Y;S_d) \\ \Gamma &\geq \mathbb{E}\phi(X) \end{aligned}$$

for some (U, S_d) such that $(U, S_d) \multimap (S, X) \multimap Y$. Then

Theorem 9 (Steinberg ITW 2006) *For any discrete memoryless state-dependent channel, with full noncausal SI at the transmitter, and rate-limited SI at the receiver, a triple (R, R_d, Γ) is achievable if and only if $(R, R_d, \Gamma) \in \mathcal{R}^*$.*

Rate limited SI at the decoder (cont'd)

$\mathcal{R}^*$ – collection of all (R, R_d, Γ) satisfying

$$\begin{aligned} R &\leq I(U;Y|S_d) - I(U;S|S_d) \\ R_d &\geq I(S;S_d) - I(Y;S_d) \\ \Gamma &\geq \mathbb{E}\phi(X) \end{aligned}$$

for some (U, S_d) such that $(U, S_d) \multimap (S, X) \multimap Y$.

- S_d – A WZ rv, represents the compressed state S^n. Fully decoded, with Y^n as SI.
- U – A GP rv, represents the encoded message. Fully decoded conditioned on S_d in both sides.

Rate limited SI at the decoder *(cont'd)*

$\mathcal{R}^*$ – collection of all (R, R_d, Γ) satisfying

$$
\begin{aligned}
R &\leq I(U;Y|S_d) - I(U;S|S_d) \\
R_d &\geq I(S;S_d) - I(Y;S_d) \quad (*) \\
\Gamma &\geq \mathbb{E}\phi(X)
\end{aligned}
$$

for some (U, S_d) such that $(U, S_d) -\!\!\circ\!\!- (S, X) -\!\!\circ\!\!- Y$.

- $(U, S_d) -\!\!\circ\!\!- (S, X) -\!\!\circ\!\!- Y$ does not imply $S_d -\!\!\circ\!\!- S -\!\!\circ\!\!- Y$. Therefore $(*)$ is not equivalent to

  $$R_d \geq I(S;S_d|Y),$$

  full duality with GP.
- In classical WZ, $S_d -\!\!\circ\!\!- S -\!\!\circ\!\!- Y$ is needed to guarantee joint typicality of S_d and Y. Here it is guaranteed due to the channel.

Rate limited SI at the decoder (cont'd)

$\mathcal{R}^*$ – collection of all (R, R_d, Γ) satisfying

$$\begin{aligned} R &\le I(U;Y|S_d) - I(U;S|S_d) \\ R_d &\ge I(S;S_d) - I(Y;S_d) \\ \Gamma &\ge \mathbb{E}\phi(X) \end{aligned}$$

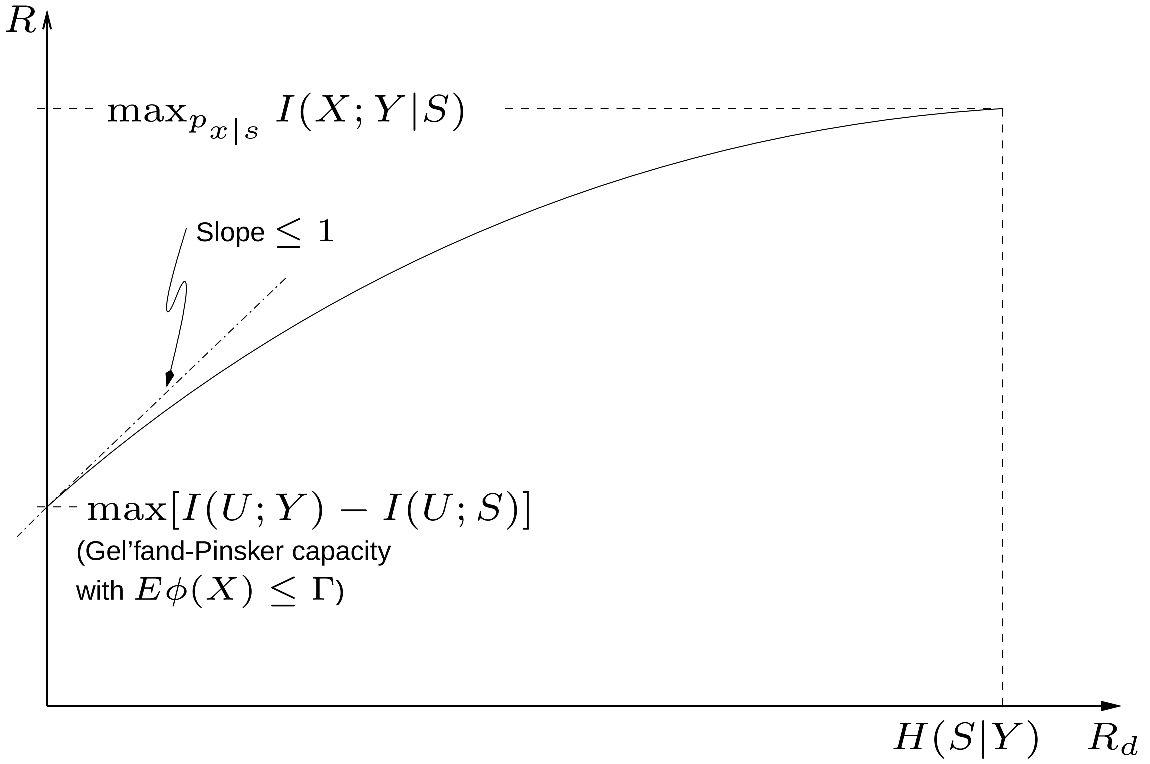
for some (U, S_d) such that $(U, S_d) \multimap (S, X) \multimap Y$.

Properties of $\mathcal{R}^*$

- $\mathcal{R}^*$ is convex
- $X = f(U, S_d, S)$, f deterministic, suffices to exhaust $\mathcal{R}^*$.

Rate limited SI at the decoder (cont'd)

A typical (R, R_d) curve, for fixed Γ:



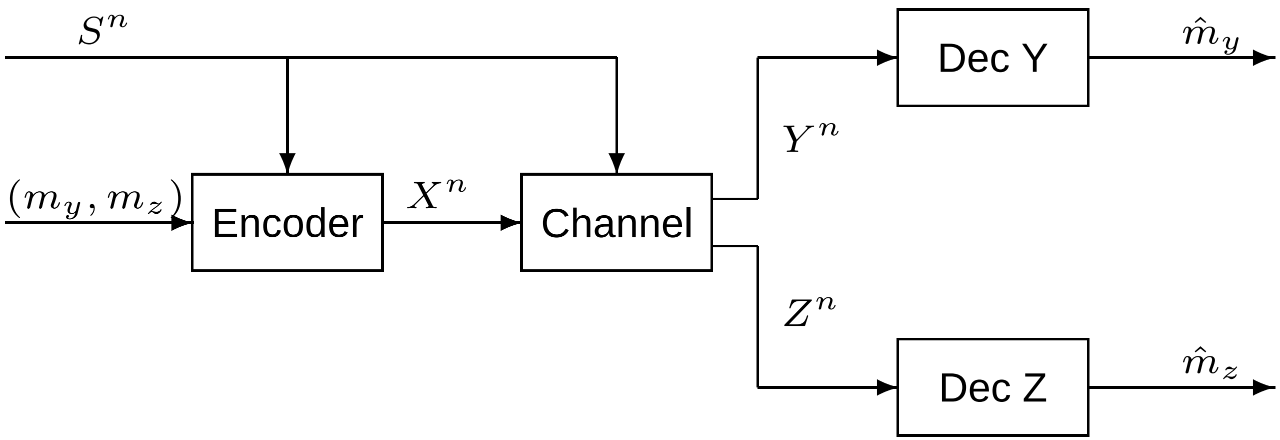
Rate limited SI at the decoder *(cont'd)*

- The rate allocated to provide the decoder with SI, is always at least as high as the gain in the forward rate.
- Provide SI to the decoder when the wayside channel cannot be used to transmit data – e.g.
  - Remotely located physical channel
  - WM, where a compressed host is kept in memory at the decoder, for future use.

Multi-user models

The state dependent broadcast channel

General BC with random parameters:



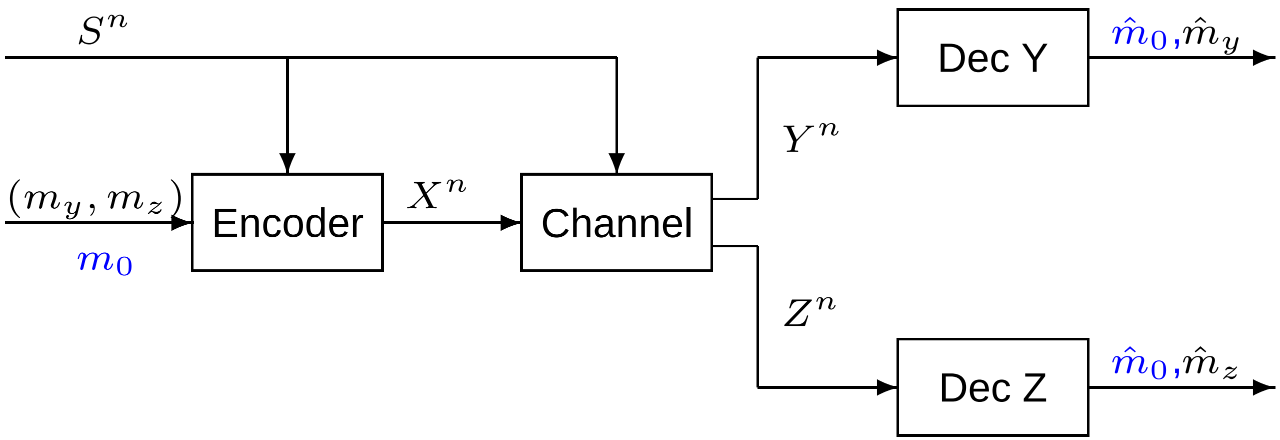
- Memoryless channel $P_{Y,Z|X,S}(y,z|x,s)$ and state $P_S(s)$
- State sequence S^n known a priori at the encoder

We are interested in (capacity?) region of achievable rates (R_Y, R_Z)

$$R_Y = \frac{\log|\mathcal{M}_Y|}{n}, \quad R_Z = \frac{\log|\mathcal{M}_Z|}{n}$$

The state dependent broadcast channel

General BC with random parameters:



- Memoryless channel $P_{Y,Z|X,S}(y,z|x,s)$ and state $P_S(s)$
- State sequence S^n known a priori at the encoder

We are interested in (capacity?) region of achievable rates (R_0, R_Y, R_Z)

$$R_0 = \frac{\log|\mathcal{M}_0|}{n}, \quad R_Y = \frac{\log|\mathcal{M}_Y|}{n}, \quad R_Z = \frac{\log|\mathcal{M}_Z|}{n}$$

The state-dependent broadcast channel (cont'd)

- Performance (capacity, achievable regions) depends on $P_{Y,Z|X,S}$ only via the conditional marginals

$$F(y|x,s) = \sum_z P_{Y,Z|X,S}(y,z|x,s)$$

$$G(z|x,s) = \sum_y P_{Y,Z|X,S}(y,z|x,s)$$

- A state dependent broadcast channel is called *physically degraded* if

$$P_{Y,Z|X,S} = F_{Y|X,S} P_{Z|Y}$$

and is called *stochastically degraded* if there exists some $P'_{Z|Y}$ such that

$$G_{Z|X,S}(z|x,s) = \sum_y F_{Y|X,S}(y|x,s) P'_{Z|Y}(z|y)$$

- Since the capacity region depends only on F, G, no distinction has to be made between stochastically and physically degraded channels, and we will use the term *degraded*.

The state-dependent broadcast channel (cont'd)

Possible applications:

1. OFDM + Coding for the broadcast channel, where coding is done across frequencies
2. Watermarking systems with several stages of attack, or several possible (fixed) attack channels (e.g, coding for both public and private users).

The state-dependent broadcast channel (cont'd)

Relevant Work: (very partial lists)

General broadcast channel

- Introduced by Cover 1972, then contributions by Bergmans, Bergmans & Cover, Cover, Van der Meulen, El Gamal, Körner & Marton, Marton,.....
- Marton 1979 – Inner (achievability) and outer bounds.
  No explicit mention of common rate, but bounds can be deduced (El Gamal).

Single user channels with random states known non causally at the transmitter

- Gel'fand & Pinsker, 1980 – Capacity formula.
- Costa, 1983 – Capacity of Gaussian channel with additive Gaussian state. $C = 1/2 \log(1 + P/N)$. "WDP property" (term coined by Kim, Sutivong, Sigurjónsson, 2004).

The state-dependent broadcast channel (cont'd)

Broadcast channels with random states, no common rate

- Gel'fand & Pinsker, 1984 – Gaussian BC and MAC satisfy WDP property. For BC, solved

$$(a) \qquad Y_i = X + S + Z_i, \quad \text{(using Costa's coding arguments)}$$

$$(b) \qquad Y_i = X + S_i + Z_i, \quad \text{(commented on, with conditions.)}$$

The state-dependent broadcast channel (cont'd)

Broadcast channels with random states, no common rate

- Gel'fand & Pinsker, 1984 – Gaussian BC and MAC satisfy WDP property. For BC, solved

$$(a) \qquad Y_i = X + S + Z_i, \quad \text{(using Costa's coding arguments)}$$
$$(b) \qquad Y_i = X + S_i + Z_i, \quad \text{(commented on, with conditions.)}$$

- Steinberg, 2002, 2004 – Inner and outer bounds for the *degraded* BC

$$P_{Y,Z|X,S} = P_{Y|X,S} P_{Z|Y} \quad \text{(physically or stochastically)}$$

  Tight for the case of informed Y decoder.
  Single letter characterization of capacity region for the causal case.

- Kim, Sutivong, & Sigurjónsson, 2004 – WDP property for the Gaussian BC, MAC, and physically degraded relay. For BC, solved

$$Y_i = X + S + Z_i, \quad i = 1, 2.$$

- Steinberg & Shamai, 2005 – Inner bound for general BC. Studied also common rate.

The state-dependent broadcast channel *(cont'd)*

Broadcast channels with random states, no common rate

- Gel'fand & Pinsker, 1984 – Gaussian BC and MAC satisfy WDP property. For BC, solved

$$(a) \qquad Y_i = X + S + Z_i, \quad \text{(using Costa's coding arguments)}$$
$$(b) \qquad Y_i = X + S_i + Z_i, \quad \text{(commented on, with conditions.)}$$

- Steinberg, 2002, 2004 – Inner and outer bounds for the *degraded* BC

$$P_{Y,Z|X,S} = P_{Y|X,S} P_{Z|Y} \quad \text{(physically or stochastically)}$$

  Tight for the case of informed Y decoder.
  Single letter characterization of capacity region for the causal case.

- Kim, Sutivong, & Sigurjónsson, 2004 – WDP property for the Gaussian BC, MAC, and physically degraded relay. For BC, solved

$$Y_i = X + S + Z_i, \quad i = 1, 2.$$

- Steinberg & Shamai, 2005 – Inner bound for general BC. Studied also common rate. It solves case (b) above, which is a non-degraded channel.

The state-dependent broadcast channel (cont'd)

Broadcast channel with random states, common rate

- Khisti, Erez, & Wornell, 2004 – Suggested the model of common rate only (i.e., broadcasting the same message to all users). Capacity of a certain class of binary channels.
- Steinberg & Shamai, ISIT 2005.
- Khisti, Wornell, Erez, Lapidoth – proved a conjecture by Steinberg and Shamai on maximal common rate in Gaussian BC with infinite power additive known interference.

Degraded BC

Inner bound

- $\mathcal{P}$ – collection of all RVs $(\tilde{U}, S, X, Y, Z)$ such that $\tilde{U} \multimap (S, X) \multimap (Y, Z)$ is Markov.
- Define $\mathcal{R}_i$ to be the set of all rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K;Z) - I(K;S) \\ R_Y &\leq I(U;Y|K) - I(U;S|K) \\ &\qquad \text{for some } ((K,U), X, S, Y, Z) \in \mathcal{P}. \end{aligned}$$

Theorem 10 (Steinberg 2002) *For any discrete memoryless degraded broadcast channel with random parameters*

$$\mathcal{R}_i \subseteq \mathcal{C}$$

Proof based on combination of the BC superposition coding, and GP binning technique.

Degraded BC *(cont'd)*

Inner bound (cont'd)

$\mathcal{R}_i$ - the set of all rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K;Z) - I(K;S) \\ R_Y &\leq I(U;Y|K) - I(U;S|K) \end{aligned}$$

for some $((K,U), X, S, Y, Z) \in \mathcal{P}$.

Proposition 2 (Steinberg 2002) *1. The set $\mathcal{R}_i$ is convex*

2. *To exhaust $\mathcal{R}_i$, it is enough to take X to be a deterministic function of the triple (K, U, S)*
3. *To exhaust $\mathcal{R}_i$, it is enough to restrict $\mathcal{K}$ and $\mathcal{U}$ to satisfy*

$$\begin{aligned} |\mathcal{K}| &\leq |\mathcal{S}||\mathcal{X}| + 1 \\ |\mathcal{U}| &\leq |\mathcal{S}||\mathcal{X}|(|\mathcal{S}||\mathcal{X}| + 1). \end{aligned}$$

Degraded BC *(cont'd)*

Outer bound

- Define $\mathcal{R}_o$ to be the set of all rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K;Z) - I(K;S) \\ R_Y &\leq I(U;Y|K,V) - I(U;S|K,V) \\ R_Y + R_Z &\leq I(K,V,U;Y) - I(K,V,U;S) \end{aligned}$$

for some $((K,V,U),S,X,Y,Z) \in \mathcal{P}$.

Theorem 11 (Steinberg 2002) *For any discrete memoryless degraded broadcast channel with random parameters*

$$\mathcal{C} \subseteq \mathcal{R}_o$$

Proof based on techniques similar to Gel'fand & Pinsker, 1980, and Csiszár & Körner, 1978.

As with $\mathcal{R}_i$, $\mathcal{R}_o$ is convex. Bounds on the alphabets are derived via the support lemma.

Degraded BC *(cont'd)*

The non-degraded user is informed

- Define $\mathcal{R}$ to be the set of rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K;Z) - I(K;S) \\ R_Y &\leq I(X;Y|K,S) \\ &\text{for some } (K,S,X,Y,Z) \in \mathcal{P}. \end{aligned}$$

Theorem 12 (Steinberg 2002) *For any discrete memoryless degraded broadcast channel with random parameters and informed non-degraded user*

$$\mathcal{C} = \mathcal{R}$$

Proof based on Theorem 10 and Theorem 11.

$\mathcal{R}$ is convex.

Degraded BC with causal SI

When the side information is provided to the encoder in a causal manner, a single letter characterization of the capacity region can be obtained.

- $\mathcal{P}_c$ – collection of all RV's $(\tilde{K}, S, X, Y, Z) \in \mathcal{P}$ satisfying $P_{\tilde{K},S} = P_{\tilde{K}} P_S$.
- Define $\mathcal{R}_c$ to be the set of rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K; Z) \\ R_Y &\leq I(U; Y|K) \end{aligned}$$

for some $((K, U), S, X, Y, Z) \in \mathcal{P}_c$.

Theorem 13 (Steinberg 2002) *For any discrete memoryless degraded broadcast channel with random parameters and causal side information at the encoder*

$$\mathcal{C} = \mathcal{R}_c$$

Proof of direct part – based on Theorem 10, with (K, U) independent on S.

Proof of converse – based on tightening of Theorem 11

Degraded BC with causal SI *(cont'd)*

- $\mathcal{P}_c$ – collection of all RV's $(\tilde{K}, S, X, Y, Z) \in \mathcal{P}$ satisfying $P_{\tilde{K},S} = P_{\tilde{K}} P_S$.
- $\mathcal{R}_c$ – the set of rate pairs (R_Y, R_Z) such that

$$\begin{aligned} R_Z &\leq I(K;Z) \\ R_Y &\leq I(U;Y|K) \end{aligned}$$

for some $((K,U), S, X, Y, Z) \in \mathcal{P}_c$.

Proposition 3 (Steinberg 2002) *1. The set $\mathcal{R}_c$ is convex*

2. To exhaust $\mathcal{R}_c$, it is enough to take X to be a deterministic function of the triple (K, U, S)

3. To exhaust $\mathcal{R}_c$, it is enough to restrict $\mathcal{K}$ and $\mathcal{U}$ to satisfy

$$\begin{aligned} |\mathcal{K}| &\leq |\mathcal{S}||\mathcal{X}| + 1 \\ |\mathcal{U}| &\leq |\mathcal{S}||\mathcal{X}|(|\mathcal{S}||\mathcal{X}| + 1). \end{aligned}$$

Degraded BC with causal SI (cont'd)

Formulation with strategies

Due to Proposition 3, the region $\mathcal{R}_c$ can be expressed in terms of strategies:

$$t : \mathcal{S} \to \mathcal{X}$$

Now the *random* strategies are drawn independently of S, but they can depend on K:

$$\mathcal{R}_c = \bigcup_{P_K P_{T|K}} \{(R_Y, R_Z) : \begin{array}{rcl} R_Z & \leq & I(K;Z) \\ R_Y & \leq & I(T;Y|K)\} \end{array}$$

where

K – external RV
$P_{T|K}(\cdot|k)$ – conditional distribution on the set of Shannon strategies $\mathcal{T}$, conditioned on $K = k$
The pair (T, K) is independent of S.

The general BC with SI

Gel'fand & Pinsker (GP) capacity formula:

$$C = \max_{P_{V,X|S}} I(V;Y) - I(V;S), \quad V \multimap (X,S) \multimap Y.$$

Marton's inner bound:

$$\begin{aligned} R_Y &\le I(WV;Y) \\ R_Z &\le I(WU;Z) \\ R_Y + R_Z &\le \min\{I(W;Y), I(W;Z)\} + I(V;Y|W) + I(U;Z|W) \\ &\quad - I(U;V|W) \end{aligned}$$

The general BC with SI

Gel'fand & Pinsker (GP) capacity formula:

$$C = \max_{P_{V,X|S}} I(V;Y) - I(V;S), \quad V \multimap (X,S) \multimap Y.$$

Marton's inner bound:

$$\begin{aligned}
R_Y &\le I(WV;Y) \\
R_Z &\le I(WU;Z) \\
R_Y + R_Z &\le \min\{I(W;Y), I(W;Z)\} + I(V;Y|W) + I(U;Z|W) \\
&\quad - I(U;V|W) \\
R_Y + R_Z &\le I(WV;Y) + I(WU;Z) - I(U;V|W) \\
&\quad - \max\{I(W;Y), I(W;Z)\}.
\end{aligned}$$

The general BC with SI

Gel'fand & Pinsker (GP) capacity formula:

$$C = \max_{P_{V,X|S}} I(V;Y) - I(V;S), \quad V \multimap (X,S) \multimap Y.$$

Marton's inner bound:

$$\begin{aligned}
R_Y &\leq I(WV;Y) \\
R_Z &\leq I(WU;Z) \\
R_Y + R_Z &\leq \min\{I(W;Y), I(W;Z)\} + I(V;Y|W) + I(U;Z|W) \\
&\quad - I(U;V|W) \\
R_Y + R_Z &\leq I(WV;Y) + I(WU;Z) - I(U;V|W) \\
&\quad - \max\{I(W;Y), I(W;Z)\}.
\end{aligned}$$

We would like to combine one of these with the GP formula, to get a good (?) achievable region for the GPBC.

The general BC with SI (cont'd)

An inner bound

$\mathcal{R}_i$ – collection of all (R_Y, R_Z) satisfying

$$
\begin{aligned}
R_Y &\leq I(WV;Y) - I(WV;S) \\
R_Z &\leq I(WU;Z) - I(WU;S) \\
R_Y + R_Z &\leq I(WV;Y) - I(WV;S) + I(WU;Z) - I(WU;S) \\
&\quad - I(U;V|WS) \\
&\quad - [\max\{I(W;Y), I(W;Z)\} - I(W;S)]_+
\end{aligned}
$$

for some (W,V,U) s.t. $(W,V,U) \multimap (X,S) \multimap (Y,Z)$. Then

Theorem 14 (Steinberg & Shamai 2005) *For any discrete memoryless broadcast channel with random parameters, CH*$(\mathcal{R}_i)$ *is achievable.*

This region reduces to Marton's for degenerate S.

The general BC with SI (cont'd)

Some caution:

$\mathcal{R}_i$ – collection of all (R_Y, R_Z) satisfying

$$\begin{aligned} R_Y &\leq I(WV;Y) - I(WV;S) \quad \text{no (pos.) restrictions here} \\ R_Z &\leq I(WU;Z) - I(WU;S) \\ R_Y + R_Z &\leq I(WV;Y) - I(WV;S) + I(WU;Z) - I(WU;S) \\ &\quad - I(U;V|WS) \\ &\quad - [\max\{I(W;Y), I(W;Z)\} - I(W;S)]_+ \end{aligned}$$

for some (W, V, U) s.t. $(W, V, U) \multimap (X, S) \multimap (Y, Z)$.

The general BC with SI *(cont'd)*

Some caution:

$\mathcal{R}_i$ – collection of all (R_Y, R_Z) satisfying

$$
\begin{aligned}
R_Y &\leq I(WV;Y) - I(WV;S) \qquad \text{no (pos.) restrictions here} \\
R_Z &\leq I(WU;Z) - I(WU;S) \\
R_Y + R_Z &\leq I(WV;Y) - I(WV;S) + I(WU;Z) - I(WU;S) \\
&\quad - I(U;V|WS) \\
&\quad - [\max\{I(W;Y), I(W;Z)\} - I(W;S)]_+
\end{aligned}
$$

for some (W,V,U) s.t. $(W,V,U) \multimap (X,S) \multimap (Y,Z)$.

Note: if we decompose

$$
\begin{aligned}
R_Y &\leq I(WV;Y) - I(WV;S) \\
&= I(W;Y) - I(W;S) \;+\; I(V;Y|W) - I(V;S|W),
\end{aligned}
$$

one of the pairs can be negative.

The general BC with SI *(cont'd)*

Private and Common messages

In fact, one can deduce achievability of the CH of all triples (R_0, R_Y, R_Z) satisfying

$$\begin{aligned}
R_0 &\le [\min\{I(W;Y), I(W;Z)\} - I(W;S)]_+ \\
R_Y + R_0 &\le I(WV;Y) - I(WV;S) \\
R_Z + R_0 &\le I(WU;Z) - I(WU;S) \\
R_Y + R_Z + R_0 &\le I(WV;Y) - I(WV;S) + I(WU;Z) - I(WU;S) \\
&\quad - I(U;V|WS) \\
&\quad - [\max\{I(W;Y), I(W;Z)\} - I(W;S)]_+
\end{aligned}$$

for some (W, V, U) s.t. $(W, V, U) \multimap (X, S) \multimap (Y, Z)$.

The general BC with SI (cont'd)

Common messages only (Khisti, Erez, & Wornell problem)

If interested in conveying one message to all users, the transmitter can utilize the independent rate for that purpose. Thus, the following is achievable

$$R_{common} = \max_{(W,V,U)} \{\min\{R_Y, R_Z\} + R_0\}$$

where (R_0, R_Y, R_Z) are characterized as before.

- Here R_Y and R_Z do not carry independent information.

The general BC with SI *(cont'd)*

Example: The Gaussian BC. Let

$$\begin{aligned} Y &= X + S_1 + N_1 \\ Z &= X + S_2 + N_2 \end{aligned}$$

where

- $N_i \sim \mathcal{N}(0, Q_{N_i})$, $S_i \sim \mathcal{N}(0, Q_{S_i})$, $i = 1, 2$.
- N_1, N_2, and (S_1, S_2) are independent of each other, but S_1, S_2 can be correlated.

Gelfand & Pinsker, 1984: For the case of independent rates ($R_0 = 0$), the capacity region of this channel coincides with that of Gaussian BC without states.
Can be deduced by proper choice of (W, V, U).

The general BC with SI (cont'd)

W.l.o.g, assume $Q_{N_1} \geq Q_{N_2}$. Define $S = (S_1, S_2)$, and set $\beta \in [0, 1]$. Decompose

$$X = X_1 + X_2, \quad X_1, X_2 \text{ indep., of powers } \beta Q_X, \ (1-\beta)Q_X.$$

Let W be a null r.v., and define

$$\begin{aligned} V &= X_1 + \alpha_1 S_1 \\ U &= X_2 + \alpha_2 (S_2 + X_1) \\ \alpha_1 &= \frac{\beta Q_X}{Q_X + Q_{N_1}} \\ \alpha_2 &= \frac{(1-\beta)Q_X}{(1-\beta)Q_X + Q_{N_2}} \end{aligned}$$

The general BC with SI (cont'd)

Substituting in the formula for $\mathcal{R}_i$, we get

$$R_Y = \frac{1}{2}\log\left(1+\frac{\beta Q_X}{Q_{N_1}+(1-\beta)Q_X}\right)$$

$$R_Z = \frac{1}{2}\log\left(1+\frac{(1-\beta)Q_X}{Q_{N_2}}\right)$$

which is the capacity region of the Gaussian BC without states.

$\Rightarrow$ Additive interference (possibly different at each user) does not reduce the capacity region of the Gaussian BC, provided it is known non causally at the encoder.

The general BC with SI *(cont'd)*

Common messages only For simplicity, assume

- Symmetric Gaussian BC: $Q_{N_1} = Q_{N_2} = Q_N$, $Q_{S_1} = Q_{S_2} = Q_S$.
- S_1, S_2 independent of each other, with $Q_S \to \infty$

With the current definition of (W, V, U), the following common rate is achievable

$$R_{common} = \max_{0 \le \beta \le 1} \min\{R_Y, R_Z\} = \frac{1}{4} \log(1 + Q_X/Q_N).$$

Achievable by time sharing, applying Costa's coding to each of the users separately, in its own time slot.

- Note: Here (W, V, U) were optimized for the capacity region with $R_0 = 0$, not for R_{common}. For finite Q_S, higher R_{common} may be obtained.

The general BC with SI *(cont'd)*

$$R_{common} = \max_{(W,V,U)} \{\min\{R_Y, R_Z\} + R_0\}$$

The general BC with SI (cont'd)

$$R_{common} = \max_{0 \le \beta \le 1} \min\{R_Y, R_Z\} = \frac{1}{4} \log(1 + Q_X / Q_N)$$

Although obtained under optimization for independent rates, this strategy is optimal with infinite interference power.

Theorem 15 *For the symmetric Gaussian broadcast channel with additive, independent, infinite power interference S_1, S_2 known a priori at the encoder, the maximal common rate is $(1/4) \log(1 + Q_X / Q_N)$, and is achieved by time sharing.*

Conjectured in [Steinberg & Shamai 2005].

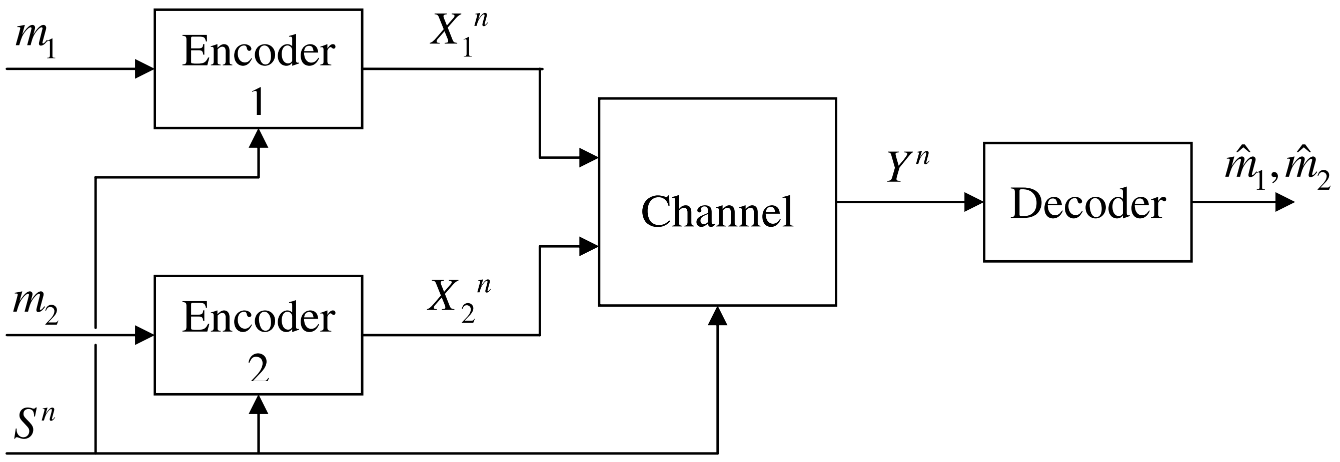
Proved in [Khisti, Erez, Lapidoth, Wornell 2006].

The general BC with SI (cont'd)

$$R_{common} = \max_{0 \le \beta \le 1} \min\{R_Y, R_Z\} = \frac{1}{4}\log(1 + Q_X/Q_N)$$

- No penalty is incurred due to the presence of (S_1, S_2) when broadcasting *independent* messages. (True for any power of (S_1, S_2).)
- When transmitting common information only, the current achievability results show noticeable penalty. The penalty affects the pre-log for infinite power (S_1, S_2).

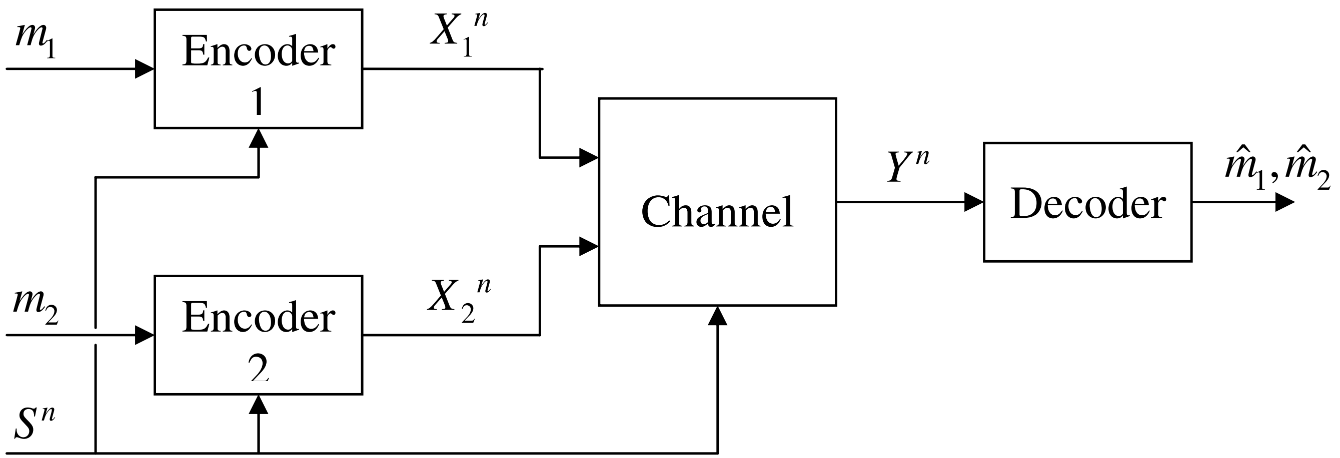
The state dependent multiple access channel



- Memoryless channel $P_{Y|X_1,X_2,S}$ and state P_S
- State sequence known non-causally at the two encoders.

A MAC version of the GP problem. Still unsolved. Inner and outer bounds are known.

State dependent MAC *(cont'd)*



An inner bound

The convex hull of the set of all rate pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &\leq I(U_1;Y|U_2) - I(U_1;S|U_2) \\ R_2 &\leq I(U_2;Y|U_1) - I(U_2;S|U_1) \\ R_1 + R_2 &\leq I(U_1,U_2;Y) - I(U_1,U_2;S) \end{aligned}$$

for some $P_{X_1,X_2,U_1,U_2|S} = P_{X_1,U_1|S} P_{X_2,U_2|S}$, is achievable.

State dependent MAC (cont'd)

Relevant work

- Gel'fand and Pinsker, 1984 – Derived the capacity region of

$$Y_i = X_{1,i} + X_{2,i} + S_i + V_i$$

where

  - $\{V_i\}_i$ is additive white Gaussian noise,
  - $\{S_i\}_i$ is additive white Gaussian interference, known non-causally at the two encoders, and independent of $\{V_i\}_i$
  - Input power constraints: $\frac{1}{n}\sum_{i=1}^{n} x_{j,i}^2 \leq \Gamma_j$.

It is shown that this channel satisfies the WDP property: no loss in rate in incurred due to the interference, provided it is known non causally at the two encoders.

Solved independently by Kim, Sutivong, & Sigurjonsson, ISIT 2004.

Method of proof –

  - Specific substitution in the achievable region
  - Successive application of Costa's WDP coding technique, to achieve points on the boundary of the capacity region.

State dependent MAC *(cont'd)*

Relevant work (cont'd)

- Das & Narayan, 2002 – Derived non single-letter expression for the capacity region of the general state dependent MAC, with various degrees of of SI at the encoders and decoder.

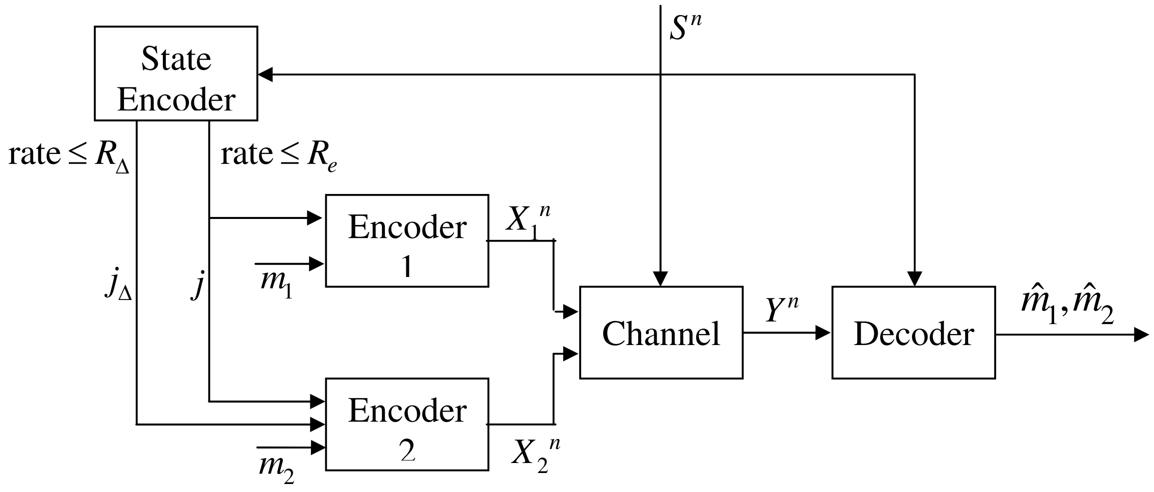
  Provided results for Gaussian fading MAC with causal SI at the encoders and fully informed decoder. Fading process need not be memoryless.

- Kotagiri & Laneman, 2005 – Provided bounds for special cases: only part of the encoders have SI.

- Cemal & Steinberg, 2005 – Investigated MAC with rate-limited SI at the encoders, and full SI at the decoder. Bounds.

  Single letter expression for capacity region when the SI at the encoders is *degraded*, or *nested*.

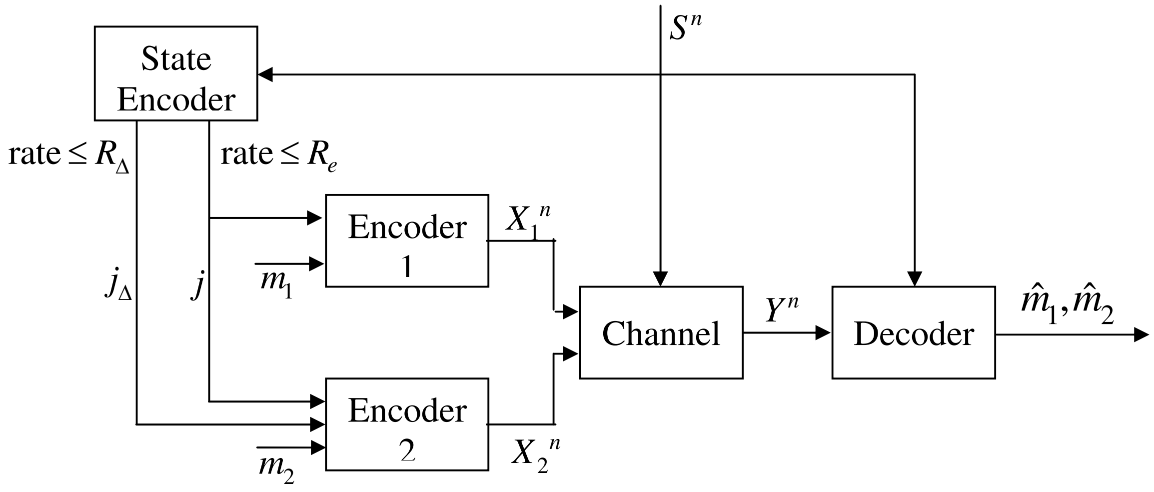
MAC with rate-limited SI, fully informed decoder



- State encoder generates two rate limited descriptions of the state S^n:
  - Common SI, $j = f_s(s^n)$, of rate R_e
  - Additional SI, $j_\triangle$, of rate $R_\triangle$, provided only to user 2.

  $\implies \quad \text{SI}_{\text{user1}} \subseteq \text{SI}_{\text{user2}}$.
- j and $j_\triangle$ are sent to the intended users via noiseless links (wayside channels)

MAC, rate-limited SI, informed dec *(cont'd)*



- Encoder 1: $x_1^n = f_1(m_1, j)$
- Encoder 2: $x_2^n = f_2(m_2, j, j_\Delta)$
- Each transmitted block, x_i^n $(i = 1, 2)$, is subject to input constraint:

$$\phi_i(x_i^n) \triangleq \frac{1}{n}\sum_{l=1}^{n} \phi_i(x_{i,l}) \leq \Gamma_i .$$

- The decoder is fully informed

MAC, rate-limited SI, informed dec *(cont'd)*

Definition:
An $(n, R_1, R_2, R_e, R_\Delta, \Gamma_1, \Gamma_2, \epsilon)$ code for the MAC with degraded SI at the encoders and fully informed decoder consists of two state-encoder mappings

$$f_s : \mathcal{S}^n \rightarrow \left\{1, 2, \ldots, 2^{nR_e}\right\}$$

$$f_{s,\Delta} : \mathcal{S}^n \rightarrow \left\{1, 2, \ldots, 2^{nR_\Delta}\right\}$$

two channel encoders

$$f_1 : \left\{1, 2, \ldots, 2^{nR_1}\right\} \times \left\{1, 2, \ldots, 2^{nR_e}\right\} \rightarrow \mathcal{X}_1^n$$

$$f_2 : \left\{1, 2, \ldots, 2^{nR_2}\right\} \times \left\{1, 2, \ldots, 2^{nR_e}\right\} \times \left\{1, 2, \ldots, 2^{nR_\Delta}\right\} \rightarrow \mathcal{X}_2^n$$

and a decoding function

$$g : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \left\{1, 2, \ldots, 2^{nR_1}\right\} \times \left\{1, 2, \ldots, 2^{nR_2}\right\}$$

MAC, rate-limited SI, informed dec (cont'd)

Definition – cont'd

such that

$$\phi_1(x_1^n(m_1,j)) = \frac{1}{n}\sum_{l=1}^{n}\phi_1(x_{1,l}^n(m_1,j)) \leq \Gamma_1$$

$$\phi_2(x_2^n(m_2,j,)) = \frac{1}{n}\sum_{l=1}^{n}\phi_2(x_{2,l}^n(m_2,j,j_\triangle)) \leq \Gamma_2$$

where $m_i \in \{1,2,\ldots,2^{nR_i}\}$ $i=1,2$, $j \in \{1,2,\ldots,2^{nR_e}\}$, $j_\triangle \in \{1,2,\ldots,2^{nR_\triangle}\}$, and where the probability of error P_e satisfies

$$P_e = 2^{-n(R_1+R_2)}\sum_{m_1=1}^{2^{nR_1}}\sum_{m_2=1}^{2^{nR_2}}\sum_{s^n} P_{S^n}(s^n)\sum_{y^n:g(y^n,s^n)\neq(m_1,m_2)}$$

$$P_{Y|X_1,X_2,S}(y^n|f_1(m_1,f_s(s^n)),f_2(m_2,f_s(s^n),f_{s,\triangle}(s^n)),s^n)$$

$$\leq \epsilon.$$

MAC, rate-limited SI, informed dec (cont'd)

The sextuple $(R_1, R_2, R_e, R_\triangle, \Gamma_1, \Gamma_2)$ is achievable if for any $\epsilon > 0$ and sufficiently large n, there exits an $(n, R_1, R_2, R_e, R_\triangle, \Gamma_1, \Gamma_2, \epsilon)$ code.

Capacity region $\mathcal{C}(\mathbf{\Gamma}, R_e, R_\triangle)$ – For a given side information rate pair $(R_e, R_\triangle)$ and input constraints $\mathbf{\Gamma} = (\Gamma_1, \Gamma_2)$, the capacity region $\mathcal{C}(\mathbf{\Gamma}, R_e, R_\triangle)$ is the closure of the set of all rate pairs (R_1, R_2) such that $(R_1, R_2, R_e, R_\triangle, \Gamma_1, \Gamma_2)$ is achievable.

MAC, rate-limited SI, informed dec (cont'd)

To give the main result for this channel, define the region $\mathcal{C}^*(\mathbf{\Gamma}, R_e, R_\triangle)$

For given $\mathbf{\Gamma} = (\Gamma_1, \Gamma_2)$ and $(R_e, R_\triangle)$, let $\mathcal{C}^*(\mathbf{\Gamma}, R_e, R_\triangle)$ be the set of all rate pairs (R_1, R_2) for which there exist random variables $(S_0, S_\triangle, X_1, X_2)$ such that the following conditions hold simultaneously:

1. The rates R_e and $R_\triangle$ satisfy

$$\begin{aligned} R_e &\geq I(S_0; S) \\ R_\triangle &\geq I(S_\triangle; S|S_0) \end{aligned}$$

2. The forward channel rates R_1 and R_2 satisfy

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, S, S_0, S_\triangle) \\ R_2 &\leq I(X_2; Y|X_1, S, S_0, S_\triangle) \\ R_1 + R_2 &\leq I(X_1, X_2; Y|S, S_0, S_\triangle) \end{aligned}$$

3. Input constraint functions satisfy

$$E[\phi_i(X_i)] \leq \Gamma_i, \ i = 1, 2$$

MAC, rate-limited SI, informed dec *(cont'd)*

4. The following Markov chains hold

$$X_1 \leftrightarrow S_0 \leftrightarrow (S, S_\Delta, X_2)$$
$$X_2 \leftrightarrow (S_0, S_\Delta) \leftrightarrow (S, X_1)$$
$$(S_0, S_\Delta) \leftrightarrow (S, X_1, X_2) \leftrightarrow Y$$

5. The auxiliary alphabets $\mathcal{S}_0$ and $\mathcal{S}_\Delta$ satisfy

$$|\mathcal{S}_0| \leq |\mathcal{S}| + 6$$
$$|\mathcal{S}_\Delta| \leq |\mathcal{S}|(|\mathcal{S}| + 6) + 5$$

MAC, rate-limited SI, informed dec (cont'd)

Theorem 16 (Cemal & Steinberg, 2005) *For any state dependent discrete memoryless MAC, with degraded rate-limited CSIT, and full CSIR,* $\mathcal{C}(\mathbf{\Gamma}, R_e, R_\Delta) = \mathcal{C}^*(\mathbf{\Gamma}, R_e, R_\Delta)$.

- Direct and Converse Proofs in the spirit of

  1. Hierarchical source coding problem ([Koshelev, 1980], [Rimoldi, 1994])
  2. Coding for single user channels with rate limited SI and informed decoder ([Heegard & El Gamal 1983], [Rosenzweig, Steinberg, Shamai 2003])
  3. Coding for multiple-access channel

MAC, rate-limited SI, informed dec (cont'd)

- **Special Case**: The capacity region with *common* rate-limited CSIT and full CSIR, ($R_\Delta = 0$), $\underline{\mathcal{C}_c(\mathbf{\Gamma}, R_e)}$, is given by the set of all rate pairs (R_1, R_2) for which there exist random variables (S_0, X_1, X_2) such that

1. $R_e \geq I(S_0; S)$
2. $R_1 \leq I(X_1; Y|X_2, S, S_0)$
   $R_2 \leq I(X_2; Y|X_1, S, S_0)$
   $R_1 + R_2 \leq I(X_1, X_2; Y|S, S_0)$
3. $E[\phi_i(X_i)] \leq \Gamma_i, \ i = 1, 2$
4. Markov chains $X_1 \leftrightarrow S_0 \leftrightarrow (S, X_2)$, $X_2 \leftrightarrow S_0 \leftrightarrow (S, X_1)$ and $S_0 \leftrightarrow (S, X_1, X_2) \leftrightarrow Y$ hold
5. $|\mathcal{S}_0| \leq |\mathcal{S}| + 5$