Source-Channel Communication in Networks: Separation Theorems and Beyond

Jun Chen
Electrical and Computer Engineering
McMaster University

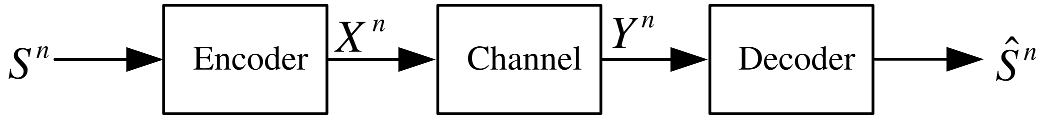
Joint work with Kia Khezeli, Lin Song, Chao Tian, Suhas Diggavi, and Shlomo Shamai

Outline

- Optimality of the source-channel separation architecture for lossy source coding in general networks
- The source broadcast problem: Application of the source-channel separation theorem as a converse method
- Other converse methods for the source broadcast problem

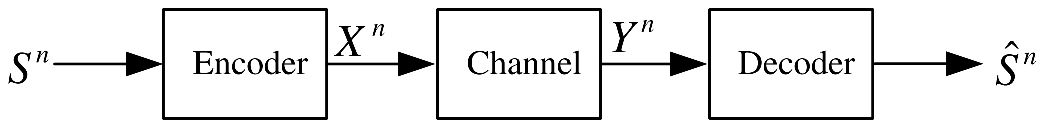
Source-Channel Separation Theorem

- Source-channel communication

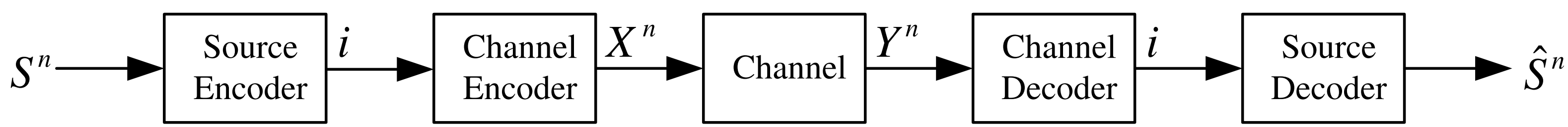


Source-Channel Separation Theorem

- Source-channel communication



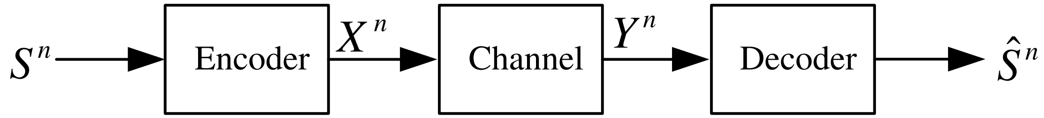
- Separation theorem (Shannon 48)



For any achievable end-to-end distortion D,

$$R(D) \leq C.$$

Two Proofs



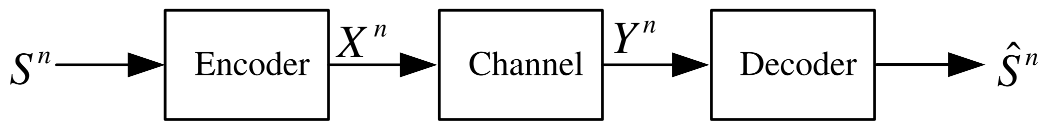
- Standard proof
  - Information-theoretic definition of channel capacity and rate-distortion function

$$C = \max_{p_X} I(X;Y)$$

$$R(D) = \min_{p_{\hat{S}|S}:\mathbb{E}[d(S,\hat{S})]\leq D} I(S;\hat{S})$$

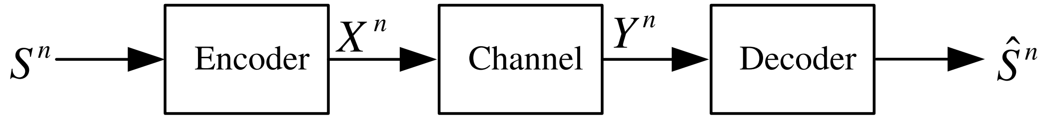
  - Converse theorem of channel coding: $I(X^n;Y^n) \leq nC$
  - Converse theorem of lossy source coding: $I(S^n;\hat{S}^n) \geq nR(D)$
  - Data processing inequality: $I(X^n;Y^n) \geq I(S^n;\hat{S}^n)$

Two Proofs

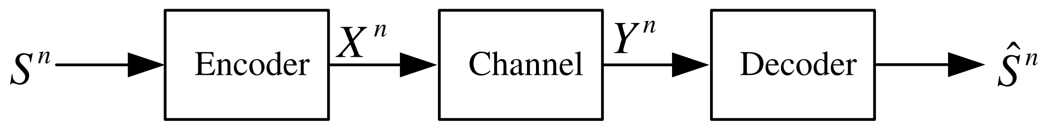


- Standard proof
  - Information-theoretic definition of channel capacity and rate-distortion function
  $$C = \max_{p_X} I(X;Y)$$
  $$R(D) = \min_{p_{\hat{S}|S}:\mathbb{E}[d(S,\hat{S})]\leq D} I(S;\hat{S})$$
  - Converse theorem of channel coding: $I(X^n;Y^n) \leq nC$
  - Converse theorem of lossy source coding: $I(S^n;\hat{S}^n) \geq nR(D)$
  - Data processing inequality: $I(X^n;Y^n) \geq I(S^n;\hat{S}^n)$
- Alternative proof
  - Operational definition of channel capacity and rate-distortion function
  - Achievability theorem of channel coding: $I(X^n;Y^n) \leq nC$
  - Achievability theorem of lossy source coding: $I(S^n;\hat{S}^n) \geq nR(D)$
  - Data processing inequality: $I(X^n;Y^n) \geq I(S^n;\hat{S}^n)$

More Proofs



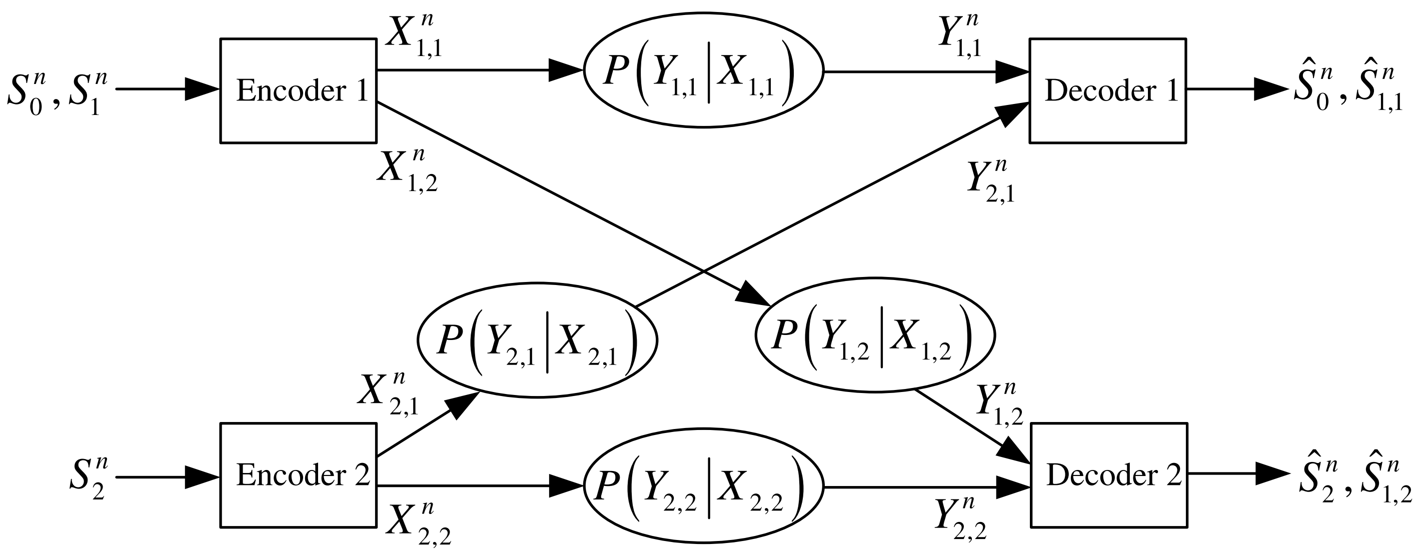
- Channel-centered proof
  - View $p_{Y^n|X^n}$ as a communication channel: $I(X^n;Y^n) \leq nC$
  - View $p_{Y^n|X^n}$ as a test channel: $I(X^n;Y^n) \geq R(p_{X^n,Y^n})$
  - $R(p_{X^n,Y^n}) \geq nR(D)$



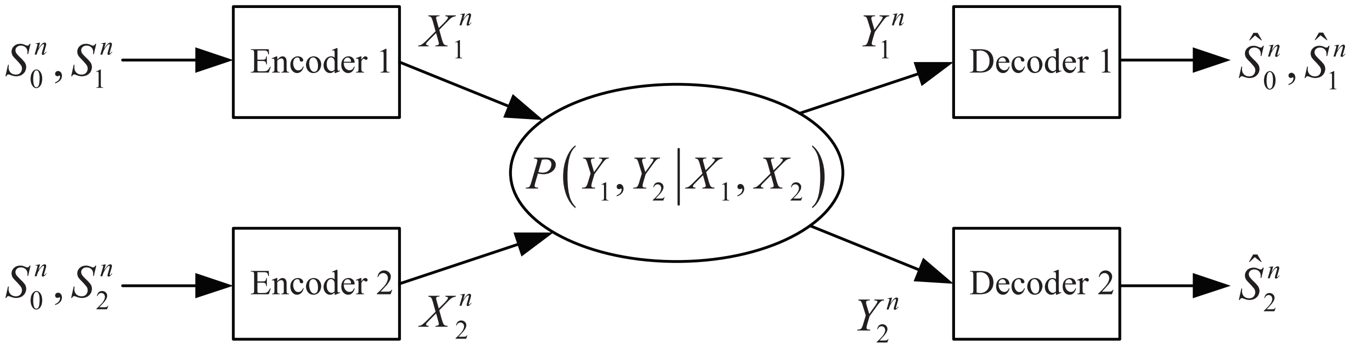
- Channel-centered proof
  - View $p_{Y^n|X^n}$ as a communication channel: $I(X^n;Y^n) \le nC$
  - View $p_{Y^n|X^n}$ as a test channel: $I(X^n;Y^n) \ge R(p_{X^n,Y^n})$
  - $R(p_{X^n,Y^n}) \ge nR(D)$

- Source-centered proof
  - View $p_{\hat{S}^n|S^n}$ as a test channel: $I(S^n;\hat{S}^n) \ge nR(D)$
  - View $p_{\hat{S}^n|S^n}$ as a communication channel: $I(S^n;\hat{S}^n) \le C(p_{\hat{S}^n|S^n})$
  - $C(p_{\hat{S}^n|S^n}) \le nC$

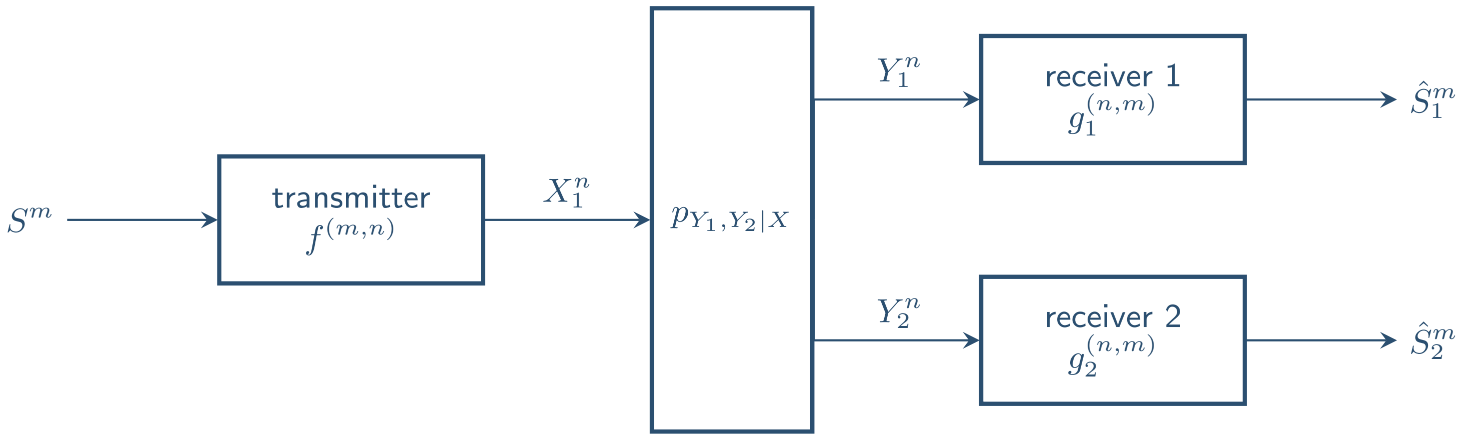
- Optimality: The memoryless sources at source nodes are arbitrarily correlated, each of which is to be reconstructed at possibly multiple destinations within certain distortions, but the channels in this network are synchronized, orthogonal and memoryless point-to-point channels.



- Optimality: The memoryless sources are mutually independent, each of which is to be reconstructed only at one destination within a certain distortion, but the channels are general, including multi-user channels such as multiple access, broadcast, interference and relay channels, possibly with feedback.



The Source Broadcast Problem



- We say $(\kappa, \mathcal{D}_1, \mathcal{D}_2)$ is achievable if

$$\frac{n}{m} \leq \kappa,$$

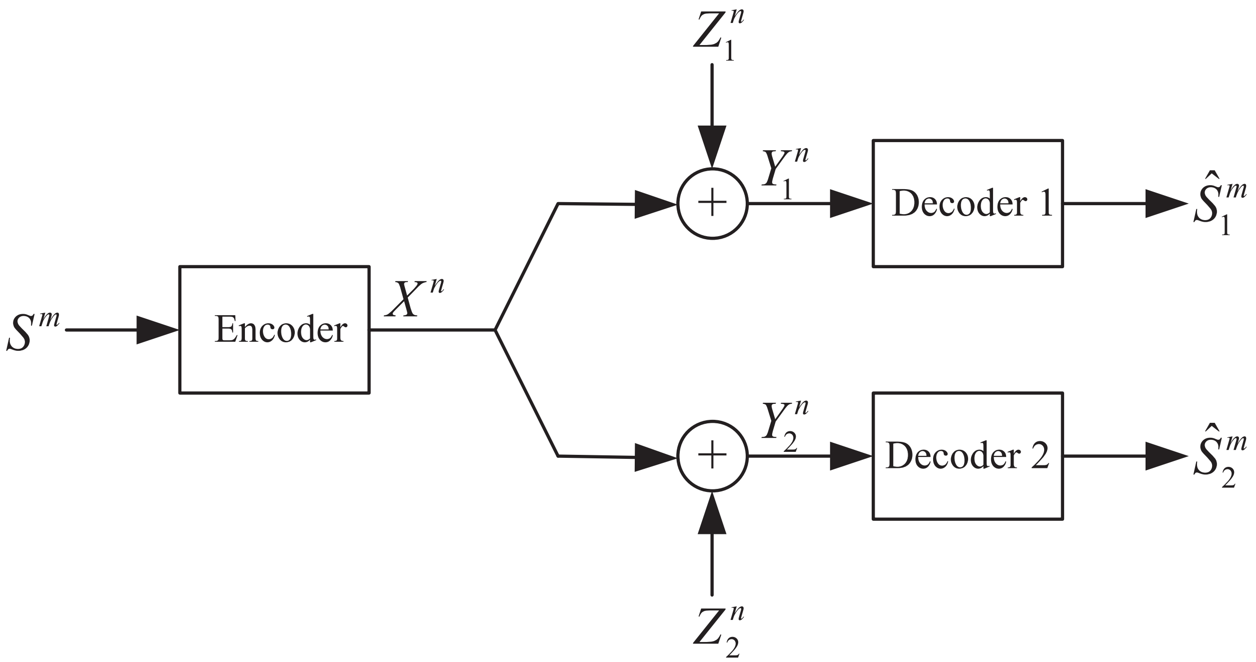
$$\frac{1}{m}\sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)} \in \mathcal{D}_i, \quad i = 1, 2. \quad (*)$$

Remark: $(*)$ is more general than conventional distortion constraints since

$$\frac{1}{m}\sum_{t=1}^{m} \mathbb{E}[d_i(S(t), S_i(t))] = \mathbb{E}[d(S, \hat{S}_i)],$$

where $p_{S,\hat{S}_i} = \frac{1}{m}\sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)}$, $i = 1, 2$.

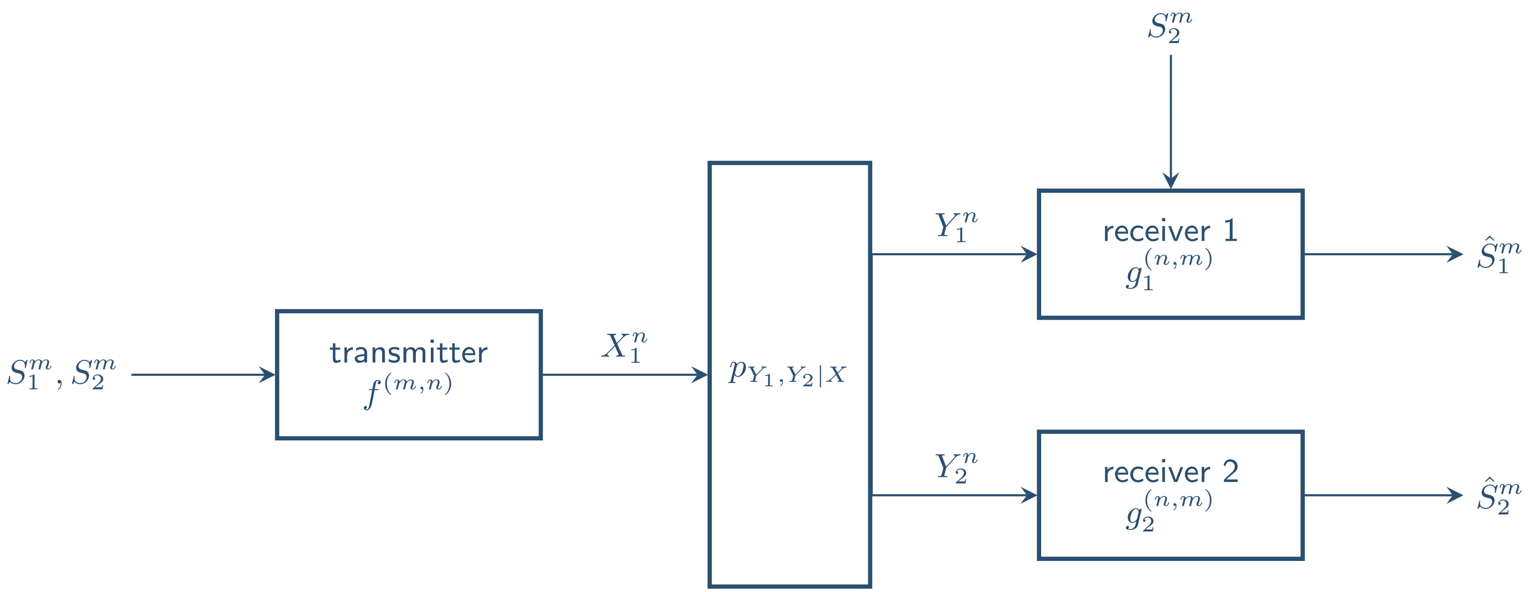
Gaussian Source over Gaussian Broadcast Channel



- Tradeoff between the transmit power P, the bandwidth mismatch factor κ, and the achievable reconstruction distortion pair (d_1, d_2)
- A mysterious auxiliary random variable (Reznic, Feder, and Zamir 06): $S+U$, where U is independent of everything else.

$$P \geq \sup_{\sigma_U^2>0} N_1\left(\frac{\sigma_S^2(d_1+\sigma_U^2)}{d_1(d_2+\sigma_U^2)}\right)^{\frac{1}{\kappa}} + (N_2-N_1)\left(\frac{\sigma_S^2+\sigma_U^2}{d_2+\sigma_U^2}\right)^{\frac{1}{\kappa}} - N_2$$

Source Broadcast with Receiver Side Information

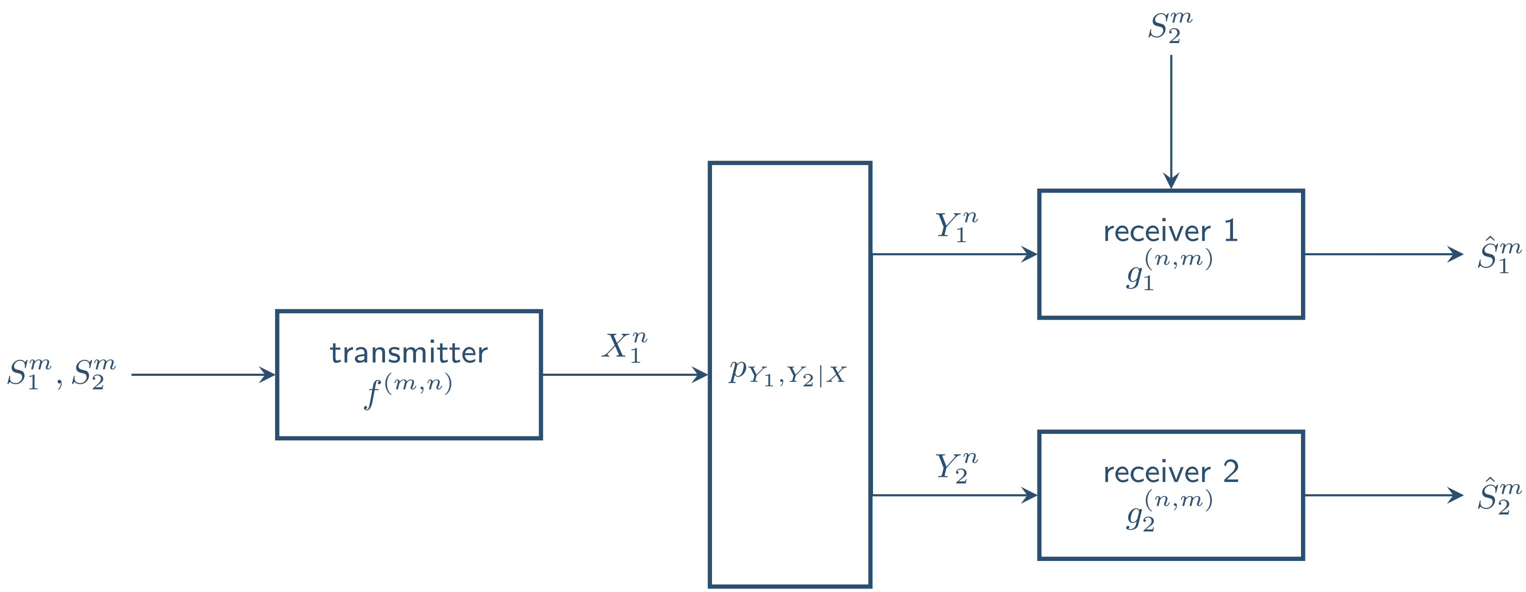


- We say $(\kappa, \mathcal{D}_1, \mathcal{D}_2)$ is achievable if

$$\frac{n}{m} \leq \kappa,$$

$$\frac{1}{m}\sum_{t=1}^{m} p_{S_1(t),S_2(t),\hat{S}_1(t)} \in \mathcal{D}_1,$$

$$\frac{1}{m}\sum_{t=1}^{m} p_{S_2(t),\hat{S}_2(t)} \in \mathcal{D}_2.$$



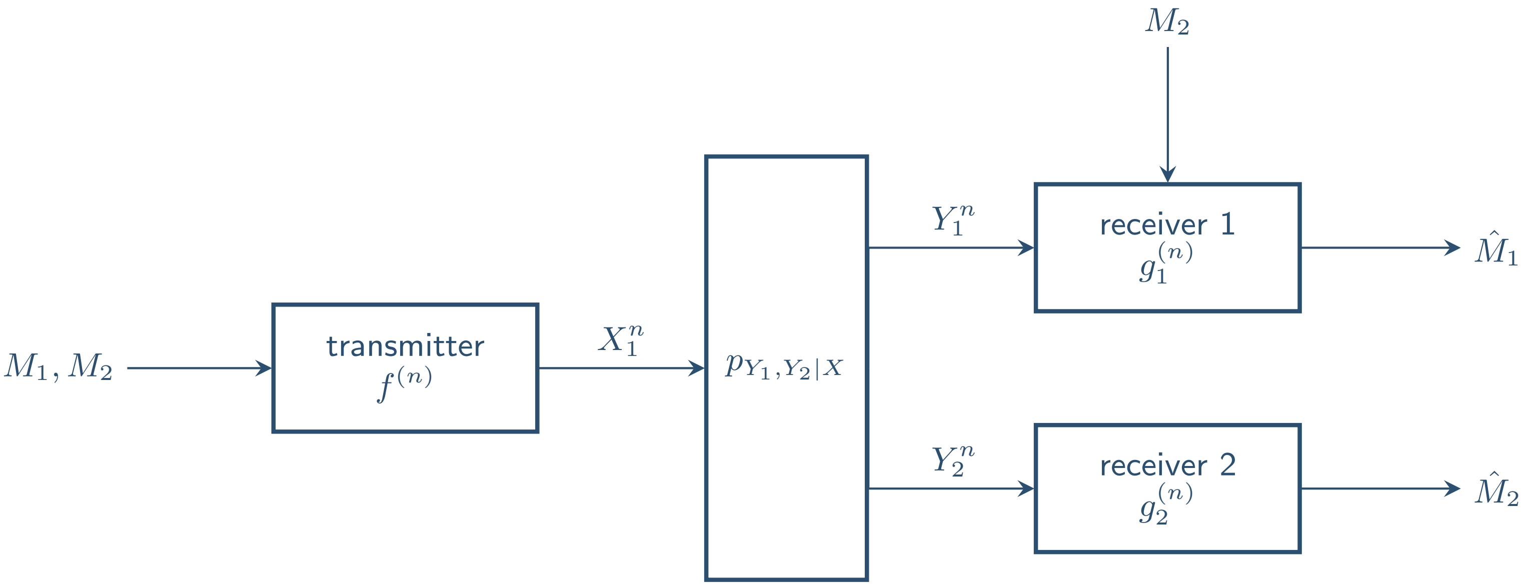
- $(\kappa, \mathcal{D}_1, \mathcal{D}_2)$ is achievable $\Longleftrightarrow$ $(R_{S_1|S_2}(\mathcal{D}_1), R_{S_2}(\mathcal{D}_2)) \in \kappa\mathcal{C}_{1|2}(p_{Y_1,Y_2|X})$, where

$$R_{S_1|S_2}(\mathcal{D}_1) = \min_{p_{S_1,S_2,\hat{S}_1} \in \mathcal{D}_1} I(S_1;\hat{S}_1|S_2),$$

$$R_{S_2}(\mathcal{D}_2) = \min_{p_{S_2,\hat{S}_2} \in \mathcal{D}_2} I(S_2;\hat{S}_2),$$

and $C_{1|2}(p_{Y_1,Y_2|X})$ is the capacity region of broadcast channel $p_{Y_1,Y_2|X}$ when the message intended for receiver 2 is available at receiver 1.

Broadcast Channel with Receiver Side Information



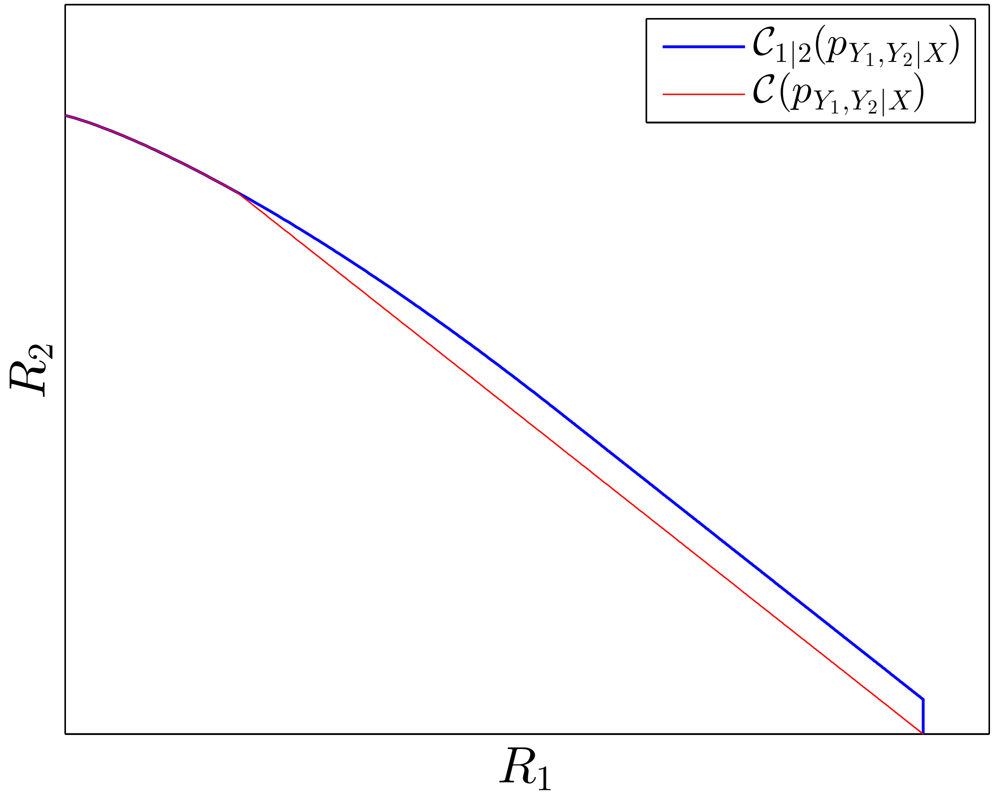
- Capacity region $\mathcal{C}_{1|2}(p_{Y_1,Y_2|X})$ (Kramer and Shamai 07)

$$\begin{aligned} R_1 &\leq I(X;Y_1), \\ R_2 &\leq I(V;Y_2), \\ R_1 + R_2 &\leq I(X;Y_1|V) + I(V;Y_2) \end{aligned}$$

for some $p_{V,X}$.

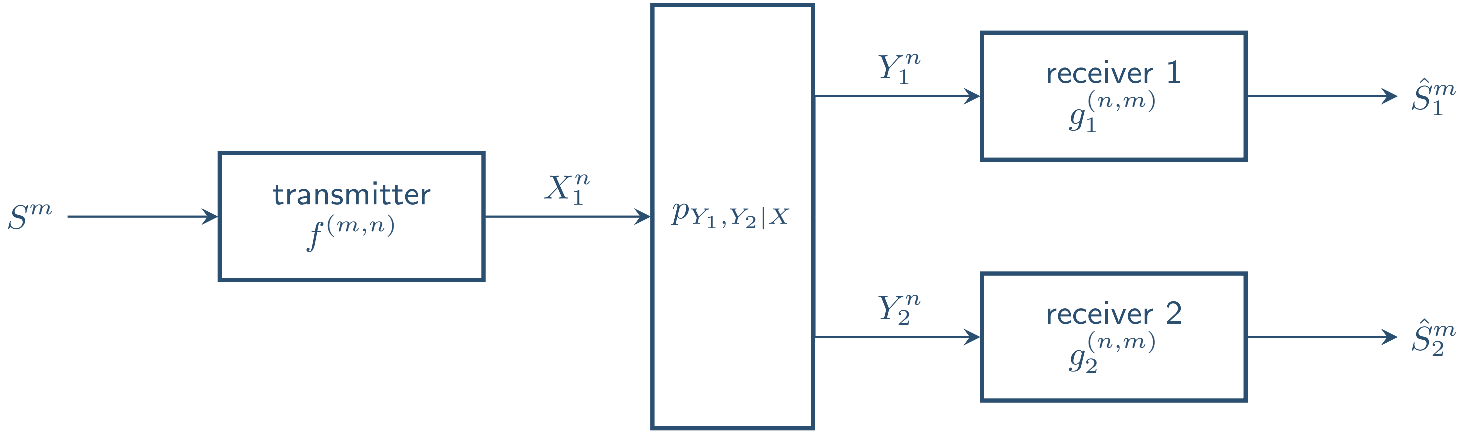
- $\mathcal{C}_{1|2}(p_{Y_1,Y_2|X}) = \mathcal{C}(p_{Y_1,Y_2|X})$ if Y_1 is less noisy than Y_2, but not necessarily so if Y_1 is more capable than Y_2.

BEC-BSC



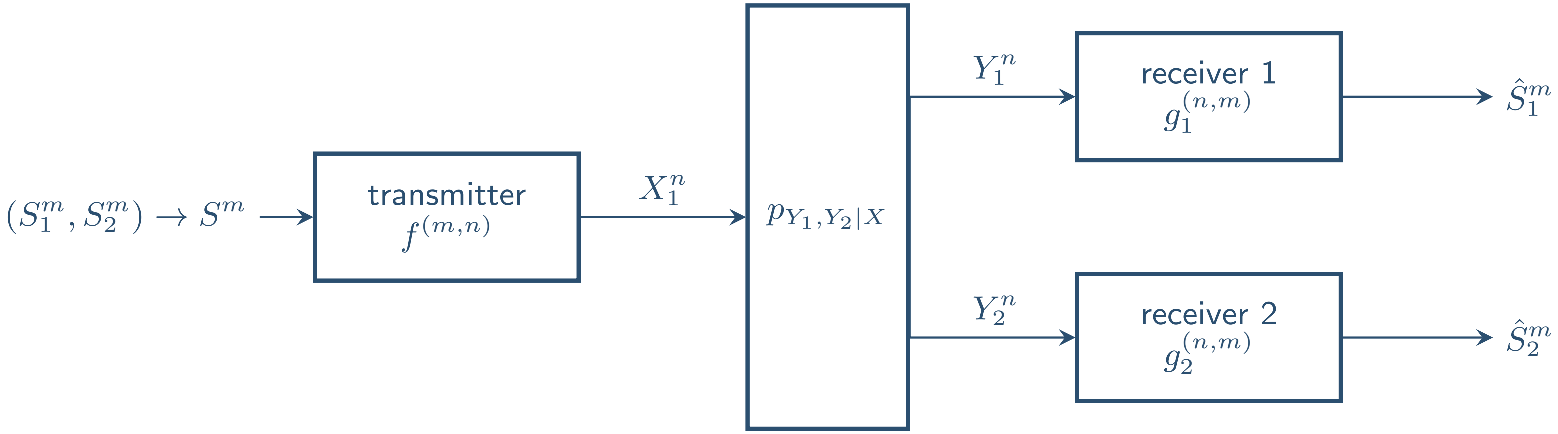
A Reduction Argument

- Introduction of remote source $\{(S_1(t), S_2(t))\}_{t=1}^{\infty}$
  Given $\frac{1}{m}\sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)} \in \mathcal{D}_i$, $i = 1, 2$, one can compute the induced $\frac{1}{m}\sum_{t=1}^{m} p_{S_1(t),S_2(t),\hat{S}_1(t)}$ and $\frac{1}{m}\sum_{t=1}^{m} p_{S_2(t),\hat{S}_2(t)}$. So the separation theorem can be applied.



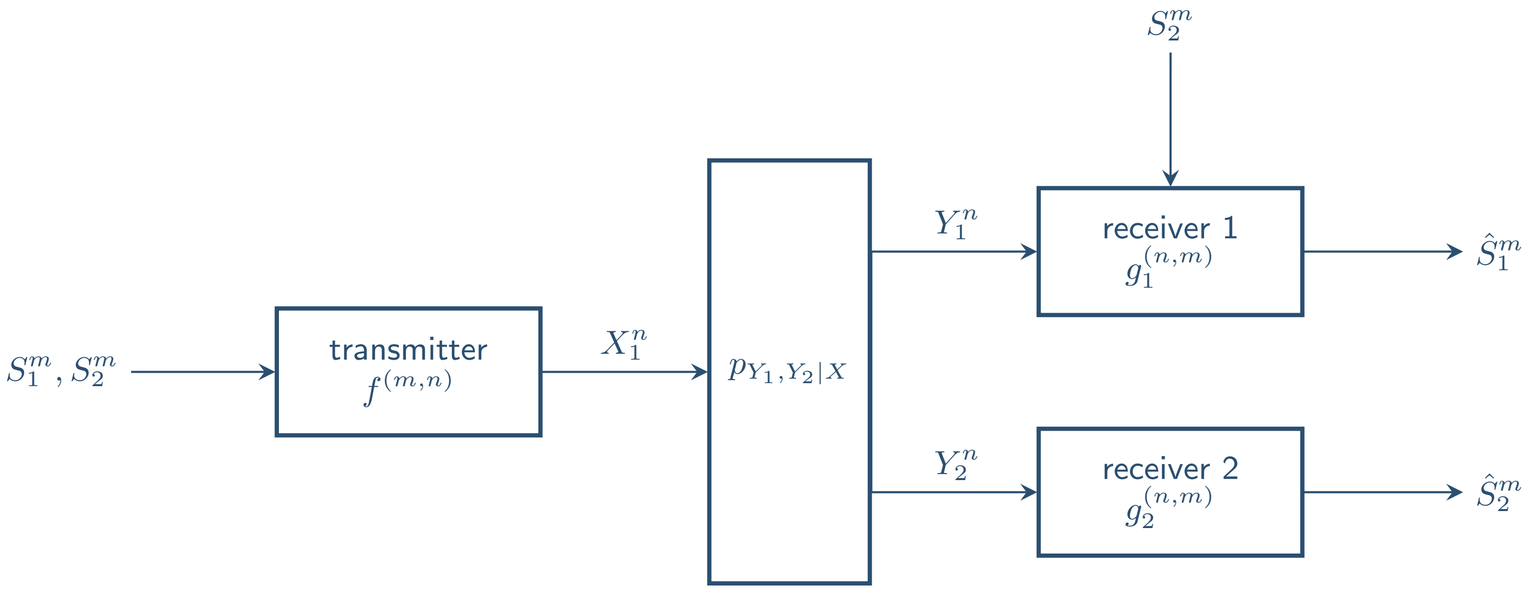
A Reduction Argument

- Introduction of remote source $\{(S_1(t), S_2(t))\}_{t=1}^{\infty}$
  Given $\frac{1}{m}\sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)} \in \mathcal{D}_i$, $i = 1, 2$, one can compute the induced $\frac{1}{m}\sum_{t=1}^{m} p_{S_1(t),S_2(t),\hat{S}_1(t)}$ and $\frac{1}{m}\sum_{t=1}^{m} p_{S_2(t),\hat{S}_2(t)}$. So the separation theorem can be applied.



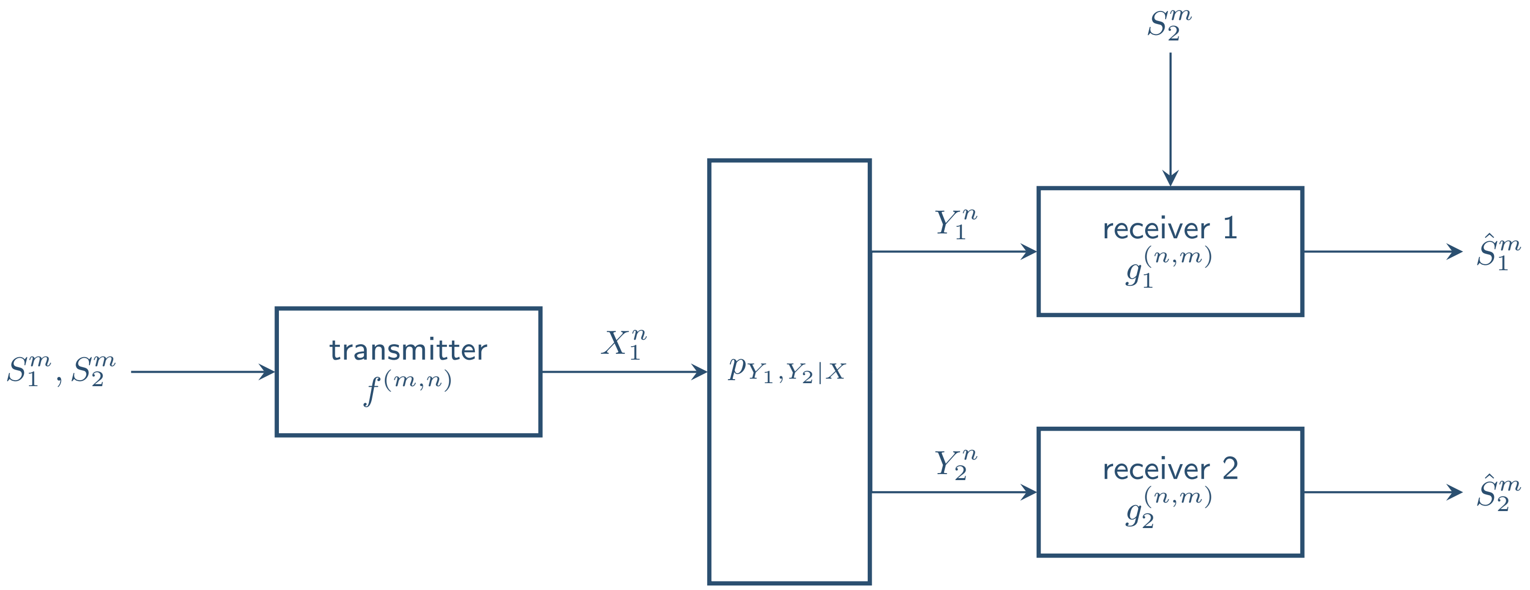
A Reduction Argument

- Introduction of remote source $\{(S_1(t), S_2(t))\}_{t=1}^{\infty}$
  Given $\frac{1}{m}\sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)} \in \mathcal{D}_i$, $i = 1, 2$, one can compute the induced $\frac{1}{m}\sum_{t=1}^{m} p_{S_1(t),S_2(t),\hat{S}_1(t)}$ and $\frac{1}{m}\sum_{t=1}^{m} p_{S_2(t),\hat{S}_2(t)}$. So the separation theorem can be applied.



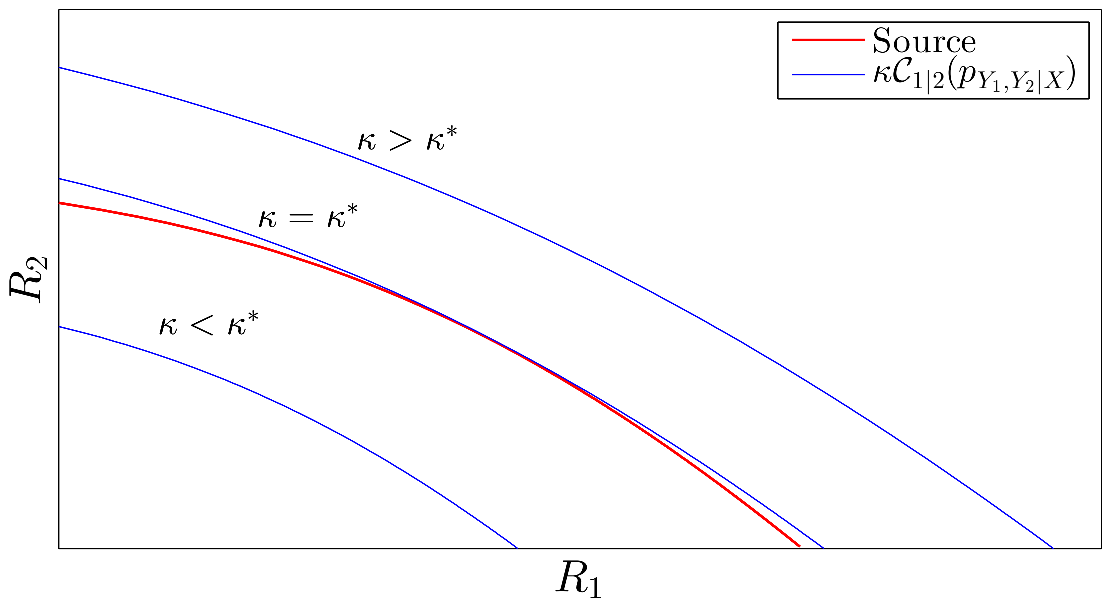
A Reduction Argument

- Introduction of remote source $\{(S_1(t), S_2(t))\}_{t=1}^{\infty}$
  Given $\frac{1}{m}\sum_{t=1}^{m} p_{S(t),\hat{S}_i(t)} \in \mathcal{D}_i$, $i = 1, 2$, one can compute the induced $\frac{1}{m}\sum_{t=1}^{m} p_{S_1(t),S_2(t),\hat{S}_1(t)}$ and $\frac{1}{m}\sum_{t=1}^{m} p_{S_2(t),\hat{S}_2(t)}$. So the separation theorem can be applied.

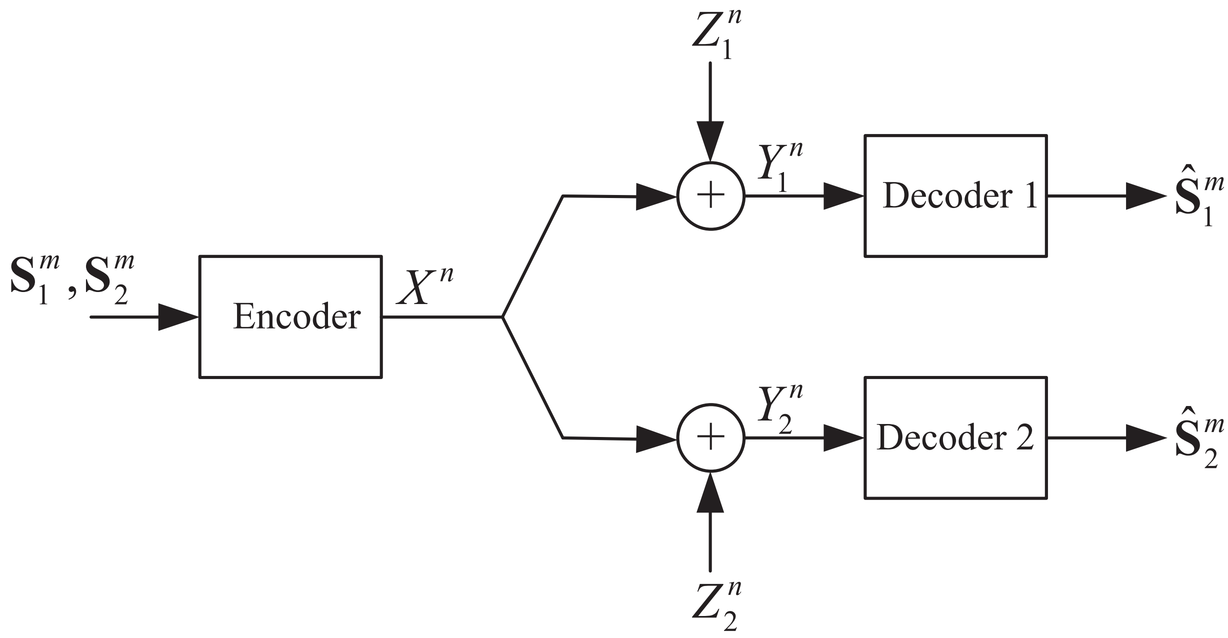


- There exists some $p_{S,\hat{S}_1,\hat{S}_2} = p_S p_{\hat{S}_1,\hat{S}_2|S}$ with $p_{S,\hat{S}_i} \in \mathcal{D}_i$, $i = 1, 2$ such that $(I(S_1; \hat{S}_1|S_2), I(S_2; \hat{S}_2)) \in \kappa\mathcal{C}_{1|2}(p_{Y_1,Y_2|X})$ for all $p_{S_1,S_2|S}$. Moreover, there is no loss of generality in choosing $S_1 = S$.

- Let $S_1 = S$ and $S_2 = S + U$, where $U \sim \mathcal{N}(0, \sigma_U^2)$.



- Tradeoff between the transmit power P, the bandwidth mismatch factor κ, and the achievable reconstruction distortion pair $(\mathbf{D}_1, \mathbf{D}_2)$



- Scalar case without bandwidth mismatch
  - Source-channel separation is suboptimal (Gao and Tuncel 11).
  - Uncoded scheme is optimal at low SNR (Bross, Lapidoth, and Tinguely 10).
  - Hybrid scheme is optimal (Tian, Diggavi, and Shamai 11).

Characterization of the Power-Bandwidth-Distortion Tradeoff

- Source: $(\mathbf{S}_1, \mathbf{S}_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{S}})$ with $\mathbf{S}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{S}_i})$, $i = 1, 2$
- Channel: $Z_i \sim \mathcal{N}(0, N_i)$, $i = 1, 2$, with $N_1 < N_2$

- Source: $(\mathbf{S}_1, \mathbf{S}_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{S}})$ with $\mathbf{S}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{S}_i})$, $i = 1, 2$
- Channel: $Z_i \sim \mathcal{N}(0, N_i)$, $i = 1, 2$, with $N_1 < N_2$
- A necessary condition:

$$P \geq \min_{\mathbf{\Theta}_1, \mathbf{\Theta}_2} \sup_{\mathbf{\Sigma}_{\mathbf{U}} \succ \mathbf{0}} N_1 \left(\frac{|\mathbf{\Sigma}_{\mathbf{S}}||\mathbf{\Theta}_1 + \mathbf{\Sigma}_{\mathbf{U}}|}{|\mathbf{\Theta}_1||\mathbf{\Theta}_2 + \mathbf{\Sigma}_{\mathbf{U}}|} \right)^{\frac{1}{\kappa}} + (N_2 - N_1) \left(\frac{|\mathbf{\Sigma}_{\mathbf{S}} + \mathbf{\Sigma}_{\mathbf{U}}|}{|\mathbf{\Theta}_2 + \mathbf{\Sigma}_{\mathbf{U}}|} \right)^{\frac{1}{\kappa}} - N_2,$$

where the minimization is over $\mathbf{\Theta}_i = \begin{pmatrix} \mathbf{\Theta}_{i,i} & * \\ * & * \end{pmatrix}$, $i = 1, 2$, subject to $\mathbf{\Sigma}_{\mathbf{S}} \succeq \mathbf{\Theta}_2 \succeq \mathbf{\Theta}_1 \succ \mathbf{0}$ and $\mathbf{\Theta}_{i,i} \preceq \mathbf{D}_i$, $i = 1, 2$.

- Source: $(\mathbf{S}_1, \mathbf{S}_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_S})$ with $\mathbf{S}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{S}_i})$, $i = 1, 2$
- Channel: $Z_i \sim \mathcal{N}(0, N_i)$, $i = 1, 2$, with $N_1 < N_2$
- A necessary condition:

$$P \geq \min_{\mathbf{\Theta}_1, \mathbf{\Theta}_2} \sup_{\mathbf{\Sigma_U} \succ \mathbf{0}} N_1 \left(\frac{|\mathbf{\Sigma_S}||\mathbf{\Theta}_1 + \mathbf{\Sigma_U}|}{|\mathbf{\Theta}_1||\mathbf{\Theta}_2 + \mathbf{\Sigma_U}|} \right)^{\frac{1}{\kappa}} + (N_2 - N_1) \left(\frac{|\mathbf{\Sigma_S} + \mathbf{\Sigma_U}|}{|\mathbf{\Theta}_2 + \mathbf{\Sigma_U}|} \right)^{\frac{1}{\kappa}} - N_2,$$

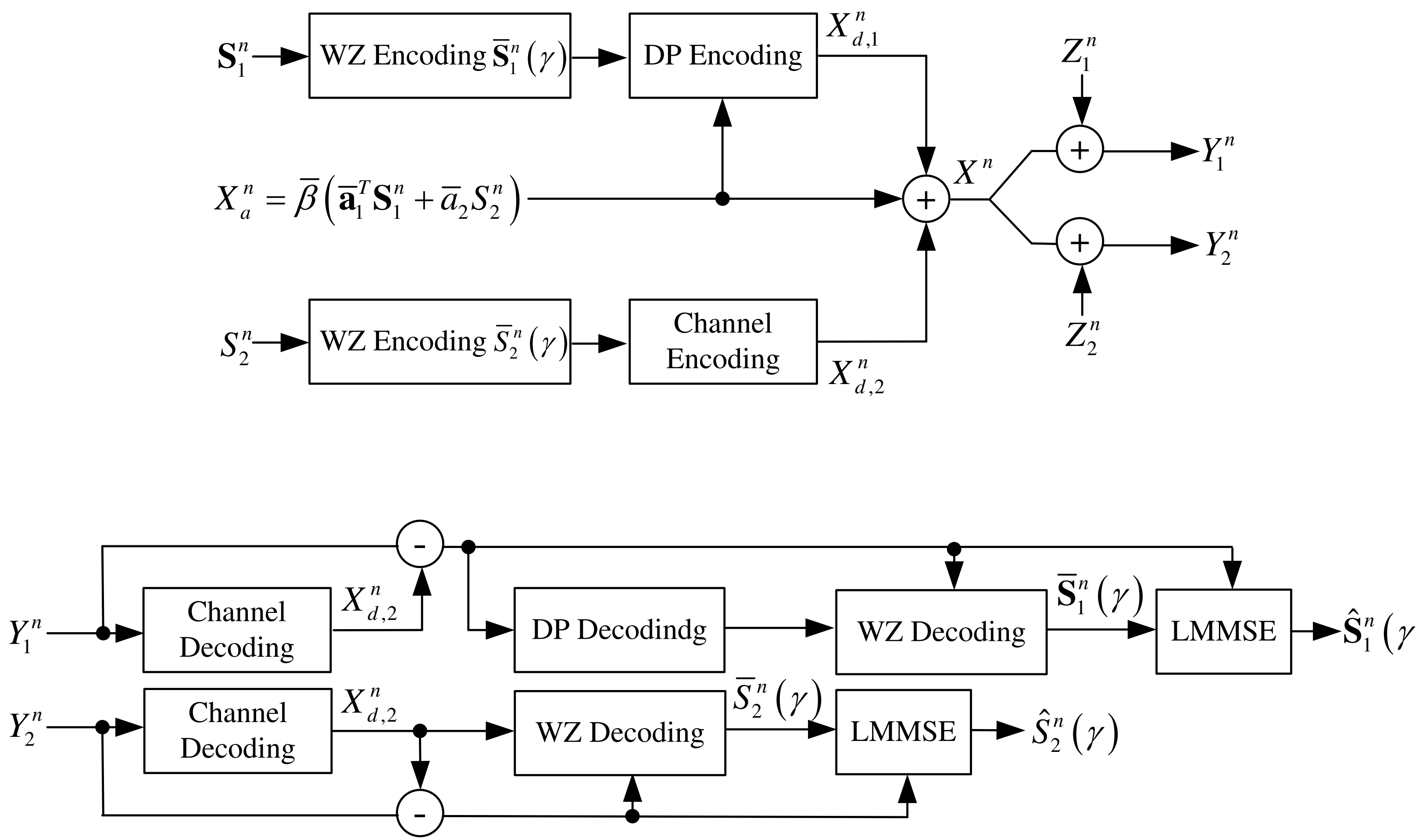
where the minimization is over $\mathbf{\Theta}_i = \begin{pmatrix} \mathbf{\Theta}_{i,i} & * \\ * & * \end{pmatrix}$, $i = 1, 2$, subject to $\mathbf{\Sigma_S} \succeq \mathbf{\Theta}_2 \succeq \mathbf{\Theta}_1 \succ \mathbf{0}$ and $\mathbf{\Theta}_{i,i} \preceq \mathbf{D}_i$, $i = 1, 2$.

- This bound is tight when S_2 is a scalar and $\kappa = 1$:

$$P \geq \sup_{\mathbf{\Sigma_U} \succ \mathbf{0}} N_1 \frac{|\mathbf{\Sigma_S} + \mathbf{\Sigma_U}|}{|\mathbf{D}_1 + \mathbf{\Sigma}_{\mathbf{U}_1}|(d_2 + \sigma_{U_2}^2)} + (N_2 - N_1) \frac{\sigma_{S_2}^2 + \sigma_{U_2}^2}{d_2 + \sigma_{U_2}^2} - N_2,$$

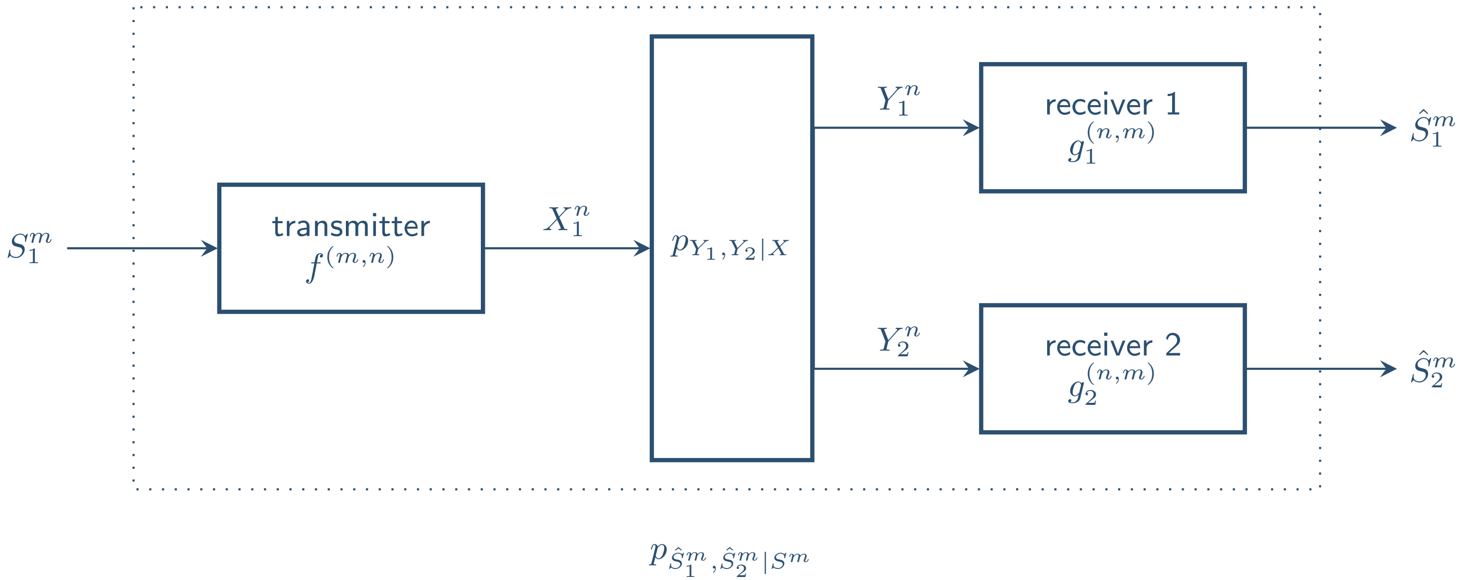
where $\mathbf{\Sigma_U} = \begin{pmatrix} \mathbf{\Sigma}_{\mathbf{U}_1} & * \\ * & \sigma_{U_2}^2 \end{pmatrix}$.

System Diagram



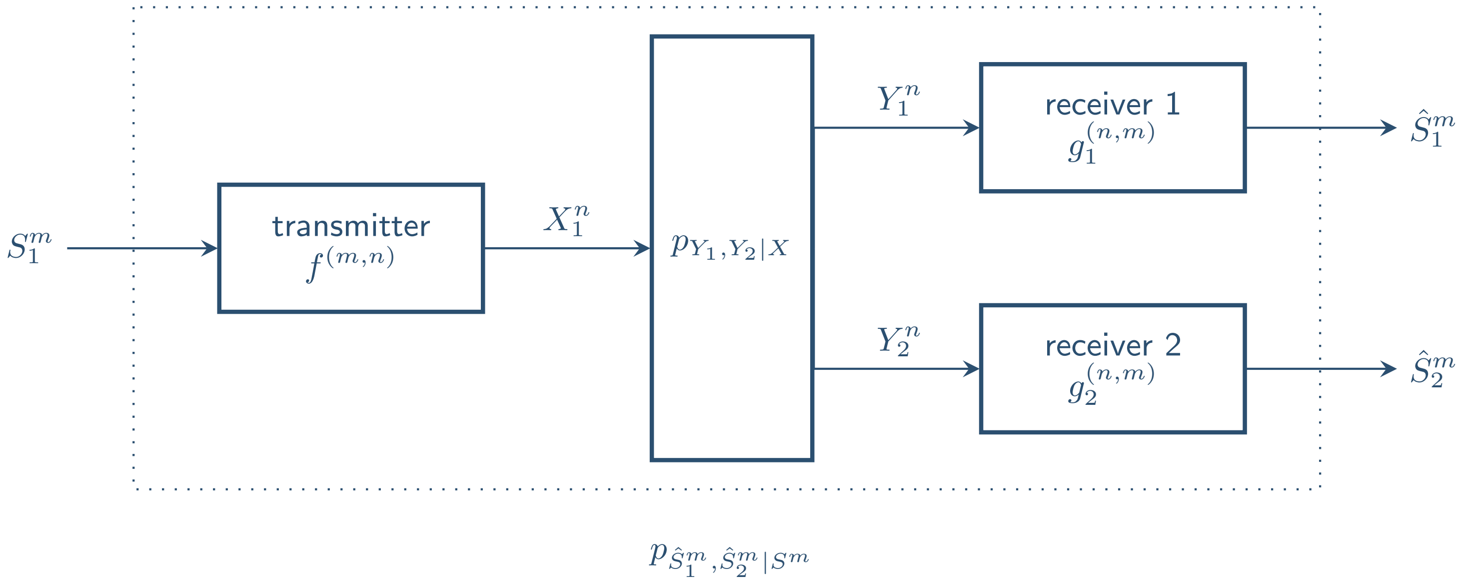
Source-channel separation theorems can be used to prove the optimality of non-separation based schemes (e.g., hybrid coding schemes) and determine performance limits even in scenarios where the separation architecture is suboptimal!

A Converse Method Based on Channel Comparison



$$\mathcal{C}(p_{\hat{S}_1^m,\hat{S}_2^m|S_1^m}) \subseteq n\mathcal{C}(p_{Y_1,Y_2|X})$$

A Converse Method Based on Channel Comparison



$$\mathcal{C}(p_{\hat{S}_1^m,\hat{S}_2^m|S_1^m}) \subseteq n\mathcal{C}(p_{Y_1,Y_2|X})$$

- A single-letter version: If $(\kappa, \mathcal{D}_1, \mathcal{D}_2)$ is achievable, then

$$\mathcal{C}(p_S, p_{\hat{S}_1,\hat{S}_2|S}) \subseteq \kappa\mathcal{C}(p_{Y_1,Y_2|X})$$

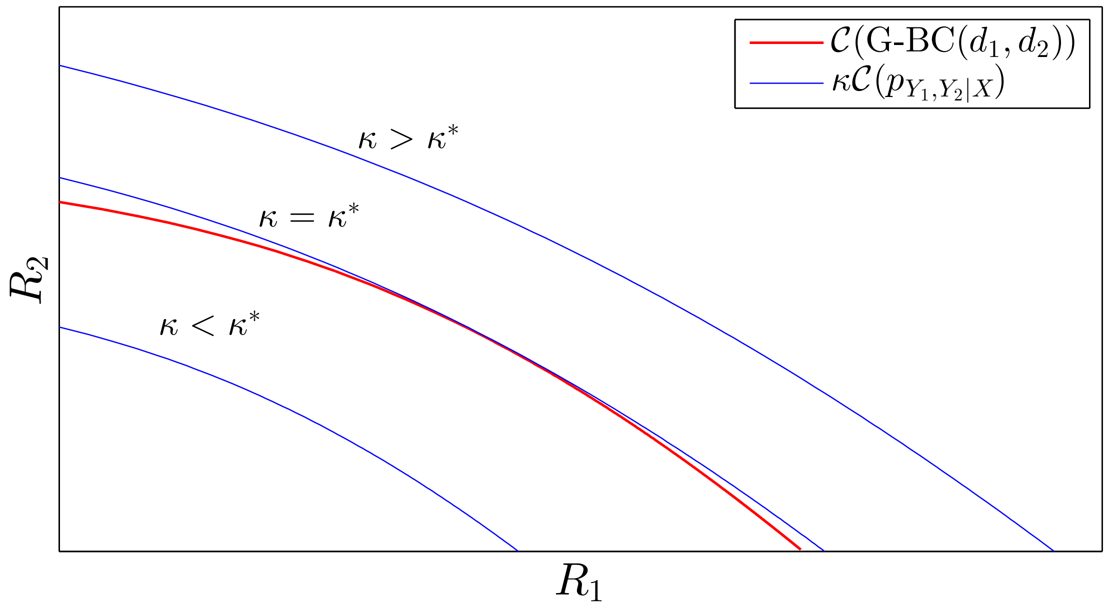
for some $p_{S,\hat{S}_1,\hat{S}_2} = p_S p_{\hat{S}_1,\hat{S}_2|S}$ with $p_{S,\hat{S}_i} \in \mathcal{D}_i$, $i = 1, 2$. A proper definition of $\mathcal{C}(p_S, p_{\hat{S}_1,\hat{S}_2|S})$ is needed. For simplicity, we can replace $\mathcal{C}(p_S, p_{\hat{S}_1,\hat{S}_2|S})$ with Marton's inner bound $\mathcal{C}_{\text{in}}(p_S, p_{\hat{S}_1,\hat{S}_2|S})$.

Gaussian Source with Squared Error Distortion Measure

- For any $p_{S,\hat{S}_1,\hat{S}_2} = p_S p_{\hat{S}_1,\hat{S}_2|S}$ with $\mathbb{E}[(S-\hat{S}_i)^2] \leq d_i$, $i=1,2$,

$$\mathcal{C}(\text{G-BC}(d_1,d_2)) \subseteq \mathcal{C}(p_S, p_{\hat{S}_1,\hat{S}_2|S}).$$

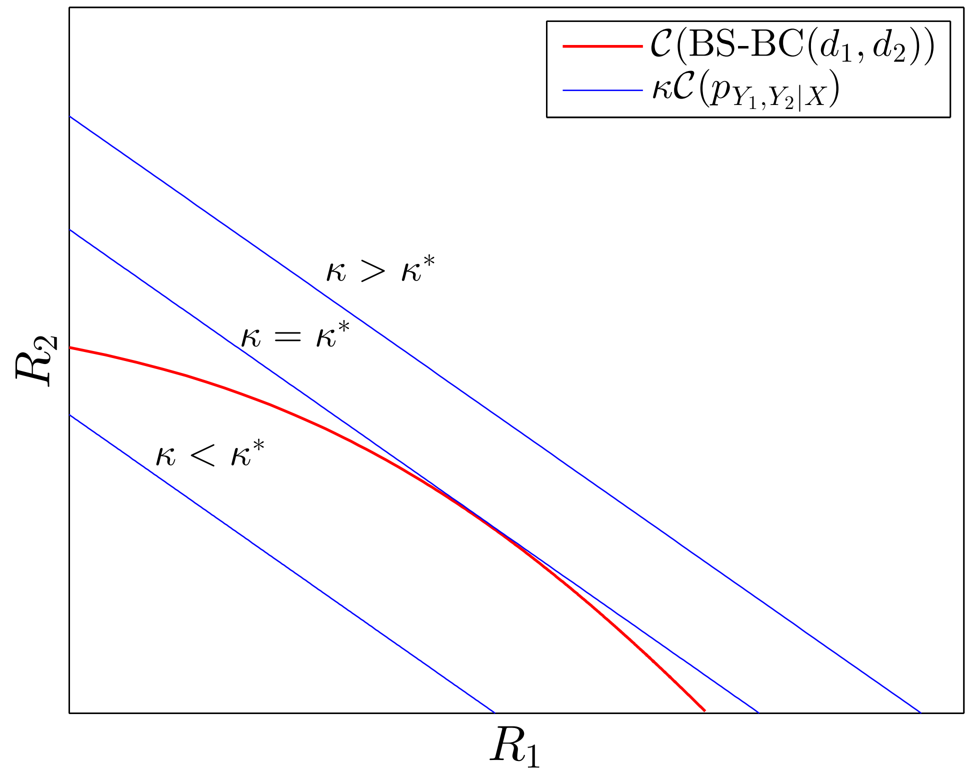
- If (κ, d_1, d_2) is achievable, then $\mathcal{C}(\text{G-BC}(d_1,d_2)) \subseteq \kappa\mathcal{C}(p_{Y_1,Y_2|X})$.



- For any $p_{S_1,\hat{S}_1,\hat{S}_2} = p_S p_{\hat{S}_1,\hat{S}_2|S}$ with $\mathbb{E}[S \oplus \hat{S}_i] \leq d_i$, $i = 1, 2$,

$$\mathcal{C}(\text{BS-BC}(d_1, d_2)) \subseteq \mathcal{C}(p_S, p_{\hat{S}_1,\hat{S}_2|S}).$$

- If (κ, d_1, d_2) is achievable, then $\mathcal{C}(\text{BS-BC}(d_1, d_2)) \subseteq \kappa\mathcal{C}(p_{Y_1,Y_2|X})$.



Another Converse Method

- A general ordering: Given any $U_1, \cdots, U_L$, there exist $V_1, \cdots, V_L$ such that $I(U_{\mathcal{A}_1}; \hat{S}_1) + I(U_{\mathcal{A}_2}; \hat{S}_2 | U_{\mathcal{A}_1}) + \cdots + I(U_{\mathcal{A}_k}; \hat{S}_2 | U_{\cup_{j=1}^{k-1} \mathcal{A}_j})$
$\leq \kappa[I(V_{\mathcal{A}_1}; Y_1) + I(V_{\mathcal{A}_2}; Y_2 | V_{\mathcal{A}_1}) + \cdots + I(V_{\mathcal{A}_k}; Y_2 | V_{\cup_{j=1}^{k-1} \mathcal{A}_j})]$
for any $\mathcal{A}_1, \cdots, \mathcal{A}_k \subseteq \{1, \cdots, L\}$.

- A general ordering: Given any $U_1, \cdots, U_L$, there exist $V_1, \cdots, V_L$ such that $I(U_{\mathcal{A}_1}; \hat{S}_1) + I(U_{\mathcal{A}_2}; \hat{S}_2 | U_{\mathcal{A}_1}) + \cdots + I(U_{\mathcal{A}_k}; \hat{S}_2 | U_{\cup_{j=1}^{k-1} \mathcal{A}_j})$
$\leq \kappa[I(V_{\mathcal{A}_1}; Y_1) + I(V_{\mathcal{A}_2}; Y_2 | V_{\mathcal{A}_1}) + \cdots + I(V_{\mathcal{A}_k}; Y_2 | V_{\cup_{j=1}^{k-1} \mathcal{A}_j})]$
for any $\mathcal{A}_1, \cdots, \mathcal{A}_k \subseteq \{1, \cdots, L\}$.

- A subset of inequalities
$I(U_0; \hat{S}_i) \leq \kappa I(V_0; Y_i)$, $i = 1, 2$
$I(U_0, U_i; \hat{S}_i) \leq \kappa I(V_0, V_i; Y_i)$, $i = 1, 2$
$I(U_0; \hat{S}_1) + I(U_2; \hat{S}_2 | U_0) \leq \kappa[I(V_0; Y_1) + I(V_2; Y_2 | V_0)]$
$I(U_0; \hat{S}_2) + I(U_1; \hat{S}_1 | U_0) \leq \kappa[I(V_0; Y_2) + I(V_1; Y_1 | V_0)]$
$I(U_0, U_1; \hat{S}_1) + I(S; \hat{S}_2 | U_0, U_1) \leq \kappa[I(V_0, V_1; Y_1) + I(X; Y_2 | V_0, V_1)]$
$I(U_0, U_2; \hat{S}_2) + I(S; \hat{S}_1 | U_0, U_2) \leq \kappa[I(V_0, V_2; Y_2) + I(X; Y_1 | V_0, V_2)]$
$I(U_0; \hat{S}_1) + I(U_2; \hat{S}_2 | U_0) + I(S; \hat{S}_1 | U_0, U_2)$
$\leq \kappa[I(V_0; Y_1) + I(V_2; Y_2 | V_0) + I(X; Y_1 | V_0, V_2)]$
$I(U_0; \hat{S}_2) + I(U_1; \hat{S}_1 | U_0) + I(S; \hat{S}_2 | U_0, U_1)$
$\leq \kappa[I(V_0; Y_2) + I(V_1; Y_1 | V_0) + I(X; Y_2 | V_0, V_1)]$
Therefore, if $(\kappa, \mathcal{D}_1, \mathcal{D}_2)$ is achievable, then

$$\mathcal{C}_{\text{out}}(p_S, p_{\hat{S}_1, \hat{S}_2 | S}) \subseteq \kappa \mathcal{C}_{\text{out}}(p_{Y_1, Y_2 | X})$$

for some $p_{S, \hat{S}_1, \hat{S}_2} = p_S p_{\hat{S}_1, \hat{S}_2 | S}$ with $p_{S, \hat{S}_i} \in \mathcal{D}_i$, $i = 1, 2$.

Conclusion

- The source-channel separation theorem can be useful even in the scenarios where the source-channel separation architecture is strictly suboptimal!
- From source-channel separation to source-channel correspondence

Conclusion

- The source-channel separation theorem can be useful even in the scenarios where the source-channel separation architecture is strictly suboptimal!
- From source-channel separation to source-channel correspondence

Thank you!