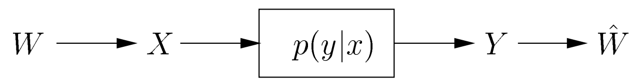
The Role of Side Information in Communication and Data Compression

Tom Cover
Stanford University

Combination of joint work with Mung Chiang, Arak Sutivong, Young-Han Kim, and David Julian.

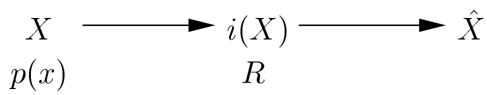
Introduction

Channel



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

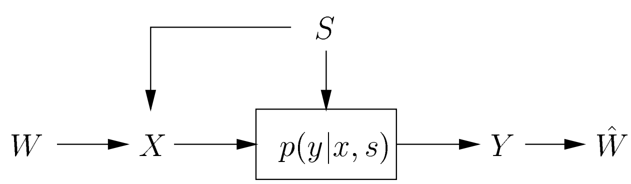
Data Compression



$$R = \min_{p(\hat{x}|x)} I(X;\hat{X})$$

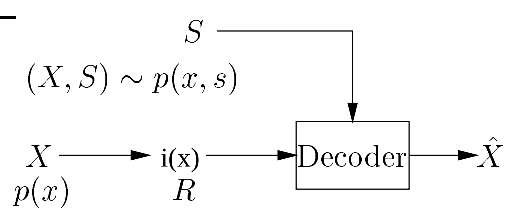
Now consider state information:

Channel



$$C \stackrel{??}{=} \max_{p(x|s)} I(X;Y|S)$$

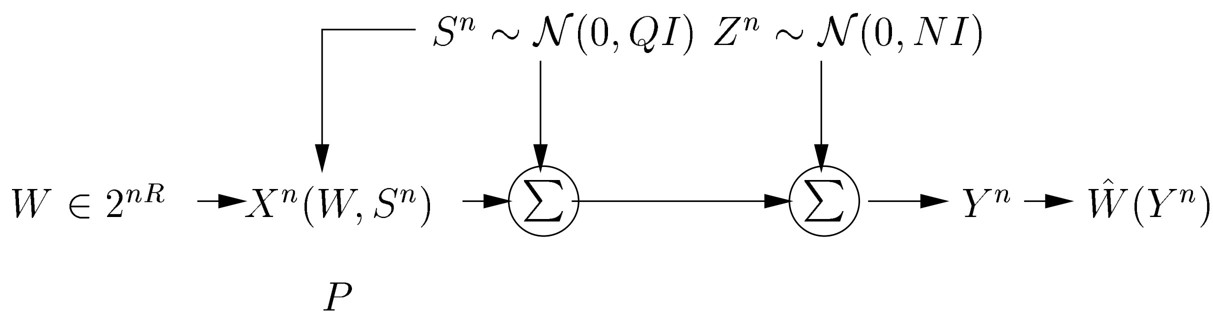
Data Compression



$$R \stackrel{??}{=} \min_{p(\hat{x}|x,s)} I(X;\hat{X}|S)$$

Example 1

Gaussian



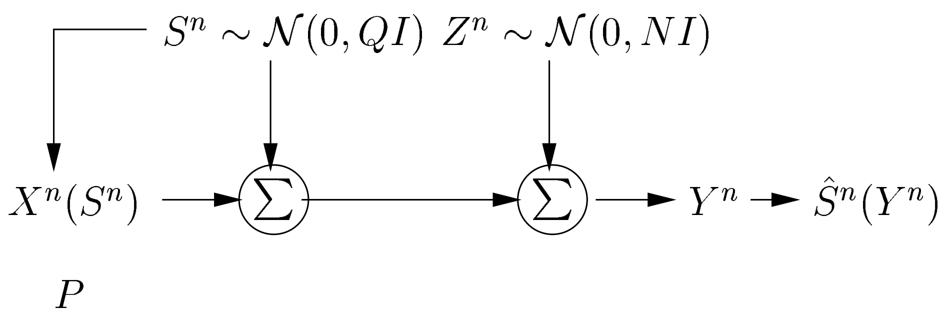
$$C = \frac{1}{2}\log(1 + \frac{P}{N})$$

(Costa, Writing on Dirty Paper)

$$E(S - \hat{S})^2 = Q\frac{P + N}{Q + P + N}$$

Example 2

State estimation



Let $X = \sqrt{\frac{P}{Q}} S$ (amplifying state).

Then

$$\begin{aligned} Y &= X + S + Z \\ &= \left(1 + \sqrt{\frac{P}{Q}}\right) S + Z. \end{aligned}$$

$$E(S - \hat{S})^2 = Q \frac{N}{\left(\sqrt{Q} + \sqrt{P}\right)^2 + N}$$

Example 3

<u>5-card channel</u>

- 52 card deck
- Deal 5 cards
- Hold one back
- View other 4
- Infer the missing card

Example

$$\begin{array}{rccccc} & \text{5C} & \text{2H} & \text{5H} & \text{JH} & \text{10S} \\ i = & 5 & 28 & 31 & 37 & 49 \end{array}$$

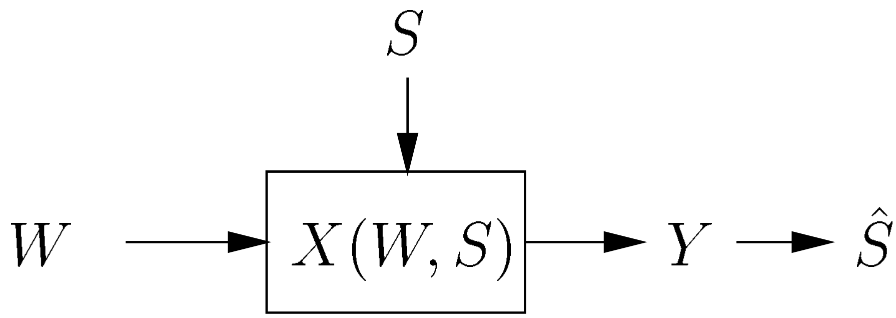
Hold out: 2H
Send: 5C 5H JH 10S
Infer missing card: 2H

<u>Conflict of state with info</u>
What you say conflicts with how you say it.
It's what you don't do that conveys the information.

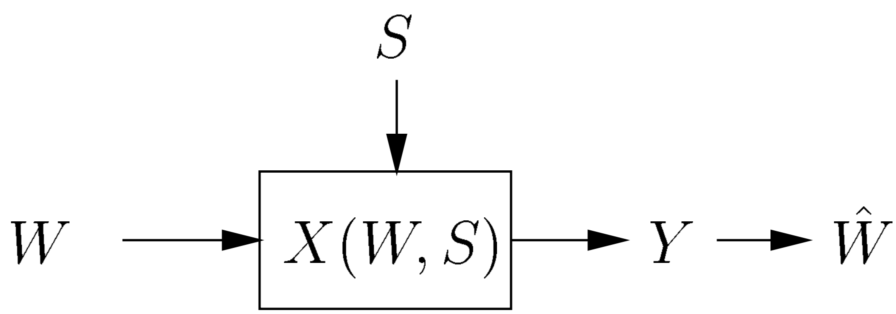
Channel with known state

$$S \longrightarrow X = Y \longrightarrow \hat{S}$$

5 cards   4 cards   5 cards



Pure state info

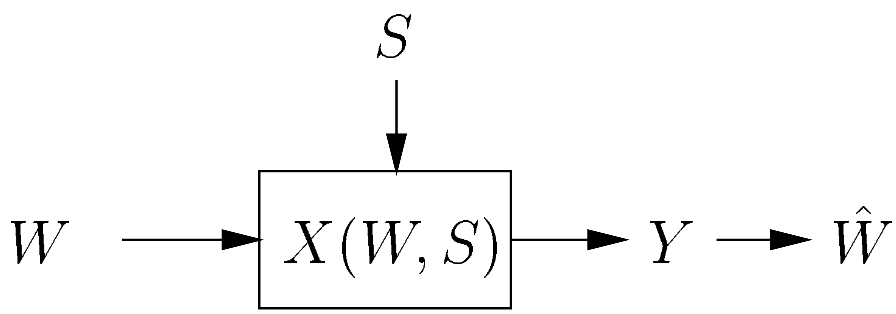


Pure info

Conflict of $\hat{S}$ and $\hat{W}$.

5-card channel: pure information

m = 52-card deck
S = 5 cards
X = 4-card subset
Y = X



<u>Claim</u> Capacity $C = \log 5$ bits

Pinsker, Gel'fand, El Gamal, Heegard $C = I(U;Y) - I(U;S)$.

5-card channel: state estimation

How large a deck m?

$C = \log 5$ bits
4! orderings of 4 cards
(5-card hand) $\rightarrow$ (5)4! $=$ 120 actions

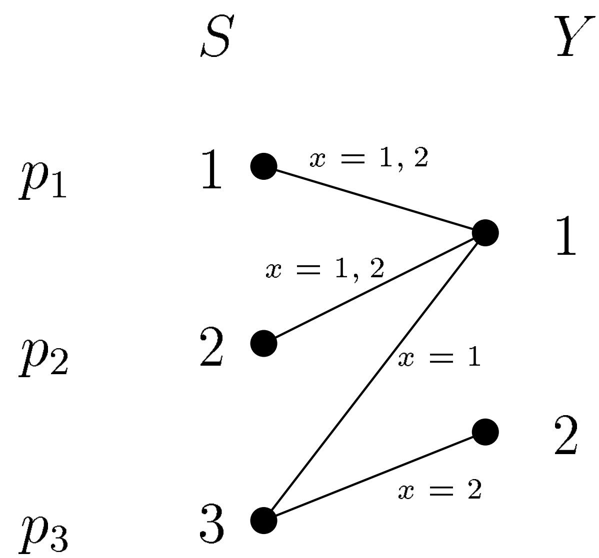
Can determine 120 missing cards.

$$m = 124$$

Example 4

Deterministic channel

$$y = g(x,s)$$



Pure info

$$C_{10} = E_S \log |\mathcal{Y}_S| = \max_{p(x|s)} H(Y|S) = p_3$$

$$\alpha^* = \tfrac{1}{2}$$

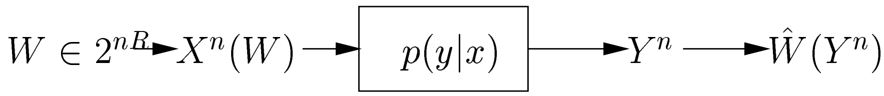
State estimation Decrease in entropy of S:

$$\Delta H(S) = H(S) - (H(S|Y) - H(Y|S)) = H(Y)$$

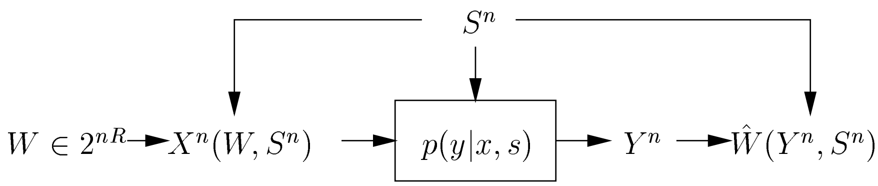
$$\max_{p(x|s)} H(Y) = \max_{\alpha} H(\alpha p_3) = \begin{cases} 1, & p_3 \geq \frac{1}{2} \\ H(p_3), & p_3 \leq \frac{1}{2} \end{cases}$$

$$\alpha^* = \frac{1}{2p_3}$$

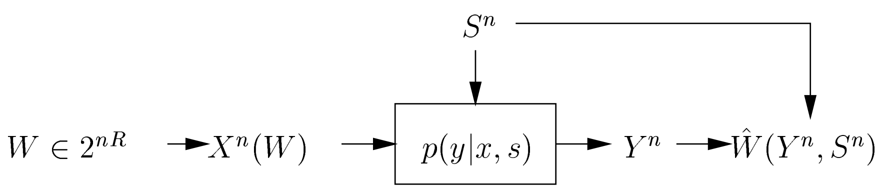
Channel capacity with side information



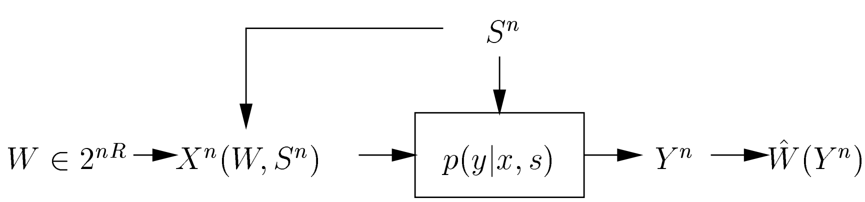
$$C_{00} = \max_{p(x)} I(X;Y)$$



$$C_{11} = \max_{p(x|s)} I(X;Y|S)$$



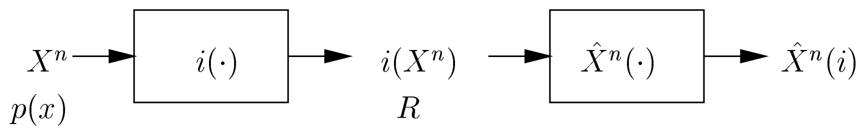
$$C_{01} = \max_{p(x)} I(X;Y|S)$$



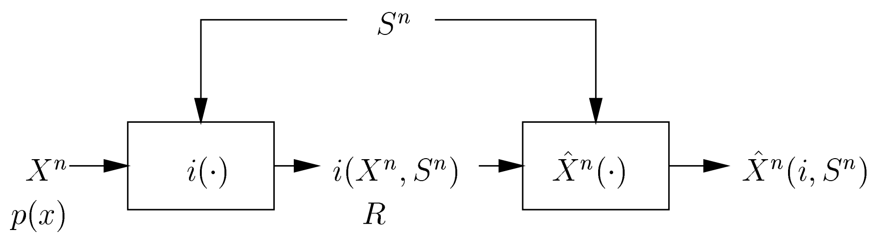
$$C_{10} = \max_{p(u,x|s)} (I(U;Y) - I(U;S))$$

Gel'fand, Pinsker, El Gamal, and Heegard

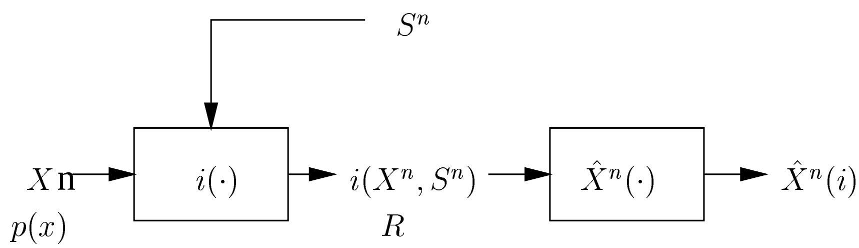
Rate-distortion with side information



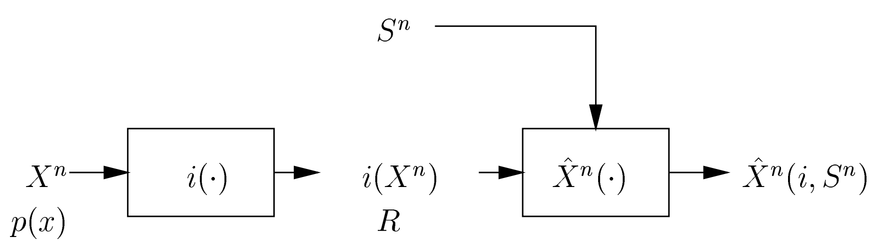
$$R_{00} = \min_{p(\hat{x}|x)} I(X;\hat{X})$$



$$R_{11} = \min_{p(\hat{x}|x,s)} I(X;\hat{X}|S)$$



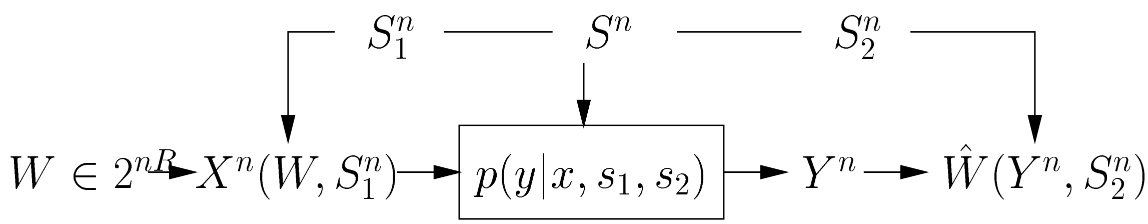
$$R_{10} = \min_{p(\hat{x}|x)} I(X;\hat{X})$$



$$R_{01} = \min_{p(u|x)p(\hat{x}|u,s)} (I(U;X) - I(U;S))$$

Wyner and Ziv

Generalized capacity with state information



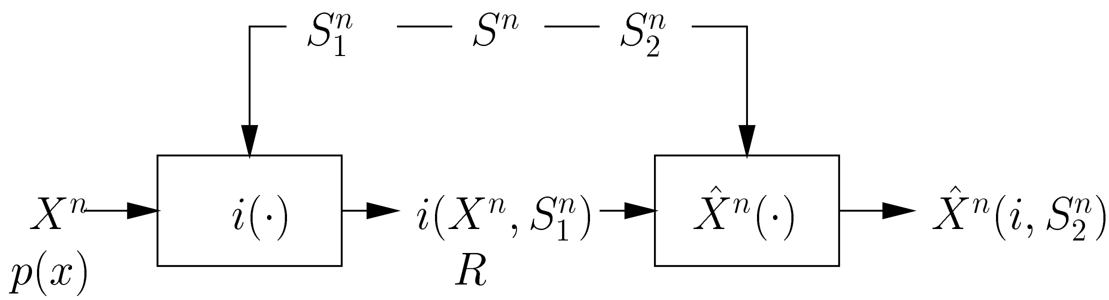
Channel: $p(y|x, s_1, s_2)$
State: $(S_{1i}, S_{2i}) \sim$ i.i.d. $p(s_1, s_2)$

Theorem 1*

$$C = \max_{p(u,x|s_1)} (I(U; Y, S_2) - I(U; S_1))$$

* M. Chiang

Generalized rate-distortion with side information



Source: $(X_i, S_{1i}, S_{2i}) \sim$ i.i.d. $p(x, s_1, s_2)$
Distortion: $d(x, \hat{x})$

<u>Theorem 2</u>

$$R = \min_{p(u|x,s_1)p(\hat{x}|u,s_2)} \left(I(U; X, S_1) - I(U; S_2) \right)$$

Duality

Channel capacity

$$C_{S_1 S_2} = \max_{p(u|s_1)p(x|u,s_1)} (I(U;Y,S_2) - I(U;S_1))$$

Rate-distortion function

$$R_{S_1 S_2}(D) = \min_{p(u|x,s_1)p(\hat{x}|u,s_2): Ed(X,\hat{X}) \le D} (I(U;X,S_1) - I(U;S_2))$$

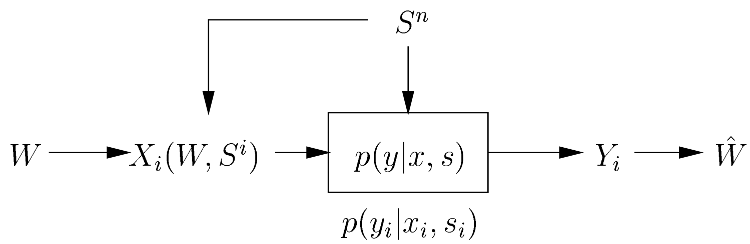
Duality:

- Isomorphism
- Complementarity

C	$\Leftrightarrow$	$R(D)$
maximization	$\Leftrightarrow$	minimization
C_{00}	$\Leftrightarrow$	R_{00}
C_{11}	$\Leftrightarrow$	R_{11}
C_{10}	$\Leftrightarrow$	R_{01}
Y (received symbol)	$\Leftrightarrow$	X (source)
S (state)	$\Leftrightarrow$	S (state)
U (auxiliary)	$\Leftrightarrow$	U (auxiliary)
X (transmitted symbol)	$\Leftrightarrow$	$\hat{X}$ (estimation)

Note: Pradhan, Chou, and Ramchandran have some recent results on functional duality between C and R.

Causal side information (Shannon)



$$C = \max_{p(t)} I(T;Y), \quad T \in \{f : \mathcal{S} \to \mathcal{X}\}$$

Can recast as

$$C = \max_{p(x(\cdot))} I(X(\cdot);Y)$$

(Shannon, Channels with state information at the transmitter, *IBM J. Res. Dev.*, 1958, pp 289-293.)

New characterization

Shannon (Causal)

$$C_{10} = \max_{p(x(\cdot))} I(X(\cdot);Y)$$

Can recast (YHK) as

$$C_{10} = \max_{\substack{p(u)\,p(x|u,s)\\ x=f(u,s)}} (I(U;Y) - I(U;S))$$

Generalized capacity with two-sided state information:

Noncausal

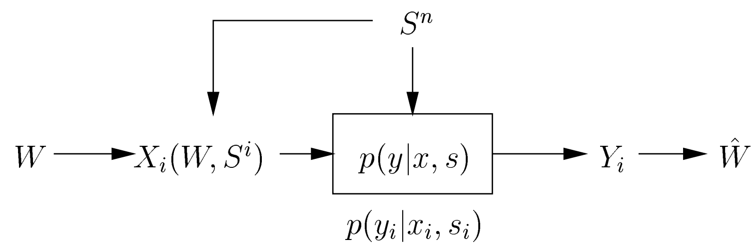
$$C = \max_{p(u|s_1)\,p(x|u,s_1)} (I(U;Y,S_2) - I(U;S_1))$$

Causal

$$C = \max_{p(u)\,p(x|u,s_1)} (I(U;Y,S_2) - I(U;S_1))$$

(Young-Han Kim)

Converse for Shannon causal side information channel



$$
\begin{aligned}
C_{10} &= \max_{p(u)p(x|u,s)} \left(I(U;Y) - I(U;S) \right) \\
&= \max_{p(u)p(x|u,s)} I(U;Y)
\end{aligned}
$$

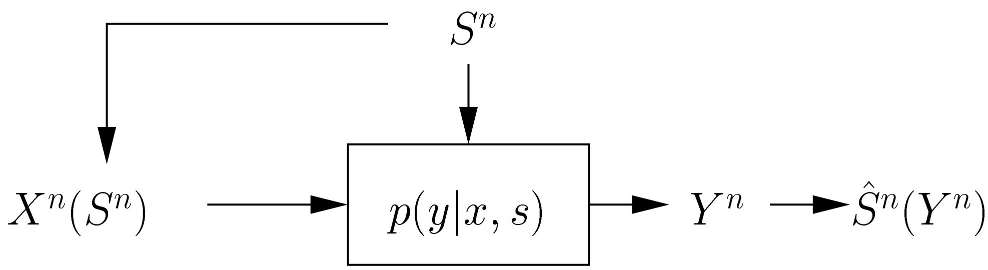
Converse:

Given code $X_i(W, S^i)$, with Pr$\{\hat{W} \neq W\} \to 0$.
Then

$$
\begin{aligned}
nR &= H(W) \\
&\leq I(W;Y^n) + n\epsilon_n \\
&= H(Y^n) - H(Y^n|W) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} H(Y_i) - H(Y_i|\underbrace{W, Y^{i-1}}_{U_i}) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(U_i;Y_i) + n\epsilon_n \\
&\leq nC + n\epsilon.
\end{aligned}
$$

Sending state information

Deterministic channel. "Amplify" state information.



Let $y = g(x, s)$ (i.e., a deterministic function).

Note: Capacity $C = \max_{p(x|s)} H(Y|S)$ bits

Reduction in uncertainty about S:

Original uncertainty: $H(S)$
Uncertainty after observing Y: $H(S|Y)$
Uncertainty after refinement (using C): $H(S|Y) - H(Y|S)$

Note: Reduction in uncertainty $\Delta H(S) = \max_{p(x|s)} H(Y)$

(Arak Sutivong)

Concluding remarks

Channel with noncausal state information

$$C_{S_1S_2} = \max_{p(u|s_1)p(x|u,s_1)} (I(U;Y,S_2) - I(U;S_1))$$

Rate-distortion function

$$R_{S_1S_2} = \min_{p(u|x,s_1)p(\hat{x}|u,s_2)} (I(U;X,S_1) - I(U;S_2))$$

Channel with causal state information

$$C_{S_1S_2} = \max_{p(u)p(x|u,s_1)} (I(U;Y,S_2) - I(U;S_1))$$

Amplification of state

$$\Delta H(S) = \max_{p(x|s)} H(Y)$$