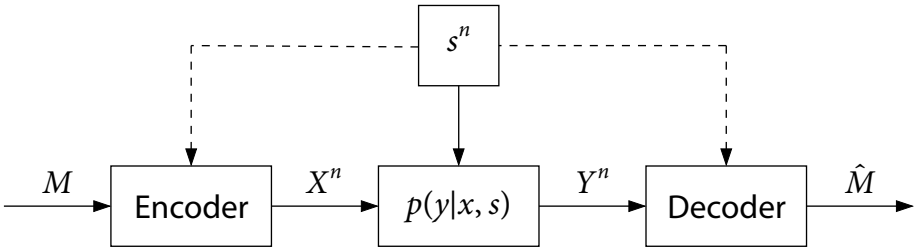
Lecture #7 Channels with State

(Reading: NIT 3.7, 7.1, 7.3–7.7, 9.5)

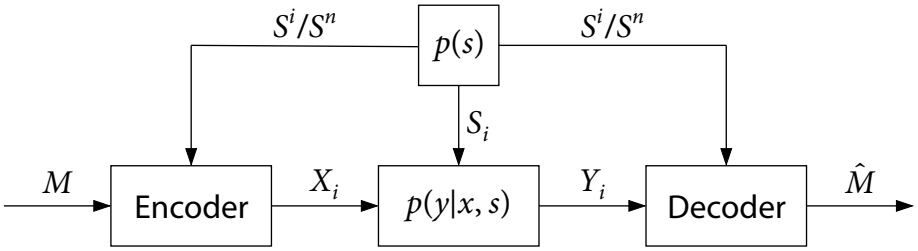
- DMC with state
- DMC with DM state
- Causal state information available at the encoder
- Noncausal state information available at the encoder
- Writing on dirty paper

DMC with state



- DMC with state $(\mathcal{X} \times \mathcal{S}, p(y|x,s), \mathcal{Y})$
- State: Channel uncertainty, jamming, fading, memory faults, host image
- Three general classes:
  - Compound channel: State is fixed throughout transmission
  - Arbitrarily varying channel: s^n is an arbitrary sequence
  - Random state

DMC with DM state



- DMC with DM state $(\mathcal{X} \times \mathcal{S}, p(y|x,s)p(s), \mathcal{Y})$
- State information availability:
  - Encoder, decoder, neither, both
  - Noiseless, noisy, coded
  - Causal (S^i known before transmission i), noncausal (S^n known before transmission)
- For each setup, $(2^{nR}, n)$ code, achievability, and capacity defined in the usual way

Simple special cases

- No state information available at either the encoder or the decoder:

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

  where $p(y|x) = \sum_s p(s)p(y|x,s)$
- State information available (causally or noncausally) at the decoder ($\hat{m}(y^n, s^n)$):

$$C_{\text{SI-D}} = \max_{p(x)} I(X;Y,S) = \max_{p(x)} I(X;Y|S),$$

  and is achieved by treating (Y,S) as the channel output
- State information available at both encoder and decoder ($x^n(m,s^i)$, $\hat{m}(y^n,s^n)$):

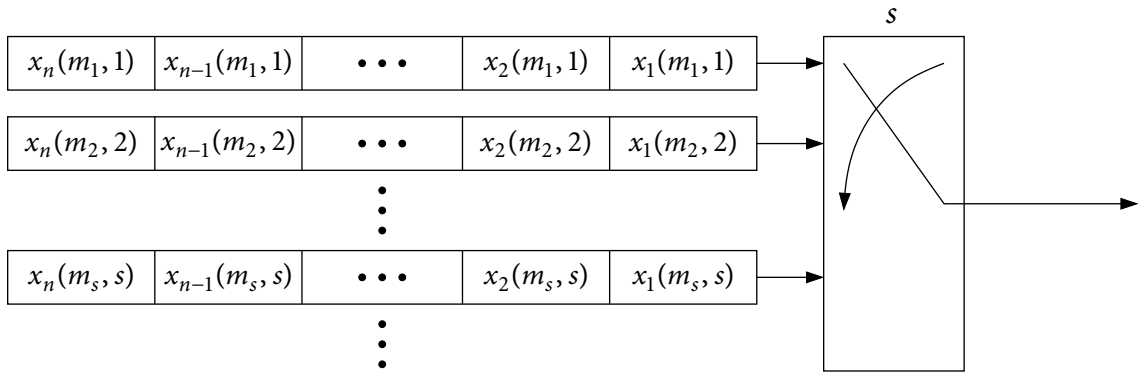
$$C_{\text{SI-ED}} = \max_{p(x|s)} I(X;Y|S)$$

  for both causal and noncausal cases

  Key achievability idea: Treat S^n as a time-sharing sequence

Proof of achievability (Goldsmith–Varaiya 1997)

- Split M into independent messages with rates R_s, $s \in \mathcal{S}$; hence $\sum_s R_s = R$
- Codebook generation:
  - For each s, generate 2^{nR_s} sequences $x^n(m_s, s) \sim \prod_{i=1}^n p_{X|S}(x_i|s)$, $m_s \in [1 : 2^{nR_s}]$
- Encoding:
  - To send message $m = (m_s : s \in \mathcal{S})$, store each of $x^n(m_s, s)$ in a FIFO buffer for s
  - In time i, transmit the first untransmitted symbol from the FIFO buffer for s_i

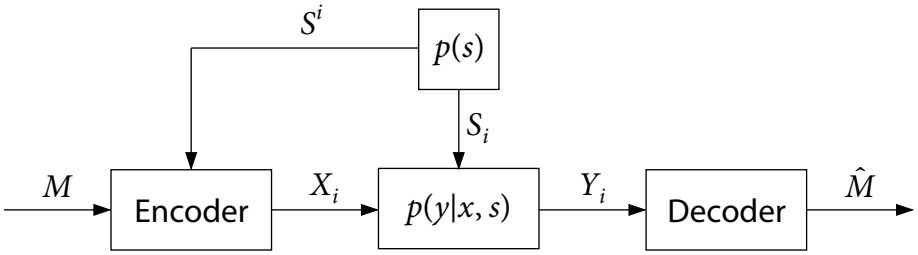


Proof of achievability (Goldsmith–Varaiya 1997)

- Split M into independent messages with rates R_s, $s \in \mathcal{S}$; hence $\sum_s R_s = R$
- Codebook generation:
  - For each s, generate 2^{nR_s} sequences $x^n(m_s, s) \sim \prod_{i=1}^n p_{X|S}(x_i|s)$, $m_s \in [1 : 2^{nR_s}]$
- Encoding:
  - To send message $m = (m_s : s \in \mathcal{S})$, store each of $x^n(m_s, s)$ in a FIFO buffer for s
  - In time i, transmit the first untransmitted symbol from the FIFO buffer for s_i
- Decoding and the analysis of the probability of error:
  - Demultiplex the received sequence into subsequences $(y^{n_s}(s), s \in \mathcal{S})$, $\sum_s n_s = n$
  - If $s^n \in \mathcal{T}_\epsilon^{(n)}$, then $n_s \geq n(1-\epsilon)p(s)$ for every $s \in \mathcal{S}$

    Find a unique $\hat{m}_s$ for each s such that $(x^{n(1-\epsilon)p(s)}(\hat{m}_s, s), y^{n(1-\epsilon)p(s)}(s)) \in \mathcal{T}_\epsilon^{(n)}$
  - By the LLN and packing lemma, $P_e^{(n)}(s) \to 0$ if $R_s < (1-\epsilon)p(s)I(X; Y|S = s) - \delta(\epsilon)$
  - Hence, $P_e^{(n)} \to 0$ if $R < (1-\epsilon)I(X; Y|S) - \delta(\epsilon)$

Causal state information available at the encoder



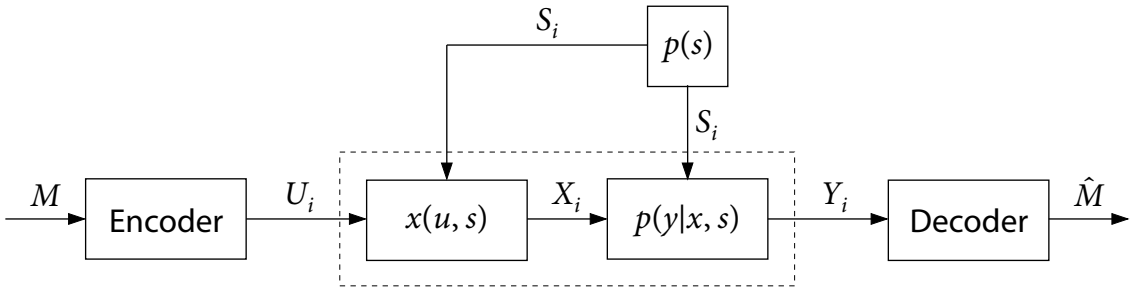
Theorem 7.2 (Shannon 1958)

$$C_{\text{CSI-E}} = \max_{p(u),\, x(u,s)} I(U;Y),$$

where U is independent of S with $|\mathcal{U}| \le \min\{(|\mathcal{X}|-1)|\mathcal{S}|+1,\ |\mathcal{Y}|\}$

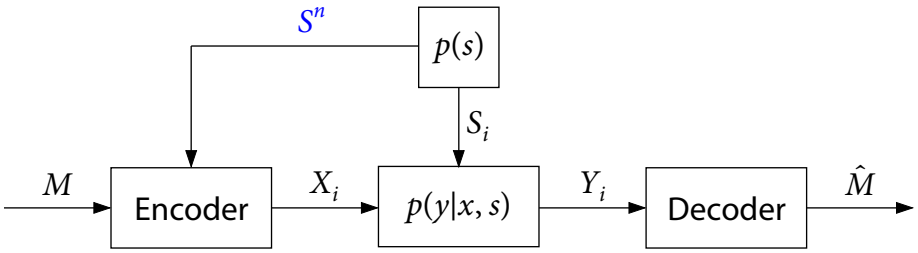
- Proof of the converse: Read NIT 7.5

Proof of achievability



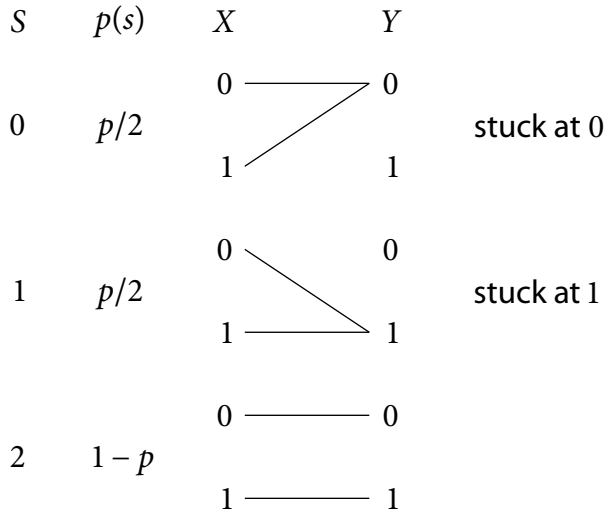
- Fix $p(u)$ and $x(u,s)$ that achieve $C_{\text{SI-E}}$
- Shannon strategy: Attach a "physical device" $x(u,s)$ in front of the actual channel
- This induces a DMC $p(y|u) = \sum_s p(y|x(s,u),s)p(s)$ with input U and output Y
- Now code for the induced DMC $p(y|u)$ to achieve $I(U;Y)$
  Encoding: To send m, transmit $x_i = x(u_i(m), s_i)$, $i \in [1:n]$
- Can be viewed as coding over all functions $\{x_u(s) : \mathcal{S} \to \mathcal{X}\}$ (u: function index)

Noncausal state information available at the encoder



- Motivation for noncausal state information:
  - Memory with defects
  - Write-once memory
  - Digital watermarking
  - General broadcast channel

Memory with stuck-at faults



- If the reader knows the fault locations: $C_{\text{SI-D}} = C_{\text{SI-ED}} = 1 - p$
- If neither the writer nor the reader knows: $C = 1 - H(p/2)$
- If the writer knows the fault locations: $C_{\text{SI-E}} = ?$
- Kuznetsov–Tsybakov (1974) showed: $C_{\text{SI-E}} = 1 - p$

Multicoding

- Codebook generation: Randomly partition $\{0,1\}^n$ into 2^{nR} subcodebooks

$\mathcal{C}(1)$	1	1	1	0	1	0	1
	1	0	1	0	0	0	1
	0	1	0	1	1	1	0
	0	0	0	1	0	1	0
	1	1	1	0	0	0	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathcal{C}(m)$	0	0	1	0	1	0	1
	1	0	0	1	0	0	1
	1	1	0	0	1	1	0
	0	0	1	1	0	1	0
	0	1	1	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathcal{C}(2^{nR})$	0	0	0	0	1	1	1
	1	1	1	1	0	0	1
	0	0	0	0	1	1	0
	0	1	1	1	1	1	0
	1	0	0	0	0	0	0

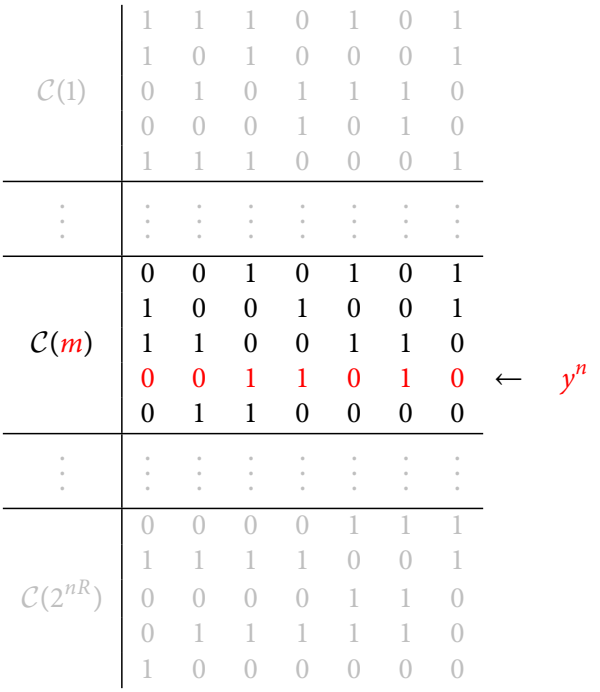
Multicoding

- Writing: Store m with s^n: 0 2 1 2 2 2 0

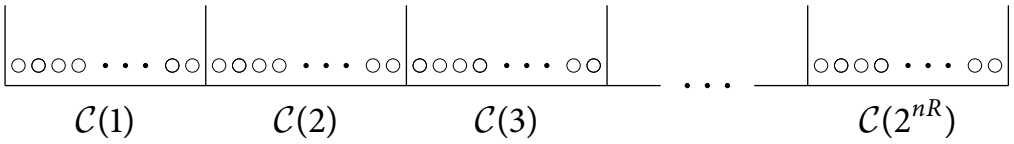
$\mathcal{C}(1)$	1	1	1	0	1	0	1	
	1	0	1	0	0	0	1	
	0	1	0	1	1	1	0	
	0	0	0	1	0	1	0	
	1	1	1	0	0	0	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\mathcal{C}(m)$	0	0	1	0	1	0	1	
	1	0	0	1	0	0	1	
	1	1	0	0	1	1	0	
	0	0	1	1	0	1	0	$\leftarrow$ x^n
	0	1	1	0	0	0	0	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\mathcal{C}(2^{nR})$	0	0	0	0	1	1	1	
	1	1	1	1	0	0	1	
	0	0	0	0	1	1	0	
	0	1	1	1	1	1	0	
	1	0	0	0	0	0	0	

Multicoding

- Reading:



Analysis of the probability of error



- Error occurs iff there is no $x^n \in \mathcal{C}(m)$ that matches the fault pattern
- For n large, there are $\approx np$ faults
- Hence, there are $\doteq 2^{n(1-p)}$ that match any given fault pattern
- If $R < 1 - p$ and n large, $\mathcal{C}(m)$ has a matching sequence w.h.p.
- Hence the capacity $C_{\text{SI-E}} = 1 - p$

Gelfand–Pinsker theorem

- Gelfand–Pinsker (1980) generalized this result to arbitrary DMC with DM state

Theorem 7.3

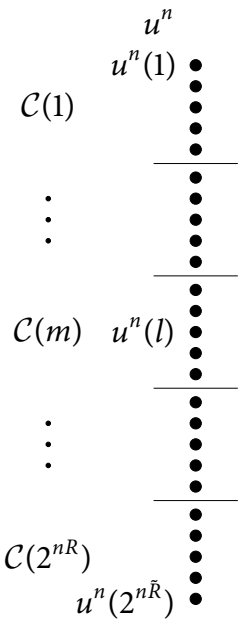
$$C_{\text{SI-E}} = \max_{p(u|s),\, x(u,s)} \big(I(U;Y) - I(U;S)\big),$$

where $|\mathcal{U}| \le \min\{|\mathcal{X}| \cdot |\mathcal{S}|,\ |\mathcal{Y}| + |\mathcal{S}| - 1\}$

- Example: Memory with defects
  - If $S = 2$, set $U = X \sim \text{Bern}(1/2)$
  - If $S = 1$ or 0, set $U = X = S$
  - Then,
$$I(U;Y) - I(U;S) = H(U|S) - H(U|Y) = 1 - p$$

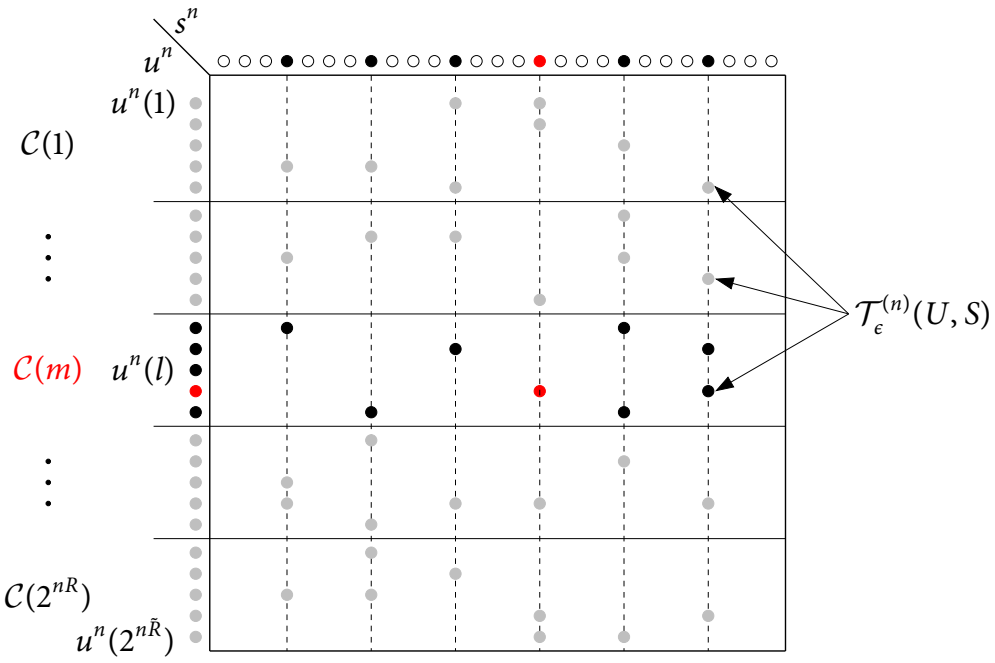
Proof of achievability

- Codebook generation: Fix $p(u|s)$ and $x(u,s)$ that achieves $C_{\text{SI-E}}$, let $\tilde{R} > R$
  - For each $m \in [1 : 2^{nR}]$, generate a subcodebook $\mathcal{C}(m)$ consisting of $2^{n(\tilde{R}-R)}$ sequences $u^n(l) \sim \prod_{i=1}^{n} p_U(u_i)$, $l \in [(m-1)2^{n(\tilde{R}-R)} + 1 : m2^{n(\tilde{R}-R)}]$



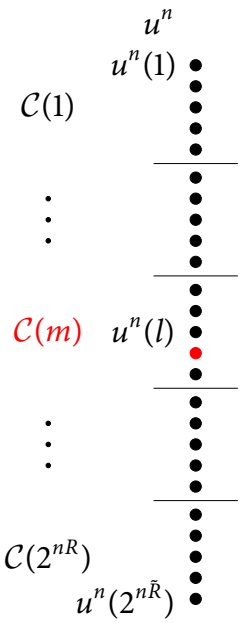
Proof of achievability

- Encoding: To send m given s^n, find $u^n(l) \in \mathcal{C}(m)$ such that $(u^n(l), s^n) \in \mathcal{T}_{\epsilon'}^{(n)}$
  - If no such $u^n(l)$ exists, set $l = 1$
  - Then transmit $x_i = x(u_i(l), s_i)$ for $i \in [1 : n]$



Proof of achievability

- Decoding:
  - Find the unique $\hat{m}$ such that $(u^n(l), y^n) \in \mathcal{T}_\epsilon^{(n)}$ for some $u^n(l) \in \mathcal{C}(\hat{m})$



Analysis of the probability of error

- Consider $\mathrm{P}(\mathcal{E})$ conditioned on $M = 1$
- Let L denote the index of the chosen U^n for S^n and $M = 1$
- Error events:

$$\begin{aligned}
\mathcal{E}_1 &= \{(U^n(l), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } U^n(l) \in \mathcal{C}(1)\}, \\
\mathcal{E}_2 &= \{(U^n(L), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}, \\
\mathcal{E}_3 &= \{(U^n(l), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } l \notin [1 : 2^{n(\tilde{R}-R)}]\}
\end{aligned}$$

Thus, by the union of events bound

$$\mathrm{P}(\mathcal{E}) \le \mathrm{P}(\mathcal{E}_1) + \mathrm{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathrm{P}(\mathcal{E}_3)$$

Conditional and joint typicality lemmas

Conditional typicality lemma

Let $(X, Y) \sim p(x, y)$ and $\epsilon > \epsilon'$. If $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$, $Y^n \sim \prod_{i=1}^n p_{Y|X}(y_i|x_i)$, then

$$\lim_{n\to\infty} \mathrm{P}\{(x^n, Y^n) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y)\} = 1$$

- If $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$, $\epsilon > \epsilon'$, then for n sufficiently large,

$$|\mathcal{T}_{\epsilon}^{(n)}(Y|x^n)| \ge 2^{n(H(Y|X)-\delta(\epsilon))}$$

Joint typicality lemma (part 2)

Let $(X, Y) \sim p(x, y)$ and $\epsilon > \epsilon'$. If $x^n \in \mathcal{T}_{\epsilon'}^{(n)}$ and $\tilde{Y}^n \sim \prod_{i=1}^n p_Y(\tilde{y}_i)$, then for some $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ and n sufficiently large,

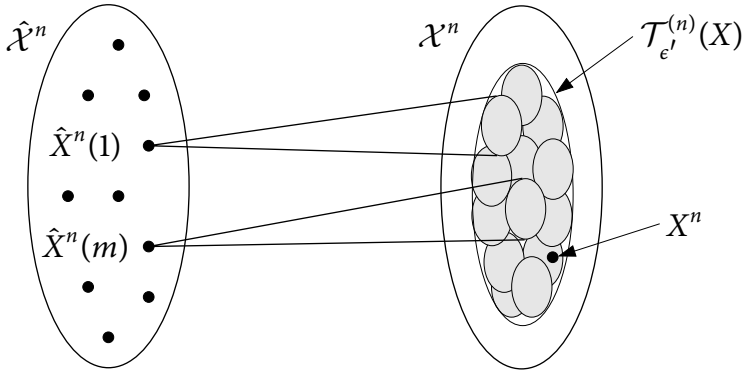
$$\mathrm{P}\{(x^n, \tilde{Y}^n) \in \mathcal{T}_{\epsilon}^{(n)}(X, Y)\} \ge 2^{-n(I(X;Y)+\delta(\epsilon))}$$

Covering lemma ($U = \emptyset$)

- Let $(X, \hat{X}) \sim p(x, \hat{x})$ and $\epsilon' < \epsilon$
- Let $X^n \sim p(x^n)$ be arbitrarily distributed such that

$$\lim_{n\to\infty} \mathsf{P}\{X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)\} = 1$$

- Let $\hat{X}^n(m) \sim \prod_{i=1}^n p_{\hat{X}}(\hat{x}_i)$, $m \in \mathcal{A}$, $|\mathcal{A}| \geq 2^{nR}$, be independent of each other and of X^n



Covering lemma ($U = \emptyset$)

- Let $(X, \hat{X}) \sim p(x, \hat{x})$ and $\epsilon' < \epsilon$
- Let $X^n \sim p(x^n)$ be arbitrarily distributed such that

$$\lim_{n\to\infty} \mathsf{P}\{X^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)\} = 1$$

- Let $\hat{X}^n(m) \sim \prod_{i=1}^n p_{\hat{X}}(\hat{x}_i)$, $m \in \mathcal{A}$, $|\mathcal{A}| \geq 2^{nR}$, be independent of each other and of X^n

Lemma 3.3 (Covering lemma)

There exists $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ such that

$$\lim_{n\to\infty} \mathsf{P}\{(X^n, \hat{X}^n(m)) \notin \mathcal{T}_{\epsilon}^{(n)} \text{ for all } m \in \mathcal{A}\} = 0,$$

if $R > I(X; \hat{X}) + \delta(\epsilon)$

Analysis of the probability of error

- Error events:

$$\mathcal{E}_1 = \{(U^n(l), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } U^n(l) \in \mathcal{C}(1)\},$$
$$\mathcal{E}_2 = \{(U^n(L), Y^n) \notin \mathcal{T}_{\epsilon}^{(n)}\},$$
$$\mathcal{E}_3 = \{(U^n(l), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } l \notin [1 : 2^{n(\tilde{R}-R)}]\}$$

- By the covering lemma ($|\mathcal{A}| = 2^{n(\tilde{R}-R)}$, $X \leftarrow S$, $\hat{X} \leftarrow U$), $\mathsf{P}(\mathcal{E}_1) \to 0$ if $\tilde{R} - R > I(U; S) + \delta(\epsilon')$
- Since $\mathcal{E}_1^c = \{(U^n(L), X^n, S^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}$ and

$$Y^n | \{U^n(L) = u^n, X^n = x^n, S^n = s^n\} \sim \prod_{i=1}^n p_{Y|U,X,S}(y_i|u_i, x_i, s_i) = \prod_{i=1}^n p_{Y|X,S}(y_i|x_i, s_i),$$

by the conditional typicality lemma, $\mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) \to 0$

- Since $U^n(l) \sim \prod_{i=1}^n p_U(u_i)$, $l \notin [1 : 2^{n(\tilde{R}-R)}]$, and Y^n are independent, by the packing lemma, $\mathsf{P}(\mathcal{E}_3) \to 0$ if $\tilde{R} < I(U; Y) - \delta(\epsilon)$
- Combining the bounds, $\mathsf{P}(\mathcal{E}) \to 0$ if $R < I(U; Y) - I(U; S) - \delta(\epsilon') - \delta(\epsilon)$

Proof of the converse (Heegard–El Gamal 1983)

- We will need the Csiszár sum identity: Let $(U, X^n, Y^n) \sim F(u, x^n, y^n)$, then

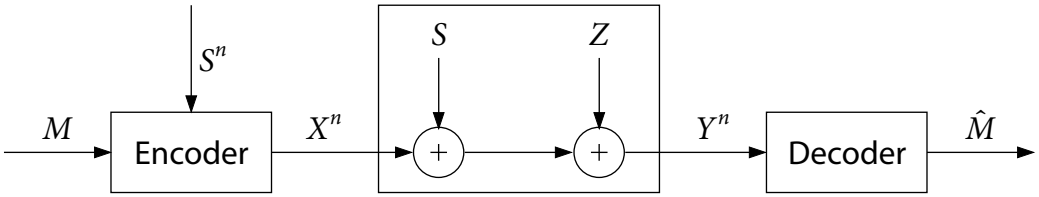
$$\sum_{i=1}^n I(X_{i+1}^n; Y_i | Y^{i-1}, U) = \sum_{i=1}^n I(Y^{i-1}; X_i | X_{i+1}^n, U)$$

- By Fano's inequality,

$$\begin{aligned}
nR &\le I(M; Y^n) + n\epsilon_n \\
&\le \sum_{i=1}^n I(M, Y^{i-1}; Y_i) + n\epsilon_n \\
&= \sum_{i=1}^n I(M, Y^{i-1}, S_{i+1}^n; Y_i) - \sum_{i=1}^n I(S_{i+1}^n; Y_i | M, Y^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^n I(M, Y^{i-1}, S_{i+1}^n; Y_i) - \sum_{i=1}^n I(Y^{i-1}; S_i | M, S_{i+1}^n) + n\epsilon_n \\
&= \sum_{i=1}^n I(M, Y^{i-1}, S_{i+1}^n; Y_i) - \sum_{i=1}^n I(M, S_{i+1}^n, Y^{i-1}; S_i) + n\epsilon_n
\end{aligned}$$

- Now, identify $U_i = (M, S_{i+1}^n, Y^{i-1})$ $(U_i \to (X_i, S_i) \to Y_i)$, ...

Gaussian channel with additive Gaussian state

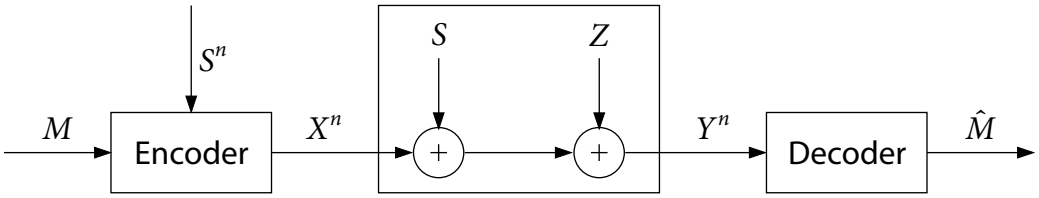


- $S \sim \mathrm{N}(0, Q)$ and $Z \sim \mathrm{N}(0, 1)$ are independent
- Average power constraint P on X
- State information not available at the encoder or decoder: $C = \mathsf{C}\left(P/(1+Q)\right)$
- State information available at the decoder: $C_{\text{SI-D}} = C_{\text{SI-ED}} = \mathsf{C}(P)$
- State information available noncausally at the encoder (Costa 1983):

Theorem 7.4 (Writing on dirty paper)

$$C_{\text{SI-E}} = \mathsf{C}(P)$$

Application: Digital watermarking



- The publisher embeds a watermark X in a host image S
- Given S^n, the authentication message M is encoded into watermark $X^n(M, S^n)$
- The watermark is added to image to generate watermarked image $X^n + S^n$
- An authenticator wishes to retrieve M from $Y^n = X^n + S^n + Z^n$, where $Z \sim \mathrm{N}(0, 1)$
- What is the optimal tradeoff between
  - Capacity C (amount of watermark information) and
  - Power of watermark X (determines fidelity of watermarked image)?
- By the writing on dirty paper, $C(D) = \mathsf{C}(D)$, where D is power of watermark

Proof of achievability

- Gelfand–Pinsker theorem for the DMC with DM state and input cost:

$$C_{\text{SI-E}} = \max_{p(u|s),\, x(u,s):\mathsf{E}(b(X))\le B} \big(I(U;Y) - I(U;S)\big)$$

- For Gaussian channel with additive Gaussian state, find optimal $F(u|s)$ and $x(u,s)$
- Let $U = X + \alpha S$, where $X \sim \mathrm{N}(0,P)$ is independent of S!
- With this choice,

$$I(U;Y) = \frac{1}{2}\log\left(\frac{(P+Q+1)(P+\alpha^2 Q)}{PQ(1-\alpha^2)+(P+\alpha^2 Q)}\right),$$
$$I(U;S) = \frac{1}{2}\log\left(\frac{P+\alpha^2 Q}{P}\right)$$

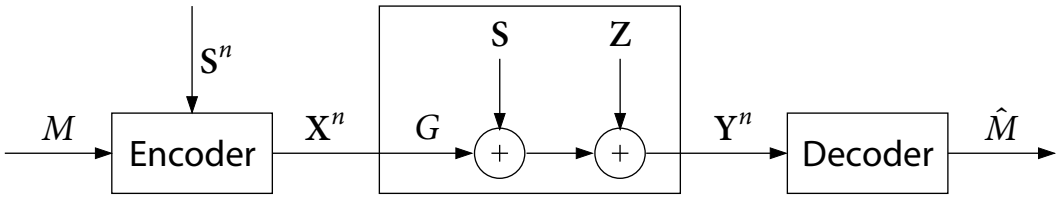
  Thus

$$R(\alpha) = I(U;Y) - I(U;S) = \frac{1}{2}\log\left(\frac{P(P+Q+1)}{PQ(1-\alpha)^2+(P+\alpha^2 Q)}\right)$$

- Maximizing w.r.t. α, we find that $\alpha^* = P/(P+1)$ and $R(\alpha^*) = \mathsf{C}(P)$

Extensions

- Non-Gaussian state (Cohen–Lapidoth 2002): $C = \mathsf{C}(P)$
- Vector writing on dirty paper: Read NIT 9.1, 9.5



  - Average power constraint: $\sum_{i=1}^{n} \mathsf{E}(\mathbf{x}^T(m,\mathrm{S},i)\mathbf{x}(m,\mathrm{S},i)) \le nP$
  - $\mathrm{S} \sim F(\mathbf{s})$ and $\mathbf{Z} \sim \mathrm{N}(0, I_r)$ are independent
  - As in the scalar case, the capacity is the same as if S were not present:

$$C = \max_{F(\mathbf{x}):\mathsf{E}(\mathbf{x}^T\mathbf{x})\le P} I(\mathbf{X};G\mathbf{X}+\mathbf{Z}) = \max_{K_{\mathbf{X}}:\mathrm{tr}(K_{\mathbf{X}})\le P} \frac{1}{2}\log|GK_{\mathbf{X}}G^T + I_r|$$

Summary

- DMC with DM state
- Channel coding with side information
- Shannon strategy
- Gelfand–Pinsker coding:
  - Multicoding (subcodebook generation)
  - Joint typicality encoding
- Covering lemma
- Writing on dirty paper
- Vector writing on dirty paper

References

Cohen, A. S. and Lapidoth, A. (2002). The Gaussian watermarking game. *IEEE Trans. Inf. Theory*, 48(6), 1639–1667.

Costa, M. H. M. (1983). Writing on dirty paper. *IEEE Trans. Inf. Theory*, 29(3), 439–441.

Gelfand, S. I. and Pinsker, M. S. (1980). Coding for channel with random parameters. *Probl. Control Inf. Theory*, 9(1), 19–31.

Goldsmith, A. J. and Varaiya, P. P. (1997). Capacity of fading channels with channel side information. *IEEE Trans. Inf. Theory*, 43(6), 1986–1992.

Heegard, C. and El Gamal, A. (1983). On the capacity of computer memories with defects. *IEEE Trans. Inf. Theory*, 29(5), 731–739.

Kuznetsov, A. V. and Tsybakov, B. S. (1974). Coding in a memory with defective cells. *Probl. Inf. Transm.*, 10(2), 52–60.

Shannon, C. E. (1958). Channels with side information at the transmitter. *IBM J. Res. Develop.*, 2(4), 289–293.