The capacity of the $(1, \infty)$-RLL Input-Constrained Erasure Channel with Feedback

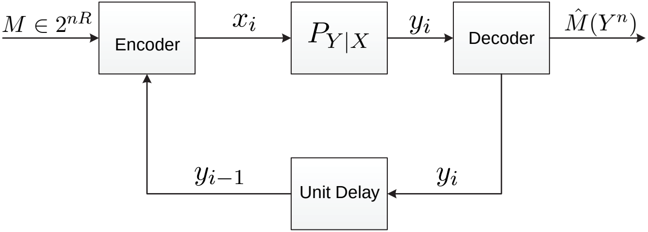
Oron Sabag, Haim Permuter and Navin Kashyap

Ben Gurion University and Indian Institute of Science

Information Forum
Stanford, Feb. 2015

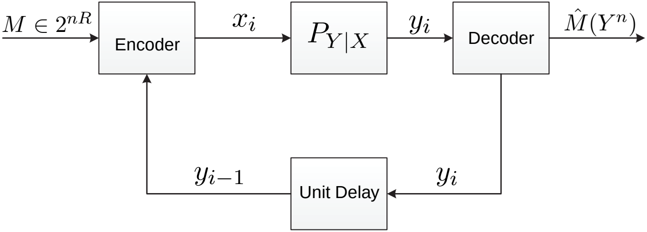
System model

- Feedback

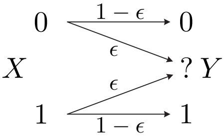


System model

- Feedback

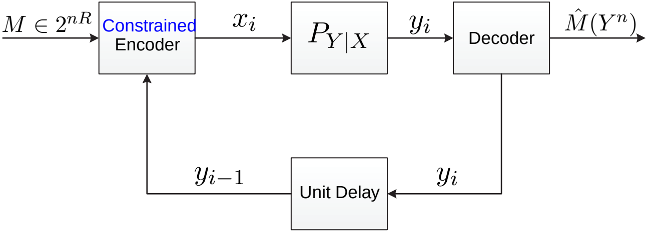


- The channel is the binary erasure channel:

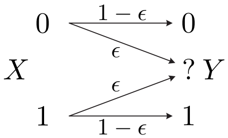


System model

- Feedback



- The channel is the binary erasure channel:



- The codes has a constraint: no two 1's in a row. (a.k.a. $(1, \infty)$-RLL constrained.)

Motivation and Questions

- Motivation:
  - Simple and fundamental problem: simple memoryless channel with simple input-constrainted.

Motivation and Questions

- Motivation:
  - Simple and fundamental problem: simple memoryless channel with simple input-constrainted.
  - The capacity of input-constraint channels without feedback is still an open problem [Zehavi/Wolf88], [Vontobel02],[Han/Marcus10], [Han/Marcus12], [Li/Han14]

Motivation and Questions

- Motivation:
  - Simple and fundamental problem: simple memoryless channel with simple input-constrainted.
  - The capacity of input-constraint channels without feedback is still an open problem [Zehavi/Wolf88], [Vontobel02],[Han/Marcus10], [Han/Marcus12], [Li/Han14]
  - Encoding for memories books by [Immink91], [Lind/Marcus96], [Marcus/Roth/Siegel98]

Motivation and Questions

- Motivation:
  - Simple and fundamental problem: simple memoryless channel with simple input-constrainted.
  - The capacity of input-constraint channels without feedback is still an open problem [Zehavi/Wolf88], [Vontobel02],[Han/Marcus10], [Han/Marcus12], [Li/Han14]
  - Encoding for memories books by [Immink91], [Lind/Marcus96], [Marcus/Roth/Siegel98]
  - Energy Harvesting and RFID [Fouladgar/Simeone/Erkip14]

Motivation and Questions

- Motivation:
  - Simple and fundamental problem: simple memoryless channel with simple input-constrainted.
  - The capacity of input-constraint channels without feedback is still an open problem [Zehavi/Wolf88], [Vontobel02],[Han/Marcus10], [Han/Marcus12], [Li/Han14]
  - Encoding for memories books by [Immink91], [Lind/Marcus96], [Marcus/Roth/Siegel98]
  - Energy Harvesting and RFID [Fouladgar/Simeone/Erkip14]
- Interesting Questions:
  - What is the capacity?
  - Can we find a simple coding scheme?
  - What is the non-causal capacity?

Our First Main Result

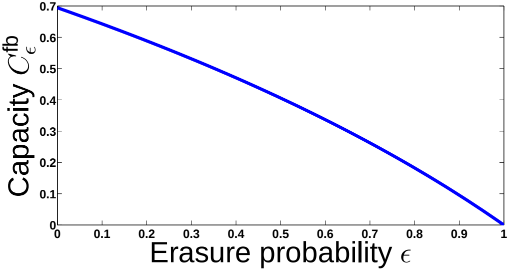
Theorem

The capacity of the $(1,\infty)$-RLL input-constrained erasure channel with feedback is

$$C_\epsilon^{\textit{fb}} = \max_{0\le p\le \frac{1}{2}} \frac{H_b(p)}{p+\frac{1}{1-\epsilon}}.$$

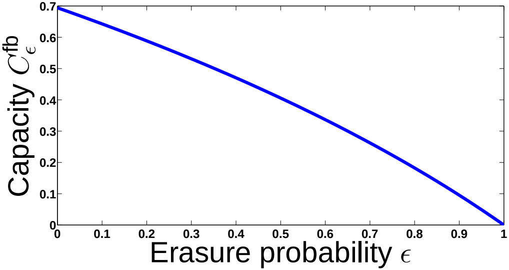
Properties of capacity expression

- The capacity $C_\epsilon^{\text{fb}} = \max_{0\leq p\leq \frac{1}{2}} \frac{H_b(p)}{p+\frac{1}{1-\epsilon}}$.



Properties of capacity expression

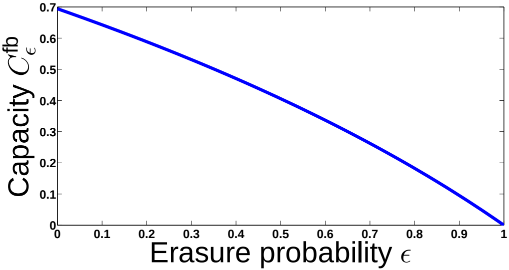
- The capacity $C_\epsilon^{\text{fb}} = \max_{0\leq p\leq \frac{1}{2}} \frac{H_b(p)}{p+\frac{1}{1-\epsilon}}$.



- Special cases:
  - When $\epsilon = 0$, we have $C_0^{\text{fb}} = \log_2 \phi$.

Properties of capacity expression

- The capacity $C_\epsilon^{\text{fb}} = \max_{0\leq p\leq \frac{1}{2}} \frac{H_b(p)}{p+\frac{1}{1-\epsilon}}$.



- Special cases:
  - When $\epsilon = 0$, we have $C_0^{\text{fb}} = \log_2 \phi$.
  - When $\epsilon = 1$, we have $C_1^{\text{fb}} = 0$.

Main Tools

1. Directed Information $I(X^n \to Y^n)$ and causal conditioning $P(x^n||y^{n-1})$
2. Dynamic Programming (infinite horizon average reward DP)

Definitions of Directed Information

$$I(X^n;Y^n) \triangleq H(Y^n) - H(Y^n|X^n)$$

$$H(Y^n|X^n) \triangleq E[-\log P(Y^n|X^n)]$$

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x^n, y^{i-1})$$

Definitions of Directed Information

Directed Information [Massey90] inspired by [Marko 73]

$$
\begin{aligned}
I(X^n \to Y^n) &\triangleq H(Y^n) - H(Y^n||X^n) \\
I(X^n; Y^n) &\triangleq H(Y^n) - H(Y^n|X^n)
\end{aligned}
$$

$$
H(Y^n|X^n) \triangleq E[-\log P(Y^n|X^n)]
$$

$$
P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x^n, y^{i-1})
$$

Definitions of Directed Information

Directed Information [Massey90] inspired by [Marko 73]

$$I(X^n \to Y^n) \triangleq H(Y^n) - H(Y^n||X^n)$$
$$I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n|X^n)$$

Causal Conditioning [Kramer98]

$$H(Y^n||X^n) \triangleq E[-\log P(Y^n||X^n)]$$
$$H(Y^n|X^n) \triangleq E[-\log P(Y^n|X^n)]$$

$$P(y^n|x^n) = \prod_{i=1}^{n} P(y_i|x^n, y^{i-1})$$

Definitions of Directed Information

Directed Information [Massey90] inspired by [Marko 73]

$$
\begin{aligned}
I(X^n \to Y^n) &\triangleq H(Y^n) - H(Y^n||X^n) \\
I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n)
\end{aligned}
$$

Causal Conditioning [Kramer98]

$$
\begin{aligned}
H(Y^n||X^n) &\triangleq E[-\log P(Y^n||X^n)] \\
H(Y^n|X^n) &\triangleq E[-\log P(Y^n|X^n)]
\end{aligned}
$$

$$
\begin{aligned}
P(y^n||x^n) &\triangleq \prod_{i=1}^{n} P(y_i|x^i, y^{i-1}) \\
P(y^n|x^n) &= \prod_{i=1}^{n} P(y_i|x^n, y^{i-1})
\end{aligned}
$$

Feedback capacity

For this channel we show:

$$C_{FB} = \lim_{n \to \infty} \frac{1}{n} \max_{Q(x^n || y^{n-1})} I(X^n \to Y^n)$$

Feedback capacity

For this channel we show:

$$
\begin{aligned}
C_{FB} &= \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n) \\
&= \sup_{\{p(x_t|x_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, X_{t-1}; Y_t|Y^{t-1})
\end{aligned}
$$

Feedback capacity

For this channel we show:

$$
\begin{aligned}
C_{FB} &= \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n) \\
&= \sup_{\{p(x_t|x_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, X_{t-1}; Y_t|Y^{t-1})
\end{aligned}
$$

This reminds us infinite-horizon average-reward dynamic programming.

Feedback capacity

For this channel we show:

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$
$$= \sup_{\{p(x_t|x_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, X_{t-1}; Y_t|Y^{t-1})$$

This reminds us infinite-horizon average-reward dynamic programming.

Feedback capacity

For this channel we show:

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$
$$= \sup_{\{p(x_t|x_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, X_{t-1}; Y_t|Y^{t-1})$$

This reminds us infinite-horizon average-reward dynamic programming.

Feedback capacity

For this channel we show:

$$\begin{aligned} C_{FB} &= \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n) \\ &= \sup_{\{p(x_t|x_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, X_{t-1}; Y_t|Y^{t-1}) \end{aligned}$$

This reminds us infinite-horizon average-reward dynamic programming.

Dynamic programming (DP) formulation

- state: β_{t-1}
- action: a_t
- disturbance: w_t

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(x_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: a_t
- disturbance: w_t

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(x_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|x_{t-1}), \quad a_t \in [0,1]$
- disturbance: w_t

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(x_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|x_{t-1}), \quad a_t \in [0,1]$
- disturbance: $w_t = y_{t-1}, \quad w_t \in \{0,1\}$

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(x_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|x_{t-1}), \quad a_t \in [0,1]$
- disturbance: $w_t = y_{t-1}, \quad w_t \in \{0,1\}$

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t) = I(X_t, X_{t-1}; Y_t|y^{t-1}).$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(x_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|x_{t-1}), \quad a_t \in [0,1]$
- disturbance: $w_t = y_{t-1}, \quad w_t \in \{0,1\}$

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t) = I(X_t, X_{t-1}; Y_t|y^{t-1}).$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

$$= \sup_{\{p(x_t|x_{t-1}, y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, X_{t-1}; Y_t|Y^{t-1}) = C_{FB}$$

Dynamic programming operator, T

- The dynamic programming operator T is given by

$$(T \circ J)(\beta) = \sup_{a \in \mathcal{A}} \left(g(\beta, a) + \sum_{w} P(w|\beta, a) J(F(\beta, a, w)) \right)$$

Dynamic programming operator, T

- The dynamic programming operator T is given by

$$(T \circ J)(\beta) = \sup_{a \in \mathcal{A}} \left(g(\beta, a) + \sum_{w} P(w|\beta, a) J(F(\beta, a, w)) \right)$$

- In our case: $J : [0,1] \to \mathbb{R}$ and

$$(T \circ J)(\beta) = \sup_{0 \leq \delta \leq z} \bar{\epsilon} H_b(\delta) + (1-\delta)\bar{\epsilon} J(1) + \epsilon J(1-\delta) + \delta \bar{\epsilon} J(0)$$

Dynamic programming operator, T

- The dynamic programming operator T is given by

$$(T \circ J)(\beta) = \sup_{a \in \mathcal{A}} \left(g(\beta, a) + \sum_{w} P(w|\beta, a) J(F(\beta, a, w)) \right)$$

- In our case: $J : [0, 1] \to \mathbb{R}$ and

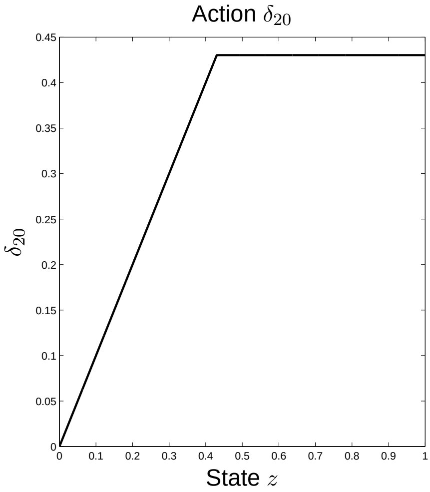
$$(T \circ J)(\beta) = \sup_{0 \leq \delta \leq z} \bar{\epsilon} H_b(\delta) + (1 - \delta)\bar{\epsilon} J(1) + \epsilon J(1 - \delta) + \delta \bar{\epsilon} J(0)$$

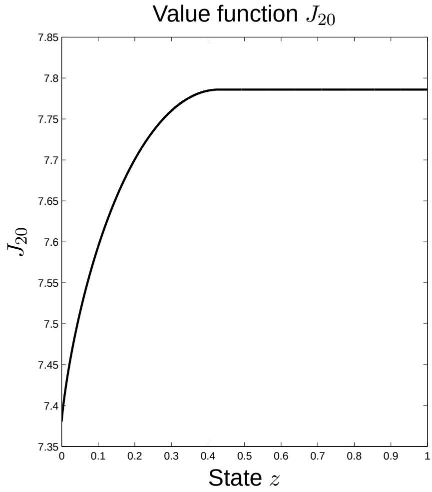
- Value iteration algorithms: executed 20 iterations

$$J_{k+1} = T \circ J_k$$

$$C_{FB} \approx 0.405 \text{ bits}$$

Result of value iteration





Dynamic programming- Bellman equation

Theorem

(Bellman Equation.) If there exist a function $J(\beta)$ and a constant ρ that satisfy

$$T \circ J(\beta) = J(\beta) + \rho$$

then ρ is the optimal infinite horizon average reward.

Dynamic programming- Bellman equation

Theorem

(Bellman Equation.) If there exist a function $J(\beta)$ and a constant ρ that satisfy

$$T \circ J(\beta) = J(\beta) + \rho$$

then ρ is the optimal infinite horizon average reward.

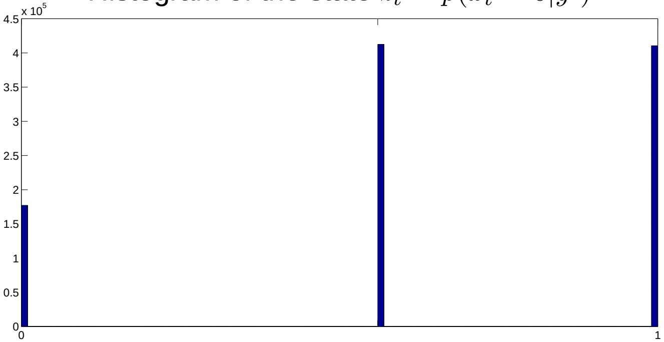
$$J^*(z) = \begin{cases} \bar{\epsilon} H_b(z) - z\bar{\epsilon}\frac{H_b(p_\epsilon)}{p_\epsilon + \frac{1}{1-\epsilon}} & \text{if } 0 \leq z \leq p_\epsilon \\ \frac{H_b(p_\epsilon)}{p_\epsilon + \frac{1}{1-\epsilon}} & \text{if } p_\epsilon \leq z \leq 1. \end{cases}$$

$$\rho_\epsilon^* = \max_{0 \leq p \leq 1} \frac{H_b(p)}{p + \frac{1}{1-\epsilon}},$$

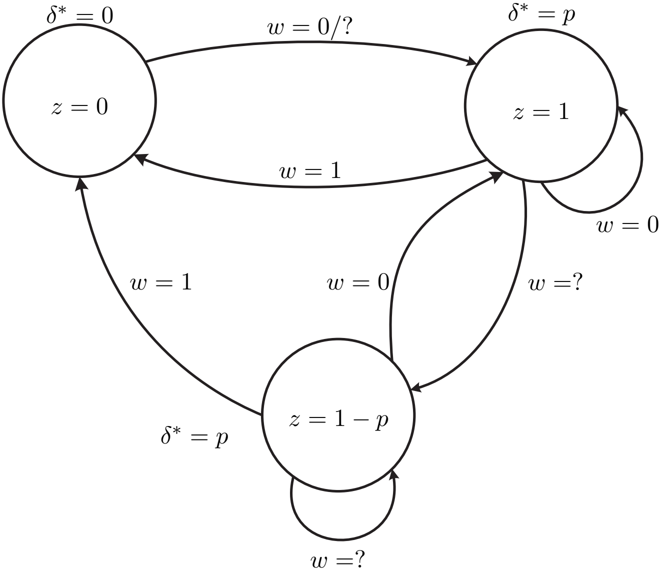
We showed that $J^*(z)$ and ρ_ϵ^* solve the Bellman equation.

Result of Value iteration

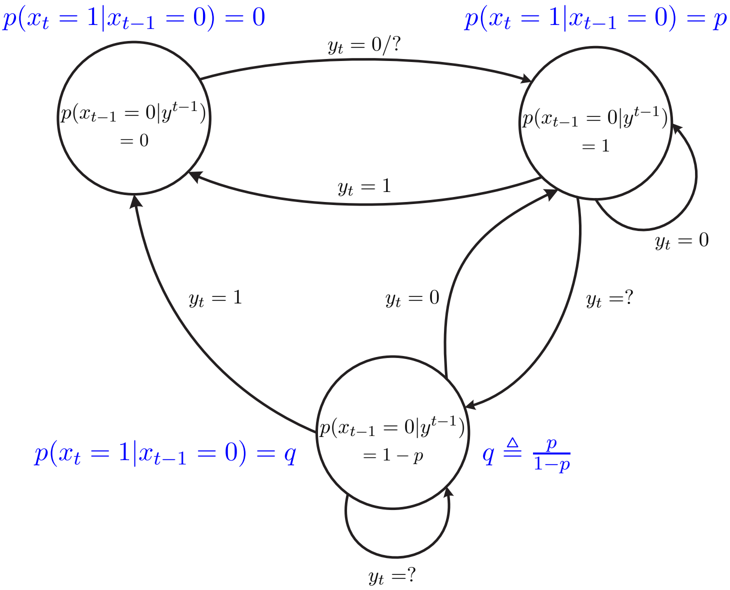
Histogram of the state $z_t \triangleq p(x_t = 0|y^t)$



The DP optimal policy

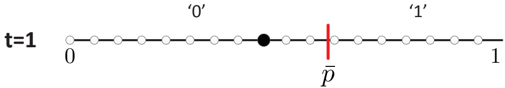


Optimal encoding procedure



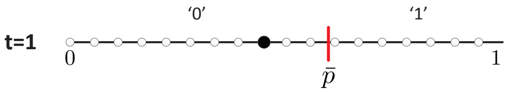
How do we transmit a message?

$M \sim$ points $[0,1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]



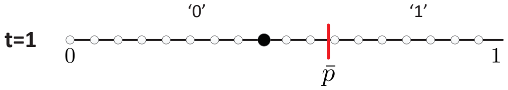
How do we transmit a message?

$M \sim$ points $[0,1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]
We partition the unit interval according to p or $q = \frac{p}{\bar{p}}$.



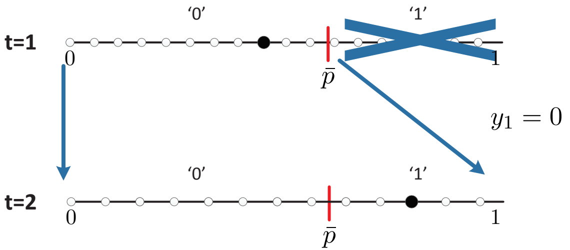
How do we transmit a message?

$M \sim$ points $[0, 1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]
We partition the unit interval according to p or $q = \frac{p}{\bar{p}}$.
The input to the channel is simply the label.



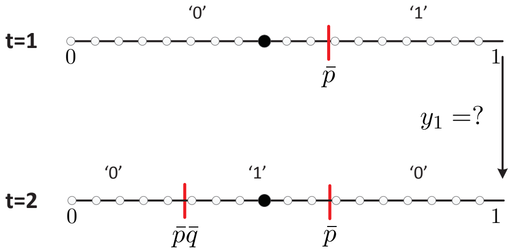
How do we transmit a message?

$M \sim$ points $[0,1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]
We partition the unit interval according to p or $q = \frac{p}{\bar{p}}$.
The input to the channel is simply the label.



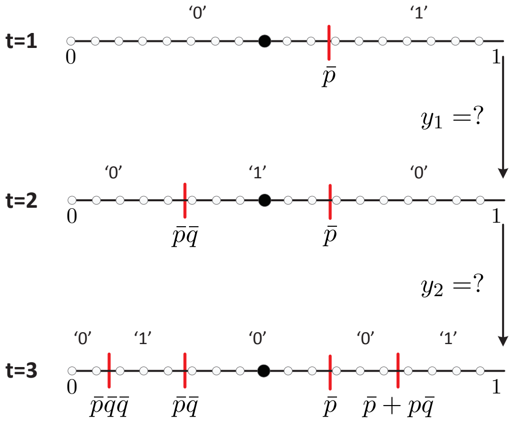
How do we transmit a message?

$M \sim$ points $[0,1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]
We partition the unit interval according to p or $q = \frac{p}{\bar{p}}$.
The input to the channel is simply the label.



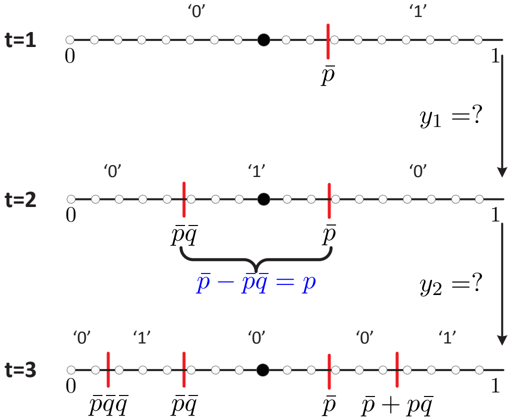
How do we transmit a message?

$M \sim$ points $[0, 1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]
We partition the unit interval according to p or $q = \frac{p}{\bar{p}}$.
The input to the channel is simply the label.



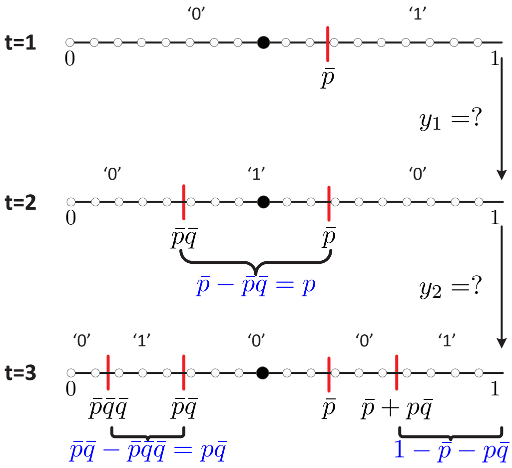
How do we transmit a message?

$M \sim$ points $[0,1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]
We partition the unit interval according to p or $q = \frac{p}{\bar{p}}$.
The input to the channel is simply the label.



How do we transmit a message?

$M \sim$ points $[0,1)$ [Horstein03 Schalkwijk/Kailath66 Kim06 Shayevitz/Feder11]
We partition the unit interval according to p or $q = \frac{p}{\bar{p}}$.
The input to the channel is simply the label.



Rate analysis

Lemma

At any step of the message transmission process, the lengths of the sub-intervals that are labelled by $'1'$ *sum up to* p*.*

- Therefore, any successful transmission provides $H_b(p)$ bits of the message.

Rate analysis

Lemma

At any step of the message transmission process, the lengths of the sub-intervals that are labelled by $'1'$ sum up to p.

- Therefore, any successful transmission provides $H_b(p)$ bits of the message.
- The expected number of channel uses for a successful transmission is
$$\frac{1}{1-\epsilon}+p.$$

Rate analysis

Lemma

At any step of the message transmission process, the lengths of the sub-intervals that are labelled by $'1'$ sum up to p.

- Therefore, any successful transmission provides $H_b(p)$ bits of the message.
- The expected number of channel uses for a successful transmission is
$$\frac{1}{1-\epsilon} + p.$$
- The rate of the scheme can be arbitrary close to $\frac{H_b(p)}{p+\frac{1}{1-\epsilon}}$.

Rate analysis

Lemma

At any step of the message transmission process, the lengths of the sub-intervals that are labelled by $'1'$ *sum up to* p*.*

- Therefore, any successful transmission provides $H_b(p)$ bits of the message.
- The expected number of channel uses for a successful transmission is

$$\frac{1}{1-\epsilon}+p.$$

- The rate of the scheme can be arbitrary close to $\frac{H_b(p)}{p+\frac{1}{1-\epsilon}}$.
- In particular choose $p=\arg\max_p \frac{H_b(p)}{p+\frac{1}{1-\epsilon}}$.

Summary and future work

- Obtained the capacity of fundamental feedback problem.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Near future work:
  - (d, k)-RLL constraints

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Near future work:
  - (d, k)-RLL constraints
  - Binary Symmetric Channel(BSC)

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Near future work:
  - (d,k)-RLL constraints
  - Binary Symmetric Channel(BSC)
  - general memoryless channel with general constrained coding.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Near future work:
  - (d,k)-RLL constraints
  - Binary Symmetric Channel(BSC)
  - general memoryless channel with general constrained coding.
- In the past we solved several unifilar FSC with feedback: Trapdoor channel, the Ising channel, POST channel and now erasure channel with constrained input.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Near future work:
  - (d, k)-RLL constraints
  - Binary Symmetric Channel(BSC)
  - general memoryless channel with general constrained coding.
- In the past we solved several unifilar FSC with feedback: Trapdoor channel, the Ising channel, POST channel and now erasure channel with constrained input.
  Long term goal: Find a unified solution for all FSC with feedback.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Near future work:
  - (d, k)-RLL constraints
  - Binary Symmetric Channel(BSC)
  - general memoryless channel with general constrained coding.
- In the past we solved several unifilar FSC with feedback: Trapdoor channel, the Ising channel, POST channel and now erasure channel with constrained input.

  Long term goal: Find a unified solution for all FSC with feedback.

More results at ITW, Jerusalem.

Thank you !

Does feedback increase capacity?

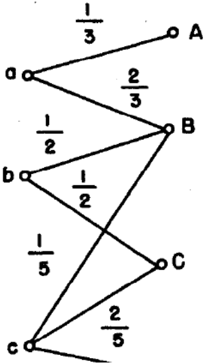
Does feedback increase capacity?

THE ZERO ERROR CAPACITY OF A NOISY CHANNEL

Claude E. Shannon
Bell Telephone Laboratories, Murray Hill, New Jersey
Massachusetts Institute of Technology, Cambridge, Mass.

Abstract

The zero error capacity C_0 of a noisy channel is defined as the least upper bound of rates at which it is possible to transmit information with zero probability of error. Various properties of C_0 are studied; upper and lower bounds and methods of evaluation of C_0 are given. Inequalities are obtained for the C_0 relating to the "sum" and "product" of two given channels. The analogous problem of zero error capacity C_{0F} for a channel with a feedback link is considered. It is shown that while the ordinary capacity of a memoryless channel with feedback is equal to that of the same channel without feedback, the zero error capacity may be greater. A solution is given to the problem of evaluating C_{0F}.



<u>Theorem 6</u>: In a memoryless discrete channel with feedback, the forward capacity is equal to the ordinary capacity C (without feedback). The average change in mutual information I_{vm} between received sequence v and message m for a letter of text is not greater than C.

Theorem 6: In a memoryless discrete channel with feedback, the forward capacity is equal to the ordinary capacity C (without feedback). The average change in mutual information I_{vm} between received sequence v and message m for a letter of text is not greater than C.

It is interesting that the first sentence of Theorem 6 can be generalized readily to channels with memory provided they are of such a nature that the internal state of the channel can be calculated at the transmitting point from the initial state and the sequence of letters that have been transmitted. If this is not the case, the conclusion of the theorem will not always be true, that is, there exist channels of a more complex sort for which the forward capacity with feedback exceeds that without feedback. We shall not, however, give the details of these generalizations here.

<u>Theorem 6</u>: In a memoryless discrete channel with feedback, the forward capacity is equal to the ordinary capacity C (without feedback). The average change in mutual information I_{vm} between received sequence v and message m for a letter of text is not greater than C.

It is interesting that the first sentence of Theorem 6 can be generalized readily to channels with memory provided they are of such a nature that the internal state of the channel can be calculated at the transmitting point from the initial state and the sequence of letters that have been transmitted. If this is not the case, the conclusion of the theorem will not always be true, that is, there exist channels of a more complex sort for which the forward capacity with feedback exceeds that without feedback. We shall not, however, give the details of these generalizations here.

__Theorem 6__: In a memoryless discrete channel with feedback, the forward capacity is equal to the ordinary capacity C (without feedback). The average change in mutual information I_{vm} between received sequence v and message m for a letter of text is not greater than C.

It is interesting that the first sentence of Theorem 6 can be generalized readily to channels with memory provided they are of such a nature that the internal state of the channel can be calculated at the transmitting point from the initial state and the sequence of letters that have been transmitted. If this is not the case, the conclusion of the theorem will not always be true, that is, there exist channels of a more complex sort for which the forward capacity with feedback exceeds that without feedback. We shall not, however, give the details of these generalizations here.

Question: Is Shannon's claim correct?