

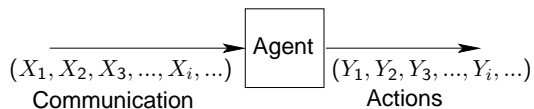
# Finite State Channels With and Without Feedback via Directed Information

Haim Permuter

Stanford University

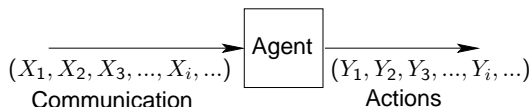
Information Theory and Applications  
University of California, San Deigo  
Jan 2008

# Measuring Causality



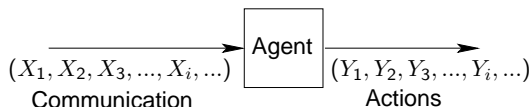
# Measuring Causality

**Question:** Are the actions  $Y^i$  *caused* by the communication  $X^i$ ?  
**not caused** if  $P(y_i|y^{i-1}, x^i) = P(y_i|y^{i-1})$ , e.g.,  $Y^i$  independent  $X^i$ .  
**caused** if  $P(y_i|y^{i-1}, x^i) \neq P(y_i|y^{i-1})$ , e.g.,  $Y_i = X_i$  &  $X_i \sim \text{i.i.d.}$



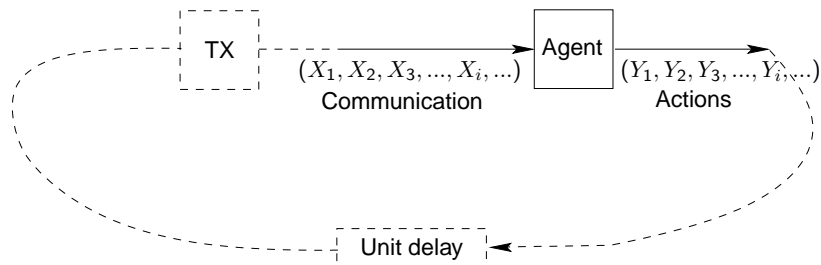
# Measuring Causality

**Question:** Are the actions  $Y^i$  *caused* by the communication  $X^i$ ?  
**not caused** if  $P(y_i|y^{i-1}, x^i) = P(y_i|y^{i-1})$ , e.g.,  $Y^i$  independent  $X^i$ .  
**caused** if  $P(y_i|y^{i-1}, x^i) \neq P(y_i|y^{i-1})$ , e.g.,  $Y_i = X_i$  &  $X_i \sim$  i.i.d.  
Does the mutual information  $I(X^n; Y^n)$  measure causality?



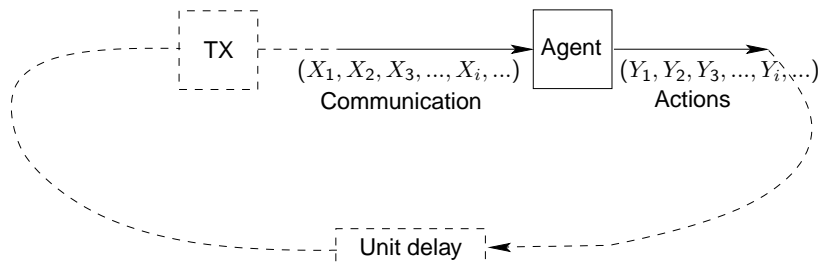
# Measuring Causality

**Question:** Are the actions  $Y^i$  *caused* by the communication  $X^i$ ?  
**not caused** if  $P(y_i|y^{i-1}, x^i) = P(y_i|y^{i-1})$ , e.g.,  $Y^i$  independent  $X^i$ .  
**caused** if  $P(y_i|y^{i-1}, x^i) \neq P(y_i|y^{i-1})$ , e.g.,  $Y_i = X_i$  &  $X_i \sim \text{i.i.d.}$   
Does the mutual information  $I(X^n; Y^n)$  measure causality?



# Measuring Causality

Does the mutual information  $I(X^n; Y^n)$  measure causality?

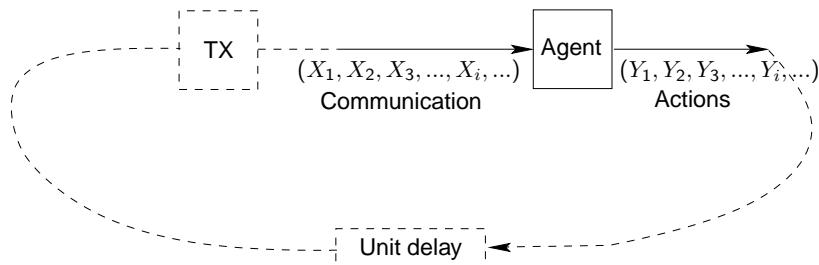


Example:  $Y_i$  i.i.d Bernoulli( $\frac{1}{2}$ ) and  $X_i = Y_{i-1}$ .

$$\frac{1}{n} I(X^n; Y^n) = \frac{n-1}{n} \rightarrow 1$$

# Measuring Causality

Does the mutual information  $I(X^n; Y^n)$  measure causality? **No.**



Example:  $Y_i$  i.i.d Bernoulli( $\frac{1}{2}$ ) and  $X_i = Y_{i-1}$ .

$$\frac{1}{n} I(X^n; Y^n) = \frac{n-1}{n} \rightarrow 1$$

# Directed Information and Causal Conditioning

*Directed Information*

[Marko73, Massey90]

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

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

# Directed Information and Causal Conditioning

## *Directed Information*

[Marko73, Massey90]

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

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

## *Causal Conditioning*

[Kramer98]

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

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

# Directed Information and Causal Conditioning

## *Directed Information*

[Marko73, Massey90]

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

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

## *Causal Conditioning*

[Kramer98]

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

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

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

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

# Directed Information and Causal Conditioning

## *Directed Information*

[Marko73, Massey90]

$$\begin{aligned} I(X^n \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n || X^n) \\ I(X^{n-1} \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n || X^{n-1}) \end{aligned}$$

## *Causal Conditioning*

[Kramer98]

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

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

# Directed Information and Causal Conditioning

## *Directed Information*

[Marko73, Massey90]

$$\begin{aligned} I(X^n \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n \| X^n) \\ I(X^{n-1} \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n \| X^{n-1}) \end{aligned}$$

## *Causal Conditioning*

[Kramer98]

$$\begin{aligned} H(Y^n \| X^n) &\triangleq E[\log P(Y^n \| X^n)] \\ H(Y^n \| X^{n-1}) &\triangleq E[\log P(Y^n \| X^{n-1})] \end{aligned}$$

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

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

# Directed Information and Causal Conditioning

## *Directed Information*

[Marko73, Massey90]

$$\begin{aligned} I(X^n \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n || X^n) \\ I(X^{n-1} \rightarrow Y^n) &\triangleq H(Y^n) - H(Y^n || X^{n-1}) \end{aligned}$$

## *Causal Conditioning*

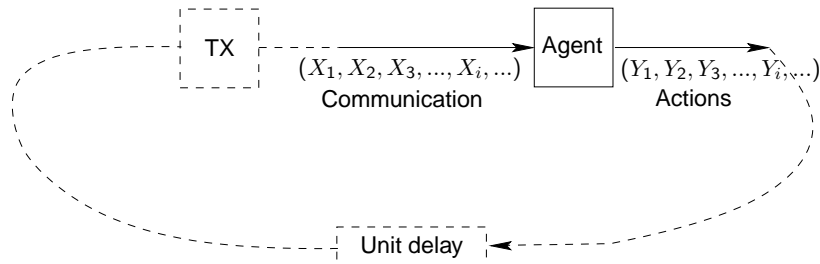
[Kramer98]

$$\begin{aligned} H(Y^n || X^n) &\triangleq E[\log P(Y^n || X^n)] \\ H(Y^n || X^{n-1}) &\triangleq E[\log P(Y^n || X^{n-1})] \end{aligned}$$

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

# Measuring Causality

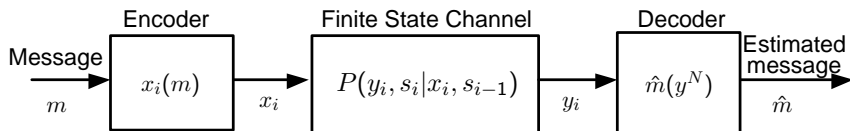
Consider  $I(X^n \rightarrow Y^n)$  for measuring causality.



Example:  $Y_i$  i.i.d Bernoulli( $\frac{1}{2}$ ) and  $X_i = Y_{i-1}$ .

$$\begin{aligned} I(X^n \rightarrow Y^n) &= H(Y^n) - H(Y^n || X^n) \\ &= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1}, X^i) = 0 \end{aligned}$$

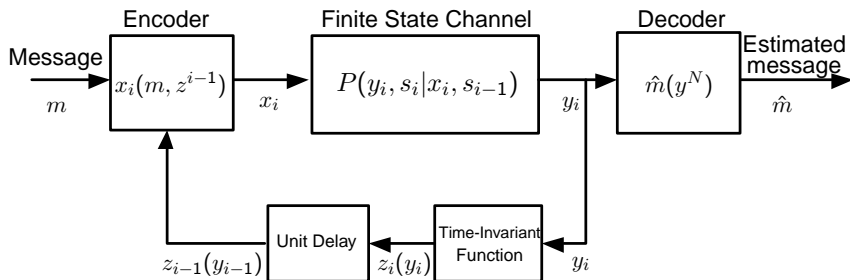
# Finite State Channels



Finite State Channel(FSC) property:

$$P(y_i, s_i | x^i, s^{i-1}, y^{i-1}) = P(y_i, s_i | x_i, s_{i-1})$$

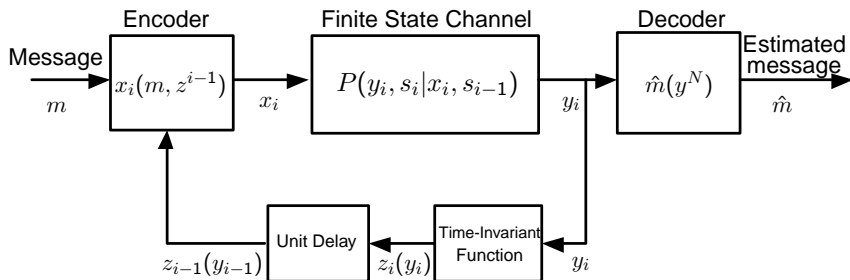
# Finite State Channels with and without feedback



Finite State Channel(FSC) property:

$$P(y_i, s_i | x^i, s^{i-1}, y^{i-1}) = P(y_i, s_i | x_i, s_{i-1})$$

# Finite State Channels with and without feedback



Finite State Channel(FSC) property:

$$P(y_i, s_i | x^i, s^{i-1}, y^{i-1}) = P(y_i, s_i | x_i, s_{i-1})$$

**Question:** What is the channel capacity in this setting?

## Theorem

For any FSC without feedback

[Gallager68]

$$C_{FB} \geq \frac{1}{n} \max_{P(x^n)} \min_{s_0} I(X^n; Y^n | s_0) - \frac{\log |\mathcal{S}|}{n}$$

$$C_{FB} \leq \frac{1}{n} \max_{P(x^n)} \max_{s_0} I(X^n; Y^n | s_0) + \frac{\log |\mathcal{S}|}{n}$$

## Theorem

For any FSC without feedback

[Gallager68]

$$C_{FB} \geq \frac{1}{n} \max_{P(x^n)} \min_{s_0} I(X^n; Y^n | s_0) - \frac{\log |\mathcal{S}|}{n}$$

$$C_{FB} \leq \frac{1}{n} \max_{P(x^n)} \max_{s_0} I(X^n; Y^n | s_0) + \frac{\log |\mathcal{S}|}{n}$$

## Theorem

For any FSC with feedback

[P.& Weissman& Goldsmith06]

$$C_{FB} \geq \frac{1}{n} \max_{P(x^n || z^n)} \min_{s_0} I(X^n \rightarrow Y^n | s_0) - \frac{\log |\mathcal{S}|}{n}$$

$$C_{FB} \leq \frac{1}{n} \max_{P(x^n || z^n)} \max_{s_0} I(X^n \rightarrow Y^n | s_0) + \frac{\log |\mathcal{S}|}{n}$$

# Key properties

$$P(x^n, y^n) = P(x^n || y^{n-1})P(y^n || x^n)$$

$$P(x^n, y^n) = P(x^n)P(y^n | x^n)$$

# Key properties

$$P(x^n, y^n) = P(x^n || y^{n-1})P(y^n || x^n)$$

$$P(x^n, y^n) = P(x^n)P(y^n | x^n)$$

$$\frac{1}{n} |I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n | S)| \leq \frac{H(S)}{n}$$

$$\frac{1}{n} |I(X^n; Y^n) - I(X^n; Y^n | S)| \leq \frac{H(S)}{n}$$

# Gallager's proof for the case of non-feedback

- 1 Randomly chosen codewords with distribution

$$P^*(x^n) = \arg \max_{P(x^n)} \min_{s_0} I(X^n; Y^n | s_0).$$

- 2 Maximum likelihood decoder :

$$m^* = \arg \max_m P(y^n | m) = \arg \max_m P(y^n | x^n(m)).$$

- 3 Show that if

$$R < \sum_{y^n} \sum_{x^n} P(x^n) \cdot P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} P(x^n) P(y^n | x^n)},$$

then  $P_e \rightarrow 0$ .

# Proof for the case of time-invariant feedback

- 1 Randomly chosen codewords , with distribution

$$P^*(x^n) = \arg \max_{P(x^n)} \min_{s_0} I(X^n; Y^n | s_0).$$

# Proof for the case of time-invariant feedback

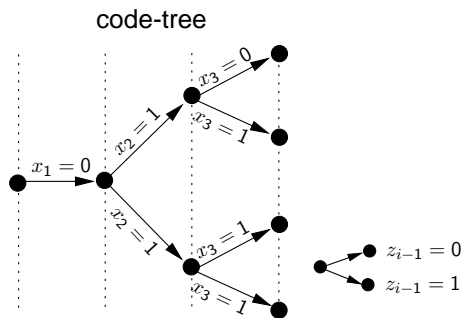
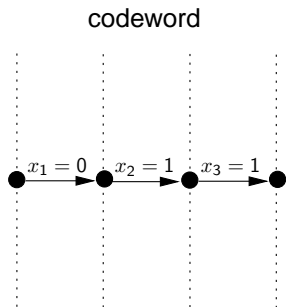
- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees**, with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$



# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

- 2 Maximum likelihood decoder :

$$m^* = \arg \max_m P(y^n | m) = \arg \max_m P(y^n | x^n).$$

# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

- 2 Maximum likelihood decoder :

$$m^* = \arg \max_m P(y^n | m) = \arg \max P(y^n || x^n).$$

# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

- 2 Maximum likelihood decoder :

$$m^* = \arg \max_m P(y^n | m) = \arg \max_m P(y^n || x^n).$$

- 3 If

$$R < \sum_{y^n} \sum_{x^n} P(x^n) \cdot P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} P(x^n) P(y^n | x^n)},$$

then  $P_e \rightarrow 0$ .

# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

- 2 Maximum likelihood decoder :

$$m^* = \arg \max_m P(y^n | m) = \arg \max_m P(y^n || x^n).$$

- 3 If

$$R < \sum_{y^n} \sum_{x^n} P(x^n || y^{n-1}) \cdot P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} P(x^n) P(y^n | x^n)},$$

then  $P_e \rightarrow 0$ .

# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

- 2 Maximum likelihood decoder :

$$m^* = \arg \max_m P(y^n | m) = \arg \max P(y^n || x^n).$$

- 3 If

$$R < \sum_{y^n} \sum_{x^n} P(x^n || y^{n-1}) \cdot P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} P(x^n || y^{n-1}) P(y^n | x^n)},$$

then  $P_e \rightarrow 0$ .

# Proof for the case of time-invariant feedback

- 1 Randomly chosen **codetrees** , with distribution

$$P^*(x^n || z^{n-1}) = \arg \max_{P(x^n || z^{n-1})} \min_{s_0} I(X^n \rightarrow Y^n | s_0).$$

- 2 Maximum likelihood decoder :

$$m^* = \arg \max_m P(y^n | m) = \arg \max P(y^n || x^n).$$

- 3 If

$$R < \sum_{y^n} \sum_{x^n} P(x^n || y^{n-1}) \cdot P(y^n || x^n) \ln \frac{P(y^n || x^n)}{\sum_{x^n} P(x^n || y^{n-1}) P(y^n || x^n)},$$

then  $P_e \rightarrow 0$ .

# Did we get directed information?

Recall first property:

$$P(x^n, y^n) = P(x^n || y^{n-1})P(y^n || x^n)$$

$$\sum_{x^n, y^n} P(x^n || y^{n-1}) \cdot P(y^n || x^n) \ln \frac{P(y^n || x^n)}{\sum_{x^n} P(x^n || y^{n-1}) \cdot P(y^n || x^n)} \stackrel{?}{=} I(X^n \rightarrow Y^n)$$

# Did we get directed information?

Recall first property:

$$P(x^n, y^n) = P(x^n || y^{n-1})P(y^n || x^n)$$

$$\sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n || x^n)}{\sum_{x^n} P(x^n, y^n)} = I(X^n \rightarrow Y^n).$$

# Did we get directed information?

Recall first property:

$$P(x^n, y^n) = P(x^n || y^{n-1})P(y^n || x^n)$$

$$\sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n || x^n)}{\sum_{x^n} P(x^n, y^n)} = I(X^n \rightarrow Y^n).$$

Does always switching between  $P(x^n) \leftrightarrow P(x^n || y^{n-1})$  and  $P(y^n | x^n) \leftrightarrow P(y^n || x^n)$  work?

# Did we get directed information?

Recall first property:

$$P(x^n, y^n) = P(x^n || y^{n-1})P(y^n || x^n)$$

$$\sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n || x^n)}{\sum_{x^n} P(x^n, y^n)} = I(X^n \rightarrow Y^n).$$

Does always switching between  $P(x^n) \leftrightarrow P(x^n || y^{n-1})$  and  $P(y^n | x^n) \leftrightarrow P(y^n || x^n)$  work?

No! For instance we have

$$\sum_u P(y^n, u | x^n) = P(y^n | x^n)$$

but in general

$$\sum_u P(y^n, u || x^n) \neq P(y^n || x^n)$$

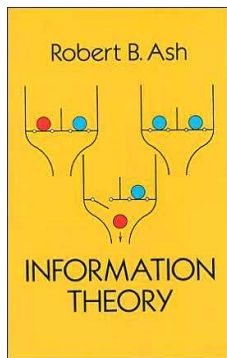
# Capacity Results

For some cases the upper bound and the lower bound coincide.

- 1 Initial state has positive probability for all states.
- 2 Compound channel setting (infinitely many FSCs  $P_{\theta}(y_i, s_i|x_i, s_{i-1}), \theta \in \Theta$ ).
- 3 Indecomposable FSC and no ISI.
- 4 The state is computable at the the encoder, and all states are connected.

# The trapdoor channel

Introduced by David Blackwell in 1961. [Ash65], [Ahlsvede & Kaspi 87], [Ahlsvede 98], [Kobayashi 02].



(a) Ash book

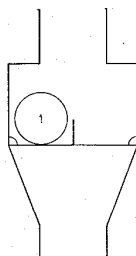
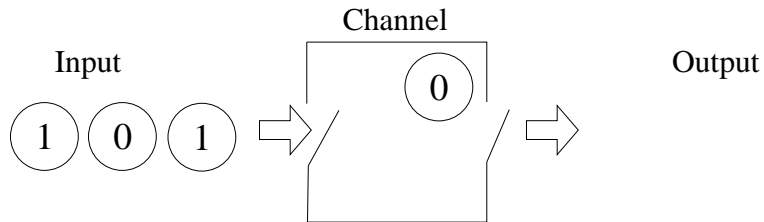


Fig. 7.1 A simple two-state channel.

(b) D. Blackwell

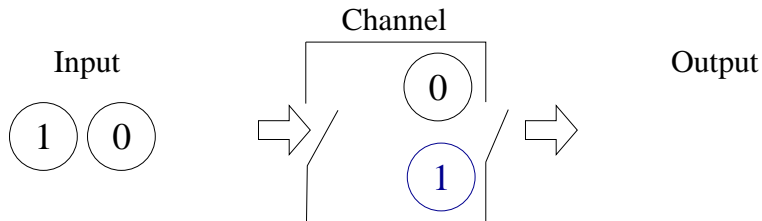
# The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$$s_0 = 0$$

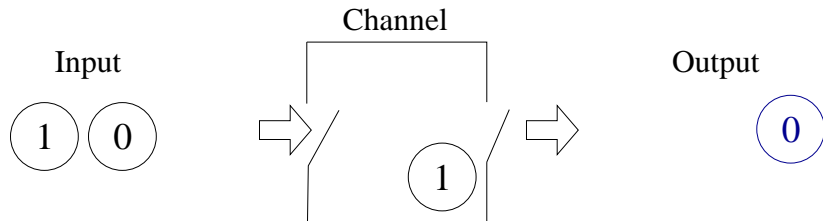
# The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$$s_0 = 0$$
$$x_1 = 1,$$

# The Trapdoor Channel

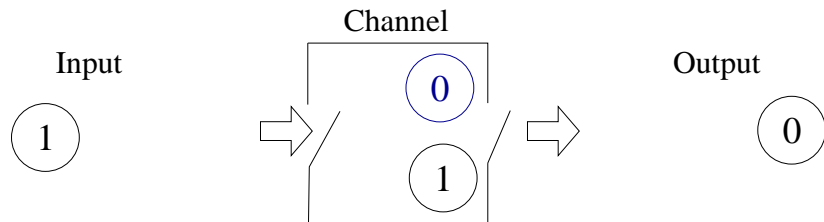


$$s_t = s_{t-1} + x_t - y_t$$

$$s_0 = 0$$

$$x_1 = 1, s_1 = 1, y_1 = 0,$$

# The Trapdoor Channel



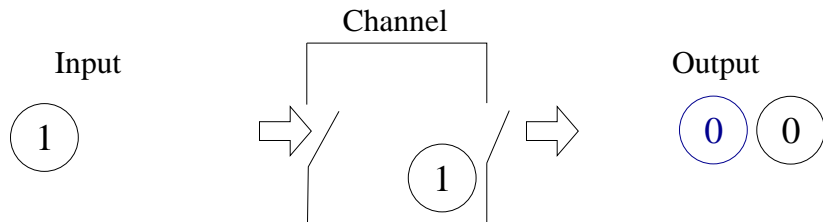
$$s_t = s_{t-1} + x_t - y_t$$

$$s_0 = 0$$

$$x_1 = 1, s_1 = 1, y_1 = 0,$$

$$x_2 = 0,$$

# The Trapdoor Channel



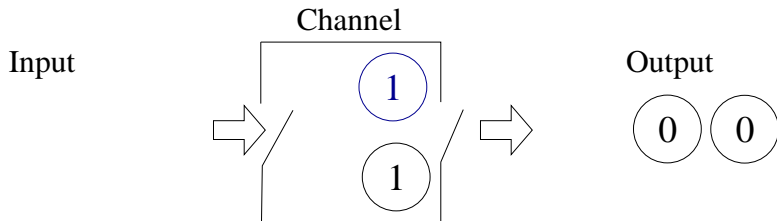
$$s_t = s_{t-1} + x_t - y_t$$

$$s_0 = 0$$

$$x_1 = 1, s_1 = 1, y_1 = 0,$$

$$x_2 = 0, s_2 = 1, y_2 = 0,$$

# The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

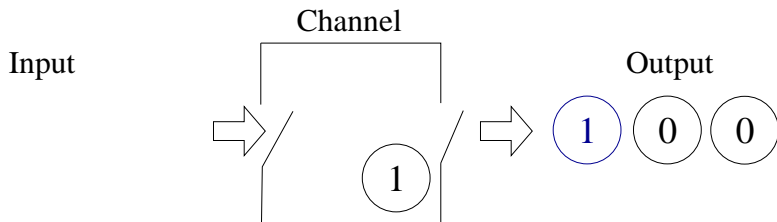
$$s_0 = 0$$

$$x_1 = 1, s_1 = 1, y_1 = 0,$$

$$x_2 = 0, s_2 = 1, y_2 = 0,$$

$$x_3 = 1,$$

# The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

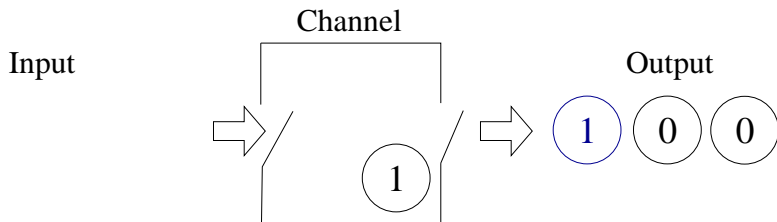
$$s_0 = 0$$

$$x_1 = 1, s_1 = 1, y_1 = 0,$$

$$x_2 = 0, s_2 = 1, y_2 = 0,$$

$$x_3 = 1, s_3 = 1, y_3 = 1.$$

# The Trapdoor Channel



$$s_t = s_{t-1} + x_t - y_t$$

$$s_0 = 0$$

$$x_1 = 1, s_1 = 1, y_1 = 0,$$

$$x_2 = 0, s_2 = 1, y_2 = 0,$$

$$x_3 = 1, s_3 = 1, y_3 = 1.$$

Biochemical Interpretation [Berger 71]

# Solving the trapdoor channel

- Initial state doesn't matter; upper and lower bounds become equal.

$$C_{FB} = \lim_{N \rightarrow \infty} \frac{1}{N} \max_{P(x^N || y^{N-1})} I(X^N \rightarrow Y^N)$$

- Converting it into a dynamic program. [Yang, Kavčić and Tatikonda05].
- Use value iteration algorithm to solve numerically the dynamic program.
- Verify the optimal solution through Bellman equation.

# Solving the dynamic programming

Executed 20 value iterations:  $J_{k+1} = T \circ J_k$

$$C_{FB} \approx 0.694 \text{ bits}$$

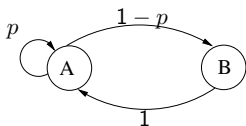
# Solving the dynamic programming

Executed 20 value iterations:  $J_{k+1} = T \circ J_k$

$$C_{FB} \approx 0.694 \text{ bits}$$

HW question posed in Prof. Cover's class

*Entropy rate.* Find the maximum entropy rate of the following two-state Markov chain:



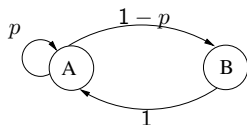
# Solving the dynamic programming

Executed 20 value iterations:  $J_{k+1} = T \circ J_k$

$$C_{FB} \approx 0.694 \text{ bits}$$

HW question posed in Prof. Cover's class

*Entropy rate.* Find the maximum entropy rate of the following two-state Markov chain:



*Solution:*  $H(\mathcal{X}) = \log \phi = 0.6942\dots$  bits (Golden Ratio:  $\phi = \frac{\sqrt{5}+1}{2}$ )

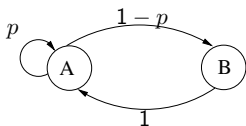
# Solving the dynamic programming

Executed 20 value iterations:  $J_{k+1} = T \circ J_k$

$$C_{FB} \approx 0.694 \text{ bits}$$

HW question posed in Prof. Cover's class

*Entropy rate.* Find the maximum entropy rate of the following two-state Markov chain:



*Solution:*  $H(\mathcal{X}) = \log \phi = 0.6942\dots \text{ bits}$  (Golden Ratio:  $\phi = \frac{\sqrt{5}+1}{2}$ )

## Theorem

*(Bellman Equation.) If there exists a function  $J(\beta)$  and a constant  $\rho$  that satisfy*

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

*then  $\rho$  is the optimal infinite horizon average reward.*

# Dynamic programming- Bellman equation

## Theorem

*(Bellman Equation.) If there exists a function  $J(\beta)$  and a constant  $\rho$  that satisfy*

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

*then  $\rho$  is the optimal infinite horizon average reward.*

By constructing  $J(\beta)$  and  $\rho = \log \frac{\sqrt{5}+1}{2}$  that satisfies the Bellman equation we conclude that

$$C_{fb} = \log \frac{\sqrt{5} + 1}{2}.$$

# A scheme that achieves capacity

Let  $r^n$  denotes a sequences of length  $n$  with no two consecutive 1's.

00101010100101...

## A simple scheme

*Encoder:* Map each message  $m$  to a sequence  $[r^n(m)]$ .

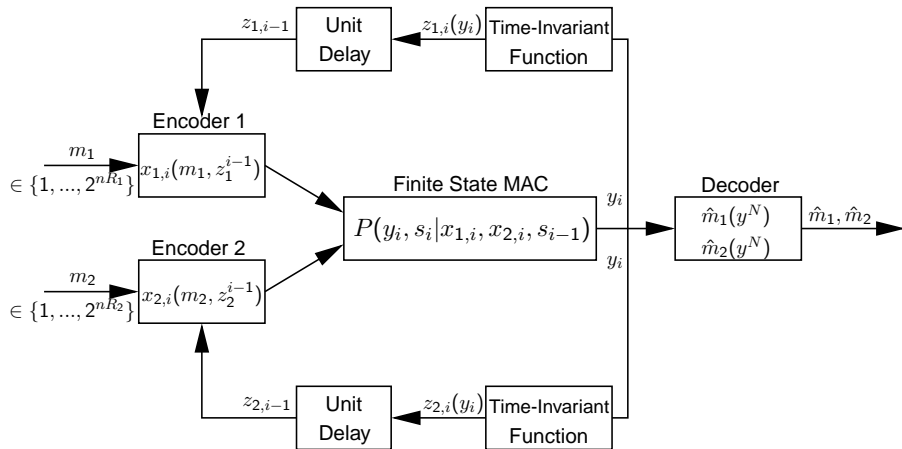
*Decoder:* The decoder decodes the sequence backward!

$$r_i = 1 \quad \Rightarrow \quad r_{i-1} = 0$$

$$r_i = 0 \quad \Rightarrow \quad r_{i-1} = y_i \oplus y_{i-1}$$

[P.&Cuff&Van-Roy&Weissman06]

# MAC with time-invariant feedback



Let

$$\underline{\mathcal{R}}_n = \bigcup \left\{ \begin{array}{l} R_1 \leq \min_{s_0} \frac{1}{n} I(X_1^n \rightarrow Y^n || X_2^n, s_0) - \frac{\log |\mathcal{S}|}{n}, \\ R_2 \leq \min_{s_0} \frac{1}{n} I(X_2^n \rightarrow Y^n || X_1^n, s_0) - \frac{\log |\mathcal{S}|}{n}, \\ R_1 + R_2 \leq \min_{s_0} \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n | s_0) - \frac{\log |\mathcal{S}|}{n}, \end{array} \right.$$

the union is over input distribution  $P(x_1^n || z_1^{n-1})P(x_2^n || z_2^{n-1})$ .

## Theorem

*For any FS-MAC with time invariant feedback,  $\underline{\mathcal{R}}_n$  is an inner bound.*

*The outer bound is given in terms of limit.*

*P.&Weissman&Chen07*

# An Ingredient of the Proof: Sup-additivity

Let  $a_n$  be a bounded sequence of real numbers.

If  $na_n$  is sup-additive, i.e., for all  $n > k$

$$na_n \geq ka_k + (n - k)a_{n-k},$$

then

$$\lim_{n \rightarrow \infty} a_n = \sup_n a_n.$$

# An Ingredient of the Proof: Sup-additivity

Let  $a_n$  be a bounded sequence of real numbers.  
If  $na_n$  is sup-additive, i.e., for all  $n > k$

$$\begin{aligned}na_n &\geq ka_k + (n - k)a_{n-k}, \\a_n &\geq \frac{k}{n}a_k + \frac{(n - k)}{n}a_{n-k},\end{aligned}$$

then

$$\lim_{n \rightarrow \infty} a_n = \sup_n a_n.$$

# An Ingredient of the Proof: Sup-additivity

Let  $A_n$  be a bounded sequence of **2D sets**.

If  $nA_n$  is sup-additive, i.e., for all  $n > k$

$$nA_n \supseteq kA_k + (n - k)A_{n-k},$$

$$A_n \supseteq \frac{k}{n}A_k + \frac{(n - k)}{n}A_{n-k},$$

then

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{n \geq 1} A_n.$$

# An Ingredient of the Proof: Sub-additivity

Let  $A_n$  be a bounded sequence of **2D convex sets**.

If  $nA_n$  is sub-additive, i.e., for all  $n > k$

$$nA_n \subseteq kA_k + (n - k)A_{n-k},$$

$$A_n \subseteq \frac{k}{n}A_k + \frac{(n - k)}{n}A_{n-k},$$

then

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n \geq 1} A_n.$$

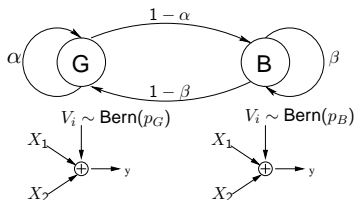
# Applications

- $\mathcal{R} = 0 \iff \mathcal{R}_{fb} = 0$

The proof is based on the fact that

$$\max_{Q(x^n|y^{n-1})} I(X^n \rightarrow Y^n) = 0 \iff \max_{Q(x^n)} I(X^n; Y^n) = 0$$

- Gillbert-Elliot MAC.



- feedback does not increase capacity.
- source-channel separation theorem holds for the lossless case.

Consider a horse-race market

- $X_i$  - the horse that wins at time  $i$ .
- $Y_i$  - side information available at time  $i$ .

Consider a horse-race market

- $X_i$  - the horse that wins at time  $i$ .
- $Y_i$  - side information available at time  $i$ .

$(X_i, Y_i)$ , i.i.d

Kelly[56]

The optimal strategy is to invest the capital proportional to  $P(x|y)$ . The increase in the growth rate due to side information  $Y$  is

$$\Delta W = nI(X; Y).$$

# Portfolio Theory

Consider a horse-race market

- $X_i$  - the horse that wins at time  $i$ .
- $Y_i$  - side information available at time  $i$ .

$(X_i, Y_i)$ , i.i.d

Kelly[56]

The optimal strategy is to invest the capital proportional to  $P(x|y)$ . The increase in the growth rate due to side information  $Y$  is

$$\Delta W = nI(X; Y).$$

$(X_i, Y_i)$  general processes

[P.&Kim&Weissman08]

The optimal strategy is to invest the capital proportional to  $P(x_i|x^{i-1}, y^i)$ . The increase in the growth rate due to side information is

$$\Delta W = I(Y^n \rightarrow X^n).$$

## Intuition

$I(X^n; Y^n)$  amount of uncertainty about  $Y^n$  reduced by knowing  $X^n$

$I(X^n \rightarrow Y^n)$  amount of uncertainty about  $Y^n$  reduced by knowing  $X^n$  **causally**.

## Results

- capacity of point-to-point FSC (e.g., trapdoor channel)
- capacity of FS-MAC (e.g., Gilbert-Elliot MAC)
- the increase in growth rate due to side information

# Other results with directed information

[Marko73]

[Kramer98]

[Chen/Berger05]

[Tatikonda/Mitter07]

[Venkataramanan/Pradhan07]

[Massey90]

[Tatikonda00]

[Yang/Kavcic/Tatikonda05]

[Kim07]

## Academic advisors

Tsachy Weissman

Tom Cover

## Collaborators

Jun Chen, Paul Cuff, Young-Han Kim, Andrea Goldsmith,  
Brooke Shrader, Benjamin Van-Roy.

# Thank You

Academic advisors

Tsachy Weissman

Tom Cover

Collaborators

Jun Chen, Paul Cuff, Young-Han Kim, Andrea Goldsmith,  
Brooke Shrader, Benjamin Van-Roy.

*Thank you for attending the talk!*