Efficient Communication for Control in Games and Networks

Paul Cuff

Princeton University
ISS Seminar

Oct. 8, 2009

Control of UAV's in Formation

Soon-Jo Chung
University of Illinois (UIUC)

Randy Beard
Brigham Young University

Control of UAV's in Formation

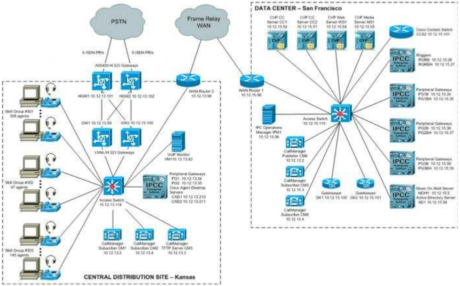
Soon-Jo Chung
University of Illinois (UIUC)

Randy Beard
Brigham Young University

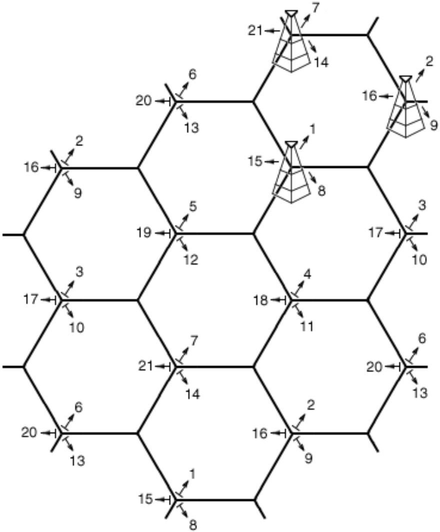
Network Traffic Control

Data Center (Cisco)

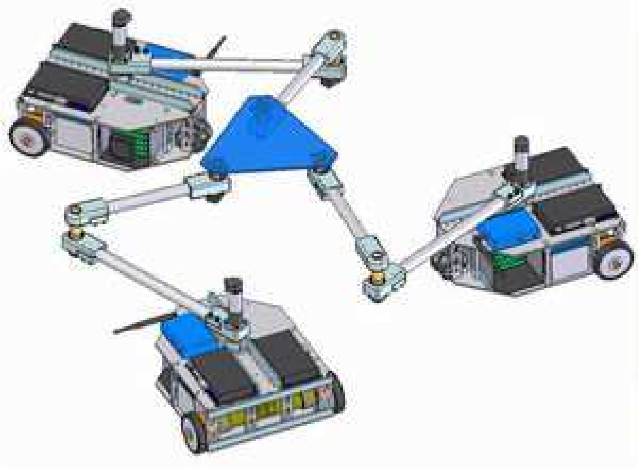
Balaji Prabhakar
Stanford University



Cellular Communication Systems

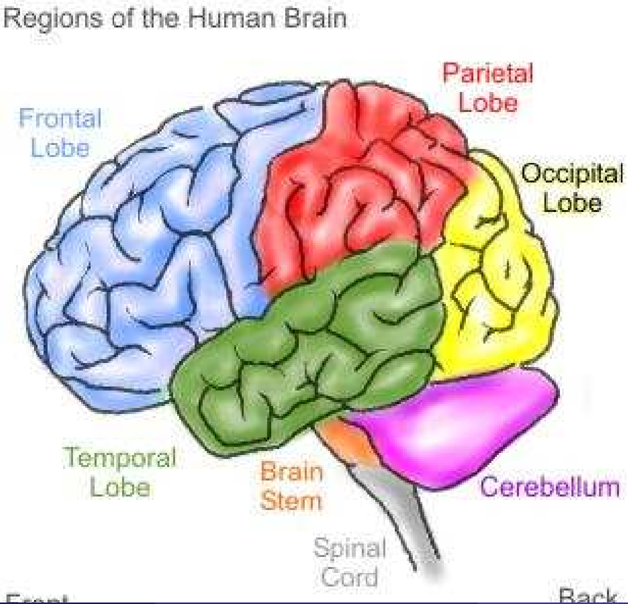


Distributed Sensing and Control



Competitive Settings

Parallel Processing in the Brain



Feedback (not part of this talk)

Control theory has many complexities.
Set them all aside for today.

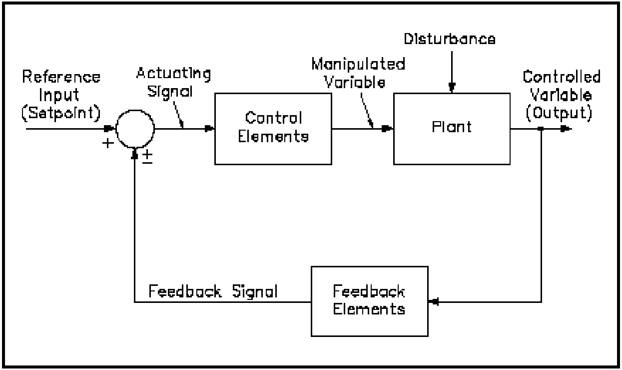


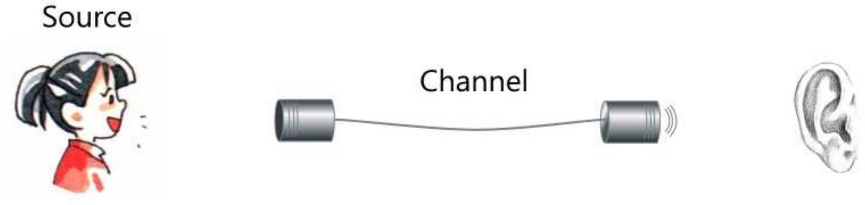
Figure 8 Feedback Control System Block Diagram

Focus of this Talk

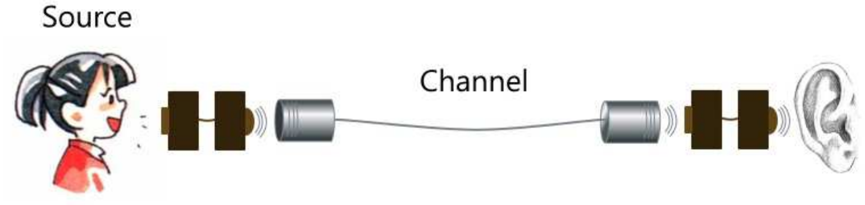
Apply Information Theory

- Rate-Distortion Theory
- Bayesian Games
- Strong Coordination and Common Information
- Networks

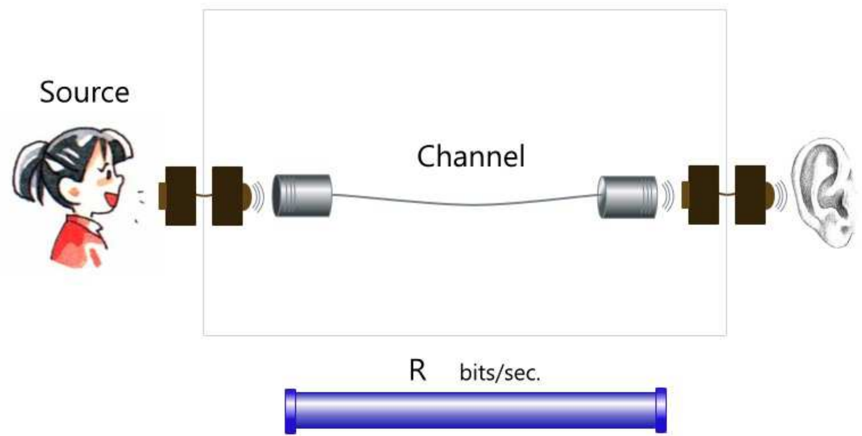
Source-Channel Separation



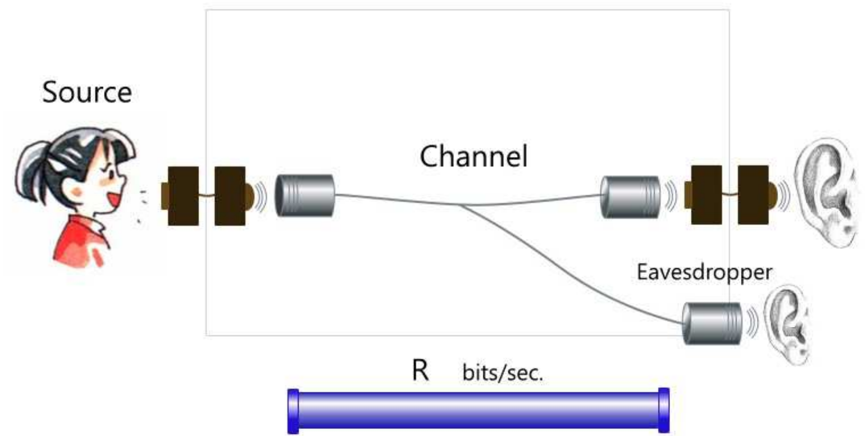
Source-Channel Separation



Source-Channel Separation



Source-Channel Separation



Compression

Lossless compression:

- Huffman codes
- Arithmetic coding
- Entropy

"Brevity is the Soul of Wit"

Lossy Compression

Rate-Distortion Problem:

Information: $X \sim p_X(x)$.
Reconstruction: $\hat{X}$.
Distortion measurement: $d(x, \hat{x})$.
Constraints: R bits per symbol, D distortion allowed.

Lossy Compression

Rate-Distortion Problem:

Information: $X \sim p_X(x)$.
Reconstruction: $\hat{X}$.
Distortion measurement: $d(x, \hat{x})$.
Constraints: R bits per symbol, D distortion allowed.

Problem setup:

X_i is a sequence of independent samples from $p_X(x)$.
Describe n symbols using nR bits.
Based on the description, create a reconstruction $\hat{X}_1, \hat{X}_2$... that satisfies $\mathbb{E}\frac{1}{n}\sum_{i=1}^{n} d(x, \hat{x}) \leq D$.

Lossy Compression

Rate-Distortion Problem:

Information: $X \sim p_X(x)$.
Reconstruction: $\hat{X}$.
Distortion measurement: $d(x, \hat{x})$.
Constraints: R bits per symbol, D distortion allowed.

Problem setup:

X_i is a sequence of independent samples from $p_X(x)$.
Describe n symbols using nR bits.
Based on the description, create a reconstruction $\hat{X}_1, \hat{X}_2$... that satisfies $\mathbb{E}\frac{1}{n}\sum_{i=1}^{n} d(x, \hat{x}) \leq D$.

For which R and D is this possible?

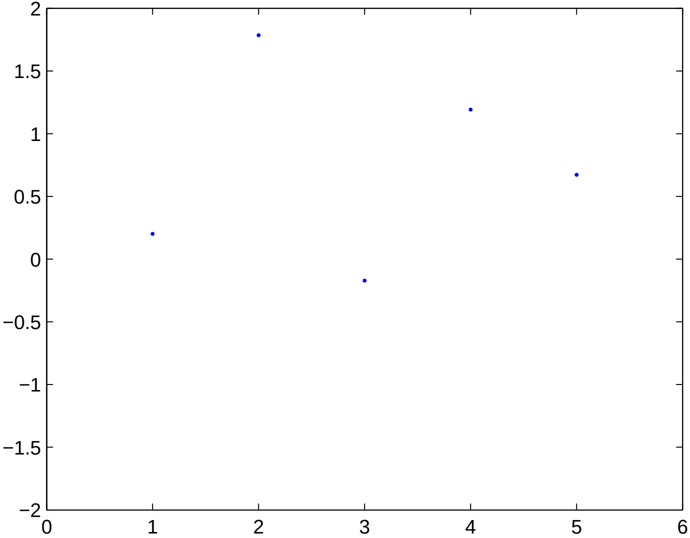
Rate-Distortion Result

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

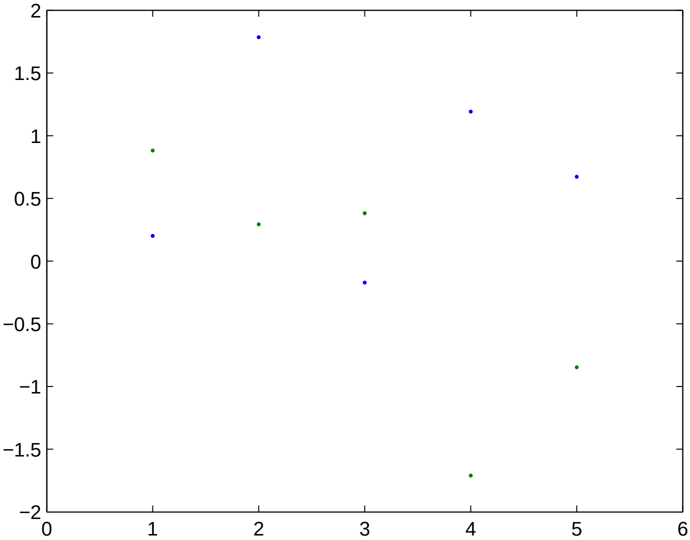
such that $\mathbb{E}d(X,\hat{X}) \leq D$.

$$\begin{aligned} I(X;\hat{X}) &= H(X) + H(\hat{X}) - H(X,\hat{X}), \\ H(X) &= \mathbb{E}\log\frac{1}{p(X)}. \end{aligned}$$

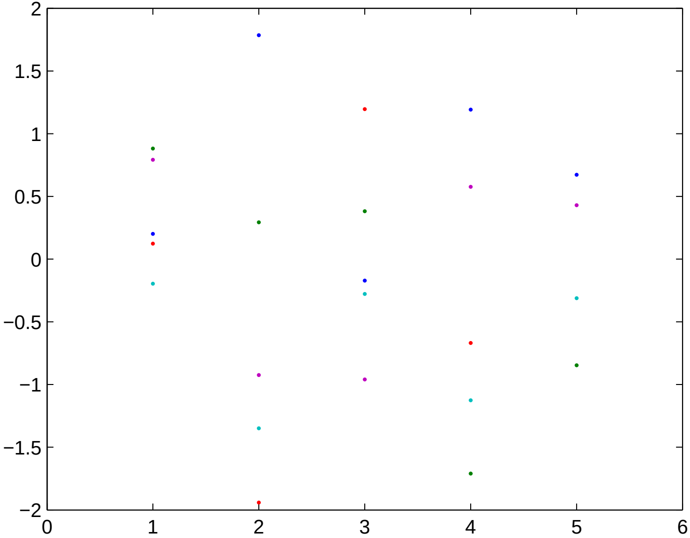
Rate-Distortion Method



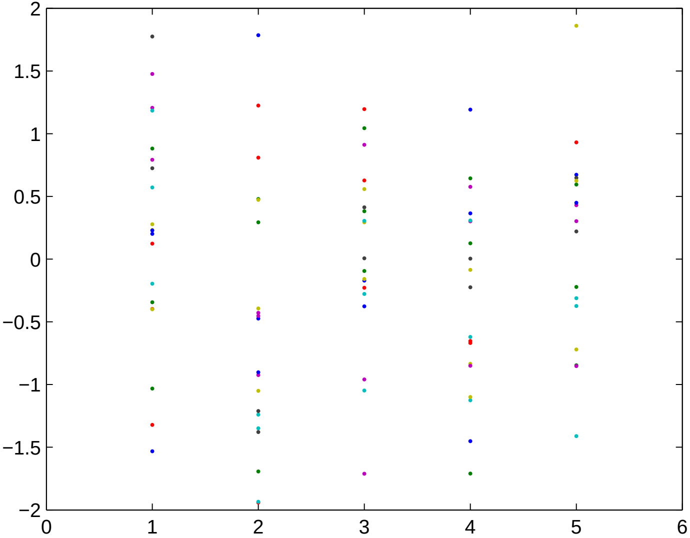
Rate-Distortion Method



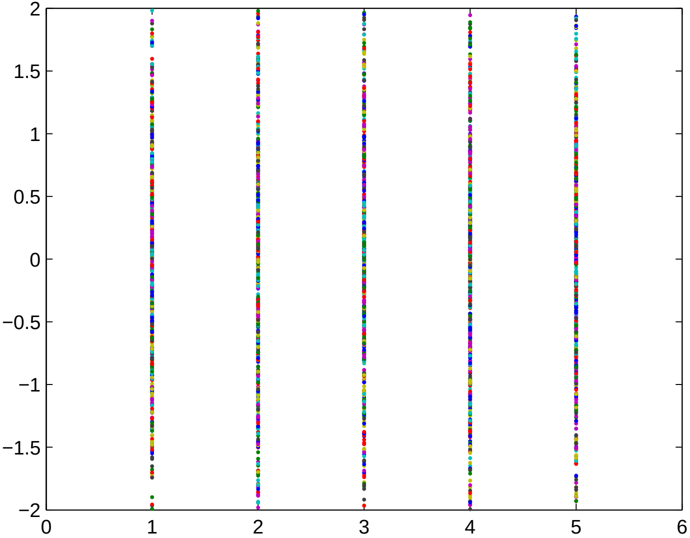
Rate-Distortion Method



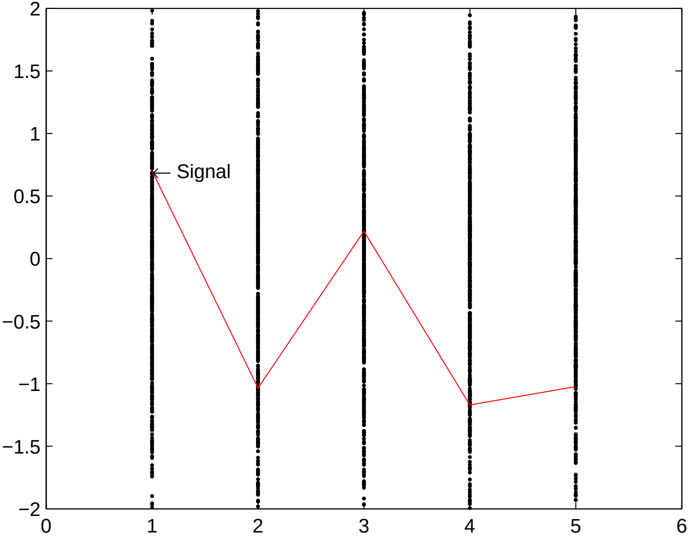
Rate-Distortion Method



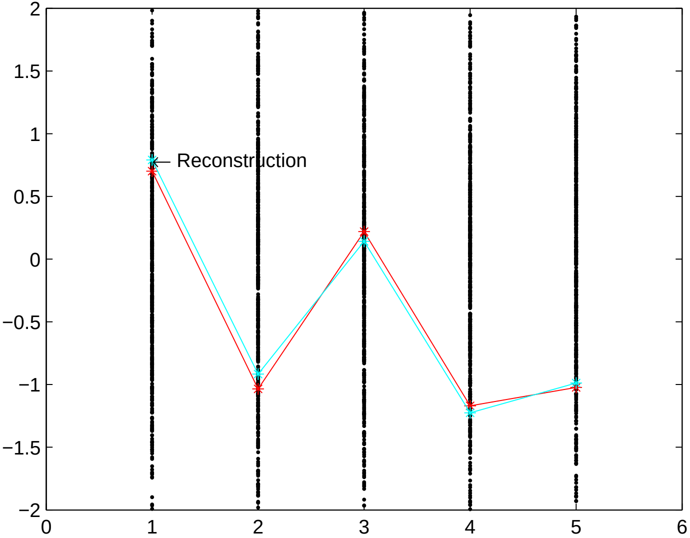
Rate-Distortion Method



Rate-Distortion Method



Rate-Distortion Method



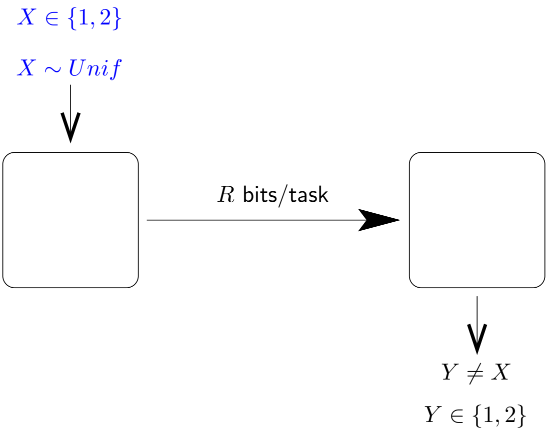
What does JPEG do?

Do practical compression schemes follow the method prescribed my rate-distortion theory?

20	−7	−1	1	−2	1	0	0
1	0	0	0	1	0	0	0
−1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

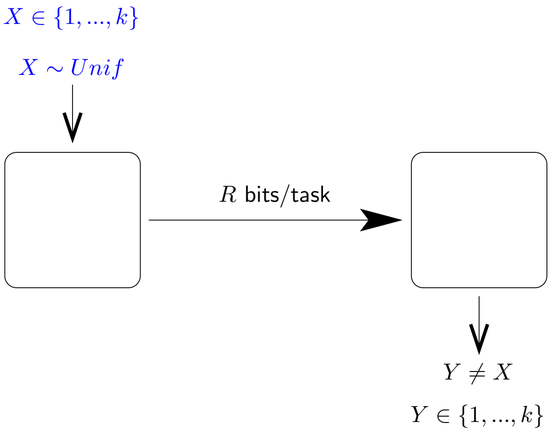
Coordinating tasks using R-D Theory

Tasks are identified by numbers.



Coordinating tasks using R-D Theory

Tasks are identified by numbers.



Buffer of Tasks

Example ($R = \frac{1}{4}$, $k = 5$):

Tasks assigned to X: X_1 X_2 ... each independent

Sample realization: 3 1 2 5 3 5 2 4

Buffer of Tasks

Example ($R = \frac{1}{4}$, $k = 5$):

Tasks assigned to X:	X_1	X_2	...		each independent			
Sample realization:	3	1	2	5	3	5	2	4
Message bits:			$b_1 b_2 = 01$					

Buffer of Tasks

Example ($R = \frac{1}{4}$, $k = 5$):

Tasks assigned to X: X_1 X_2 ... each independent

Sample realization: $3 \quad 1 \quad 2 \quad 5 \quad 3 \quad 5 \quad 2 \quad 4$

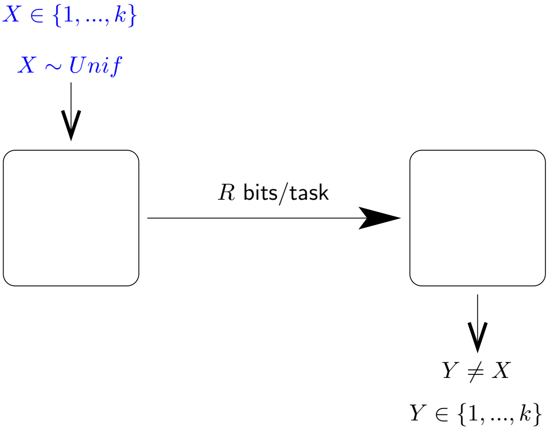
Message bits: $b_1 b_2 = 01$

Codebook:

b_1 b_2	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
0 0	1	2	3	4	5	1	2	3
0 1	2	4	1	3	5	2	4	1
1 0	...							
1 1	...							

Two Nodes

Tasks are identified by numbers.



Rate-Distortion Result

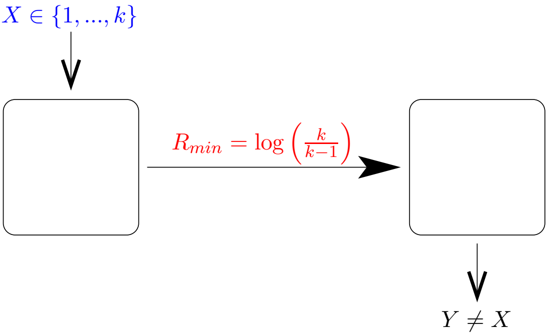
$$R_{min} = \min_{p(y|x)} I(X;Y)$$

such that $X \neq Y$ with probability 1.

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y), \\ H(X) &= \mathbb{E} \log \frac{1}{p(X)}. \end{aligned}$$

Two Node Result

Optimal two node task assignment rate:



Communication of State in Bayesian Games

Zero-Sum Game

Payoff Matrix Π for Player A:

B

A $p_A(a)$	0	1
0	3	0
1	0	1

$\square$

Zero-Sum Game

Payoff Matrix Π for Player A:

		B: 0	B: 1
A $p_A(a)$	0	3	0
	1	0	1

$$\text{Value of game} = \max_{p_A} \min_{p_B} \mathbb{E}\, \Pi(A,B) = 3/4.$$

$$p_A^*(a) = [1/4, 3/4].$$

General Zero-Sum Game

Payoff Matrix Π for Player A:

B

A

$p_A(a)$

	0	1
0	$\Pi(0,0)$	$\Pi(0,1)$
1	$\Pi(1,0)$	$\Pi(1,1)$

General Zero-Sum Game

Payoff Matrix Π for Player A:

A, $p_A(a)$ \ B	0	1
0	$\Pi(0,0)$	$\Pi(0,1)$
1	$\Pi(1,0)$	$\Pi(1,1)$

Allow payoff Π to be random, determined by a state S.

(In the literature, S is called the "type")

Erasure Game (two states)

S is equally likely to be 0 or 1:

$S = 0$

A: $p_{A|S}(a|0)$, B

	B = 0	B = 1
A = 0	3	0
A = e	0	1
A = 1	$-\infty$	$-\infty$

$S = 1$

A: $p_{A|S}(a|1)$, B

	B = 0	B = 1
A = 0	$-\infty$	$-\infty$
A = e	1	0
A = 1	0	3

Bayesian Games

In a Bayesian game, the players each may or may not have some information about the stochastic state.

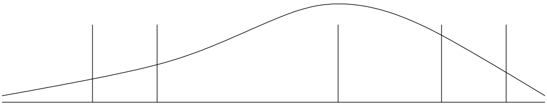
Recent related references:

- Gossner and Mertens (2001). *The value of information in zero-sum games.*
- Lehrer and Rosenberg (2004). *What restrictions do Bayesian games impose on the value of information?*
- Provan (2008). *The use of spies in strategic situations.*

Information Structure

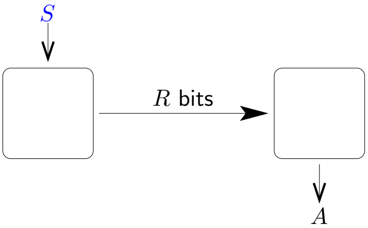
Information structure: the partition by which the state is quantized before being observed by a player of the game.

Distribution of S.



Quantization into bins.

Communication of State Information



Questions that arise:

- What is the best "information structure?" (scalar quantization)
- How about vector quantization?

Communication Tools

Tools and references from information theory:

- Han and Verdú (1993). *Approximation of output statistics.*
- Cuff (2008). *Communication requirements for generating correlated random variables.*
- Cuff, Permuter, and Cover (2009). *Coordination Capacity.*

Erasure Game (two states)

S is equally likely to be 0 or 1:

$S = 0$

B

A

	0	1
0	3	0
e	0	1
1	$-\infty$	$-\infty$

$S = 1$

B

A

	0	1
0	$-\infty$	$-\infty$
e	1	0
1	0	3

Neither know the state: Value $= 1/2$.
A knows the state: Value $= 3/2$.
B knows the state: Value $= 0$.
Both know the state: Value $= 3/4$.

Erasure Game (two states)

S is equally likely to be 0 or 1:

$S = 0$

A \ B	0	1
0	3	0
e	0	1
1	$-\infty$	$-\infty$

$S = 1$

A \ B	0	1
0	$-\infty$	$-\infty$
e	1	0
1	0	3

Neither know the state: Value $= 1/2$.
A knows the state: Value $= 3/2$.
B knows the state: Value $= 0$.
Both know the state: Value $= 3/4$.

Reminder: Concept from Rate-distortion Theory

To generate correlated actions $\sim p(a|s)$,

$$R \geq I(S;A) \text{ is required.}$$

To generate correlated actions $\sim p(a|s)$,

$R \geq I(S;A)$ is required.

This does not produce independent actions in the sequence.

To generate correlated actions $\sim p(a|s)$,

$R \geq I(S;A)$ is required.

This does not produce independent actions in the sequence.
Decode the action sequence after observing $k = n\frac{R}{H(A)}$ actions.

To generate correlated actions $\sim p(a|s)$,

$$R \geq I(S;A) \text{ is required.}$$

This does not produce independent actions in the sequence.
Decode the action sequence after observing $k = n\frac{R}{H(A)}$ actions.

What is the price of independence?

Erasure Challenge

Person A

0 1 0 0 1 1 1 1

Erasure Challenge

Person A	Person B
0 1 0 0 1 1 1 1	0 e e e e 1 e e

Erasure Challenge

Person A	Person B
0 1 0 0 1 1 1 1	0 e e e e 1 e e

How much must Person A tell Person B?

Erasure Challenge

Person A								Person B							
0	1	0	0	1	1	1	1	0	e	e	e	e	1	e	e

How much must Person A tell Person B?

- Tell all the bits
  8 bits

Erasure Challenge

Person A								Person B							
0	1	0	0	1	1	1	1	0	e	e	e	e	1	e	e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits

Erasure Challenge

Person A								Person B							
0	1	0	0	1	1	1	1	0	e	e	e	e	1	e	e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits $= \log_2 112 = 6.81$ bits

Erasure Challenge

Person A							
0	1	0	0	1	1	1	1

Person B

0 e e e e 1 e e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits $= \log_2 112 = 6.81$ bits
- Split the randomization

Erasure Challenge

Person A							
0	1	0	0	1	1	1	1

Person B

0 e e e e 1 e e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits $= \log_2 112 = 6.81$ bits
- Split the randomization

Erasure Challenge

Person A

0	1	0	0	1	1	1	1

Person B

0	1	e	e	1	1	e	e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits $= \log_2 112 = 6.81$ bits
- Split the randomization

Erasure Challenge

Person A

0 1	0 0	1 1	1 1

Person B

0 e	e e	e 1	e e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits $= \log_2 112 = 6.81$ bits
- Split the randomization

Erasure Challenge

Person A

0	1	0	0	1	1	1	1

Person B

0	e	e	e	e	1	e	e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits $= \log_2 112 = 6.81$ bits
- Split the randomization
  $\log_2 \binom{4}{2} + 4$ bits

Erasure Challenge

Person A

0 1	0 0	1 1	1 1

Person B

0 e	e e	e 1	e e

How much must Person A tell Person B?

- Tell all the bits
  8 bits
- Choose the sequence for B and tell it
  $\log_2 \binom{8}{2} + 2$ bits $= \log_2 112 = 6.81$ bits
- Split the randomization
  $\log_2 \binom{4}{2} + 4$ bits $= \log_2 96 = 6.58$ bits

Wyner's Common Information

[Wyner 75]:

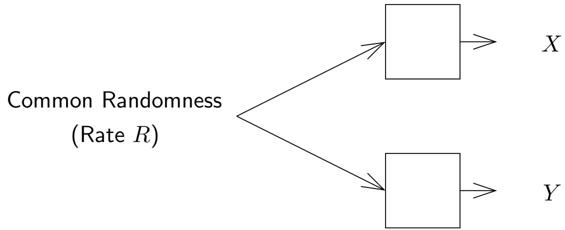
$$C(X;Y) \triangleq \min_{X-U-Y} I(X,Y;U).$$

Wyner's Common Information

[Wyner 75]:

$$C(X;Y) \triangleq \min_{X-U-Y} I(X,Y;U).$$

How much common randomness is needed to generate X and Y?

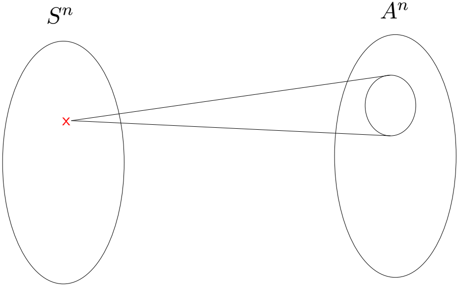


Result: $R > C(X;Y)$.

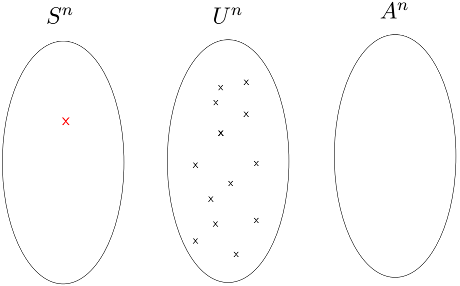
Generating Correlated Sequences

Generating Correlated Sequences

Generating Correlated Sequences

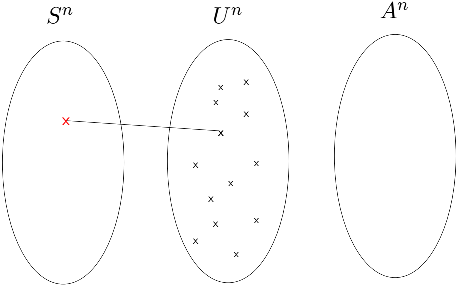


Generating Correlated Sequences



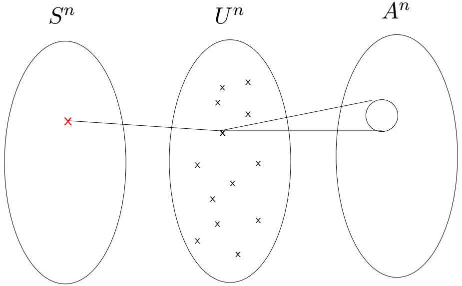
- Generate a codebook with extra u^n sequences $\sim \prod_{i=1}^{n} p(u_i)$.

Generating Correlated Sequences



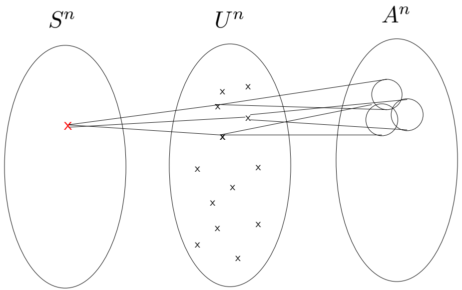
- Generate a codebook with extra u^n sequences $\sim \prod_{i=1}^{n} p(u_i)$.
- Encoder randomly chooses among u^n sequence that are correlated with x^n and sends the index i.

Generating Correlated Sequences



- Generate a codebook with extra u^n sequences $\sim \prod_{i=1}^{n} p(u_i)$.
- Encoder randomly chooses among u^n sequence that are correlated with x^n and sends the index i.
- Decoder generates y^n randomly conditioned on $u^n(i)$.

Generating Correlated Sequences



$$R \geq I(X;U) + I(U;Y|X).$$

Resolvability: [Wyner 75] [Han, Verdú 93]

Common Entropy Example

Jointly Gaussian with correlation $\rho \geq 0$:

$$
\begin{aligned}
I(X;Y) &= \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right). \\
C(X;Y) &= \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) \\
&= I(X;Y) + \log(1+\rho).
\end{aligned}
$$

Degenerate Game (counter-example)

$S = 0$

A \ B	0
0	0
1	-1

$S = 1$

A \ B	0
0	-1
1	0

The expected payoff is simply the negative Hamming distortion.

No need for randomizing.

Simple idea first:

Choose U such that $S - U - A$ form a Markov chain and $R > I(S;U)$.

$$\text{B doesn't know } S\text{: Payoff} \geq \frac{R}{I(S,A;U)} \, \underline{\Pi}_{p_{A|S}} + \frac{I(S,A;U) - R}{I(S,A;U)} \, \underline{\Pi}^{(U)}_{p_{A|U}}.$$

$$\text{B knows } S\text{: Payoff} \geq \frac{R - I(S;U)}{I(A;U|S)} \, \underline{\Pi}^{(S)}_{p_{A|S}} + \frac{I(S,A;U) - R}{I(A;U|S)} \, \underline{\Pi}^{(S,U)}_{p_{A|U}}$$

More complexity:

Choose U_1 and U_2 such that $S - (U_1, U_2) - A$ form a Markov chain and $R > I(S; U_1, U_2)$.

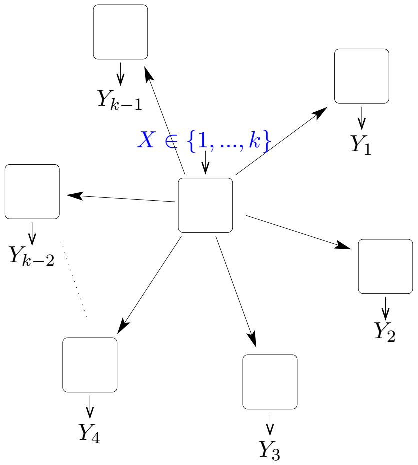
Generate a U_1 codebook and a U_2 codebook for each U_1 sequence.

The opponent learns U_1 early and U_2 late.

Communication in Networks

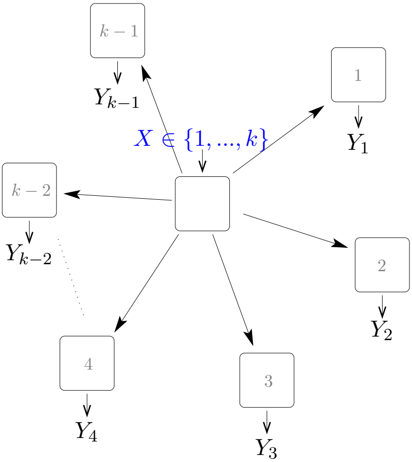
Networks

Star Network



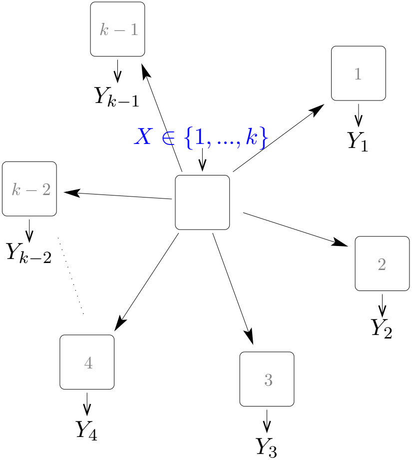
Try $R_i = \log \frac{k}{k-1}$ for all i. (Doesn't work)

Star Network



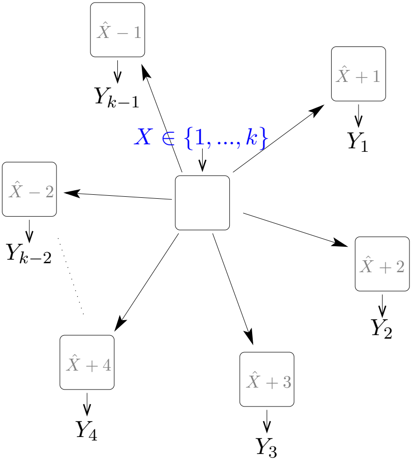
Assign Default Tasks: $R_i = h\left(\frac{1}{k}\right) \approx \frac{\log k}{k}$. Sum rate: $R \approx \log k + \log e$.

Star Network



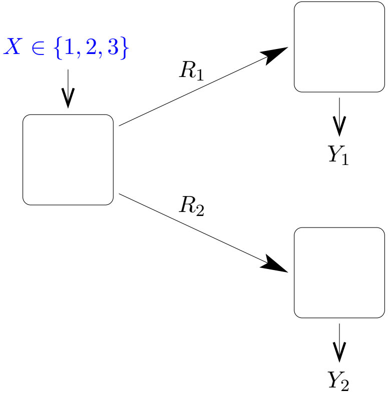
Lower bound: $R \geq I(X; Y_1, ..., Y_{k-1}) = H(X) = \log k.$

Star Network



Two phase: Specify low-rate estimate $\hat{X}$. Choose defaults to exclude $\hat{X}$.

Three Node Star (Golden Ratio)



Rates for optimized estimate quality $\hat{X}$.

Two stage Communication rate: $R_i = \log 3 - \log \phi$.

The golden ratio $\phi = \frac{\sqrt{5}+1}{2}$.

The Cost of Randomized Actions

In Star Network with one message (three node task assignment).

Not randomized: Sum rate $= \log \frac{k}{k-2}$.
Randomized: Sum rate $\approx 3 \log 3$.

Bottomline

Summary:

- Rate-distortion type coding is not suited for games.
- Generating i.i.d. sequences plays a partial role.
- Causality of decisions creates a time-varying result even with i.i.d. codebooks.
- Mutual information and common information find roles in control settings.