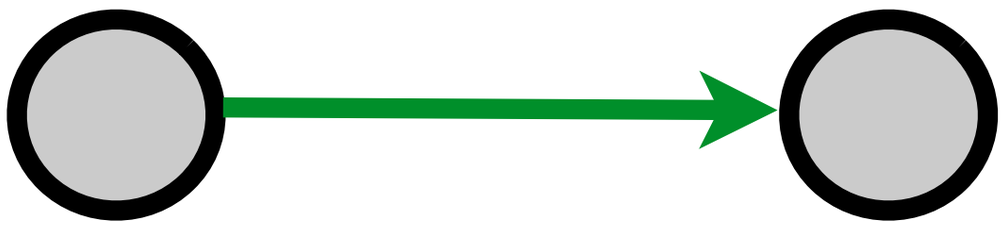
Communicating dialogues

Natasha Devroye
Assistant Professor
University of Illinois at Chicago
http://www.ece.uic.edu/~devroye

33rd IEEE SARNOFF SYMPOSIUM 2010
Princeton, New Jersey
April 12-14, 2010

Monologues vs. Dialogues

Monologues vs. Dialogues

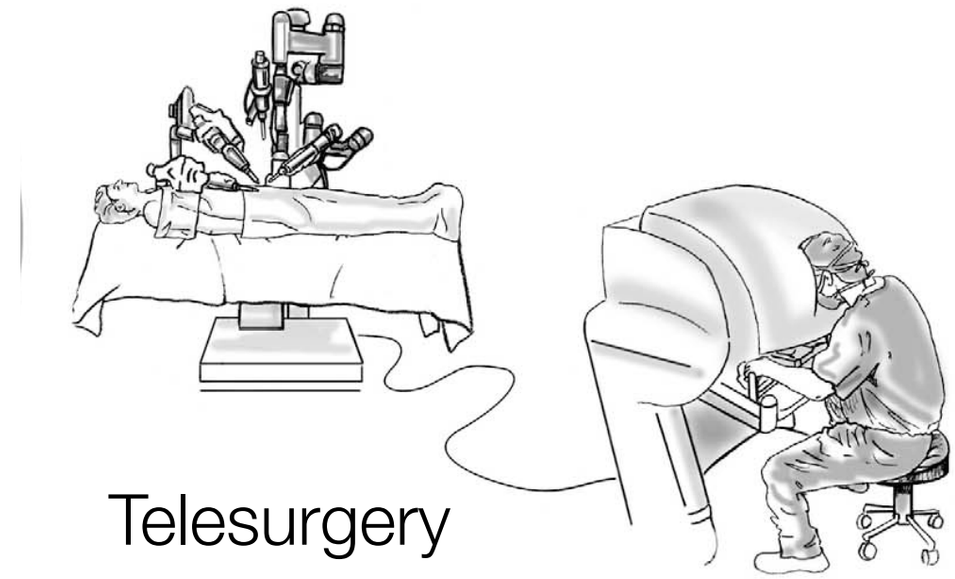


Monologue = one-way = uni-directional

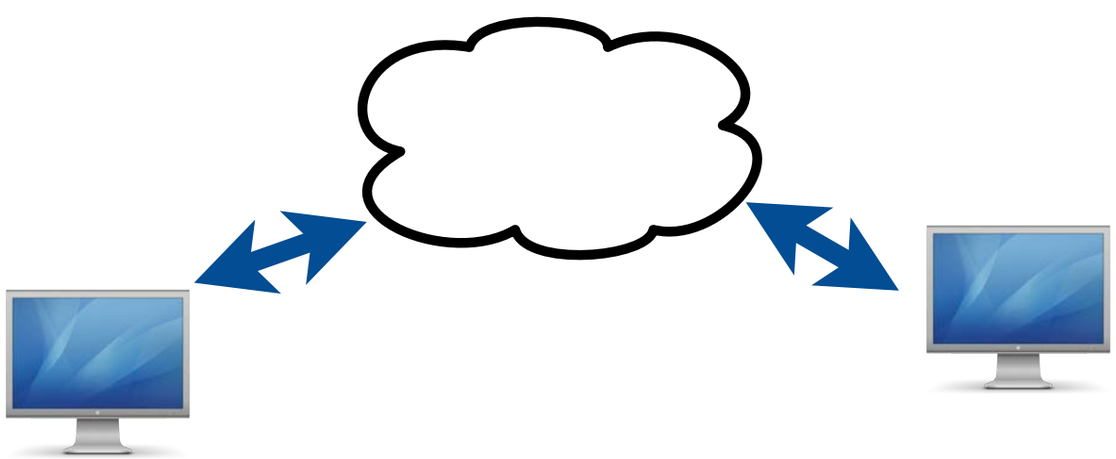
Dialogue = two-way = bi-directional

Two-way communication applications - wired

Video conferencing

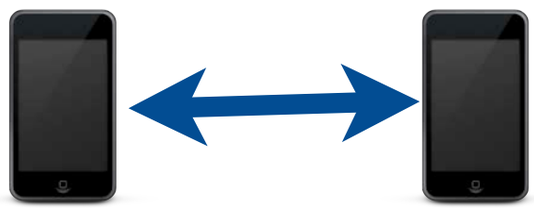


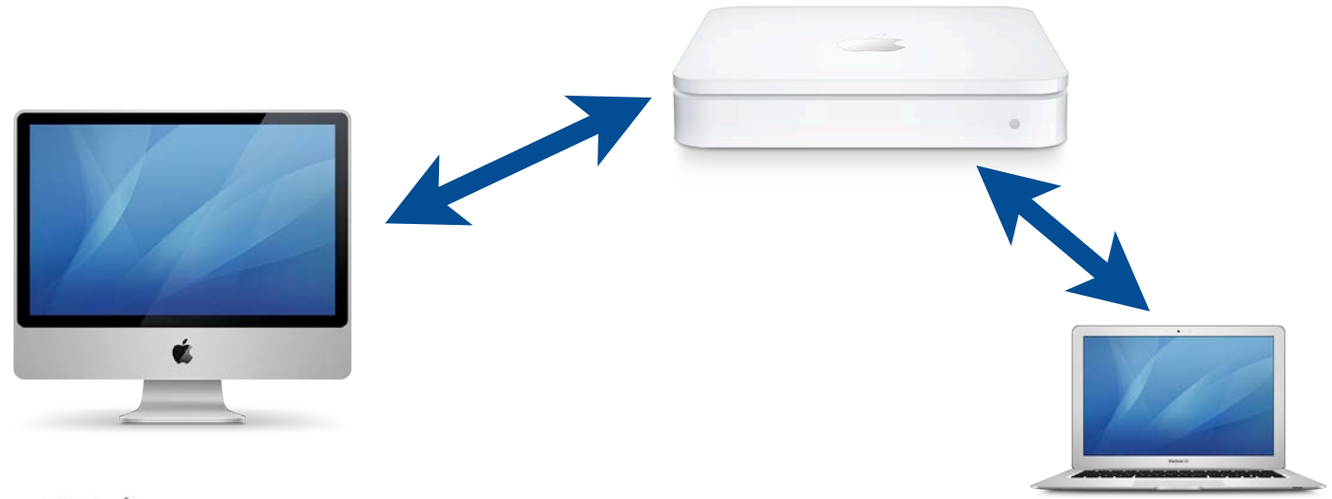
Telesurgery

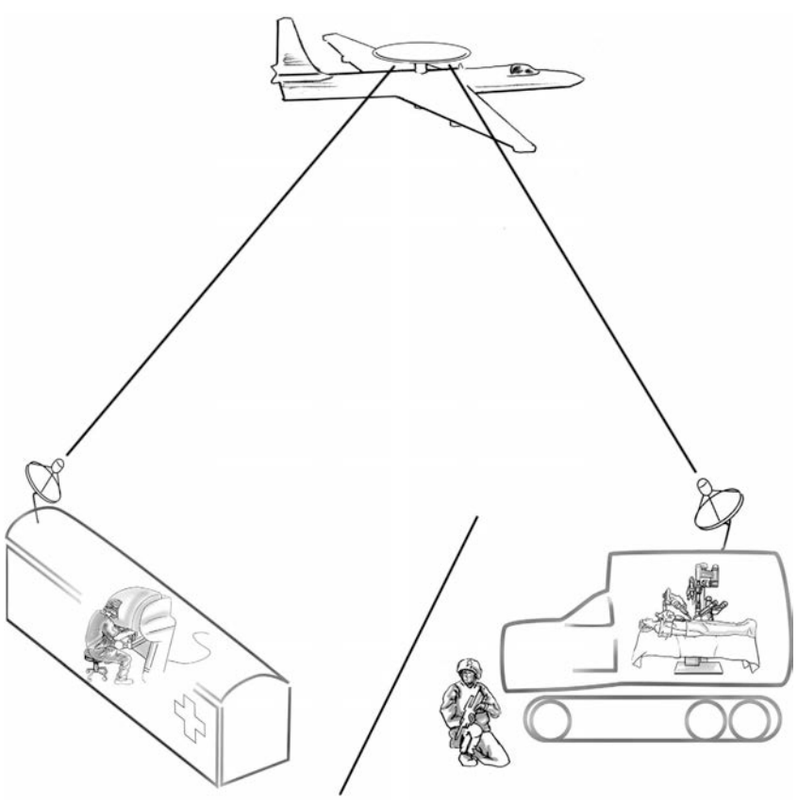


Data synchronization

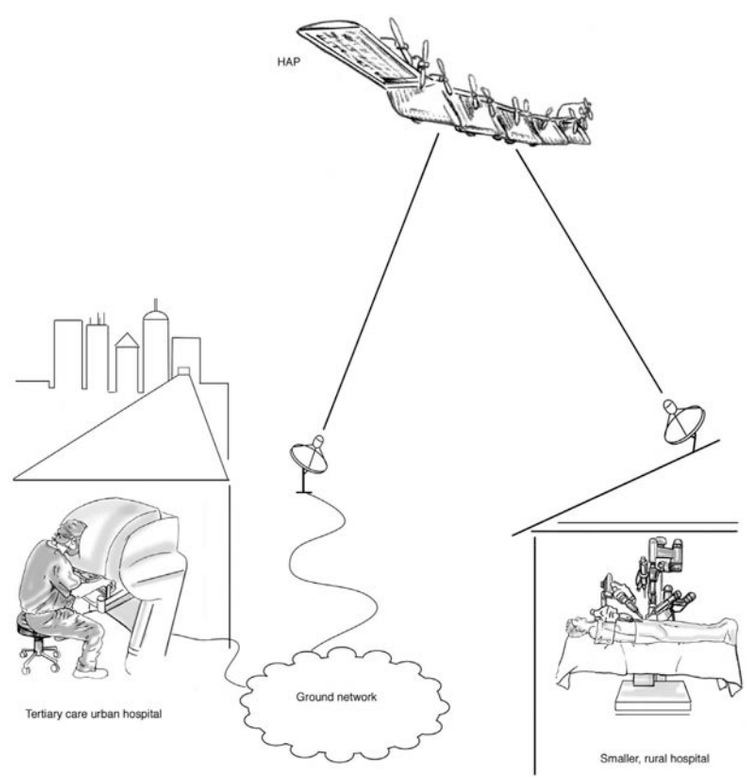
Two-way communication applications - wireless







Battlefield telesurgery



Rural telesurgery

Video conferencing

1 Dialogue **≠ or =** 2 Monologues ?

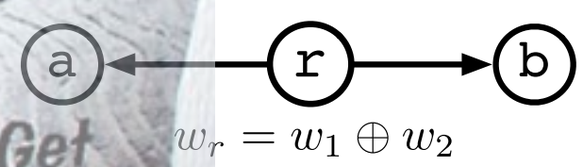
It depends....

we will use information theory to find out.

Outline

- Information theory - *what, why, when*
- Two-way channel - *channel coding*
- Wireless channels and networks
- Two-way cellular-like networks
- Two-way relay channels - *canonical example of wireless network coding*

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

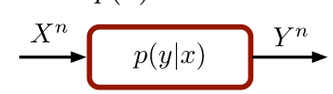


Overall - much is still unknown

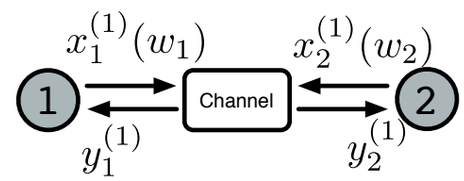
Outline

- **Information theory - *what, why, when***

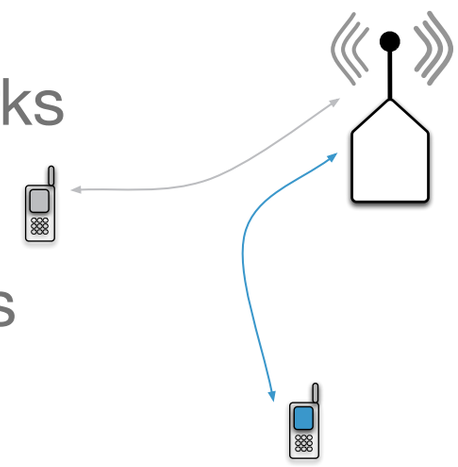
$$C = \max_{p(x)} I(X;Y)$$

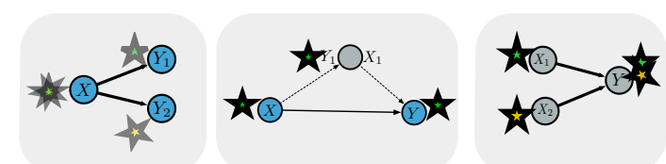


- Two-way channel - *channel coding*



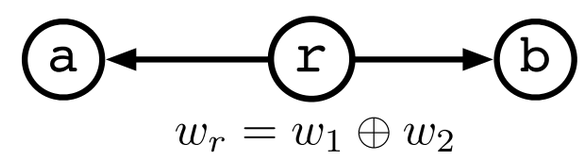
- Wireless channels and networks





- Two-way cellular-like networks

- Two-way relay channels - *canonical example of wireless network coding*



Information theory - *what, why, when*

A Mathematical Theory of Communication. **Bell System Technical Journal*, 27, 379–423 & 623–656, 1948.***

What is information?

How fast can we communicate?

How much can we compress information?

What is communication?

RANDOMNESS

33rd IEEE SARNOFF SYMPOSIUM 2010
Princeton, New Jersey
April 12-14, 2010

Information theory's claims to fame

Source coding

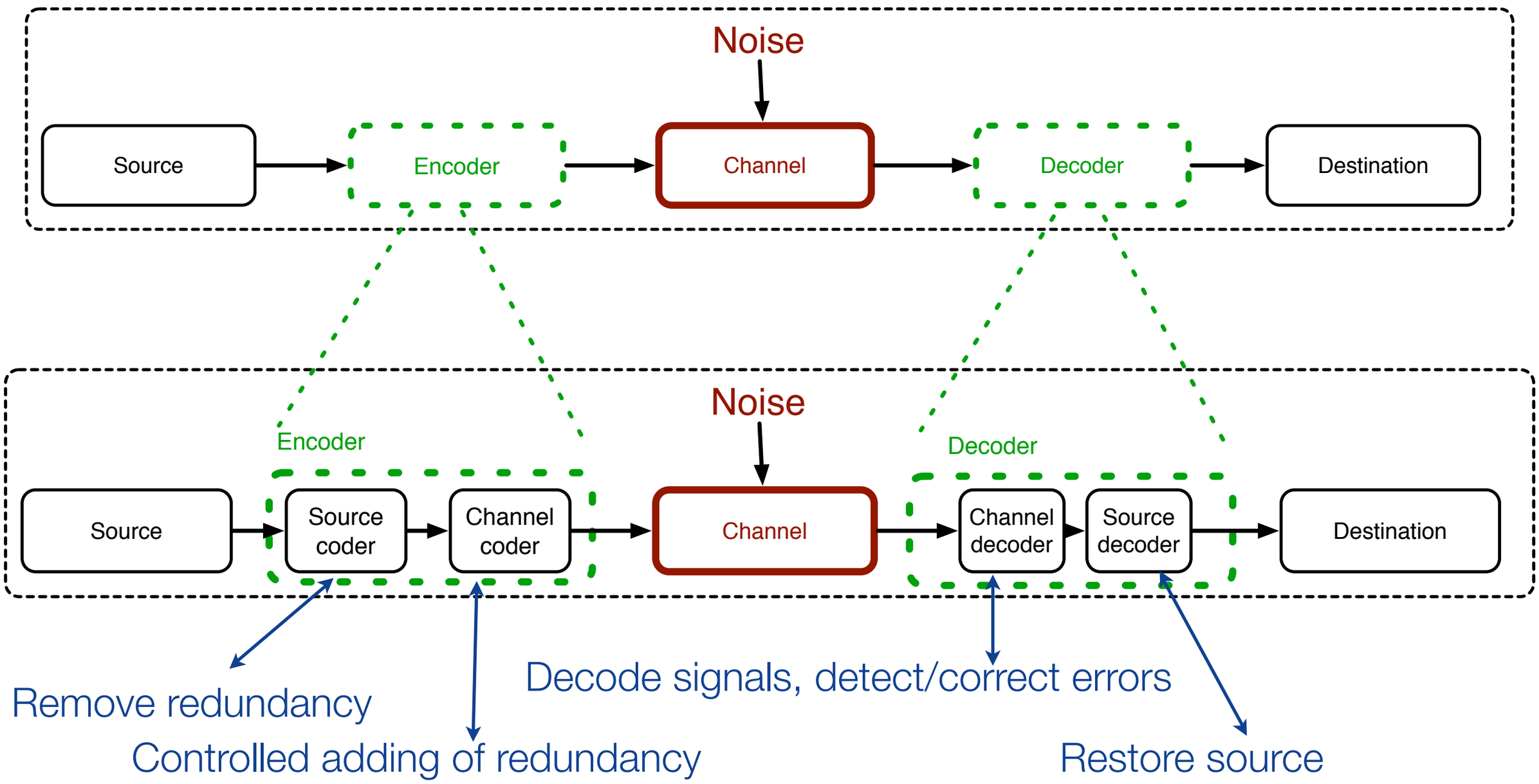
- Source = random variable
- Ultimate data compression limit is the source's entropy H

Channel coding

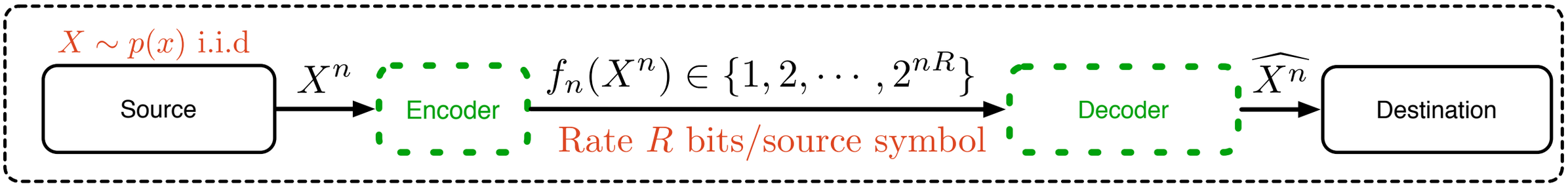
- Channel = conditional distributions
- Ultimate transmission rate is the channel capacity C

Reliable communication possible ↔ H<C

Source vs. channel coding



Source coding = data compression

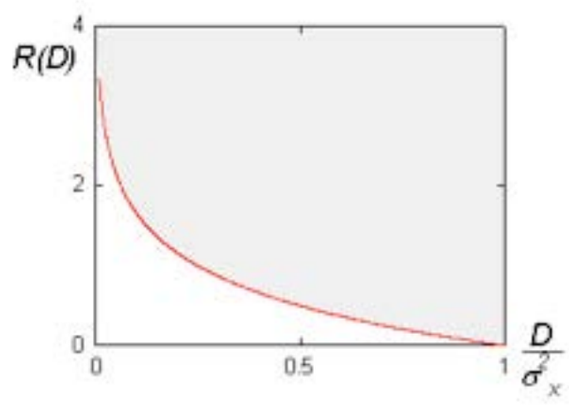


- Lossless: $\widehat{X^n} = X^n$

Minimum R needed is $H(X) = -\sum_x p(x) \log_2(p(x))$

- Lossy: $d(\widehat{X^n}, X^n) \leq D$

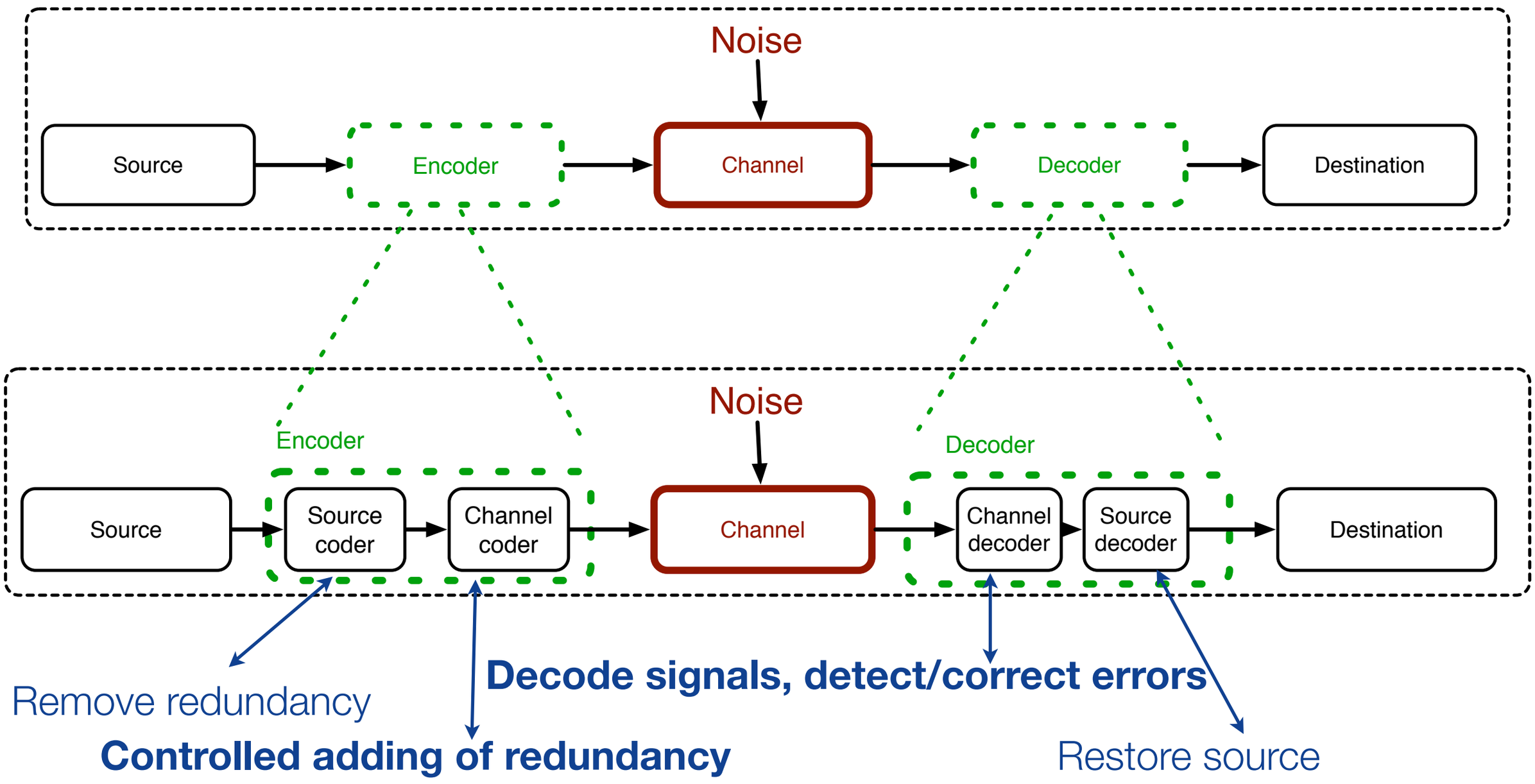
Minimum R needed is described by the Rate-Distortion function $R(D)$



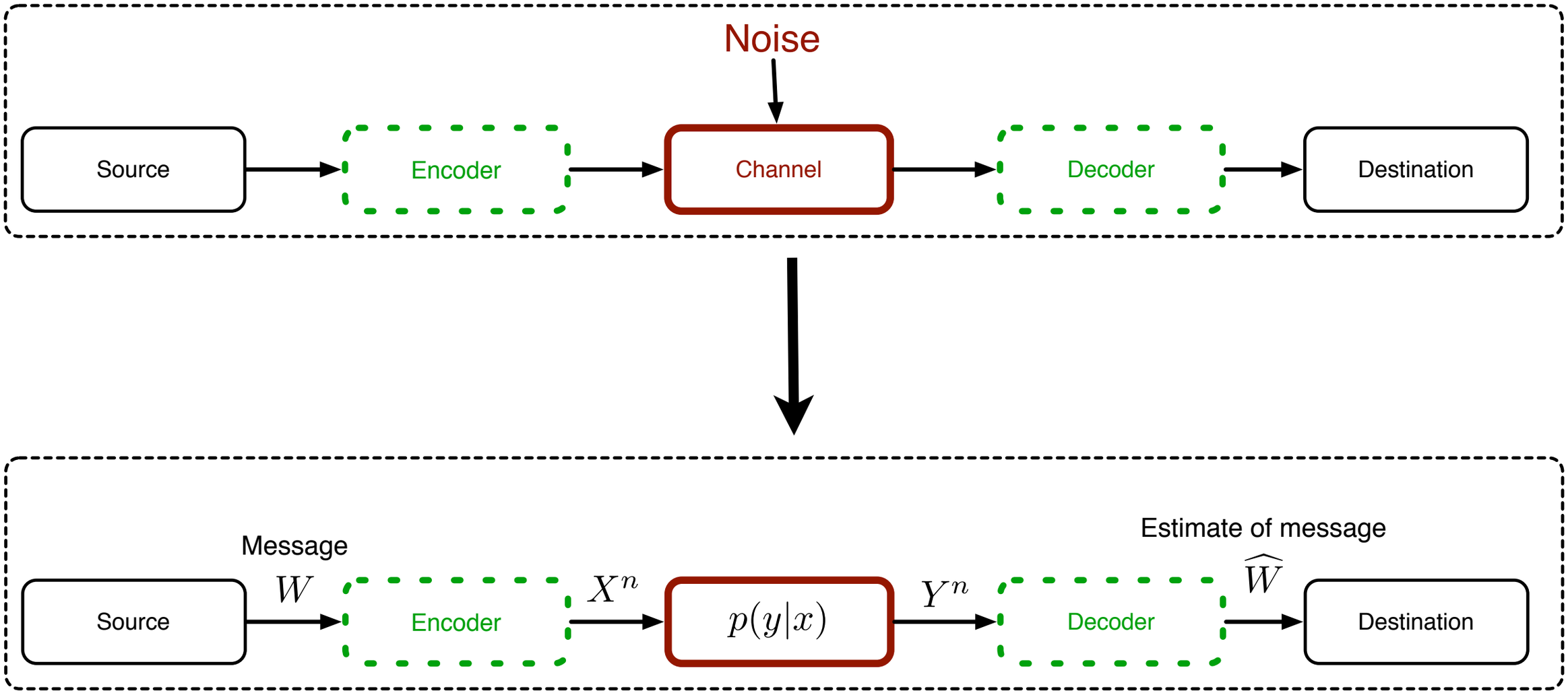
$$R(D) = \min_{p(\hat{x}|x)} I(X; \hat{X})$$

s.t. $d(\widehat{X^n}, X^n) \leq D$

Source vs. channel coding

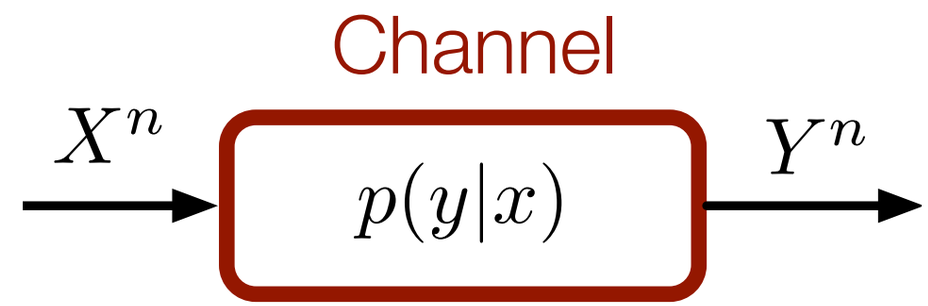


Communication system model



What is the capacity of this channel?

Channel capacity



- Information channel capacity:

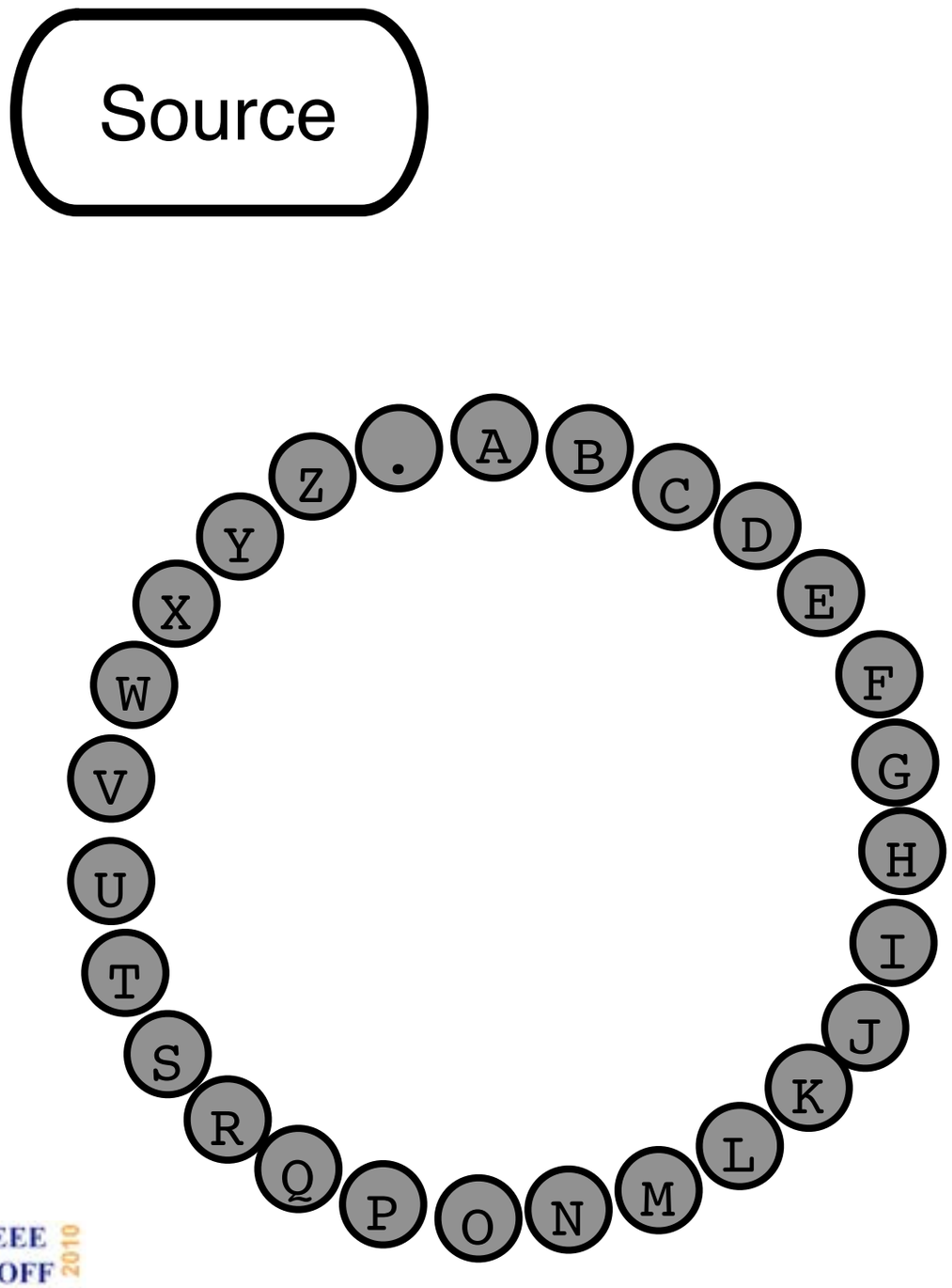
$$C = \max_{p(x)} I(X;Y)$$

- Operational channel capacity:

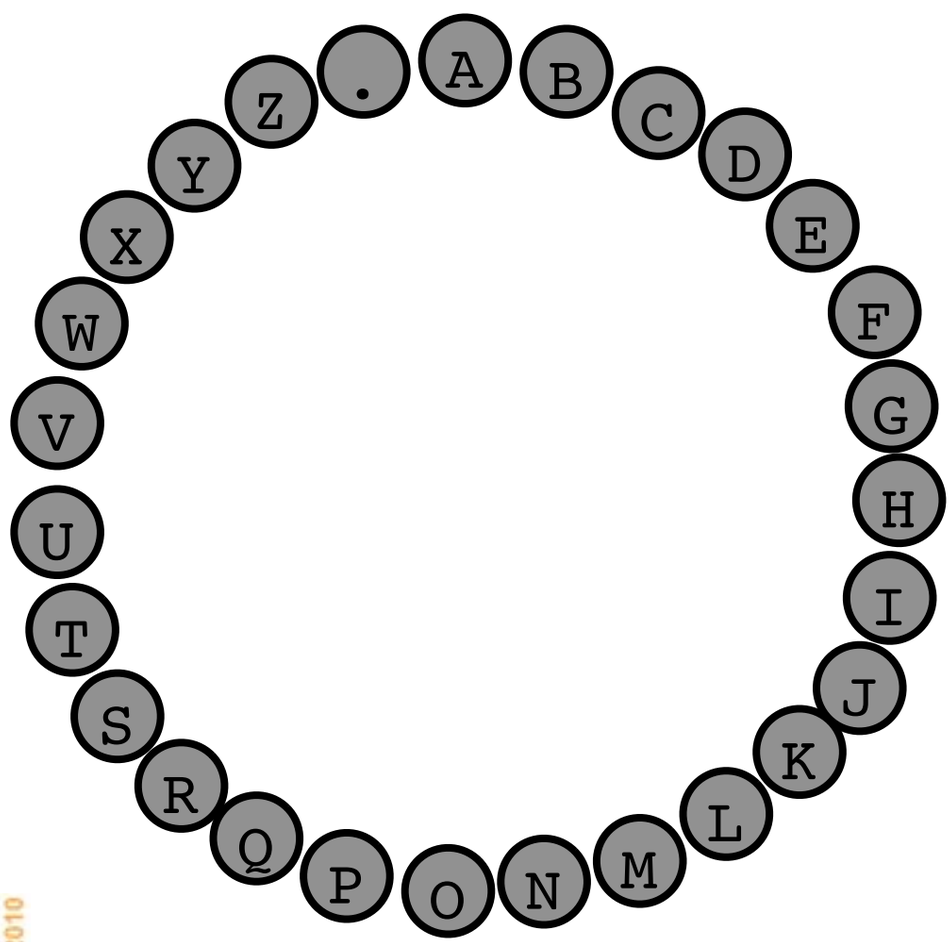
```
Highest rate (bits/channel use) that can
communicate at reliably
```

- Channel coding theorem says: information capacity = operational capacity

Channel capacity: a cute example



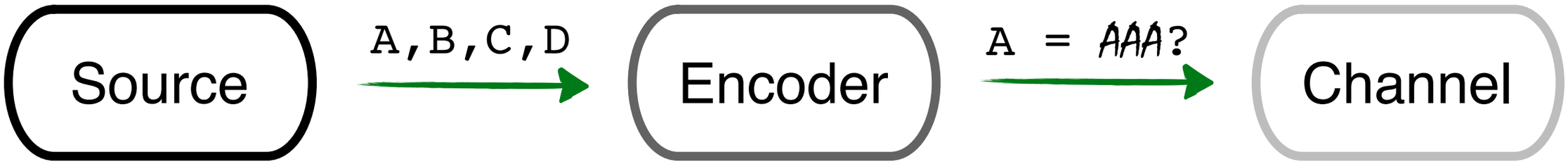
Channel capacity: a cute example



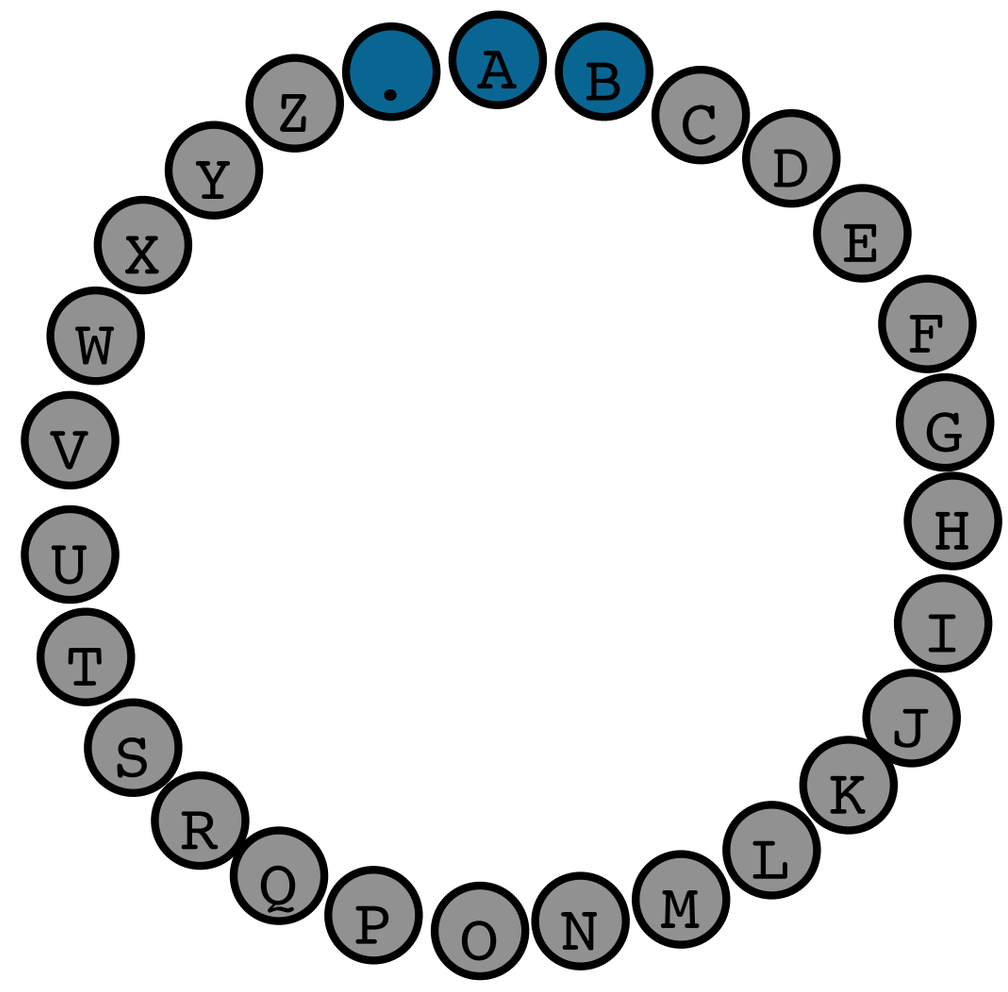
A = A?

A = AAA?

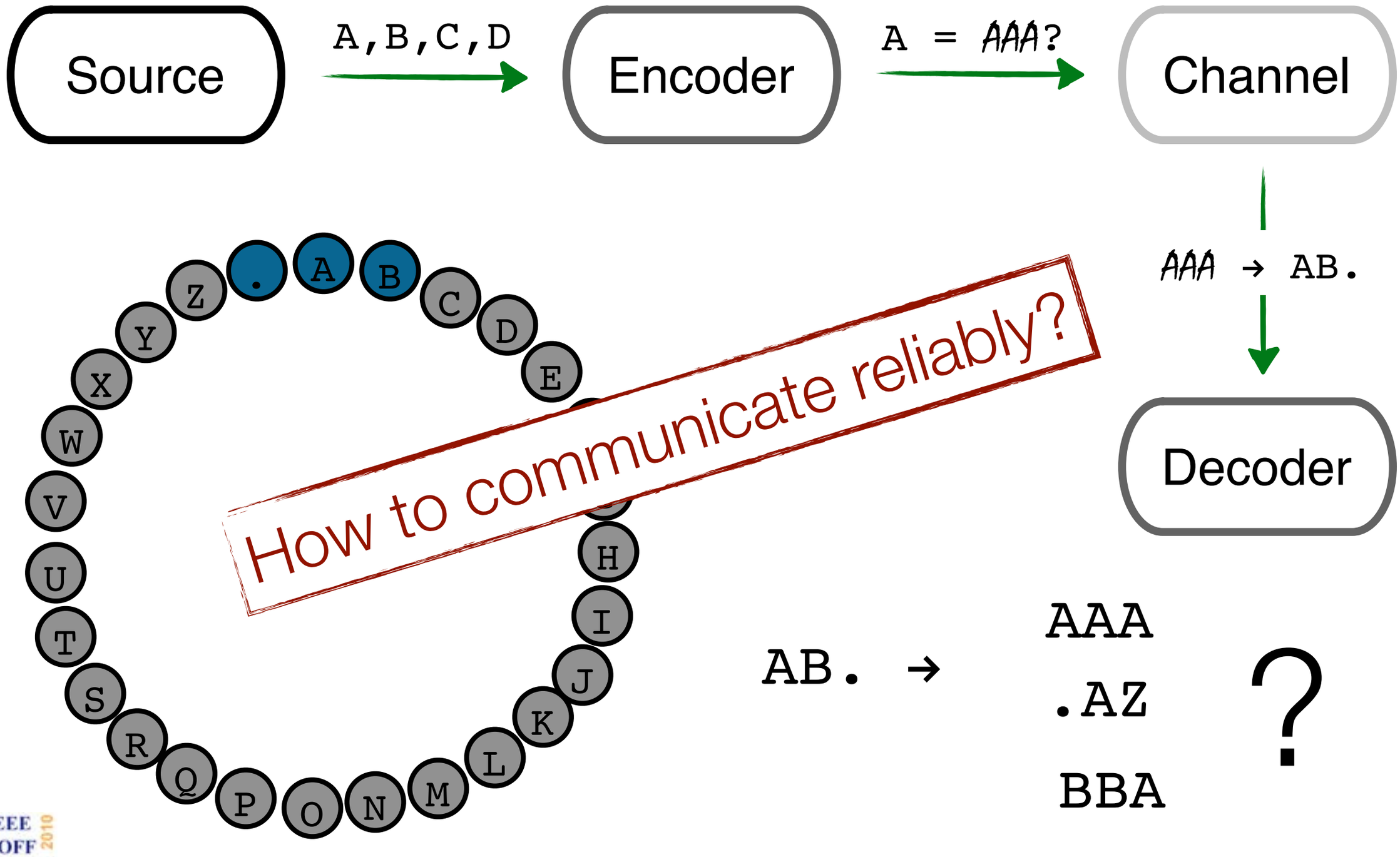
Channel capacity: a cute example



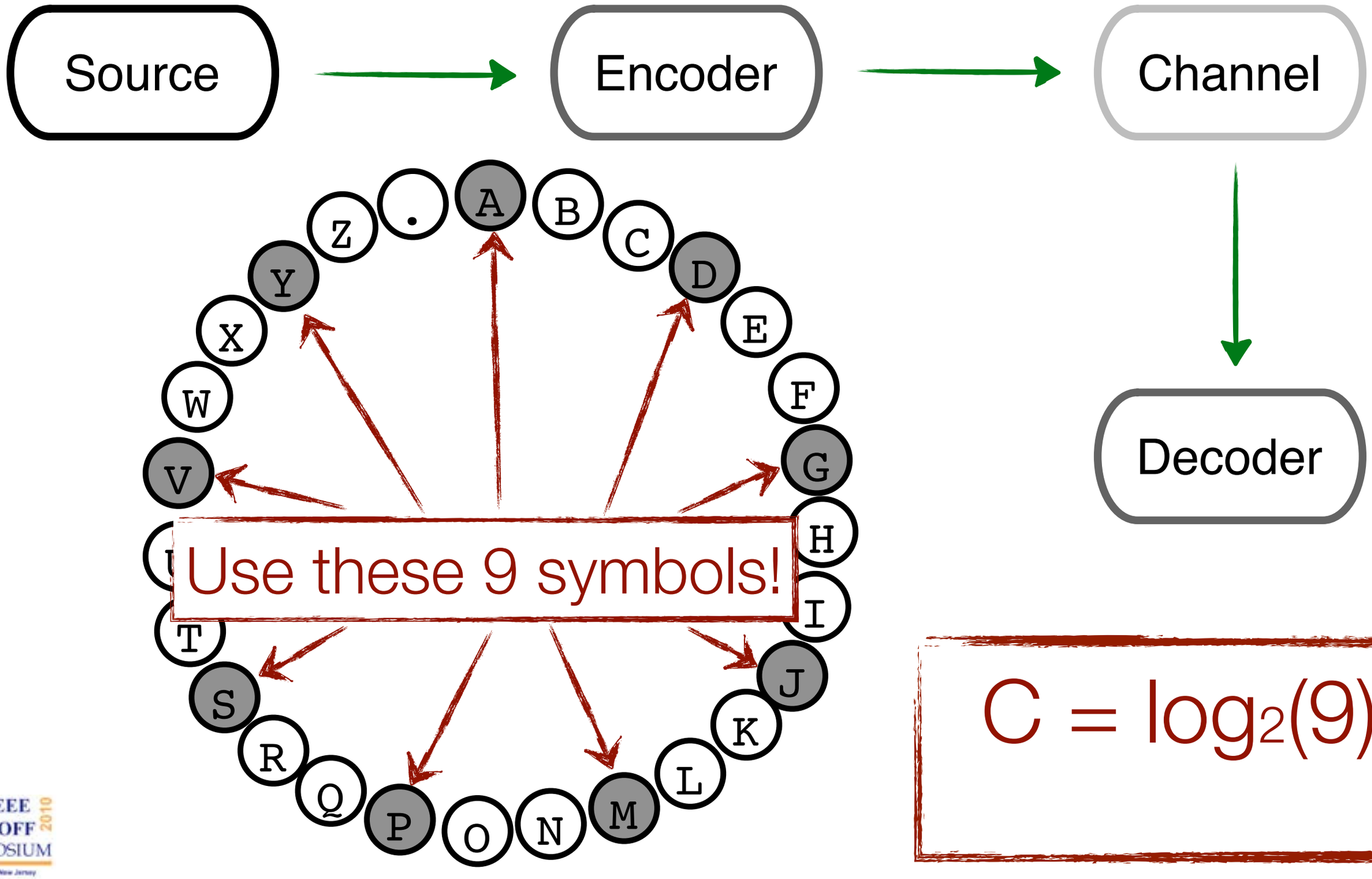
AAA → AB.



Channel capacity: a cute example



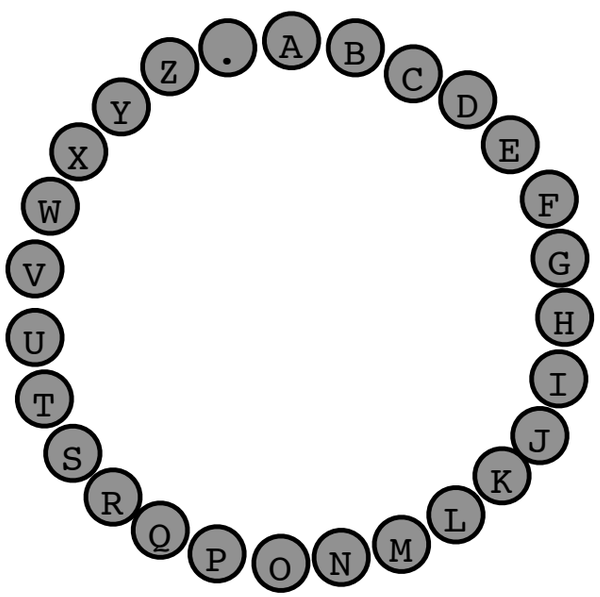
Channel capacity: a cute example



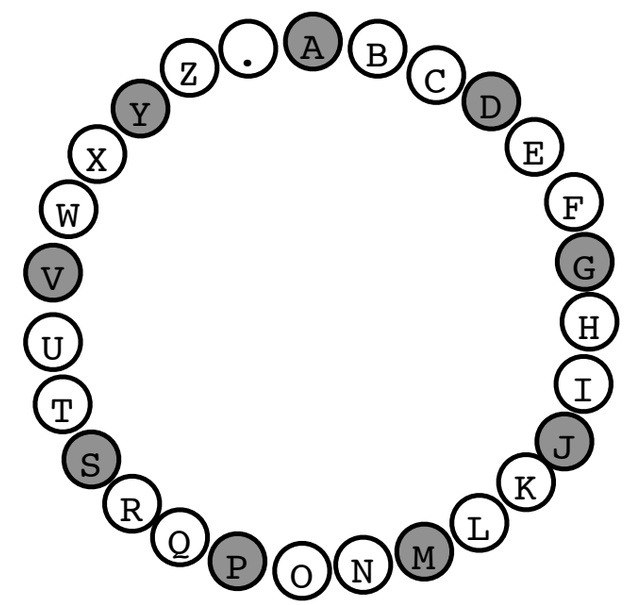
$$C = \log_2(9)$$

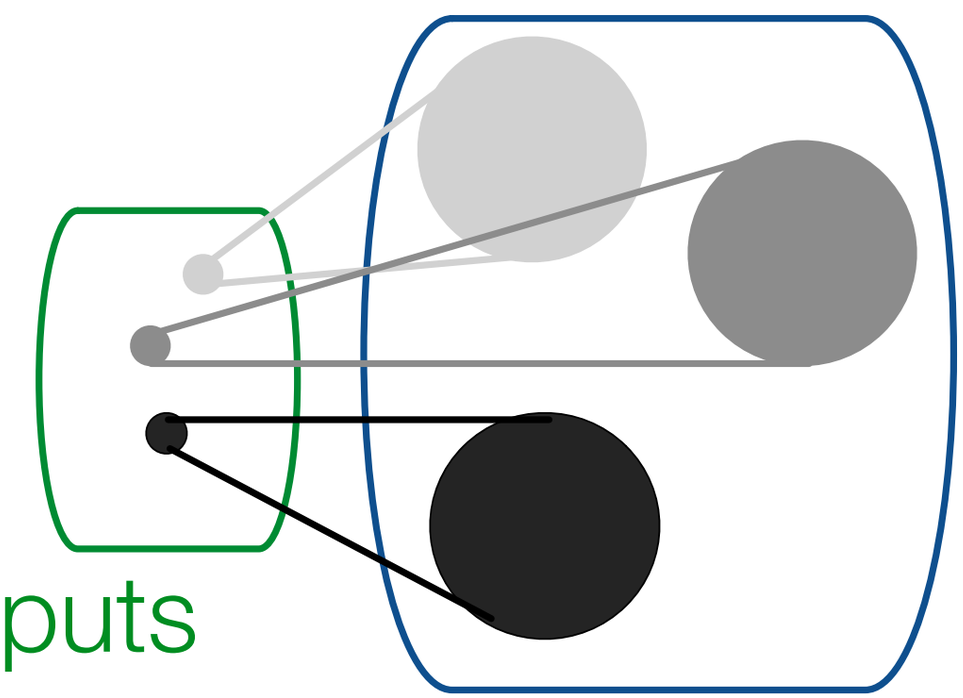
Capacity in general

- Main idea was to reduce the rate (from a 27-letter input per channel use to a 9-letter input per channel use) so as to produce



Non-overlapping outputs!





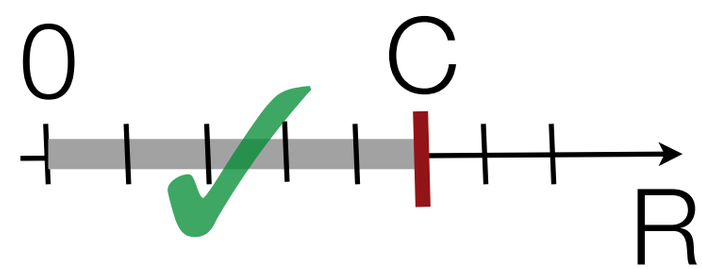
Inputs

Outputs

Mathematical description of capacity

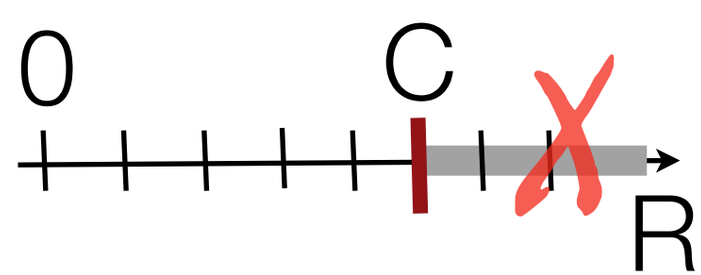
- Can achieve reliable communication for all transmission rates R:

$$R < C$$

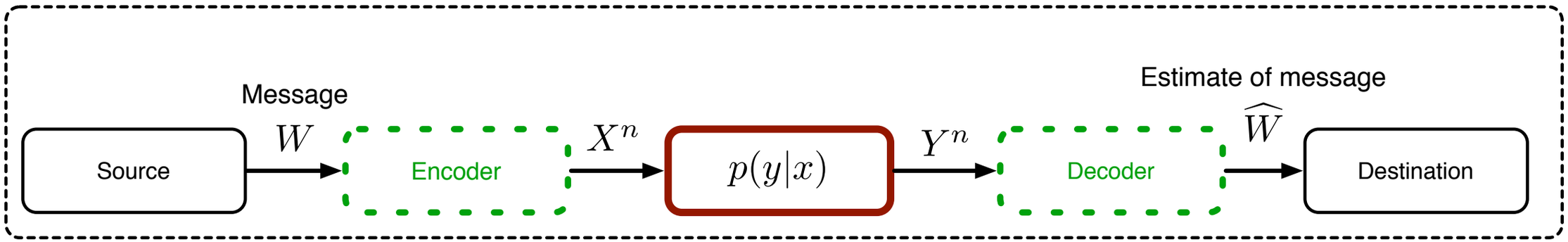


- BUT, probability of decoding error always bounded away from zero if

$$R > C$$

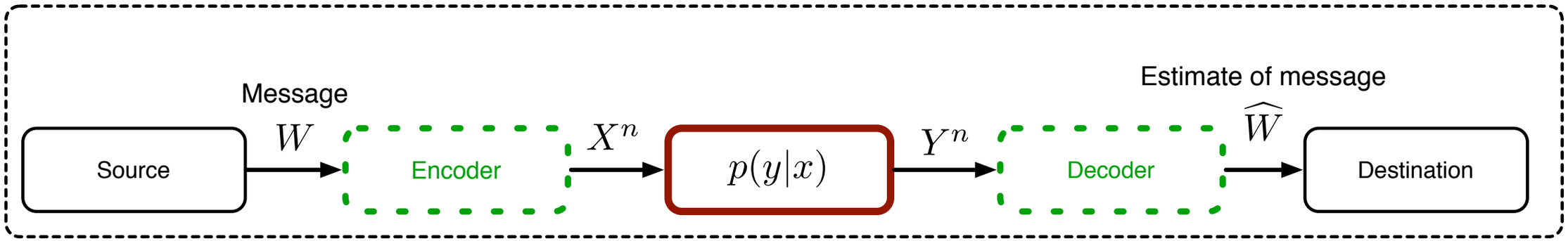


Capacity: key ideas



- choose input set of codewords so they are “non-confusable” at the output
- number of these that we can chose will determine the channel’s capacity
- number that we can choose will depend on the distribution p(y|x) which characterizes the channel

One-way channel capacity



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

bits/channel use

"mutual information" between X and Y

$$I(X;Y) = \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

A few examples

What is capacity / mutual information?

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

Entropy of a random variable

(A) entropy is the measure of **average uncertainty** in the random variable

(B) entropy is the **average number of bits** needed to describe the random variable

(C) entropy is measured in bits?

(D) $H(X) = -\sum_x p(x) \log_2(p(x))$

(E) entropy of a deterministic value is 0

Entropy of a random variable

(A) entropy is the measure of **average uncertainty** in the random variable

(B) entropy is the **average number of bits** needed to describe the random variable

(C) entropy is measured in bits?

(D) $H(X) = -\sum_x p(x) \log_2(p(x))$

(E) entropy of a deterministic value is 0

Entropy of a random variable

(A) entropy is the measure of **average uncertainty** in the random variable

(B) entropy is the **average number of bits** needed to describe the random variable

(C) entropy is measured in bits?

(D) $H(X) = -\sum_x p(x) \log_2(p(x))$

(E) entropy of a deterministic value is 0

Entropy of a random variable

(A) entropy is the measure of **average uncertainty** in the random variable

(B) entropy is the **average number of bits** needed to describe the random variable

(C) entropy is measured in bits?

(D) $H(X) = -\sum_x p(x) \log_2(p(x))$

(E) entropy of a deterministic value is 0

Entropy of a random variable

(A) entropy is the measure of **average uncertainty** in the random variable

(B) entropy is the **average number of bits** needed to describe the random variable

(C) entropy is measured in bits?

(D) $H(X) = -\sum_x p(x) \log_2(p(x))$

(E) entropy of a deterministic value is 0

Entropy of a uniform distribution

- Let X be uniformly distributed over 8 outcomes. What is the entropy of X?

$$H(X) = \sum_{x=1}^{8} p(x) \log_2(p(x)) = -\sum_{x=1}^{8} \frac{1}{8} \log_2\left(\frac{1}{8}\right) = \log_2(8) = 3 \text{ (bits)}$$

- This is the number of bits needed to describe X!

- By extension, for a discrete random variable taking on K outcomes, the **maximal entropy** is attained by a uniform distribution and is equal to the number of bits needed to describe K:

$$H(X) = \log_2(K)$$

Entropy of a non-uniform distribution

- Suppose X represents the outcome of a horse race with 8 horses, which win with probabilities $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right)$

$$
\begin{aligned}
H(X) &= -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{1}{16}\log_2\left(\frac{1}{16}\right) - 4\frac{1}{64}\log_2\left(\frac{1}{64}\right) \\
&= 2 \text{ (bits)}
\end{aligned}
$$

- 8 outcomes, 3 bits? But on average can represent with 2 bits!

$$\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right)$$

$$(000,001,010,011,100,101,110,111)$$

3 bits

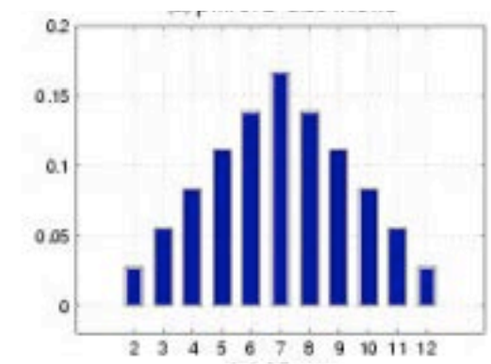
$$(0,10,110,1110,111100,111101,111110,111111)$$

2 bits (on average!)

Entropy of a continuous random variable

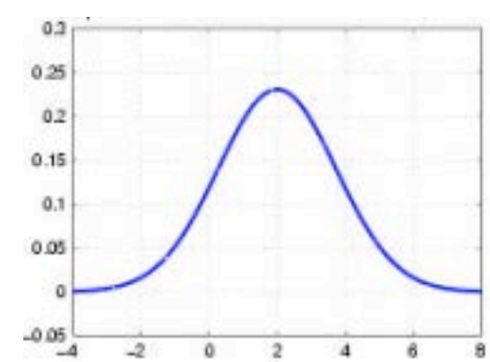
- entropy:

$$H(X) = -\sum_x p(x) \log_2(p(x))$$



- differential entropy:

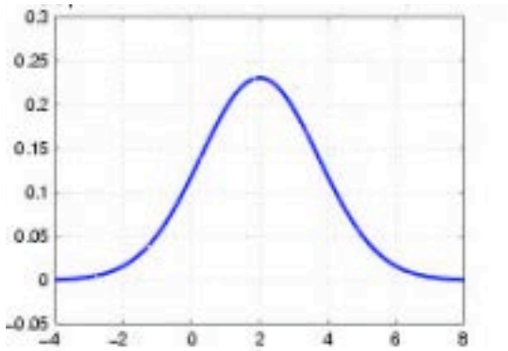
$$h(X) = -\int f(x) \log(x)\, dx$$



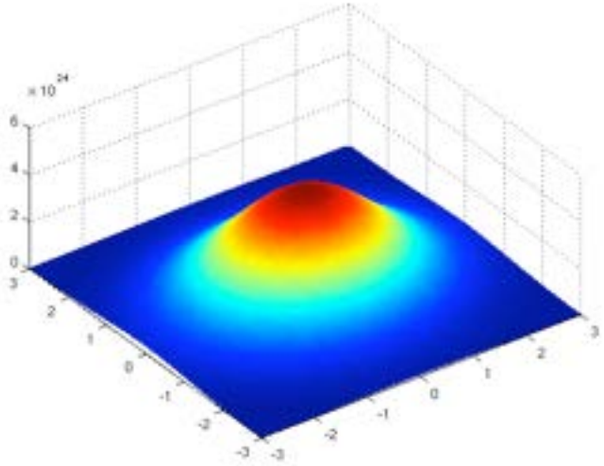
Entropy of a Gaussian random variable

- differential entropies of Gaussian distributions:

$$h\left(\mathcal{N}(0,\sigma^2)\right) = \frac{1}{2}\log\left(2\pi e\sigma^2\right)$$



$$h\left(\mathcal{N}_n(\mu,K)\right) = \frac{1}{2}\log\left((2\pi e)^n|K|\right)$$



Entropy maximization

- **Uniform** distribution maximizes entropy for a given # outcomes

$$\max_{X:|X|=K} H(X) = \log_2(K)$$

- **Gaussian** maximizes entropy for a given covariance constraint

$$\max_{E[XX^T]=K} h(X) = \frac{1}{2}\log\left((2\pi e)^n |K|\right)$$

Mutual information between 2 random variables:

$$
\begin{aligned}
I(X;Y) &= \sum p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X)
\end{aligned}
$$

Mutual information between 2 random variables:

$$
\begin{aligned}
I(X;Y) &= \sum p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X)
\end{aligned}
$$

(A) I(X;Y) is the reduction in the uncertainty about X due to knowledge of Y

(B) if X, Y are independent I(X;Y) = 0

(C) if X=Y then I(X;Y) = H(X)

(D) I(X;Y) is non-negative

Mutual information between 2 random variables:

$$
\begin{aligned}
I(X;Y) &= \sum p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X)
\end{aligned}
$$

(A) I(X;Y) is the reduction in the uncertainty about X due to knowledge of Y

(B) if X, Y are independent I(X;Y) = 0

(C) if X=Y then I(X;Y) = H(X)

(D) I(X;Y) is non-negative

Mutual information between 2 random variables:

$$
\begin{aligned}
I(X;Y) &= \sum p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X)
\end{aligned}
$$

(A) I(X;Y) is the reduction in the uncertainty about X due to knowledge of Y

(B) if X, Y are independent I(X;Y) = 0

(C) if X=Y then I(X;Y) = H(X)

(D) I(X;Y) is non-negative

Mutual information between 2 random variables:

$$
\begin{aligned}
I(X;Y) &= \sum p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X)
\end{aligned}
$$

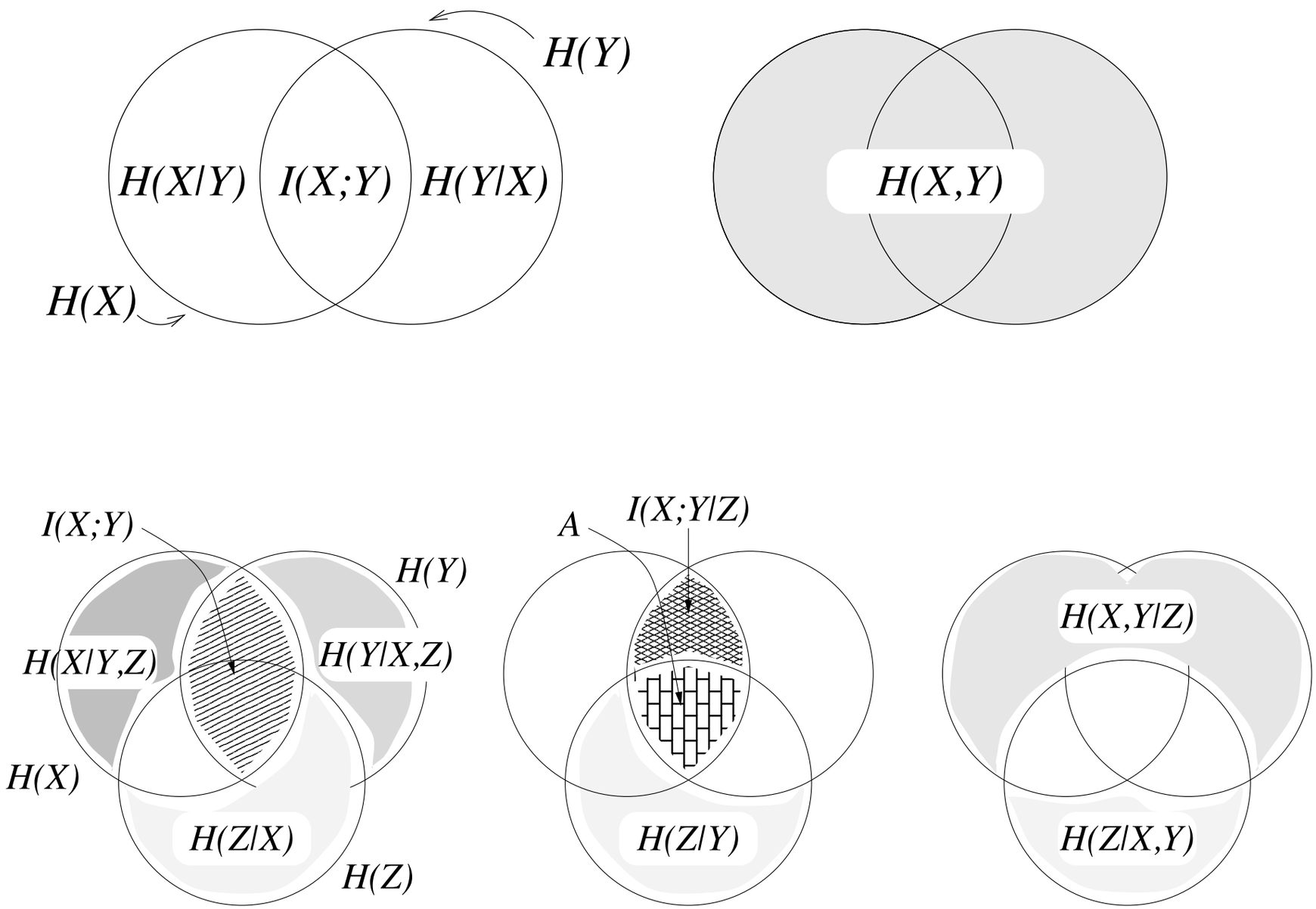
(A) I(X;Y) is the reduction in the uncertainty about X due to knowledge of Y

(B) if X, Y are independent I(X;Y) = 0

(C) if X=Y then I(X;Y) = H(X)

(D) I(X;Y) is non-negative

Properties of mutual information



With permission from David J.C. Mackay

Channel capacity

Capacity $C = \max_{p(x)} I(X;Y)$ bits/channel use

$$I(X;Y) = \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Channel capacity

Noiseless channel

Capacity?

1 bit/channel use

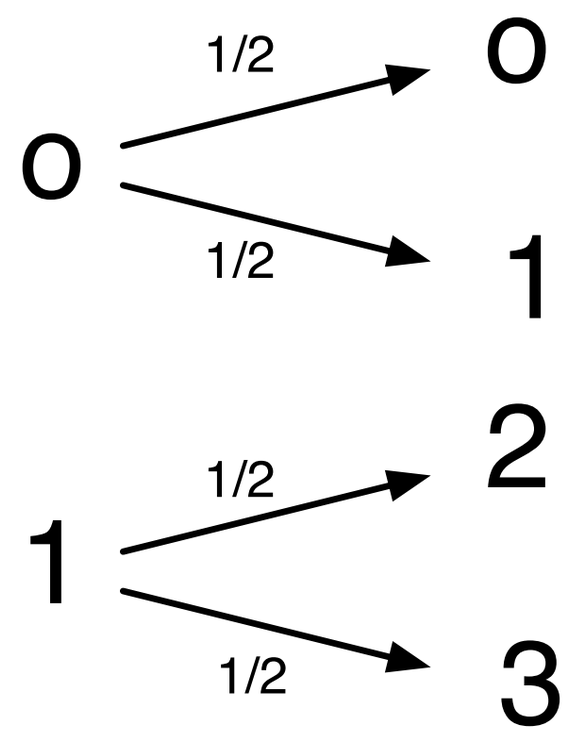
$$0 \longrightarrow 0$$

$$1 \longrightarrow 1$$

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(X) - H(X|Y) \\ &= \max_{p(x)} H(X) - 0 \\ &= 1 \end{aligned}$$

Channel capacity

Non-overlapping outputs



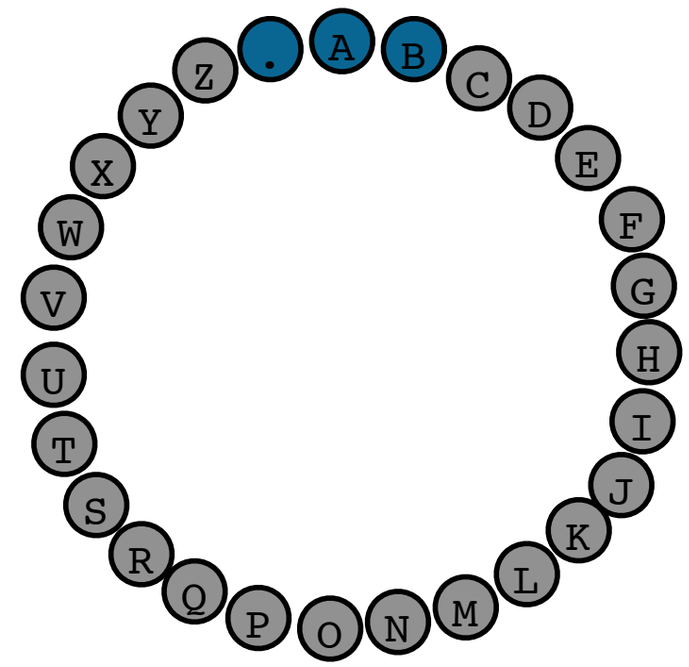
Capacity?

1 bit/channel use

$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} H(X) - H(X|Y) \\
&= \max_{p(x)} H(X) - 0 \\
&= 1
\end{aligned}
$$

Channel capacity

Noisy typewriter



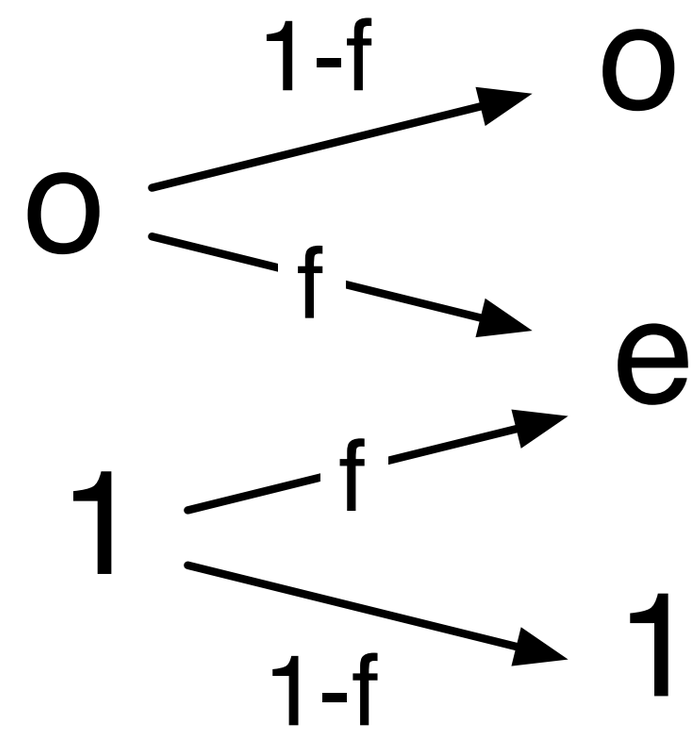
Capacity?

$\log_2(9)$ bits/channel use

$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} H(X) - H(X|Y) \\
&= \max_{p(x)} H(X) - \log_2(3) \\
&= \log_2(27) - \log_2(3) = \log_2(9)
\end{aligned}
$$

Channel capacity

Binary erasure channel



Capacity?

1-f bits/channel use

$$
\begin{aligned}
C &= \max_{p(x)} \; I(X;Y) \\
&= \max_{p(x)} \; H(X) - H(X|Y) \\
&= \max_{p(x)} \; H(X) - \sum p(y) H(X|Y=y) \\
&= \max_{p(x)} \; H(X) - p \cdot (1-f) \cdot H(X|Y=0) + (1-p) \cdot (1-f) \cdot H(X|Y=1) + p \cdot f \cdot H(X|Y=e) \\
&= \max_{p(x)} \; H(X) - p \cdot (1-f) \cdot 0 + (1-p) \cdot (1-f) \cdot 0 + f \cdot 1 \\
&= \max_{p(x)} \; H(X) - f \\
&= 1 - f
\end{aligned}
$$

Discrete memoryless channel capacity

Capacity $C = \max_{p(x)} I(X;Y)$ bits/channel use

"mutual information" between X and Y

$$I(X;Y) = \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Continuous alphabet channel capacity

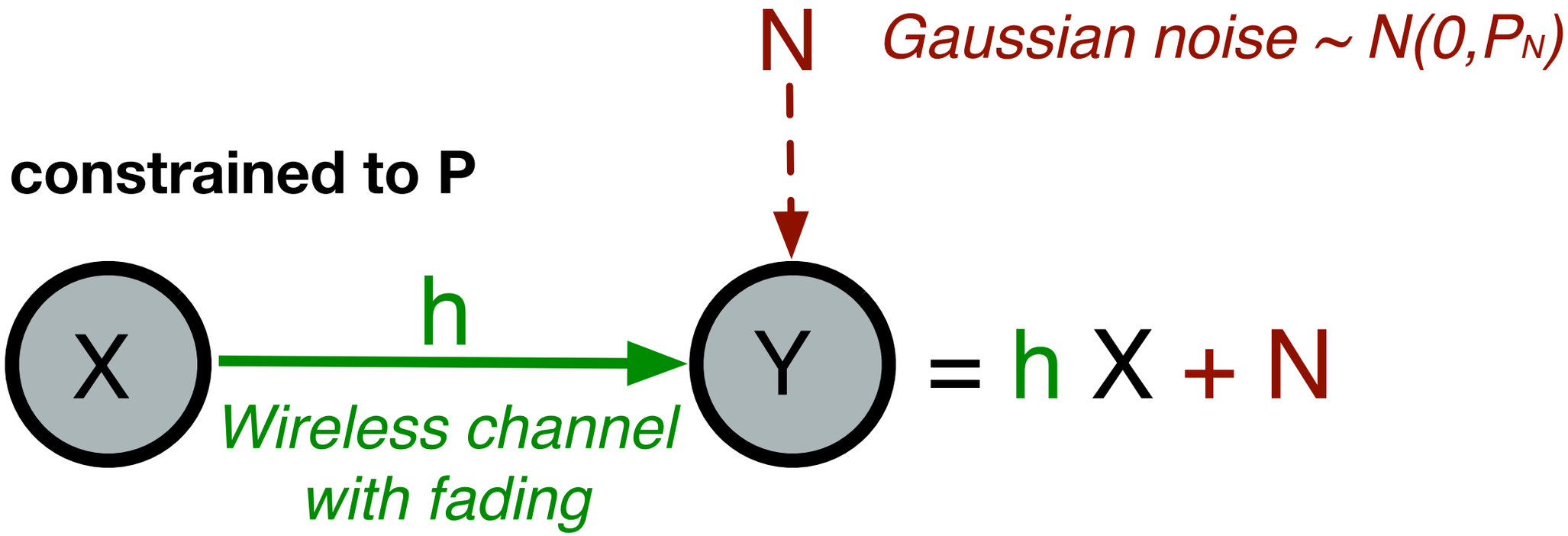
Capacity $C = \max_{p(x)} I(X;Y)$ bits/channel use

"mutual information" between X and Y

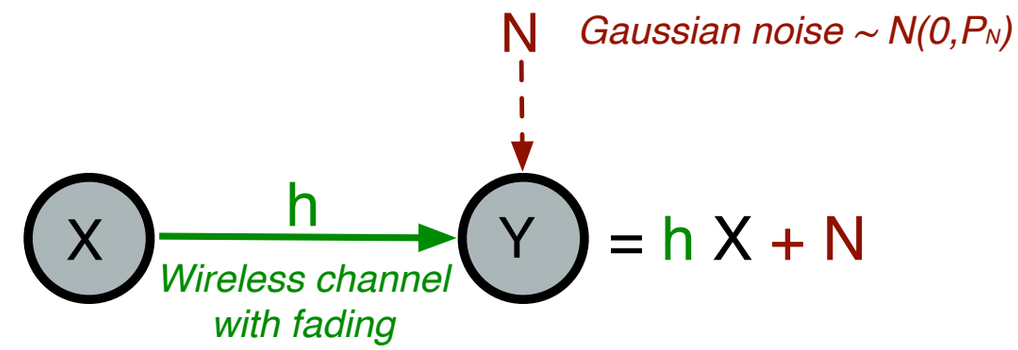
What if

X and Y are not bits, but real numbers?

AWGN channel capacity

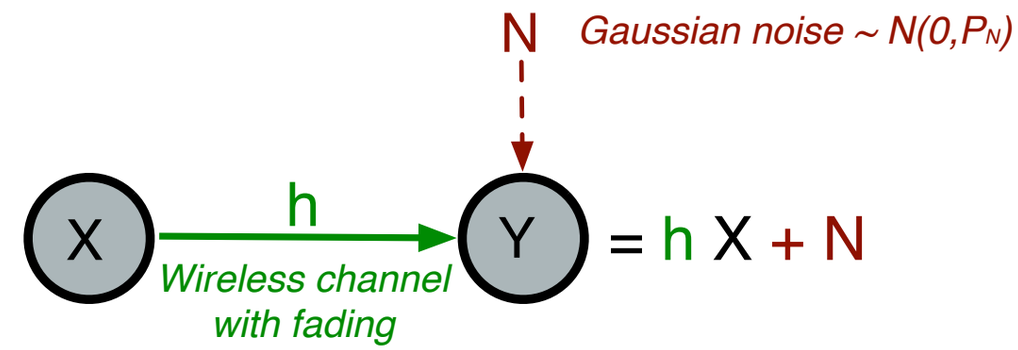


AWGN channel capacity



$$
\begin{aligned}
C &= \max_{p(x):E[XX^T]\leq P} I(X;Y) \\
&= \max_{p(x):E[XX^T]\leq P} h(X) - h(X|Y) \\
&= \max_{p(x):E[XX^T]\leq P} h(Y) - h(Y|X) \\
&= \max_{p(x):E[XX^T]\leq P} h(hX+N) - h(hX+N|X) \\
&= \max_{p(x):E[XX^T]\leq P} h(hX+N) - h(N) \\
&= \frac{1}{2}\log(2\pi e(|h|^2 P + P_N)) - \frac{1}{2}\log(2\pi e P_N) \\
&= \frac{1}{2}\log\left(\frac{|h|^2 P + P_N}{P_N}\right)
\end{aligned}
$$

AWGN channel capacity

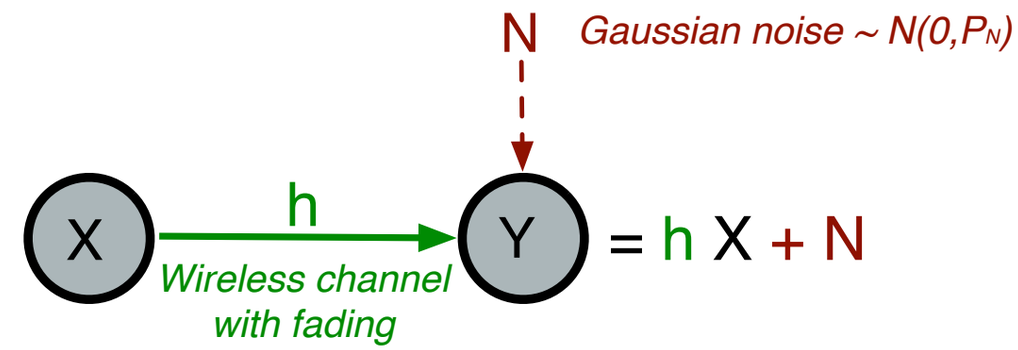


$$
\begin{aligned}
C &= \max_{p(x):E[XX^T]\leq P} I(X;Y) \\
&= \max_{p(x):E[XX^T]\leq P} h(X) - h(X|Y) \\
&= \max_{p(x):E[XX^T]\leq P} h(Y) - h(Y|X) \\
&= \max_{p(x):E[XX^T]\leq P} h(hX+N) - h(hX+N|X) \\
&= \max_{p(x):E[XX^T]\leq P} h(hX+N) - h(N) \\
&= \frac{1}{2}\log(2\pi e(|h|^2P + P_N)) - \frac{1}{2}\log(2\pi e P_N) \\
&= \frac{1}{2}\log\left(\frac{|h|^2P + P_N}{P_N}\right)
\end{aligned}
$$

signal power at Rx

noise power at Rx

AWGN channel capacity



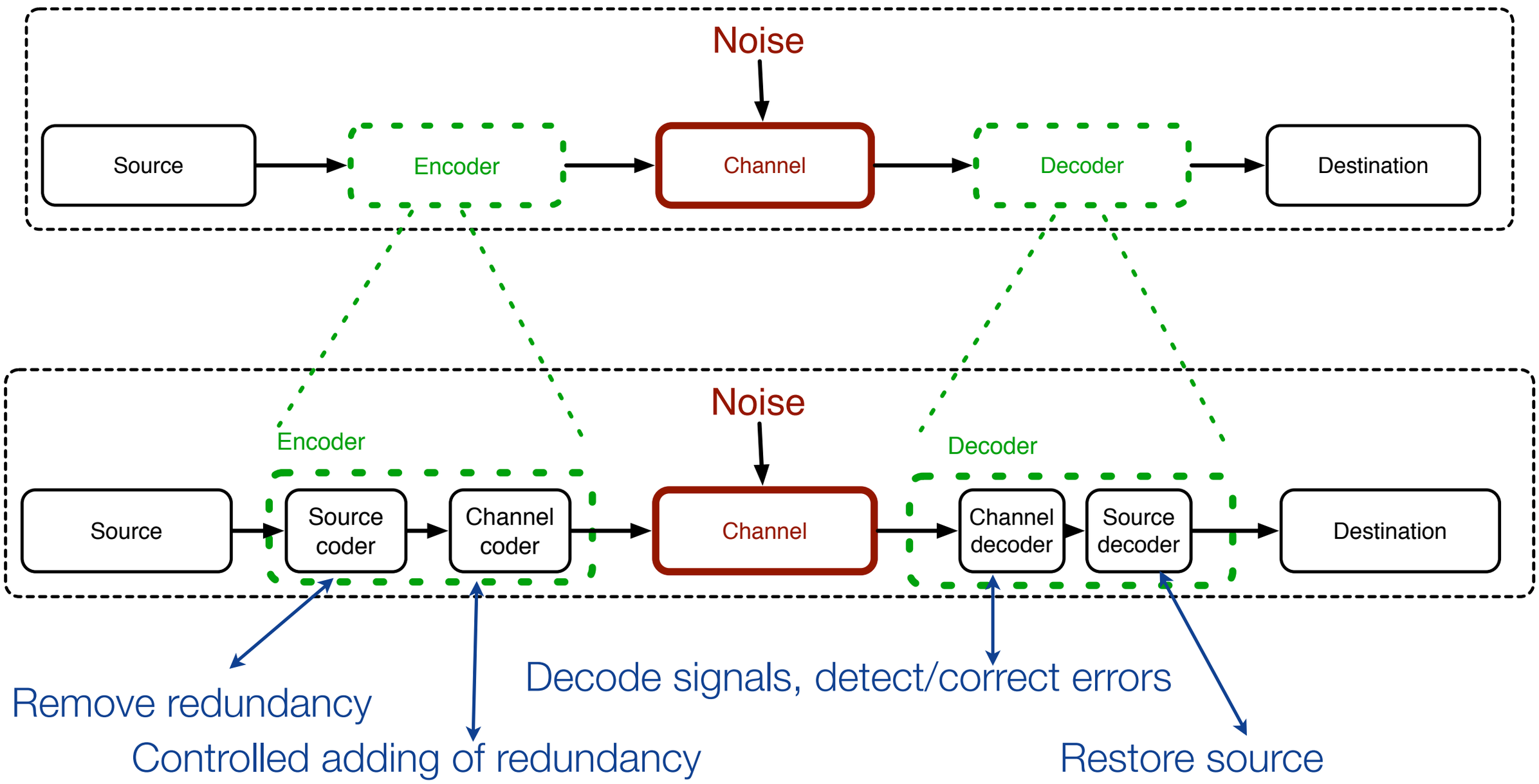
$$C = \frac{1}{2}\log\left(\frac{|h|^2 P + P_N}{P_N}\right)$$
$$= \frac{1}{2}\log\left(1 + SNR\right) \ \ \text{(bits/channel use)}$$

What about bits/second and bandwidth of the channel?

$$C = W\log_2\left(1 + \frac{P}{WN_0}\right) \ \ \text{(bits/second)}$$

[Bandwidth W, h=1, spectral density $N_0/2$]

Source vs. channel coding



Use?

[Forney, G.D. and Costello, D.J., ``Channel Coding: The Road to Channel Capacity,'' Proceedings of the IEEE, Volume 95, Issue 6, pp.1150-1177, June 2007.]

Claude Shannon — *Born on the planet Earth (Sol III) in the year 1916 A.D. Generally regarded as the father of the Information Age, he formulated the notion of channel capacity in 1948 A.D. Within several decades, mathematicians and engineers had devised practical ways to communicate reliably at data rates within 1% of the Shannon limit . . .*

Encyclopedia Galactica, 166th ed.

- algebraic codes
- convolutional codes
- iterative codes (LDPC, turbo)

-Robert J. McEliece, Allerton 2000

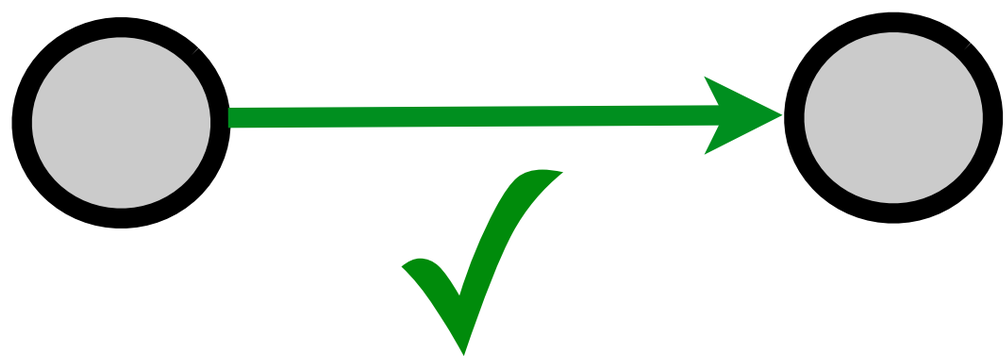
Use?

- Benchmark for performance of practical systems

- Guideline in designing systems - what’s worth shooting for?

- Theoretical insights can lead to practical insights

So now what?



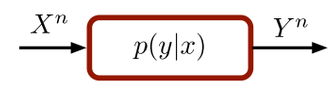
Unsolved

Fundamental

Outline

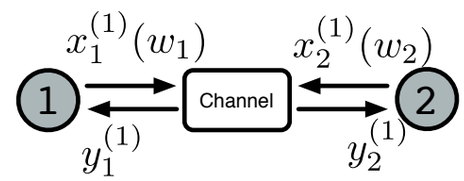
- Information theory - *what, why, when*

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

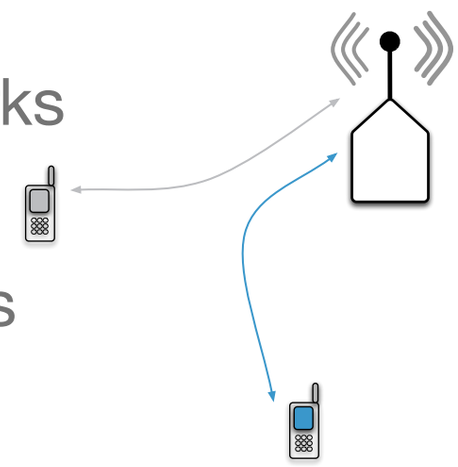


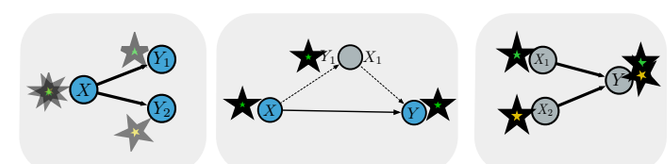
Source coding, channel coding, entropy and mutual information, capacity, Gaussian noise channel

- **Two-way channel - *channel coding***



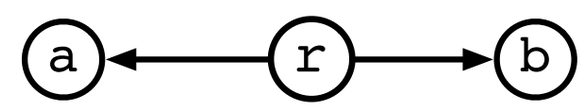
- Wireless channels and networks





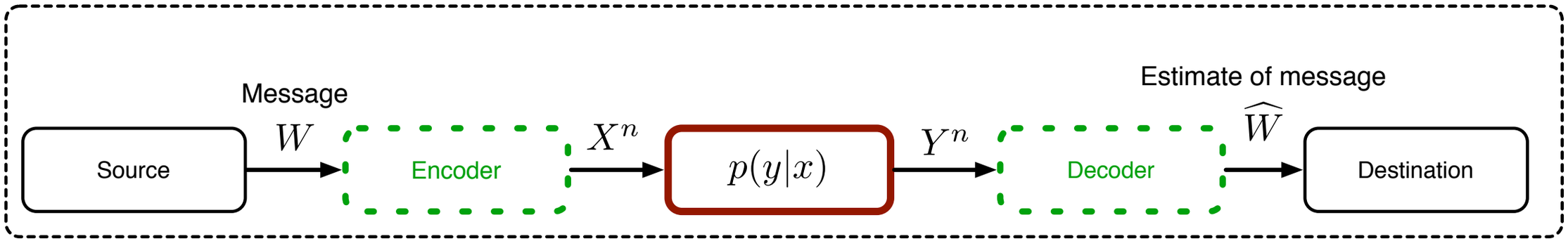
- Two-way cellular-like networks

- Two-way relay channels - *canonical example of wireless network coding*



$$w_r = w_1 \oplus w_2$$

One-way channel capacity



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

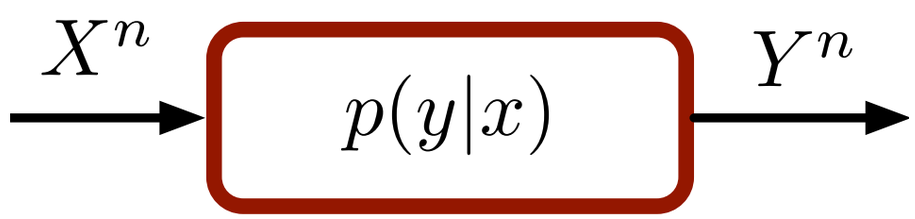
bits/channel use

"mutual information" between X and Y

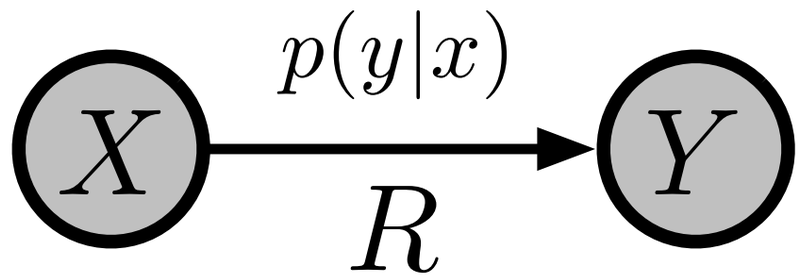
$$I(X;Y) = \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

symmetric in its arguments!

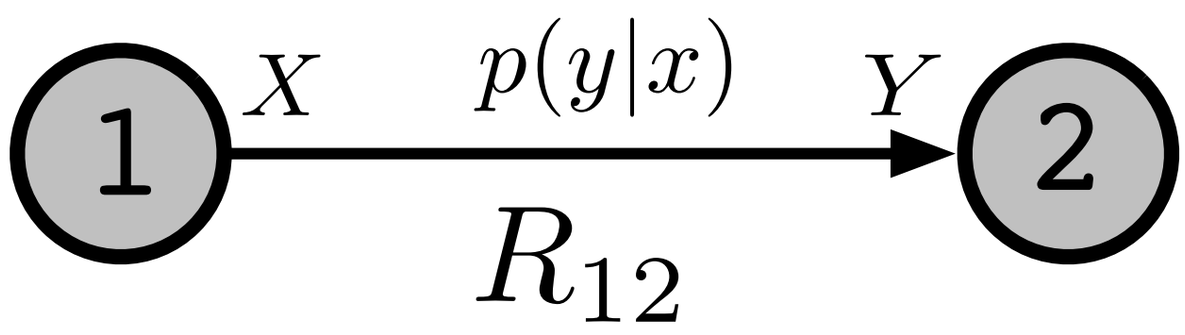
One-way channel capacity - notation



$\Updownarrow$



$\Updownarrow$



Two-way channel (historical)

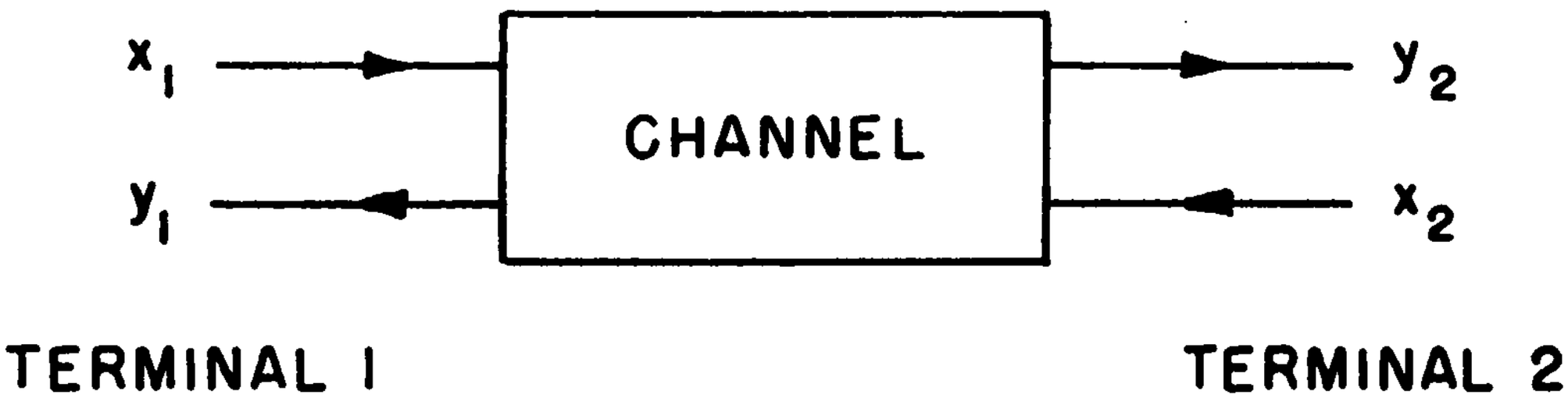
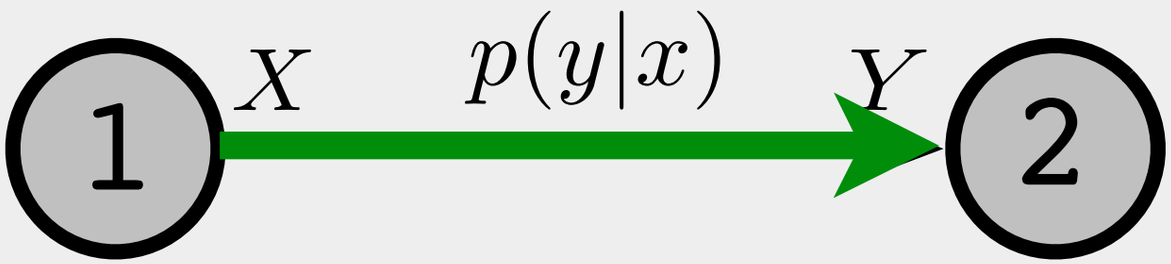


FIGURE 1

[Shannon '61]

Two-way channel capacity

One-way



Two-way

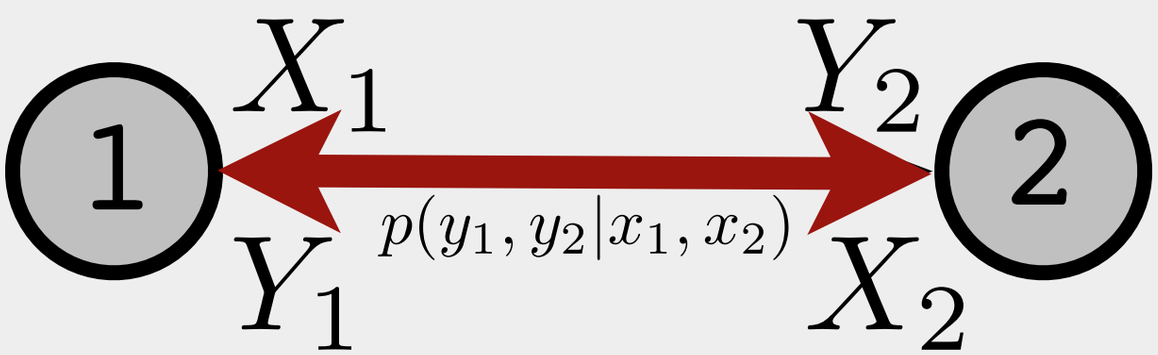
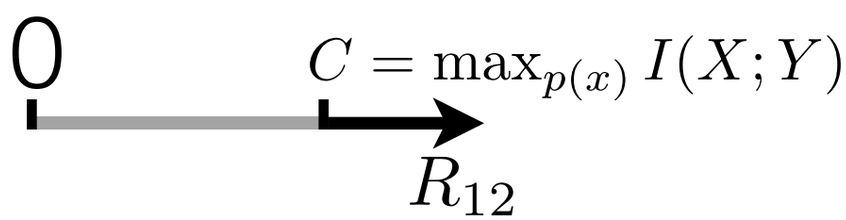
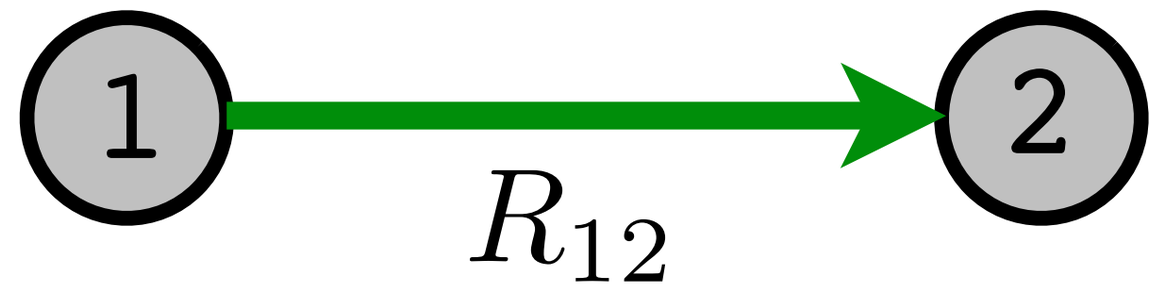


FIGURE 1

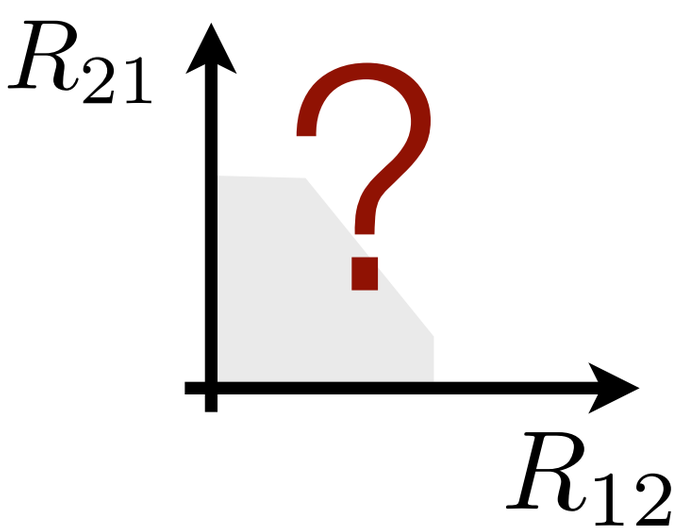
Two-way channel capacity region

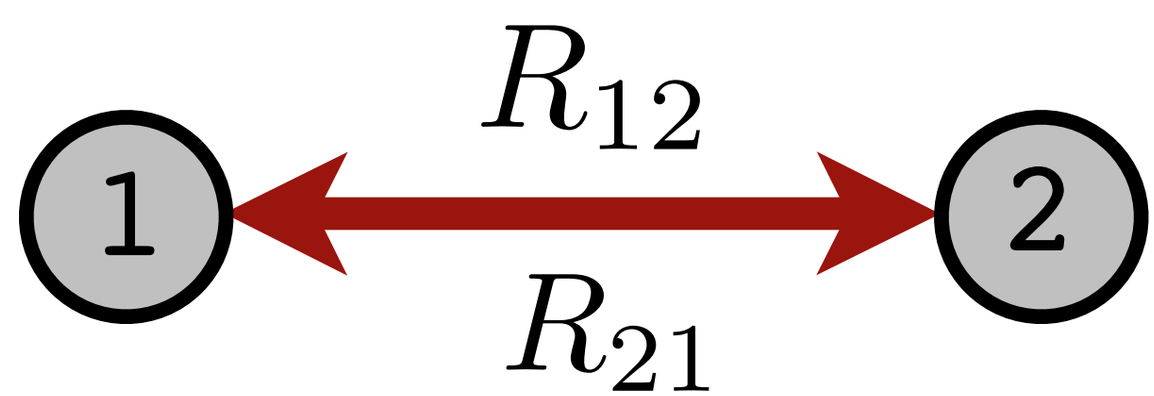
One-way ***Capacity*** ✓



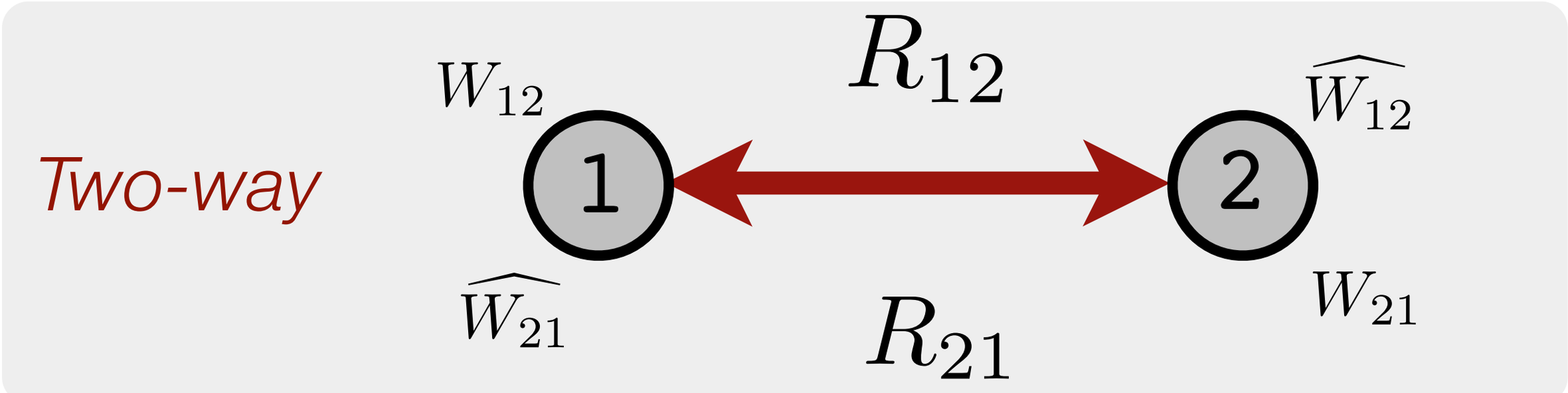


Two-way
Capacity Region

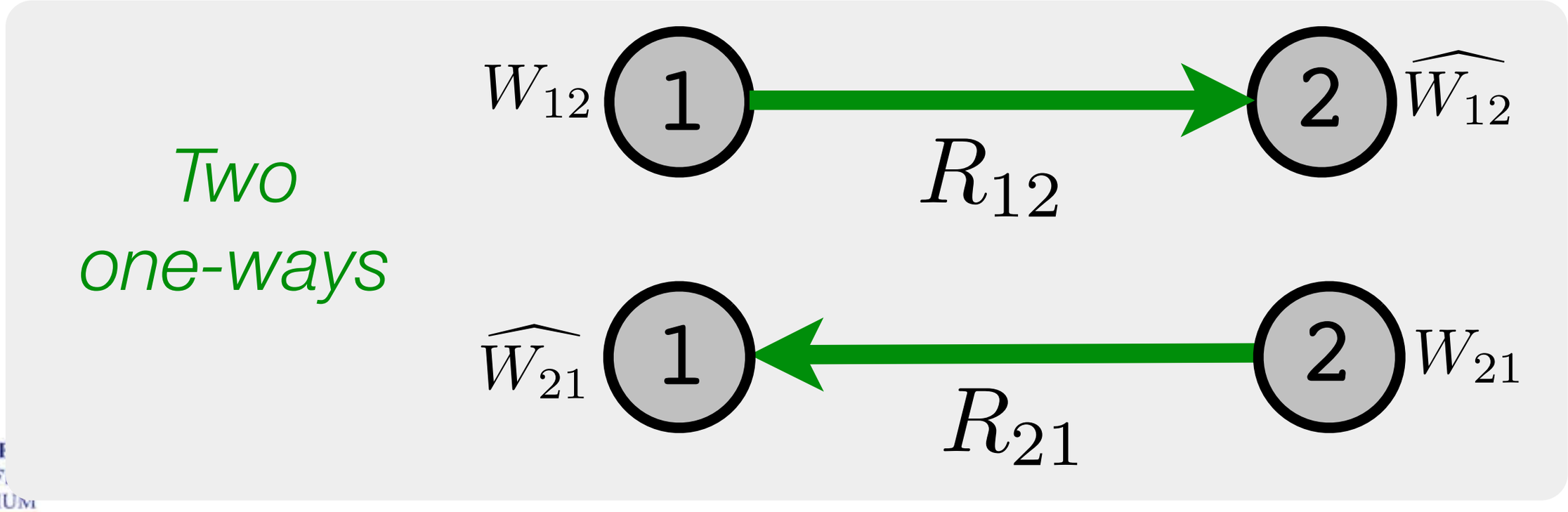




When is



equal to

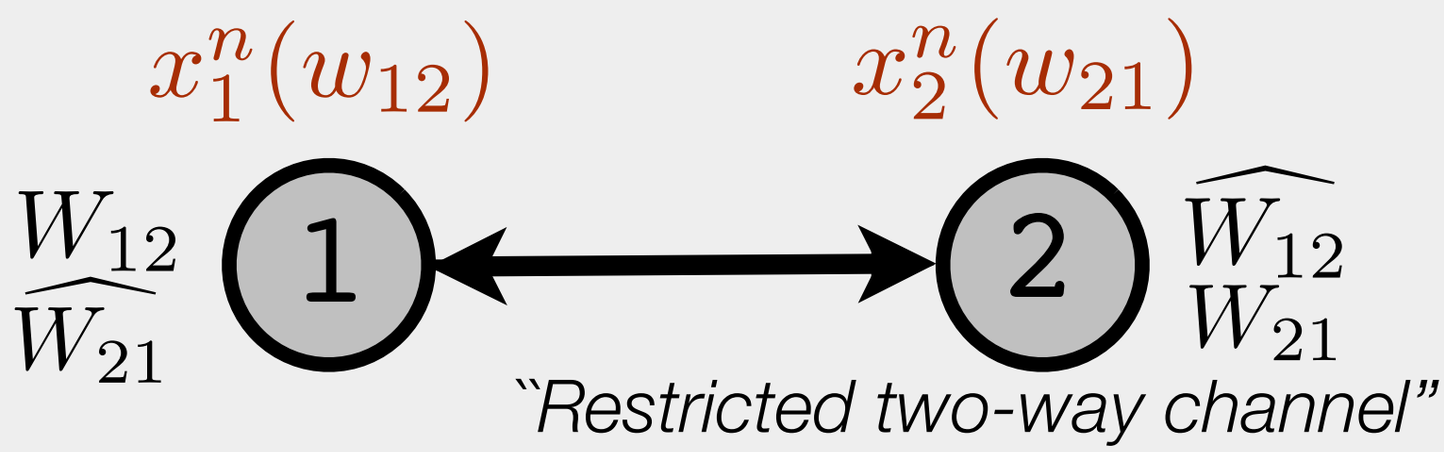


?

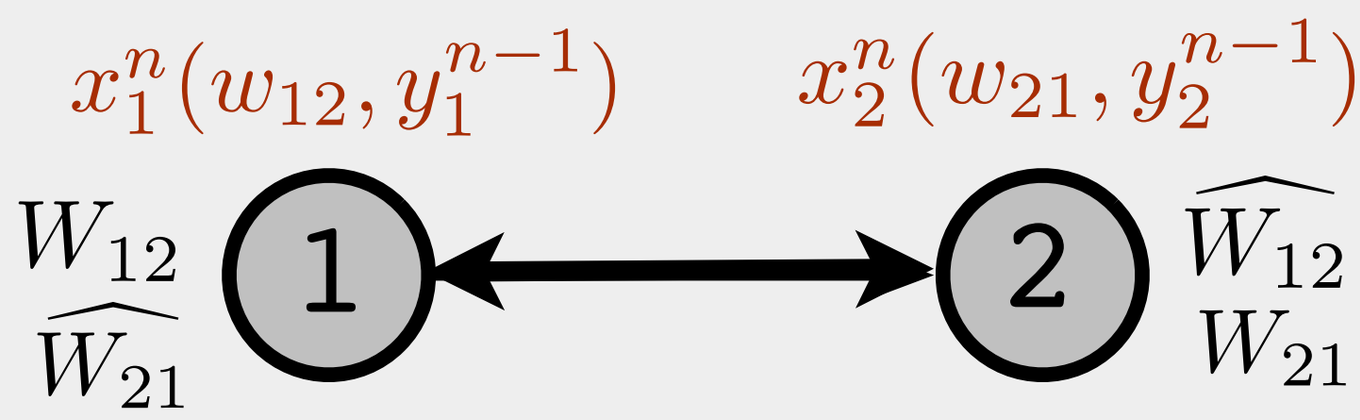
Models for two-way adaptation

One-way: no adaptation possible

Two-way: no adaptation

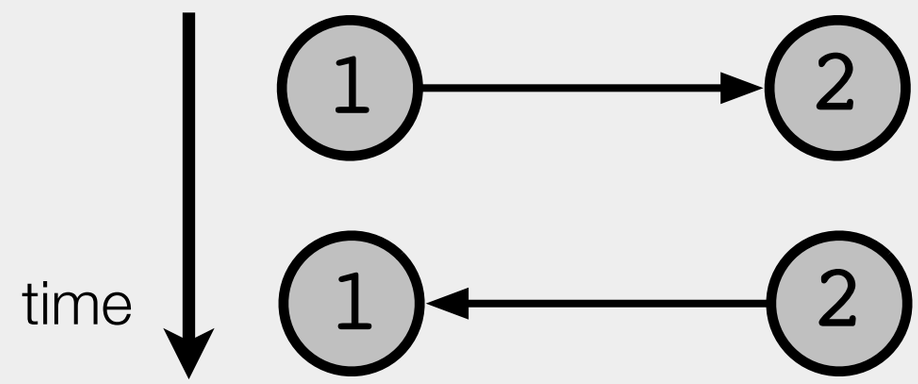


Two-way: full adaptation



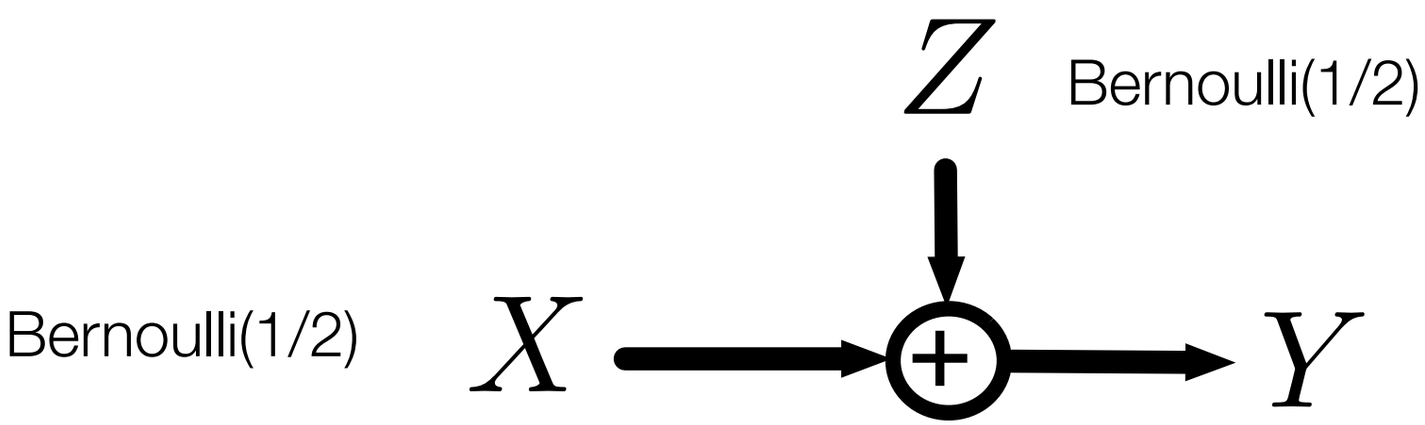
Duplex

Two-way: half duplex



Two-way: full duplex

(Aside: I(X;Y) and I(X;Y|Z))

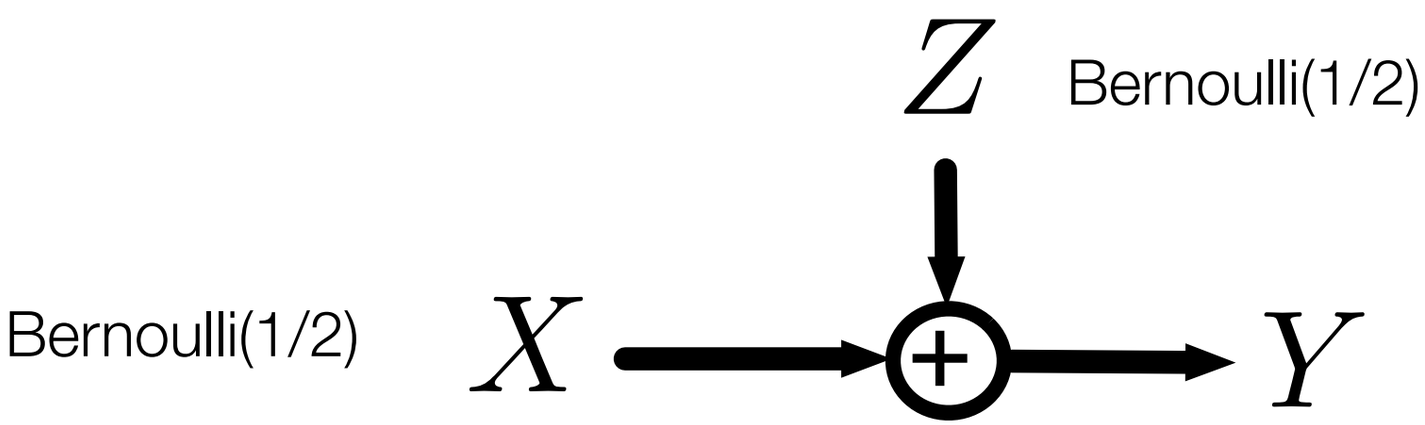


X \ Y \ Z	0	1
0	0	1
1	1	2

$$I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$
$$= E_{p(x,y)} \log \frac{p(X,Y)}{p(X)p(Y)}$$

$$I(X;Y|Z) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)}$$
$$= E_{p(x,y,z)} \log \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)}$$

(Aside: I(X;Y) and I(X;Y|Z))



X \ Y \ Z	0	1
0	0	1
1	1	2

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= \log_2(4)/4 + \log_2(2)/2 + \log_2(4)/4 - \log_2(2) \\
&= 0.5 \text{ bits}
\end{aligned}
$$

$$
\begin{aligned}
I(X;Y|Z) &= H(Y|Z) - H(Y|X,Z) \\
&= \log_2(2) - 0 \\
&= 1 \text{ bit}
\end{aligned}
$$

General results

Inner bound

$$R_1 \leq I(X_1; Y_2|X_2)$$
$$R_2 \leq I(X_2; Y_1|X_1)$$

where X_1 and X_2 follow the joint distribution $p(x_1, x_2) = p(x_1)p(x_2)$.

Not in general equal!

Outer bound

$$R_1 \leq I(X_1; Y_2|X_2)$$
$$R_2 \leq I(X_2; Y_1|X_1)$$

where the joint distribution of random variables X_1 and X_2 is $p(x_1, x_2)$.

[Shannon '61]

When is capacity known

- Parallel two-way channel
- Mod-2 adder
- Two-way restricted channel
- Two-way “push-to-talk” channel
- Two-way Gaussian noise channel (full & half duplex, restricted & unrestricted)

When is capacity unknown

- General unrestricted discrete memoryless channels
- Binary multiplier channel (BMC)

Capacity: two parallel channels

[Shannon '61]

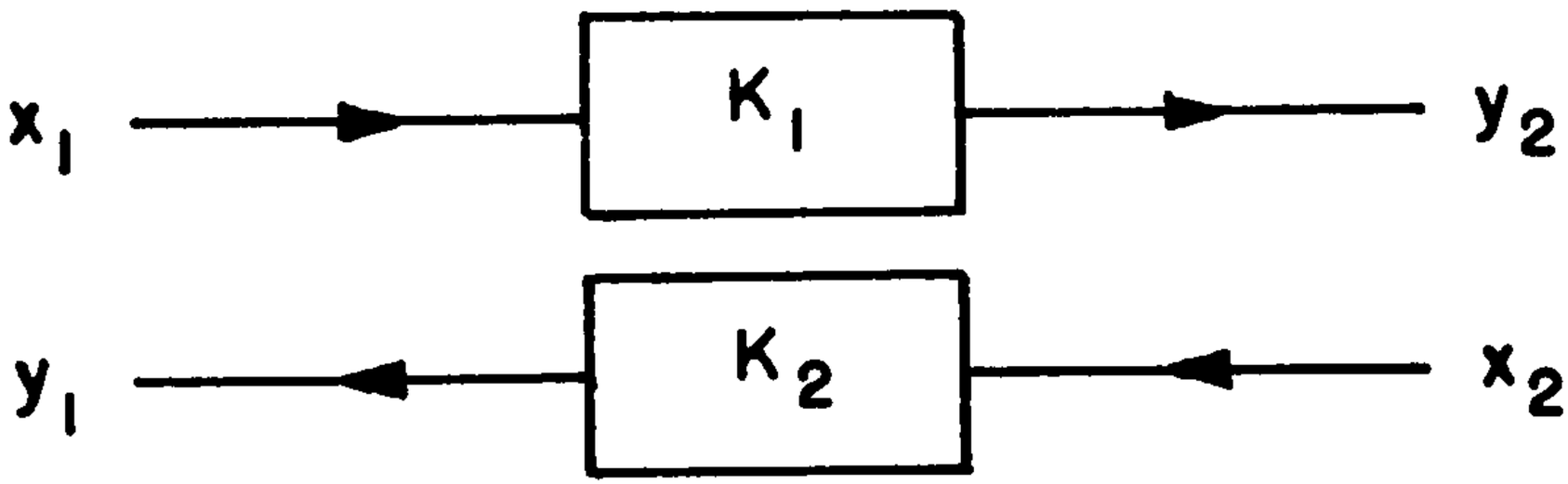
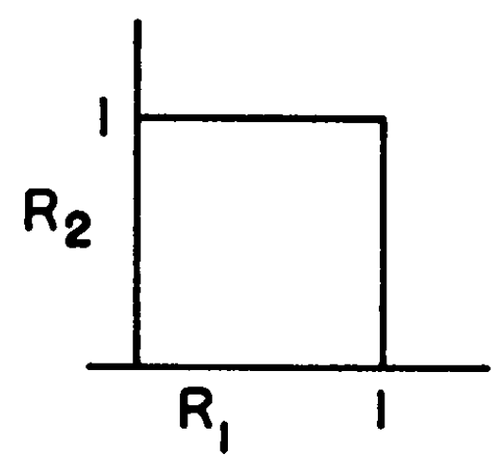


FIGURE 2

$$R_{12} \leq C_{K_1}$$
$$R_{21} \leq C_{K_2}$$



Capacity: binary mod-2 adder channel

[Shannon '61]

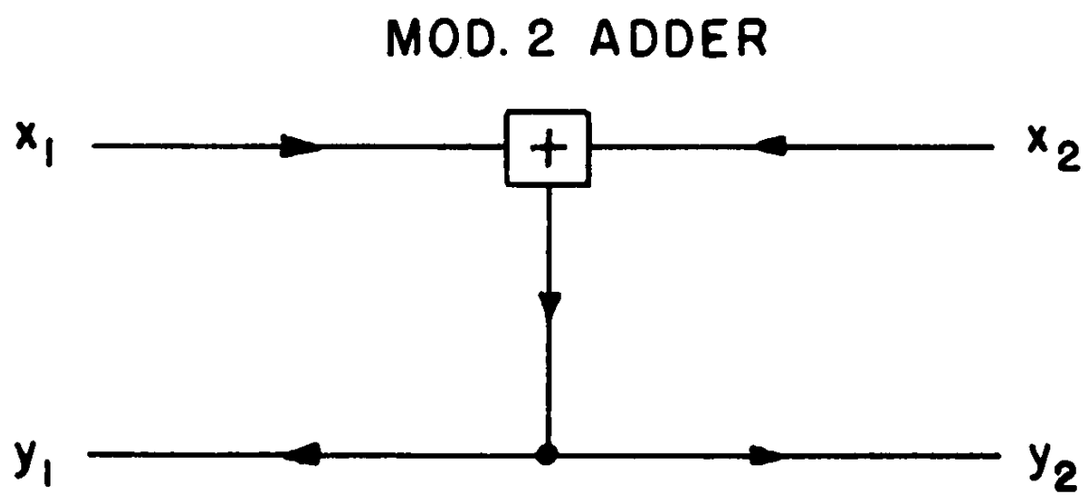


FIGURE 4

$$y_1 = y_2 = x_1 + x_2 \pmod 2.$$

How to achieve capacity region?

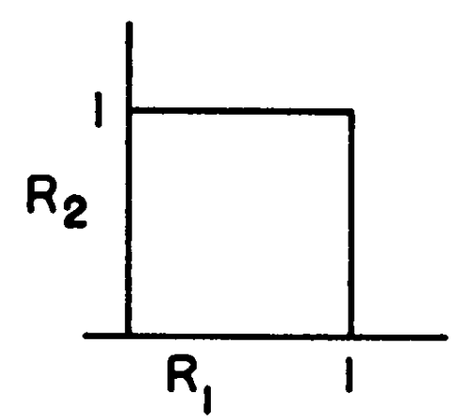


FIGURE 3

Achieving mod 2 adder channel capacity

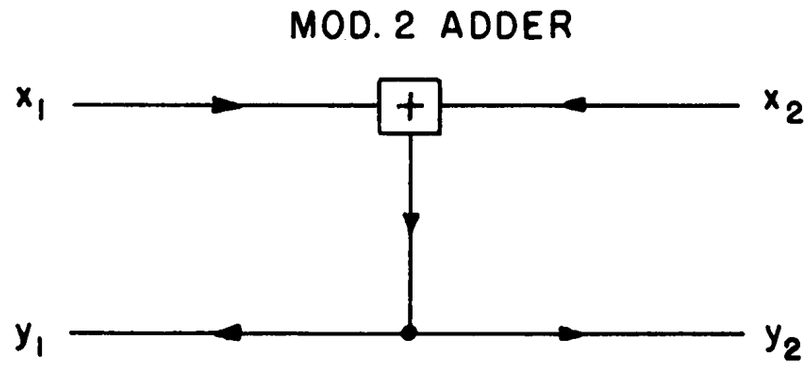


FIGURE 4

$$y_1 = y_2 = x_1 + x_2 \pmod 2.$$

Receiver 1:

$$\hat{x_2} = y_1 \oplus x_1$$
$$= x_1 \oplus x_2 \oplus x_1$$
$$= x_2$$

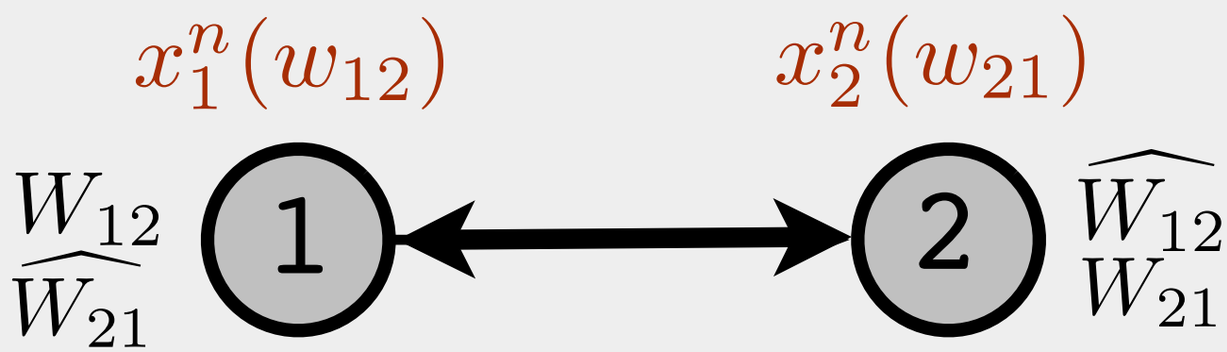
$Y_1=Y_2$	X_2: 0	X_2: 1
X_1: 0	0	1
X_1: 1	1	0

Receiver 2:

$$\hat{x_1} = y_2 \oplus x_2$$
$$= x_1 \oplus x_2 \oplus x_2$$
$$= x_1$$

EXPLOIT TWO-WAY!

Capacity: restricted channel



Capacity region:

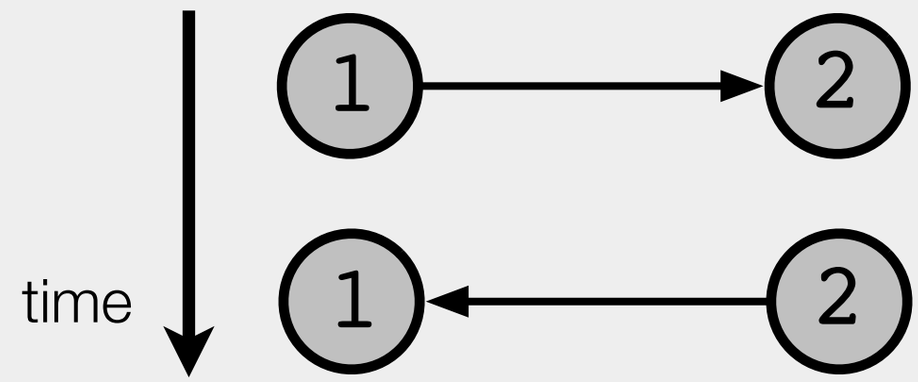
$$R_1 \leq I(X_1; Y_2 | X_2)$$
$$R_2 \leq I(X_2; Y_1 | X_1)$$

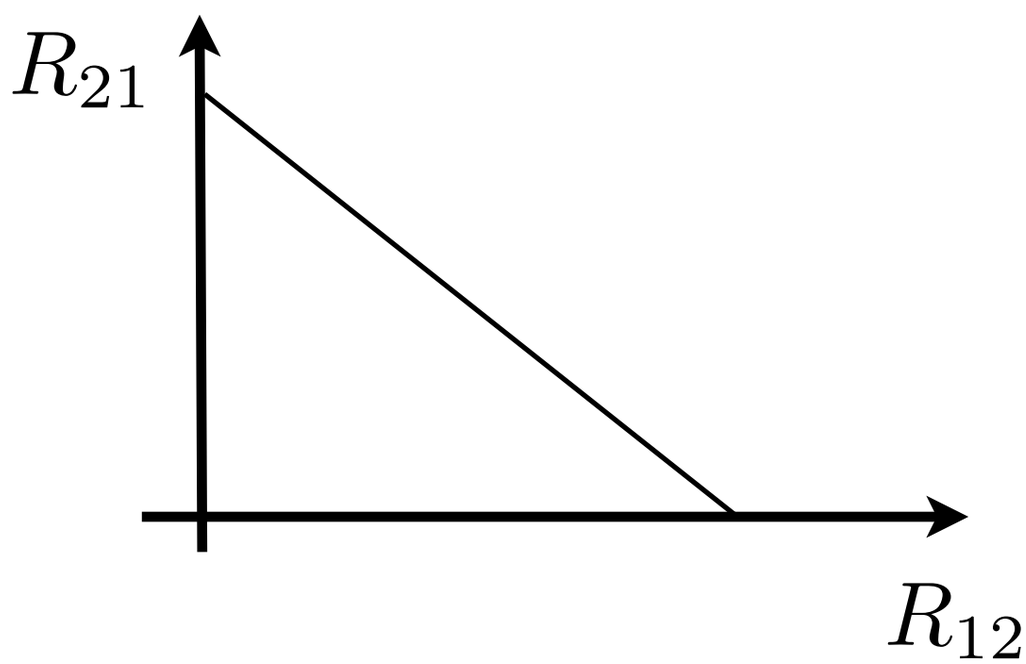
where X_1 and X_2 follow the joint distribution $p(x_1, x_2) = p(x_1)p(x_2)$.

[Shannon '61]

Capacity: ``push-to-talk" channel

Two-way:
half duplex





Capacity: Gaussian noise channel

$$Y_1 = aX_1 + bX_2 + N_1 \sim \mathcal{N}(0, \sigma_1^2)$$

$$Y_2 = cX_1 + dX_2 + N_2 \sim \mathcal{N}(0, \sigma_2^2)$$

Capacity region:

$$R_1 \leq (1/2)\log(1 + c^2 P_1/\sigma_2^2)$$

$$R_2 \leq (1/2)\log(1 + b^2 P_2/\sigma_1^2)$$

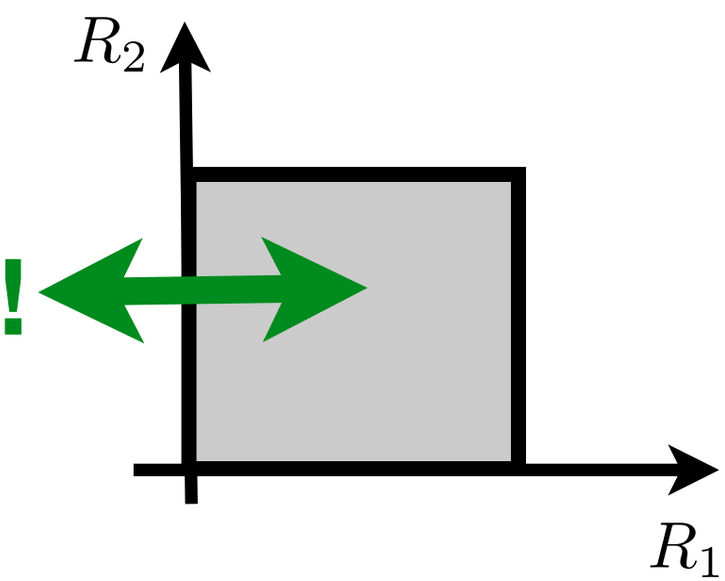
No dependence on ``a" or ``d"

[Han '84]

Capacity: Gaussian noise channel

$$Y_1 = aX_1 + bX_2 + N_1$$
$$Y_2 = cX_1 + dX_2 + N_2$$

$$R_1 \leq (1/2)\log(1 + c^2 P_1/\sigma_2^2)$$
$$R_2 \leq (1/2)\log(1 + b^2 P_2/\sigma_1^2)$$



- **TWO PARALLEL CHANNELS!!**
- Achieved by Gaussian inputs
- ``Feedback" does not help here

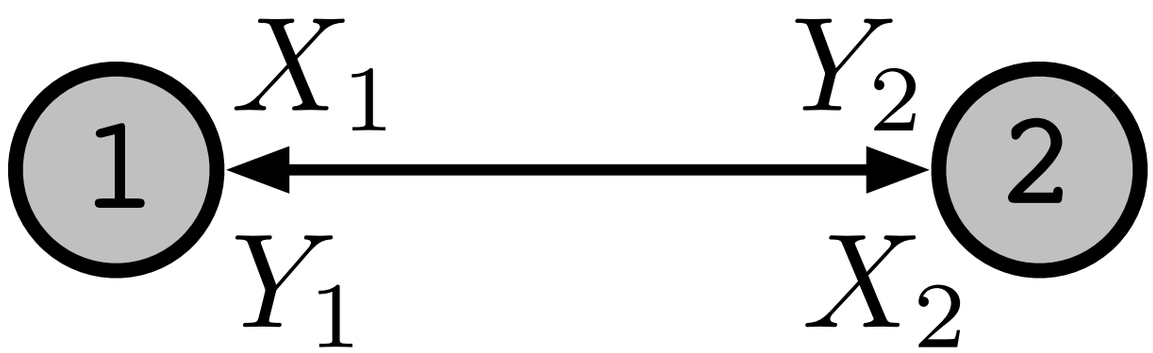
When is capacity known

- Parallel two-way channel
- Mod-2 adder
- Two-way restricted channel
- Two-way “push-to-talk” channel
- Two-way Gaussian noise channel (full & half duplex, restricted & unrestricted)

When is capacity unknown

- General unrestricted discrete memoryless channels
- Binary multiplier channel (BMC)

Capacity unknown: Binary Multiplier Channel



$Y_1=Y_2$ \ X_2	0	1
X_1 0	0	0
1	0	1

$$y_1 = y_2 = x_1 x_2, \text{ where } x_1, x_2 \in \{0, 1\}.$$

lower/upper bounds	Rate (bits)
Shannon's lower bound	0.61695
Hagelbarger's lower bound	0.593
Schalkwijk's lower bound	0.63056
Shannon's upper bound	0.69424
Zheng's upper bound	0.6390

C. E. Shannon, *Two-way communication channels*, in Proc. 4th Berkeley Symp. Math. Satist. Probab., vol. 1, 1961, pp. 611-644.

J.P. M. Schalkwijk, *The binary multiplying channel - a coding scheme that operates beyond the Shannon inner bound*, IEEE Trans. Inform. Theory, vol. IT-28(1), pp. 107-110, jan 1982.

J.P. M. Schalkwijk, *On an extension of an achievable rate region for the binary multiplying channel*, IEEE Trans. Inform. Theory, vol. IT-29, pp. 445-448, May 1983.

J.P. M. Schalkwijk, *Extending the achievable rate region of the binary multiplying channel*, In Proc. Int. Symp. on Information Theory, page 302, Budapest, Hungary, Jun 1991.

Z. Zheng, T. Berger,J.P. M. Schalkwijk, *New outer bounds to capacity region of two-way channels*, IEEE Trans. Inform. Theory, vol. IT-32, pp. 383-386, May. 1986.

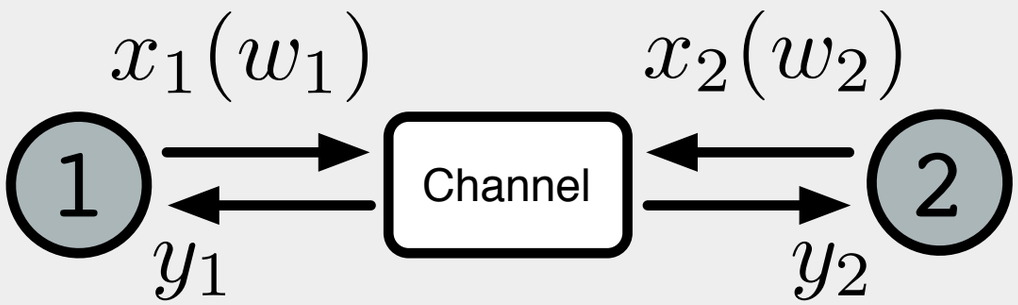
General techniques for two-way channels

Adaptation

Adaptive codewords

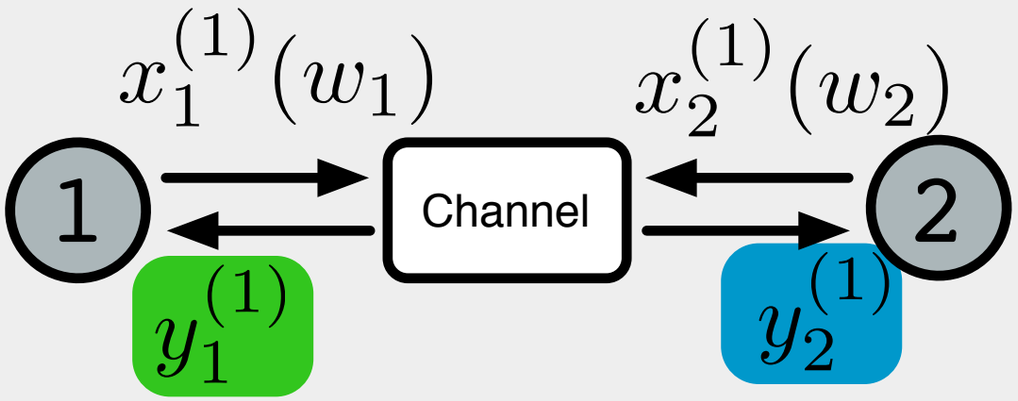
Single-letter

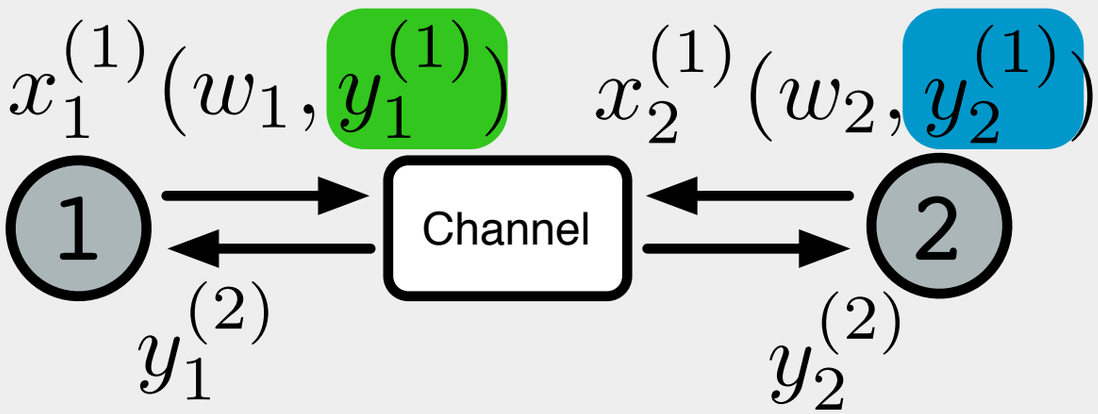
$C = f(X_1, X_2)$



Two-letter

$C = f(X_1^{(1)}, X_1^{(2)}, X_2^{(1)}, X_2^{(2)})$



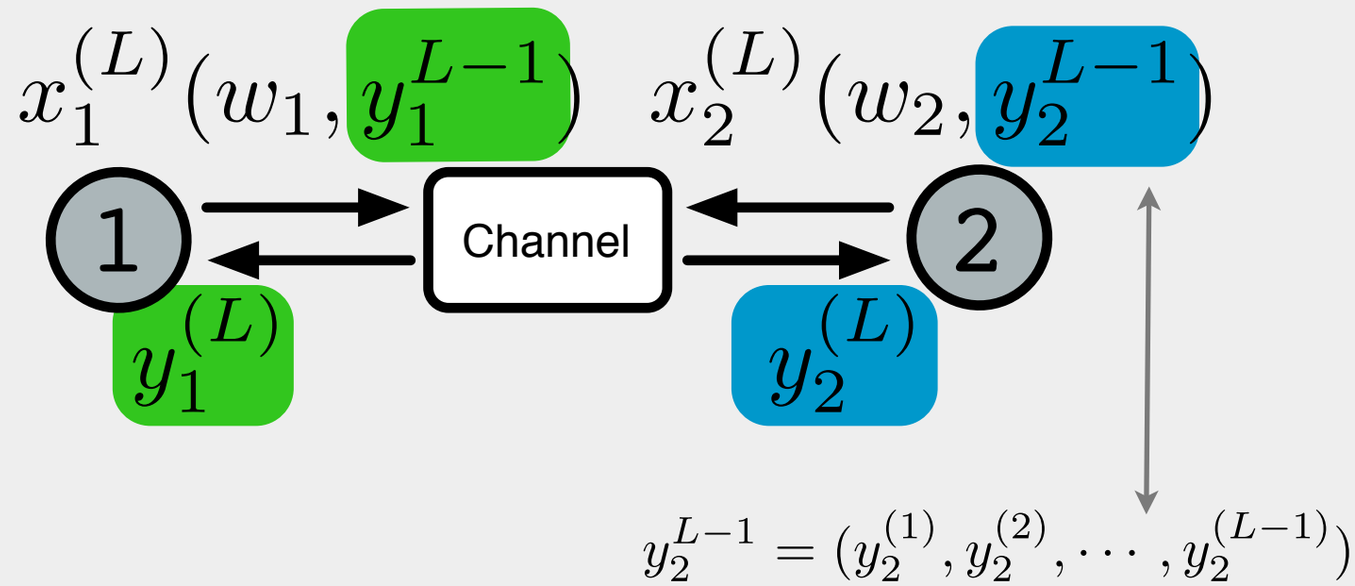


G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

S. M. S. Tatikonda, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

Adaptive codewords

L-letter



The space over which we can code (x's) is enormous!

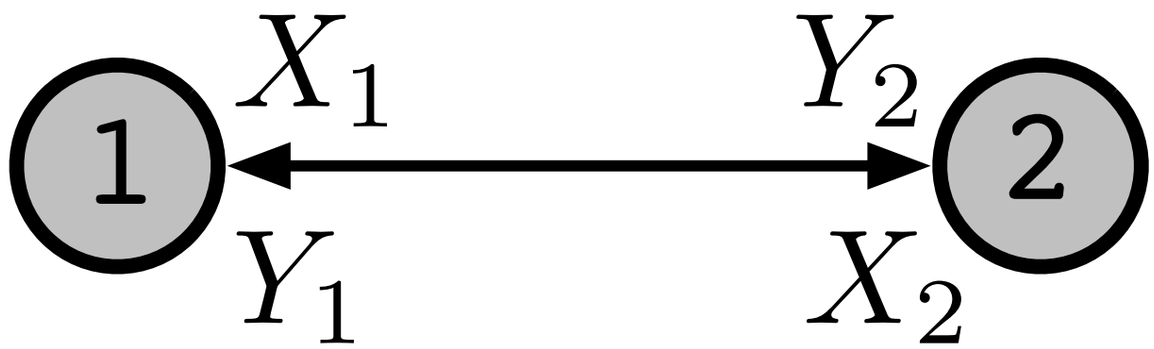
G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

S. M. S. Tatikonda, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

Non-adaptive codewords:

$X_1 \in \{0,1\}$

$Y_1 = X_1 X_2$



$Y_2 = Y_1 \mod 2$

$X_2 \in \{0,1,2\}$

Code of user 1

$u_{11} = 0$ 0 0 0

$u_{11} = 1$ 1 1 1

Code of user 2

$u_{21} = 0$ 1 0 2

$u_{21} = 1$ 0 2 0

33rd IEEE SARNOFF SYMPOSIUM 2010
Princeton, New Jersey
April 12-14, 2010

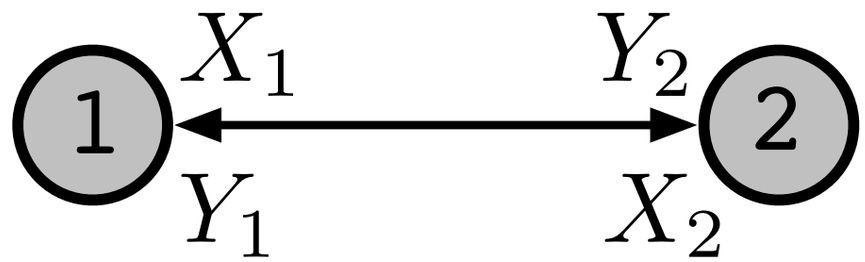
L=N=3 channel uses

G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology Zurich, 1998.

Adaptive codewords:

$X_1 \in \{0,1\}$

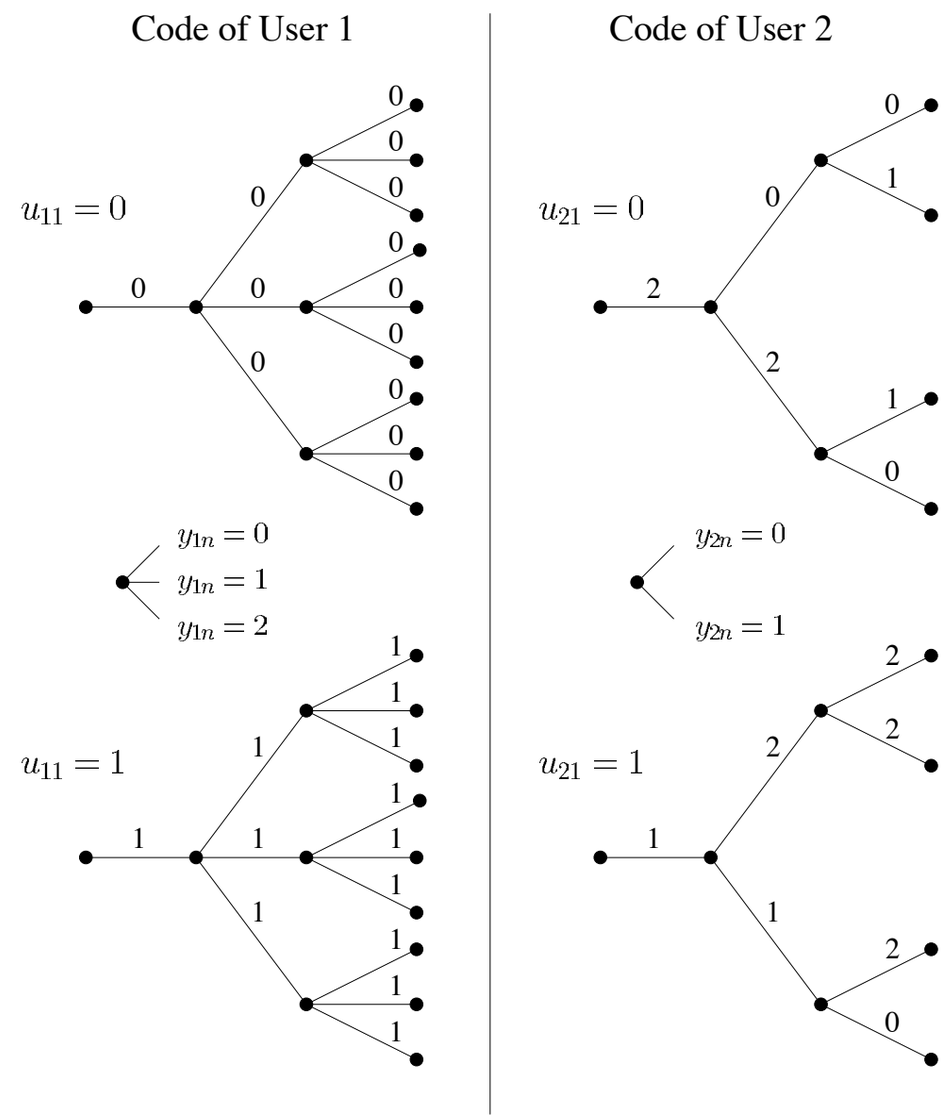
$Y_1 = X_1 X_2$



$Y_2 = Y_1 \mod 2$

$X_2 \in \{0,1,2\}$

Adaptive codewords $\mathbf{A}_1^3$



Adaptive codewords $\mathbf{A}_2^3$

L=N=3 channel uses

G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology Zurich, 1998.

Can we take this adaptation into account?

CAUSAL adaptation

From mutual information to directed information

$X^N := (X_1, X_2, \cdots, X_N)$
$Y^N := (Y_1, Y_2, \cdots, Y_N)$

Need to extend the symmetric $I(X^N; Y^N)$
to account for causally adaptive codewords

Marko/Massey's ``Directed Information"

$$I(X^N \to Y^N) := \sum_{n=1}^{N} I(X^n; Y_n | Y^{n-1})$$

Kramer's ``Causally conditioned Directed Information"

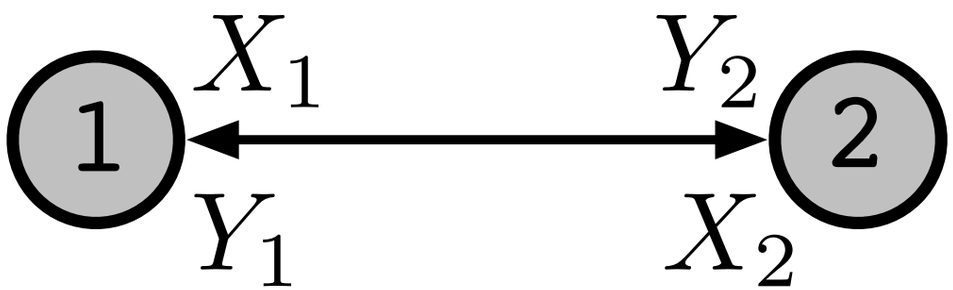
$$I(X^N \to Y^N \| Z^N) := \sum_{n=1}^{N} I(X^n; Y_n | Y^{n-1} Z^n)$$

G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology Zurich, 1998.

H. Marko, "The bidirectional communication theory- a generalization of information theory," *IEEE Trans. Commun.*, vol. COM-21, no. 12, pp. 1345–1351, Dec. 1973.

J. Massey, "Causality, feedback and directed information," in *Proc. IEEE Int. Symp. Inform. Theory and its Applications*, Honolulu, Nov. 1990.

Aside - could this be what we need?



$$I(X^N; Y^N) = I(Y^N; X^N)$$

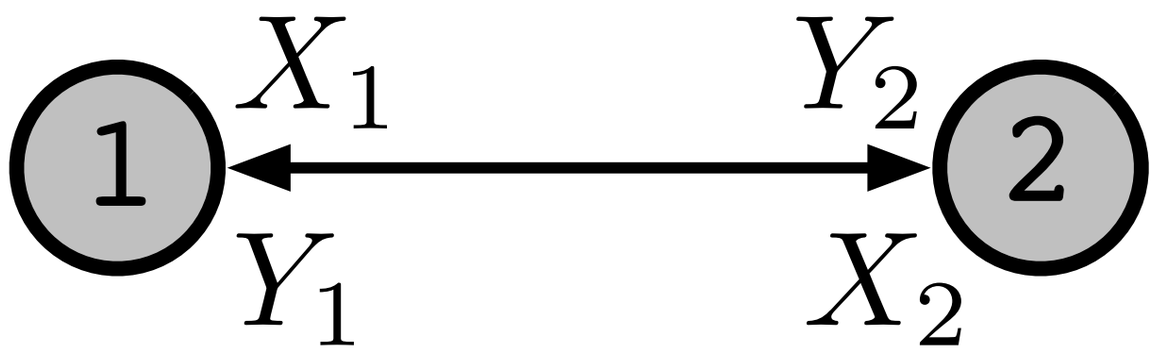
BUT

$$I(X^n \to Y^N) \neq I(Y^n \to X^N)$$

Usage: capacity of two-way channels

$X_1 \in \{0,1\}$

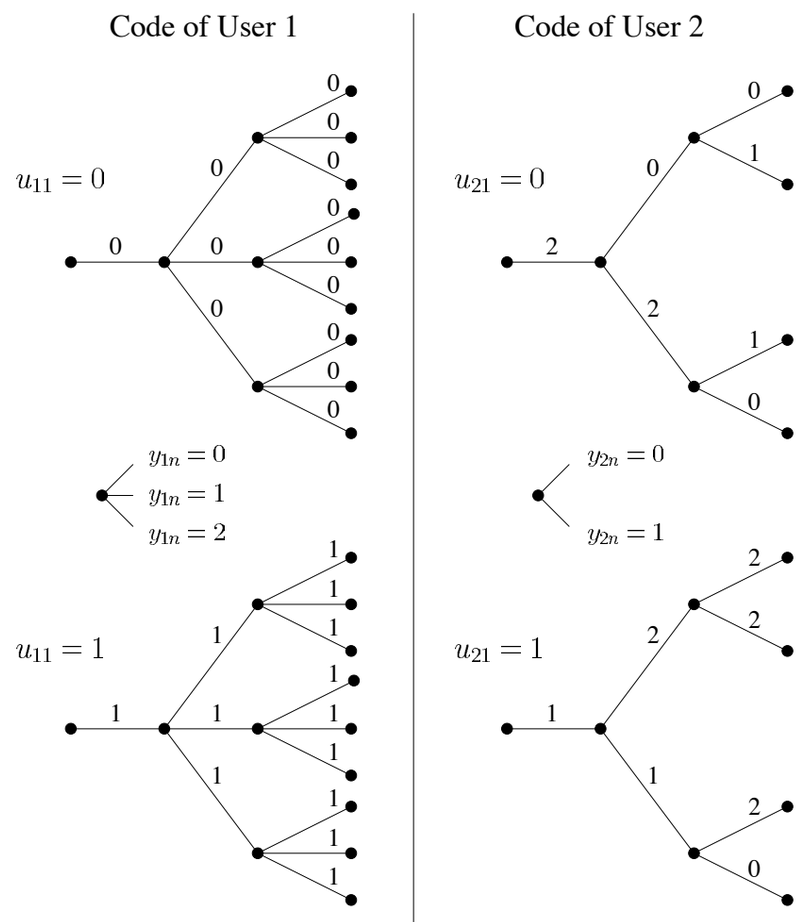
$Y_1 = X_1 X_2$



$Y_2 = Y_1 \mod 2$

$X_2 \in \{0,1,2\}$

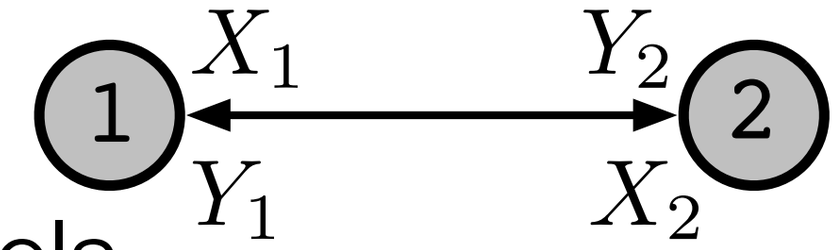
Adaptive codewords $\mathbf{A}_1^3$



Adaptive codewords $\mathbf{A}_2^3$

L=N=3 channel uses

G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, Swiss Federal Institute of Technology Zurich, 1998.



Usage: capacity of two-way channels

The capacity region of two-way channel is given by the limit as $L \to \infty$ of the regions

$$R_1 = I(\mathbf{A_1^L} \to Y_2^L || X_2^L)$$
$$R_2 = I(\mathbf{A_2^L} \to Y_1^L || X_1^L)$$

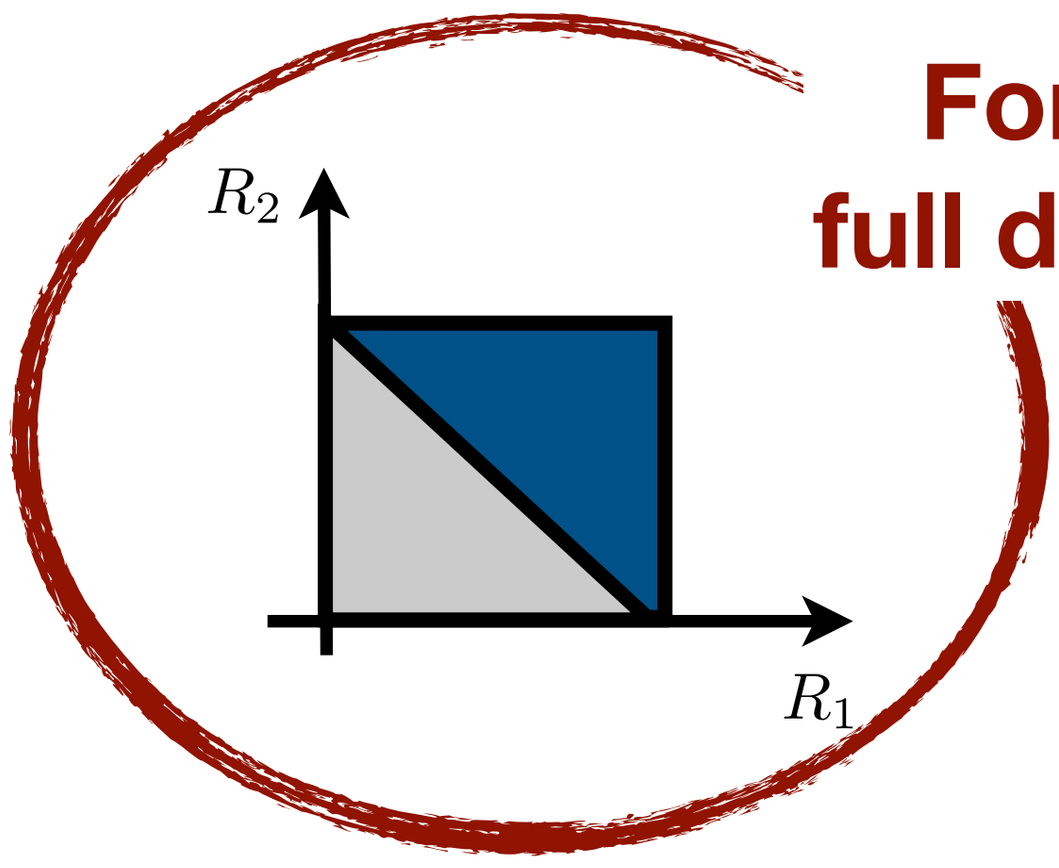
where $\mathbf{A_1^L}, \mathbf{A_2^L}$ are adaptive independent codewords.

Difficulty: space of codewords hard to compute!

Take away points - AWGN two-way channel

- If have half-duplex constraint and memoryless channels, time-share

- If have full-duplex - obtain two parallel clean channels



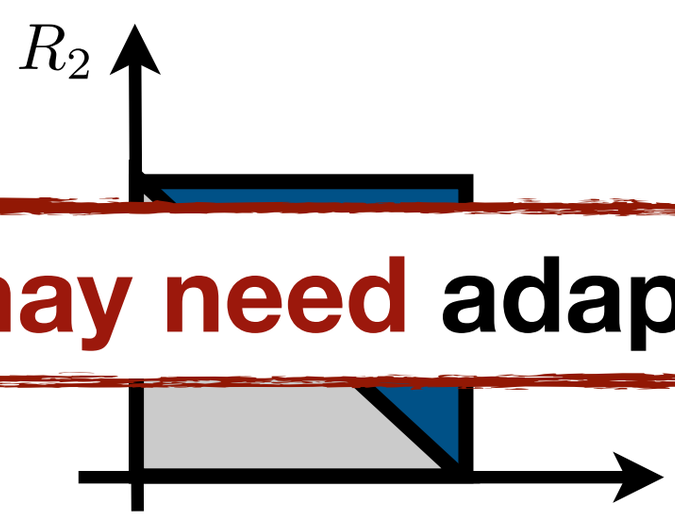
For applications - full duplex gains a lot!

Take away points - Discrete memoryless two-way channel

- If have half-duplex constraint (`push-to-talk”), time-share
- If have parallel two-way channels, mod-2 adder
- If have restricted channel

$$R_1 \leq I(X_1; Y_2|X_2)$$
$$R_2 \leq I(X_2; Y_1|X_1)$$

where X_1 and X_2 follow the joint distribution $p(x_1, x_2) = p(x_1)p(x_2)$.

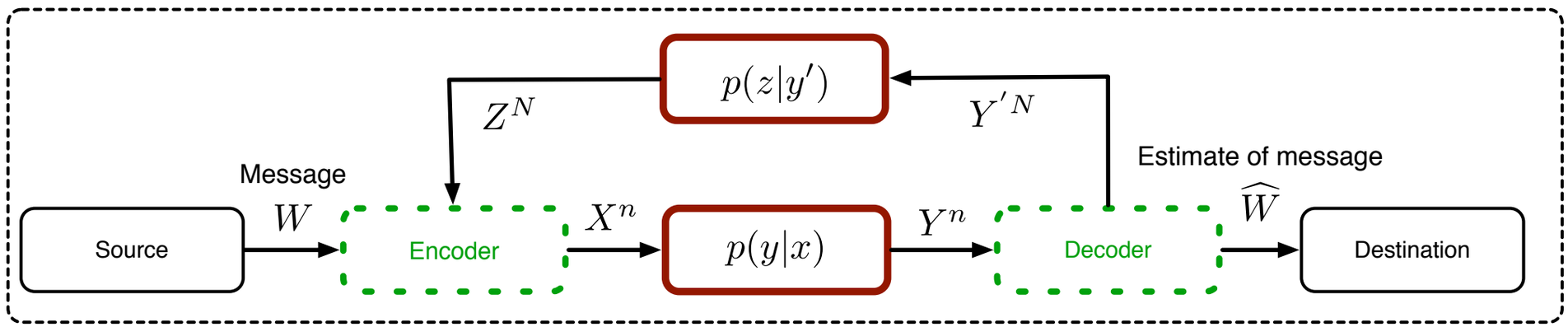


In general may need adaptive codewords

In general OPEN PROBLEM

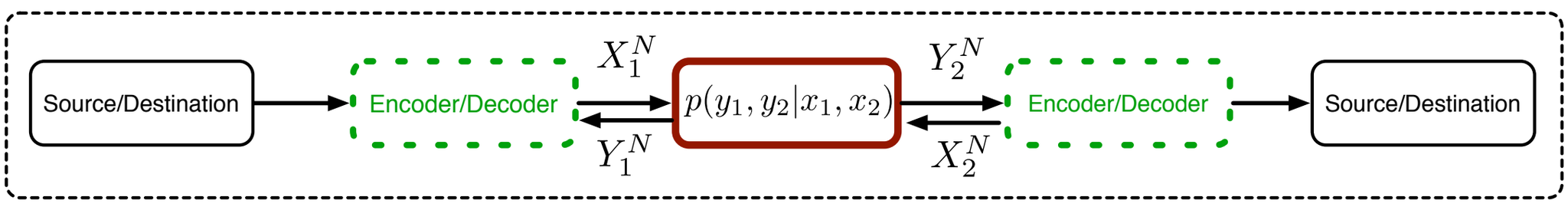
Relationship to feedback channels

- Feedback channel



Information is still one-way!!

- Two-way channel



Set rate in ← direction to 0

Properly choose $p(y_1, y_2|x_1, x_2)$

Two-way source coding = what skipping

- Two-way (lossy) source coding using block-coding and protocols of K rounds

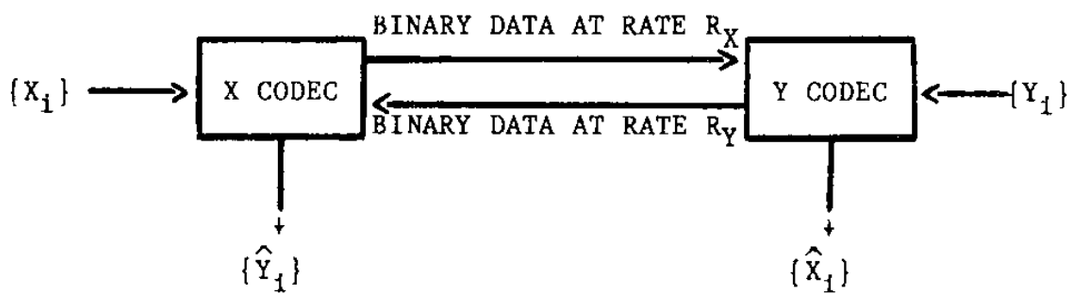
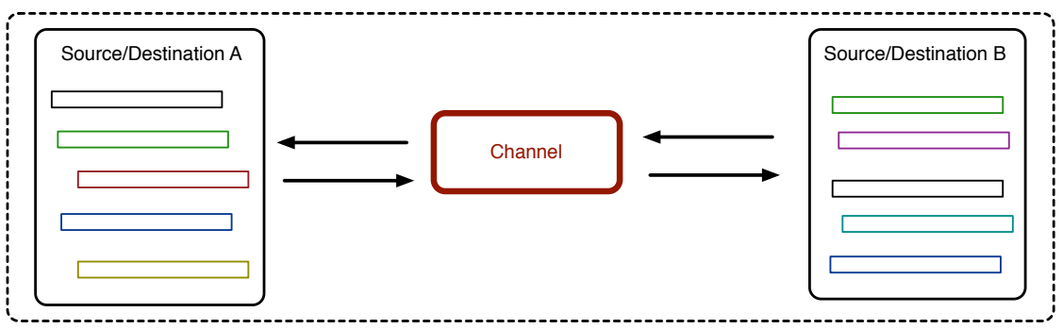


Fig. 1. Two-way source coding scheme.

A. Kaspi, "Two-Way Source Coding with a Fidelity Criterion," *IEEE Trans. Inform. Theory*, vol. IT–31, no. 6. pp. 735–740, November 1985.

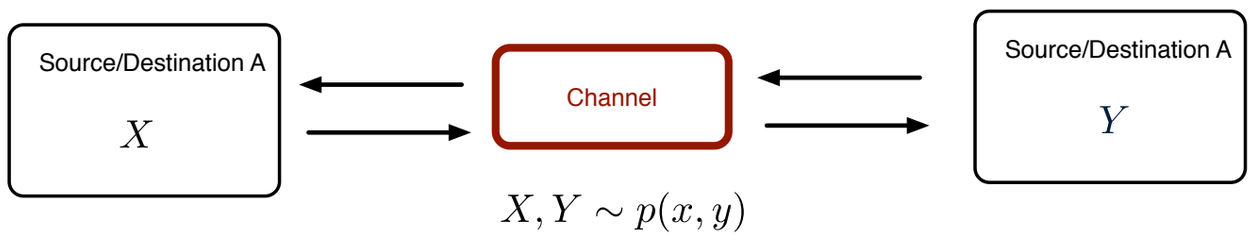
N. Merhav. Two-way successively refined joint source-channel coding. *IEEE Trans. Inf. Theory*, 52(4):1483–1494, 2006.

- Set reconciliation framework



Y. Minsky, A. Trachtenberg, and R. Zippel, "Set reconciliation with nearly optimal communication complexity," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2213–2218, 2003.

- Interactive communication framework / communication complexity



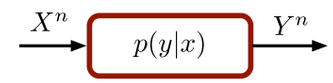
A. Orlitsky, "Worst-case interactive communication I: Two messages are almost optimal," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1111–1126, Sept. 1990.

A. Orlitsky, "Worst-case interactive communication II: Two messages are not optimal," *IEEE Trans. Inform. Theory*, vol. 37, pp. 995–1005, July 1991.

Outline

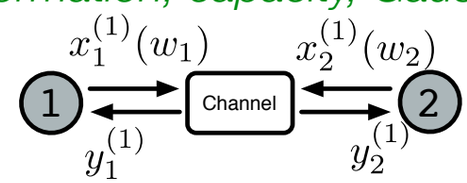
- Information theory - *what, why, when*

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



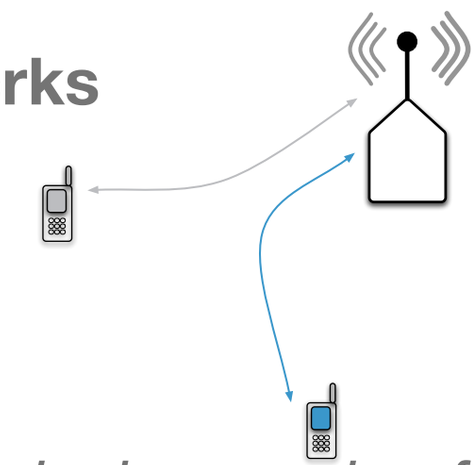
Source coding, channel coding, entropy and mutual information, capacity, Gaussian noise channel

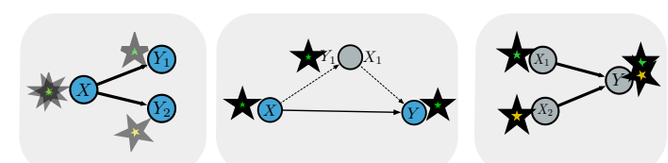
- Two-way channel - *channel coding*



Adaptive codewords, capacity in Gaussian noise = two parallel channels

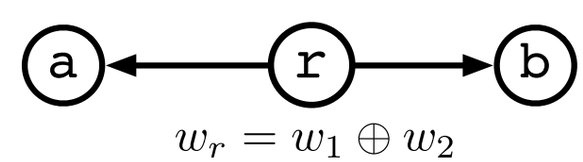
- **Wireless channels and networks**



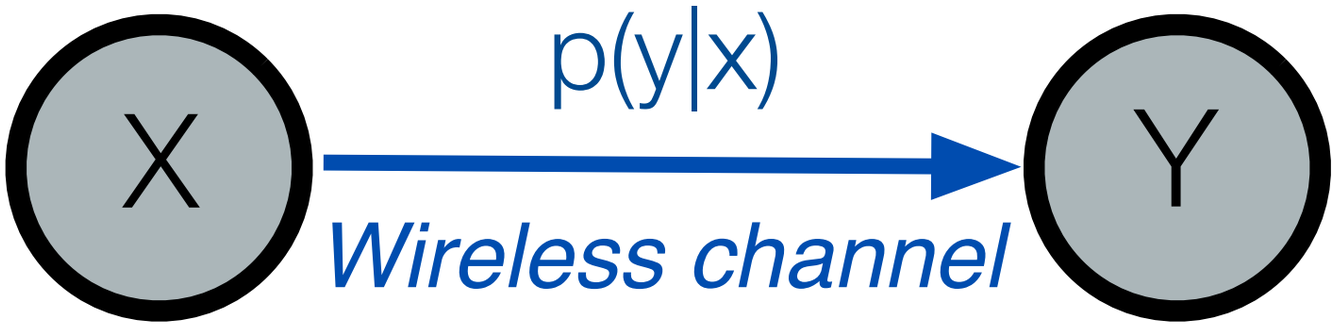


- Two-way cellular-like networks

- Two-way relay channels - *canonical example of wireless network coding*



Point-to-point



- Channel capacity ✓
- How to approach it for memoryless Gaussian noise channels ✓

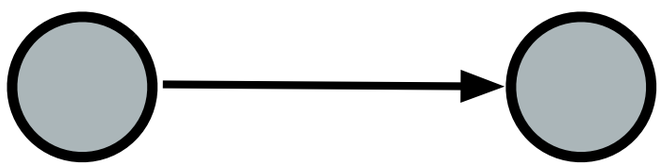
Is that the end of the story?

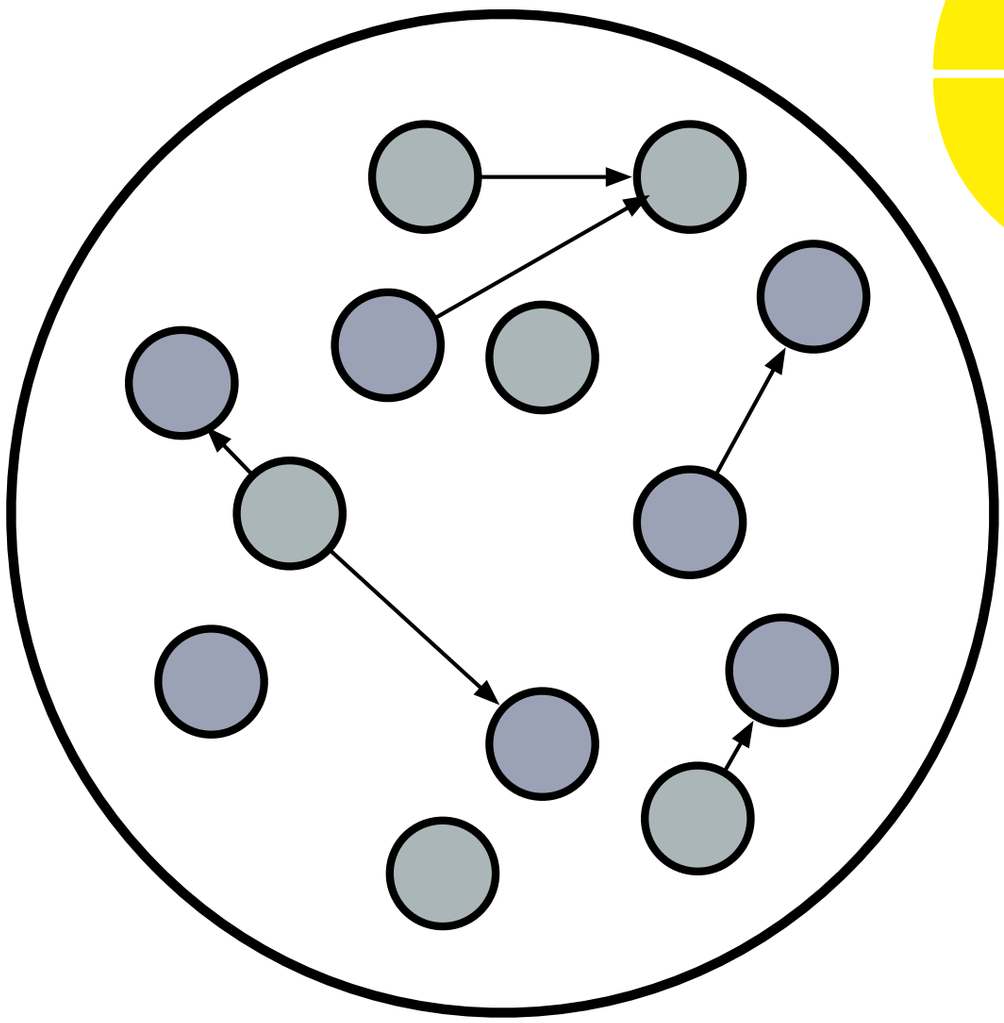
NO! Motivation 1: two-way channels

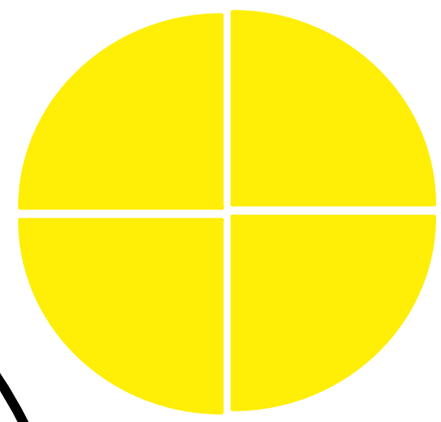
Unsolved

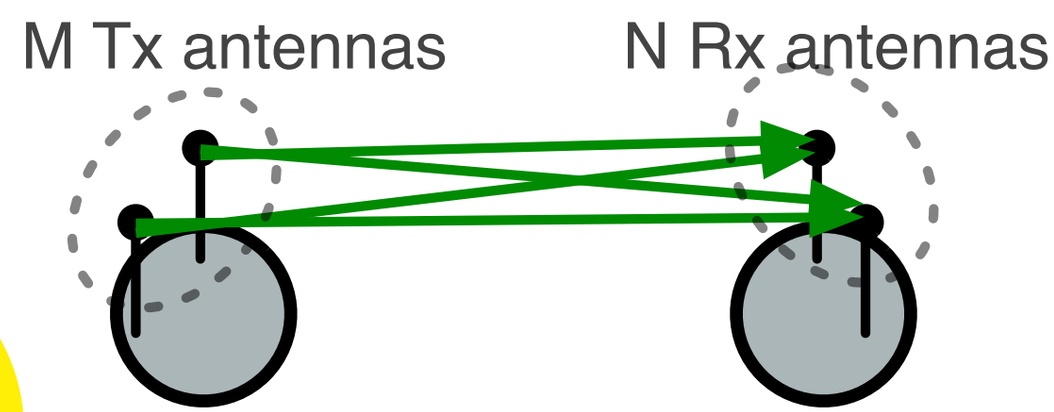
Fundamental

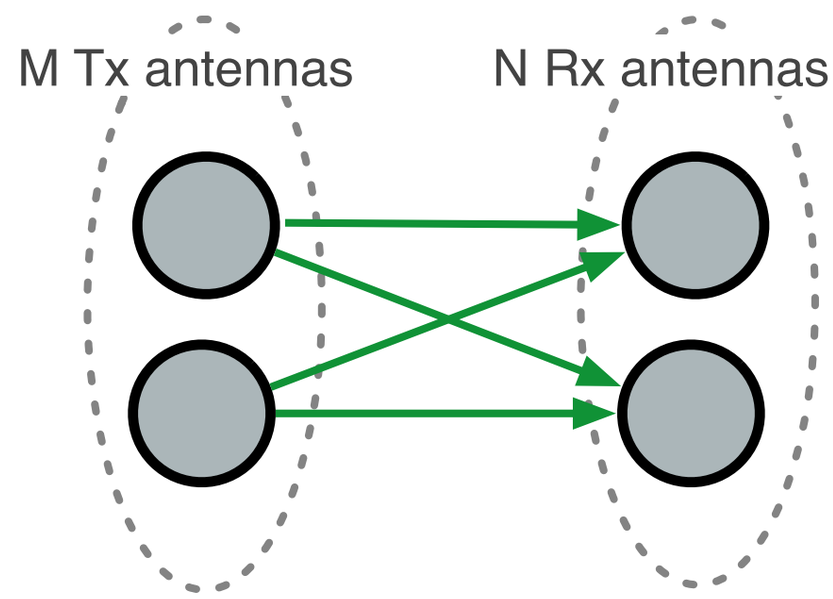
NO! Motivation 2 - networks



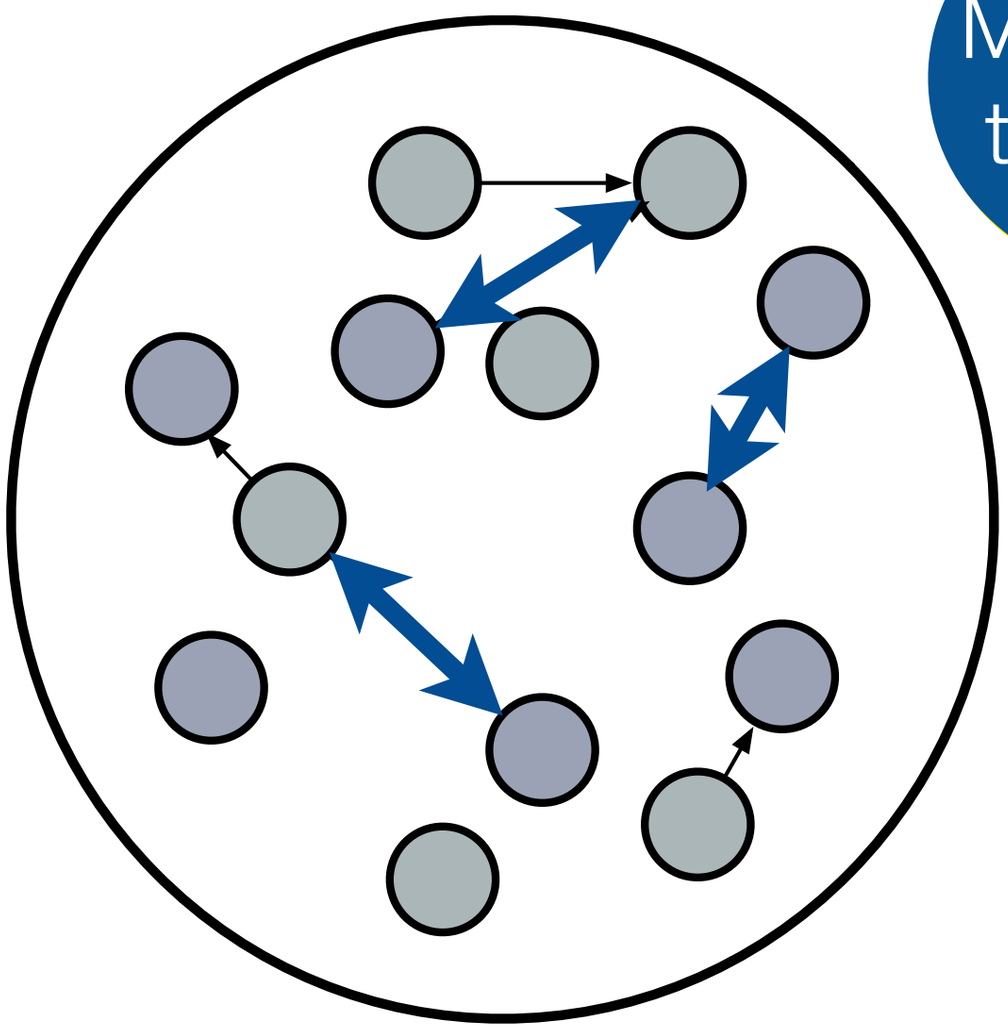




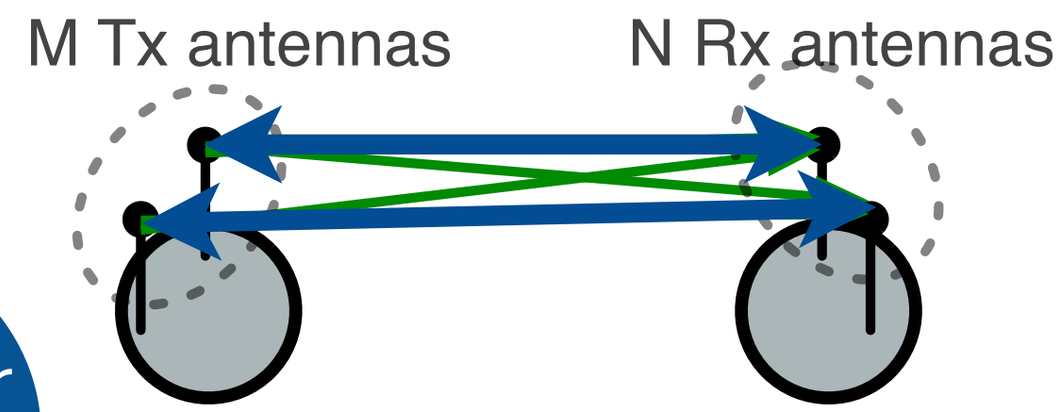


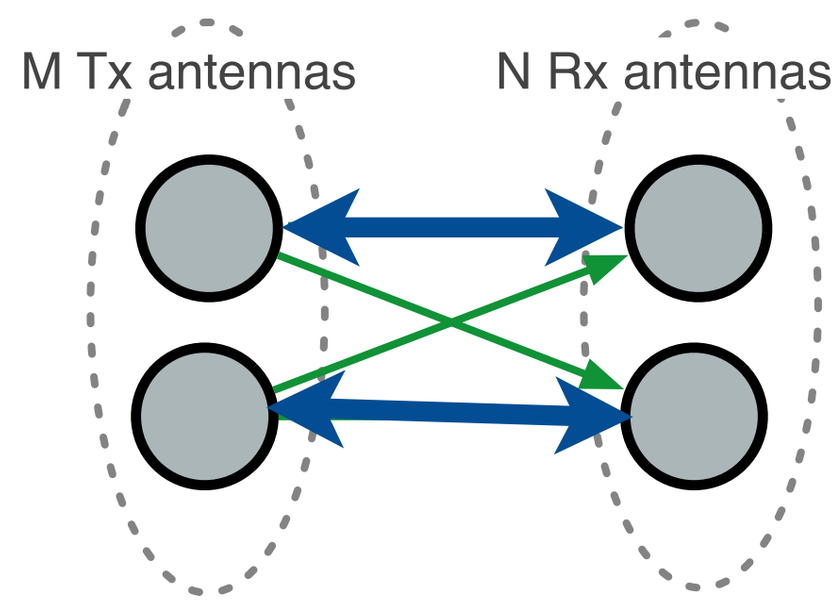


Two-way networks!



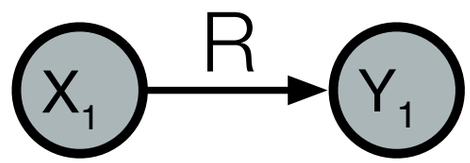
Multi-user two-way

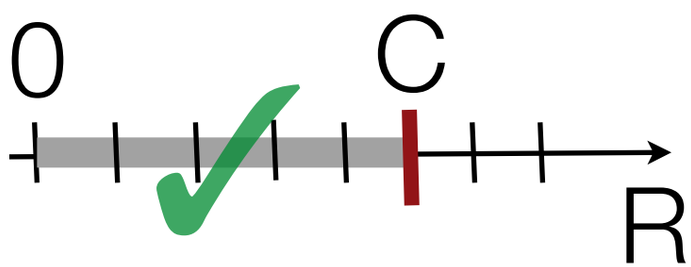


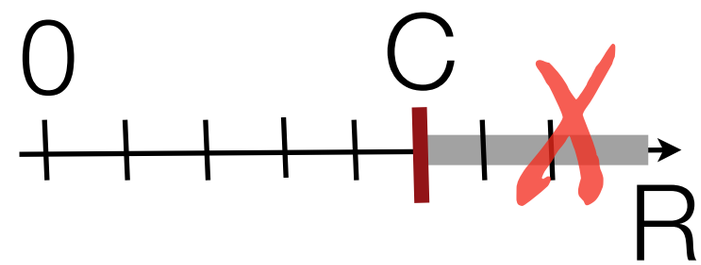


Capacity and capacity regions

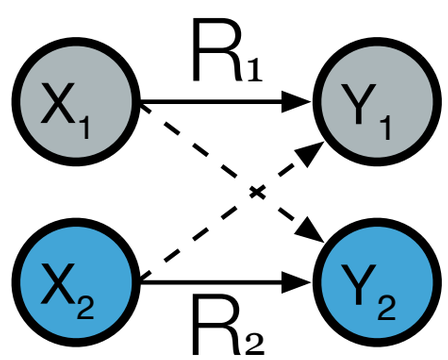
- Point to point **capacity**

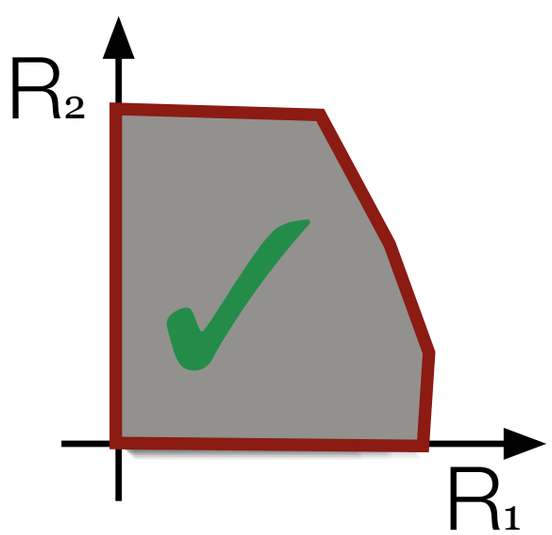


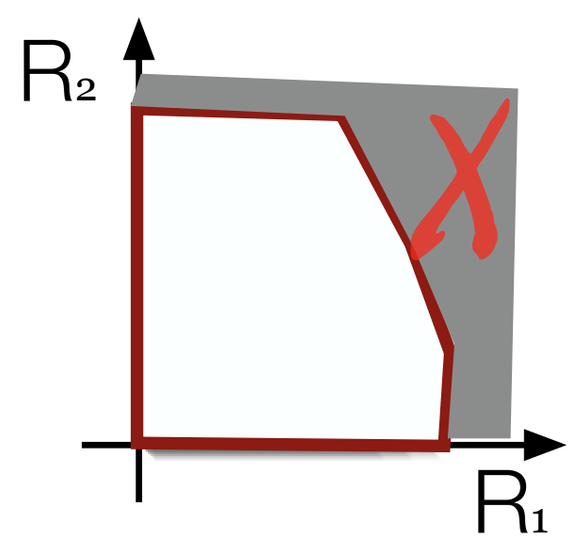




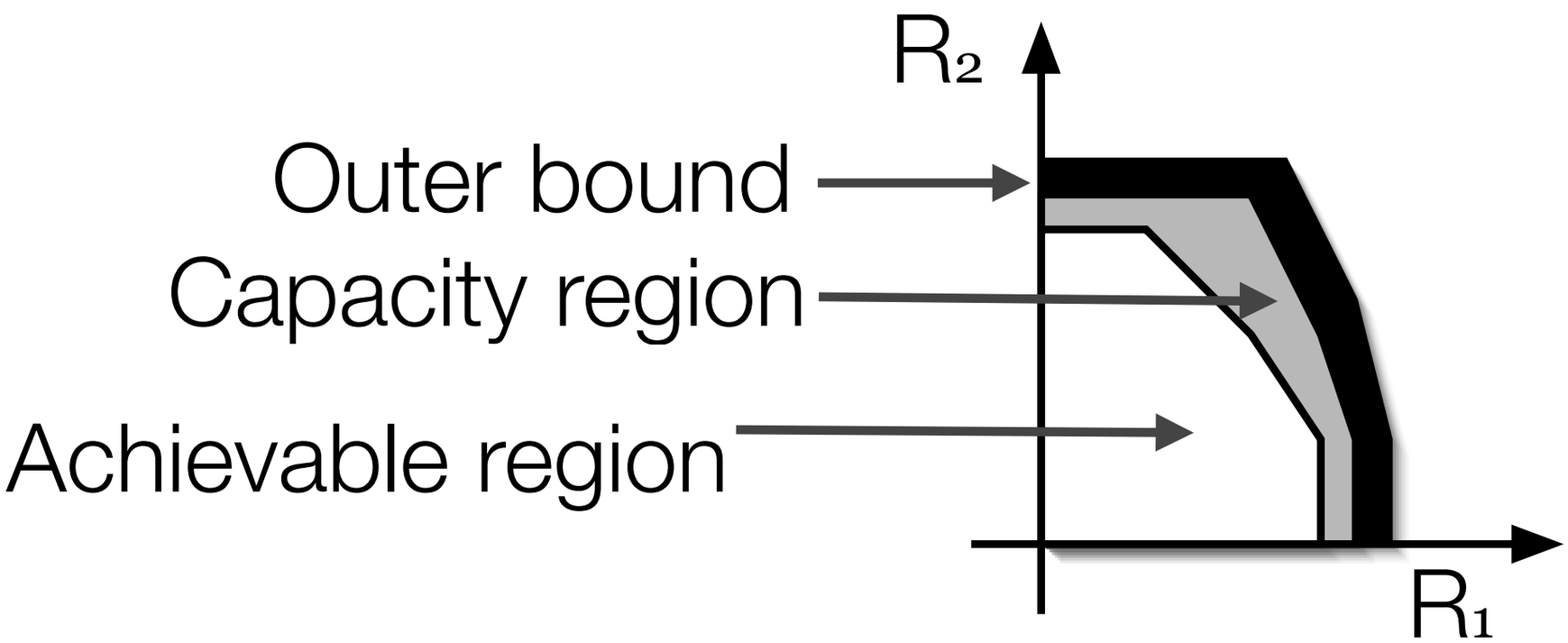
- Multi-user **capacity region**

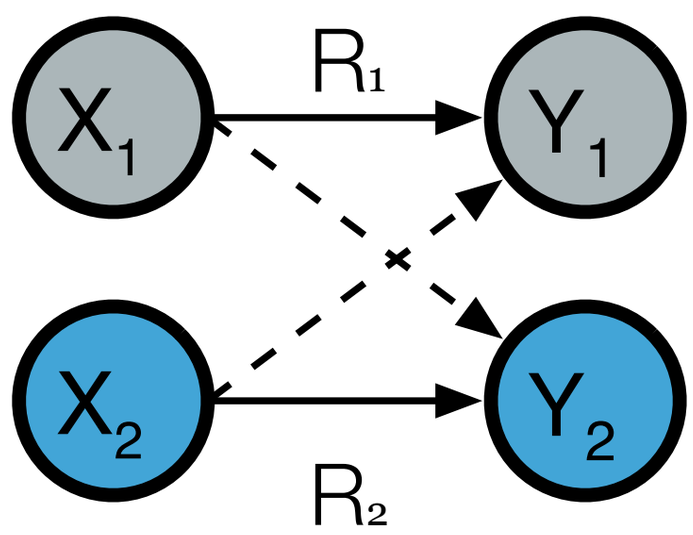




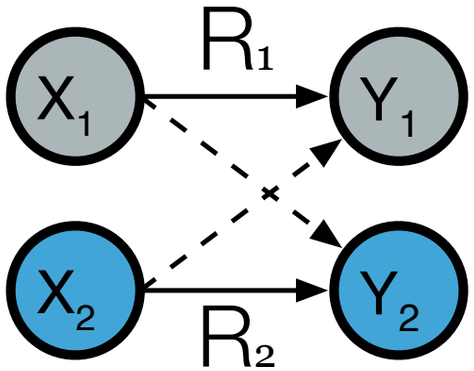


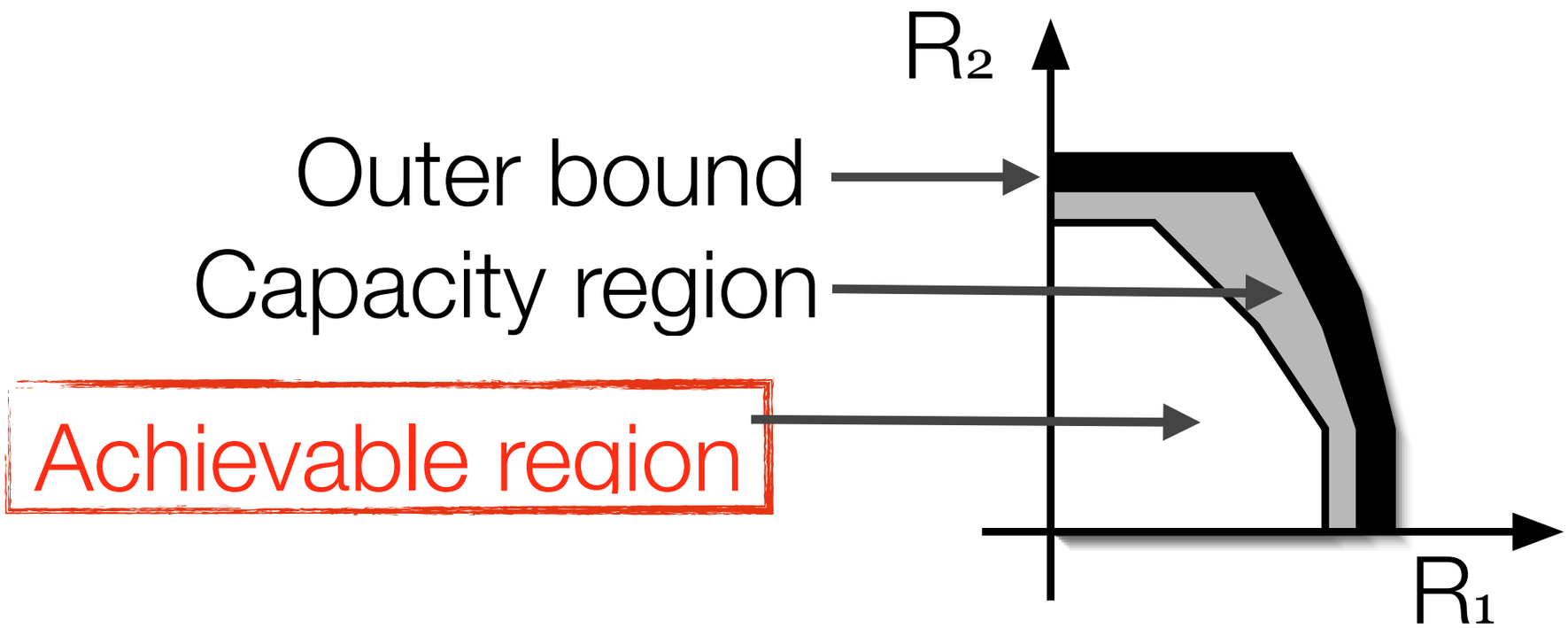
Capacity regions





Achievable rate region

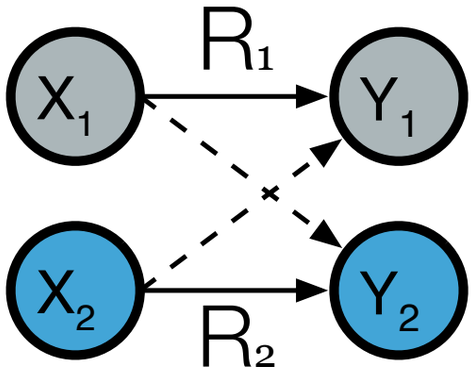


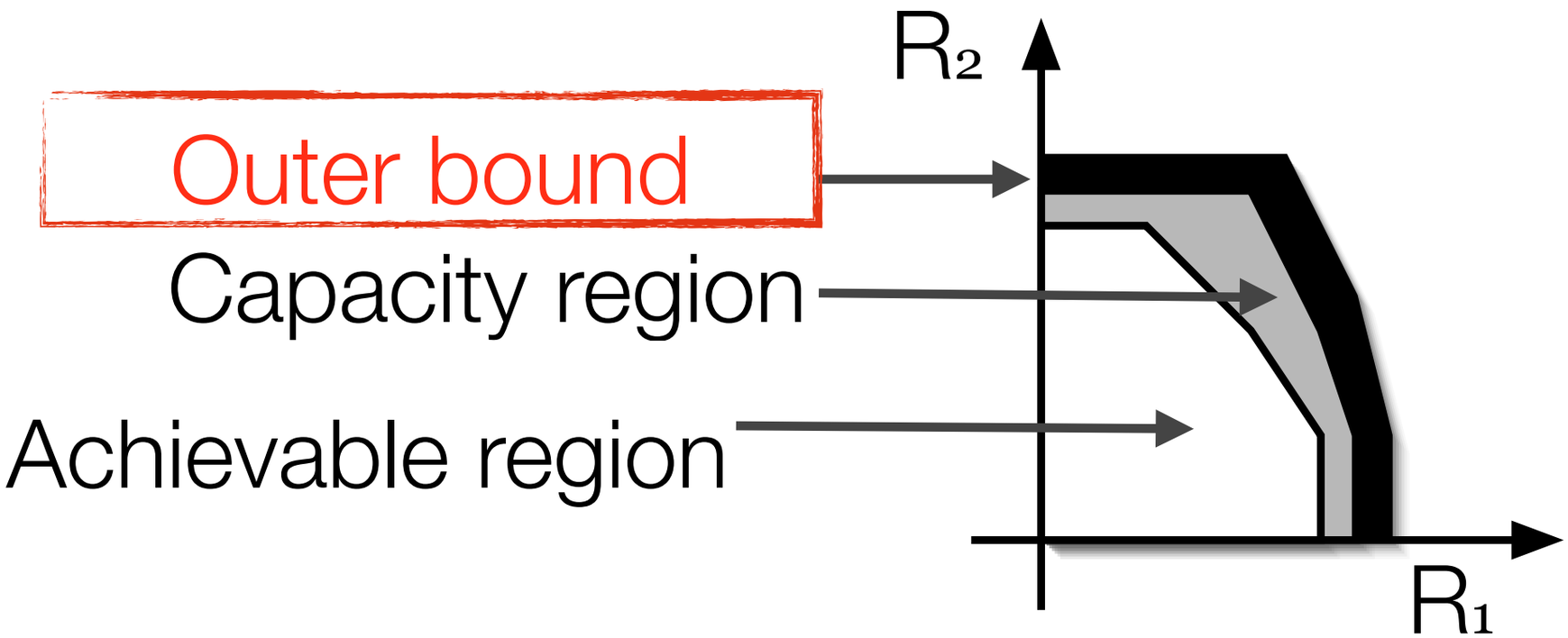


- Propose a coding scheme (random codes!)
- Prove that as long as ⇒ holds, reliable communication possible

$$R_1 \leq I(X_1; Y|X_2)$$
$$R_2 \leq I(X_2; Y|X_1)$$
$$R_1 + R_2 \leq I(X_1, X_2; Y)$$

Outer bound

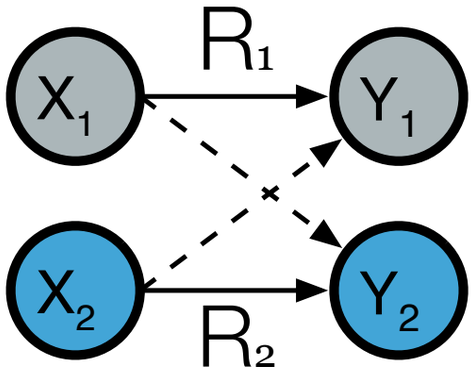


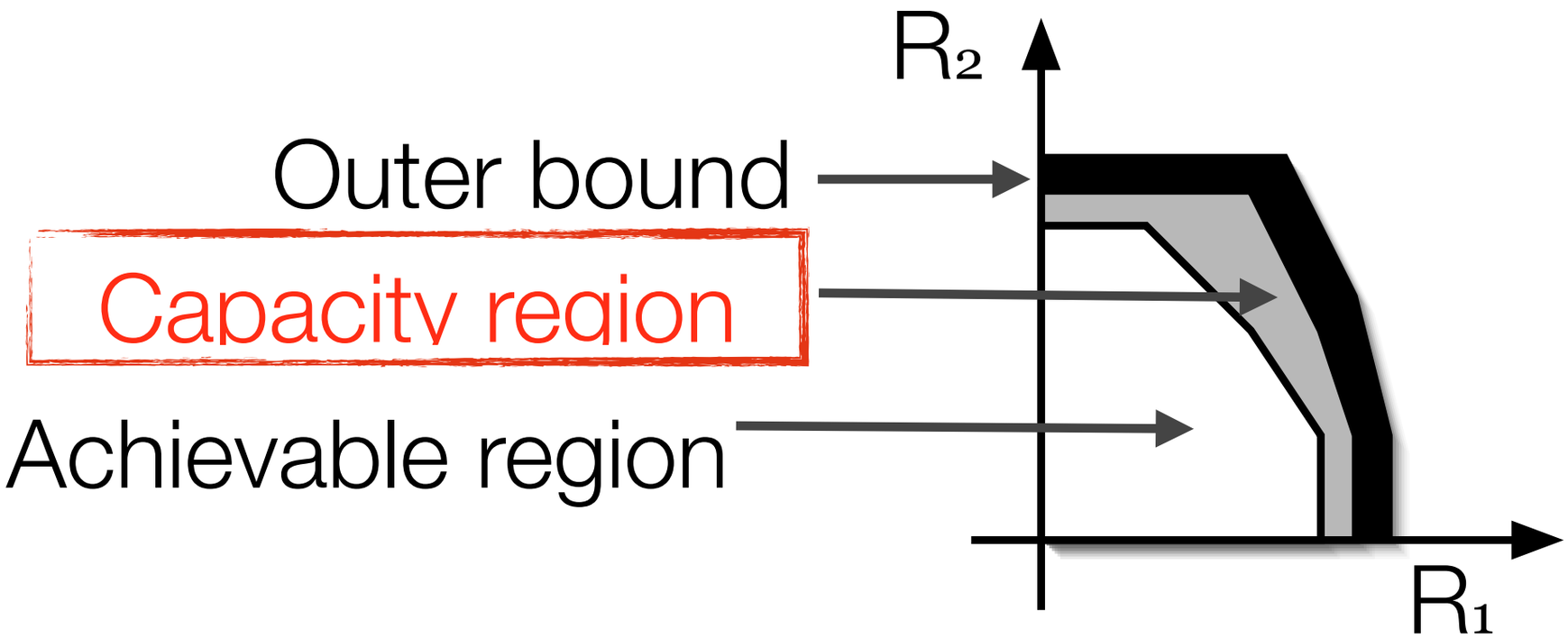


$$R_1 \leq I(X_1; Y|X_2)$$
$$R_2 \leq I(X_2; Y|X_1)$$
$$R_1 + R_2 \leq I(X_1, X_2; Y)$$

- Prove that error is bounded away from 0 when $\Uparrow$ not satisfied
- Find a more capable channel whose capacity is known
- Be creative!

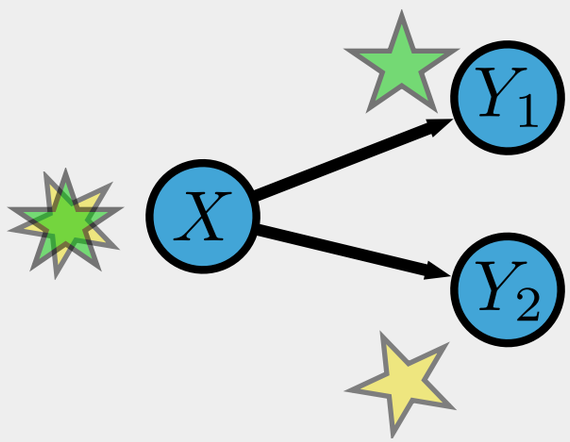
Capacity regions



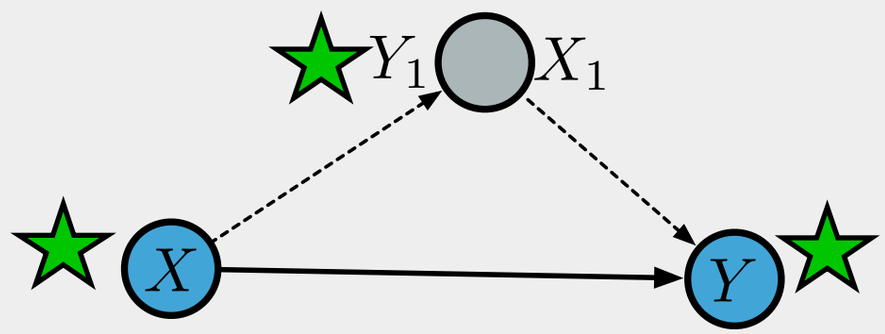


- Limit of communication, NOT how to achieve it in practice necessarily!
- However, benchmark and guidance in practical designs

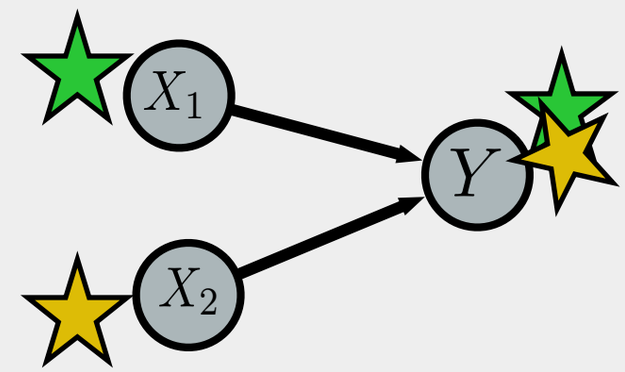
Three key multi-user channels



Broadcast channel



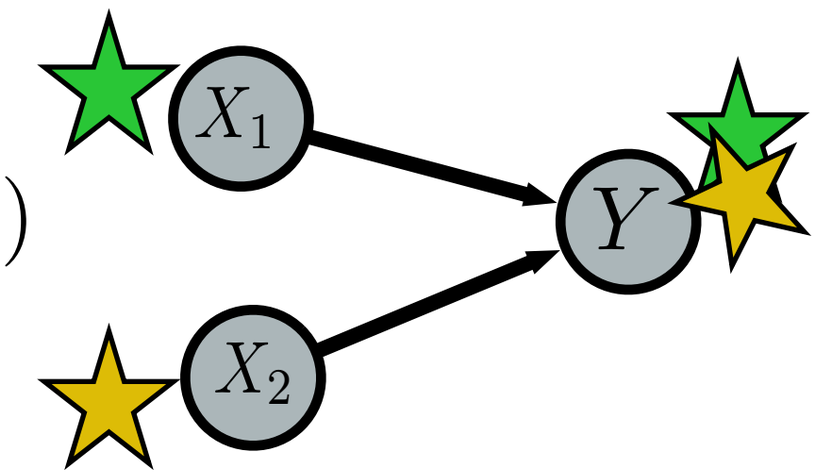
Relay channel



Multiple-access channel

Multiple-access channel (MAC)

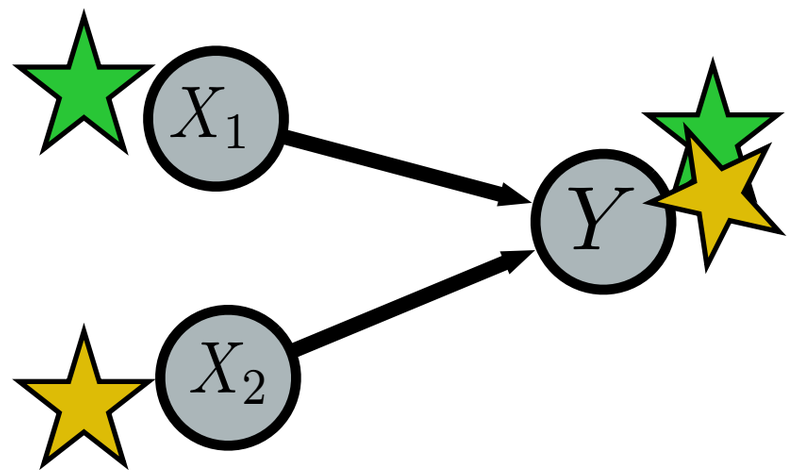
$$p(x_1, x_2) = p(x_1)p(x_2)$$



$$p(y|x_1, x_2)$$

- Introduced by Shannon in 1961
- Capacity known for discrete and Gaussian noise channels
    - Capacity [Ahlswede ’71, Liao ’72]
    - MIMO [Telatar ’99]
    - Fading [Gallager ’94, Shamai+Wyner ’97, Tse+Hanly ’98]

Multiple-access channel (MAC)

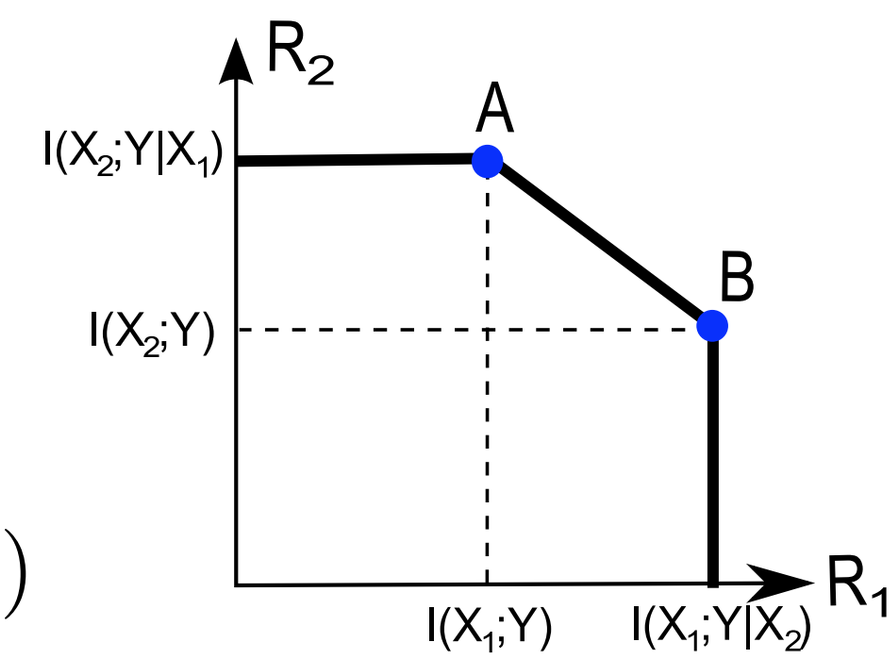


- Capacity region is the closure of the convex hull of all rate pairs (R1, R2) satisfying

$$R_1 \leq I(X_1; Y|X_2)$$
$$R_2 \leq I(X_2; Y|X_1)$$
$$R_1 + R_2 \leq I(X_1, X_2; Y)$$

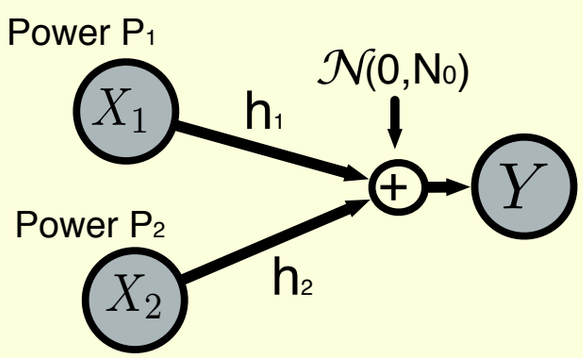
for some distribution

$$p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2)$$



Gaussian MAC

Optimal multiple access



$$R_1 < \frac{1}{2}\log_2(1+P_1)$$

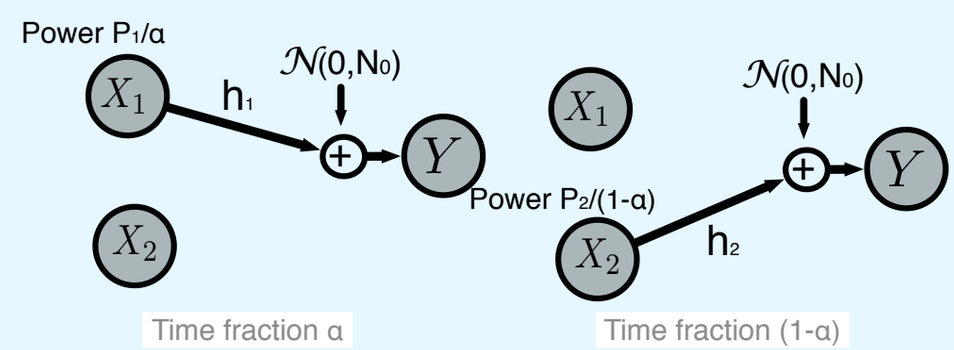
$$R_2 < \frac{1}{2}\log_2(1+P_2)$$

$$R_1 + R_2 < \frac{1}{2}\log_2(1+P_1+P_2)$$

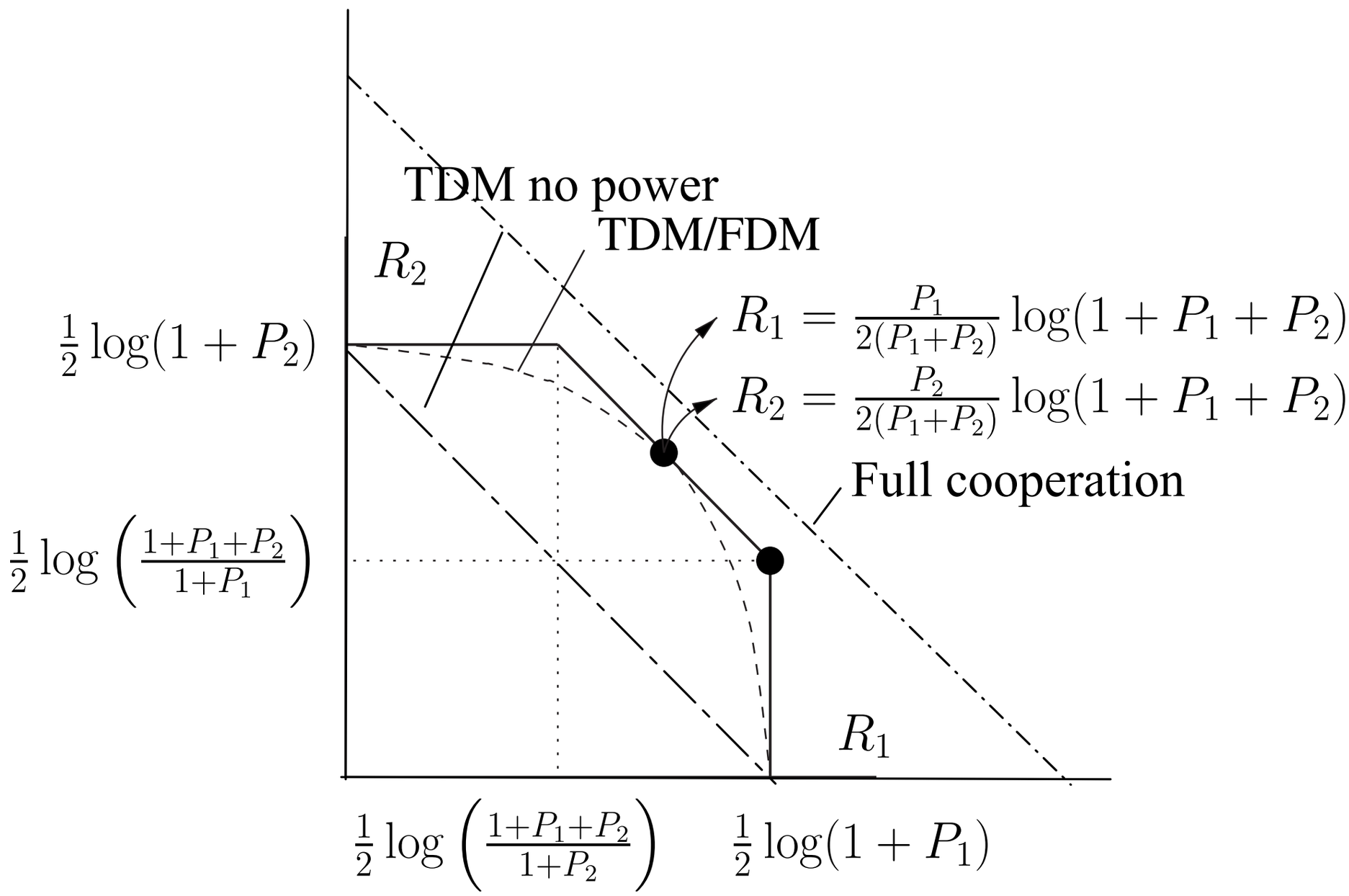
TDM/FDM multiple access

$$R_1 = \frac{\alpha}{2}\log\left(1+\frac{P_1}{\alpha}\right) \qquad R_1 \leq \frac{\alpha}{2}\log_2(1+P_1)$$

$$R_2 = \frac{1-\alpha}{2}\log\left(1+\frac{P_2}{1-\alpha}\right) \qquad R_2 \leq \frac{1-\alpha}{2}\log_2(1+P_2)$$

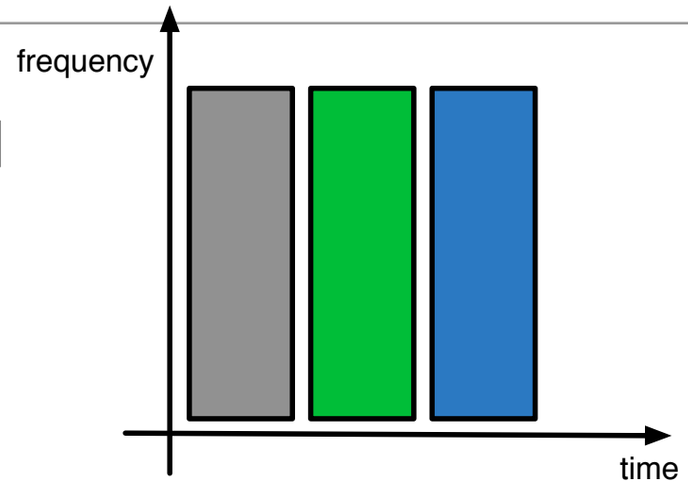


Multiple access channel

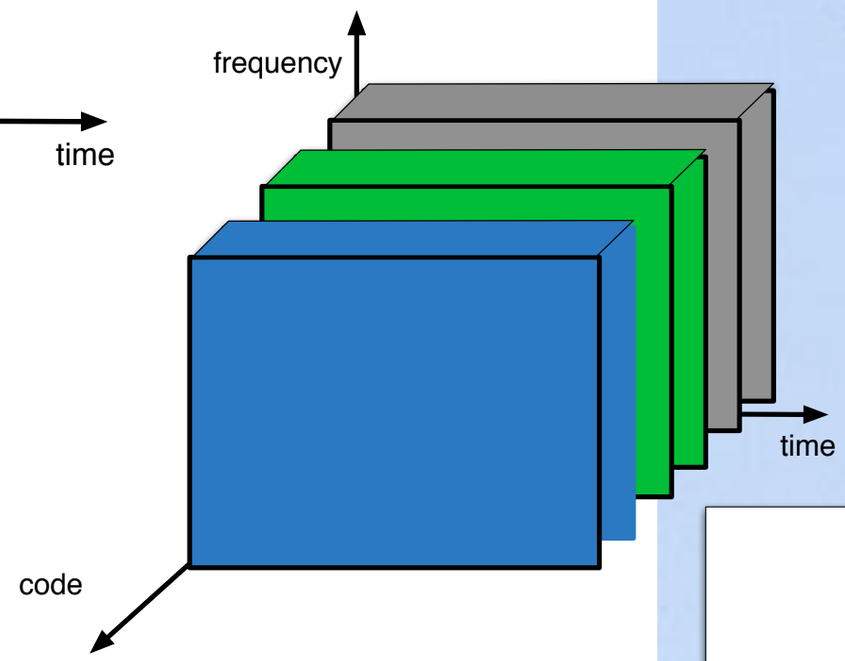


Multiple access channels in practice

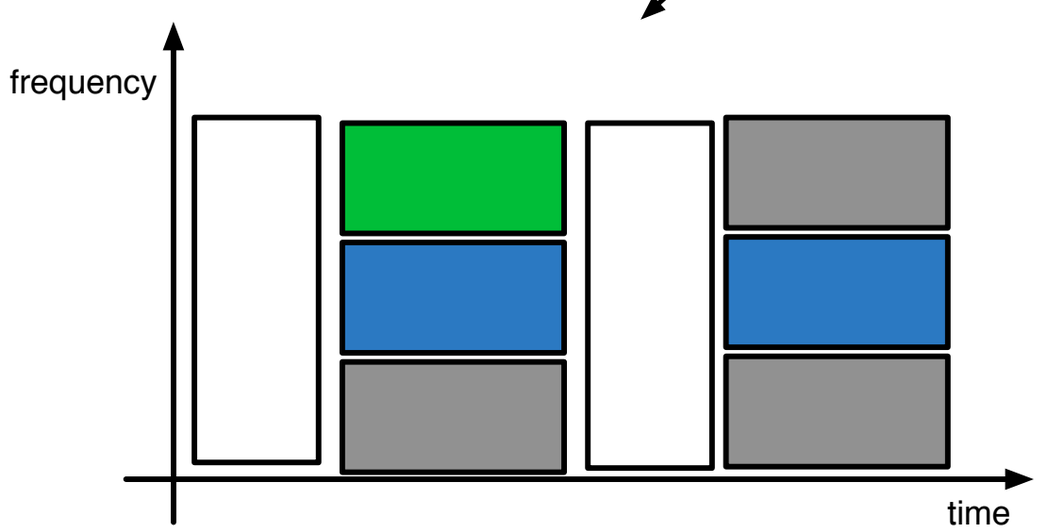
- GSM

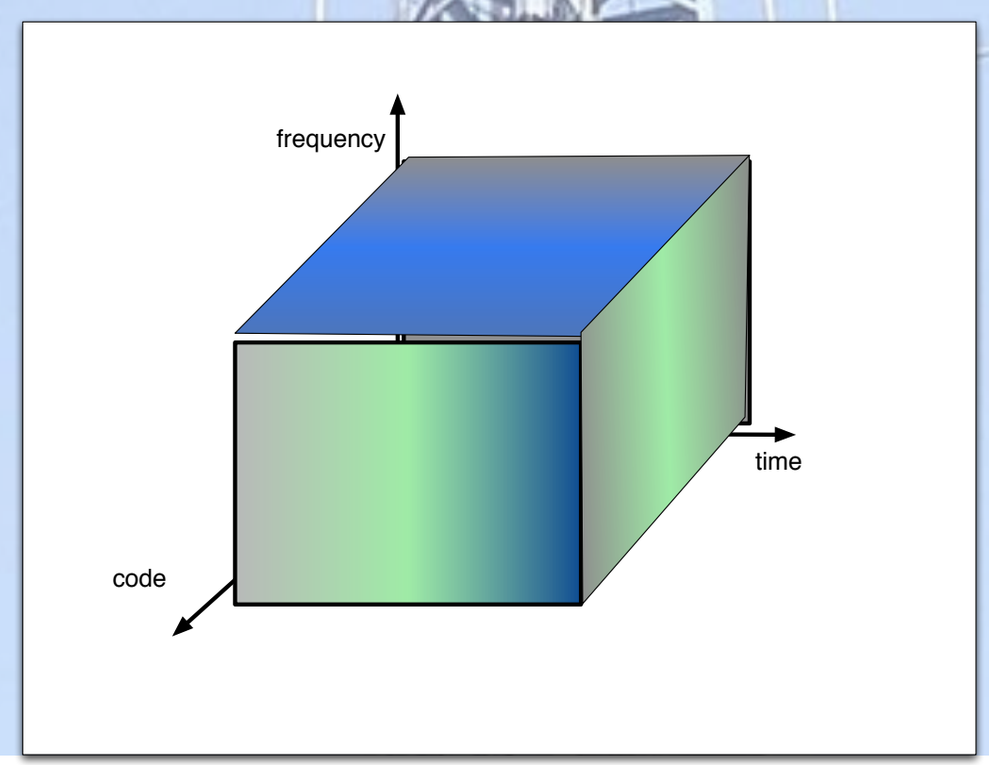


- CDMA

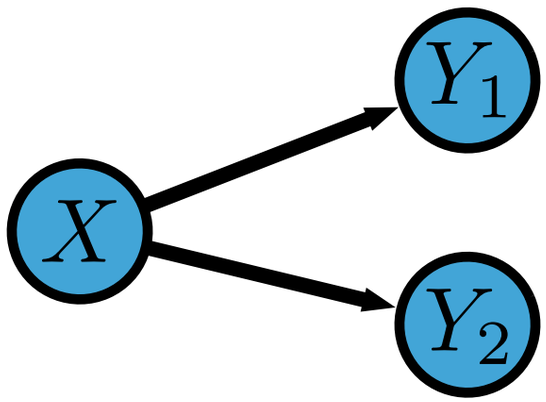


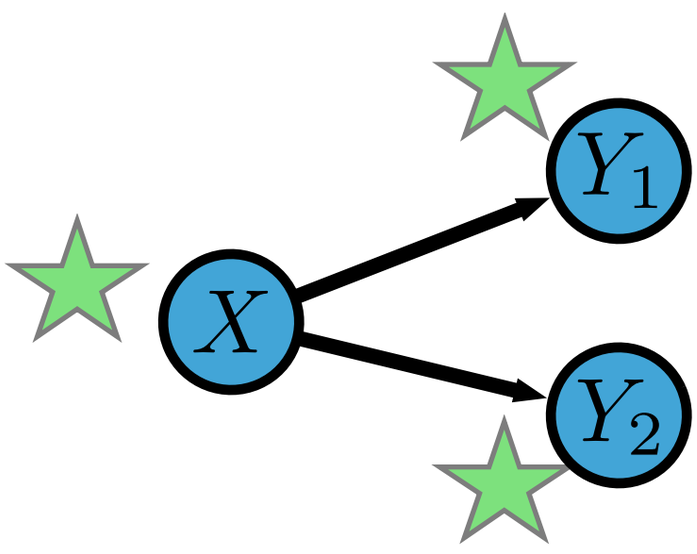
- WiFi

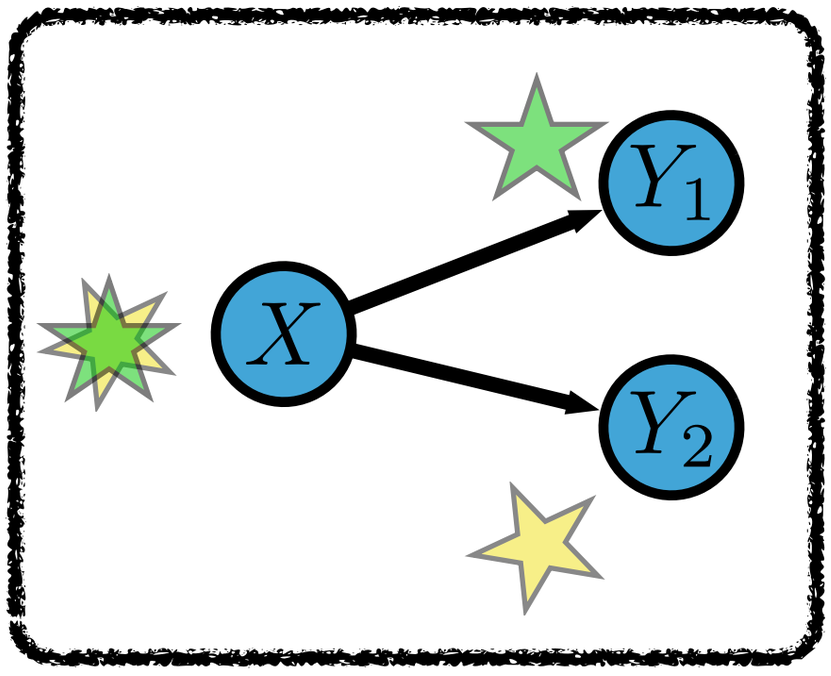


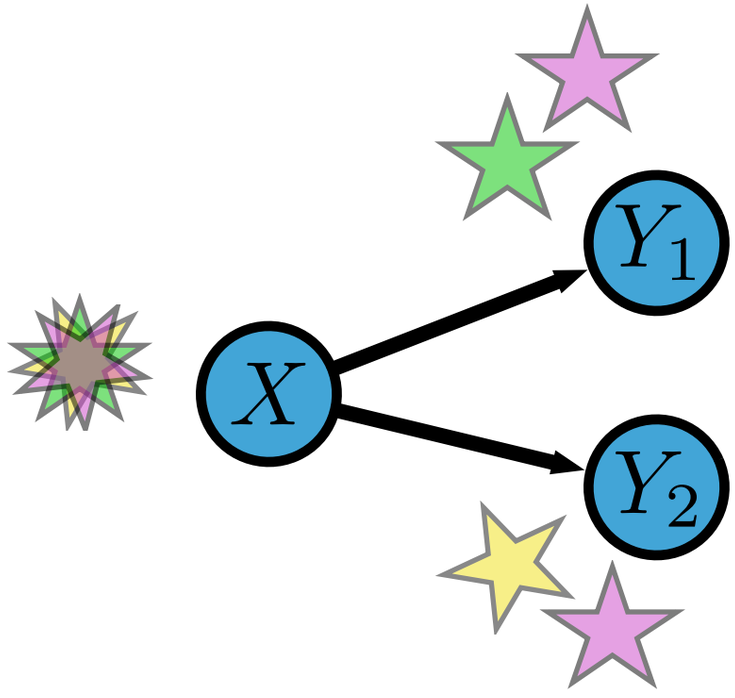


Broadcast channel

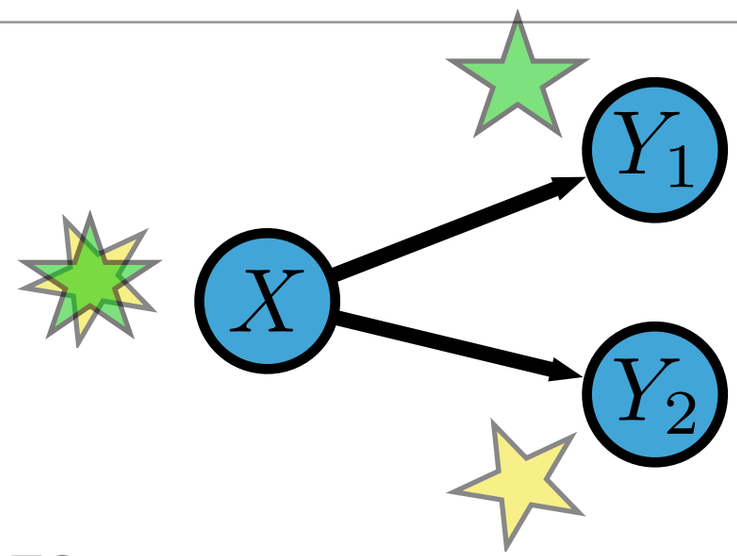








Broadcast channel



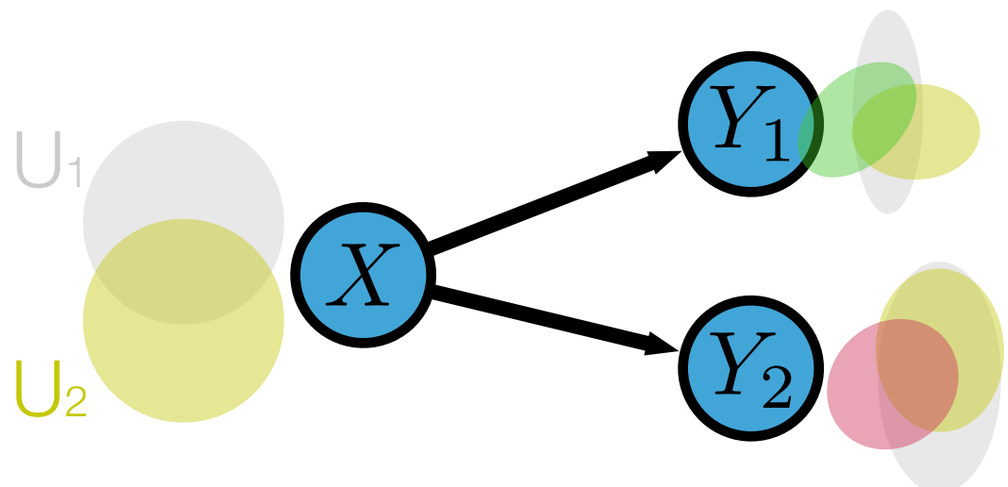
$$p(y_1, y_2|x)$$

- Introduced by Cover in 1972

- Capacity known special cases:
  - Degraded broadcast channels [Bergmans ’73’74, Gallager ’74]
  - General BC with degraded message sets [Korner + Marton ’77]
  - Gaussian MIMO broadcast channel [Weingarten, Steinberg, Shamai ’06]

- Best achievable rate region [Marton ’79]

Best achievable rate region: Marton's region



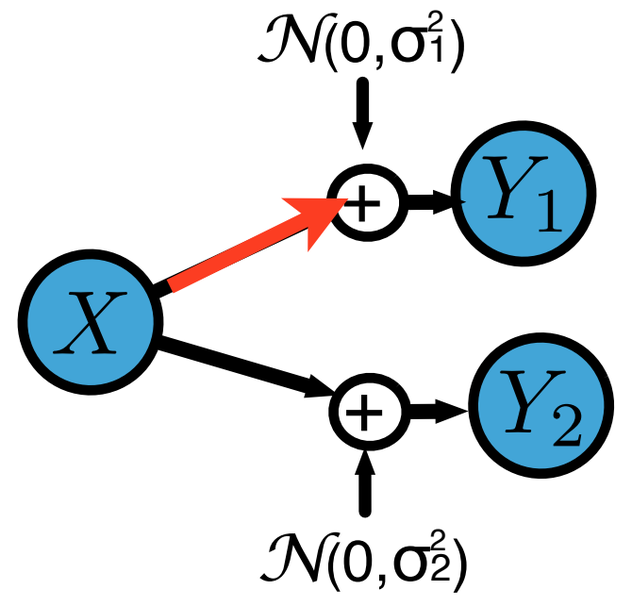
- Capacity region is the closure of the convex hull of all rate pairs (R_1, R_2) satisfying

$$
\begin{aligned}
&0 \leq R_1 \leq I(U_1;Y_1) \\
&0 \leq R_2 \leq I(U_2;Y_2) \\
&R_1 + R_2 \leq I(U_1;Y_1) + I(U_2;Y_2) - I(U_1;U_2)
\end{aligned}
$$

for some distribution

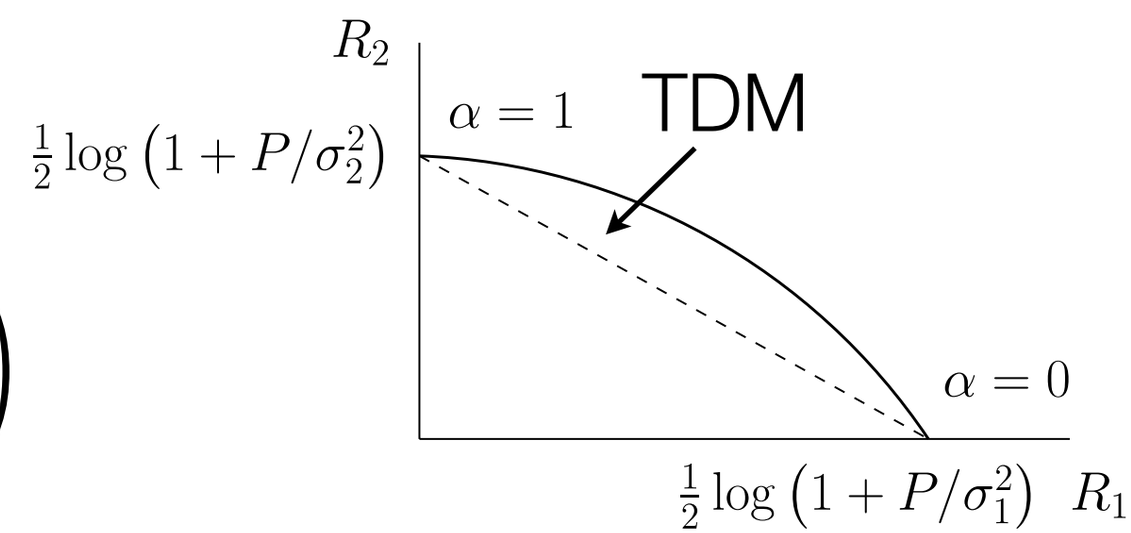
$$p(u_1, u_2, x, y_1, y_2) = p(u_1, u_2)p(x|u_1, u_2)p(y_1, y_2|x)$$

Gaussian broadcast channel

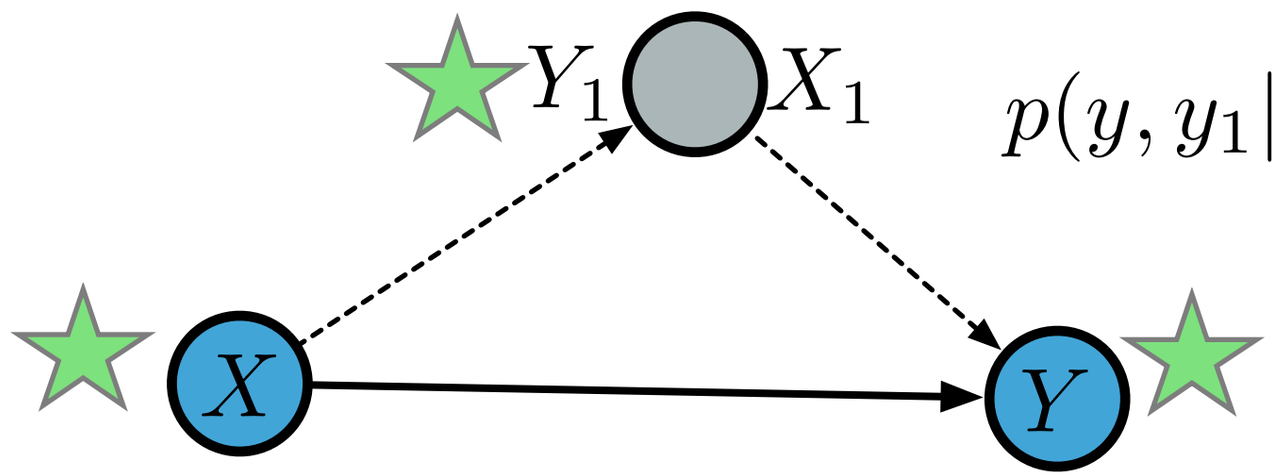


$$R_1 < \frac{1}{2}\log_2\left(1 + \frac{(1-\alpha)P}{\sigma_1^2}\right)$$

$$R_2 < \frac{1}{2}\log_2\left(1 + \frac{\alpha P}{(1-\alpha)P + \sigma_2^2}\right)$$



Relay channel



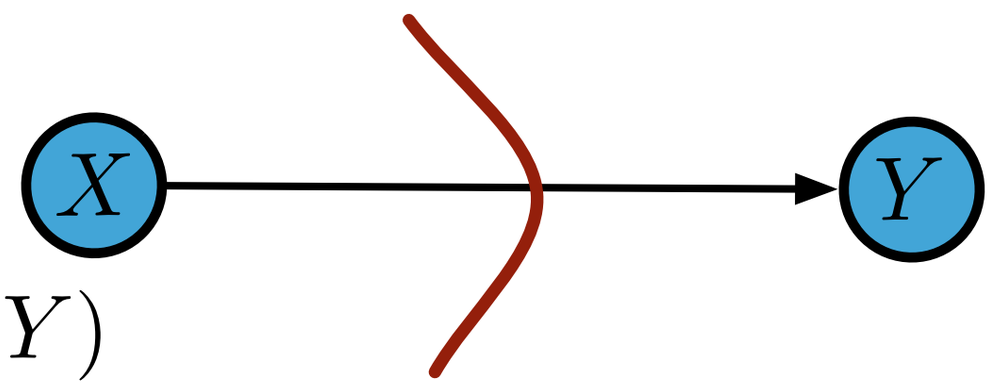
$$p(y, y_1 | x, x_1)$$

- Introduced by Van der Meulen in 1968
- Capacity known special cases:
  - Physically degraded relay channels [Cover, El Gamal ’79]
- 3 types of forwarding:
  - Decode+forward (DF), Compress+forward (CF), Amplify+forward (AF)

Relay channel: outer bounds

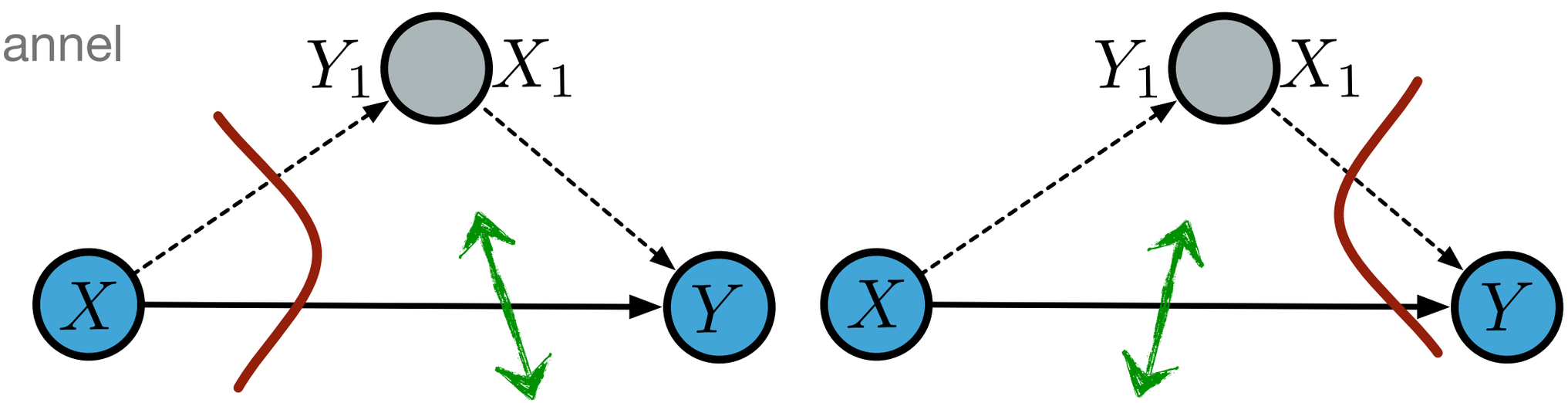
- Cut set outer bound intuitive + useful!

- Point-to-point



$$R_{out} = \cup_{p(x)} I(X;Y)$$

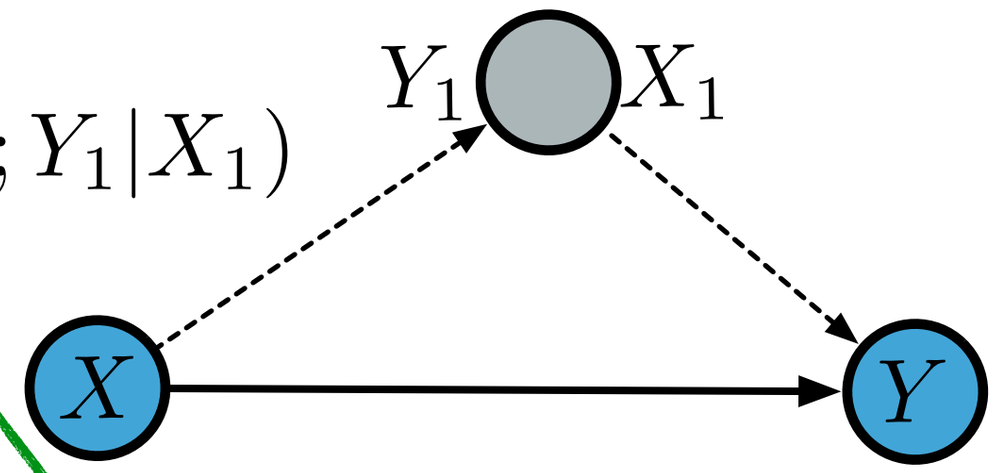
- Relay channel



$$R_{out} = \cup_{p(x,x_1)} \min\{I(X;Y_1,Y|X_1), I(X,X_1;Y)\}$$

Relay channel: Decode + forward (DF)

$$R \leq I(X; Y_1|X_1)$$



$$R \leq I(X_1; Y) + I(X; Y|X_1) = I(X, X_1; Y)$$

$$R \leq \max_{p(x,x_1)} \min\{I(X; Y_1|X_1), I(X, X_1; Y)\}$$

- Exploit broadcast transmission of source
- Source + relay transmit simultaneously (full duplex)
- Create joint codebooks at source + relay