Communicating dialogues

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

Monologue

\[ W_{12} \quad 1 \quad 2 \quad \overline{W}_{12} \]

Dialogue

\[ W_{12} \quad \overline{W}_{21} \quad 1 \quad 2 \quad \overline{W}_{12} \quad W_{21} \]
Monologues vs. Dialogues

Monologue = one-way = uni-directional

Dialogue = two-way = bi-directional
Two-way communication applications - wired

Video conferencing

Telesurgery

Data synchronization

Necessary Elements of a Telerobotic Surgery Team

In its current state, the pieces that make up telesurgery are complex, expensive, and relatively fragmented. To success...
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*

Overall - much is still unknown
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

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

\[
x_1^{(1)}(w_1) \quad x_2^{(1)}(w_2)
\]

\[
y_1^{(1)} \quad y_2^{(1)}
\]

\[
w_r = w_1 \oplus w_2
\]
Information theory - what, why, when


What is information? How fast can we communicate? How much can we compress information?
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 $\iff H < C$
Source vs. channel coding

- **Source**
  - Encoder
  - Channel coder
  - Channel decoder
  - Channel

- **Noise**
  - Decoder
  - Source decoder
  - Destination

- **Destination**

**Key Processes**:
- **Remove redundancy**
- **Controlled adding of redundancy**
- **Decode signals**: detect/correct errors
- **Restore source**
Source coding = data compression

Source: $X \sim p(x)$ i.i.d

Encoder: $X^n \xrightarrow{f_n(X^n) \in \{1, 2, \ldots, 2^{nR}\}} \hat{X}^n$

Rate $R$ bits/source symbol

Decoder: $\hat{X}^n \rightarrow \hat{X}$

Destination

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

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

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

  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. $d(\hat{X}^n, X^n) \leq D$
Source vs. channel coding

- **Source**
- **Encoder**
- **Channel**
- **Decoder**
- **Destination**

**Noise**

**Source vs. channel coding**

- **Encoder**
- **Source coder**
- **Channel coder**
- **Source decoder**
- **Channel decoder**

**Noise**

**Decoder**

- **Channel**
- **Destination**

**Remove redundancy**

**Controlled adding of redundancy**

**Decode signals, detect/correct errors**

**Restore source**
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

Source
Channel capacity: a cute example

Source $\xrightarrow{A,B,C,D}$ Encoder

A = A?
A = AAA?
Channel capacity: a cute example

Source \[\rightarrow\] Encoder \[\rightarrow\] Channel

\[\text{AAA} \rightarrow \text{AB}.\]
Channel capacity: a cute example

How to communicate reliably?
Channel capacity: a cute example

Use these 9 symbols!

\[ C = \log_2(9) \]
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!
Mathematical description of capacity

• Can achieve reliable communication for all transmission rates $R$:

  \[ R < C \]

  \[ \checkmark \]

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

  \[ R > C \]

  \[ \times \]
Capacity: key ideas

- choose input set of codewords so they are “non-confusables” at the output
- number of these that we can choose 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

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

“mutual information” between X and Y

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

\[
H(X) = -\frac{1}{2} \log_2 \left( \frac{1}{2} \right) - \frac{1}{4} \log_2 \left( \frac{1}{4} \right) - \frac{1}{8} \log_2 \left( \frac{1}{8} \right) - \frac{1}{16} \log_2 \left( \frac{1}{16} \right) - 4 \frac{1}{64} \log_2 \left( \frac{1}{64} \right)
= 2 \text{ (bits)}
\]

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

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

\[
(000,001,010,011,100,101,110,111) \quad (0,10,110,1110,111100,111101,111110,111111)
\]

3 bits

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:

\[ h(\mathcal{N}(0, \sigma^2)) = \frac{1}{2} \log (2\pi e \sigma^2) \]

\[ h(\mathcal{N}_n(\mu, K)) = \frac{1}{2} \log ((2\pi e)^n |K|) \]
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:

\[
I(X; Y) = \sum p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \\
= H(X) - H(X|Y) \\
= H(Y) - H(Y|X)
\]
Mutual information between 2 random variables:

\[ I(X; Y) = \sum p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]

\[ = H(X) - H(X|Y) \]

\[ = H(Y) - H(Y|X) \]

(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:

\[ I(X;Y) = \sum p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]

\[ = H(X) - H(X|Y) \]

\[ = H(Y) - H(Y|X) \]

(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:

\[ I(X;Y) = \sum p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]

\[ = H(X) - H(X|Y) \]

\[ = H(Y) - H(Y|X) \]

(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:

\[
I(X;Y) = \sum p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right)
\]

\[
= H(X) - H(X|Y)
\]

\[
= H(Y) - H(Y|X)
\]

(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

\[ H(X|Y) \quad I(X;Y) \quad H(Y|X) \]

\[ H(X) \quad H(Y) \quad H(X,Y) \]

\[ H(Y|X,Z) \quad H(X) \quad H(Z) \]

\[ I(X;Y|Z) \quad A \quad I(X;Y|Z) \]

\[ H(Z|X) \quad H(Z|Y) \quad H(Z|X,Y) \]

With permission from David J.C. Mackay
Channel capacity

Capacity

\[ C = \max_{p(x)} I(X; Y) \quad \text{bits/channel use} \]

\[ I(X; Y) = \sum_{x,y} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \]
Channel capacity

Capacity?

1 bit/channel use

\[
C = \max_{p(x)} I(X; Y)
= \max_{p(x)} \left( H(X) - H(X|Y) \right)
= \max_{p(x)} H(X) - 0
= 1
\]

Noiseless channel

0 → 0

1 → 1
Channel capacity

Capacity?

1 bit/channel use

Non-overlapping outputs

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

\[
= \max_{p(x)} H(X) - H(X|Y)
\]

\[
= \max_{p(x)} H(X) - 0
\]

\[
= 1
\]
Channel capacity

Capacity?

$\log_2(9)$ bits/channel use

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

$$= \max_{p(x)} H(X) - H(X|Y)$$

$$= \max_{p(x)} H(X) - \log_2(3)$$

$$= \log_2(27) - \log_2(3) = \log_2(9)$$
Channel capacity

Binary erasure channel

Capacity?

1-f bits/channel use

\[ C = \max_{p(x)} I(X; Y) \]
\[ = \max_{p(x)} H(X) - H(X|Y) \]
\[ = \max_{p(x)} H(X) - \sum p(y) H(X|Y = y) \]
\[ = \max_{p(x)} H(X) - p \cdot (1 - f) \cdot H(X|Y = 0) + (1 - p) \cdot (1 - f) \cdot H(X|Y = 1) + p \cdot f \cdot H(X|Y = e) \]
\[ = \max_{p(x)} H(X) - p \cdot (1 - f) \cdot 0 + (1 - p) \cdot (1 - f) \cdot 0 + f \cdot 1 \]
\[ = \max_{p(x)} H(X) - f \]
\[ = 1 - f \]
Discrete memoryless channel capacity

Channel: \( p(y|x) \)

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

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

Power constrained to $P$

Wireless channel with fading

$X$ \xrightarrow{h} Y$

$N \sim N(0, P_N)$

$Y = hX + N$
AWGN channel capacity

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

\[
C = \max_{p(x):E[XX^T] \leq P} I(X;Y)
\]

\[
= \max_{p(x):E[XX^T] \leq P} h(X) - h(X|Y)
\]

\[
= \max_{p(x):E[XX^T] \leq P} h(Y) - h(Y|X)
\]

\[
= \max_{p(x):E[XX^T] \leq P} h(hX + N) - h(hX + N|X)
\]

\[
= \max_{p(x):E[XX^T] \leq P} h(hX + N) - h(N)
\]

\[
= \frac{1}{2} \log(2\pi e(|h|^2 P + P_N)) - \frac{1}{2} \log(2\pi e P_N)
\]

\[
= \frac{1}{2} \log \left( \frac{|h|^2 P + P_N}{P_N} \right)
\]
AWGN channel capacity

\[ C = \frac{1}{2} \log \left( \frac{|h|^2 P + P_N}{P_N} \right) \]

\[ = \frac{1}{2} \log (1 + SNR) \] (bits/channel use)

What about bits/second and bandwidth of the channel?

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

[Bandwidth W, h=1, spectral density N_0/2]
Source vs. channel coding

- **Source**: The original data source
- **Encoder**: Converts the source data into a form suitable for transmission
- **Channel**: The medium through which the data is transmitted
- **Decoder**: Recovers the original data from the transmitted signal
- **Destination**: The intended recipient of the data

**Source**
- **Source coder**: Adds redundancy to the data
- **Channel coder**: Protects the data from channel errors
- **Channel decoder**: Corrects errors introduced by the channel
- **Source decoder**: Removes the redundancy added by the source coder
- **Destination**: Receives the corrected data

**Noise**
- The channel introduces noise, which the decoder attempts to mitigate.
- The source coder adds redundancy to help in error correction.
- The channel coder protects against channel errors.

**Key Actions**
- **Remove redundancy**: Reduces unnecessary data.
- **Controlled adding of redundancy**: Adds just enough extra data to correct errors.
- **Decode signals, detect/correct errors**: Ensures accurate data transmission.
- **Restore source**: Recovers the original data from the encoded and decoded signals.
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.

- Robert J. McEliece, Allerton 2000

• algebraic codes
• convolutional codes
• iterative codes (LDPC, turbo)
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*
  
  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*
One-way channel capacity

\[ C = \max_{p(x)} I(X; Y) \text{ bits/channel use} \]

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

“mutual information” between X and Y

symmetric in its arguments!
One-way channel capacity - notation

$X^n \xrightarrow{p(y|x)} Y^n$

$p(y|x)$

$X \xrightarrow{p(y|x)} Y$

$R$

$1 \xrightarrow{p(y|x)} 2$

$R_{12}$
Two-way channel (historical)

Figure 1
Two-way channel capacity

One-way

Two-way

Before making these notions precise, we give some simple examples. In Figure 2 the two-way channel decomposes into two independent one-way noiseless binary
Two-way channel capacity region

One-way Capacity

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

Two-way Capacity Region
When is $\text{Two-way}$ equal to $\text{Two one-ways}$?
Models for two-way adaptation

One-way: no adaptation possible

Two-way: no adaptation

“Restricted two-way channel”

Two-way: full adaptation
Duplex

**Two-way:**
*half duplex*

![Diagram showing half duplex communication](chart)

**Two-way:**
*full duplex*

![Diagram showing full duplex communication](chart)
(Aside: $I(X;Y)$ and $I(X;Y|Z)$)

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

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

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

\[
\begin{align*}
I(X;Y|Z) &= H(Y|Z) - H(Y|X, Z) \\
&= \log_2(2) - 0 \\
&= 1 \text{ bit}
\end{align*}
\]
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 \( p(x_1, x_2) = p(x_1)p(x_2) \).

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 \( p(x_1, x_2) \).

Not in general equal!
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

\[ R_{12} \leq C_{K_1} \]
\[ R_{21} \leq C_{K_2} \]
Capacity: binary mod-2 adder channel

\[ y_1 = y_2 = x_1 + x_2 \pmod{2}. \]

How to achieve capacity region?
Achieving mod 2 adder channel capacity

\[ y_1 = y_2 = x_1 + x_2 \ (\text{mod} \ 2). \]

Receiver 1:

\[
\hat{x}_2 = y_1 \oplus x_1 \\
= x_1 \oplus x_2 \oplus x_1 \\
= x_2
\]

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:

\[ R_1 \leq \frac{1}{2} \log \left( 1 + \frac{c^2 P_1}{\sigma_2^2} \right) \]
\[ R_2 \leq \frac{1}{2} \log \left( 1 + \frac{b^2 P_2}{\sigma_1^2} \right) \]

No dependence on "a" or "d"
Capacity: Gaussian noise channel

\[ R_1 \leq \frac{1}{2} \log \left(1 + \frac{c^2 P_1}{\sigma_2^2} \right) \]
\[ R_2 \leq \frac{1}{2} \log \left(1 + \frac{b^2 P_2}{\sigma_1^2} \right) \]

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

\[ Y_1 = aX_1 + bX_2 + N_1 \]
\[ Y_2 = cX_1 + dX_2 + N_2 \]
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

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

<table>
<thead>
<tr>
<th>lower/upper bounds</th>
<th>Rate (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon’s lower bound</td>
<td>0.61695</td>
</tr>
<tr>
<td>Hagelbarger’s lower bound</td>
<td>0.593</td>
</tr>
<tr>
<td>Schalkwijk’s lower bound</td>
<td>0.63056</td>
</tr>
<tr>
<td>Shannon’s upper bound</td>
<td>0.69424</td>
</tr>
<tr>
<td>Zheng’s upper bound</td>
<td>0.6390</td>
</tr>
</tbody>
</table>

General techniques for two-way channels

Adaptation
Adaptive codewords

Single-letter
\[ C = f(X_1, X_2) \]

\[
\begin{array}{c}
1 \\
\text{Channel} \\
2
\end{array}
\]

\[
\begin{array}{c}
x_1(w_1) \\
y_1 \\
x_2(w_2) \\
y_2
\end{array}
\]

Two-letter
\[ C = f(X_1^{(1)}, X_1^{(2)}, X_2^{(1)}, X_2^{(2)}) \]

\[
\begin{array}{c}
1 \\
\text{Channel} \\
2
\end{array}
\]

\[
\begin{array}{c}
x_1^{(1)}(w_1) \\
y_1^{(1)} \\
x_2^{(1)}(w_2) \\
y_2^{(1)}
\end{array}
\]

\[
\begin{array}{c}
x_1^{(1)}(w_1, y_1^{(1)}) \\
y_1^{(2)} \\
x_2^{(1)}(w_2, y_2^{(1)}) \\
y_2^{(2)}
\end{array}
\]


Adaptive codewords

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


Non-adaptive codewords:

\[ X_1 \in \{0, 1\} \]
\[ Y_1 = X_1 X_2 \]
\[ Y_2 = X_2 \mod 2 \]
\[ X_2 \in \{0, 1, 2\} \]

**Code of user 1**

\[ u_{11} = 0 \]
\[ 0 \quad 0 \quad 0 \]
\[ u_{11} = 1 \]
\[ 1 \quad 1 \quad 1 \]

**Code of user 2**

\[ u_{21} = 0 \]
\[ 1 \quad 0 \quad 2 \]
\[ u_{21} = 1 \]
\[ 0 \quad 2 \quad 0 \]

\( L=N=3 \) channel uses

Adaptive codewords:

\[ X_1 \in \{0, 1\} \]
\[ Y_1 = X_1 X_2 \]
\[ X_2 \in \{0, 1, 2\} \]

Adaptive codewords \( A_1^3 \)

Adaptive codewords \( A_2^3 \)

\( L=N=3 \) channel uses
Can we take this adaptation into account?

CAUSAL adaptation
From mutual information to directed information

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

Marko/Massey’s “Directed Information”

\[
I(X^N \rightarrow Y^N) := \sum_{n=1}^{N} I(X^n; Y_n | Y^{n-1})
\]

Kramer’s “Causally conditioned Directed Information”

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


Aside - could this be what we need?

\[
1 \quad X_1 \quad Y_2 \quad 2 \\
Y_1 \quad \quad \quad X_2
\]

\[
I(X^N; Y^N) = I(Y^N; X^N)
\]

BUT

\[
I(X^n \rightarrow Y^N) \neq I(Y^n \rightarrow X^N)
\]
Usage: capacity of two-way channels

\[ X_1 \in \{0, 1\} \]
\[ Y_1 = X_1 X_2 \]
\[ X_2 = \{0, 1, 2\} \]

Adaptive codewords \( A_1^3 \)

Adaptive codewords \( A_2^3 \)

\[ L = N = 3 \text{ channel uses} \]
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

\[
\begin{align*}
R_1 & \leq I(X_1; Y_2 | X_2) \\
R_2 & \leq I(X_2; Y_1 | X_1)
\end{align*}
\]

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
  - Properly choose $p(y_1, y_2 | x_1, x_2)$
  - Set rate in $\leftarrow$ direction to 0
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.

- Set reconciliation framework

- Interactive communication framework / communication complexity
Outline

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

\[ C = \max_{\mathbf{x}} I(X; Y) \]

\[ X^n \xrightarrow{p(y|x)} y^n \]

\[ x_1^{(1)}(w_1) \xrightarrow{x_2^{(1)}(w_2)} \]

\[ y_1^{(1)}, y_2^{(1)} \]

\[ w_r = w_1 \oplus w_2 \]
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

Outer bound
Capacity region
Achievable region
Propose a coding scheme (random codes!) 
Prove that as long as \( R \) holds, reliable communication possible 

\[
\begin{align*}
R_1 & \leq I(X_1; Y | X_2) \\
R_2 & \leq I(X_2; Y | X_1) \\
R_1 + R_2 & \leq I(X_1, X_2; Y)
\end{align*}
\]
Outer bound

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

\begin{align*}
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)
\end{align*}
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)

- 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 $(R_1, R_2)$ 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

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

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

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

$$R_2 \leq \frac{1 - \alpha}{2} \log_2 (1 + P_2)$$
Consider two examples. First, consider the AWGN MAC with blockpower constraints $P_1$ and $P_2$ for the respective transmitters 1 and 2. The maximum entropy theorem ensures that
\[
\begin{align*}
C_{\text{MAC}} &= \begin{cases} 
R_1, R_2 : & \frac{1}{2} \log(1 + P_1) \\
0 \leq R_1 \leq \frac{1}{2} \log(1 + P_1) \\
0 \leq R_2 \leq \frac{1}{2} \log(1 + P_2) \\
R_1 + R_2 \leq \frac{1}{2} \log(1 + P_1 + P_2) 
\end{cases}
\end{align*}
\]
(8.31)

The resulting region is plotted in Figure 8.4. We remark that an alternative method for blockpower constraints is to use time-division multiplexing (TDM) or frequency-division multiplexing (FDM). For example, suppose that transmitters 1 and 2 use the fractions $\alpha$ and $1 - \alpha$ of the available bandwidth, respectively. The resulting rates are
\[
\begin{align*}
R_1 &= \frac{1}{2} \log \left( \frac{1 + P_1 + P_2}{1 + P_1} \right) \\
R_2 &= \frac{1}{2} \log \left( \frac{1 + P_1 + P_2}{1 + P_2} \right)
\end{align*}
\]
(8.32)

where the transmitters block their powers in their frequency bands. The resulting rate pairs are plotted in Figure 8.4. In particular, by choosing $\alpha = \frac{P_1}{P_1 + P_2}$ one achieves a boundary point with $R_1 + R_2 = \frac{1}{2} \log(1 + P_1 + P_2)$.
(8.34)

This shows that TDM and FDM can be effective techniques for the MAC.
Multiple access channels in practice

- GSM
- CDMA
- WiFi
Broadcast channel
Broadcast channel

- 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

\[
0 \leq R_1 \leq I(U_1; Y_1) \\
0 \leq R_2 \leq I(U_2; Y_2) \\
R_1 + R_2 \leq I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2)
\]

for some distribution

\[
p(u_1, u_2, x, y_1, y_2) = p(u_1, u_2)p(x|u_1, u_2)p(y_1, y_2|x)
\]
The multiple-access channel capacity region is the closure of the convex hull of all 
for some product distribution.

For the Gaussian broadcast channel $1_R R_2 R 2_R < R_2 < 1_R < R_2 < 1_R < 2_R + 1_R < 2_R < log_2 (1 + (1 - \alpha) P/\sigma_1^2)\)

\[
R_1 < \frac{1}{2} \log_2 \left( 1 + \frac{(1 - \alpha) P}{\sigma_1^2} \right) \\
R_2 < \frac{1}{2} \log_2 \left( 1 + \frac{\alpha P}{(1 - \alpha) P + \sigma_2^2} \right) \]

The diagram shows the capacity region for $\alpha = 1$ and $\alpha = 0$. The TDM (time-division multiple-access) scheme is also indicated.
Relay channel

- 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

\[ R_{out} = \bigcup_{p(x)} I(X; Y) \]

- Relay channel

\[ R_{out} = \bigcup_{p(x, x_1)} \min \{ I(X; Y_1, Y | X_1), I(X, X_1; Y) \} \]
Relay channel: Decode + forward (DF)

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