# Lecture 22: Final Review

- Nuts and bolts
- Fundamental questions and limits
- Tools
- Practical algorithms
- Future topics

Dr. Yao Xie, ECE587, Information Theory, Duke University

# Basics

# Nuts and bolts

- Entropy:

$$H(X) = -\sum_x p(x) \log_2 p(x) \text{ (bits)}$$

$$H(X) = -\int f(x) \log f(x) dx \text{ (bits)}$$

- Conditional entropy: $H(X|Y)$, joint entropy: $H(X,Y)$

- Mutual information: reduction in uncertainty due to another random variable

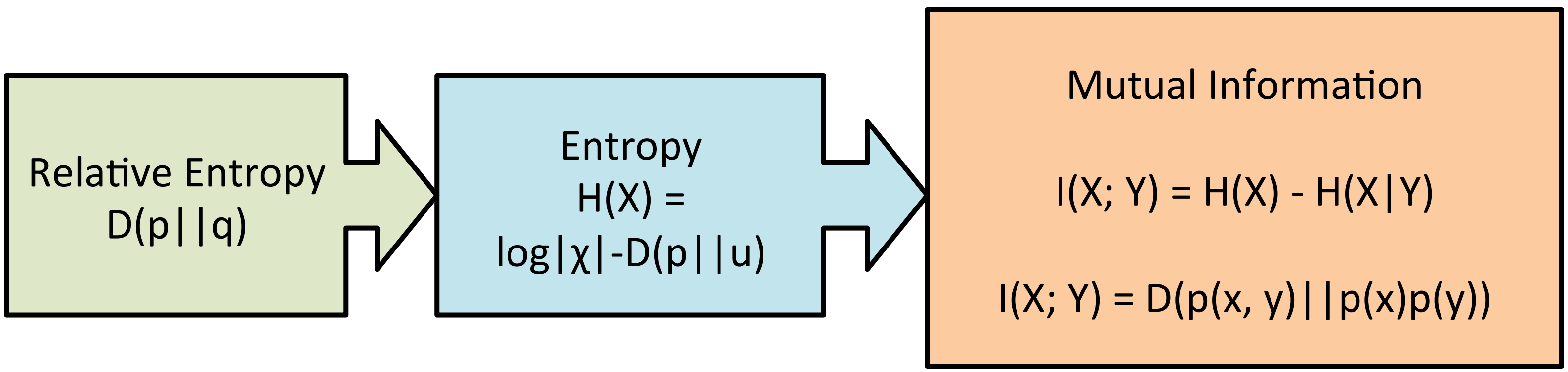
$$I(X;Y) = H(X) - H(X|Y)$$

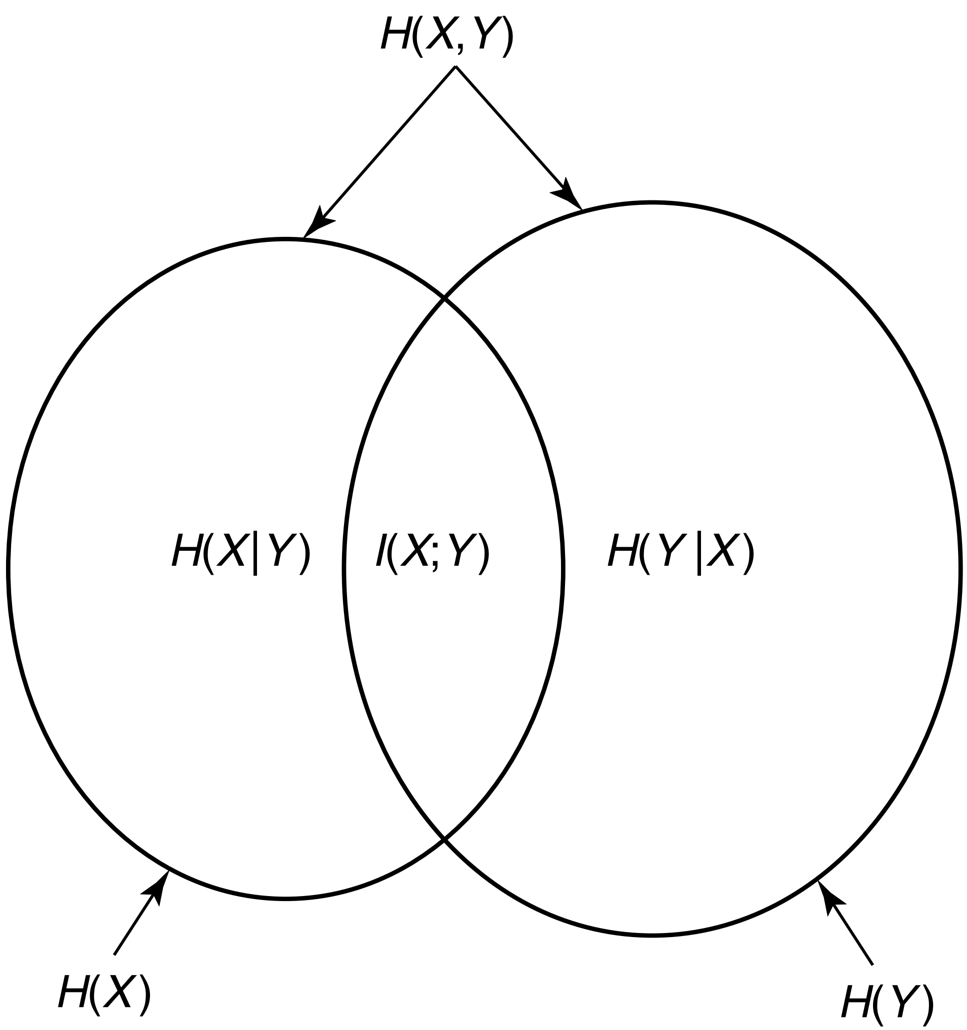
- Relative entropy: $D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$

- For stochastic processes: entropy rate

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{H(X^n)}{n} = \lim_{n\to} H(X_n|X_{n-1},\ldots,X_1)$$

$$= -\sum_{ij} \mu_i P_{ij} \log P_{ij} \text{ for 1st order Markov chain}$$




H(X,Y)
H(X|Y)
I(X;Y)
H(Y|X)
H(X)
H(Y)

# Thou Shalt Know (In)equalities

- Chain rules:
  $H(X,Y) = H(X) + H(Y|X)$,
  $H(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, \cdots, X_1)$
  $I(X_1, X_2, \cdots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, \cdots, X_1)$

- Jensen's inequality: if $f$ is a convex function, $Ef(X) \geq f(EX)$.

- Conditioning reduces entropy: $H(X|Y) \leq H(X)$

- $H(X) \geq 0$ (but differential entropy can be $< 0$), $I(X;Y) \geq 0$ (for both discrete and continuous)

- Data processing inequality: $X \to Y \to Z$

$$I(X;Z) \leq I(X;Y)$$

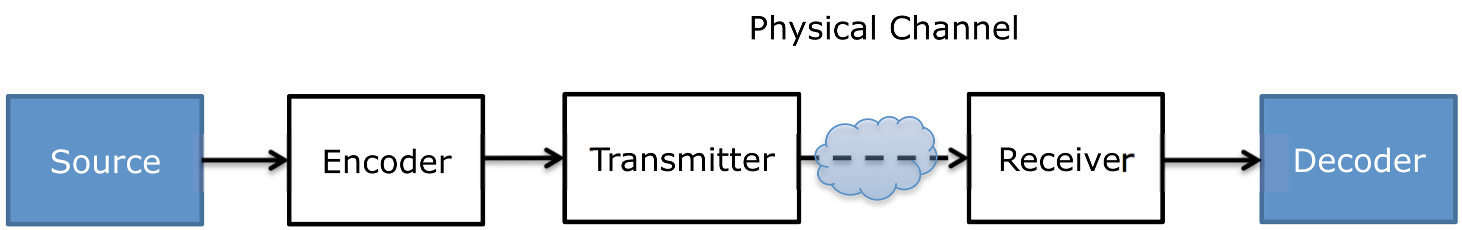
$$I(X;Z) \leq I(Y;Z)$$

- Fano's inequality:

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}$$

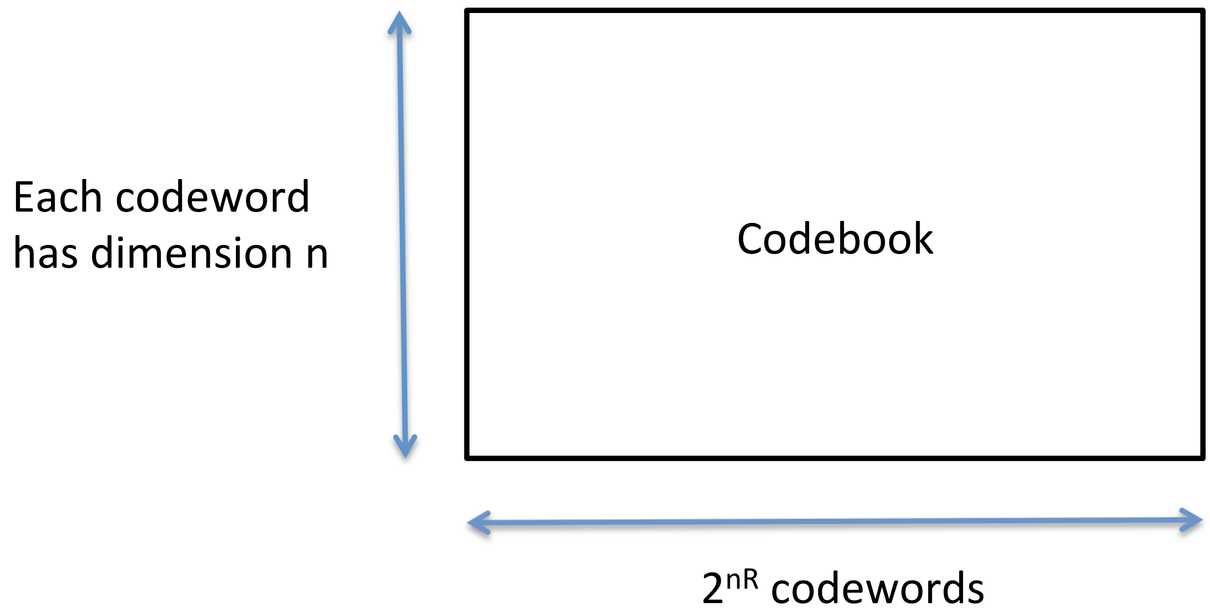
# Fundamental Questions and Limits

# Fundamental questions

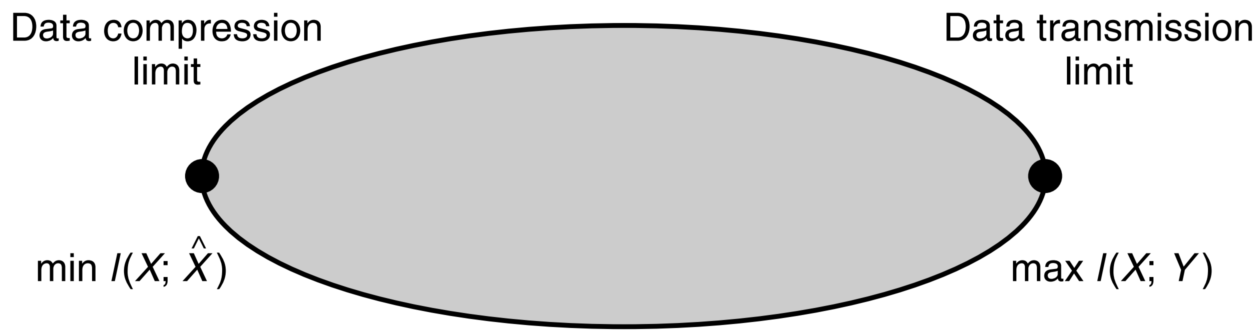



- Data compression limit (lossless source coding)
- Data transmission limit (channel capacity)
- Tradeoff between rate and distortion (lossy compression)

# Codebook



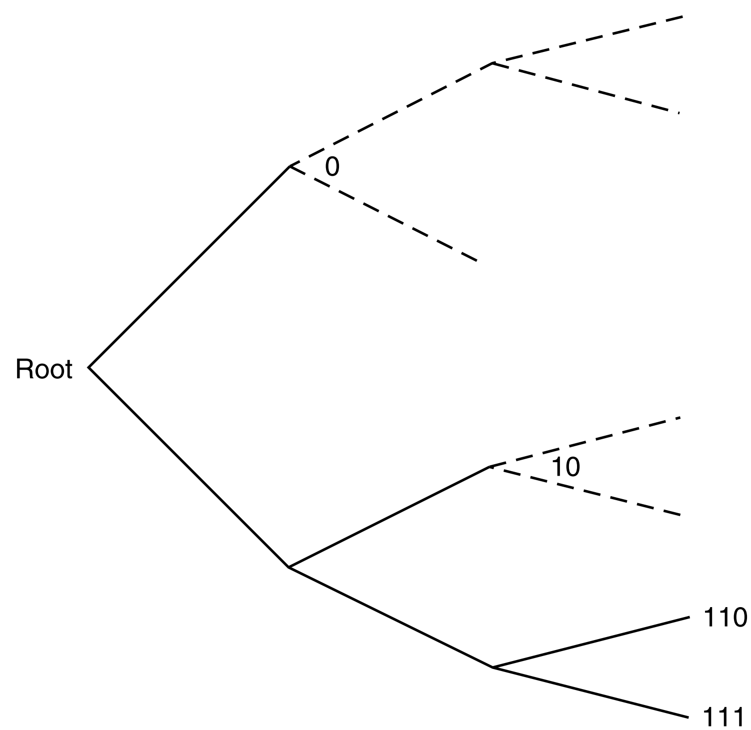
# Fundamental limits

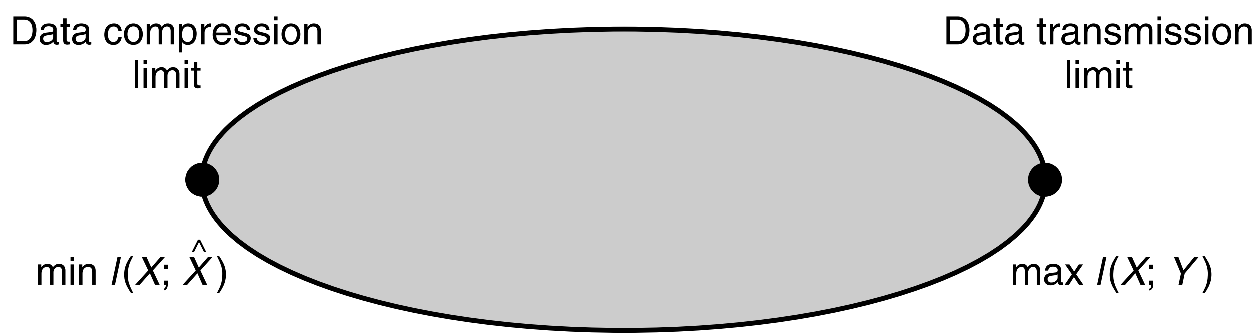



**Lossless compression**:

- $I(X;\hat{X}) = H(X) - H(X|\hat{X})$
- $H(X|\hat{X}) = 0$, $I(X;\hat{X}) = H(X)$
- Data compression limit: $\sum_x l(x)p(x) \geq H(X)$

- instantaneous code: $\sum_{i=1}^{m} D^{-l_i} \leq 1$
- optimal code length: $l_i^* = -\log_D p_i$





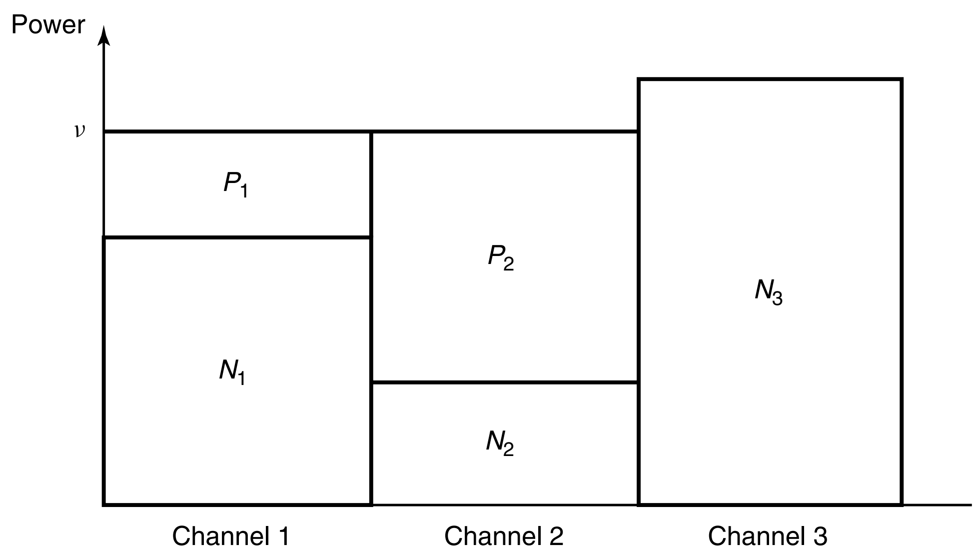

**Channel capacity**:

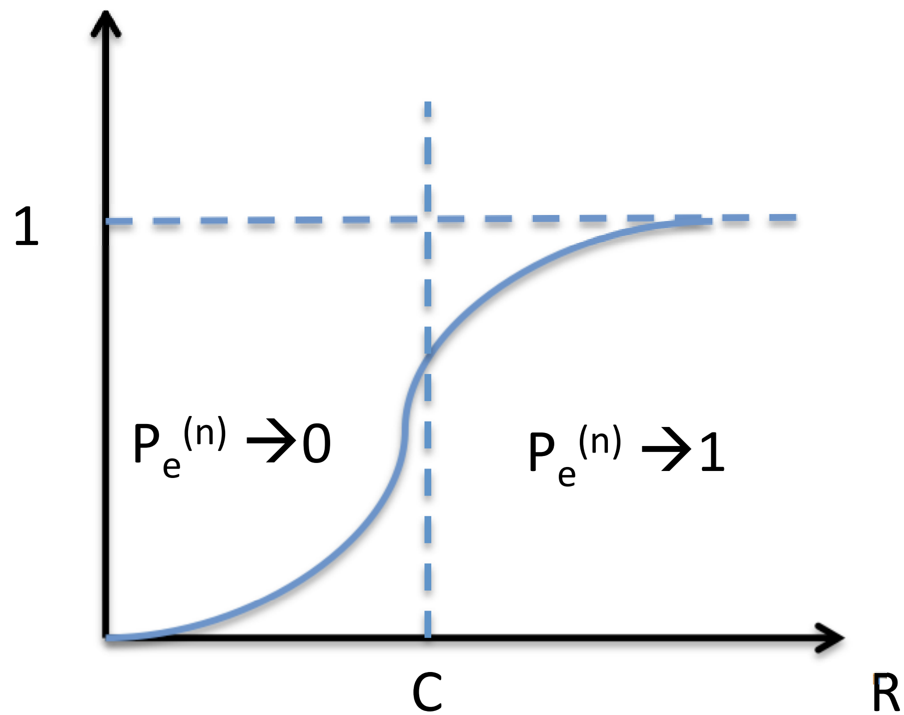
- given fixed channel with transition probability $p(y|x)$
- $C = \max_{p(x)} I(X; Y)$

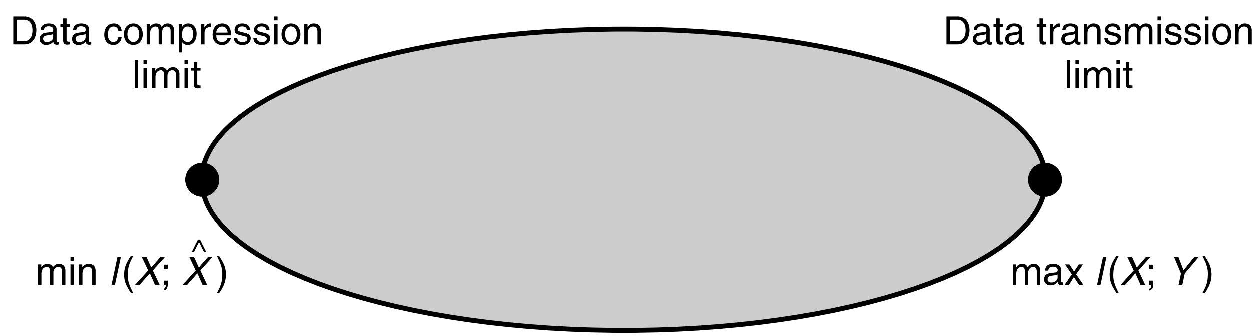
- BSC: $C = 1 - H(p)$, Erasure: $C = 1 - \alpha$

- Gaussian channel: $C = \frac{1}{2}\log(1 + \frac{P}{N})$

- water-filling




1
$P_e^{(n)} \rightarrow 0$
$P_e^{(n)} \rightarrow 1$
C
R




**Rate-distortion**:

- given source with distribution $p(x)$
- $R(D) = \min_{p(x|\hat{x}): \sum p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D} I(X; \hat{X})$
- compute $R(D)$: construct "test" channel

- Bernoulli source: $R(D) = (H(p) - H(D))^+$
- Gaussian source: $R(D) = \left(\frac{1}{2}\log\frac{\sigma^2}{D}\right)^+$

# Tools

# Tools: from probability

- Law of Large Number (LLN): If $x_n$ independent and identically distributed,

$$\frac{1}{N}\sum_{n=1}^{N} x_n \to \mathbb{E}\{X\}, \text{w.p.}1$$

- Variance: $\sigma^2 = \mathbb{E}\{(X-\mu)^2\} = \mathbb{E}\{X^2\} - \mu^2$

- Central Limit Theorem (CLT):

$$\frac{1}{\sqrt{N\sigma^2}}\sum_{n=1}^{N}(x_n - \mu) \to \mathcal{N}(0,1)$$

# Tools: AEP

- LLN for product of i.i.d. random variables

$$\sqrt[n]{\prod_{i=1}^{n} X_i} \rightarrow e^{E(\log X)}$$

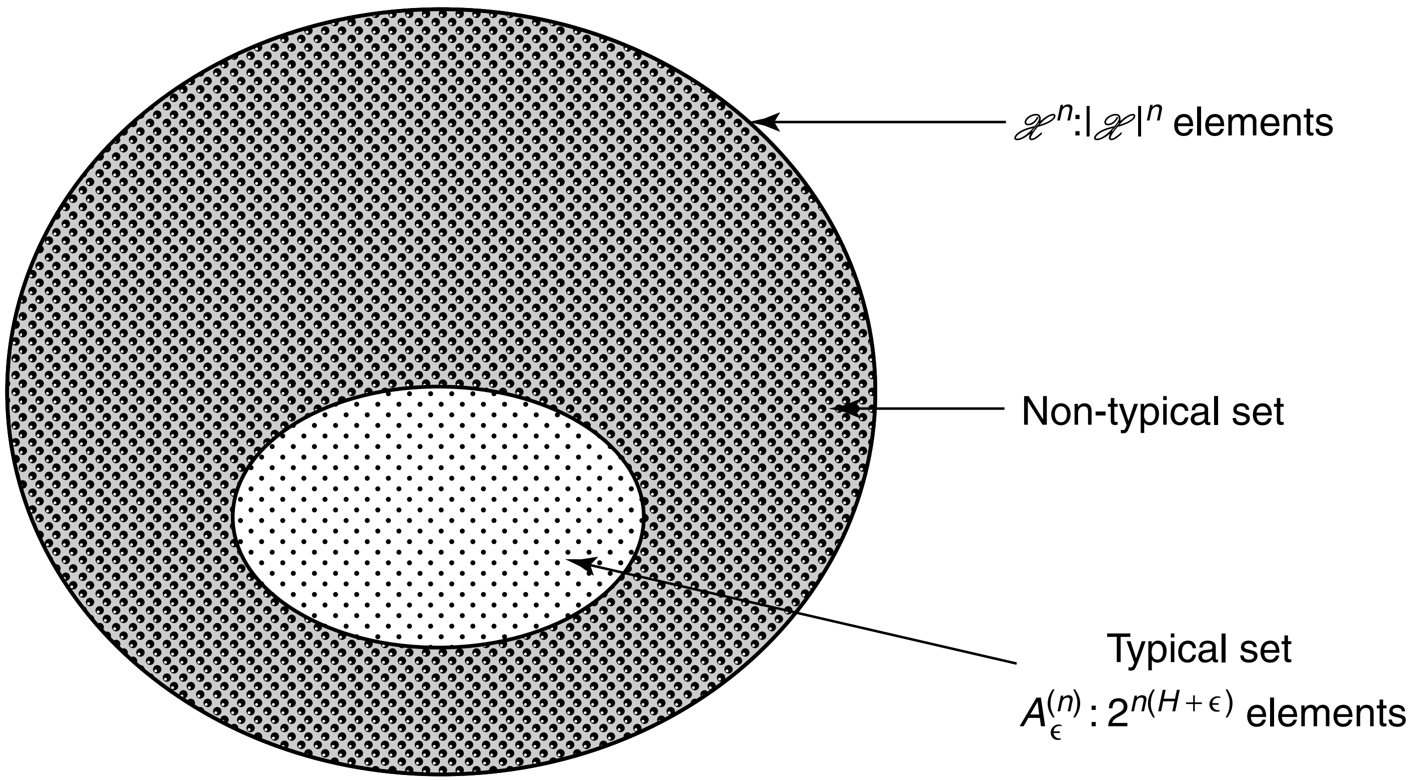
- AEP

$$\frac{1}{n}\log\frac{1}{p(X_1, X_2, \ldots, X_n)} \rightarrow H(X)$$

$$p(X_1, X_2, \ldots, X_n) \approx 2^{-nH(X)}$$

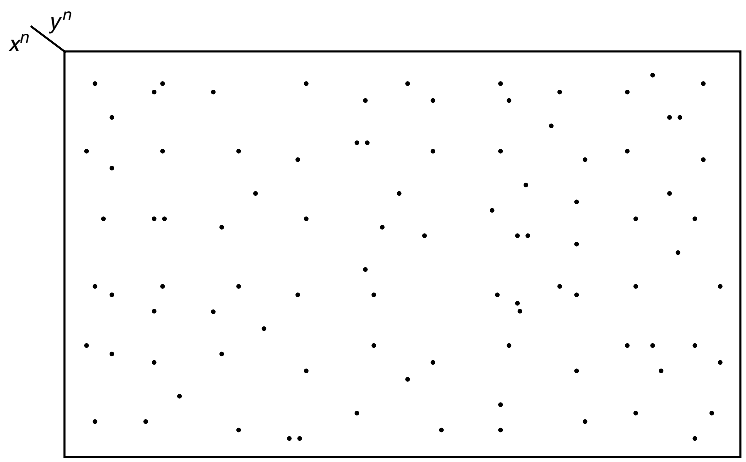
Divide all sequences in $\mathcal{X}^n$ into two sets

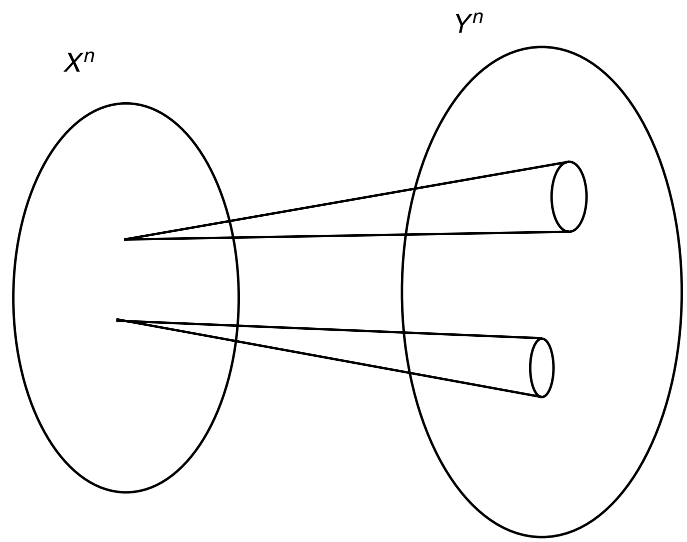
# Typicality



$\mathscr{X}^n$:$|\mathscr{X}|^n$ elements
Non-typical set
Typical set
$A_\epsilon^{(n)}: 2^{n(H+\epsilon)}$ elements

# Joint typicality



# Tools: Maximum entropy

Discrete:

- $H(X) \leq \log |\mathcal{X}|$, equality when $X$ has uniform distribution

Continuous:

- $H(X) \leq \frac{1}{2} \log(2\pi e)^n |K|$, $EX^2 = K$
  equality when $X \sim \mathcal{N}(0, K)$

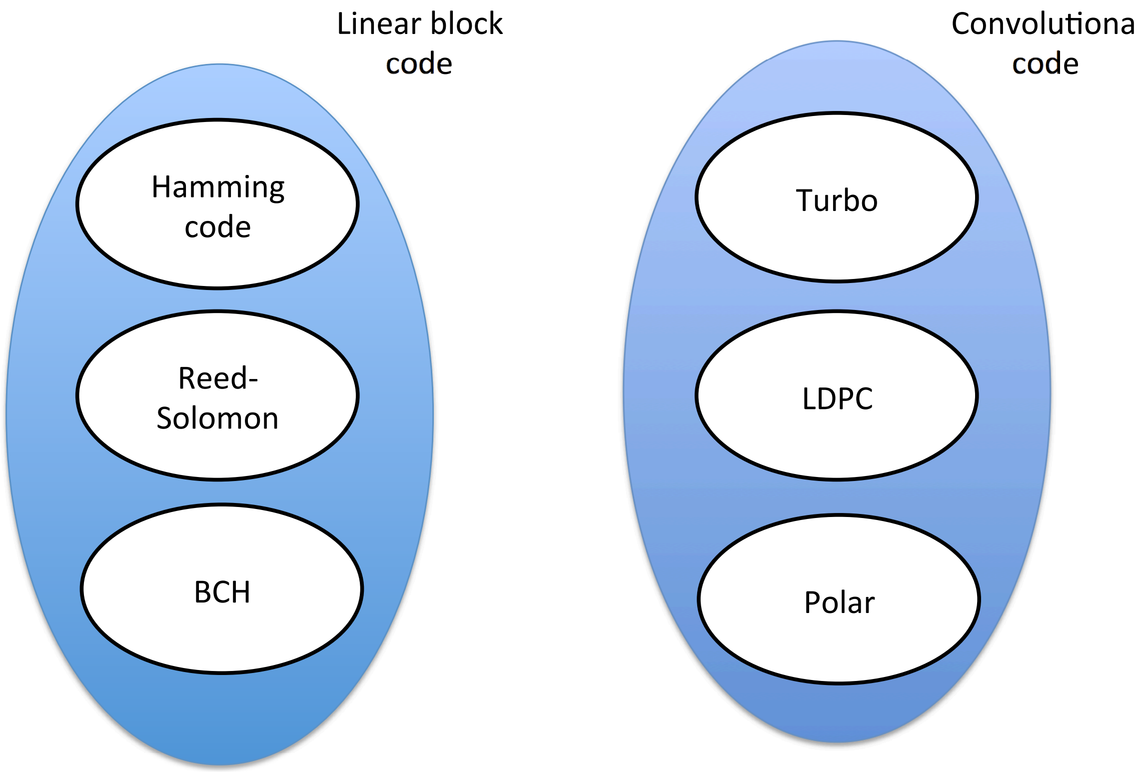
# Practical Algorithms

# Practical algorithms: source coding

- Huffman, Shannon-Fano-Elias, Arithmetic code

| Codeword Length | Codeword | $X$ | Probability | | | | |
|---|---|---|---|---|---|---|---|
| 2 | 01 | 1 | 0.25 | 0.3 | 0.45 | 0.55 | 1 |
| 2 | 10 | 2 | 0.25 | 0.25 | 0.3 | 0.45 | |
| 2 | 11 | 3 | 0.2 | 0.25 | 0.25 | | |
| 3 | 000 | 4 | 0.15 | 0.2 | | | |
| 3 | 001 | 5 | 0.15 | | | | |

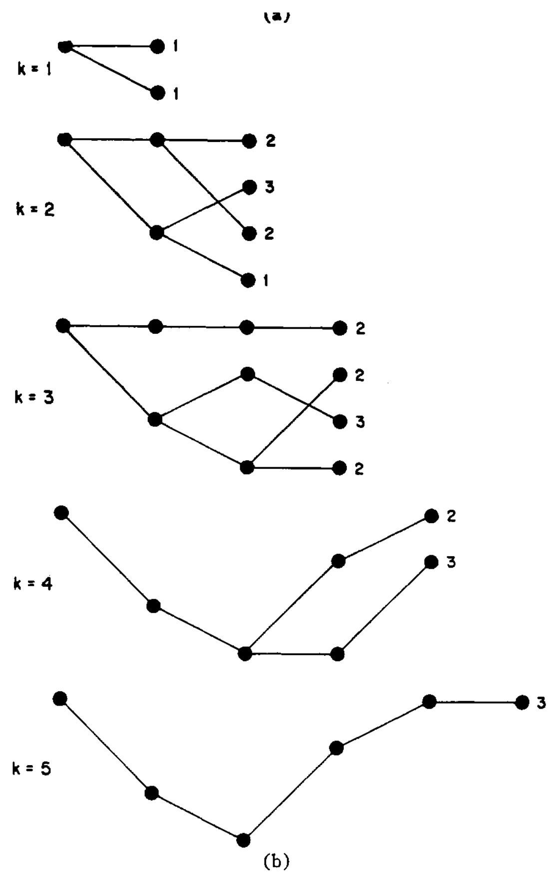
# Practical algorithms: channel coding




A (partial) diagram

# Practical algorithms: decoding

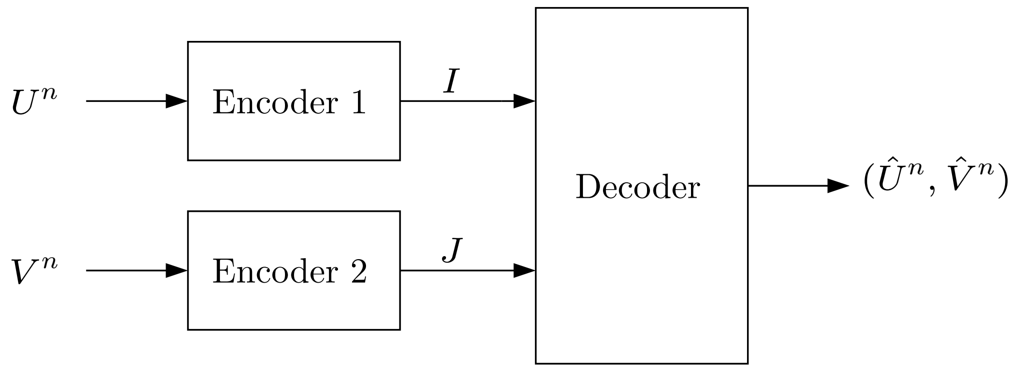
Viterbi algorithm

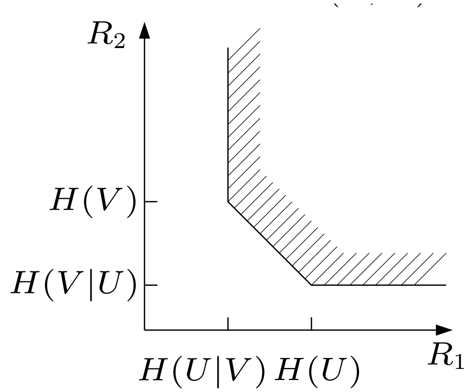


# What are some future topics

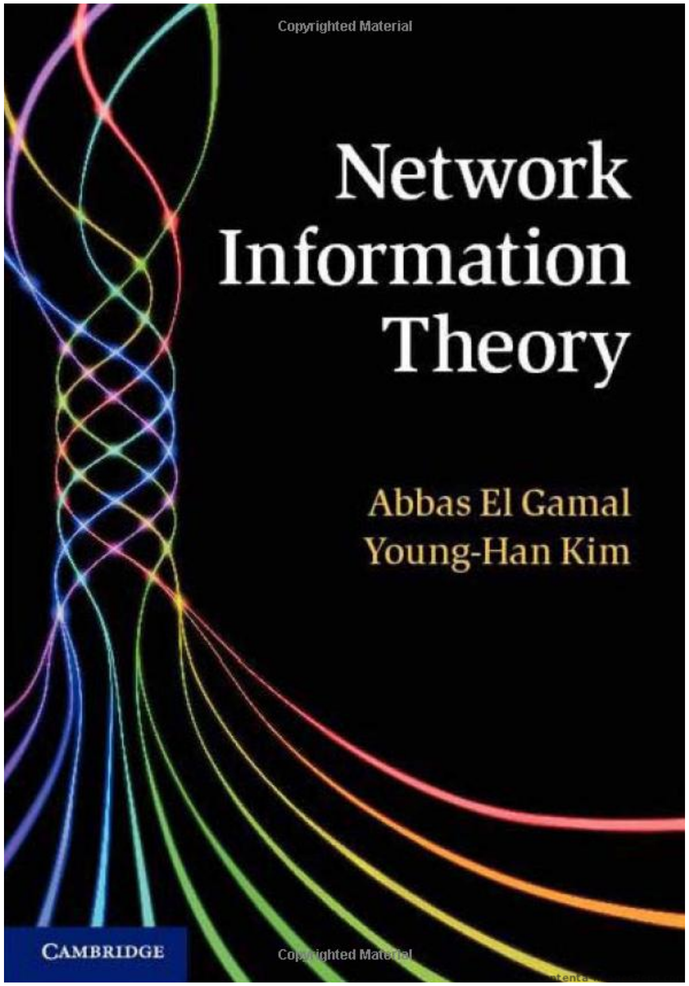
# Distributed lossless coding

Consider the two i.i.d. sources $(U, V) \sim p(u, v)$

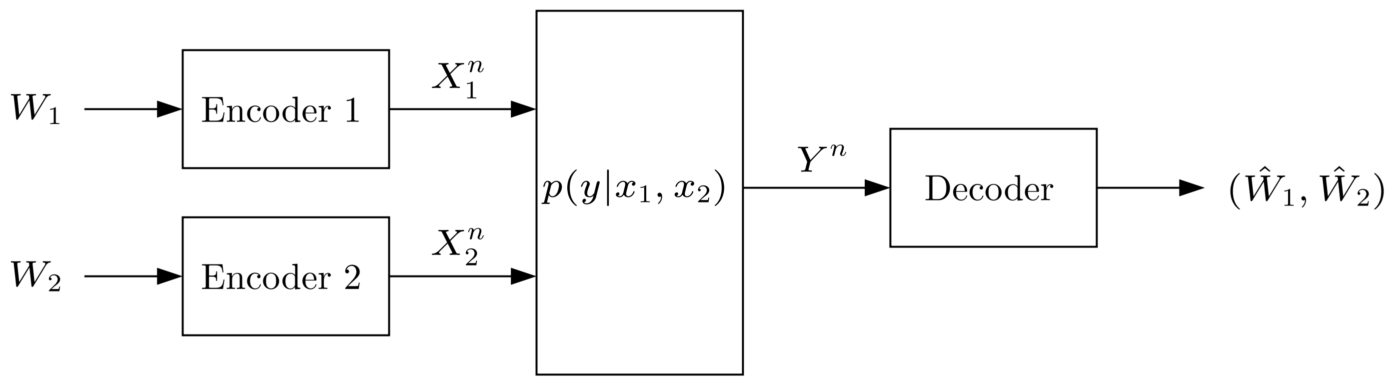


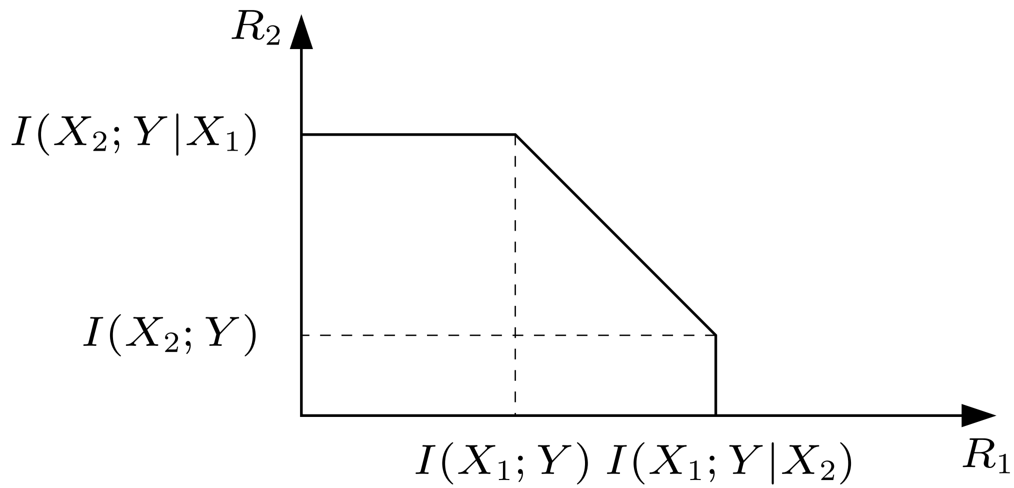



# Multi-user information theory



# Multiple access channel

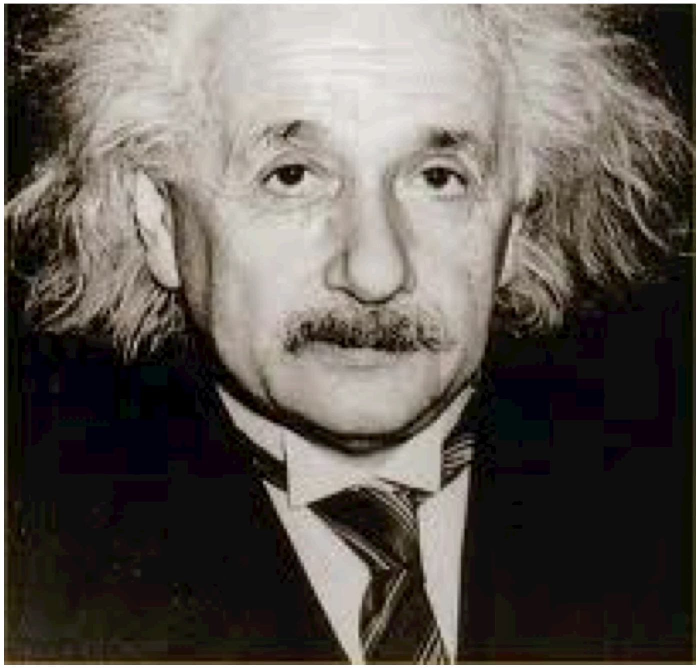


# Statistics

- Large deviation theory
- Stein's lemma...

## One last thing...



Make things as simple as possible, but not simpler.

– A. Einstein