RBE 510 – Multi-Robot Systems
Lecture 6: Sensing and Sensor Placement

Kevin Leahy

September 9, 2025

Admin

- HW 1 solutions and grades posted
  - Looks good overall. Questions?

- HW 2 due Friday midnight
  - Questions?

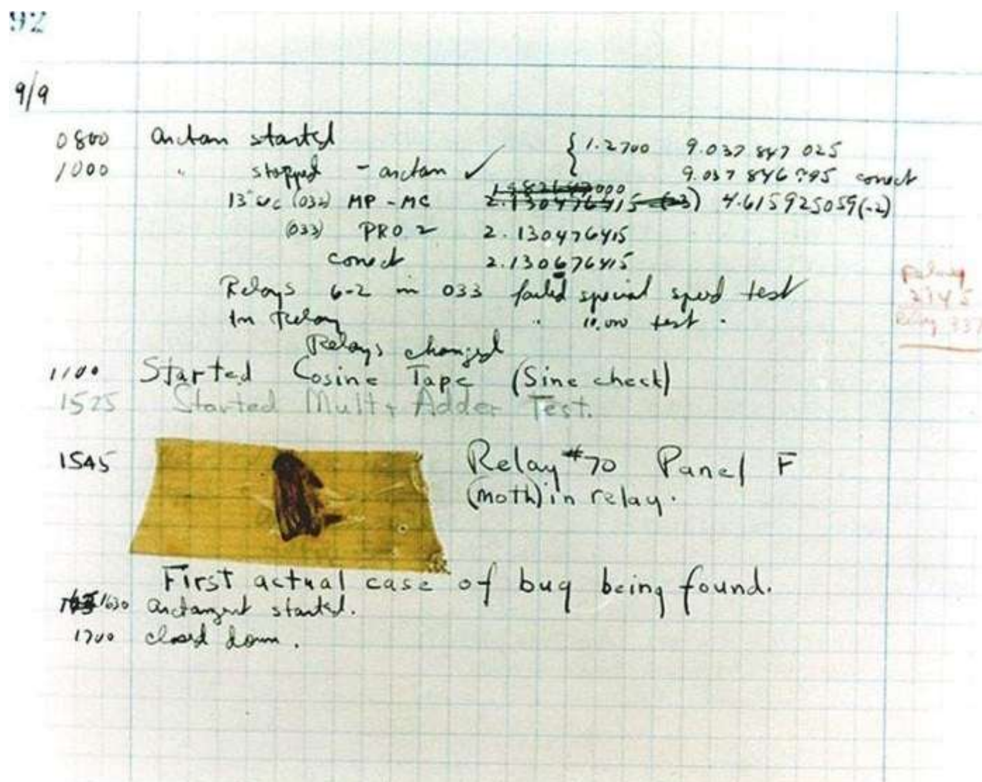
- Second paper posted this afternoon
  - Next presentation on Tuesday, September 16

- Office hours this week Thursday (virtual)

HW 2 Tips

- Your step function should update the physical state of your robot using the value computed with the **`ctl`** function.
- You should constrain the robot's state to be between **`0`** and **`2*pi`**. The modulo operator (**`%`**) should help with this.
- It is possible to encounter errors when the magnitude of the value computed by **`ctl`** gets too small. I recommend setting a minimum value for the output of **`ctl`** (say, **`0.001`**), and not allowing the output to be smaller than that value
- Play around with time horizon if agents are not quite converging

Today


9/9
0800 Arctan started
1000 stopped - arctan
Relays changed
1100 Started Cosine Tape (Sine check)
1525 Started Mult+ Adder Test.
1545 Relay #70 Panel F
(moth) in relay.
First actual case of bug being found.
1700 closed down.

Today

- Sensing introduction

- Sensing models

- Static sensor placement
  - Complexity
  - Information theory
  - Submodularity
  - Approaches to optimization

- **Goal: build up some models of sensing and understand the problems of complexity**

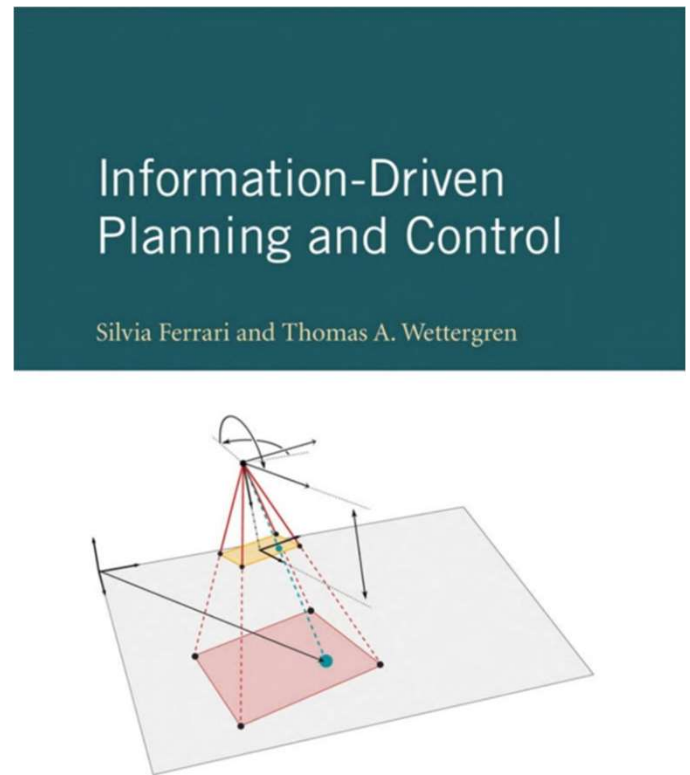
Overview: Multi-robot Sensing and Estimation

Switching Gears

- Lots of motion and control before
- Now looking at sensing and estimation
- Given a sensing objective and n robots:
  - Where should I put the robots to maximize what I sense?
  - How should I move them?
  - How do I quantify sensing quality?

Good resources

- Lecture notes will include citations for important resources
- Includes the books we’ve been using, as well as research papers
- Information-Driven Planning and Control (2021) is a great reference for diving in further



Motivation

The resulting distribution is shown in Figure 5.

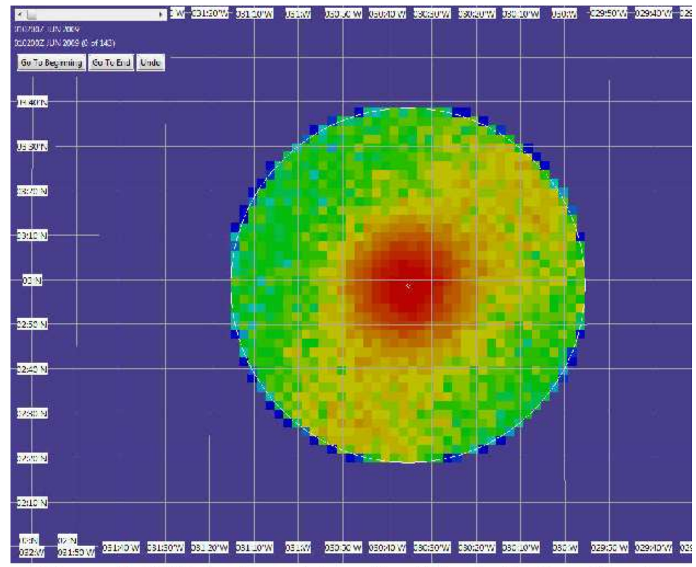


Figure 5. PDF for Impact Location Prior to Surface Search

We will focus primarily on sensing the environment and objects in it

Problem Characteristics

- Quality to be sensed:
  - Static
  - Dynamic

- Tasks:
  - Coverage
  - Search

- We are especially interested in stochasticity in sensing and how to manage it

Roadmap for the next few lectures

- Static sensing – get used to objectives, main difficulties, and prime our intuition
- Coverage – adding motion into the sensor placement problem
- Search – adding motion into the agents *and* the environment
  - This will help us segue into problems of decision-making in later lectures

Probability and Sensing

Probability Terms

- Sample space S of all possible outcomes of an experiment

- A subset of outcomes is an event $\mathcal{E}$

- Example:
  - Coin flip: $S = \{heads, tails\}$; $\mathcal{E}_1 = heads$, $\mathcal{E}_2 = tails$
  - 2 coin flips: $S = \{hh, ht, th, tt\}$; $\mathcal{E}_1 = tt, \ldots$

- Note: any subset of outcomes is an event

Axioms of Proability

1. Non-negativity: $P(\mathcal{E}) \geq 0$

2. Normalization: $P(S) = 1$

3. Additivity: $P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2)$ for $\mathcal{E}_1 \cap \mathcal{E}_2 = \{\emptyset\}$ (mutually exclusive events)

Mutually exclusive events have $P(\mathcal{E}_1 \cap \mathcal{E}_2) = 0$

Consequences of the Axioms

1. $P(\mathcal{E}^C) = 1 - P(\mathcal{E})$, where $\mathcal{E}^C = S \setminus \mathcal{E}$

2. $P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2) - P(\mathcal{E}_1 \cap \mathcal{E}_2)$

3. $P(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3) = P(\mathcal{E}_1) + P\left(\mathcal{E}_1^C \cap \mathcal{E}_2\right) + P\left(\mathcal{E}_1^C + \mathcal{E}_2^C + \mathcal{E}_3\right)$

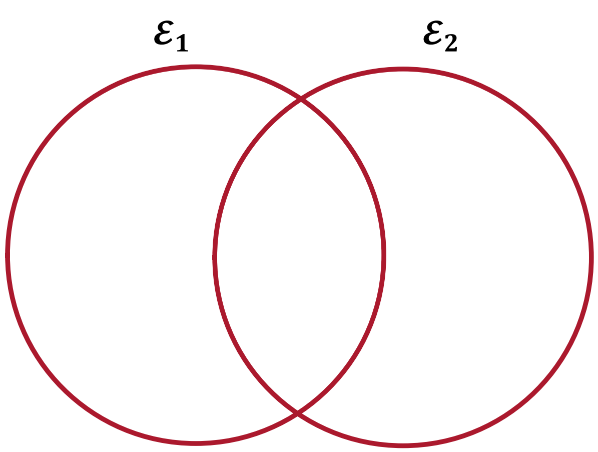
Example and Review (Ferrari and Wettergen 4.1)

Example and Review (Ferrari and Wettergen 4.1)

Conditional Probability

- Relationship among probabilities of different events can be represented using conditional probability

- For two events $\mathcal{E}_1$ and $\mathcal{E}_2$, conditional probability of $\mathcal{E}_1$ given $\mathcal{E}_2$:



$$P(\mathcal{E}_1 \mid \mathcal{E}_2) = \frac{P(\mathcal{E}_1 \cap \mathcal{E}_2)}{P(\mathcal{E}_2)}$$

Conditional Probability

- We can treat $\mathcal{E}_2$ as the *prior*

- Then, we can write

$$P(\mathcal{E}_1 \cap \mathcal{E}_2) =$$

Independence, Conditional Independence

- Consider events $\mathcal{E}_1$ and $\mathcal{E}_2$

- They are independent if

$$P(\mathcal{E}_1 \cap \mathcal{E}_2) = P(\mathcal{E}_1)P(\mathcal{E}_2)$$

- They are conditionally independent given $\mathcal{E}_3$ if

$$P(\,\mathcal{E}_1 \cap \mathcal{E}_2 \mid \mathcal{E}_3\,) = P(\,\mathcal{E}_1 \mid \mathcal{E}_3\,)P(\,\mathcal{E}_2 \mid \mathcal{E}_3\,)$$

Conditional Probability

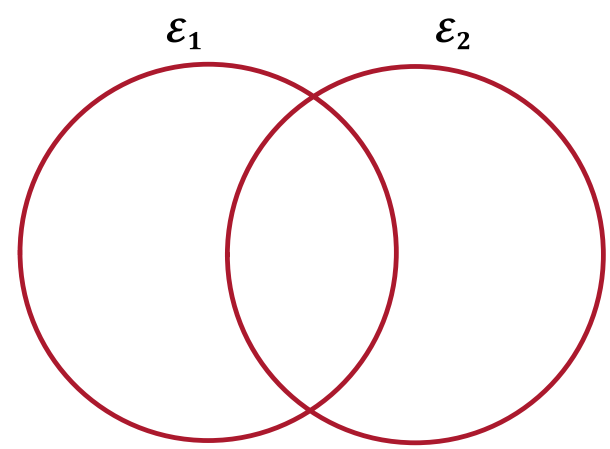
- Consider $\mathcal{E}$ occurs only if $\mathcal{E}_1, \ldots, \mathcal{E}_n$ all occur

- $P(\mathcal{E}) = P(\mathcal{E}_1 \cap \mathcal{E}_2 \cap \cdots \cap \mathcal{E}_n) =$

Conditional Probability

- If S is partitioned into $\mathcal{E}_1, \dots, \mathcal{E}_n$ disjoint events
- For any other event X in the environment
- $P(X) = P(X \cap \mathcal{E}_1) + \cdots + P(X \cap \mathcal{E}_n)$

Bayes' Rule



Bayes Rule for Multiple Events

- For $\mathcal{E}_1, \dots, \mathcal{E}_n$, mutually exclusive events, with only one true event
- Probability of event $\mathcal{E}_i$ given X can be obtained as

- $P(\mathcal{E}_i \mid X) = \frac{P(X|\mathcal{E}_i)}{P(X)} =$

Example and Review (Ferrari and Wettergen 4.2)

Example and Review (Ferrari and Wettergen 4.2)

Sensing Models and Belief

Classification

- Classification: inferring value of hidden categorical variable from multiple (possibly heterogeneous) sensor measurements

- For us

$$P(X = x \mid Y = y)$$

Sensor Model

- Let's put a sensor in location q in one of two locations, s_1 or s_2
- $S = \{s_1, s_2\}$
- There is a target in one of the two locations but we don't know which
- How to characterize the sensor?

Sensor Model

- Measurement likelihood

- If we put the sensor where the target is
  - $P(y = 1 \mid q = s) = (1 - \beta)$
  - $P(y = 0 \mid q = s) = \beta$

- If we put it somewhere else
  - $P(y = 1 \mid q \neq s) = \alpha$
  - $P(y = 0 \mid q \neq s) = (1 - \alpha)$

Representing Prior Knowledge

- If we know there is *exactly one* target, then we can represent probability it is in s_1 or s_2
- $P(s = s_1)$
- $P(s = s_2)$
- We know $P(s = s_1) + P(s = s_2) = 1$
- This is a discrete probability distribution we call the **belief**

Belief

- Belief can be updated with Bayes' rule

$$P(S = s \mid Y = y; q) = \frac{P(Y = y \mid S = s; q)P(S = s)}{P(Y = y; q)}$$

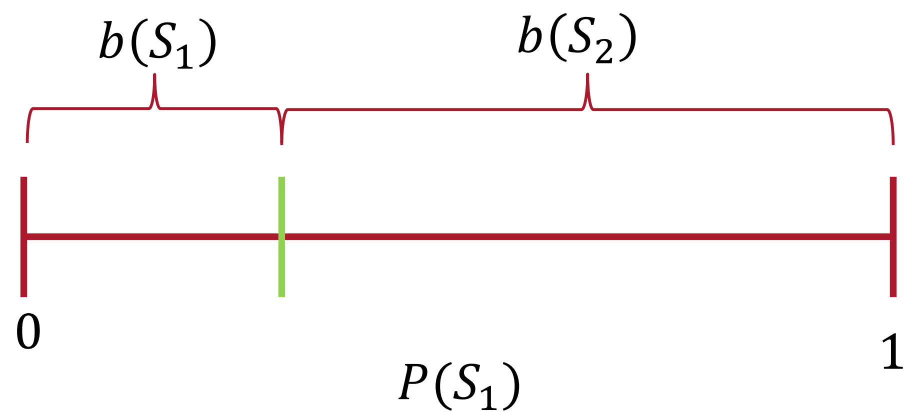
Updating Belief Example

Belief

- Target is in one of 2 locations
- $b(S_1) = P(S = S_1)$
- $b(S_2) = P(S = S_2)$
- $\sum_s b(s) = 1 \rightarrow b(S_1) + b(S_2) = 1$
- $b(S_2) = 1 - b(S_1)$
- Our entire belief is capture by $b(S_1)$

Belief

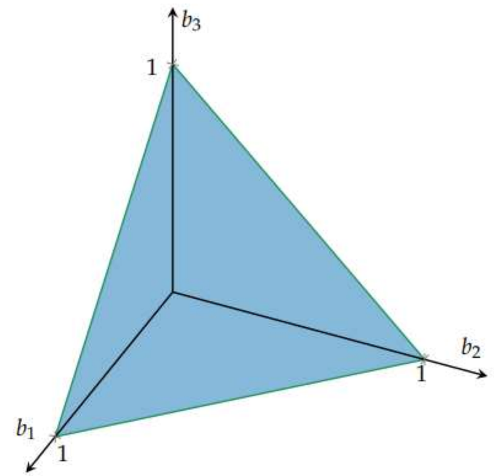
- $\sum_s b(s) = 1 \rightarrow b(S_1) + b(S_2) = 1$
- $b(S_2) = 1 - b(S_1)$



- In 2D, belief is a simple partition

Belief is a simplex

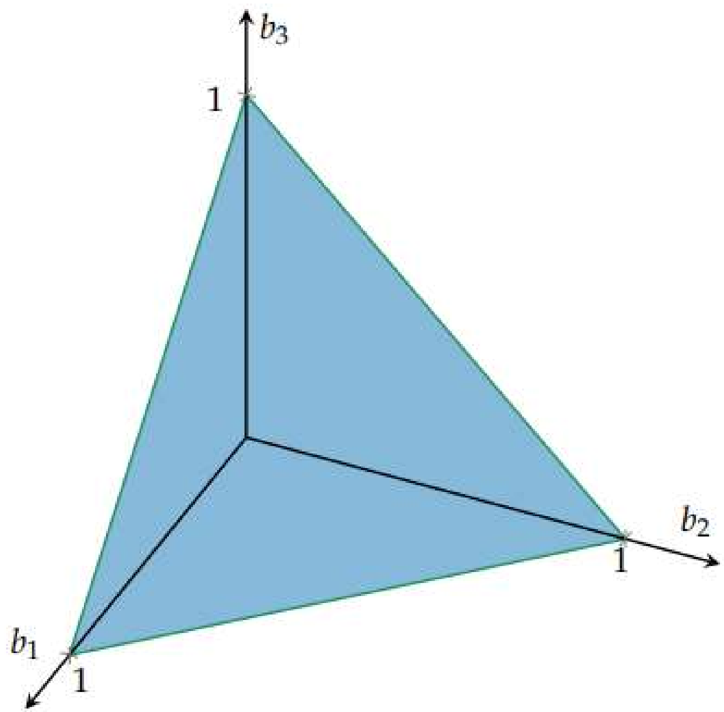
- In 3D (and in general) a point in discrete belief space lies on a ***simplex:***
  - $b(s) \geq 0$ for all $s \in S$
  - $\sum_s b(s) = 1$



- Handy vector form:
  - $\boldsymbol{b} \geq \boldsymbol{0}$
  - $\boldsymbol{1}^T \boldsymbol{b} = \boldsymbol{1}$

- For an n-dimensional space, a simplex is an $n-1$-dimensional space embedded in n dimensions

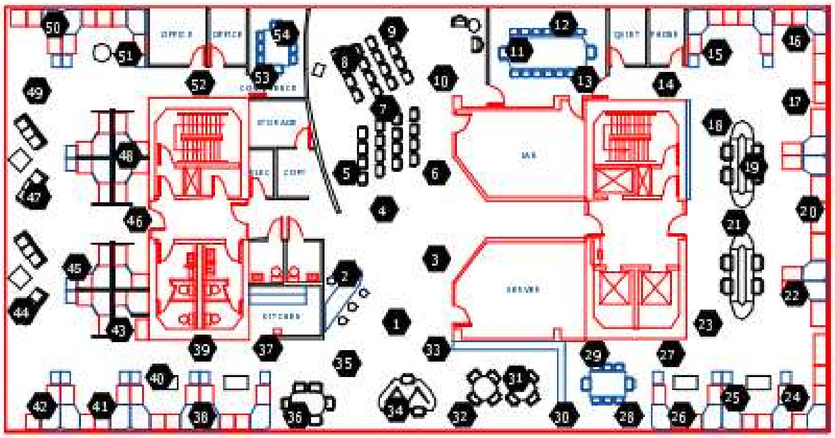
3D Simplex

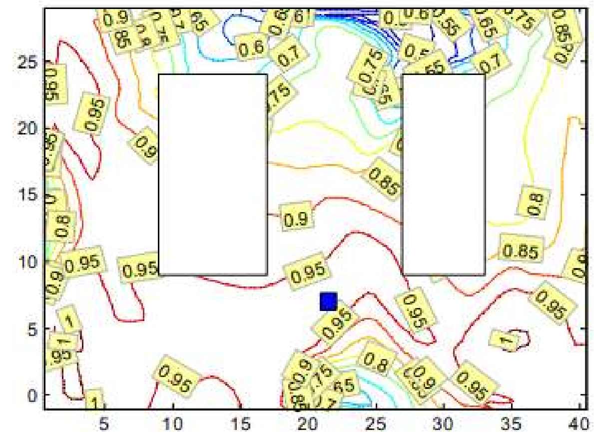


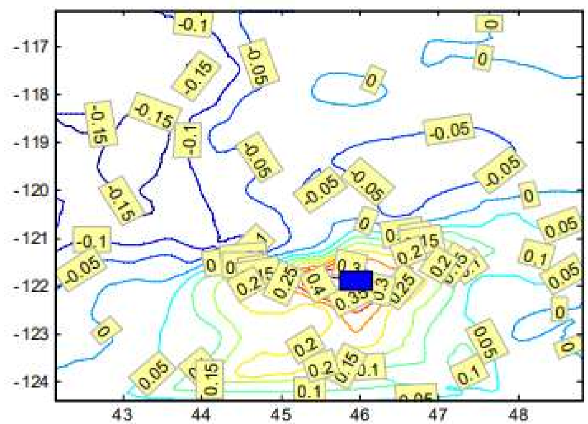
- If $b_1 = 1$ then $b_2 = b_3 = 0$
- Same if b_2 or b_3 equal 1
- If $b_1 = 0$, then it is reduced to the 2D partition between b_2 and b_3
- Same if b_2 or b_3 equal 0

Static Sensor Placement

Examples



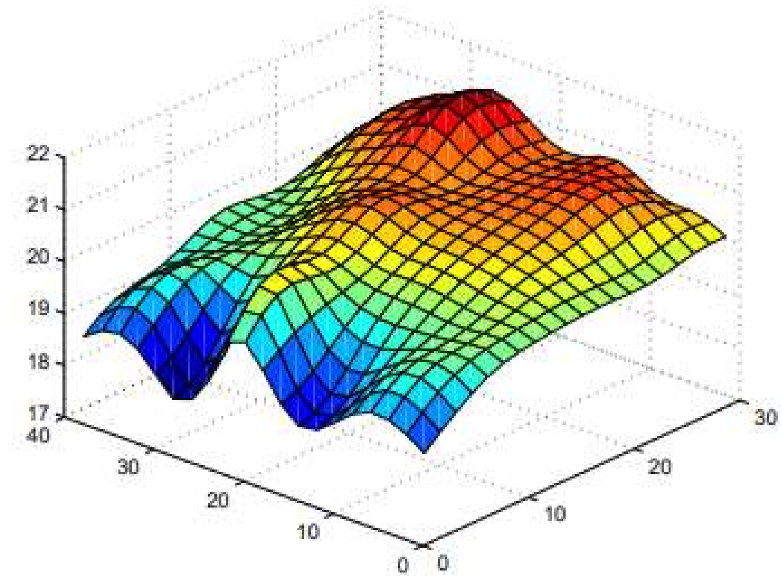

25
20
15
10
5
0
5 10 15 20 25 30 35 40


-117
-118
-119
-120
-121
-122
-123
-124
43 44 45 46 47 48

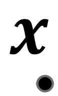
Problem Definition

- Given
  - Background probability distribution
  - Robot sensor model
  - Estimated impact of measurements on probability distribution
  - Objective function over the distribution

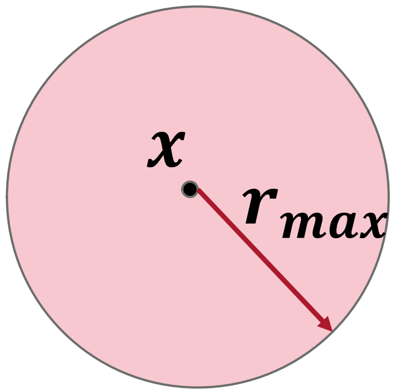
- Place a set of robots in locations to maximize your objective function



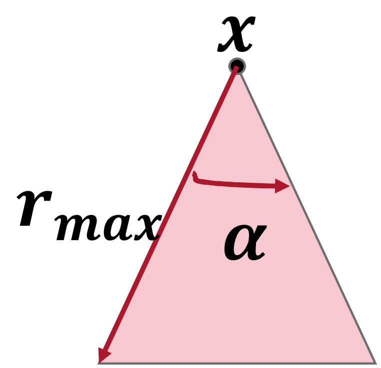
Field-of-View (FoV) Sensors



Point

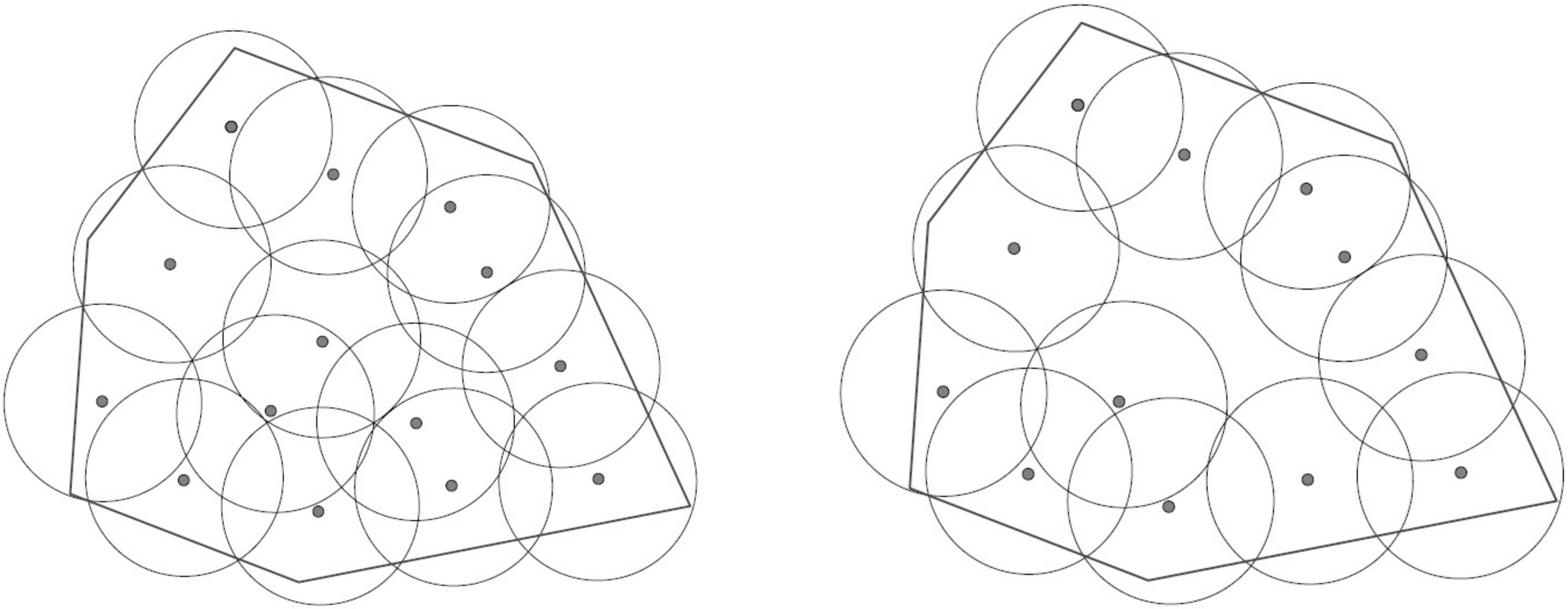


Omnidirectional



Directional

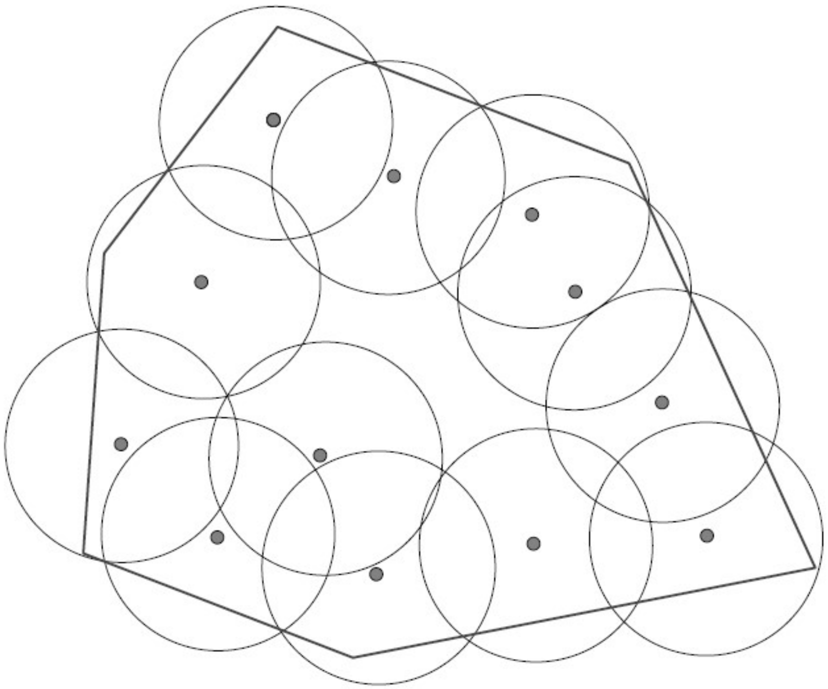
Covering the Environment



Coverage

- For uniform coverage (i.e., all areas equally “interesting”)
- Coverage criterion is $\frac{Total\ area\ of\ sensor\ footprints}{Total\ ROI}$
- Can be weighted by “interest”

Incomplete Coverage



- If we can't cover the whole environment
  - Where is the best set of places to put our limited number of sensors?
  - Is there a number of sensors that is "good enough" without placing all available sensors?
- Discretize and solve?

Combinatorics

- Given
  - n discrete locations
  - k robots
- Pick a subset of k locations that maximize your objective
- How to solve?
- Let’s try brute force…

Combinatorics

- How many options are there to try?

- $\binom{n}{k}$

- Equivalent to $\frac{n!}{k!(n-k)!}$

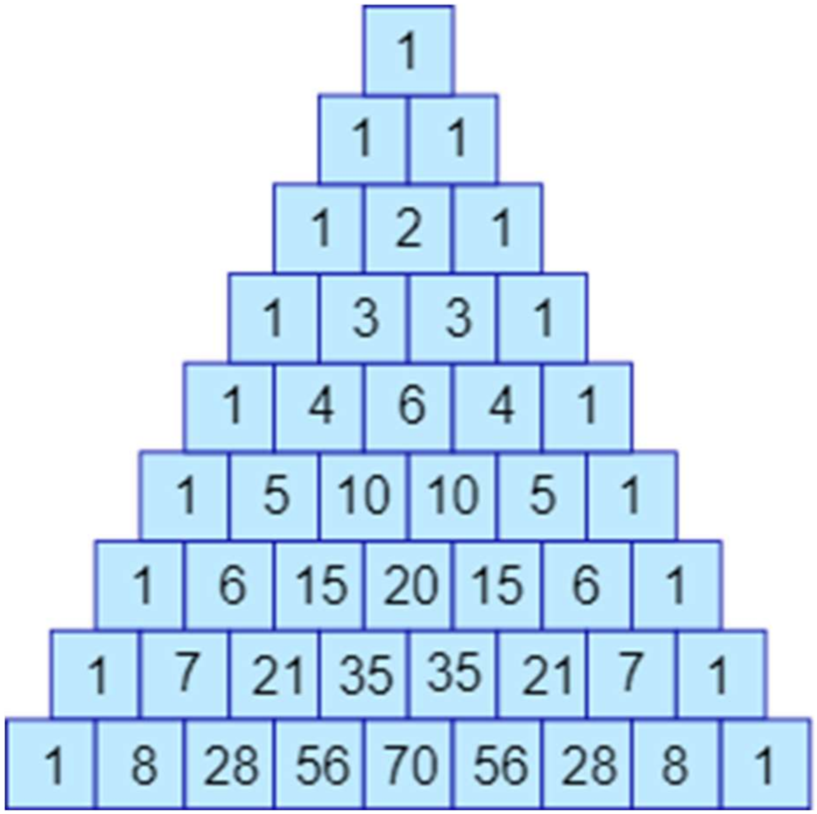
- How big is this?

Binomial Coefficient – How many combinations?

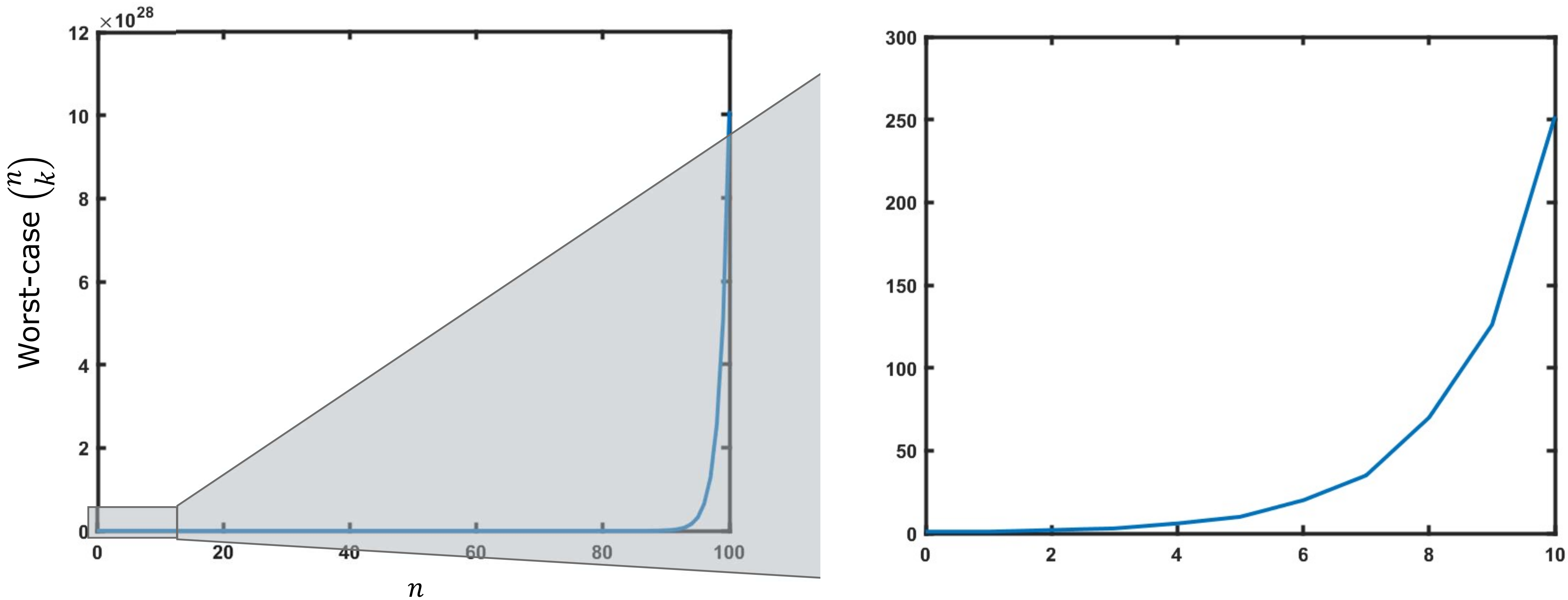
- Let’s start with $n = 2$
  - $k = 0$
  - $k = 1$
  - $k = 2$

- How about $n = 5$
  - $k = 0$
  - $k = 1$
  - $k = 2$
  - $k = 3$
  - $k = 4$
  - $k = 5$

Pascal’s Triangle



Combinatorics



A way forward

- What can we do?
  - We know that in general we can’t get to optimality
- But maybe heuristics can help?
- In an ideal world, we can bound our heuristic
  - Within a certain distance of the optimal solution
  - Let’s try!

Information Theory

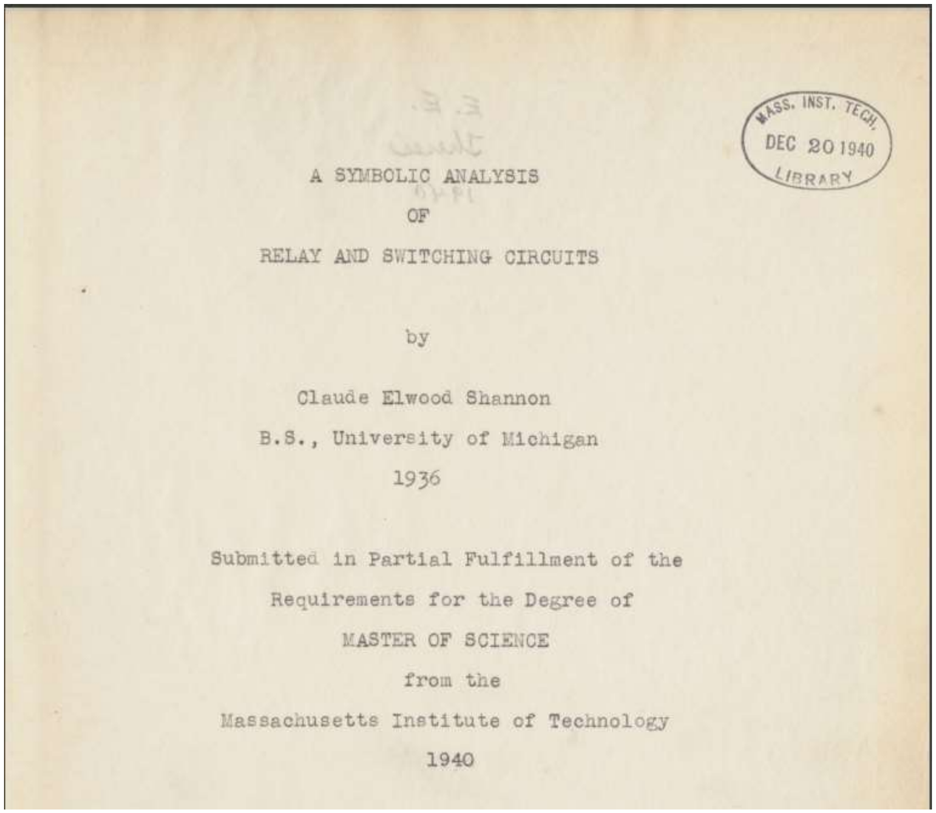
Sensing Objective and Information Theory

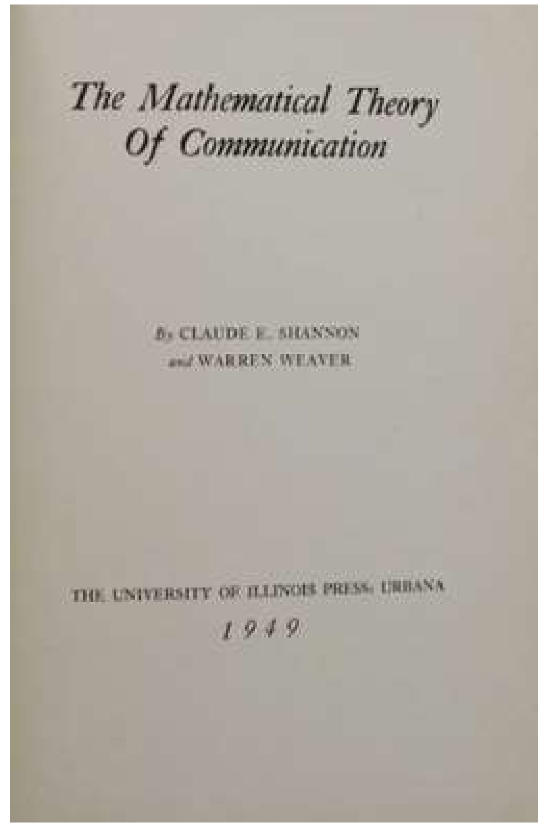
- What’s a good sensing objective? How do we know we’ve “done it”?
- Our *uncertainty* about what we’re sensing is low enough
- But we can’t change the randomness, how do we change uncertainty?

Background – Claude Shannon

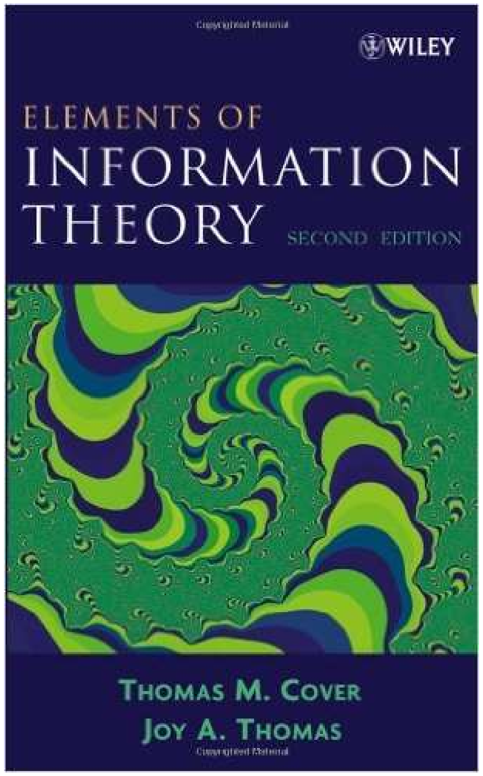
- "Father of the information age"
- Pioneer in CS, Information Theory, AI
- Data compression and decoding (codecs, zip files, etc)
- Wireless (and wired) digital communication

Works of Claude Shannon





Best Resource



Motivation: Models of Communication

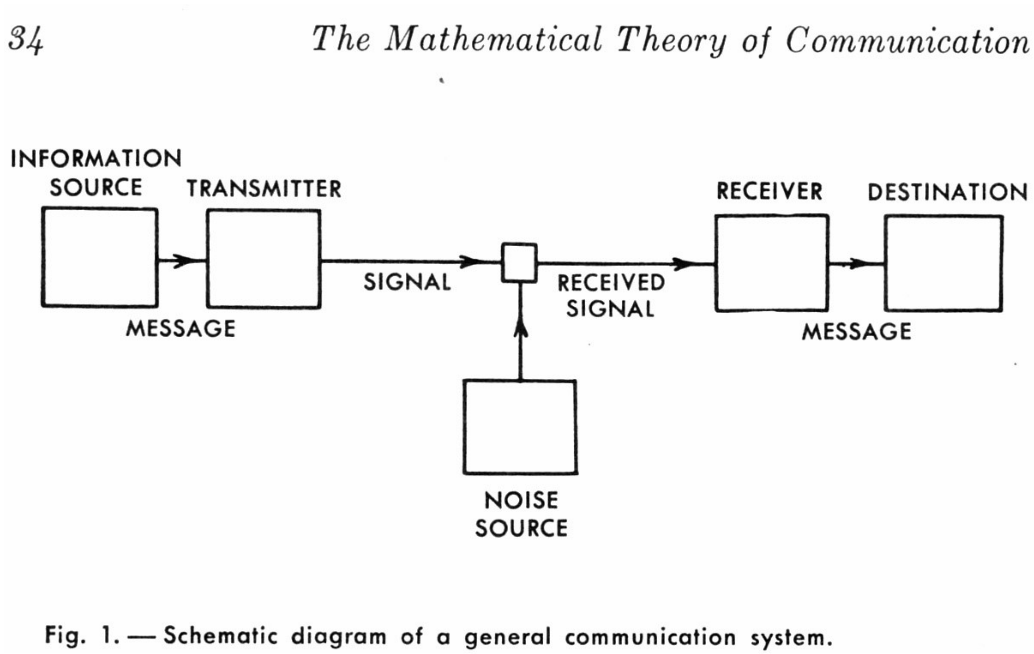
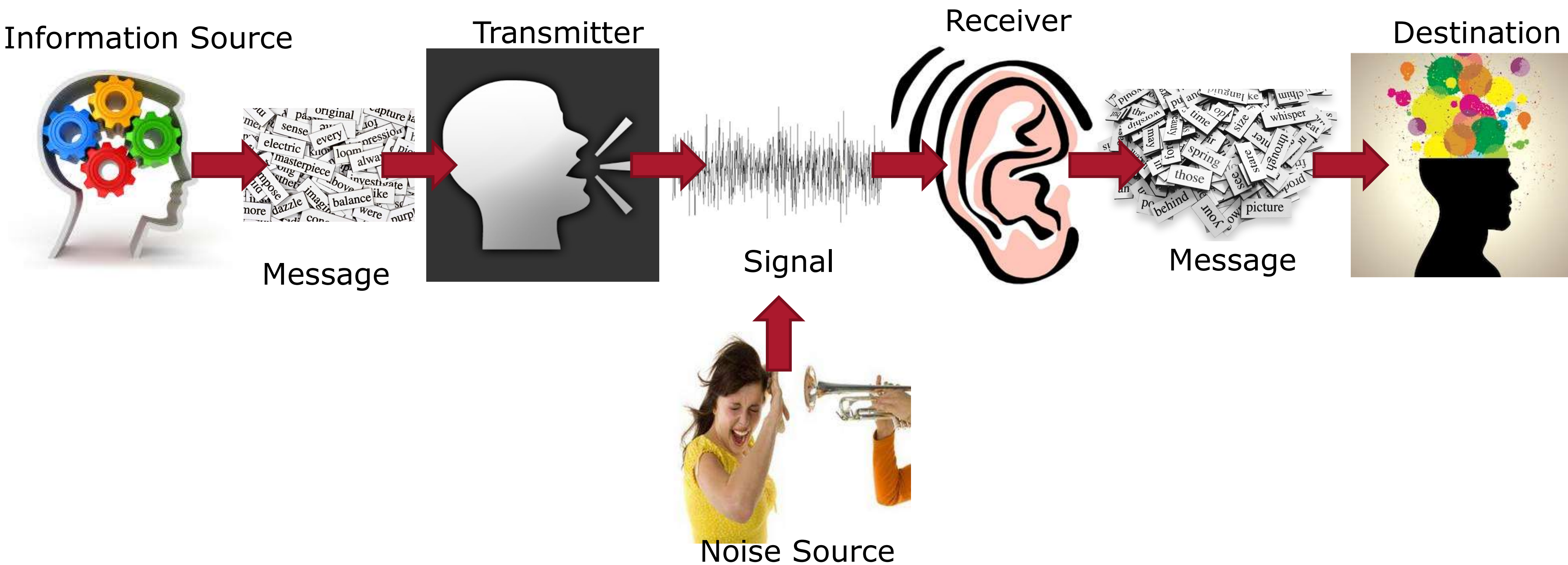

34 *The Mathematical Theory of Communication*

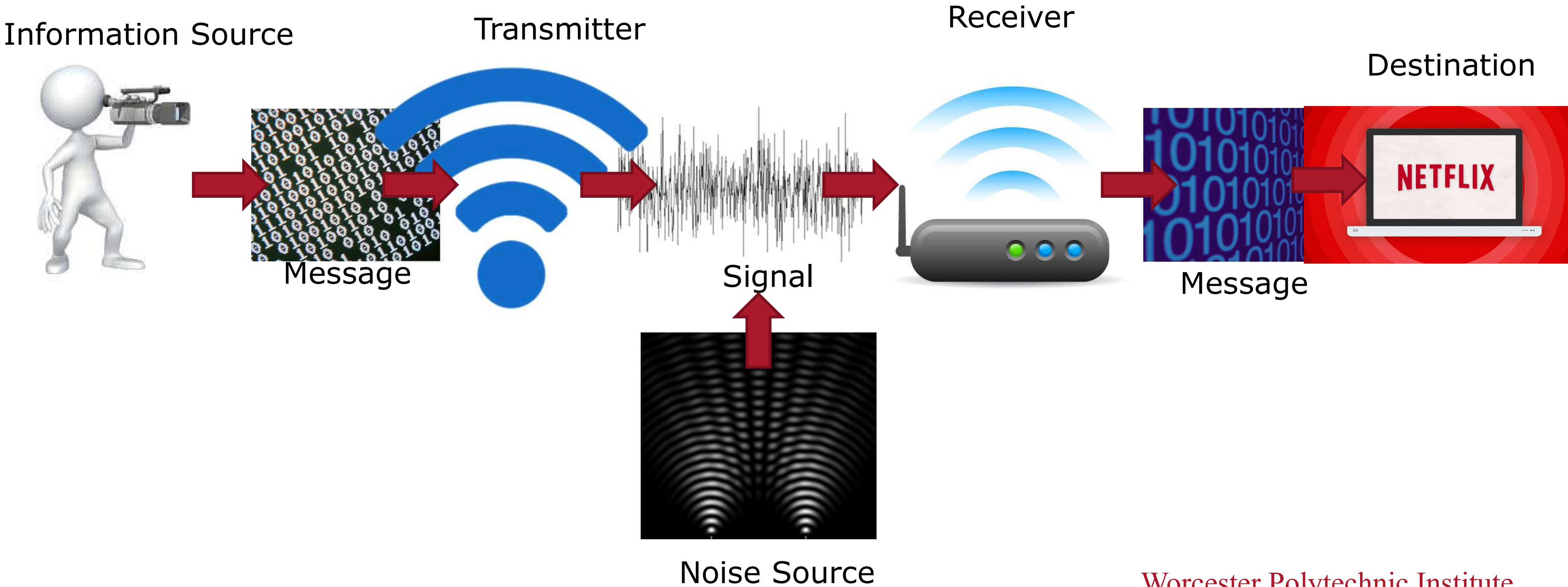
Fig. 1. — Schematic diagram of a general communication system.

Shannon 194

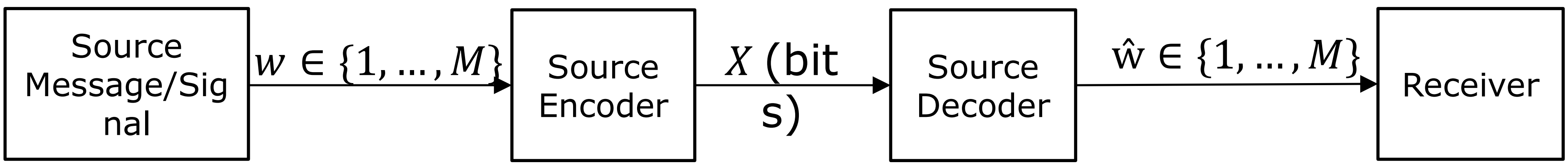
Example: Speech



Example: Video Streaming



Source Coding



Shannon 194

Motivating Example: Coin Flip

$\{0,1\}$

Motivating Example: Coin Flip

- Coin 1:
  - $P(x = 1) = 0.4; P(x = 0) = 0.6$

- Coin 2:
  - $P(x = 1) = 0.8; P(x = 0) = 0.2$

- Is one coin *more random* than the other?

- Can we generalize the difference between more than just biased coins?

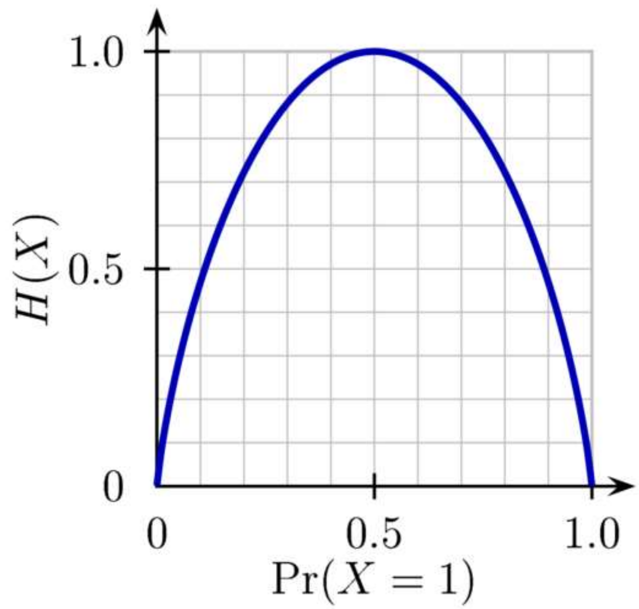
Entropy

- For a discrete R.V. $X \sim p_X(x)$ **entropy** is defined as:

$$H(X) = E[-\log_2(p_{X(x)})] = -\Sigma_{x \in X} p(x) \log_2 p(x)$$

- Quantifies "randomness" or "uncertainty" or "information content" of PMF
- Measured in bits
- Can also be measured in nats for $\log_e$ or other bases
- N.B. entropy is a function of a *probability distribution*
- N.B.2 this and other information measures can be extended to continuous RVs

Binary entropy



Some interesting properties

1. $H(X) \geq 0$
2. Maximized for uniform distribution: $H(X) = \log_2 |X|$
3. For uniform distribution, $H(X)$ increases monotonically in $|X|$
4. Any move toward uniformity increases $H(X)$

An additional aside about entropy

Some (hopefully) Helpful Intuition

- 1. Zero-order approximation (symbols independent and equiprobable).
  - XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.
- 2. First-order approximation (symbols independent but with frequencies of English text).
  - OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.
- 3. Second-order approximation (digram structure as in English).
  - ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

Some (hopefully) Helpful Intuition

- 4. Third-order approximation (trigram structure as in English).
  - IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.
- 5. First-order word approximation. Rather than continue with tetragram, … , n-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.
  - REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.
- 6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.
  - THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Some (hopefully) Helpful Intuition

- Zero-order approximation is uniformly random – highest entropy
  - XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.
- Exactly deterministic is very boring – lowest entropy
  - AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
- "Interesting" entropy is usually in the middle
  - THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Some (hopefully) Helpful Intuition

Increasing Entropy

The Big Bang
Deterministic
All matter condensed to a single point
Lowest Entropy
"Boring"

In Between
Interestingly random
Pretty good, all things considered
Medium Entropy
"Interesting"

The Heat Death of the Universe
Uniform
Highest Entropy
"Boring"

Back to the matter at hand

Joint and Conditional Entropy

- For two RVs, X and Y
- $H(X, Y) \ = \ -\sum_{x \in \mathrm{X}} \sum_{y \in Y} p(x, y) \log p(x, y)$
- $H(Y|X) = \ \sum_{x \in X} p(x) H(Y|X = x)$
- $\mathrm{H(Y|X)} = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log p(y|x)$
- $H(Y|X) \ \leq \ H(Y)$ – Conditioning reduces entropy! (on average)

Mutual Information

- Mutual Information is the relative entropy between joint and marginal PMFs of two RVs, X and Y
- $I(X;Y) = \sum_{x\in X} \sum_{y\in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$
- This measure always exists and is non-negative

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

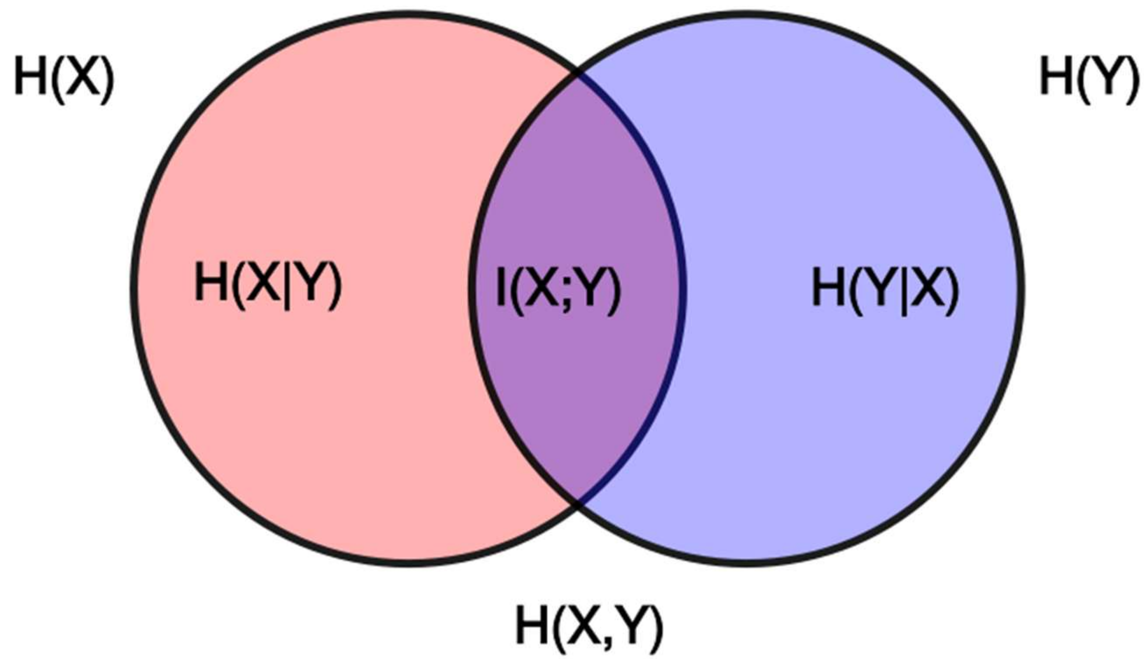
$$I(X;Y) = H(Y) - H(Y|X)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

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

$$I(X;X) = H(X).$$

Information Diagram



Numerical Example

$H(X)$

$Y \backslash X$	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

Numerical Example

$Y \backslash X$	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

$H(Y)$

Numerical Example

Y \ X	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

$H(X,Y)$

Numerical Example

$Y \backslash X$	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

$H(X \mid Y = 1)$

Numerical Example

$Y \backslash X$	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

$H(X \mid Y = 2)$

Numerical Example

$Y \backslash X$	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

$H(X \mid Y)$

Numerical Example

$Y \backslash X$	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

$I(X;Y)$

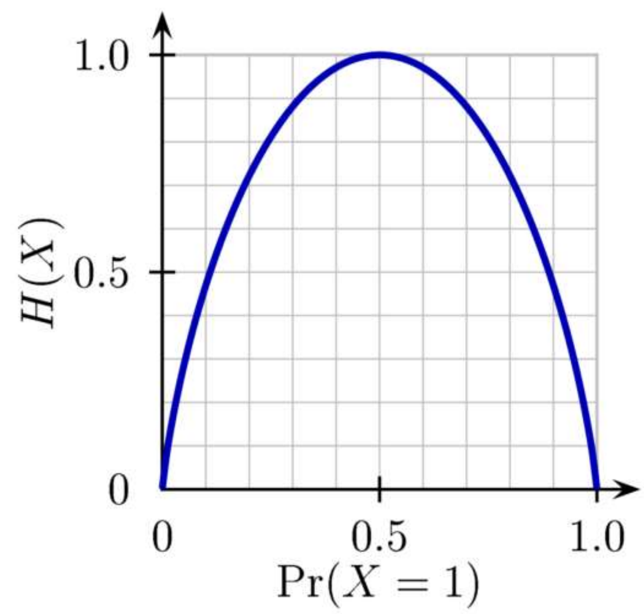
Why do we care?

- Sensor placement can be viewed as a problem of mutual information
- Given a random variable we want to measure, how much we can reduce the uncertainty
- Let’s return to sensor placement

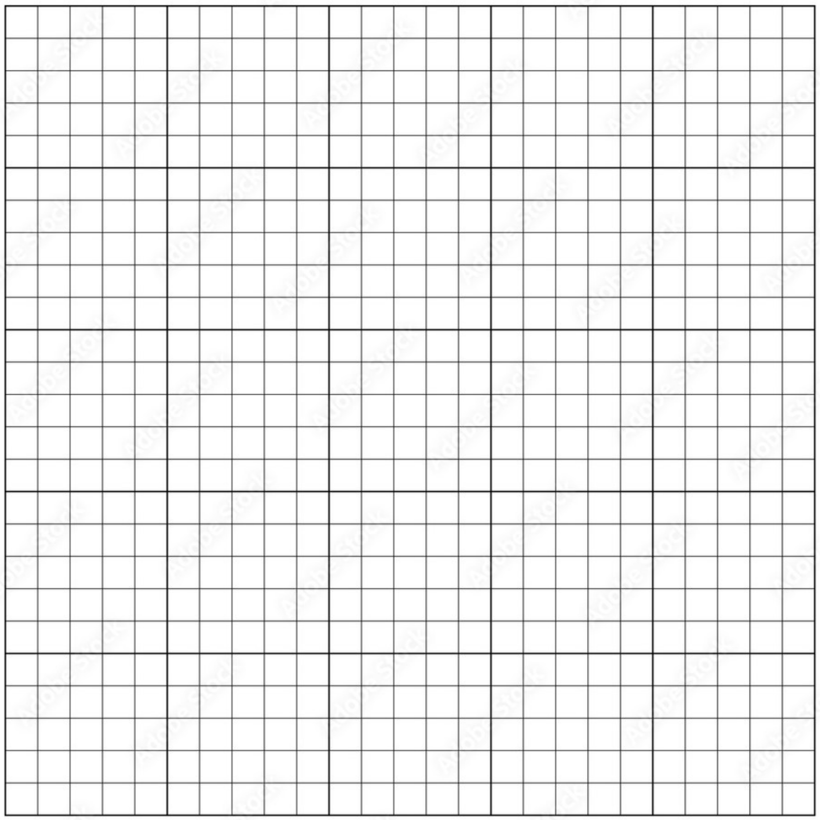
Recall

- Belief forms a simplex
- PMF of a random variable *also* forms a simplex
- Let’s use our belief like a probability distribution

Entropy and Belief



Recasting Our Problem



- Environment with finite set of possible locations
- Random variables Ω associated with each location
- Set of k sensors to place in the environment

$$X_k^* = \arg \max_{X \subseteq \Omega, |X| = k} I(X; \Omega \setminus X)$$

- Selecting the "most informative" locations

Entropy Reduction

- A good goal – reduce entropy via our measurements

$$H(X \mid \lambda) - H(X \mid Y, \lambda)$$

- Similarly

$$I(X; Y \mid \lambda)$$

- If we have the right probabilities, can compute *expected* entropy reduction given the sensor placement

Submodularity

Submodularity

- Popularized by Krause and Guestrin; Also Stefanie Jegelka, among others

Submodularity

- For a finite set Ω, a set function $f: 2^{\Omega} \to \mathbb{R}$ is a **submodular set function** if
  - $\forall X, Y \subseteq \Omega$,
  - $X \subseteq Y$,
  - $\forall x \in \Omega \setminus Y$,

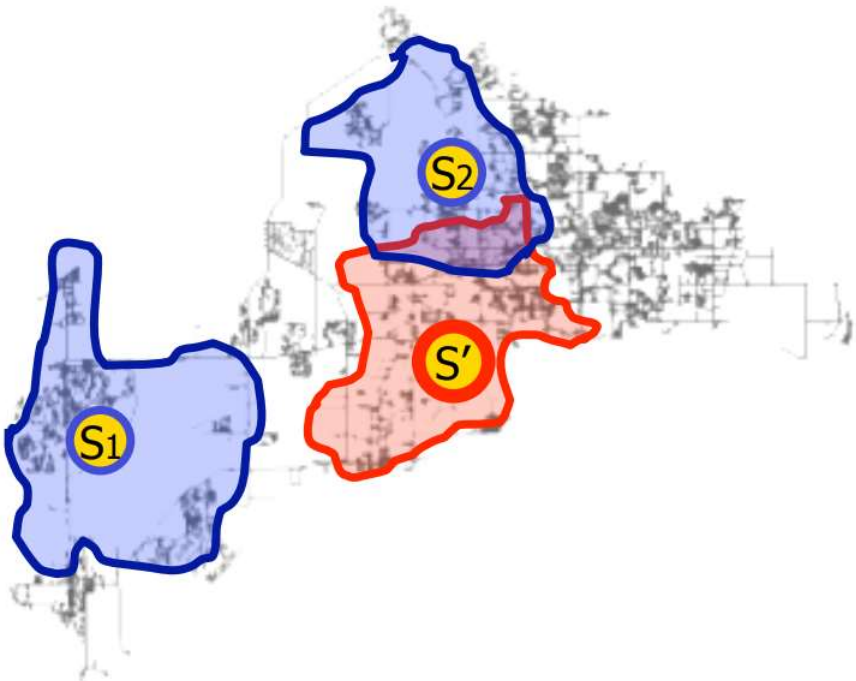
$$f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y)$$

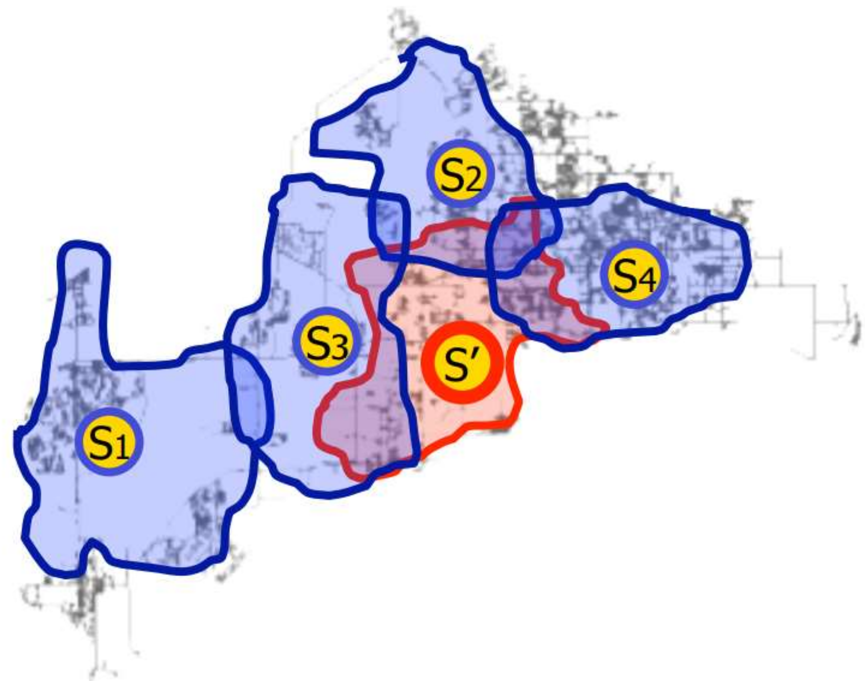
Submodularity (Equivalent Def.)

- For a finite set Ω, a set function $f: 2^{\Omega} \to \mathbb{R}$ is a **submodular set function** if
  - $\forall X \subseteq \Omega$,
  - $\forall x_1, x_2 \in \Omega \setminus X$,

$$f(X \cup \{x_1\}) + f(X \cup \{x_2\}) \geq f(X \cup \{x_1, x_2\}) - f(X)$$

Submodularity Examples





 Source: Krause and Golovin, “Submodular Function Maximization”

Greedy Maximization

- Start with the empty set $S_0 = \emptyset$

- At i^{th} iteration, select

$$S_i = S_{i-1} \cup \{\text{argmax}_e \Delta(\, e \mid S_{i-1}\,)\}$$

Submodularity and the Greedy Approach

- Let f be:
  - Submodular over Ω
  - Monotone ($\forall X \subseteq Y \subseteq \Omega, \quad f(Y) \geq f(X)$)
  - $f(\emptyset) = 0$

- Then, choosing A_k (set of k elements chosen greedily) results in

$$f(A_k) \geq \left(1 - \frac{1}{e}\right) f\left(A_{opt}\right)$$

- Where A_{opt} is the globally optimal selection

- Credited to Nemhauser et al. 1978

Submodular Maximization Intuition

- Proof sketch of greedy suboptimality – details in Nemhauser and Krause and Guestrin

Our objective function

- Is our objective function submodular?
  - Yes!

A tiny little asterisk

- As shown by Krause and Guestrin, entropy, mutual information, etc. are not monotone
- Compromise with a tiny fudge factor

$$f(A_k) \geq \left(1 - \frac{1}{e}\right)\left(f\left(A_{opt}\right) = k\epsilon\right)$$

Analysis

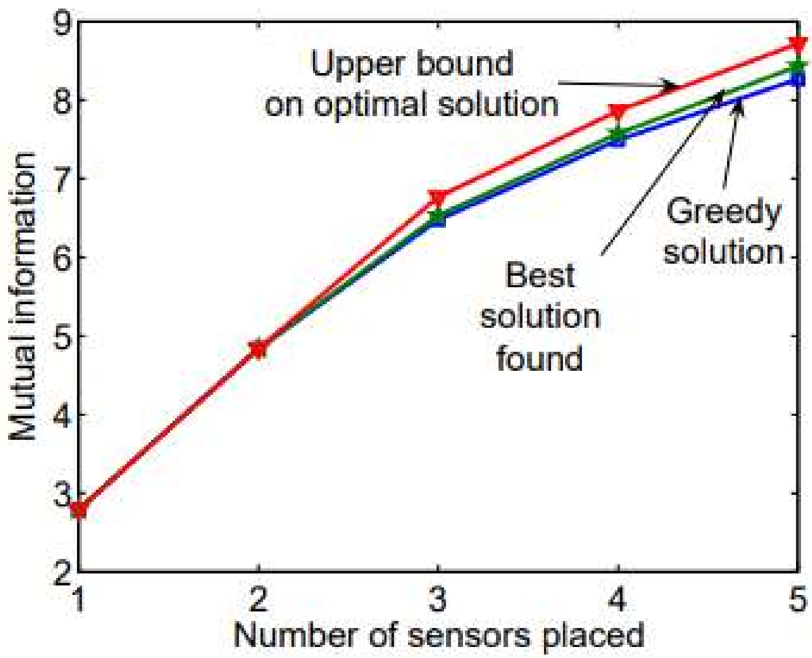
Complexity vs Optimalty

- Brute force approach
  - $\binom{\Omega}{k}$ combinations to try
  - But, guaranteed optimal
  - Only feasible for small combinations

- Heuristic approach
  - k locations selected greedily ($O(\Omega k)$, depending on greedy mechanism)
  - Suboptimal within factor of $\left(1-\frac{1}{e}\right) \approx 63\%$ of optimal
  - Very scalable

- Both approaches are *centralized*

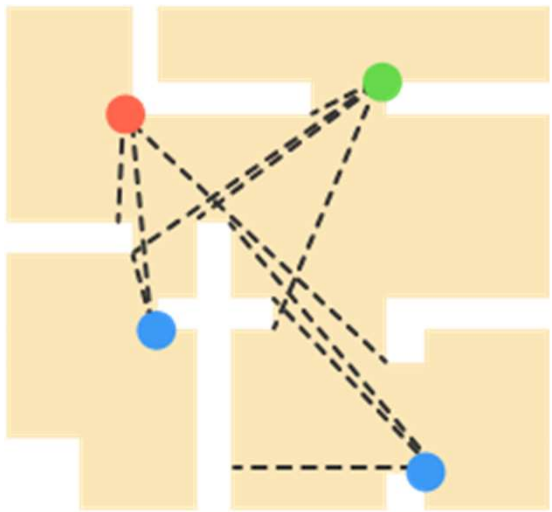
Optimality


Upper bound
on optimal solution
Greedy
solution
Best
solution
found
Mutual information
Number of sensors placed
9
8
7
6
5
4
3
2
1
2
3
4
5

Similar Problems

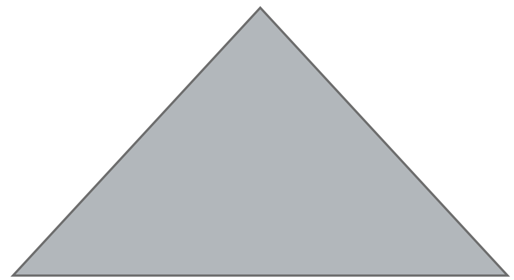
Art Gallery Problem

- Given a floor plan of an "art gallery"
- Agent model with line-of-sight visibility
- What is the minimum number of "guards" you can place that covers the whole area?



Art Gallery Problem

- Any floore plan with n walls
- At most $\left\lfloor \frac{n}{3} \right\rfloor$ guards
- Proved by Chvatal in 1975
- Useful for maintaining connectivity in non-convex environments
- Many variants – range-limited, moving guards (“patrolling”), etc.



Wrap Up

Recap

- Introduced sensing problems
  - Probability review
  - Sensor models
  - Belief
  - Information theory

- Static sensor placement
  - Problem formulation
  - Sensor model
  - Useful heuristic

Next Time

- Coverage and deployment
  - Mobile robots executing controllers for sensing objectives
  - Distributed controllers