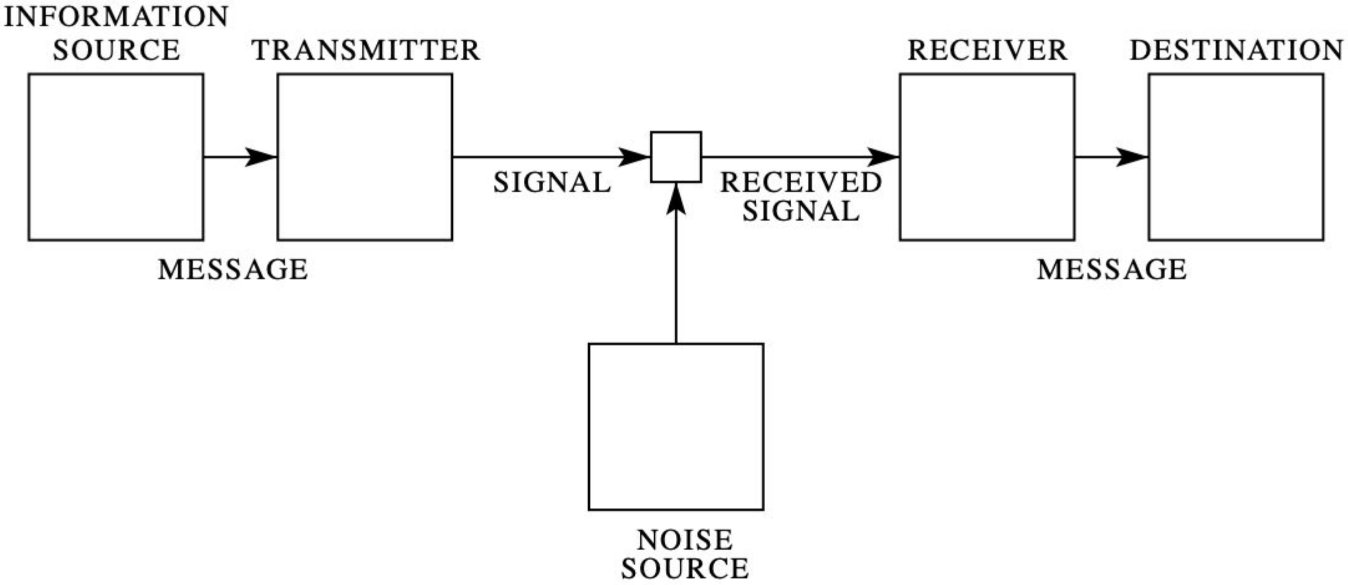
A Mathematical Theory of Communication

Discrete Noiseless Systems

Dongxu Liu, Sahar Moghtaderi, Debarghya Barik,
Rikuto Momoi

Communication system



Communication System Classify

- Discrete

  Both message and the signal are a sequence of discrete symbols, e,g, Telegraphy

- Continuous

  Both message and the signal are treated as continuous functions, e.g., Radio

- Mixed

  Both discrete and continuous variables appear, e.g., Speech transmissed to bits

Measure of information

Number of possible messages, and its monotonic function

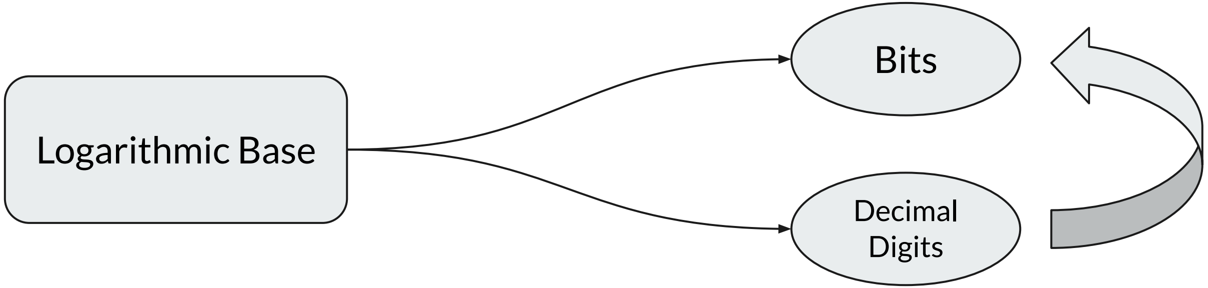
$$I = \log_2 N \qquad\qquad I = -\log_2 p$$

logarithmic function:

1. Useful, linearly with time, bandwidth…
2. Intuitive; we like linear comparison
3. Mathematically suitable

Teletype: 32 symbols set, 5bits.
ASCII: 2^7 bits

Unit of information measure



Capacity of Channel

- Channel: Medium used to transmit signal, e.g., wires, a band of radio frequencies.
- Discrete Channel: Transmit a sequence of choices from a finite set of elementary symbols, e.g. Tletype, Telegraphy
- Noiseless? Only consider the maximum capacity

Capacity of Channel

How many information can the channel transmit in a given time period?

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T}$$

$$N(T) = \text{number of sequences of duration } T$$

. THE DISCRETE SOURCE OF INFORMATION

information source produces messages

Information source have probabilistic structure.

THE SERIES OF APPROXIMATIONS TO ENGLISH

- **zero-order approximation,**

  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZL

- **first-order approximation,**

  OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTT

THE SERIES OF APPROXIMATIONS TO ENGLISH

- **second-order approximation,**

  ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TU-COOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

- **third-order approximation,**

  IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONS-TURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

- **second-order word approximation,**

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

⇨ **natural language is not random**, but has a **statistical structure**.

From Random Signals to Structured Sources

Shannon's goal: mathematically describe how information is generated and transmitted.

Up to Chapter 3: definition about sources and channels.

In Chapters 4 and 5, source as a mathematical object - *a probabilistic machine that emits symbols — a Markov model.*

Chapter 4 – Markoff Process as a Source

The Source as a Finite-State Machine

- Each 'state' remembers the previous symbol.
- Arrows -> next symbols probability.

Fig. 1: A single state machine – a completely random generator where every letter is independent.

Fig. 2: Dependent letters (multi-state Markov chain).

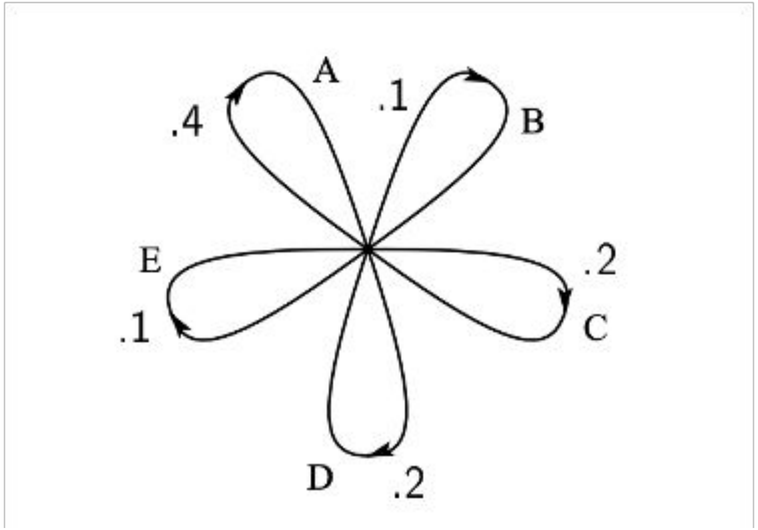


Fig. 1

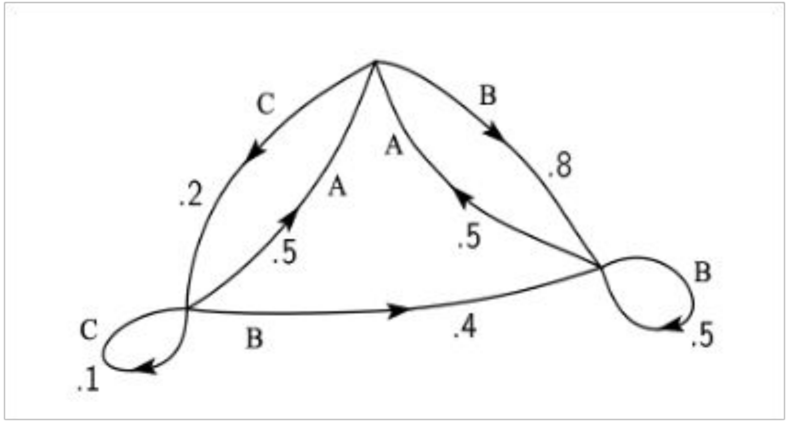


Fig. 2

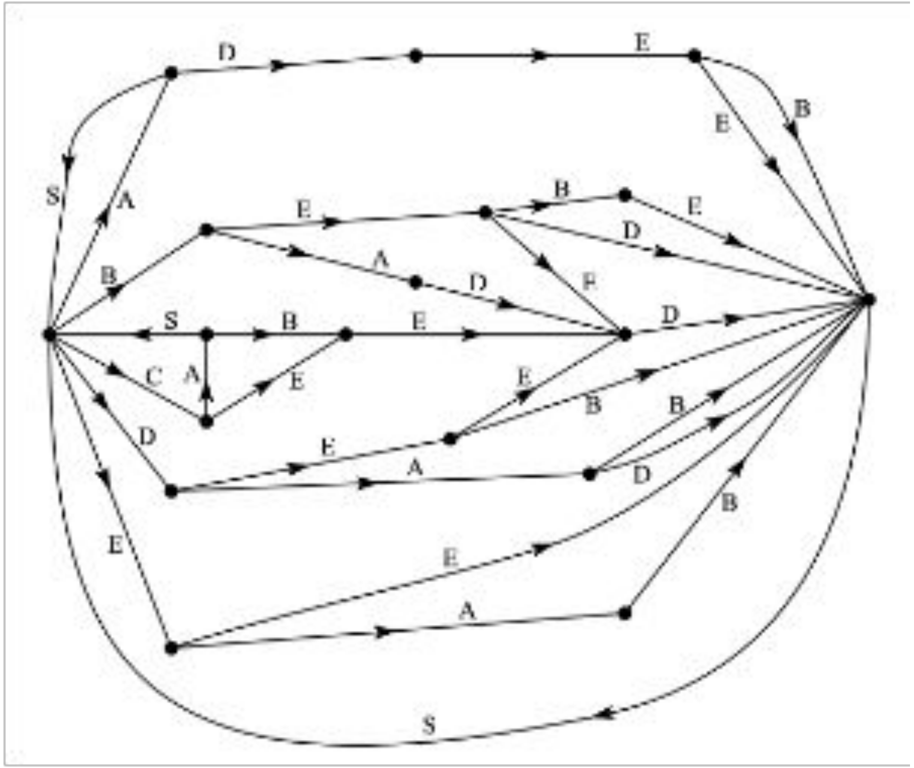
Chapter 5 – Ergodic Sources

When Is a Source Statistically Stable?

Ergodic -> long sequences always produces the same statistics — the same letter frequencies, the same transition counts.

- Graph must be:
  - **Connected** (can reach any state from any other)
  - **Aperiodic** (no fixed repeating cycle)

Non-ergodic → “mixed source” = combination of several independent subgraphs



An Ergodic Graph

Mixed Sources

When a Source Has Multiple Behaviors

If graph splits into separate disconnected parts i.e. multiple ergodic components -> source is mixed.

Consists of several mini languages or modes (self contained)

Mathematically:

- $L = p_1 L_1 + p_2 L_2 + \dots$
    - Li,…-> The **component sources**, each of which is **ergodic** — meaning statistically stable and self-contained.

      Pi,…-> The **probabilities (weights)** associated with each component source Li

So, Why It Matters? (Link to AI & Entropy)

Thus,

- entire logic behind probabilistic modeling.
- The finite-state source became the template for **HMMs** and **sequence decoders**.
- And once the source is stable, Shannon can define a single measure of uncertainty — the **entropy H** in Chapter 6.

Choice, Uncertainty and Entropy

Section 6

Measuring Information and Surprise

The Problem

Can we measure how much "surprise" or "uncertainty" there is in a situation?

Examples:

- Coin flip: 50/50 uncertainty chance
- Weather: 80% chance of rain
- Lottery: 1 in millions chance

More choices = More uncertainty?

THREE REQUIREMENTS FOR OUR "SURPRISE METER"

1. CONTINUOUS (Smooth)
   - No sudden jumps
   - Small changes in probability → small changes in uncertainty
2. MORE OPTIONS = MORE UNCERTAINTY
   - Fair coin (2 options) < Fair dice (6 options) < Lottery (millions)
   - When all probabilities equal
3. CAN BREAK DOWN COMPLEX CHOICES
   - Total uncertainty = sum of step-by-step uncertainties
   - Like breaking a big decision into smaller steps

THE ENTROPY FORMULA

$H = -\sum(p_i \times \log p_i)$

Breaking it down:

- H = Entropy (uncertainty measure)
- p_i = Probability of outcome i
- log = Logarithm (surprise level)
- $\sum$ = Sum all outcomes

Simple Example - Fair Coin:

$H = -[0.5\times\log(0.5) + 0.5\times\log(0.5)] = 1$ bit

Higher H = More uncertainty/surprise

Lower H = More predictable

FIGURE 7: ENTROPY CURVE

When p = 0 or 1: H = 0 (you know the outcome)

When p = 0.5: H = 1 (maximum uncertainty)

Real Examples:

- p = 0.1 (10% chance): H = 0.47 bits
- p = 0.5 (50% chance): H = 1.00 bits
- p = 0.9 (90% chance): H = 0.47 bits

Shape: Upside-down U (symmetric)

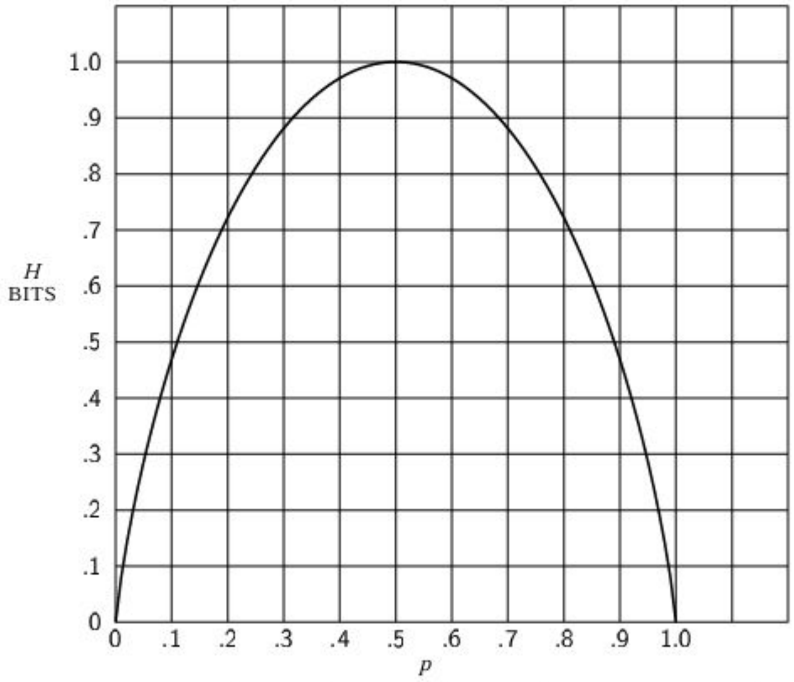


Fig. 7—Entropy in the case of two possibilities with probabilities p and $(1-p)$.

The Sweet Summary: Shannon's Revolution

How much information is in this sentence?

Before Shannon (1948):

Engineers: "Let's try this wire thickness..."

Communication: Trial and error

After Shannon:

One equation governs ALL communication

Information became measurable

The Sweet Summary:

How did he do it?

He separated information from meaning. That was the breakthrough.

RANDOM:	ENGLISH:
XFOML RXKHR	THE QUICK BROWN FOX
❌ No patterns	✅ Patterns everywhere
❌ Can't compress	✅ Can compress
❌ High uncertainty	✅ Low uncertainty

real messages have structure.

The Sweet Summary:

The Impact

1948 → Shannon's Paper

↓

1960s → Error Correction

1970s → Digital storage (compression)

1980s → CDs & digital music

1990s → Internet & modems

2000s → Smartphones & streaming

2020s → AI & everything digital

"Information is the resolution of uncertainty"

\- Claude Shannon

Thank you! / Questions?