A Brief Introduction to Information and Communication Theory

Ayfer Ozgur,

Stanford University

Disclaimer:

What is “science of information”?

knowledge of communication of knowledge

communication

Communication: information in action

The Bell System Technical Journal

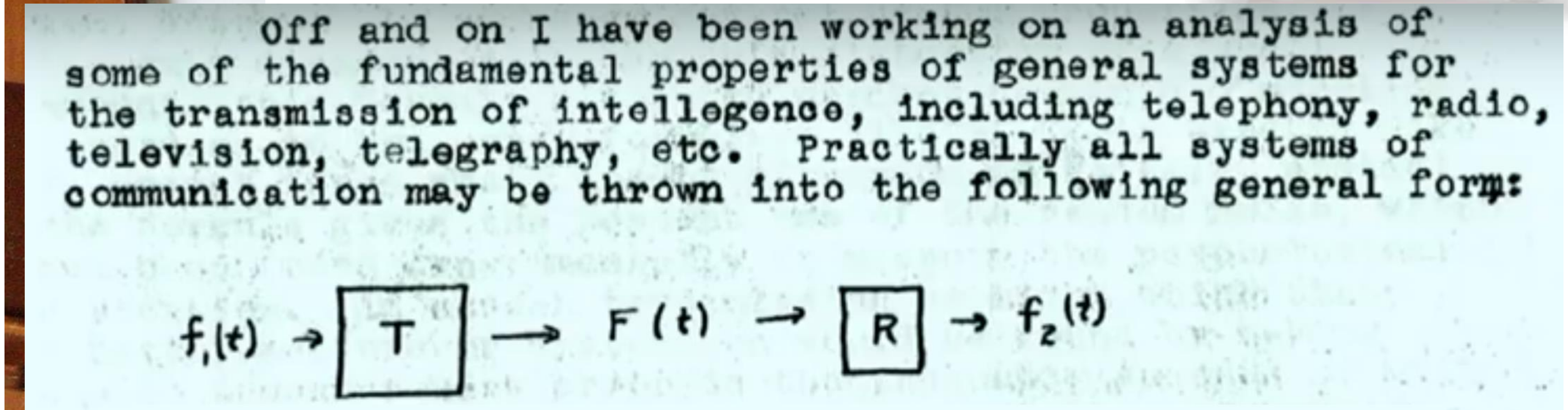
Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

Introduction

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2]



Off and on I have been working on an analysis of some of the fundamental properties of general systems for the transmission of intellegence, including telephony, radio, television, telegraphy, etc. Practically all systems of communication may be thrown into the following general form:

Excerpt of a letter from Shannon to Bush. Feb. 16, 1939. From Library of Congress

A Mathematical Theory of Communication

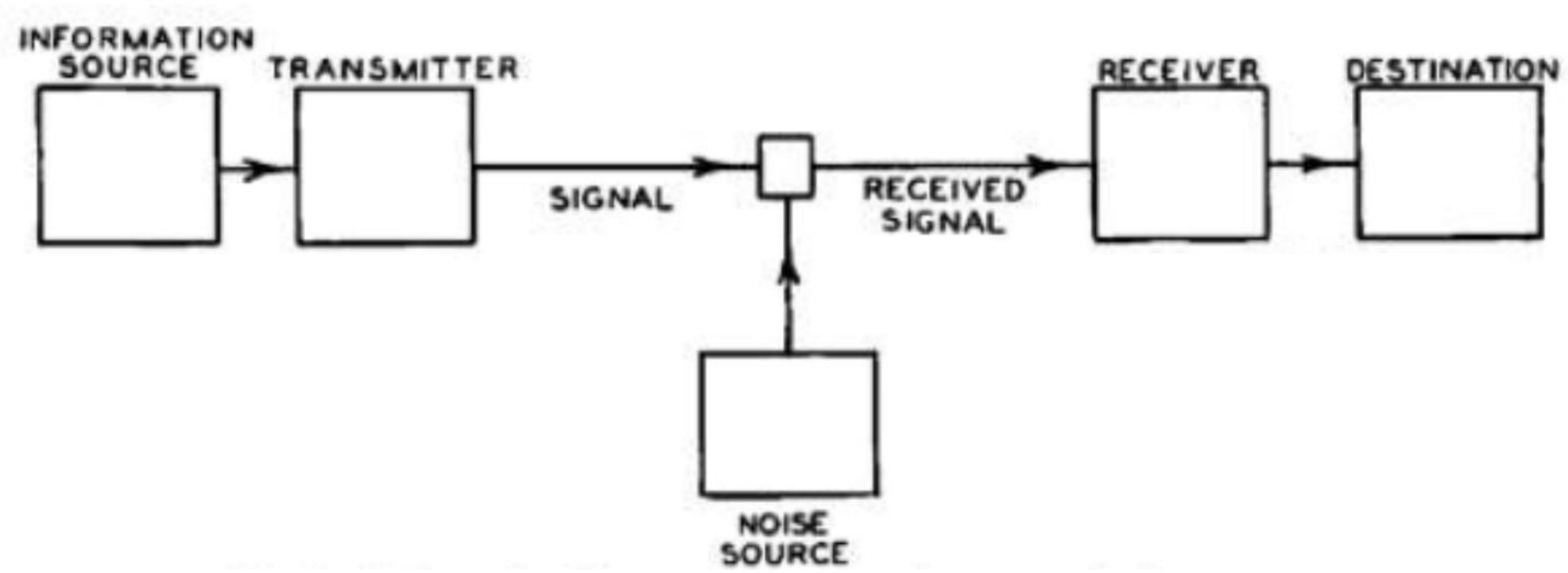
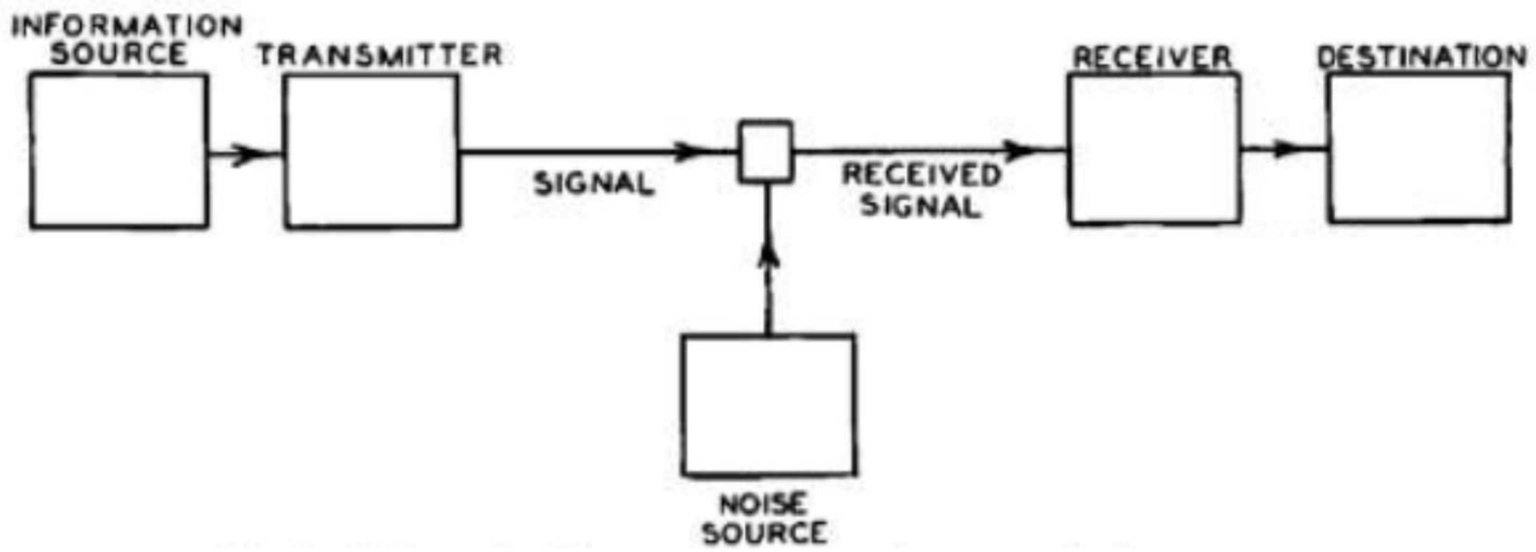


Fig. 1 —Schematic diagram of a general communication system.

A Mathematical Theory of Communication



- What is information? How can we measure information?
- How do we mathematically model a noisy channel?

C.E. Shannon (1916-2001)

- Born on April 30, 1916 in Northern Michigan.
- Bachelor’s degrees in Electrical Engineering and Mathematics from University of Michigan, 1936.
- Joins MIT as a research assistant to work on a differential analyzer under the supervision of Professor Vannevar Bush.
- His MS thesis “A Symbolic Analysis of Relay and Switching Circuits” forms the foundation of digital circuit design and the modern computer, 1937.
- Completes PhD Thesis *“An Algebra for Theoretical Genetics” in 1940.*
- Bell Labs, 1941-1956.
- Faculty at MIT, 1956-1978.

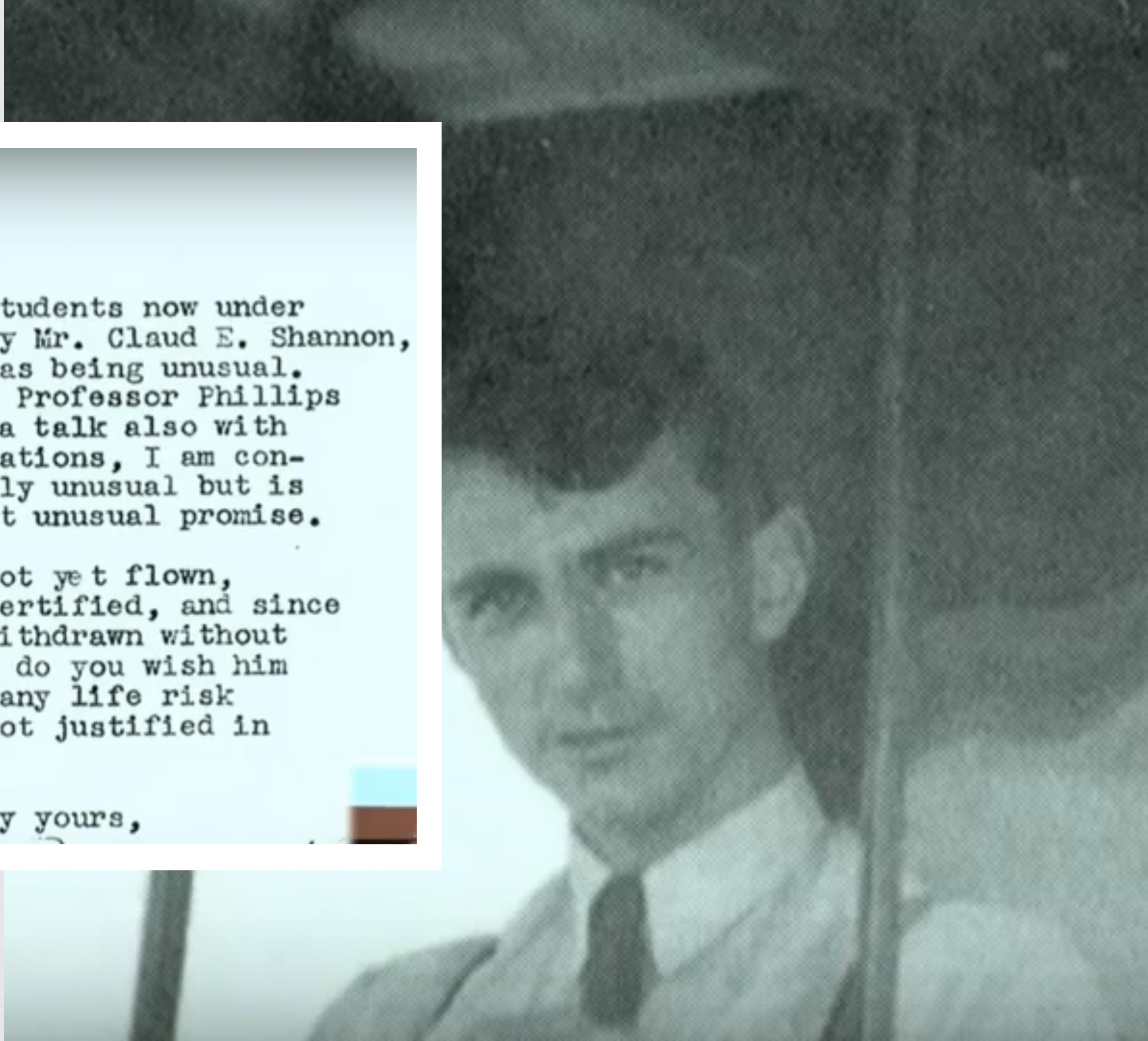
President
M.I.T.

Dear Doctor Compton:

One of our twenty students now under C.A.A. Flight Training, namely Mr. Claud E. Shannon, XVIII G, rather impressed me as being unusual. I, therefore, had a talk with Professor Phillips about him and a little later a talk also with Caldwell. From these conversations, I am convinced that Shannon is not only unusual but is in fact a near-genious of most unusual promise.

Since Shannon has not yet flown, being one of the last to be certified, and since he may, therefore, still be withdrawn without obligation to the Government, do you wish him withdrawn on the ground that any life risk whatever, however small, is not justified in his case?

Very truly yours,



How do we measure information?

Suppose you ask a question and get an answer. How do you tell if the answer involves a small amount or large amount of information?

- Depends on the semantics and context.
- Depends on the length of the answer.
- Depends on how many bits we need to represent/store the answer.

Shannon: Information is the ability to distinguish reliably among possible alternatives.

How do we measure information?

M: number of possibilities

- If M is small, the amount of information in the message/answer is small.

  Ex. Did you travel abroad over the summer? Yes/No

- If M is large, the amount of information in the message/answer is large.

  Ex. Which country did you travel to over the summer? ~200

How do we measure information?

Take the logarithm to make information additive:

$$\textbf{Entropy: } \boldsymbol{H} = \log_{\mathbf{2}} \boldsymbol{M}$$

Entropy can be also viewed as a measure of uncertainty:
it represents our uncertainty about the message before we receive it.

- If M=1, H=0 bits.
- If M=2, H=1 bits.
- If we have two binary messages, M=4, H=2 bits.

Entropy and Representation

The number of bits n we need to represent a message:

n ≥ H

Ex: How about when M=5? $H = \log_2 5 \ = \ 2.32$

Block Coding: Consider encoding blocks of messages generated by this source:

- Blocks of 2: 2.5 bits per message.
- Blocks of 3: 7/3=2.33 bits per message

Entropy and Representation

Conclusion:

The entropy H of a source equals the minimum number of bits required to represent a message from this source.

Refining the definition of entropy

Very High Probability

Low Probability

The surprise factor

Assume we have a source with M possible outcomes $k = 1, \ldots M$ with corresponding probabilities $p(1), \ldots p(M)$. Surprise of outcome k can be defined as

$$s(k) = \log_2 \frac{1}{p(k)}$$

Fire Alarm:

	$p(k)$	$s(k)$
No Fire	0.999	0.014
Fire	0.001	9.97

A new definition of entropy

Entropy is the average surprise:

$$H = \sum_{k=1}^{M} p(k) \log_2 \frac{1}{p(k)}$$

- Coincides with our earlier definition of entropy when messages are equally likely.
- Fire alarm example:

$$H(Fire\ Alarm) = 0.024 \text{ bits}$$

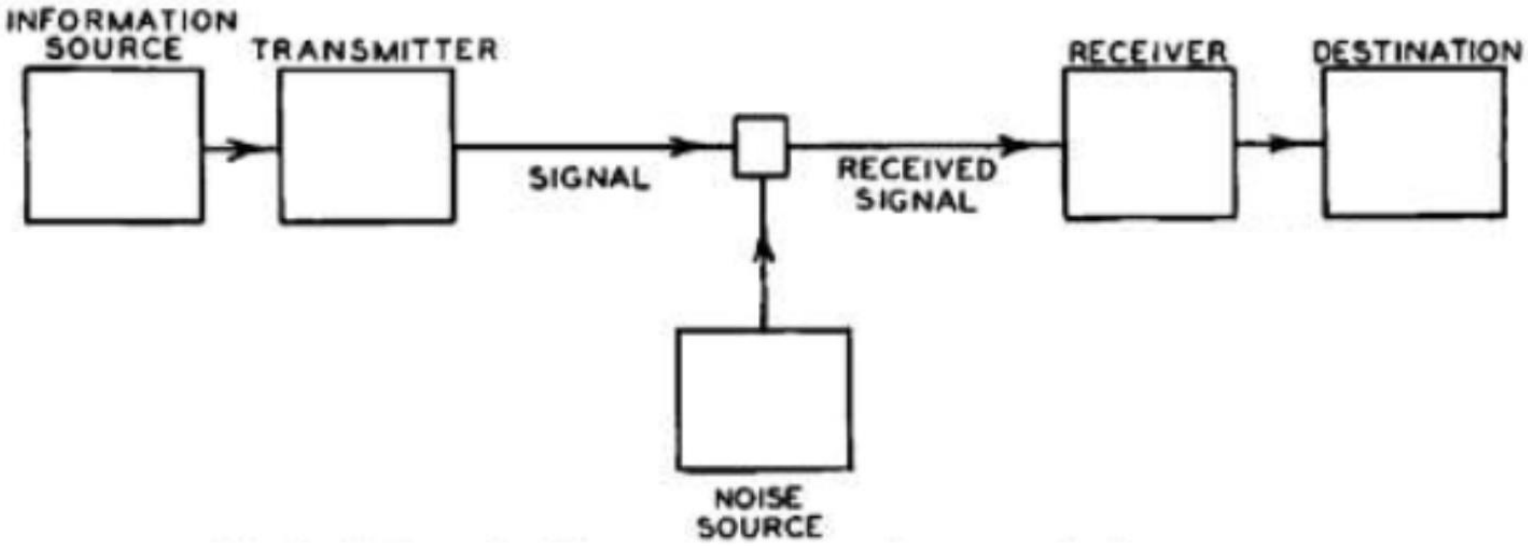
- More generally:

$$H \leq \log_2 M$$

Shannon’s First Fundamental Theorem

Source Coding Theorem: The entropy H(X) of a source (the average surprise) equals the minimum average number of bits necessary to code messages from that source.

A Mathematical Theory of Communication



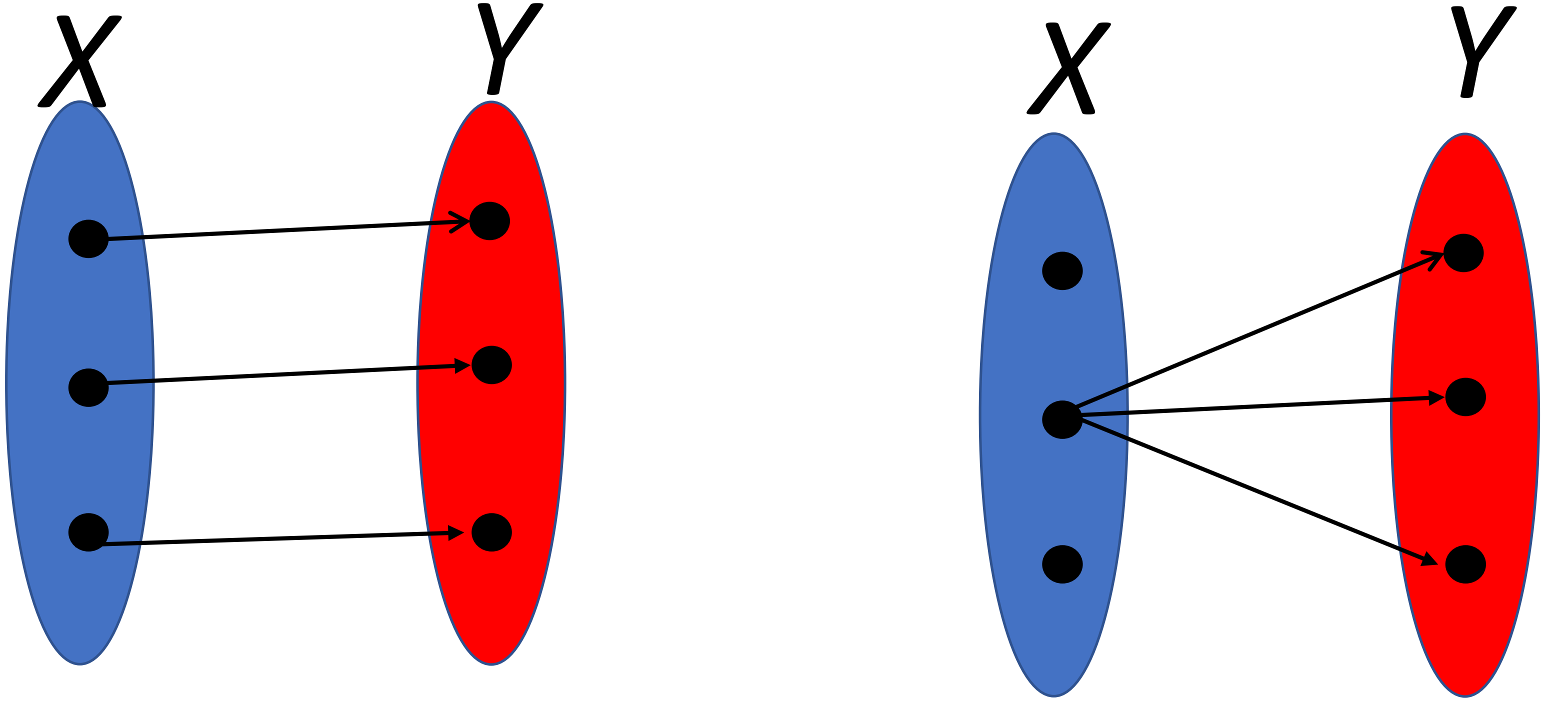
- What is information? How can we measure information?
- How do we mathematically model a noisy channel?

Communication Channel

The ordinary telegraph is like a very long cat. You pull the tail in New York, and it meows in Los Angeles. The wireless is the same, only without the cat.

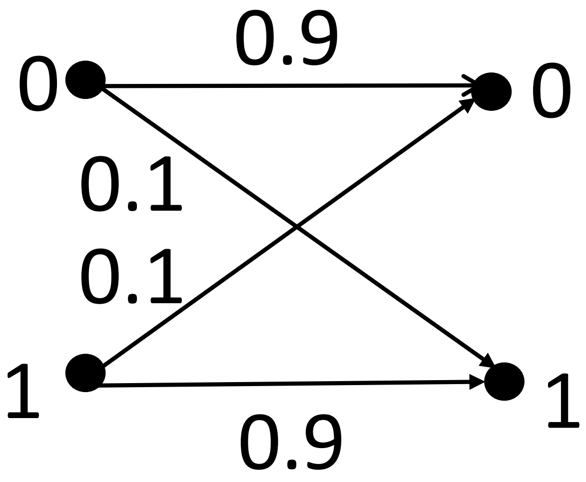
Albert Eistein

Channel Model

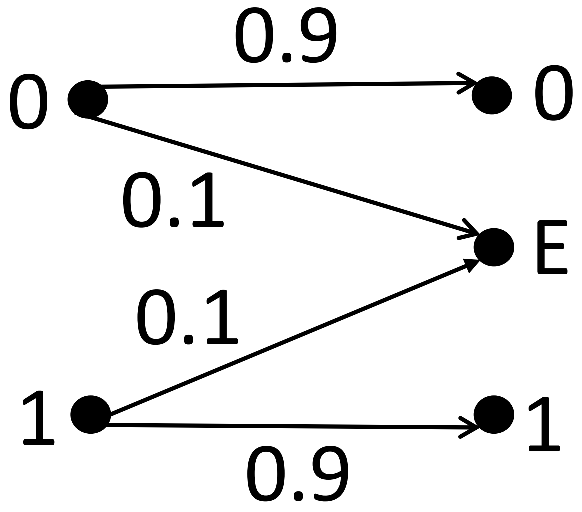


For each x in X, specify $p(y|x)$ for all y in Y

Examples



Binary Symmetric Channel (E=0.1)



Binary Erasure Channel (E=0.1)

Can you learn half a bit about a bit?

Assume we transmit a bit X (equally likely to be 0 or 1) over the binary symmetric channel, how much information does the receiver get about X?

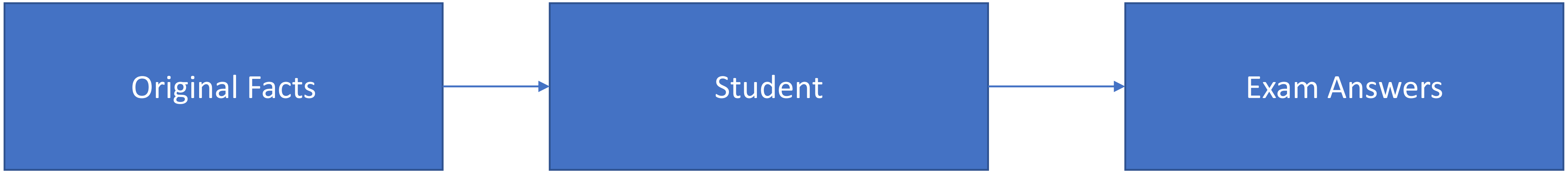
- H(X)=1 bit
- $H(X|Y) = 0.9 \log_2 \frac{1}{0.9} + 0.1 \log_2 \frac{1}{0.1} = 0.469$ bits
- $I(X;Y) = 1 - 0.469 = 0.531$ bits

How about if we communicate X over the binary erasure channel?

- H(X|Y)=0.1 bits, I(X;Y)=0.9 bits.

Learning as a communication channel

A final exam contains 100 Yes/No questions:



Which student knows more?

- Student A gets the correct answer 90% of the time and marks the wrong answer 10% of the time.
- Student B knows the correct answer also 90% of the time; but when she knows the answer, she invariably answers correctly, if she is not sure, she sure leaves the question blank.

Can we communicate reliably over noisy channels?

- If we communicate a bit over the binary symmetric channel (e=0.1),

  Probability of Error=0.1

- Repetition coding:

000

111

Probability of Error=0.028

00000

11111

Probability of Error=0.0086

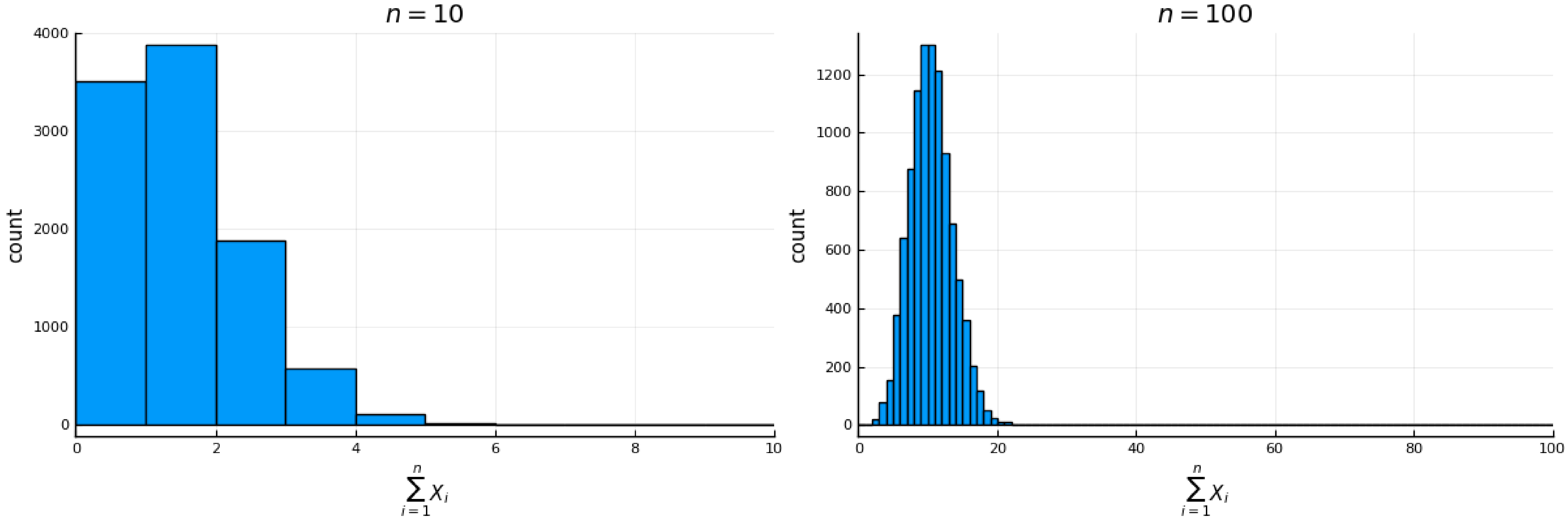
Shannon’s Second Fundamental Theorem

Channel Coding Theorem: There exists a transmitter and a receiver that make it possible to send C bits per second while keeping the error probability as small as desired. Conversely, if we try to send at a rate higher than C bits per second, then errors are inevitable. The threshold C is called the capacity of the channel and is given by

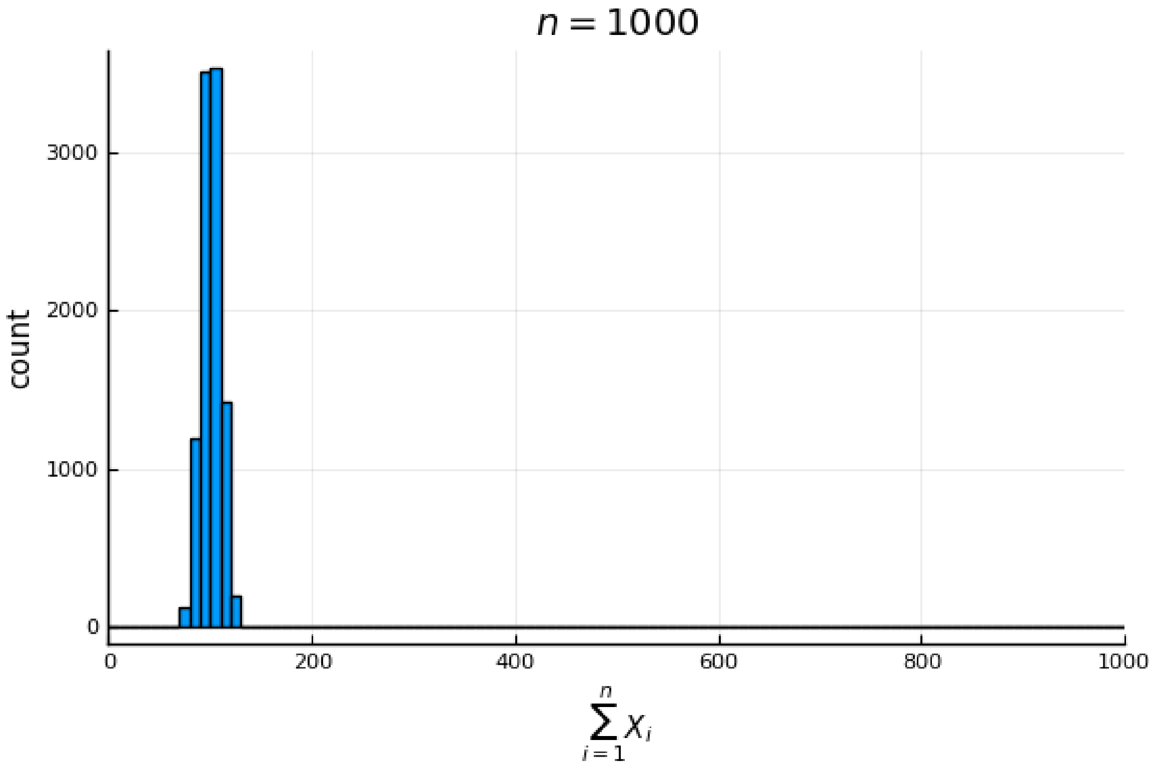
$$\mathbf{C = \max I(X;Y)}$$

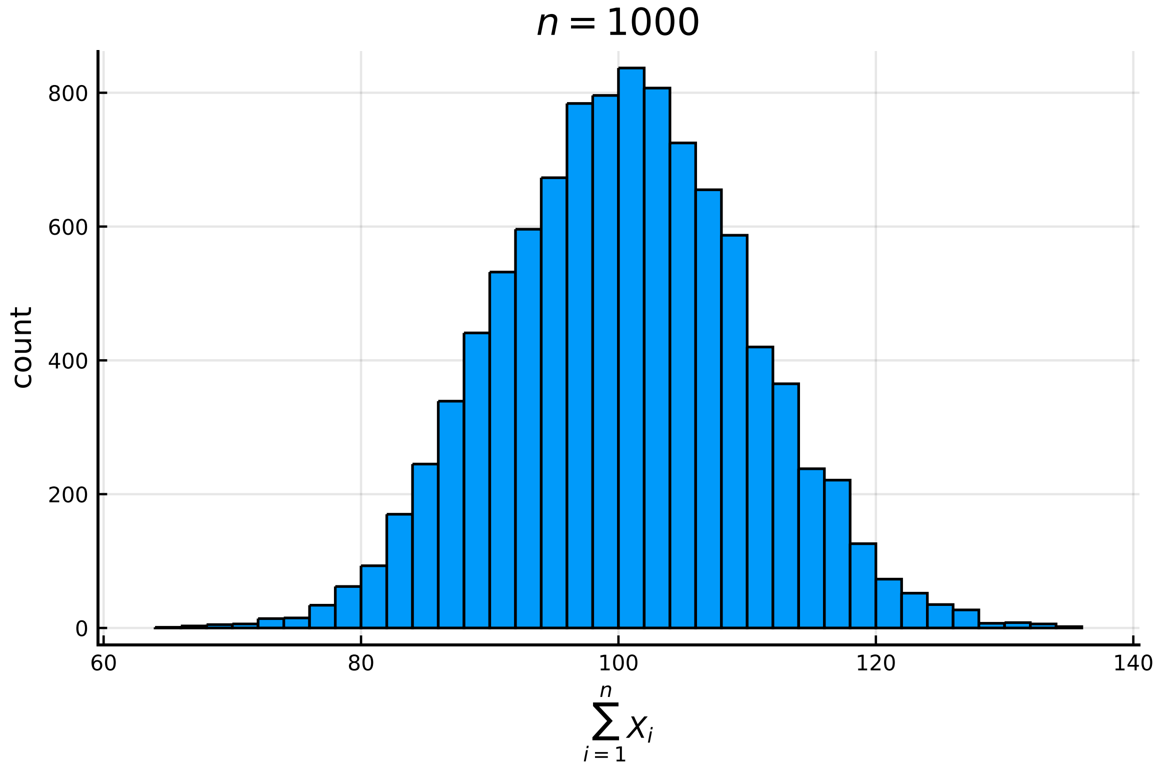
Binary Symmetric Channel (e=0.1): C=0.531 bits/channel use.
Binary Erasure Channel (e=0.1): C=0.9 bits/channel use.

Why is block coding good?

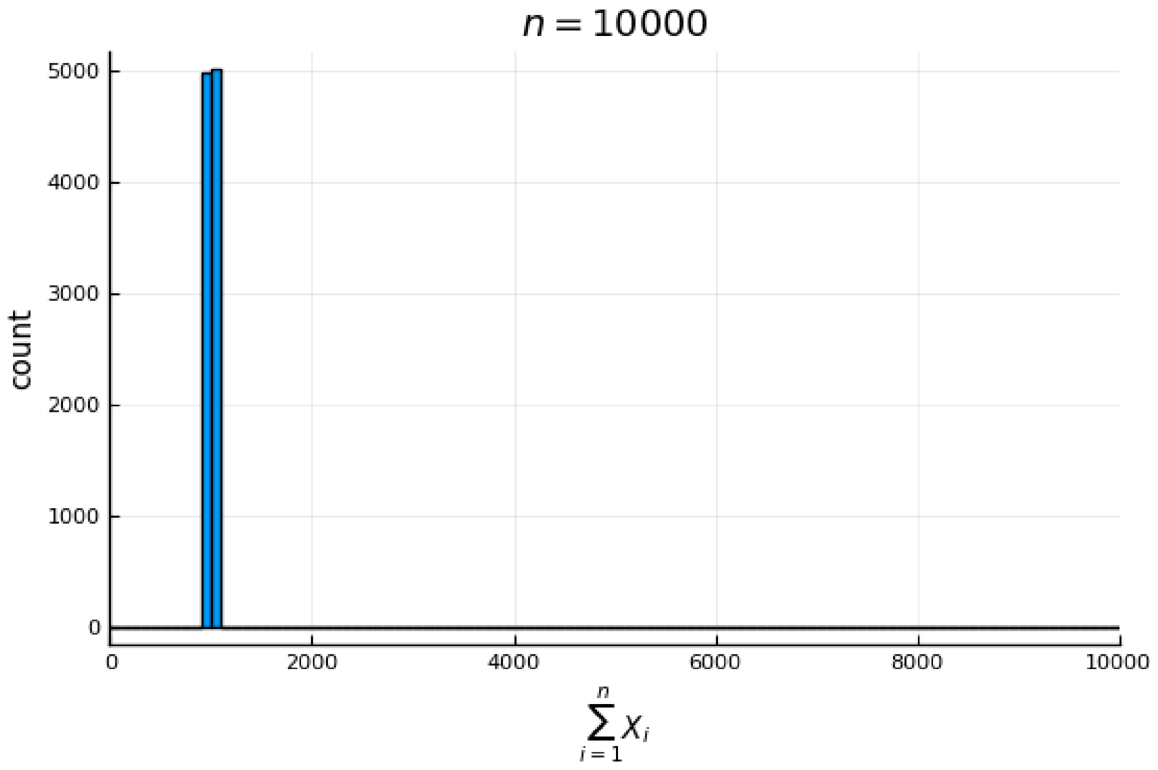


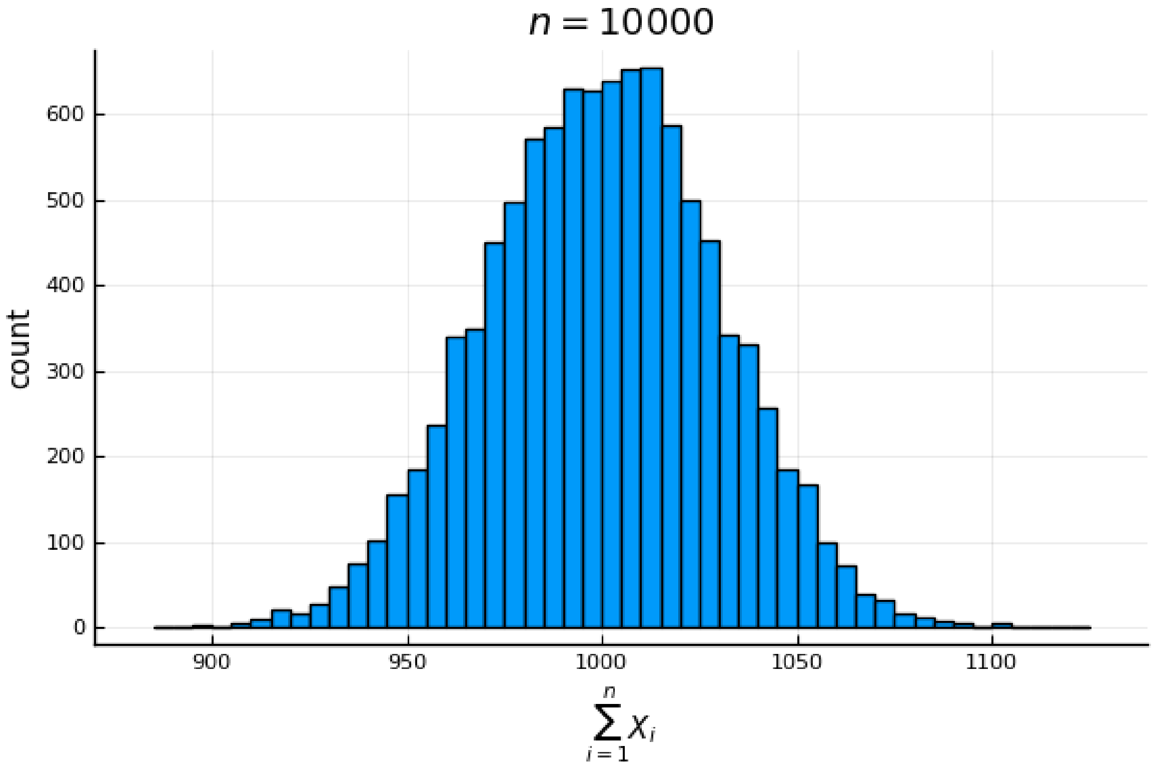
Why is block coding good?





Why is block coding good?



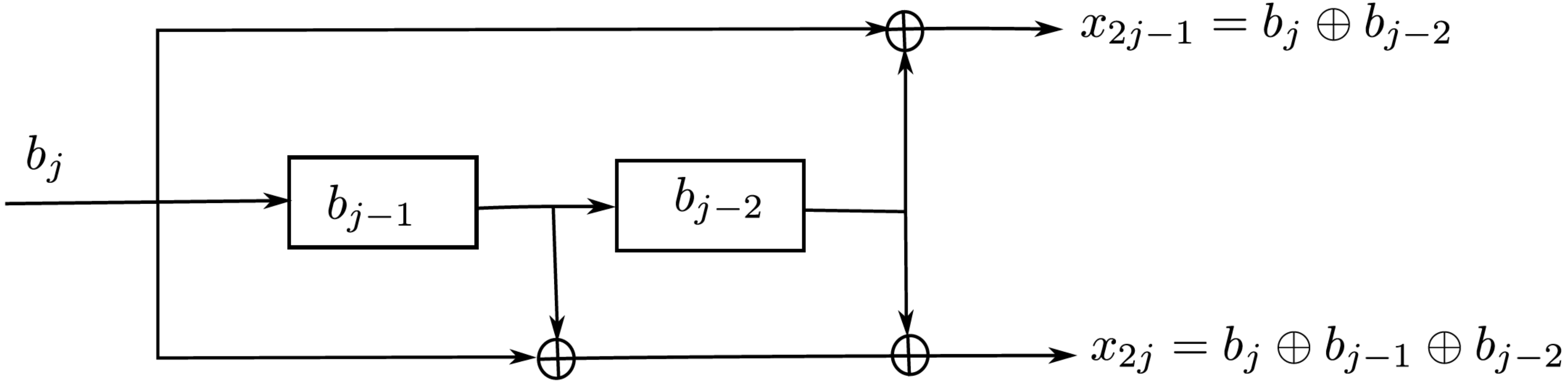


LAW OF LARGE NUMBERS

Complexity of Block Coding

- Assume $n = 100, M = 2^{100}$ is approximately 10^{30}
- Using this approximation, a VLSI chip that makes 10^9 inner products per second takes 10^{21} seconds to check all possibilities. This is roughly 4 x 10^{13} years.
- The universe is "only" roughly 2 x 10^{10} years old!

A simple convolutional code



Rate=0.5 bits/channel use

Probability of error=0.07

The bit as the universal information currency

So far we talked about

- how we can represent sources by a binary sequence.
- how we can reliably communicate long sequences of bits over noisy channels.

Information Sources:

- Discrete Sources: produce symbols taking values in a discrete set.
- Continuous Alphabet Sources: produce symbols taking a continuum of values.
- Continuous-Time Sources: produce continuous-time signals

Shannon's Third Fundamental Theorem

- **Source Channel Separation Theorem:** If a source can be transmitted over a channel in any way at all, it can be transmitted using a binary interface between the source and the channel. [More generally, if a source can be transmitted over a channel with certain fidelity, then it can be transmitted with the same fidelity using a binary interface between the source and the channel.]

Bits have become the universal information currency as result of:

- the source channel/separation theorem.
- electronic circuits becoming more and more digital.
- a standardized binary interface between the source and the channel simplifies implementation and understanding.

From 1948 to 2018

Analog vs Digital Communication

Analog Communication:

- The message to be communicated is one of a continuum of possibilities.
- Can never fully remove the effects of noise.

Digital Communication:

- The message to be communicated is one of a finite set of possible choices.
- Can remove the effects of noise induced by the channel.