University of Delaware

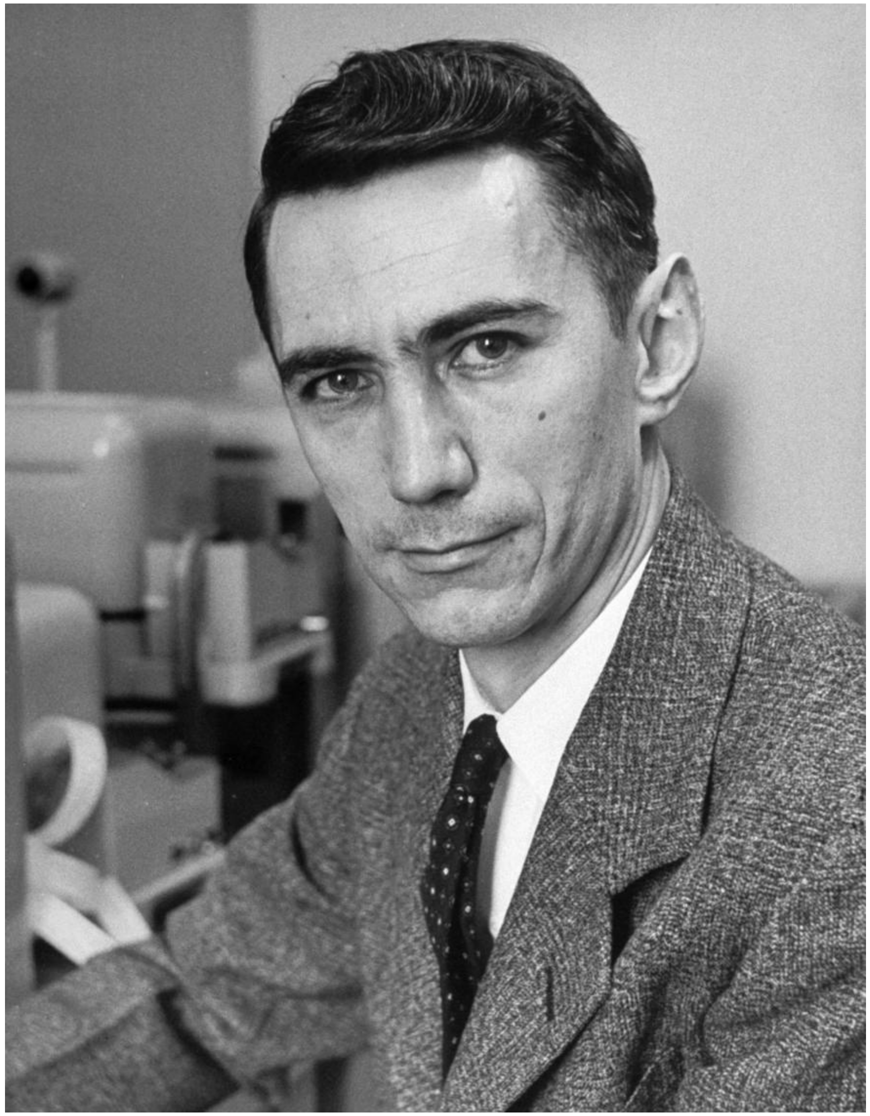
February 17, 2010

Information Theory Today

Sergio Verdú

Princeton University

Claude Shannon



1948

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] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are

Published in THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 27, pp. 379-423, 623-656, July, October, 1948

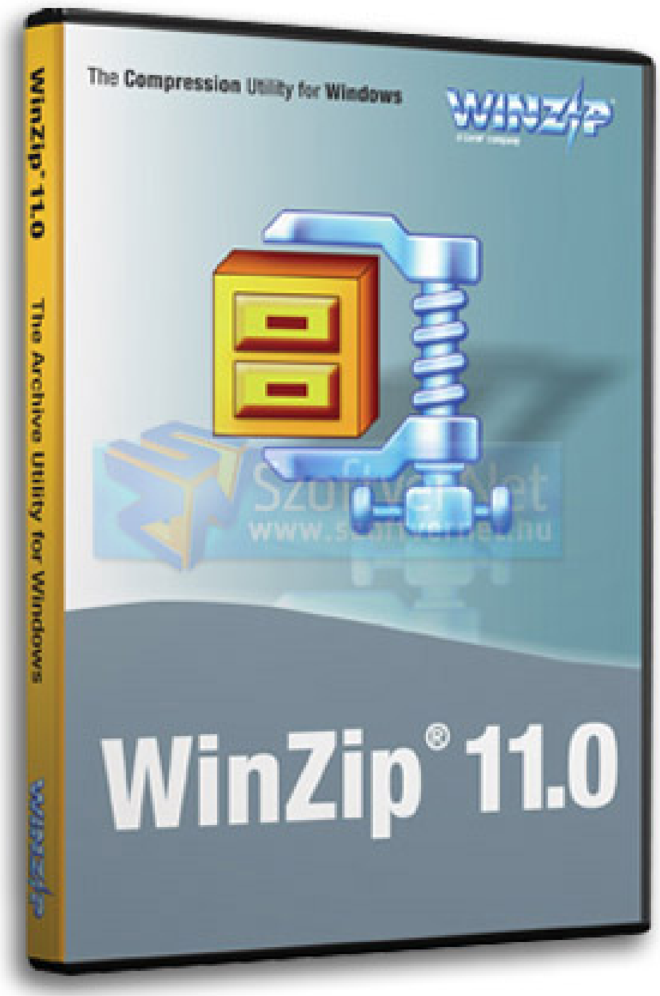
Information Theory

- **Fundamental Limits**
  - lossless data compression
  - lossy data compression
  - channel coding
  - complexity of simulation
  - portfolio allocation
  - decisions

- **Information Measures**
  - Shannon theory
  - ergodic theory
  - probability and statistics
  - physics, economics, neuroscience...

- **Engineering Design Driver**

Lossless Data Compression

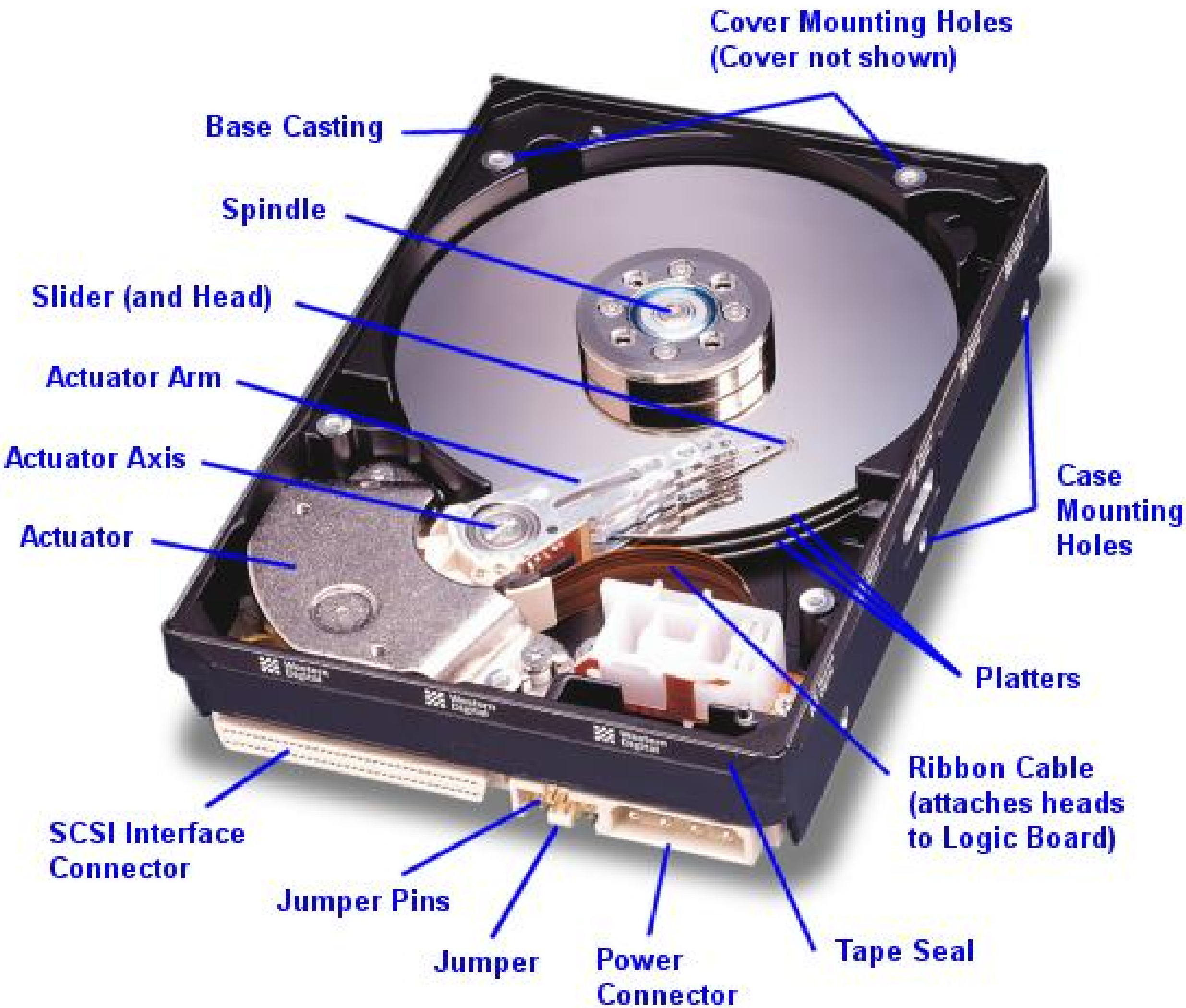


Lossy Data Compression

DVD VIDEO

Error Correction Codes: Compact Disc

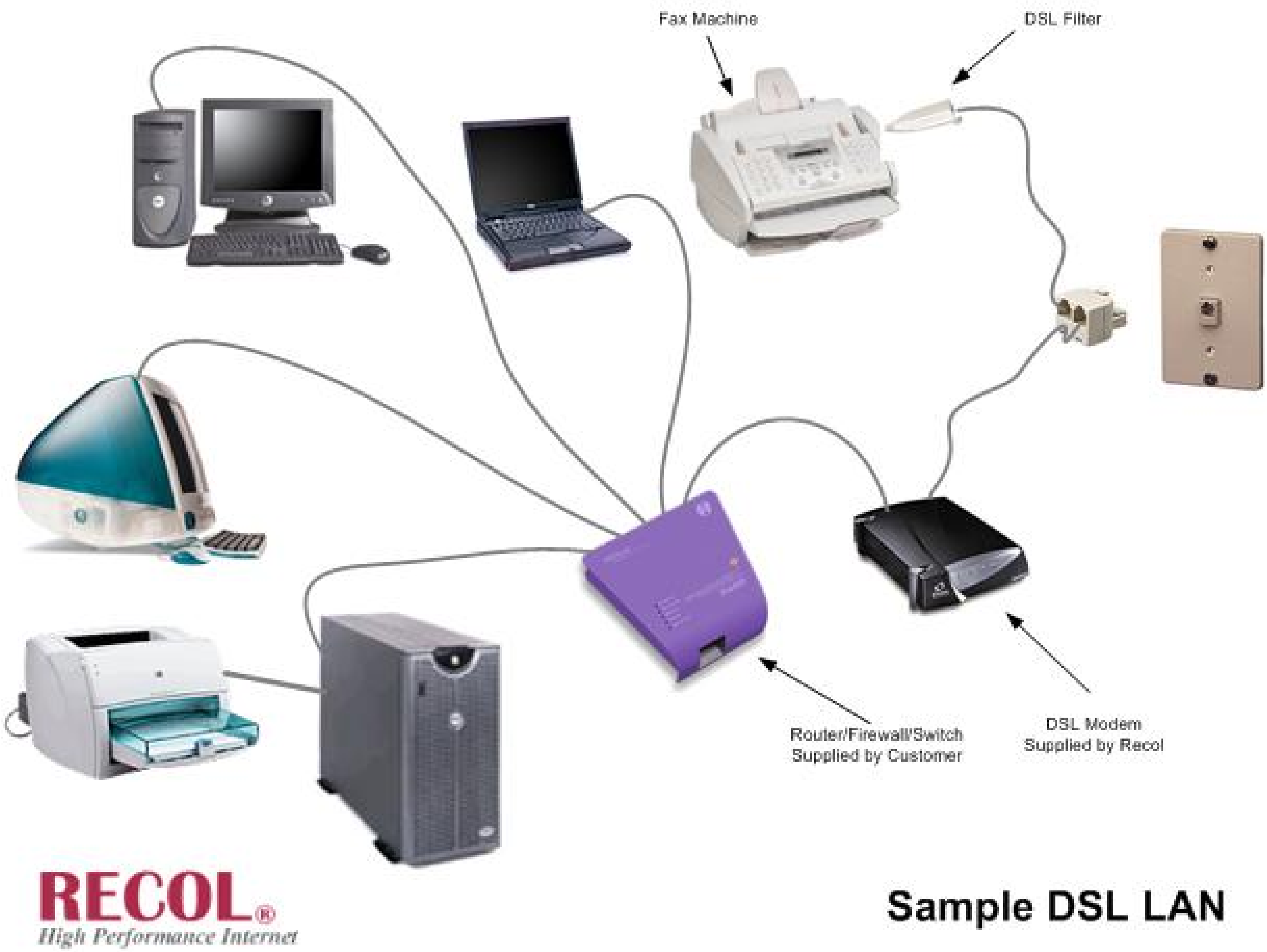
Codes for Magnetic Recording



Error Correction Codes: Satellite Communication

Modems

Data Transmission: Digital Subscriber Lines



Data Transmission: Cellular Wireless

WiFi


BELKIN
N1 Wireless Router
security
wireless
wired
router
modem
internet

Information Theory as a Design Driver

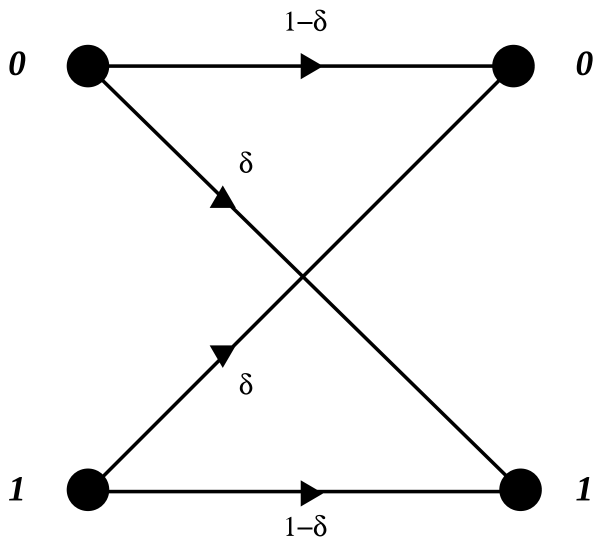
- Universal data compression
- Sparse-graph codes
- Voiceband modems
- Discrete multitone modulation
- CDMA
- Multiuser detection
- Flash Signaling
- Multiantenna
- Space-time codes
- Dirty-paper coding
- Opportunistic signaling
- Network coding
- Discrete denoising
- Secrecy

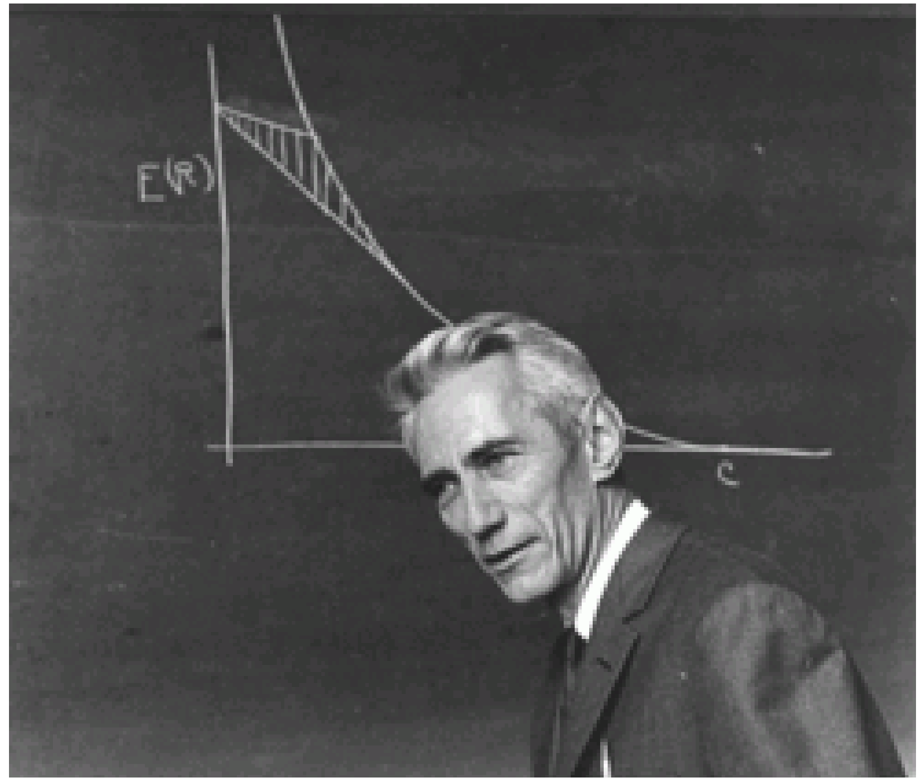
Open Problems: Single-User Channels

Open Problems: Single-User Channels

- Reliability Function

Reliability function





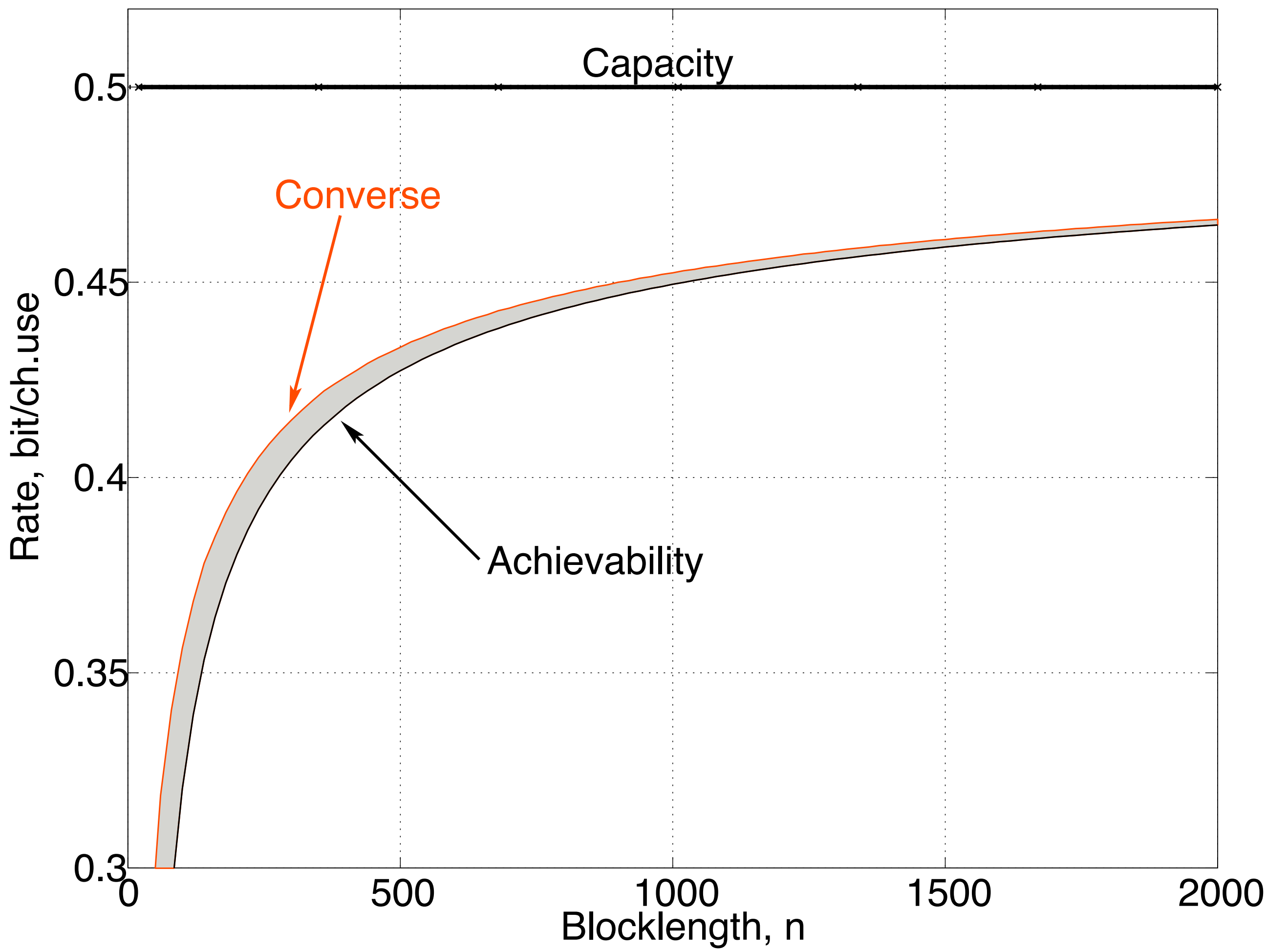
Open Problems: Single-User Channels

- Reliability Function
- Delay – Error Probability Tradeoff

Delay – Error Probability Tradeoff: Non-asymptotic regime

how much do we need to back off from channel capacity when blocklength = 1000?

Delay – Error Probability Tradeoff: Non-asymptotic regime



Y. Polyanskiy, H. V. Poor, S. Verdú, "New Channel Coding Achievability Bounds"
***2008 IEEE Int. Symposium on Information Theory,* Toronto, Ontario, Canada, July 6-11, 2008**

Delay – Error Probability Tradeoff: Non-asymptotic regime

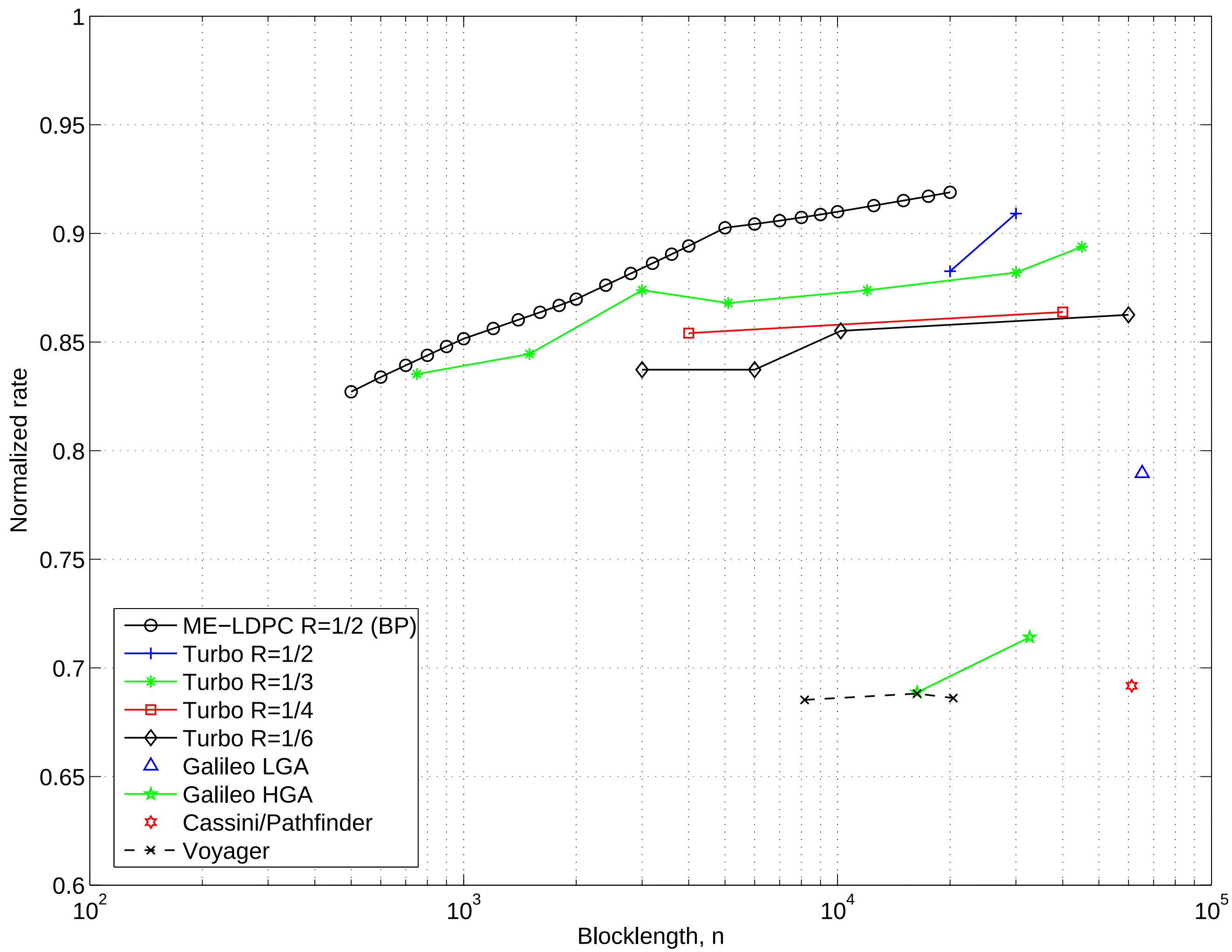


Figure 15: Normalized rates for various practical codes over AWGN, probability of block error $\epsilon = 10^{-4}$.

Open Problems: Single-User Channels

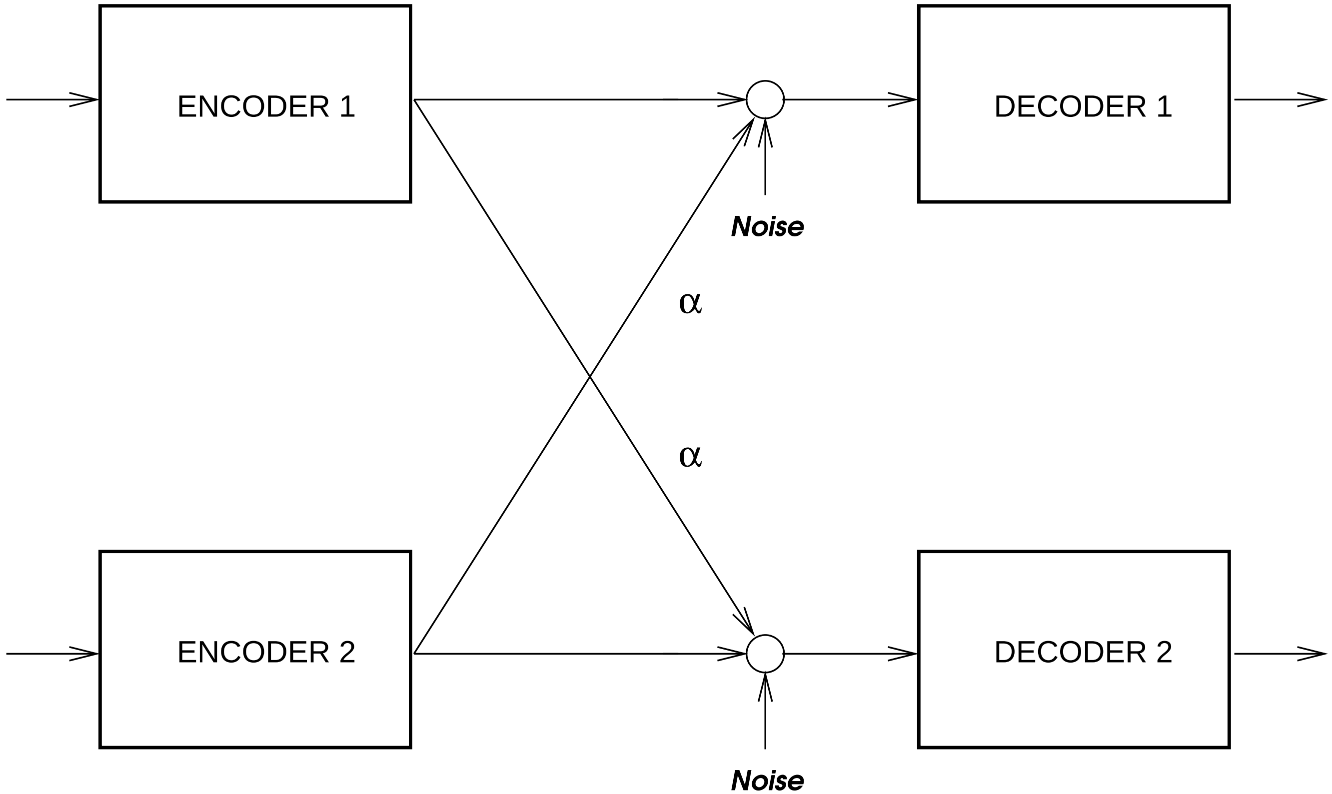
- Reliability Function
- Delay – Error Probability Tradeoff
- Feedback
  - partial/noisy feedback
  - delay-performance tradeoff;
  - constructive schemes
  - Gaussian channels with memory
- Deletions, Synchronization
- Zero-error Capacity

Open Problems: Multiuser Channels

Open Problems: Multiuser Channels

- Interference Channels

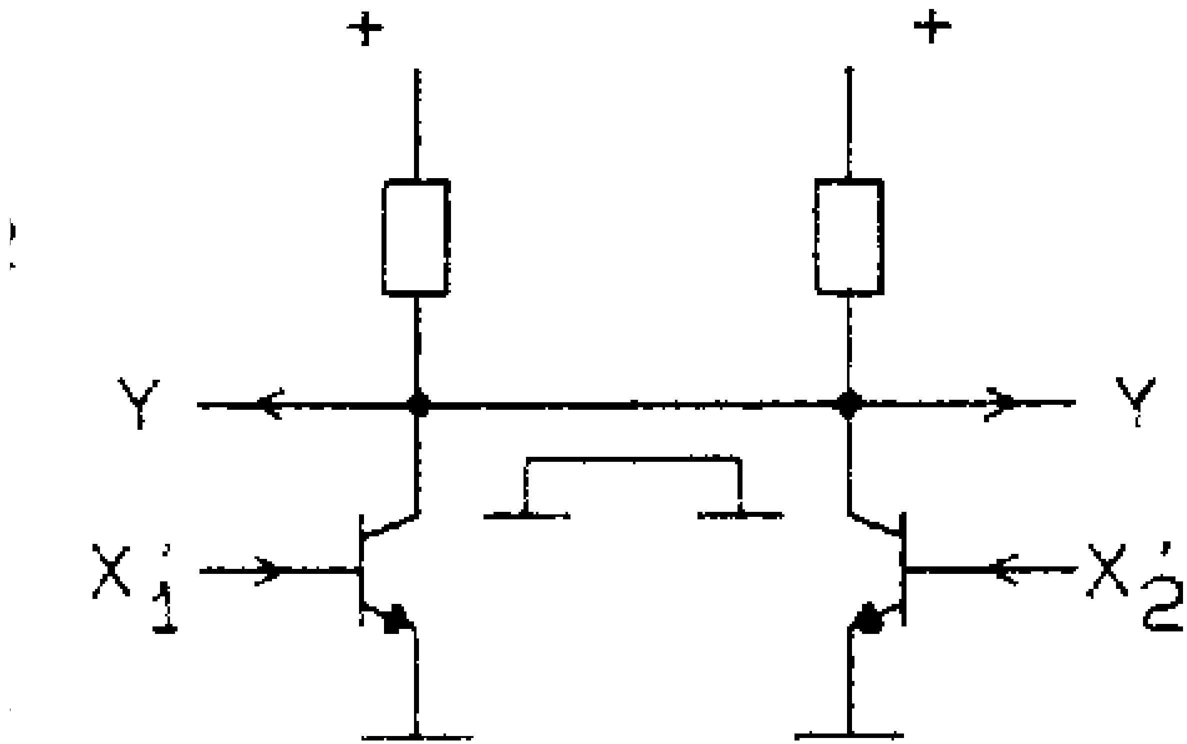
Interference Channels



Open Problems: Multiuser Channels

- Interference Channels
- Two-way Channels

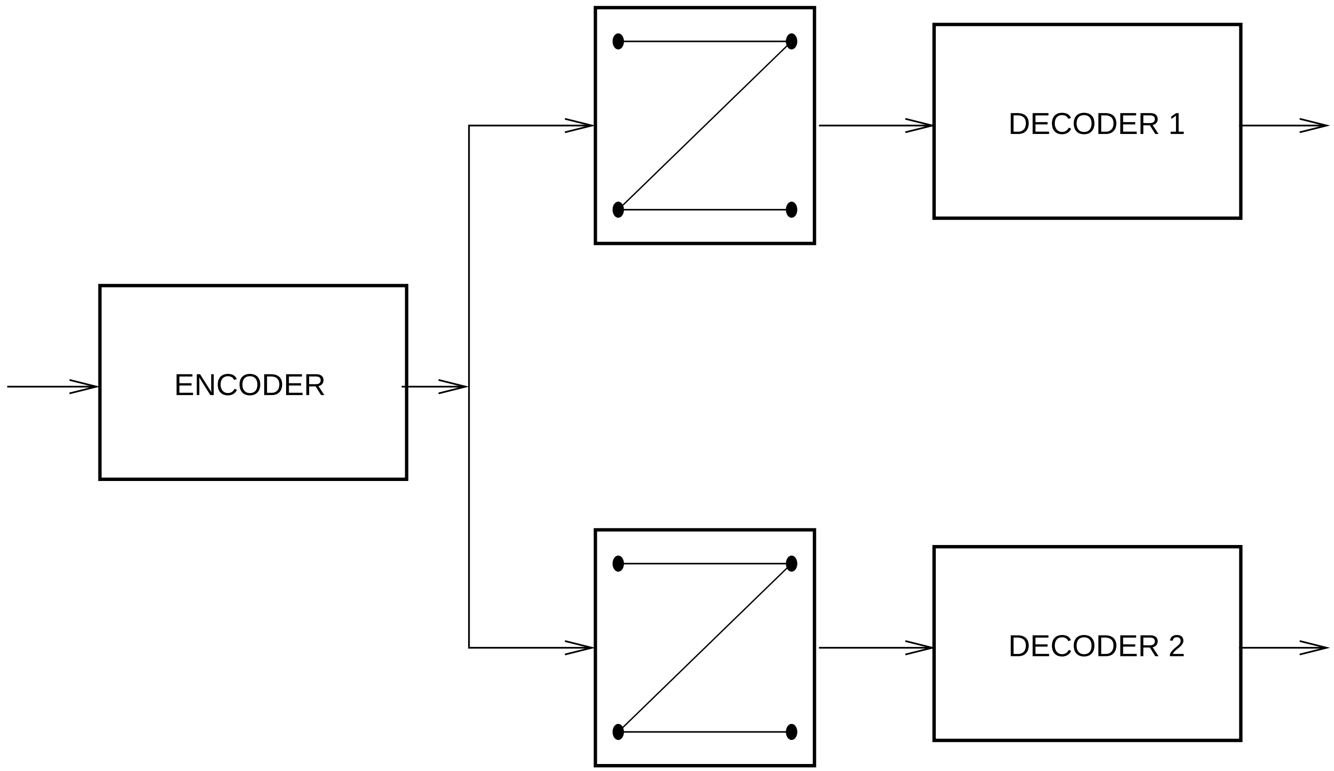
Two-Way Channels



Open Problems: Multiuser Channels

- Interference Channels
- Two-way Channels
- Broadcast Channels

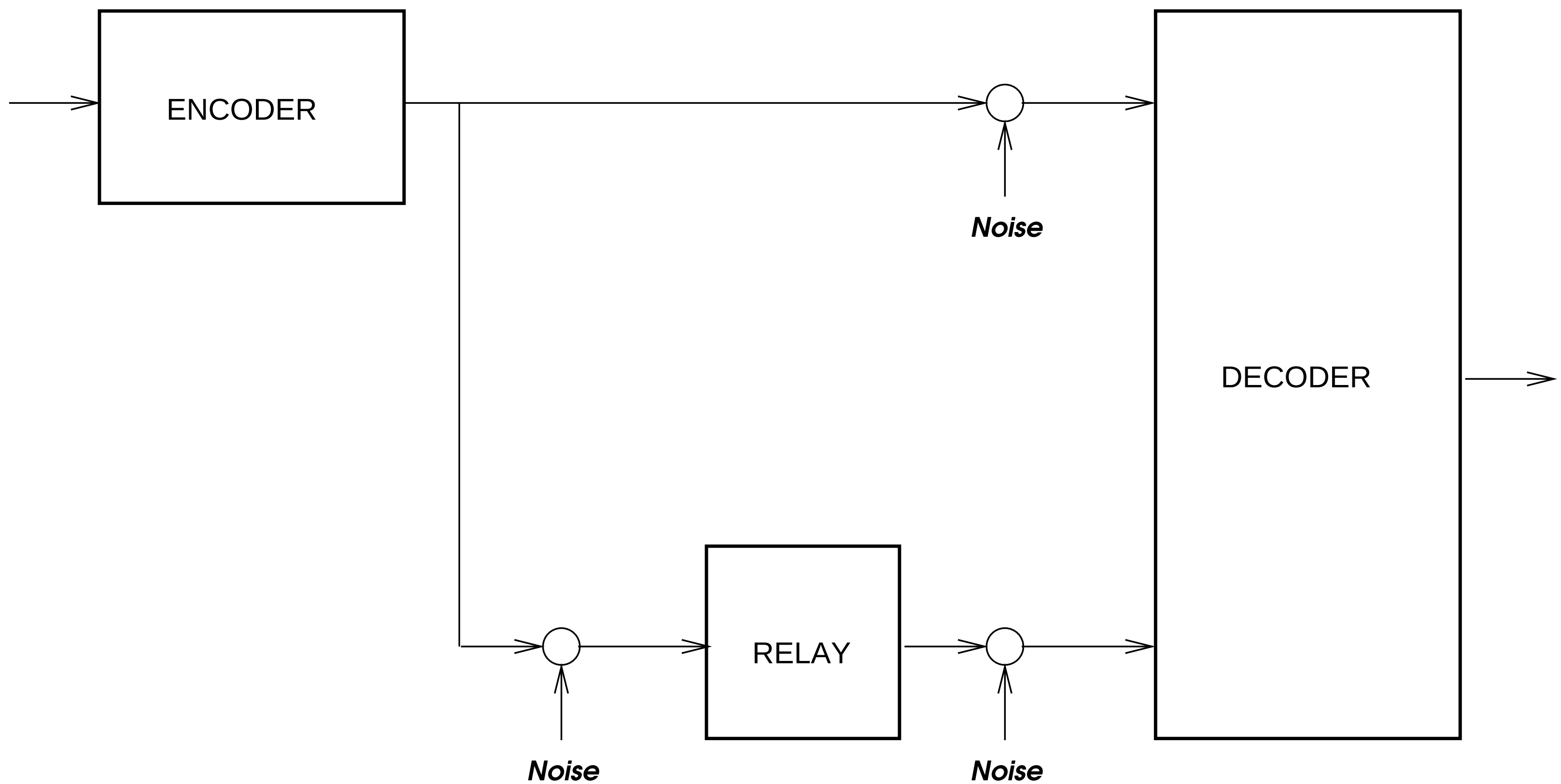
Broadcast Channels



Open Problems: Multiuser Channels

- Interference Channels
- Two-way Channels
- Broadcast Channels
- Relay Channels

Relay Channels



Open Problems: Multiuser Channels

- Interference Channels
- Two-way Channels
- Broadcast Channels
- Relay Channels
- Compression-Transmission

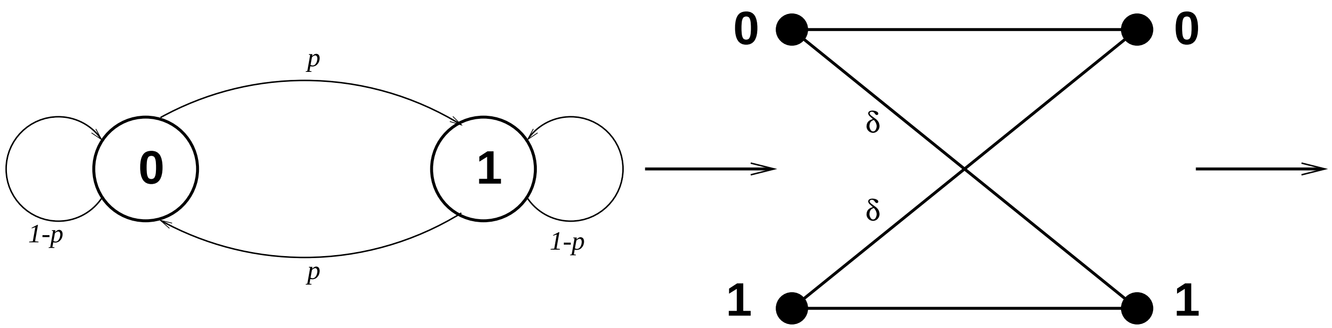
Open Problems: Data Compression: Non-asymptotics

how many bits do we
need to compress a
140-character
twitter message?

Open Problems: Lossless Data Compression

- Non-asymptotic regime
- Joint source/channel coding
- Images
- Implementing Slepian-Wolf:
  - Backup hard-disks with dialup modems?
- $\Longleftarrow$ Artificial intelligence
- Entropy rate of sources with memory

Entropy Rate of Sources with Memory



Open Problems: Lossy Data Compression

- Theory $\leftrightarrow$ $\leftrightarrow$ Practice

Open Problems: Lossy Data Compression

- Theory $\leftrightarrow$ $\leftrightarrow$ Practice
- Constructive schemes
  - memoryless sources
  - universal lossy data compression

Open Problems: Lossy Data Compression

- Theory $\leftrightarrow$ $\leftrightarrow$ Practice
- Constructive Schemes
  - memoryless sources
  - universal lossy data compression
- Multi-source Fundamental Limits

Multi-source Fundamental Limits

DIGITAL · STEREO

453 478-2 GH

PY 925

Deutsche Grammophon

a PolyGram company

4D DDD

AUDIO RECORDING

A new dimension in clarity and realism

Klarheit und Klangtreue in neuer Dimension

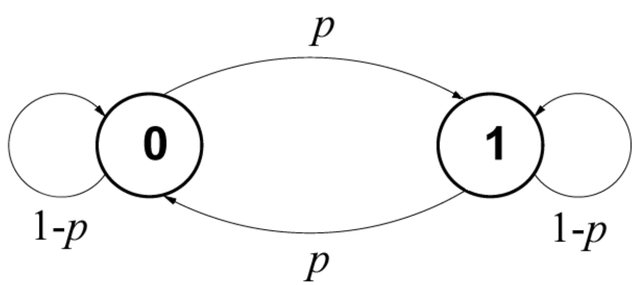
Une nouvelle dimension dans la clarté et le naturel

La nuova dimensione di un suono limpido e naturale

Open Problems: Lossy Data Compression

- Theory $\leftrightarrow$ $\leftrightarrow$ Practice
- Constructive Schemes
  - memoryless sources
  - universal lossy data compression
- Multi-source Fundamental Limits
- Rate-Distortion Functions

Binary Markov chain; Bit Error Rate



$0 \le p \le \frac{1}{2}$

$$R(D) = \begin{cases} h(p) - h(D) & \text{for } 0 \le D \le D_* \\ \text{UNKNOWN} & \text{otherwise.} \end{cases}$$

Gradient

↗ Constructive ↘ Combinatorics

↗ Applied ↘ Continuous Time

↗ Multiuser ↘ Ergodic Theory

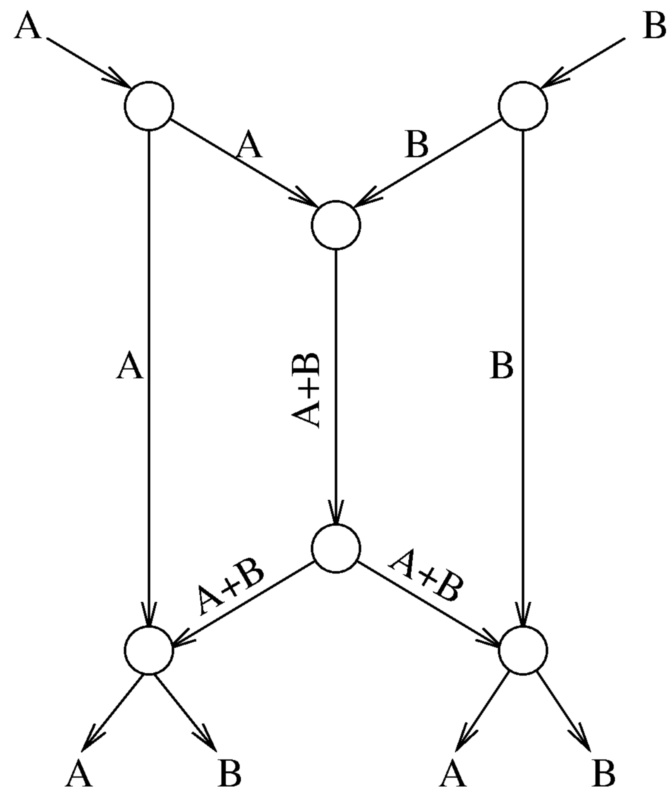
↗ Universal Methods ↘ Error Exponents

↗ Intersections

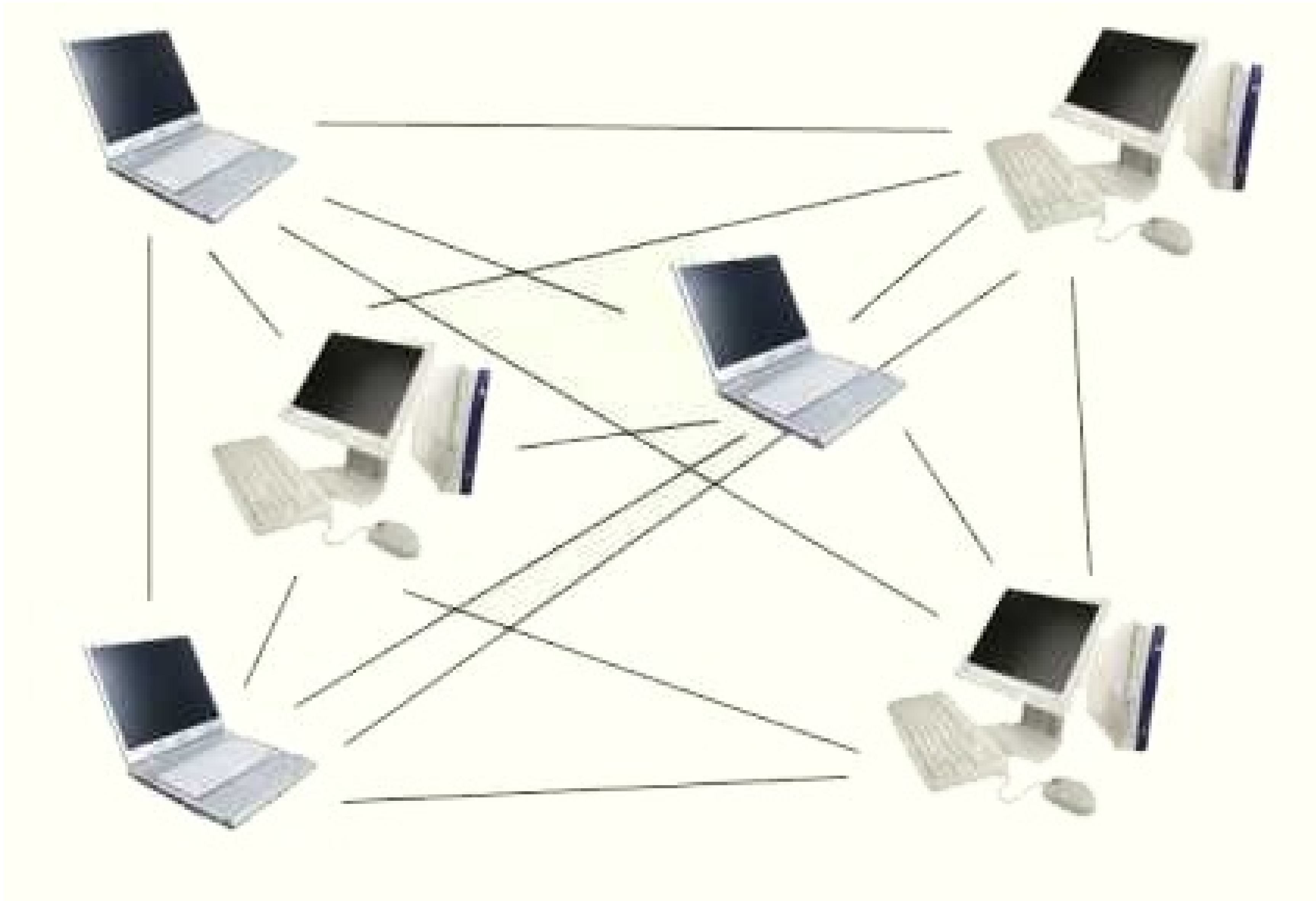
Intersections

- Networks
  - Network coding
  - Scaling laws

Network Coding



Scaling Laws



Intersections

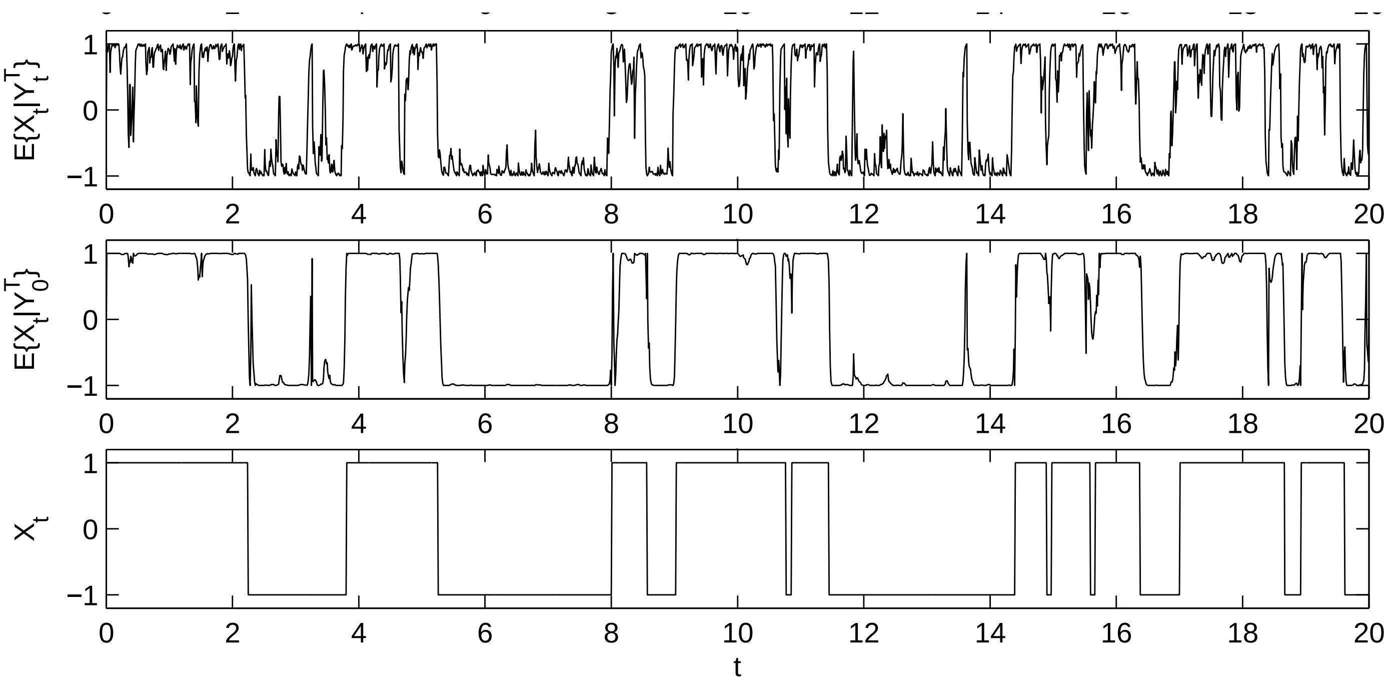
- Networks
  - Network coding
  - Scaling laws
- Signal Processing
  - Estimation theory
  - Compressed sensing
  - Discrete denoising
  - Finite-alphabet

Information Theory ⇔ Estimation Theory

$$\frac{\mathsf{d}}{\mathsf{dsnr}} I\left(\boldsymbol{X}; \sqrt{\mathsf{snr}} \cdot \boldsymbol{H}\,\boldsymbol{X} + \boldsymbol{W}\right) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr})$$

- Entropy power inequality
- Monotonicity of nonGaussianness
- Mercury-Waterfilling
- Continuous-time Nonlinear Filtering

Information Theory $\Leftrightarrow$ Nonlinear Filtering



$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\, \mathsf{d}\gamma$$

Intersections

- Networks
  - Network coding
  - Scaling laws
- Signal Processing
  - Estimation theory
  - Compressed sensing
  - Discrete denoising
  - Finite-alphabet

Text Denoising: Don Quixote de La Mancha

Noisy Text (21 errors, 5% error rate):

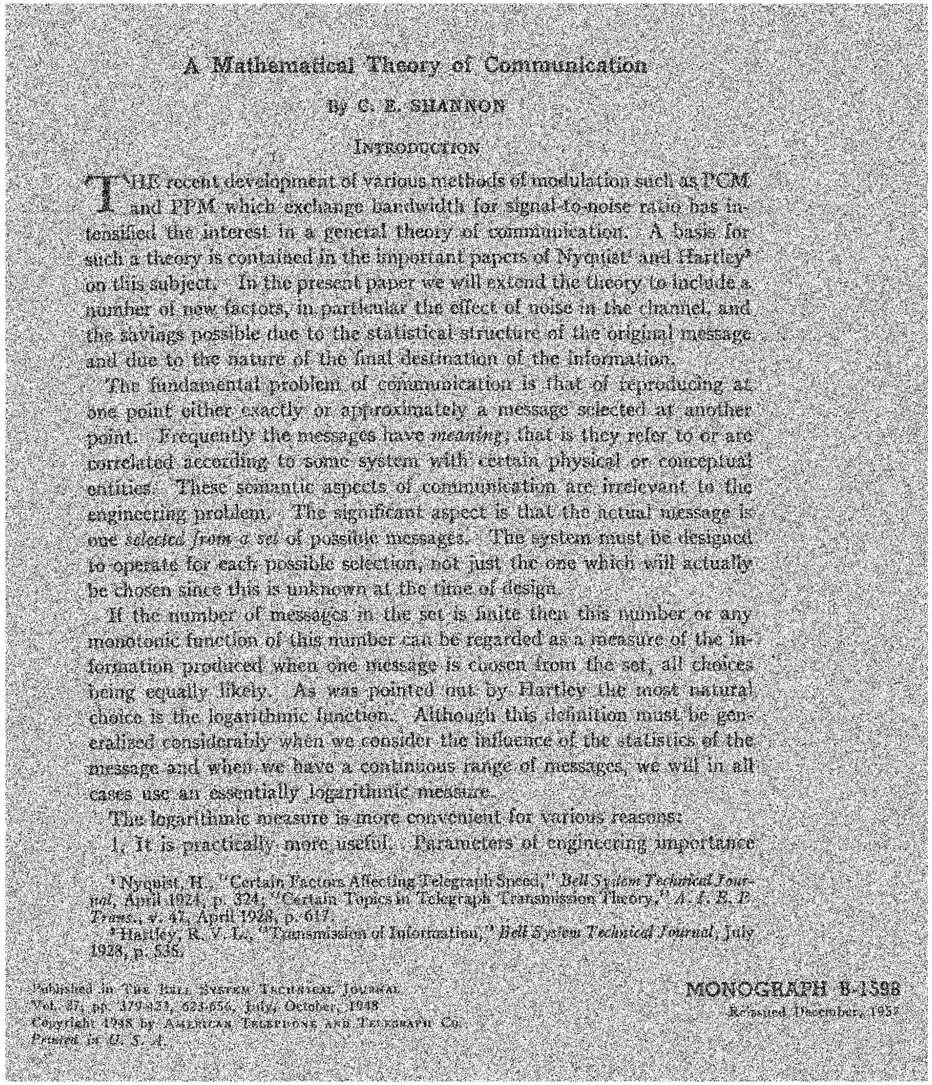
”Whar giants?” said Sancho Panza. ”Those thou seest theee,” snswered yis master, ”with the long arms, and spne have tgem ndarly two leagues long.” ”Look, ylur worship,” sair Sancho; ”what we see there zre not gianrs but windmills, and what seem to be their arms are the sails that turned by the wind make rhe millstpne go.” ”Kt is easy to see,” replied Don Quixote, ”that thou art not used to this business of adventures; fhose are giantz; and if thou arf wfraod, away with thee out of this and betake thysepf to prayer while I engage them in fierce and unequal combat.”

DUDE output, $k = 2$ (7 errors):

”What giants?” said Sancho Panza. ”Those thou seest there,” answered his master, ”with the long arms, and spne have them nearly two leagues long.” ”Look, your worship,” said Sancho; ”what we see there are not giants but windmills, and what seem to be their arms are the sails that turned by the wind make the millstone go.” ”It is easy to see,” replied Don Quixote, ”that thou art not used to this business of adventures; fhose are giantz; and if thou arf wfraod, away with thee out of this and betake thyself to prayer while I engage them in fierce and unequal combat.”

BSC systematic output; $C = 1 - h(0.25) = 0.19$

0.25



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] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.

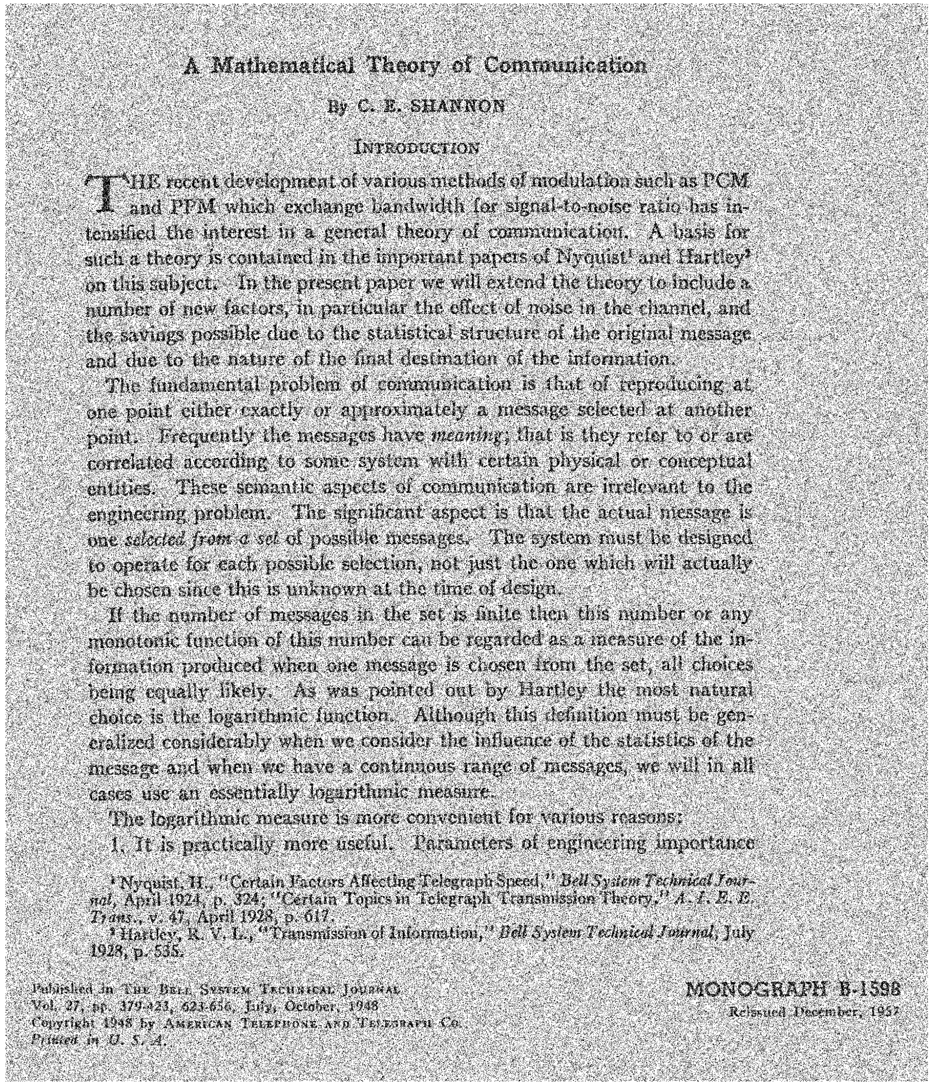
[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

Published in THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 27, pp. 379-423, 623-656, July, October, 1948

MONOGRAPH B-1598
Reissued December, 1957

BP Decoder Output (RA; Rate = 0.25; $k = 4000$; 30 iter.)

0.21



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] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.

[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

Published in THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 27, pp. 379-423, 623-656, July, October, 1948

MONOGRAPH B-1598
Reissued December, 1957

Denoising+Decoding

0.0003

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] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.
[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

Published in THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 27, pp. 379-423, 623-656, July, October, 1948

MONOGRAPH B-1598
Reissued December, 1957

Intersections

- Networks
  - Network coding
  - Scaling laws
- Signal Processing
  - Estimation theory
  - Compressed sensing
  - Discrete denoising
  - Finite-alphabet
- Control
  - Noisy [plant $\longrightarrow$ controller] channel.
  - Control-oriented feedback communication schemes.

Intersections

- Networks
  - Network coding
  - Scaling laws
- Signal Processing
  - Estimation theory
  - Compressed sensing
  - Discrete denoising
  - Finite-alphabet
- Control
  - Noisy [plant $\longrightarrow$ controller] channel.
  - Control-oriented feedback communication schemes.
- Computer Science
  - Analytic information theory
  - Interactive communication

Other Intersections

- Economics
- Quantum
- Bio
- Physics

Emerging Tools

- Optimization
- Statistical Physics
- Random Matrices