Shannon's "A Mathematical Theory of Communication"

Emre Telatar

EPFL

Kanpur — October 19, 2016

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

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

First published in two parts in the July and October 1948 issues of BSTJ.

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

First published in two parts in the July and October 1948 issues of BSTJ.

A number of republications since,

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

First published in two parts in the July and October 1948 issues of BSTJ.

A number of republications since, notably in 1949 by the UI Press (with a preface by W. Weaver) as "The Mathematical Theory of Communication".

Origins of A Mathematical Theory Communication

Shannon's 1949 paper *Communication Theory or Secrecy Systems* was already published in classified literature in 1945, and contains ideas also present in the AMTC. This leads some to believe that Shannon arrived to his Mathematical Theory of Communication via Cryptography.

Origins of A Mathematical Theory Communication

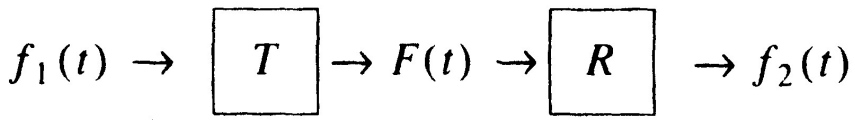
Shannon's 1949 paper *Communication Theory or Secrecy Systems* was already published in classified literature in 1945, and contains ideas also present in the AMTC. This leads some to believe that Shannon arrived to his Mathematical Theory of Communication via Cryptography.

My understanding is that this is not the case. Shannon had been working on developing his Mathematical Theory of Communication since late 1930's — and his cryptography work was built upon his communication ideas, not vice-versa.

Letter to Vannevar Bush — Feb 16, 1939

Dear Dr. Bush,

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



$f_1(t)$ is a general function of time (arbitrary except for certain frequency limitations) representing the intelligence to be transmitted. It represents, for example, the pressure-time function in radio and telephony, or the voltage-time curve output of an iconoscope in television.

T is a transmission element which operates on $f_1(t)$ through modulation, distortion, etc. to give a new function of time $F(t)$, which is actually transmitted. $F(t)$ in radio and television is the electromagnetic wave sent out by the transmitter, and in general need not be at all similar to $f_1(t)$, although, of course, they are closely related. I consider T to be a mathematical operator which transforms f_1 into F, thus $F(t) = T[f_1(t)]$.

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;* [...] These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is *selected from a set* of possible messages. The system must be designed to operate for each possible selection [...]"

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;* [...] These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is *selected from a set* of possible messages. The system must be designed to operate for each possible selection [...]"

- It is the choice (selection) that matters.

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;* [...] These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is *selected from a set* of possible messages. The system must be designed to operate for each possible selection [...]"

- It is the choice (selection) that matters.
- Net neutrality.

“Coat of arms”

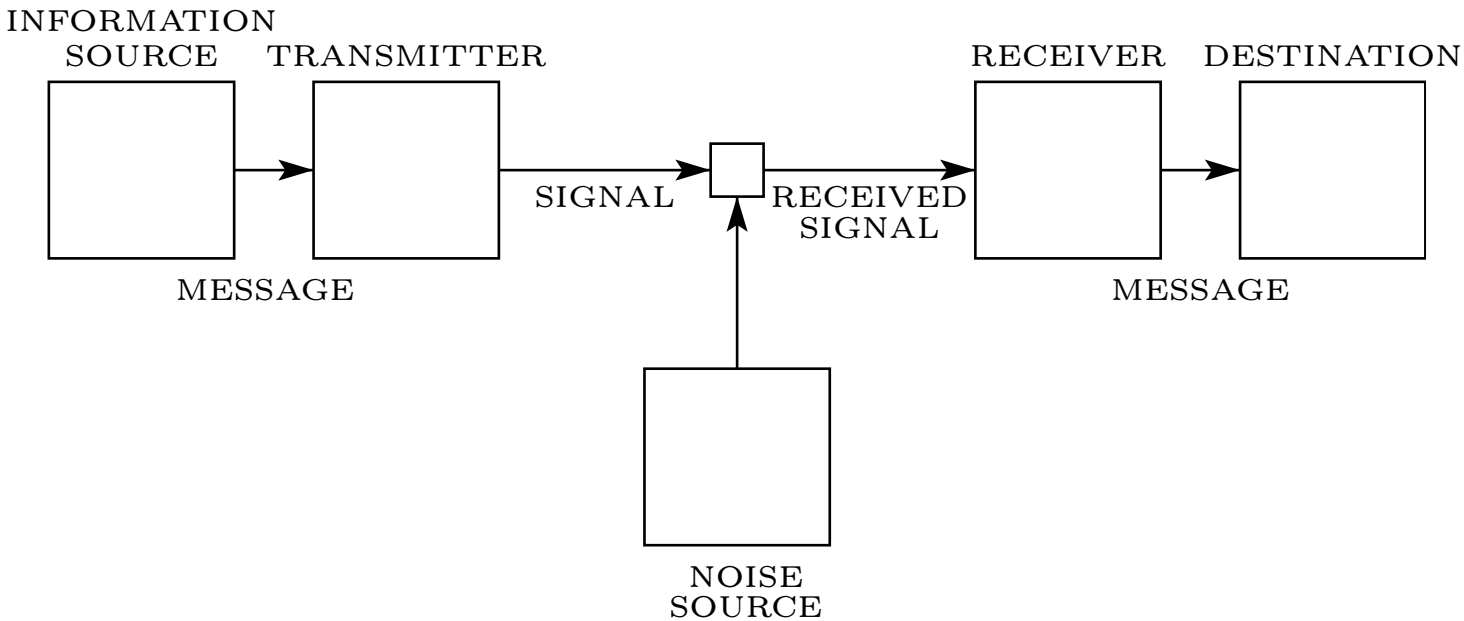


Fig 1. Schematic diagram of a general communication system

“Coat of arms”

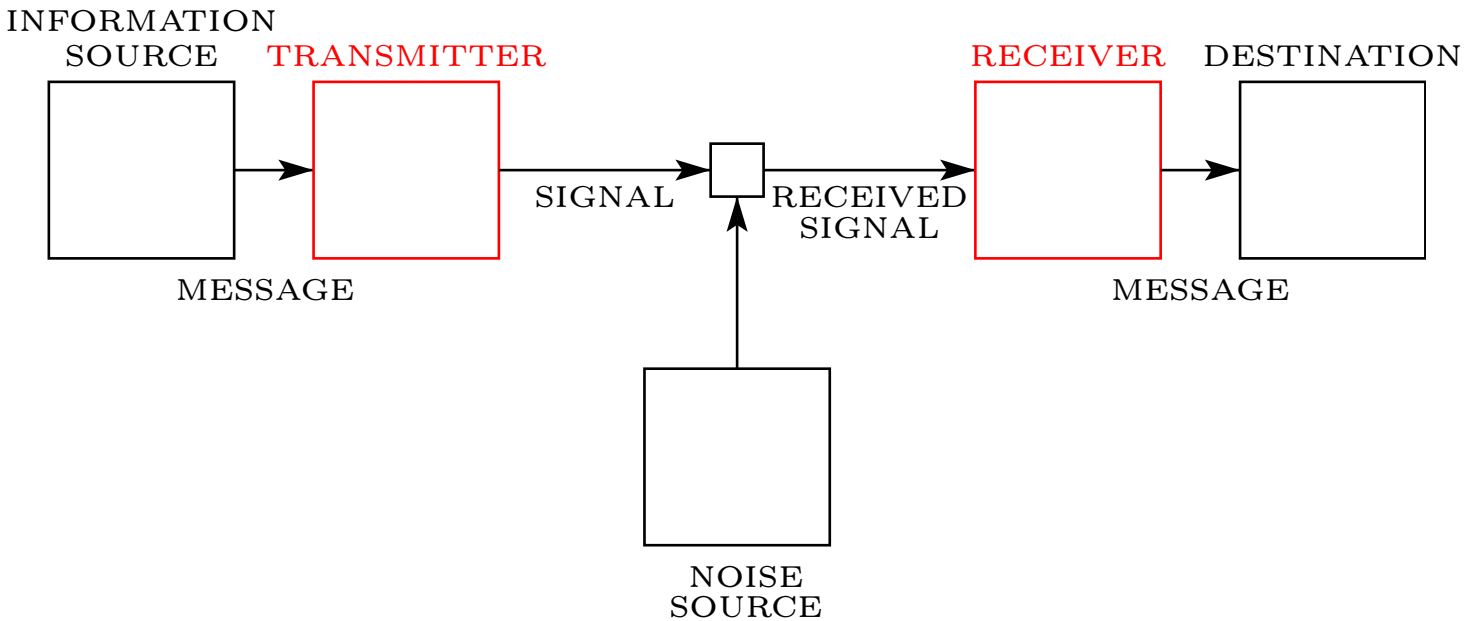


Fig 1. Schematic diagram of a general communication system

Information source

- Shannon models the information source as a probabilistic device that chooses among possible messages. A message is taken to be a sequence of symbols, this makes the message a stochastic process.

Information source

- Shannon models the information source as a probabilistic device that chooses among possible messages. A message is taken to be a sequence of symbols, this makes the message a stochastic process.
- Such a model remains non-intuitive to students even today. Shannon dedicates quite some space and gives numerous examples to motivate the model. “It appears then that a sufficiently complex stochastic process will give a satisfactory representation of a discrete source.”

Information source

- Shannon models the information source as a probabilistic device that chooses among possible messages. A message is taken to be a sequence of symbols, this makes the message a stochastic process.
- Such a model remains non-intuitive to students even today. Shannon dedicates quite some space and gives numerous examples to motivate the model. "It appears then that a sufficiently complex stochastic process will give a satisfactory representation of a discrete source."

Information source

- Shannon models the information source as a probabilistic device that chooses among possible messages. A message is taken to be a sequence of symbols, this makes the message a stochastic process.
- Such a model remains non-intuitive to students even today. Shannon dedicates quite some space and gives numerous examples to motivate the model. "It appears then that a sufficiently complex stochastic process will give a satisfactory representation of a discrete source."
- His running example is the class of Markov sources: in a ℓth order Markov source the successive symbols are generated one by one, the statistics of the next letter to be generated depends only on the ℓ most recent letters already generated.

Information source

- Shannon models the information source as a probabilistic device that chooses among possible messages. A message is taken to be a sequence of symbols, this makes the message a stochastic process.
- Such a model remains non-intuitive to students even today. Shannon dedicates quite some space and gives numerous examples to motivate the model. "It appears then that a sufficiently complex stochastic process will give a satisfactory representation of a discrete source."
- His running example is the class of Markov sources: in a ℓth order Markov source the successive symbols are generated one by one, the statistics of the next letter to be generated depends only on the ℓ most recent letters already generated. Call this the state of the source.

Information source

- Shannon models the information source as a probabilistic device that chooses among possible messages. A message is taken to be a sequence of symbols, this makes the message a stochastic process.
- Such a model remains non-intuitive to students even today. Shannon dedicates quite some space and gives numerous examples to motivate the model. "It appears then that a sufficiently complex stochastic process will give a satisfactory representation of a discrete source."
- His running example is the class of Markov sources: in a ℓth order Markov source the successive symbols are generated one by one, the statistics of the next letter to be generated depends only on the ℓ most recent letters already generated. Call this the state of the source.
- Online example?

Information source

More generally, Shannon seems to have the following type of "finite state information source" in mind:

- The source possesses an internal state. The set of possible states $\mathcal{S}$ is finite. Denote by S_i the state at instant i.

Information source

More generally, Shannon seems to have the following type of "finite state information source" in mind:

- The source possesses an internal state. The set of possible states $\mathcal{S}$ is finite. Denote by S_i the state at instant i.
- If $S_i = s$, the source generates the letter U_i from a finite alphabet $\mathcal{U}$ according to a fixed probability law $\Pr(U_i = u | S_i = s) = P(u|s)$, independent of past S's and U's.

Information source

More generally, Shannon seems to have the following type of "finite state information source" in mind:

- The source possesses an internal state. The set of possible states $\mathcal{S}$ is finite. Denote by S_i the state at instant i.
- If $S_i = s$, the source generates the letter U_i from a finite alphabet $\mathcal{U}$ according to a fixed probability law $\Pr(U_i = u | S_i = s) = P(u|s)$, independent of past S's and U's.
- The next state S_{i+1} is determined by S_i and U_i.

Information source

More generally, Shannon seems to have the following type of "finite state information source" in mind:

- The source possesses an internal state. The set of possible states $\mathcal{S}$ is finite. Denote by S_i the state at instant i.
- If $S_i = s$, the source generates the letter U_i from a finite alphabet $\mathcal{U}$ according to a fixed probability law $\Pr(U_i = u | S_i = s) = P(u|s)$, independent of past S's and U's.
- The next state S_{i+1} is determined by S_i and U_i.
- Moreover, U_i is uniquely determined by S_i and S_{i+1}.

Information source

More generally, Shannon seems to have the following type of "finite state information source" in mind:

- The source possesses an internal state. The set of possible states $\mathcal{S}$ is finite. Denote by S_i the state at instant i.
- If $S_i = s$, the source generates the letter U_i from a finite alphabet $\mathcal{U}$ according to a fixed probability law $\Pr(U_i = u | S_i = s) = P(u|s)$, independent of past S's and U's.
- The next state S_{i+1} is determined by S_i and U_i.
- Moreover, U_i is uniquely determined by S_i and S_{i+1}.
- The first three assumptions make $\{S_i\}$ a Markov process; Shannon further assumes that this process is stationary and ergodic.

Quantifying choice — Entropy

When one of K items is chosen — kth item with probability p_k — how much "choice" is there?

Quantifying choice — Entropy

When one of K items is chosen — kth item with probability p_k — how much "choice" is there?

- Shannon proposes the entropy, $H = -\sum_k p_k \log p_k$, as the natural answer to this question.

Quantifying choice — Entropy

When one of K items is chosen — kth item with probability p_k — how much "choice" is there?

- Shannon proposes the entropy, $H = -\sum_k p_k \log p_k$, as the natural answer to this question.
- The answer generalizes to two (or more) choices and conditional choices:

$$H(UV) = -\sum_{u,v} p(u,v) \log p(u,v)$$

$$H(V|U) = -\sum_{u,v} p(u,v) \log p(v|u)$$

Entropy of a source

Shannon motivates "$H = -\sum p \log p$" by an axiomatic approach — listing desirable properties of any quantification that measures "choice" and showing that entropy is the only possible answer. A more important justification is found when Shannon considers entropy in the context of an information source.

Entropy of a source

Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

Entropy of a source

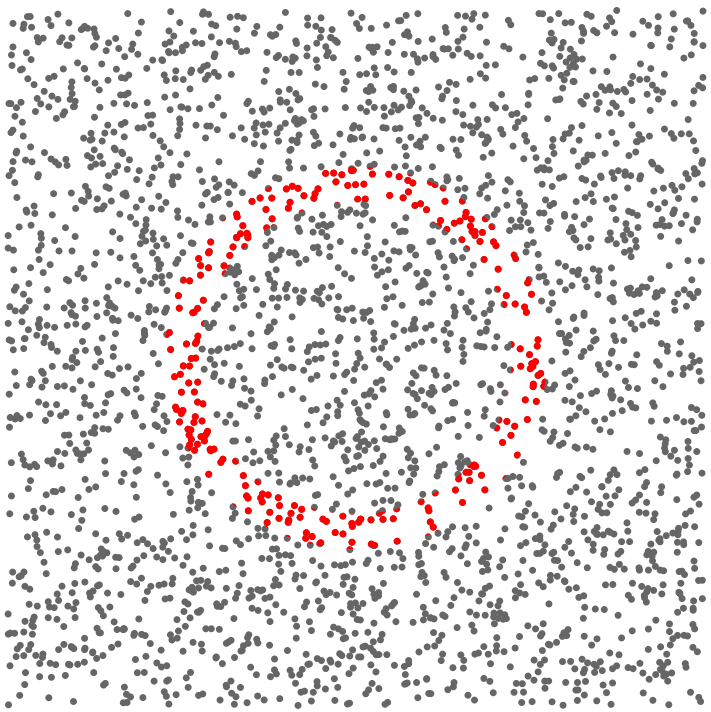
Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

- Then, for any $\epsilon > 0$ and $n > 0$, within $\mathcal{U}^n$,

Entropy of a source

Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

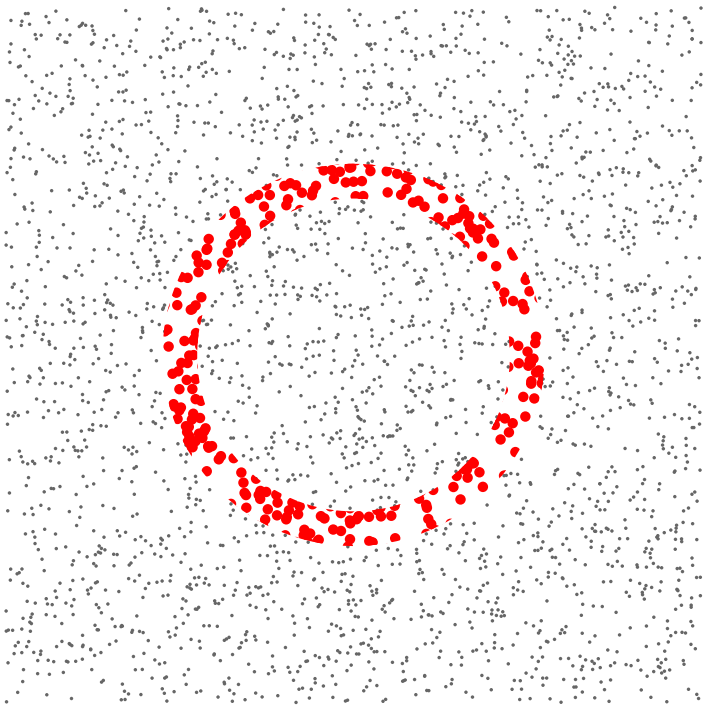
- Then, for any $\epsilon > 0$ and $n > 0$, within $\mathcal{U}^n$, there is a subset $\mathcal{T}_n$, such that $|\mathcal{T}_n| \lesssim 2^{nH}$,



Entropy of a source

Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

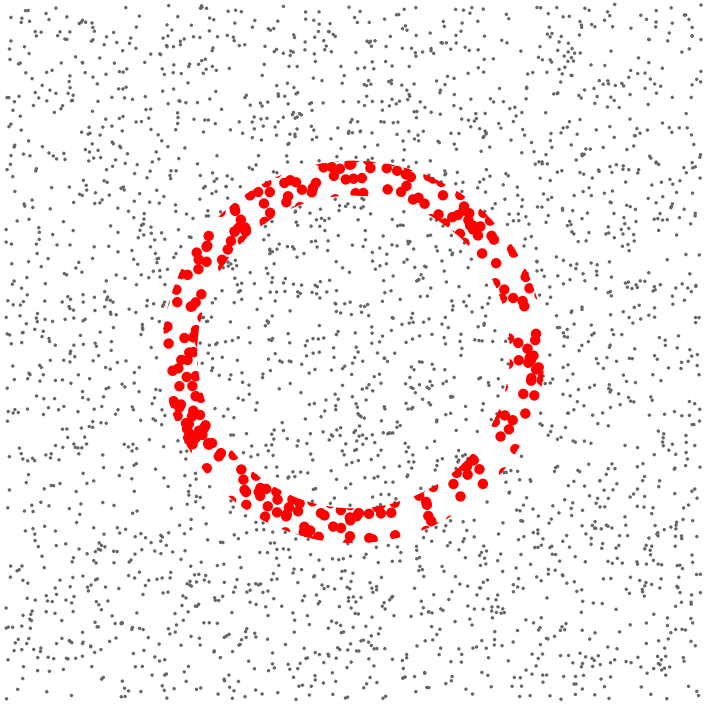
- Then, for any $\epsilon > 0$ and $n > 0$, within $\mathcal{U}^n$, there is a subset $\mathcal{T}_n$, such that $|\mathcal{T}_n| \lesssim 2^{nH}$, and yet $\Pr((U_1, \ldots, U_n) \in \mathcal{T}_n) \geq 1 - \epsilon$.



Entropy of a source

Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

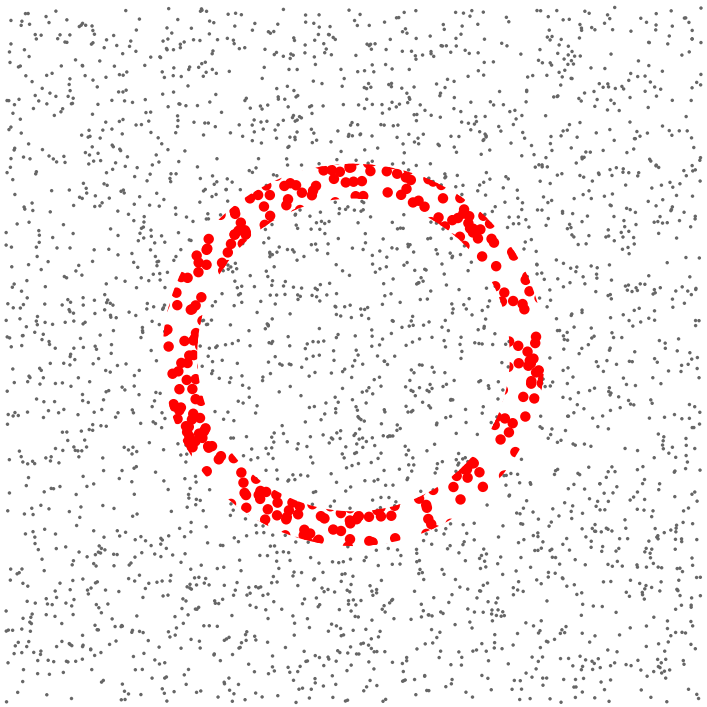
- Then, for any $\epsilon > 0$ and $n > 0$, within $\mathcal{U}^n$, there is a subset $\mathcal{T}_n$, such that $|\mathcal{T}_n| \lesssim 2^{nH}$, and yet $\Pr((U_1, \ldots, U_n) \in \mathcal{T}_n) \geq 1 - \epsilon$.
- That is: we can represent all (except a set of small total probability) source sequences of length n with nH bits.



Entropy of a source

Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

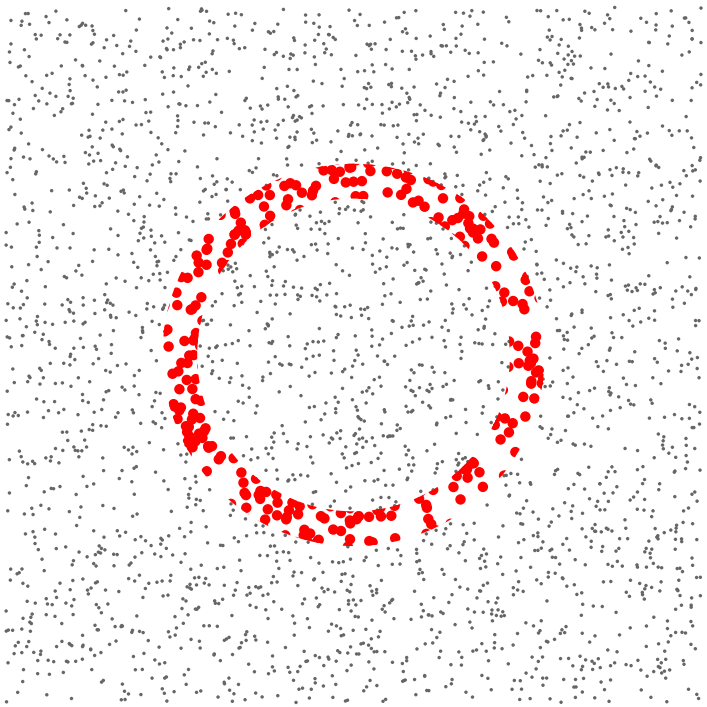
- Moreover, if $\mathcal{A}_n \subset \mathcal{U}^n$, $n = 1, 2, \ldots$ is such that $\Pr((U_1, \ldots, U_n) \in \mathcal{A}_n) > \epsilon$,



Entropy of a source

Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

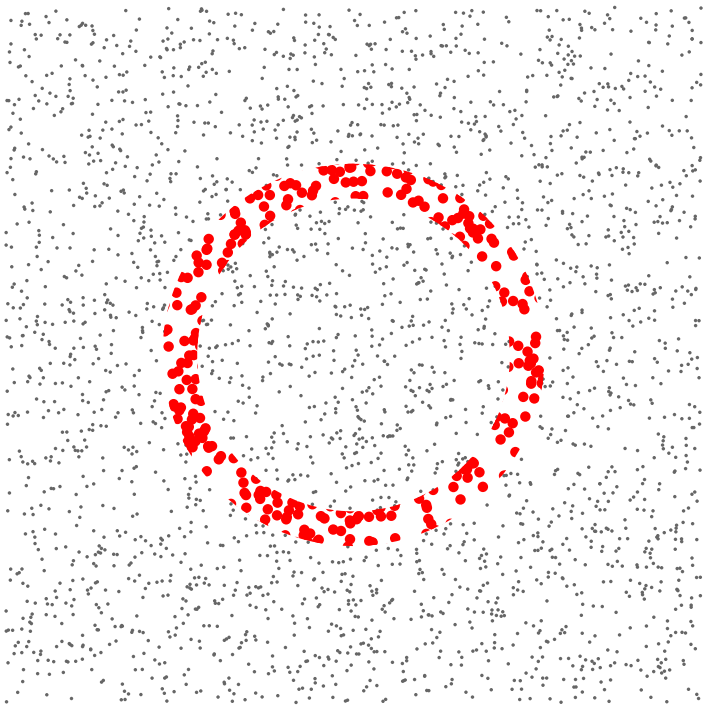
- Moreover, if $\mathcal{A}_n \subset \mathcal{U}^n$, $n = 1, 2, \ldots$ is such that $\Pr((U_1, \ldots, U_n) \in \mathcal{A}_n) > \epsilon$, then $|\mathcal{A}_n| \gtrsim 2^{nH}$.



Entropy of a source

Suppose $\ldots, U_1, U_2, \ldots$ are the letters produced by a finite state source and $\ldots, S_1, S_2, \ldots$ is the corresponding state sequence. Let $H = H(U_i|S_i)$ denote the entropy of the source letter produced conditional on the current state.

- Moreover, if $\mathcal{A}_n \subset \mathcal{U}^n$, $n = 1, 2, \ldots$ is such that $\Pr((U_1, \ldots, U_n) \in \mathcal{A}_n) > \epsilon$, then $|\mathcal{A}_n| \gtrsim 2^{nH}$.
- That is, representing any subset of source sequences of non-vanishing probability requires nH bits.



Data compression

- Having shown that $\approx nH$ bits is necessary and sufficient to represent sequences of n letters from a finite state source, Shannon also gives an explicit method of doing so.

Data compression

- Having shown that $\approx nH$ bits is necessary and sufficient to represent sequences of n letters from a finite state source, Shannon also gives an explicit method of doing so.
- Shannon's method represents all source sequences of length n using a variable number of bits, requiring on the average, less than $nH + 1$ bits.

Data compression

- Having shown that $\approx nH$ bits is necessary and sufficient to represent sequences of n letters from a finite state source, Shannon also gives an explicit method of doing so.
- Shannon's method represents all source sequences of length n using a variable number of bits, requiring on the average, less than $nH + 1$ bits.
- The method is a precursor of Huffman coding, and contains the essence of modern arithmetic coding.

Discrete Channel with Noise

Shannon's (first) model of a transmission medium (channel) is discrete in time and value:

Discrete Channel with Noise

Shannon's (first) model of a transmission medium (channel) is discrete in time and value:

- The channel accepts a sequence of letters $x_1, x_2, \ldots$, from a finite alphabet,

Discrete Channel with Noise

Shannon's (first) model of a transmission medium (channel) is discrete in time and value:

- The channel accepts a sequence of letters $x_1, x_2, \ldots$, from a finite alphabet,
- Produces a sequence of letters $Y_1, Y_2, \ldots$, from a finite alphabet.

Discrete Channel with Noise

Shannon's (first) model of a transmission medium (channel) is discrete in time and value:

- The channel accepts a sequence of letters $x_1, x_2, \ldots$, from a finite alphabet,
- Produces a sequence of letters $Y_1, Y_2, \ldots$, from a finite alphabet.
- The output sequence is related probabilistically to the input sequence.

Discrete Channel with Noise

Shannon's (first) model of a transmission medium (channel) is discrete in time and value:

- The channel accepts a sequence of letters $x_1, x_2, \ldots$, from a finite alphabet,
- Produces a sequence of letters $Y_1, Y_2, \ldots$, from a finite alphabet.
- The output sequence is related probabilistically to the input sequence.
- This model is much more readily accepted by students, that the transmission medium should be noisy, and noise is modeled by a random process is intuitive to most.

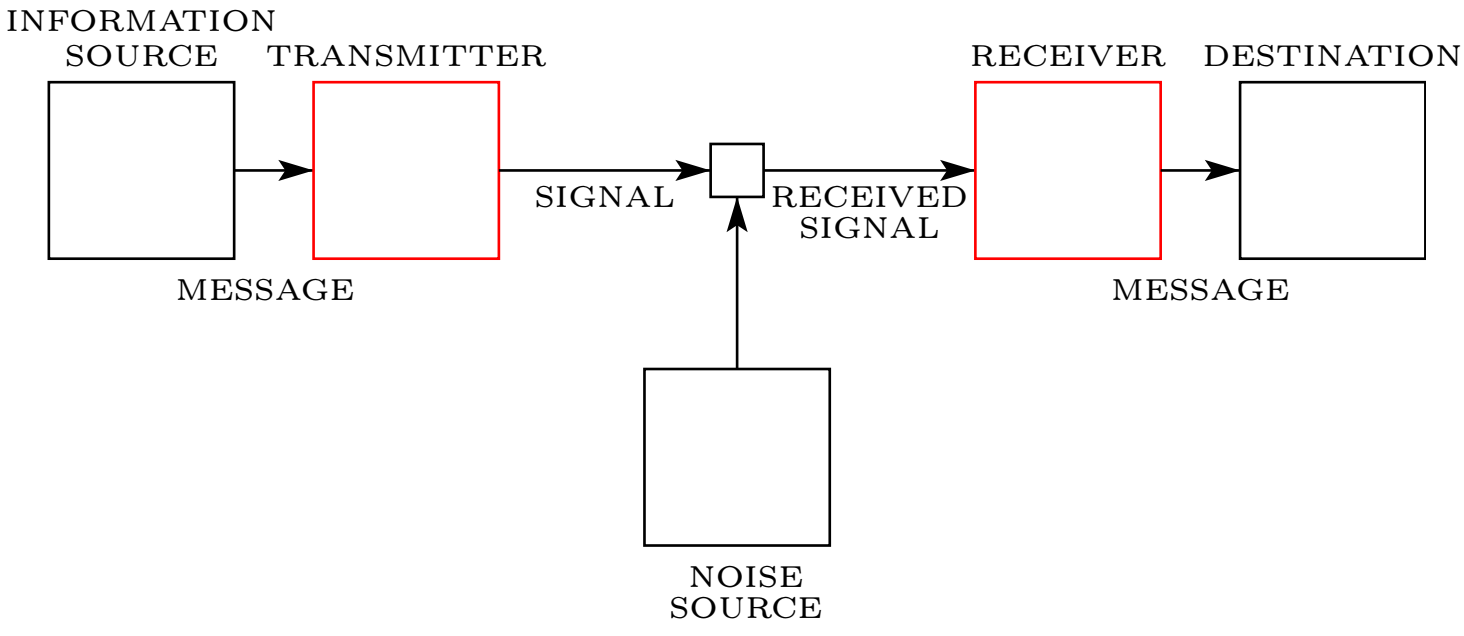
Discrete Channel with Noise

Shannon chooses to model the stochastic nature of the channel with the help of a hidden channel state that allows the channel keep some memory of the past, and a probability kernel $P(y, s'|x, s)$ that generates the current output and next state based on the current input and state

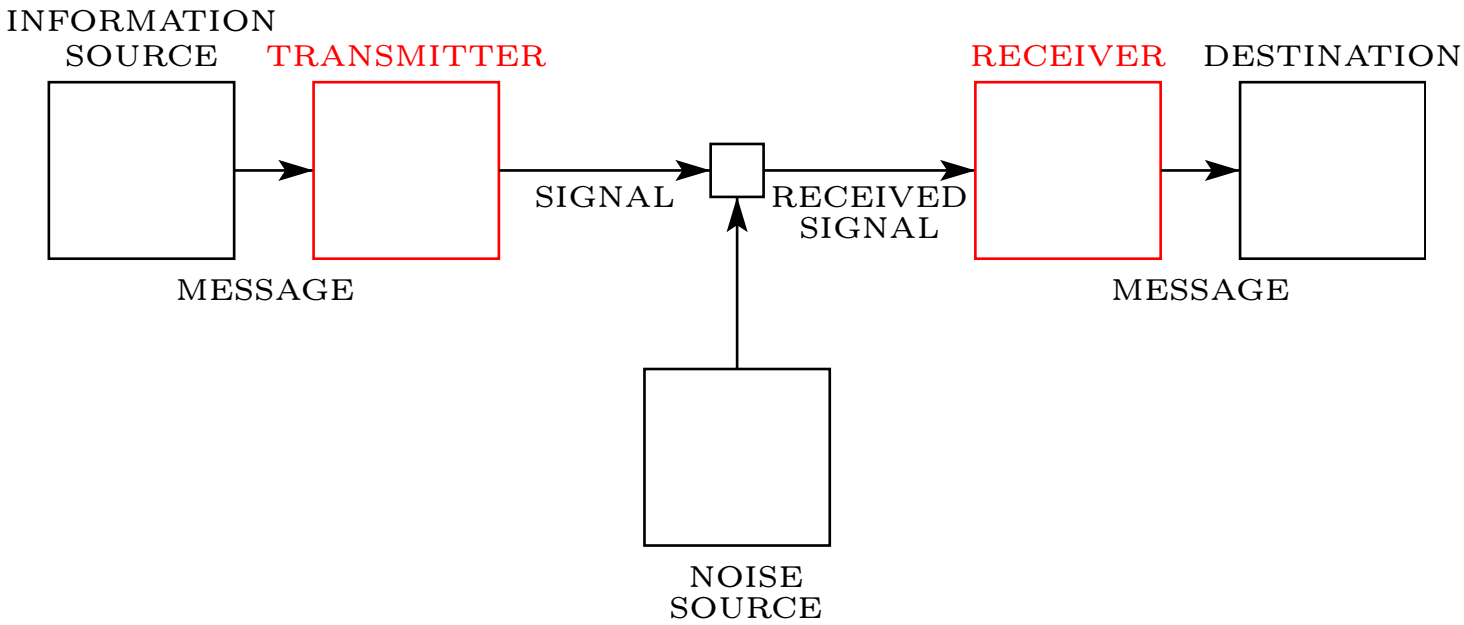
$$\Pr(Y_i = y, S_{i+1} = s'|X_i = x, S_i = s) = P(y, s'|x, s),$$

independently of the past X's, S's, Y's. A simplified model is obtained when there is a single state. The simplified model is specified by $P(y|x)$ and is known as the Discrete Memoryless Channel.

Reliable Communication



Reliable Communication



Given the source and channel, say that reliable communication is possible if for every $\epsilon > 0$ we can design the transmitter and receiver such that the probability of a source symbol is reproduced incorrectly is less than ϵ.

Discrete Channels with Noise

Shannon defines the capacity of a noisy channel as

$$C = \lim_{n} \sup_{X^n} \frac{1}{n} \left[H(X^n) - H(X^n | Y^n) \right] \qquad (?)$$

with the supremum taken over all sources X.

Discrete Channels with Noise

Shannon defines the capacity of a noisy channel as

$$C = \liminf_{n} \sup_{X^n} \frac{1}{n} \left[H(X^n) - H(X^n | Y^n) \right] \qquad (?)$$

with the supremum taken over all sources X.

Discrete Channels with Noise

Shannon defines the capacity of a noisy channel as

$$C = \sup_{X} \liminf_{n} \frac{1}{n} \left[H(X^n) - H(X^n | Y^n) \right] \qquad (?)$$

with the supremum taken over all sources X.

Discrete Channels with Noise

Shannon defines the capacity of a noisy channel as

$$C = \sup_{X} \text{p-liminf}_{n} \frac{1}{n} i(X^n; Y^n) \qquad (?)$$

with the supremum taken over all sources X.

Discrete Channels with Noise

Shannon defines the capacity of a noisy channel as

$$C = \lim_{n} \sup_{X^n} \frac{1}{n} \left[H(X^n) - H(X^n | Y^n) \right] \qquad (?)$$

with the supremum taken over all sources X.

Discrete Channels with Noise

Shannon defines the capacity of a noisy channel as

$$C = \lim_{n} \sup_{X^n} \frac{1}{n} \left[H(X^n) - H(X^n | Y^n) \right] \qquad (?)$$

with the supremum taken over all sources X.

Shannon's general channel model is slightly too general for all of his subsequent claims to hold. Nevertheless they do hold under mild regularity conditions (e.g., indecomposability) and his proof technique gives the right tools to attack more general cases.

Discrete Channels with Noise

Shannon defines the capacity of a noisy channel as

$$C = \lim_{n} \sup_{X^n} \frac{1}{n} \left[H(X^n) - H(X^n | Y^n) \right] \qquad (?)$$

with the supremum taken over all sources X.

Shannon's general channel model is slightly too general for all of his subsequent claims to hold. Nevertheless they do hold under mild regularity conditions (e.g., indecomposability) and his proof technique gives the right tools to attack more general cases.

To simplify the presentation let us proceed with the memoryless case which already illustrates the essential ideas.

Capacity of the DMC

- For a discrete memoryless channel all the expressions in the previous slide lead to the same value — the supremums are attained by a memoryless source — and are in a 'single letter' form:

$$C = \max_X H(X) - H(X|Y)$$

Capacity of the DMC

- For a discrete memoryless channel all the expressions in the previous slide lead to the same value — the supremums are attained by a memoryless source — and are in a 'single letter' form:

$$C = \max_X H(X) - H(X|Y)$$

- Shannon shows that, given an information source of entropy H and communication channel with capacity C reliable communication is possible if and only if $H \leq C$.

Transmitter/Receiver design by Probabilistic method

Shannon's proof of the 'noisy channel coding theorem' — that good transmitters and receivers exist — is not by an explicit construction but via the probabilistic method:

Transmitter/Receiver design by Probabilistic method

Shannon's proof of the 'noisy channel coding theorem' — that good transmitters and receivers exist — is not by an explicit construction but via the probabilistic method:

- Consider an auxiliary memoryless source that attains the max in the definition of C. Consider its transmission over the DMC. The pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ then form an i.i.d. sequence of random variables with entropy $H(XY)$.

Transmitter/Receiver design by Probabilistic method

Shannon's proof of the 'noisy channel coding theorem' — that good transmitters and receivers exist — is not by an explicit construction but via the probabilistic method:

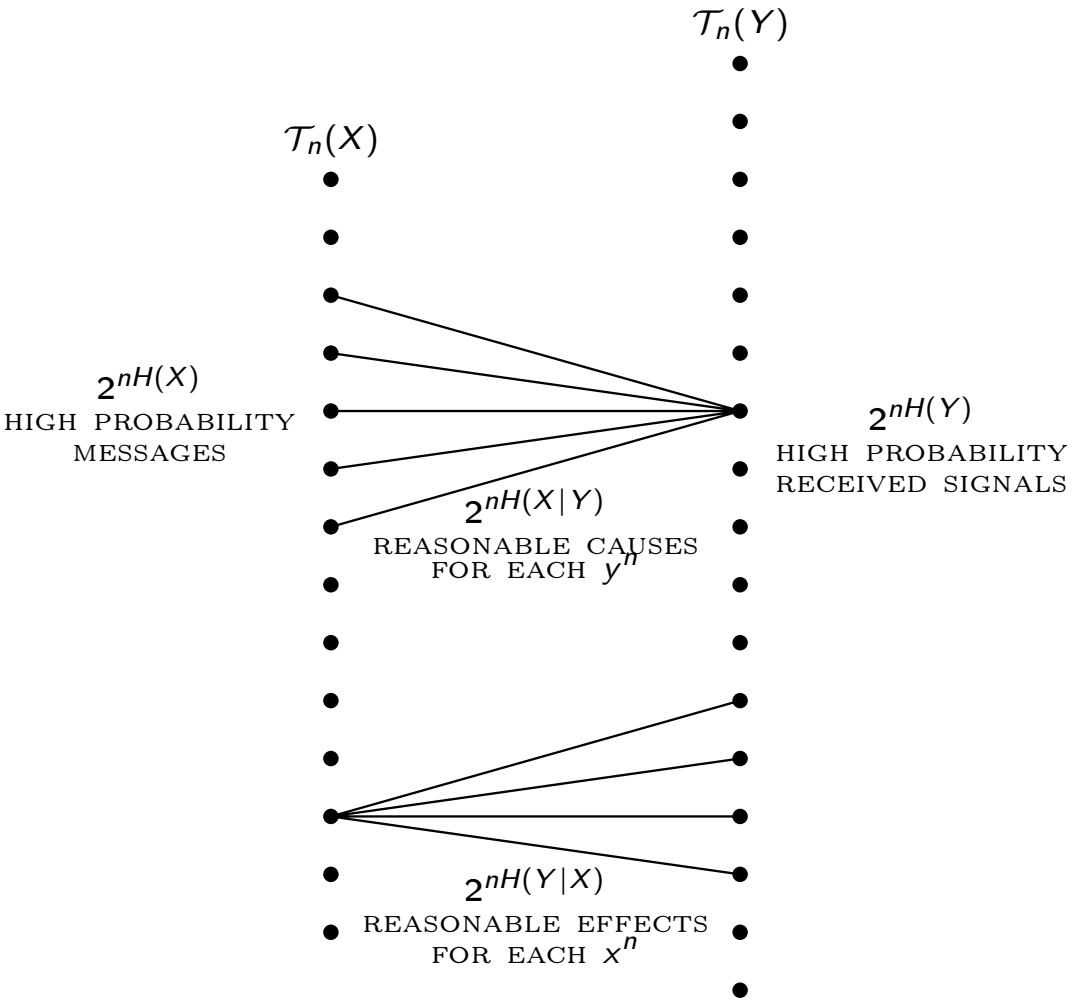
- Consider an auxiliary memoryless source that attains the max in the definition of C. Consider its transmission over the DMC. The pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ then form an i.i.d. sequence of random variables with entropy $H(XY)$.
- Thus there are $2^{nH(X)}$ typical x^n's; $2^{nH(Y)}$ typical y^n's; and $2^{nH(XY)}$ typical (x^n, y^n) pairs.

Transmitter/Receiver design by Probabilistic method

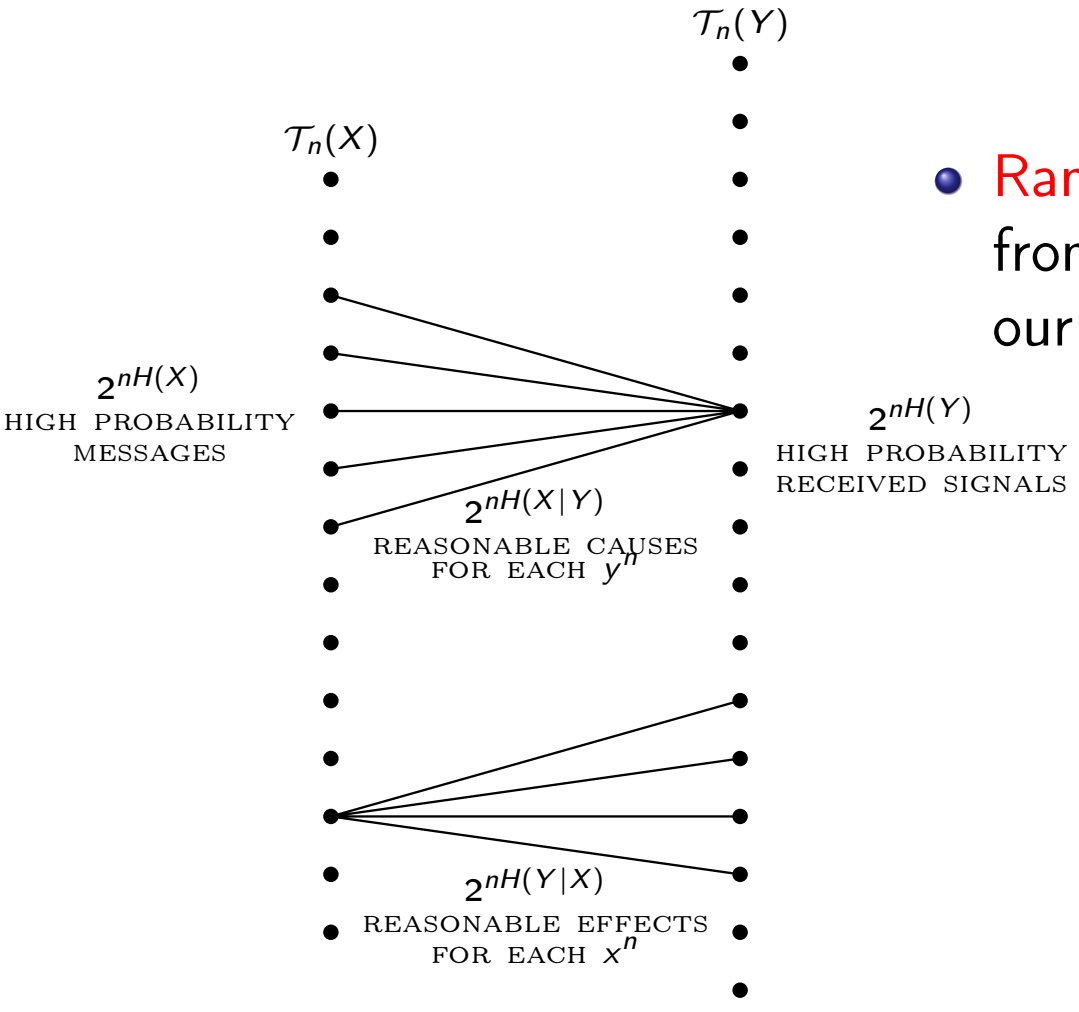
Shannon's proof of the 'noisy channel coding theorem' — that good transmitters and receivers exist — is not by an explicit construction but via the probabilistic method:

- Consider an auxiliary memoryless source that attains the max in the definition of C. Consider its transmission over the DMC. The pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ then form an i.i.d. sequence of random variables with entropy $H(XY)$.
- Thus there are $2^{nH(X)}$ typical x^n's; $2^{nH(Y)}$ typical y^n's; and $2^{nH(XY)}$ typical (x^n, y^n) pairs.
- The typicality picture now looks as follows:

Transmitter/Receiver design by Probabilistic method

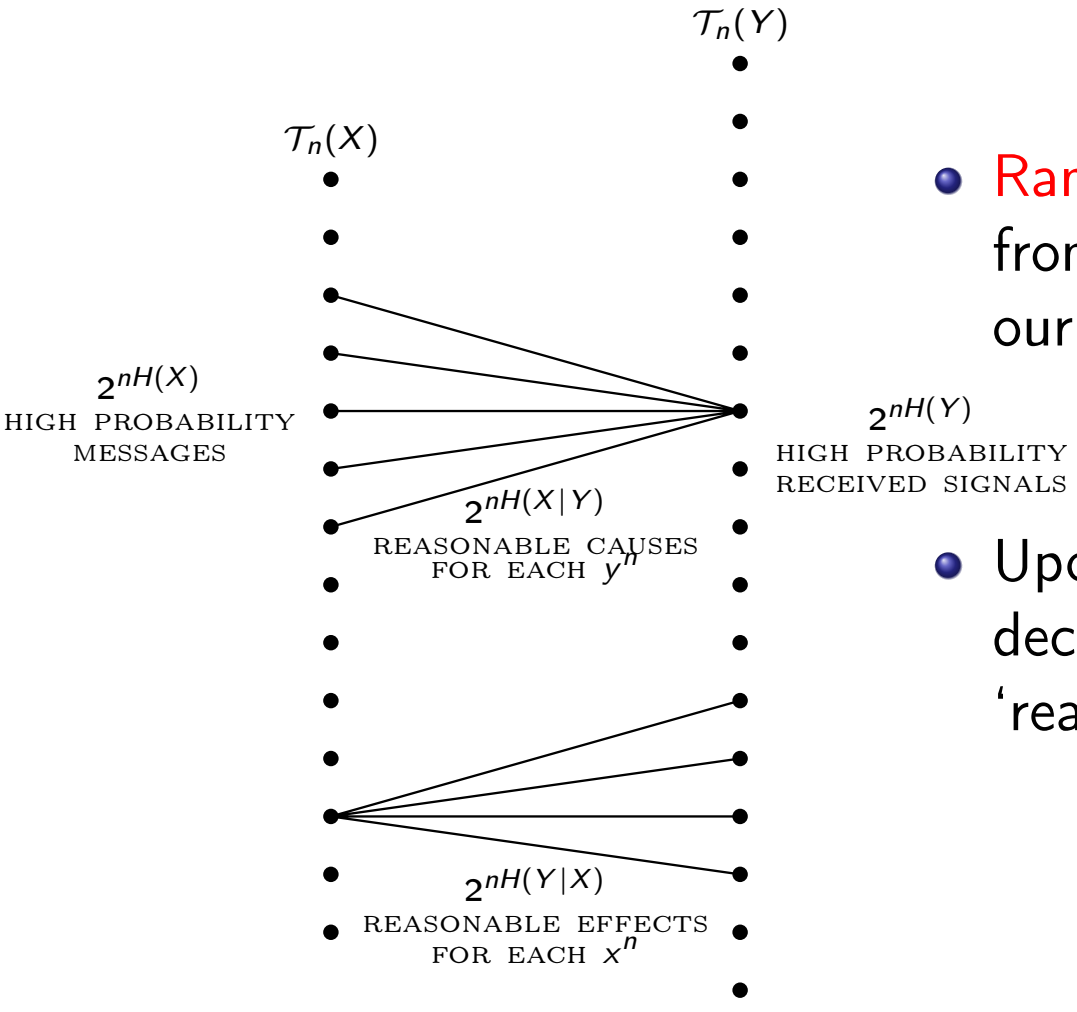


Transmitter/Receiver design by Probabilistic method



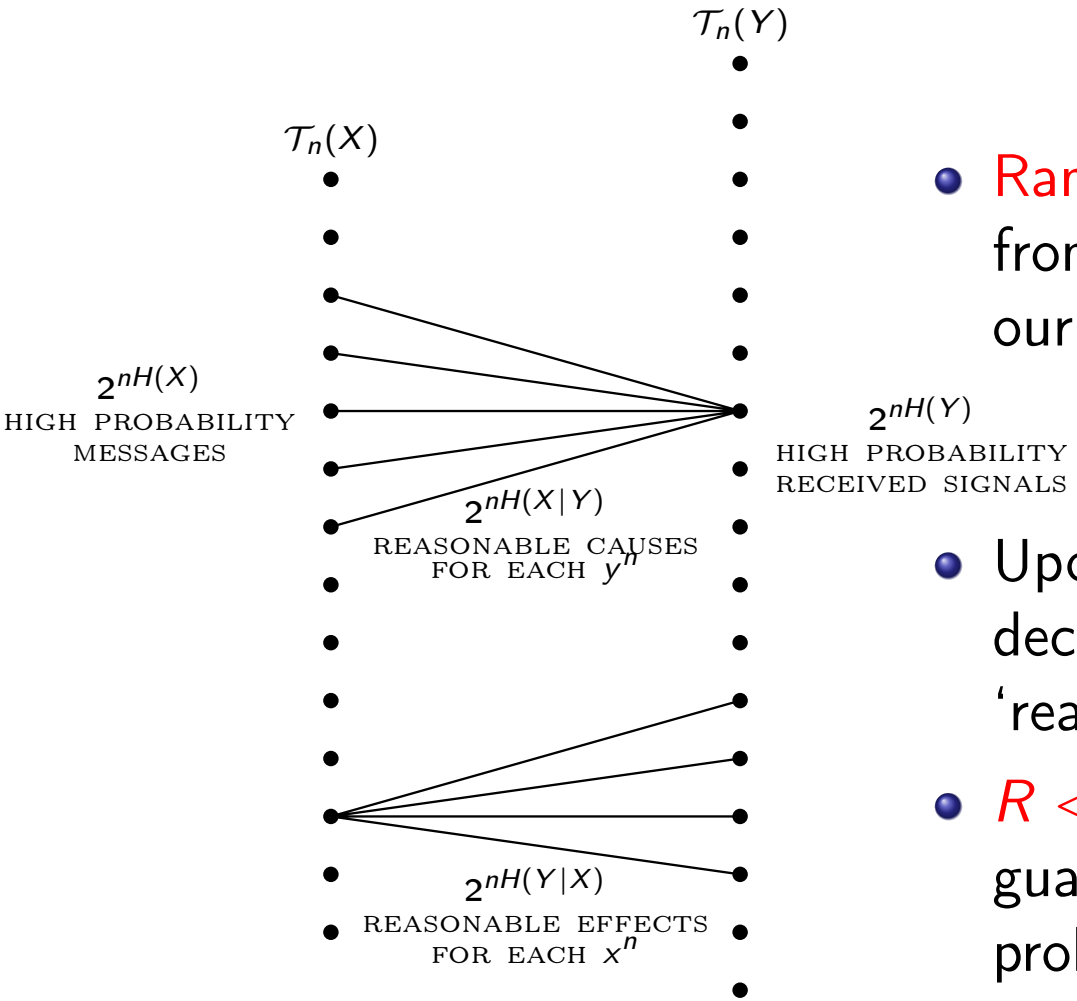
- Randomly choose 2^{nR} messages from the left column. This is our transmitter design.

Transmitter/Receiver design by Probabilistic method



- Randomly choose 2^{nR} messages from the left column. This is our transmitter design.
- Upon observing Y^n, the receiver declares a message that is a 'reasonable cause' of Y^n.

Transmitter/Receiver design by Probabilistic method



- Randomly choose 2^{nR} messages from the left column. This is our transmitter design.
- Upon observing Y^n, the receiver declares a message that is a 'reasonable cause' of Y^n.
- $R < H(X) - H(X|Y)$ guarantees that with high probability the true message will be receiver's only choice.

Converse

Shannon shows that if $H > C$ then no matter how the transmitter and receiver are designed $\frac{1}{n} H(U^n | Y^n)$ will be at least $H - C > 0$. In light of the later part of the paper that sets up rate–distortion theory, this is — in principle — sufficient to show that the symbol error probability cannot be made arbitrarily close to zero. Shannon does not state this explicitly. However he does state a 'strong converse' without proof.

Converse

Shannon shows that if $H > C$ then no matter how the transmitter and receiver are designed $\frac{1}{n}H(U^n|Y^n)$ will be at least $H - C > 0$. In light of the later part of the paper that sets up rate–distortion theory, this is — in principle — sufficient to show that the symbol error probability cannot be made arbitrarily close to zero. Shannon does not state this explicitly. However he does state a 'strong converse' without proof.

I have a preference for proving the converse for the symbol error probability. After all, the user of the communication system will interact with our design by supplying a sequence of symbols and may not even be aware of the designers notion of 'block', or whether the system works with block codes.

Modeling issues

Both for sources and channels, one can question if Shannon's model is appropriate, and if so, how accurate our knowledge of the 'true' source of channel can be.

Modeling issues

Both for sources and channels, one can question if Shannon's model is appropriate, and if so, how accurate our knowledge of the 'true' source of channel can be.

On the first question, Shannon's models have withstood the test of time — at this point we don't give much thought about the appropriateness. However, one should note that Shannon's techniques to show that the existence of good transducers depend crucially on accurate knowledge of the 'true' source or channel.

Modeling issues

Both for sources and channels, one can question if Shannon's model is appropriate, and if so, how accurate our knowledge of the 'true' source of channel can be.

On the first question, Shannon's models have withstood the test of time — at this point we don't give much thought about the appropriateness. However, one should note that Shannon's techniques to show that the existence of good transducers depend crucially on accurate knowledge of the 'true' source or channel.

Shannon's engineering instincts presumably did tell him that slightly imperfect knowledge of the model only slightly diminishes the achievable compression/transmission rates. Nevertheless, the development of universal compression schemes and coding theorems for the compound channels were a theoretical necessity.

Waveform sources and channels

Shannon next generalizes his source and channel models to those that produce and accept real-valued signals of continuous time. However, he dismisses with the continuity of time by assuming the signals are bandlimited to a certain band W, and thus, the during a time interval T (via the sampling theorem) can be specified by $n = 2WT$ samples.

Waveform sources and channels

Shannon next generalizes his source and channel models to those that produce and accept real-valued signals of continuous time. However, he dismisses with the continuity of time by assuming the signals are bandlimited to a certain band W, and thus, the during a time interval T (via the sampling theorem) can be specified by $n = 2WT$ samples.

While intuitively plausible, that this is (essentially) indeed the case took some effort to establish rigorously.

Entropy for real-valued random variables

Shannon generalizes the definition of entropy to real valued random variables as

$$H(X) = -\int f_X(x) \log f_X(x)\, dx,$$

and notes that unlike the discrete case, H is not invariant under one-to-one transformations.

Entropy for real-valued random variables

Shannon generalizes the definition of entropy to real valued random variables as

$$H(X) = -\int f_X(x) \log f_X(x)\, dx,$$

and notes that unlike the discrete case, H is not invariant under one-to-one transformations.

However, the difference $H(X) - H(X|Y)$ is invariant, and so, the notion of channel capacity is well defined. Furthermore, given a channel with real-valued inputs and outputs, one can use finer and finer quantization of its input and output to get a discrete channel that has capacity as close to the original as desired.

Gaussians and entropy power

Among random variables of given second moment σ^2 the Gaussian has the largest entropy $H(X) = \frac{1}{2}\log(2\pi e\sigma^2)$.

Gaussians and entropy power

Among random variables of given second moment σ^2 the Gaussian has the largest entropy $H(X) = \frac{1}{2}\log(2\pi e\sigma^2)$.

Shannon makes the claim that if X_1 and X_2 are independent random variables with entropies $\frac{1}{2}\log(2\pi eN_1)$ and $\frac{1}{2}\log(2\pi eN_2)$, then

$$H(X_1 + X_2) \geq \frac{1}{2}\log(2\pi e(N_1 + N_2))$$

This is the entropy power inequality. The argument Shannon gives is the only one in the paper that can't be turned into a real proof. This, if anything, is a testament to the strength Shannon's insight, and his acuity in sensing the truth.

Capacity of waveform channels

Generically, without some sort of cost constraint on the input signals the capacity of a channels with real-valued inputs/outputs will turn out to be infinity.
Shannon pays particular attention to additive noise channels. In the case of the AWGN with input power constraint, he derives its capacity as

$$C = W \log \frac{P + N}{N}$$

Rate distortion theory

The last topic Shannon touches upon is lossy compression: how many bits are required to quantize a source U if we are willing to tolerate some distortion in its reconstruction?

Rate distortion theory

The last topic Shannon touches upon is lossy compression: how many bits are required to quantize a source U if we are willing to tolerate some distortion in its reconstruction?

Shannon motivates the question from the case of continuous valued sources which generally require an infinite number of bits to represent exactly and thus distortion is a necessary evil. However, his treatment is completely general.

Rate distortion theory

Shannon makes the case that the only natural way to evaluate the quality of a quantization scheme is to measure its distortion by

$$E[\rho(U, V)]$$

where U is a source letter, V is the reconstruction of U from its quantization, ρ is some cost function giving the the penalty of representing u by v.

Rate distortion theory

Given an information source U, a distortion measure $\rho(u, v)$, a tolerable distortion ν, and a communication channel with capacity C, can we design a transmitter and receiver so that the information is reconstructed at the destination within the tolerable distortion?

Rate distortion theory

Given an information source U, a distortion measure $\rho(u, v)$, a tolerable distortion ν, and a communication channel with capacity C, can we design a transmitter and receiver so that the information is reconstructed at the destination within the tolerable distortion?

Shannon shows that the answer to this question is yes if and only if $C \geq R(\nu)$ where

$$R(\nu) = \lim_{n} \inf_{V^n} \frac{1}{n}[H(U^n) - H(U^n|V^n)]$$

where the inf is taken over all V^n jointly distributed with U^n such that $\frac{1}{n}\sum_i E[\rho(U_i, V_i)] \leq \nu$.

Rate distortion theory

Given an information source U, a distortion measure $\rho(u, v)$, a tolerable distortion ν, and a communication channel with capacity C, can we design a transmitter and receiver so that the information is reconstructed at the destination within the tolerable distortion?

Shannon shows that the answer to this question is yes if and only if $C \geq R(\nu)$ where

$$R(\nu) = \lim_{n} \inf_{V^n} \frac{1}{n}[H(U^n) - H(U^n|V^n)]$$

where the inf is taken over all V^n jointly distributed with U^n such that $\frac{1}{n}\sum_i E[\rho(U_i, V_i)] \leq \nu$.

In the case of a memoryless source this expression can be evaluated for $n = 1$. Shannon gives examples, notably when U is Gaussian and ρ is squared error.

Rate distortion theory

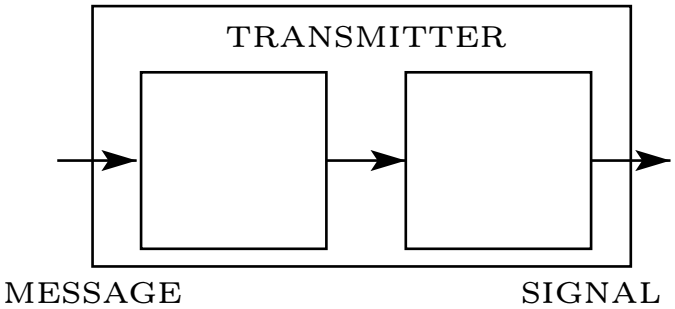
Shannon's proof of this 'rate–distortion theorem' follows the same logic as his proof of the noisy channel coding theorem and subject to similar caveats.

Rate distortion theory

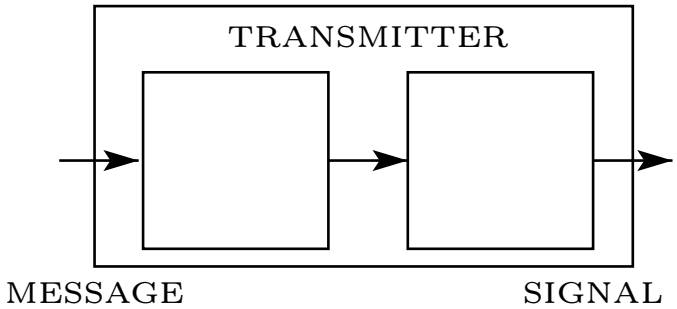
Shannon's proof of this 'rate–distortion theorem' follows the same logic as his proof of the noisy channel coding theorem and subject to similar caveats.

While proving the theorem Shannon also establishes (but does not comment on) an architectural principle: when a transmitter/receiver pair satisfying the distortion criteria can be designed, they can designed in a modular way.

Modular design – Universal digital interface

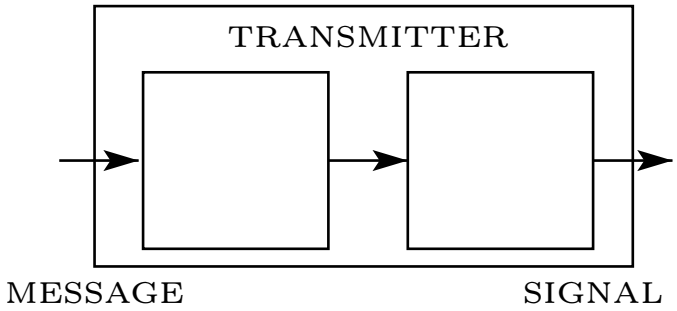


Modular design – Universal digital interface



The transmitter can be implemented in two steps: (1) a source coder (designed with the knowledge only of p_U, ρ, ν) that maps the source into $R(\nu)$ bits per source letter, (2) a channel coder (designed with the knowledge only of channel statistics) that maps bit sequences into channel input sequences. Similarly at the receiver (1) a channel decoder recovers the bits, (2) a source decoder maps the bits into the reconstruction of the source.

Modular design – Universal digital interface



The transmitter can be implemented in two steps: (1) a source coder (designed with the knowledge only of p_U, ρ, ν) that maps the source into $R(\nu)$ bits per source letter, (2) a channel coder (designed with the knowledge only of channel statistics) that maps bit sequences into channel input sequences. Similarly at the receiver (1) a channel decoder recovers the bits, (2) a source decoder maps the bits into the reconstruction of the source.

Consequently bits (or any other digital format) may be used as a universal interface between sources and channels without sacrificing feasibility. *This* is precisely what distinguishes 'digital' from 'analog' communication systems.

Remarks

Shannon's "A mathematical theory of communication" is a most eloquent elaboration of an end-to-end theory of communication systems. It never fails to amaze no matter how often one consults it. How often do we see a research field arise fully formed from the mind of a single man?

Remarks

Shannon's "A mathematical theory of communication" is a most eloquent elaboration of an end-to-end theory of communication systems. It never fails to amaze no matter how often one consults it. How often do we see a research field arise fully formed from the mind of a single man?

Despite its mathematical nature AMTC is written in a very didactic style. Criticism along the lines 'Shannon did not dot this i and cross that t' completely misses the point. A fully rigorous treatment would not have had a fraction of the impact of AMTC (and would have arrived a decade too late).

Remarks

Any mathematical theory is concerned with mathematical models of reality, not reality itself. In a mathematical theory for an engineering discipline one has to make a tradeoff: if the models are very accurate, they represent reality well, but they are hard to analyze/understand/intuit; if they are too simple, they ignore crucial parts of reality to the point of being useless to analyze/understand/.... What AMTC has done for communication engineering has no parallel in any engineering field.