

Bell Labs celebrates 50 years of Information Theory

An Overview of Information Theory

Humans are symbol-making creatures. We communicate by symbols -- growls and grunts, hand signals, and drawings painted on cave walls in prehistoric times. Later we developed languages, associating sounds with ideas. Eventually Homo Sapiens developed writing, perhaps first symbols scratched on rocks, then written more permanently on tablets, papyrus, and paper.

Today, we transmit symbols -- coded digital signals of voice, graphics, video, and data -- around the world at close to the speed of light. We're even sending signals into outer space in the hope of finding other symbol-creating species.

Beginning of Information Theory

Our ability to transmit signals at billions of bits per second is due to an inventive and innovative Bell Labs mathematician, Claude Shannon, whose "Mathematical Theory of Communications" published 50 years ago in the *Bell System Technical Journal* has guided communications scientists and engineers in their quest for faster, more efficient, and more robust communications systems. If we live in an "Information Age," Shannon is one of its founders.

Shannon's ideas, which form the basis for the field of Information Theory, are yardsticks for measuring the efficiency of communications systems. He identified problems that had to be solved to get to what he described as ideal communications systems -- a goal we have yet to reach as we push today the practical limits of communications with our commercial gigabit- and experimental terabit-per-second systems.

Shannon also told us something that we thought we intuitively knew, but really didn't -- what "information" really is -- and he permitted us to find shortcuts in communicating more effectively.

In defining information, he identified the critical relationships among the elements of a communication system -- the power at the source of a signal; the bandwidth or frequency range of an information channel through which the signal travels; and the noise of the channel, such as unpredictable static on a radio, which will alter the signal by the time it reaches the last element of the system, the receiver, which must decode the signal. In telecommunications, a channel is the path over a wire or fiber, or in wireless systems, the slice of radiospectrum used to transmit the message through "free space."

Shannon's equations told engineers how much information could be transmitted over the channels of an ideal system. He also spelled out mathematically the principles of "data compression," which recognize what the end of this sentence demonstrates, that "only information essential to understanding must be transmitted." And he showed how we could transmit information over noisy channels at error rates we could control.

Shannon's theory has been likened to a lighthouse. Its beacon tells communications scientists and engineers where they are, where they're going, how far they must go, and significantly, when they can stop.

The only thing his theory doesn't explain is how to get there. And there the challenges lie.

Growth of System Capacity

When Shannon announced his theory in the July and October issues of the *Bell System Technical Journal* in 1948, the largest communications cable in operation at that time carried 1,800 voice conversations. Twenty-five years later, the highest capacity cable was carrying 230,000 simultaneous conversations. Today a single strand of Lucent's recently announced WaveStar™ optical fiber--as thin as a human hair--can carry more than 6.4 million conversations. Or it can transmit the contents of 90,000 encyclopedias in just one second.

Even with these high speeds, today's communications systems don't approach the theoretical limits of fiber, wireless, and other systems. A single optical fiber strand, in theory, might transmit up to 100 quadrillion conversations (1 followed by seventeen 0's), each encoded at 64,000 bits per second.

Nor are communications scientists and engineers happy with the current high rates. They want more, because we need more. And Shannon's equations, 50 years later, are still showing us the way.

Understanding Information Theory

Understanding Shannon's equations, the basis of Information Theory, is not an easy matter. His work is abstract and subtle, the world of mathematicians and engineers, even though we see it has everyday consequences.

To get a high-level understanding of his theory, a few basic points should be made.

First, words are symbols to carry information between people. If one says to an American, "Let's go!", the command is immediately understood. But if we give the commands in Russian, "Pustim v xod!", we only get a quizzical look. Russian is the wrong code for an American.

Second, all communication involves three steps – coding a message at its source, transmitting the message through a communications channel, and decoding the message at its destination.

In the first step, the message has to be put into some kind of symbolic representation – words, musical notes, icons, mathematical equations, or bits. When we write "Hello," we encode a greeting. When we write a musical score, it's the same thing – only we're encoding sounds.

For any code to be useful it has to be transmitted to someone or, in a computer's case, to something. Transmission can be by voice, a letter, a billboard, a telephone conversation, a radio or television broadcast, or the now ubiquitous e-mail. At the destination, someone or something has to receive the symbols, and then decode them by matching them against his or her own body of information to extract the data.

Fourth, there is a distinction between a communications channel's *designed symbol rate* of so many bits per second and its actual *information capacity*. Shannon defines channel capacity as how many kilobits per second of user information can be transmitted over a noisy channel with as small an error rate as possible, which can be less than the channel's "raw" symbol rate.

Shannon describes the elements of communications system theory as a source--encoder--channel--decoder--destination model. What his theory does is to replace each element in the model with a mathematical model that describes that element's behavior within the system.

The Meaning of Information

"Information" has a special meaning for Shannon. For years, people deliberately compressed telegraph messages by leaving certain words out, or sending key words that stood for longer messages, since costs were determined by the number of words sent. Yet people could easily read these abbreviated messages, since they supplied these predictable words, such "a" and "the."

In the same vein, for Shannon, information is symbols that contain unpredictable news, like our sentence, "only information essential to understanding must be transmitted."

The predictable symbols that we can leave out, which Shannon calls redundancy, are not really news.

Another example is coin flipping. Each time we flip a coin, we can transmit which way it lands, "heads" or "tails," by transmitting a code of "zero" or "one." But what if the coin has two "heads" and everyone knows it? Since there is no uncertainty concerning the outcome of a flip, no message need be sent at all.

Although this view might seem like common sense today, it was not always so. Shannon made clear that uncertainty or unpredictability is the very commodity of communication.

Encoding a Message

Shannon equates information with uncertainty. For Shannon, an information source is someone or something that generates messages in a statistical fashion. Think of a speaker revealing her thoughts one letter at a time. From an observer's point of view each letter is chosen at random, although the speaker's choice may depend on what has been uttered before, while for other letters there may be a considerable amount of latitude.

The randomness of an information source can be described by its "entropy." The operational meaning of entropy is that it determines the smallest number of bits per symbol that is required to represent the total output.

As an illustration, suppose we are watching cars going past on a highway. For simplicity, suppose 50% of the cars are black, 25% are white, 12.5% are red, and 12.5% are blue. Consider the flow of cars as an information source with four words: black, white, red, and blue. A simple way of encoding this source into binary symbols would be to associate each color with two bits, that is:

black = 00, white = 01, red = 10, and blue = 11, an *average* of 2.00 bits per color.

A Better Code Using Information Theory

However, by properly using Information Theory, a better encoding can be constructed by allowing for the frequency of certain symbols, or words:

black = 0, white = 10, red = 110, blue = 111.

How is this encoding better? With this code, the average number of bits per car will be less:

$$0.50 \text{ black} \times 1 \text{ bit} = .500$$

$$0.25 \text{ white} \times 2 \text{ bits} = .500$$

$$0.125 \text{ red} \times 3 \text{ bits} = .375$$

$$0.125 \text{ blue} \times 3 \text{ bits} = \underline{.375}$$

Average-- 1.750 bits per car

Furthermore Information Theory tells us that the entropy of this information source is 1.75 bits per car and thus no encoding scheme will do better than the scheme we just described.

In general, an efficient code for a source will not represent single letters, as in our example above, but will represent strings of letters or words.

If we see three black cars, followed by a white car, a red car, and a blue car, the sequence would be encoded as 00010110111, and the original sequence of cars can readily be recovered from the encoded sequence.

The theory also says how complex a code needs to be for a given complexity. As a general rule, the closer one compresses a source to its entropy, the more complex the code will become.

Defining a Channel's Capacity

Having compressed the source output to a sequence of bits, we must transmit them.

In Information Theory the medium of transmission is called a channel, which could, for example, accept as input one of 256 symbols (i.e., 8 bits) 8,000 times per second and deliver those symbols intact to its receiver. Take, as an example, a DS0 telephone channel of 64,000 bits per second. If the output symbols are identical to the input symbols, the channel is noiseless, and its information carrying capacity is $8 \text{ bits/symbol} \times 8000 \text{ symbols/second} = 64000 \text{ bits/second}$. The channel's designed symbol rate and its capacity are the same.

Matters are more complex if the channel, as in most cases, has noise. For example, suppose that the channel accepts 8 bits 16,000 times per second for a total of 128,000 bits per second, but the bits that it delivers to its receiver are noisy: 90% of the time an output bit is identical to the corresponding input bit and 10% of the time it is not, that is, a 0 appears instead of a 1, or vice versa.

Information Theory tells us that the capacity of the channel in the above example is 67,840 bits per second. This means that for any desired data rate less than 67,840 bits per second -- no matter how close, and any desired error rate -- no matter how small, we can, by proper encoding, communicate at this

data rate over this noisy channel and make errors at a rate not exceeding the desired error rate.

For example, we can use this noisy channel to communicate at a DS0 rate of 64,000 bits per second and make only one error every billion bits (10^{-9} error rate).

Note that the channel's designed symbol rate operates at a 128,000 bits per second rate but its output at that rate is unreliable. According to Information Theory, the channel's capacity is 67,840 bits per second, which allows us to communicate reliably at a DS0 rate of 64,000 bits per second. If we devote half the channel's 128,000 bits we send to error correction, we reduce our throughput by half but achieve reliability.

Deliberately Introducing Redundancy

Information Theory tells us more about this channel -- reliable data transmission at rates above its channel capacity of 67,840 bits per second is not possible by any means whatsoever.

A simple way of combating noise is repetition -- to get a smaller probability of error, repeat the information symbol a certain number of times. One problem with this method of repetition is that we will make the effective information transmission rate smaller and smaller as we desire lower and lower error probability.

Again, however, Information Theory comes to our rescue. It says that one need not lower the transmission rate to anything below channel capacity to achieve smaller error probabilities. As long as the user's information rate is less than the capacity of the channel, it is possible to use error correction codes to achieve as small a probability of error as desired. However, in general, the smaller the desired error probability, the more complex the design of such an error correcting code.

How Fast Can We Go?

Encoding techniques for video transmission also owe a debt of gratitude to Shannon. To transmit into a home a full-motion studio-quality TV signal would require 70,000,000 bits per second, far too many bits to make it economically practical even using the high bandwidth of fiber optics

However, video compression techniques, such as Bell Labs patented perceptual audio coding (PAC) algorithm encoding scheme, has greatly reduced the number of bits necessary for transmission, now making video services economically possible.

Other encoding techniques for video conferencing permit acceptable video signals to be transmitted over channels at 368 kbps, 112 kbps, and even 56 kbps.

Continuing to Make Things Work

Research continues at Bell Labs in developing communications systems for the next century, including the Internet, wireless and fiber.

All digital transmission today, including the graphical representations that delight us on the Internet, owes a debt of gratitude to Claude Shannon, who told us it was all possible.