



Claude Shannon: Mastermind of Information Theory

George Strawn, NITRD

Claude Shannon (1916–2001) was one of the WWII-era geniuses who created the digital IT revolution. He contributed to both digital computing and digital communications and was a cryptographer during the war. He had many hobbies—some unusual, such as juggling and unicycling, and some conventional, such as chess. He also invented many devices, including rocket-powered flying discs, a motorized pogo stick, and a flame-throwing trumpet.

Here, I highlight Shannon's contributions to digital circuit theory and information theory, which he referred to as the "mathematical theory of communication."¹ I also examine the connection he initiated between information and physics.

A Formalism to Describe Digital Circuits

After studying electrical engineering and mathematics at the University of Michigan, Shannon began his graduate studies in electrical engineering at MIT in 1936. His background in mathematics had acquainted him with Boolean

Algebra, the algebra of logic. His master's thesis demonstrated that Boolean algebra provided a good formalism for expressing electrical circuits containing relays and switches. In 1987, Howard Gardner called Shannon's thesis "possibly the most important, and also the most famous, master's thesis of the century."² The results were published in Shannon's 1938 paper, "A Symbolic Analysis of Relay and Switching Circuits."³ Even though the work was motivated by telephones and electromechanical switches, after WWII, it became clear that the same formalism applied to computers and electronic switches built out of vacuum tubes (and then transistors and then chips).

Shannon's work connecting Boolean Algebra and electrical circuits opened doors for him. He spent 1940 at the Institute for Advanced Study in Princeton, where he met great scientists of the day, including Hermann Weyl, John von Neumann, Albert Einstein, and Kurt Gödel. After the US entered the war, he became a cryptanalyst and met Alan Turing, who journeyed to the US in 1943 to share Enigma

secrets. These connections, along with his own work on cryptography, deepened his thoughts about communication. In fact, according to Shannon, his wartime insights into communication theory and cryptography developed simultaneously: "they were so close together you couldn't separate them."⁴

A declassified version of his work in cryptography was published in 1948, in which he proved that, for any "perfect" (unbreakable) encryption, the key must be truly random, as large as the plaintext, never reused in whole or part, and kept secret.⁵ Given these requirements (especially the last one), perfect encryption seems unattainable in practice, which is similar to Turing's proof that "there can be no program that determines whether or not any given program will halt."⁶

Information Theory

The transmission of information across computer network channels and the storage and retrieval of information on disks can be broadly characterized as "communication," in the sense that networks communicate information across

space, and disks communicate it across time. Shannon's results apply to both these types of communication, which might be one reason it's now called "information theory."

Information theory is based on probability and statistics, and "information" here means not so much "what" you send but what you "can" send. Shannon developed a quantitative measure of information "uncertainty," which he called "entropy" because of its similarity to the concept of thermodynamic entropy. The entropy of a set of messages defines the uncertainty involved as the receiver tries to decide which of the possible messages was sent. If $M = \{m_1, m_2, \dots, m_n\}$ is a set of messages, and p_i is the probability that m_i was sent, then the entropy H of M is defined by

$$H = -(p_1 \times \log_2(p_1) + \dots + p_n \times \log_2(p_n)).$$

For example, if $n = 256 (= 2^8)$, and if each $p_i = 1/256$, then $H = 8$. This means that 8 bits of information are necessary to communicate one message out of the 256 equally probable messages. Each message is minimally represented as an 8-bit binary number (now called a byte), and any error in transmission would cause the wrong message to be selected.

At the other extreme, where one message is sent with a probability of 1, and all other messages have a probability of 0, then $H = 0$, which means there's no uncertainty and thus no real communication of new information. If the various m_i have differing probabilities, H will be greater than 0 and less than 8. If $H = 7$, for example, then only 7 bits would be required (on average) to send one of the messages. This is because more probable messages can be given shorter numbers and less probable ones longer numbers.

Shannon's fundamental theorem for a noiseless channel states that an encoding for the messages in a message set M with entropy H can always be found "very close to" the theoretical maximum of C/H messages per second, where C is the bits per second rate of the channel. A simple example is $M = \{a, b, c, d\}$, with $p(a) = 1/2$, $p(b) = 1/4$, and $p(c) = p(d) = 1/8$. In this example, $H = 7/4$. If $a = 0$, $b = 10$, $c = 110$, and $d = 111$, then the average message length is also $7/4$, making it as short as possible.

Things are more complicated when the probability of a message depends on the previous one, such as "u" having to follow "q" in English text. Creating such encodings to increase the number of messages transmitted per second is called *source coding*, and such coding can be *lossless* as in zip files or *lossy* as in JPEG and MP3 files. However, optimally short messages increase the harm of transmission errors. Shannon also studied *channel coding*, which is adding redundant information to a message to deal with transmission errors.

A naive solution to channel coding would be to send the same message multiple times and "vote," bit by bit, on the correct values. However, sending multiple copies of a message would greatly lower the rate of messages per second. Given the low bit-error rates of today's reliable hardware, checking for a single error bit in a byte is now more common. For example, adding one parity bit to each byte can detect a single bit error.

There is also a way to correct single bit errors, which Richard Hamming discovered.⁷ Hamming proposed including 3 parity bits as part of each 7-bit message (but the procedure can be generalized to n parity bits in messages of $2^n - 1$ bits). In the 7-bit message, bits 1, 2, and 4 are parity bits. Bit 4 is chosen to make the number a , determined by

bits $4 + 5 + 6 + 7$, even; bit 2 is chosen to make b , determined by bits $2 + 3 + 6 + 7$, even; and bit 1 is chosen to make c , determined by bits $1 + 3 + 5 + 7$, even. The reader can verify that the binary number $a'b'c'$, computed at the destination (where a' is set to 0 if the sum of bits $4 + 5 + 6 + 7$ is even, or 1 if that sum is odd; b' and c' are determined similarly), will be 0 if there is no error or, if there is a 1-bit error, it will be the position of that error. In general, Shannon's Noisy Channel Theorem asserts that it's always possible to find a coding scheme with redundancy and to get "arbitrarily close" to the maximum rate of transmission of messages over a given channel.

Many other error detection/correction methods are now known (such as cyclic redundancy codes and message signatures), but the discovery and development of these methods would likely have been delayed without Shannon's Noisy Channel Theorem proving which such things were possible. In addition to laying the groundwork for source coding and channel coding, Shannon's concept of information entropy also opened the door to a connection between information and physics.

The Physics of Information


As noted, Shannon's concept of information entropy was quite similar to the concept of thermodynamic entropy, developed by J. Willard Gibbs and Ludwig Boltzmann in the 1870s as they sought to combine earlier theories of energy with the atomic theory of matter. They showed that heat could be explained as a result of atoms in motion, with the faster atoms being "hotter." Thermodynamics dealt with uncertainty in heat like Shannon dealt with uncertainty in information. To illustrate this, consider a closed container with a partition dividing it into two parts, with hot atoms in one part and cold atoms in the other.

If the partition is removed, the two gases flow together and end up filling the container with a gas of intermediate temperature. The mixed gases have a higher entropy than the two gases of different temperatures (more uncertainty where the fast atoms are). The second law of thermodynamics says that this process is irreversible—that is, the hot and cold atoms can't separate themselves into a hot part and a cold part as they were before the mixing.

However, James Clerk Maxwell suggested the following thought experiment. Consider mixed gases in the container and assume that the partition has a "door" that can be opened and closed by a tiny demon. If the demon saw a fast atom approaching from the left, it would momentarily open the door to let it into the right part. Similarly, a slow atom from the right would be let into the left part. Eventually, the right part becomes hotter and the left becomes colder. Would not such a scheme violate the second law?

The answer turned out to be "no," because the information required by the demon to recognize fast and slow atoms "balances the account." Just as the separate laws of conservation of mass and energy had to be combined after it was understood that mass could be converted to energy, it is now

understood that thermodynamic entropy and information entropy are connected: an increase in thermodynamic entropy involves a loss of information and vice-versa. For example, the act of erasing a bit (losing that information) has to produce heat, which is an increase in thermodynamic entropy. Surprisingly, other computational actions that do not erase information can be accomplished without producing heat.⁸ The upshot of all this is that information is physical, not an ephemeral thing.^{9,10} Thus, we are in the early decades of the emergence of the science of information, which, if history is a guide, will provide new possibilities for the technology of information.

The range of Shannon's contribution to IT is breathtaking. He created an enduring formalism for describing computer circuits; he clarified and quantified the concept of information itself, while proving that it can be efficiently and reliably communicated; and he initiated the investigation of information as part of physics. These far-reaching contributions put Shannon among the highest ranks of IT masterminds. 

References

1. C.E. Shannon, "A Mathematical Theory of Communication," *Bell System*

Technical J., vol. 27, July and Oct. 1948, pp. 379–423 and 623–656.

2. H. Gardner, *The Mind's New Science: A History of the Cognitive Revolution*, Basic Books, 1987, p. 144.
3. C.E. Shannon, "A Symbolic Analysis of Relay and Switching Circuits," *Trans. Am. Inst. Electrical Engineers*, vol. 57, no. 12, 1938, pp. 713–723; doi:10.1109/T-AIEE.1938.5057767.
4. D. Kahn, *The Codebreakers*, Simon and Schuster, 1996, pp. 743–751.
5. C. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical J.*, vol. 28, no. 4, 1949, pp. 656–715.
6. G. Strawn, "Alan Turing," *IT Professional*, vol. 16, no. 1, 2014, pp. 5–7.
7. R.W. Hamming, "Error Detecting and Error Correcting Codes," *Bell System Technical J.*, vol. 29, no. 2, 1950, pp. 147–160; doi:10.1002/j.1538-7305.1950.tb00463.x. MR 0035935.
8. R. Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM J. Research and Development*, vol. 5, no. 3, 1961, pp. 183–191.
9. R. Landauer, "Information is Physical," *Proc. Workshop on Physics and Computation (PhysComp 92)*, 1993, pp. 1–4.
10. T. Siegfried, *The Bit and the Pendulum*, John Wiley, 2000.

George Strawn is director of the National Coordination Office for the Networking and Information Technology Research and Development Program (NITRD). Contact him at gostrawn@gmail.com.

IT Professional (ISSN 1520-9202) is published bimonthly by the IEEE Computer Society. IEEE Headquarters, Three Park Avenue, 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Circle, PO Box 3014, Los Alamitos, CA 90720-1314; voice +714 821 8380; fax +714 821 4010; IEEE Computer Society Headquarters, 1828 L St. NW, Suite 1202, Washington, DC 20036. Annual subscription: \$49 in addition to any IEEE Computer Society dues. Nonmember rates are available on request. Back issues: \$20 for members, \$143 for nonmembers.

Postmaster: Send undelivered copies and address changes to *IT Professional*, Membership Processing Dept., IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854-4141. Periodicals Postage Paid at New York, NY, and at additional mailing offices. Canadian GST #125634188. Canada Post Publications Mail Agreement Number 40013885. Return undeliverable Canadian addresses to PO Box 122, Niagara Falls, ON L2E 6S8, Canada. Printed in the USA.

Editorial: Unless otherwise stated, bylined articles, as well as product and service descriptions, reflect the author's or firm's opinion. Inclusion in *IT Professional* does not necessarily constitute endorsement by the IEEE or the Computer Society. All submissions are subject to editing for style, clarity, and space.



SOFTWARE DEVELOPER

Software Developer, Appl'ns. Dev./create/modify .NET appl'ns using C/C++/C# lang. & WPF techniques. Design software to integrate multiple video/graphics formats into .NET framework. U.S. Bach. or foreign equiv. in Comp. Sci. req'd. 5 yrs' exp. in progressively responsible post-baccalaureate software pos'ns req'd. Prior exp. must incl. software dev. for .NET app's using C/C++/C# lang. & WPF techniques. Prior exp. must incl. video/graphics integration. Stats LLC, Northbrook, IL. Resumes: Recruiting, PO Box 641152, Chicago, IL 60664.