Claude Elwood Shannon (1916–2001)

Solomon W. Golomb, Elwyn Berlekamp, Thomas M. Cover, Robert G. Gallager, James L. Massey, and Andrew J. Viterbi

Solomon W. Golomb

While his incredibly inventive mind enriched many fields, Claude Shannon's enduring fame will surely rest on his 1948 work "A mathematical theory of communication" [7] and the ongoing revolution in information technology it engendered.

Shannon, born April 30, 1916, in Petoskey, Michigan, obtained bachelor's degrees in both mathematics and electrical engineering at the University of Michigan in 1936. He then went to M.I.T., and after spending the summer of 1937 at Bell Telephone Laboratories, he wrote one of the greatest master's theses ever, published in 1938 as "A symbolic analysis of relay and switching circuits" [8], in which he showed that the symbolic logic of George Boole's nineteenth century *Laws of Thought* provided the perfect mathematical model for switching theory (and indeed for the subsequent "logic design" of digital circuits and computers). This work was awarded the prestigious Alfred Noble Prize of the combined engineering societies of the United States in 1940.

Spending the summer of 1938 at Woods Hole, Shannon decided to put the Mendelian laws of inheritance on a proper mathematical footing. His Ph.D. thesis in mathematics (M.I.T., 1940), "An algebra for theoretical genetics", was the result. Done in complete isolation from the community of population geneticists, this work went unpublished until it appeared in 1993 in Shannon's *Collected Papers* [5], by which time its results were known independently and genetics had become a very different subject. After his Ph.D. thesis Shannon wrote nothing further about genetics, and he expressed skepticism about attempts to expand the domain of information theory beyond the communications area for which he created it.

Starting in 1938 Shannon worked at M.I.T. with Vannevar Bush's "differential analyzer", the ancestral analog computer. After another summer (1940) at Bell Labs, he spent the academic year 1940–41 working under the famous mathematician Hermann Weyl at the Institute for Advanced Study in Princeton, where he also began thinking about recasting communications on a proper mathematical foundation. In 1941 he returned to Bell Labs for the next fifteen years, initially working on projects related to the war effort.

In 1945 Shannon wrote a classified report, "A mathematical theory of cryptography", which was finally declassified and published in 1949 in the *Bell System Technical Journal* (BSTJ) as the "Communication theory of secrecy systems" [6]. Perhaps it was from thinking about cryptography in terms of the set of all possible keys that might be used in the encryption of messages that Shannon was led to his breakthrough in "A mathematical theory of communication", published in two installments in the BSTJ in 1948.

At the start of this epic paper, he acknowledged the work at Bell Labs in the 1920s of Harry Nyquist

Solomon W. Golomb is professor of electrical engineering and mathematics at the University of Southern California. His e-mail address is milly@mizar.usc.edu. *His part of this article is based on his "Retrospective on C. E. Shannon" that appeared in the April 20, 2001, issue of* Science, *volume 292, page 455.*

(who contributed the "sampling theorem" and "Nyquist diagrams" to communication and control theory) and R. V. L. Hartley, who recommended a logarithmic measure of "information"; but Shannon, like Newton, "standing on the shoulders of giants", was able to see much farther than any of his predecessors. Early in the paper, he wrote "[The] 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" [Shannon's emphasis].

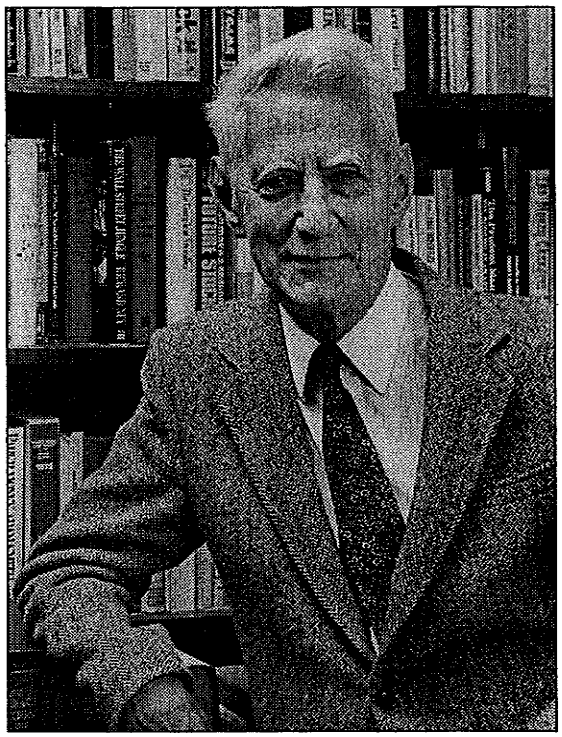
Shannon's great insight was to think in terms of statistical ensembles: the *source* as the set of all messages that might possibly be sent; and the *channel* contributing the set of possible disturbances or corruptions ("noise") to the message.

Shannon liberated the "entropy" of thermodynamics from physics and redefined it as a measure of uncertainty on probability distributions. While crediting the term "bit" (for "binary digit") to J. W. Tukey, Shannon defined his *bit* as the amount of information gained (or entropy removed) upon learning the answer to a question whose two possible answers were equally likely a priori. (When one possible answer is more likely than the other, learning the answer conveys less than one bit of information.) He derived formulas for the *information rate of a source* and for the *capacity of a channel* (in both the noiseless and noisy cases), each measured in *bits per second*, and he proved that for any information rate R less than the channel capacity C, it is possible (by suitable encoding) to send information at rate R, with an error rate less than any preassigned positive ϵ, over that channel. His ingenious proof considers the set of all possible encodings of source messages into streams of binary digits and shows that an encoding chosen "at random" from this set will have the desired property with extremely high probability.

When Shannon's paper appeared, some communications engineers found it to be too mathematical (there are twenty-three theorems!) and too theoretical, while some mathematicians criticized it as being insufficiently rigorous. In reality Shannon had almost unfailing instinct for what was actually *true* and gave outlines of proofs that other mathematicians (such as Khinchin and Wolfowitz) would make fully rigorous.

Since 1948 generations of coding theorists have struggled to find actual codes that perform as well as Shannon's "random" ones. Today there are communication systems operating over noisy channels within 0.005 dB of the Shannon limit, and *stored* information (in computers, on CDs and DVDs) is protected with the same types of "error-correcting codes" used for *transmitted* information.

Shannon also pioneered the study of "source coding" (or "data compression") to remove all



Claude Shannon

"useless" redundancy from source messages, which, if they are then to be sent over noisy channels, can have "useful" redundancy (extra symbols for error detection and correction) added back.

Shannon was grateful to Bell Labs for *tolerating* (though certainly not *encouraging*) his work on "A mathematical theory...", which seemed (at that time!) to have no practical benefit for AT&T. The name "Bell Labs" is now used by Lucent, while "AT&T Research Labs" has been renamed "Shannon Labs".

At the end of "A mathematical theory...", Shannon acknowledged the contributions of several colleagues at Bell Labs, with a special mention of the influence that Norbert Wiener's work at M.I.T. had on his own thinking.

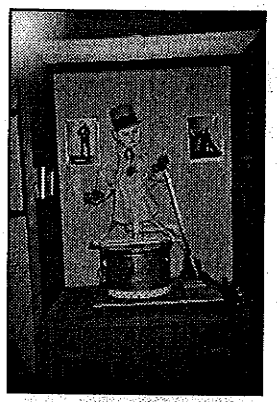
The *Shannon bit*, as the basic unit of information, though not a particle of physics, clearly has a reality of its own. There has been a serious proposal to rename this unit the *shannon*. If a message consists of N shannons, then the theoretically best "source encoding" could express it in N binary digits.

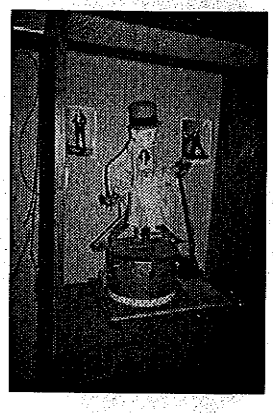
After 1948 Shannon wrote many more seminal papers in information theory. Also, in "Reliable circuits using less reliable relays" (BSTJ, 1956, in two installments coauthored with Edward F. Moore), he showed that arbitrarily reliable circuits could be built with unreliable parts, again using redundancy, akin to achieving arbitrarily reliable communication over unreliable (i.e., noisy) channels.

In 1956 Shannon left Bell Labs for M.I.T., where he was Donner Professor of Science from 1958 until his retirement in 1978. For decades, M.I.T. was

the leading university for information and communication theory.

The Information Theory Group of the Institute of Radio Engineers (IRE), founded in the early 1950s (later the "Information Theory Society" of the Institute of Electrical and Electronics Engineers (IEEE)), established the Shannon Award (originally called the "Shannon Lecture") as its highest honor. In 1973 Shannon himself delivered the first Shannon Lecture, at the International Symposium on Information Theory, in Ashkelon, Israel. When I spent most of fall 1959 visiting at M.I.T., I had gotten to know Shannon quite well, but it was an unexpected honor when Shannon attended my Shannon Lecture in 1985 in Brighton, England—the only one he attended after his own.





Shannon's juggling machine, constructed sometime around 1983. Built mostly from pieces of an erector set and dressed to look like W. C. Fields, the machine bounce juggles three small metal balls in a cascade off a drum. *Photos, courtesy of Arthur Lewbel and Betty Shannon, can be found at* http://www2.bc.edu/~lewbel/Shannon.html.

Shannon was a talented gadgeteer who built some of the earliest robotic automata, game-playing devices, and puzzle-solving machines. He could juggle while riding a unicycle and designed machines to juggle and to ride unicycle-like vehicles. Not working in any Nobel Prize field, but in the new science he had invented, he received innumerable honors and awards, including the U.S. National Medal of Science (1966), Israel's Harvey Prize (1972), and Japan's Kyoto Prize (1985). His research efforts bore bountiful fruit during his lifetime. His *Collected Papers* [5] include 127 publications from 1938 to 1982. The last few years of his life, Shannon was tragically afflicted with Alzheimer's disease. He died February 24, 2001, in Medford, Massachusetts, in his 85th year. Shannon is survived by his wife of more than fifty years, Mary Elizabeth (Betty), née Moore, a son Andrew, and a daughter Margarita. Another son, Robert, died in 1998.

Digital Communications, the title of a book I edited and coauthored with members of my JPL group, was still considered an oxymoron when the book appeared in 1964. (Dozens of similarly titled books have appeared since.) To most communications engineers, signals were quite obviously *analog*. But at Bell Labs in the late 1940s, the transistor was invented. With Shannon's remarkable theorems telling communications engineers what ultimate goals to strive for, and integrated circuits providing ever-improving hardware to realize these goals, the incredible digital communications revolution has occurred. (The theory of error-correcting codes also began in the late 1940s, largely independent of Shannon's work, with Richard W. Hamming at Bell Labs and Marcel Golay at IBM Research Labs.) It is no exaggeration to refer to Claude Shannon as the "father of the information age", and his intellectual achievement as one of the greatest of the twentieth century.

The following five contributions, all by winners of the Shannon Award, describe Shannon's influence and some subsequent developments in specific areas in which he pioneered. This presentation is a representative, but by no means exhaustive, indication of the many disciplines that Shannon's work profoundly enriched.

Elwyn Berlekamp

Shannon's Impact on Portfolio Theory

Shannon was quite interested in portfolio management. He personally gave several talks on the subject now called "financial mathematics" in the late 1960s and 1970s. He studied published data on large pension funds and other institutions to determine the net flows of cash into and out of the U.S. stock market, and he designed analog electric circuits intended to simulate the market. This work attracted little enthusiasm from financial professionals and was never published. But it did attract the interest of some within Shannon's circle, including Fano, whose unpublished work extended some of Shannon's.

John L. Kelly Jr. was a colleague of Shannon's at Bell Labs (although in a different department) who did publish. Kelly became interested in what is now called the "asset allocation problem", which is the question of how best to diversify one's total portfolio among different possible investments. Stated in the colorful terminology of horse racing, an investment opportunity may be attractive whenever the "true odds" known to the investor differ from the "betting odds" on which the payoffs are based. Kelly proved that given a long sequence of such opportunities, the maximum exponential growth rate that can be achieved with probability approaching 1 can be viewed as the capacity of the communication channel over which the investor receives his noisy tips!

The impact of this work emerged slowly but steadily over the subsequent decades. Here is an excerpt from a 1998 interview with Ed Thorpe, a very successful investor:

> A June 1998 *New York Times* Science Times article attributed the degrees of separation idea to a sociologist in 1967. Yet it was well known to Shannon

Elwyn Berlekamp is professor of mathematics at the University of California at Berkeley. His e-mail address is berlek@math.berkeley.edu.

before 1960. For bet sizing in favorable games, Shannon suggested I look at a 1956 paper by Kelly [4]. I did and adapted it as our guide for blackjack and roulette, and used it later in other favorable games, sports betting, and the stock market. The principle was to bet to maximize the expected value of the logarithm of wealth. This has desirable properties that are discussed in detail by Cover and Ordentlich [2].

The *IEEE Transactions on Information Theory* began publishing papers on portfolio theory around 1980. The topic received considerable attention in the IEEE Shannon lectures of 1990 and 1993, both later published in the *Information Theory Newsletter*. Yet all this work still attracted only skeptical attention from financial scholars in the leading business schools. Applying the Kelly criteria to their favorite model of stock price series (Brownian motion or white Gaussian noise) led to substantially more aggressive investments than portfolio managers had historically deemed prudent. So they tended to reject the Kelly criteria as unsound. However, real price movements have big swings considerably more often than predicted by the normal distribution. When the calculations are done correctly, most of the alleged overaggressiveness disappears.

Perhaps the impact Shannon and Kelly have had on finance can now best be measured by the number and quality of Wall Street firms that are actively recruiting mathematicians and information theorists, an outstanding example of which is documented in the cover story of the November 2000 issue of *Institutional Investor*.

Thomas M. Cover

Shannon's Contributions to Shannon Theory

Shannon's landmark 1948 contribution [7] initiating information theory presented a capacity theorem for the transmission of information, an entropy theorem for data compression, and an asymptotic equipartition theorem for the probability of sequences from an ergodic process. When the *Transactions on Information Theory* was formed in the mid-1950s, some areas closely related to Shannon's original work were naturally included in the purview of the journal. These other areas included prediction, estimation, filtering, modulation, and detection. Also came the quickly growing body of work in algebraic coding theory, an area that comprises roughly a third of the contributions in the journal each year. Other areas, like the mathematical theory of learning and algorithmic complexity, were soon to follow. Because of this proliferation, Aaron Wyner used the term "Shannon theory" in the early 1970s to designate those theorems growing directly out of the study of Shannon's work. Roughly speaking, Shannon theory involves problems in which mutual information and entropy play a prominent role.

Some of the initial reactions to Shannon's work were interesting. For example, the publisher insisted that Warren Weaver write an expository chapter for the book [9], presumably to make it more accessible. Since Shannon wrote as simply as possible, Weaver's task was impossible.

J. L. Doob [3] wrote in *Mathematical Reviews* in 1949, "The discussion is suggestive throughout, rather than mathematical, and it is not always clear that the author's mathematical intentions are honorable." I hasten to add that Doob has recanted this remark many times, saying that it and his naming of super martingales (processes that go down instead of up) are his two big regrets.

On the question of mathematical rigor, however, we should say that after fifty years, it is clear that Shannon was correct in each of his assertions and that his proofs, some of which might be considered outlines, could eventually be filled out along the lines of his arguments. It also must be said, given the breadth and scope of his theorems, that Shannon's intuition must have been anchored in a deep and natural theoretical understanding.

In the Soviet Union, Shannon's paper was considered to be in the field of cybernetics, which had been deemed [10] "a false science of obscurantists" (*izhenauka mrakobesov*). Even to publish its translation required special efforts. The great mathematician A. N. Kolmogorov became excited by Shannon's work and organized an informal seminar around these ideas in 1954. Those involved included I. M. Gelfand, A. M. Yaglom, M. S. Pinsker, R. L. Dobrushin, and Y. G. Sinai. Kolmogorov was eventually led to the definition of algorithmic complexity, the minimum length binary program needed for a Turing machine to print out a given sequence x. It turns out that the algorithmic complexity (pretty much simultaneously and independently put forth by Solomonoff, Chaitin, and Kolmogorov) is a very close counterpart to Shannon entropy.

Kolmogorov's attitude [1] expressed in 1983 was that "information theory must precede probability theory, and not be based on it."

Shannon's two most dominant theorems are on data compression and data expansion. In the data compression theorem, Shannon shows that there are 2^{nH} roughly equally probable sequences of length n from an ergodic stochastic process (H denotes entropy). The set of these so-called

Thomas M. Cover is professor of electrical engineering and statistics at Stanford University. His e-mail address is cover@stanford.edu.

typical sequences has probability nearly 1. Thus nH bits suffice to describe a sequence from such a process with a probability of description error arbitrarily small. Of course the general proof of this result for ergodic processes took more than thirty years. It was proved for independent identically distributed random variables by Shannon, who argued that it held for ergodic processes, but the rigorous proof was done in stages, first by McMillan for Markov processes, by Breiman for finite alphabet ergodic processes, extended to countably infinite alphabet processes by Chung, and proved in generality for arbitrary real-valued random variables forming an ergodic process by Barron and Orey.

The general asymptotic equipartition theorem, also known as the Shannon-McMillan-Breiman theorem, is that if $\{X_i\}$ is a stationary ergodic random process with probability mass function $p(\cdot)$, then

$$-(1/n)\log p(X_1,\ldots,X_n)$$

converges with probability 1 to H, where

$$H = \lim_{n\to\infty} H(X_n|X_{n-1},\ldots,X_1)$$

is the entropy rate of the process.

The other primary theorem of Shannon is the channel capacity theorem. Suppose one has a communication channel $p(y|x)$ with the understanding that the output Y is drawn according to $p(y|x)$ when x is the input. The question is how many distinguishable inputs are there? The capacity C is the logarithm of the number of distinguishable inputs. Shannon argued that if this situation is presented n times, so that $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$, then this communication channel takes on a nice structure. Let $(X, Y) \sim p(x, y)$, and define

$$H(X) = E\log(\tfrac{1}{p(X)}),$$

$$H(X|Y) = E\log(\tfrac{1}{p(X|Y)}),$$

$$I(X;Y) = H(X) - H(X|Y).$$

Fixing for a moment the type $p(x)$ of the input sequence, we can see that there are $2^{nH(X)}$ typical inputs, and for each input there are $2^{nH(Y|X)}$ roughly conditionally equally probable outputs Y^n. So we must merely count the number of distinguishable inputs in the sense that their output fans do not overlap. A simple sphere-packing argument shows that there can be no more than $2^{nH(Y)}/2^{nH(Y|X)}$ such distinguishable inputs. In fact, there are exactly that many, at least to first order in the exponent, as Shannon showed by introducing a random coding argument. He merely picked the 2^{nC} input sequences X^n at random, where $C = \max_{p(x)} I(X;Y)$.

Now I would like to comment on the research that these inquiries engendered. The first ideas in data compression were how to actually minimize the expected description length of a random variable X drawn according to a known probability mass function $p(x)$. Shannon suggested assigning a binary sequence of length $\lceil\log(\frac{1}{p(x)})\rceil$ to x. This achieves an expected description length within one bit of the entropy $H = \sum -p\log p$. Then Huffman found an algorithm for achieving the minimum. In practice today, one does not use Huffman coding but instead uses either arithmetic coding (mapping the source sequence $x_1, x_2, \ldots$ into the unit interval via the distribution function $F(.x_1x_2\ldots)$, thereby giving a uniform distribution) or Lempel-Ziv data compression in which one keeps track of each new phrase in the data sequence as it evolves and describes the next phrase by reference to the past ones. Since the phrases one is likely to see are the so-called typical ones, one has a very efficient reference library with respect to which one can describe the next phrase.

In 1961 Shannon wrote an important paper on the communication capacity of the two-way channel in which two senders interfere with each other as they try to talk over a common communication line. The simplest example of this is the binary multiplier channel in which the senders send either a 0 or a 1 and receive the product of what they send. Thus if both senders receive a 1, they know of course that they sent a 1 and that a 1 was transmitted by the other. On the other hand, if one sends a 1 and the other a 0, the first sender will know that a 0 was sent, but the second sender will not. To this day the capacity region of this channel is not known. This is one of many unsolved problems in network information theory.

I first met Shannon in 1972 in Ashkelon, Israel, a few years after he had retired from research. He had been asked to give the first Shannon Lecture and was delighted by the prospect, mostly because of the recursive aspect. The lecture was on feedback, which he illustrated with Campbell soup cans on which were pictures of Campbell soup cans, sounds built up of sounds of sounds, and lecturers receiving their own awards.

As for his place in history, Shannon blasted three fields into existence. First, switching theory, a subject that benefits from a mathematical foundation, but turns out not to be intrinsically deep. Then cryptography, where he illuminated an already existent highly mathematical subject. And finally, out of the blue, information theory, with its deep penetration of the mathematics of stochastic processes, the definition of intrinsic randomness, and the capacity relation between cause and effect—a whole beautiful field based on the ineffable idea of information. This ability to create new fields and develop their form and depth surely places Shannon in the top handful of creative minds of the century.

Robert G. Gallager

Shannon at M.I.T.

Claude Shannon spent both his graduate years and the latter half of his professional career at the Massachusetts Institute of Technology. He joined the M.I.T. electrical engineering department as a research assistant in 1936 to work on Vannevar Bush's differential analyzer, an early analog computer. While working on the analog gear mechanisms, he also became interested in the switching circuits that controlled the analyzer. He combined this experience, and a summer assignment at Bell Labs, with an undergraduate course in Boolean algebra to see that Boolean algebra was the right mathematical approach to the analysis of switching circuits. After fleshing this idea out, he wrote it up for his master's thesis.

This thesis, and the published version [8], won him both fame and the prestigious Alfred Noble Prize for the best engineering paper of the year by an author under thirty. This paper is now recognized as the foundation of modern switching theory and was crucial for the growth of both the computer industry and the telephone industry.

Partly under Vannevar Bush's advice, Shannon began to study genetics. He switched to the mathematics department to do his Ph.D. work on a mathematical foundation for genetics. He continued his interests in switching and his burgeoning interest in communication theory while doing the thesis, and quickly left the genetics field after completing his thesis in 1940. The thesis work was unpublished and remained unknown until recently. His results would have been very important if known earlier, but most of the results have since been rediscovered independently.

After a very fruitful fifteen years at Bell Labs, Claude Shannon returned to M.I.T. in 1956, first as a visiting professor, and then, in 1958, as Donner Professor of Science, with a joint appointment in electrical engineering and in mathematics. There was a very active group in information theory at M.I.T. at that time, and students and younger faculty viewed Shannon as an idol. Many of these students are now leaders in the digital communication field, some in academic careers, some in industrial laboratories, and some as entrepreneurs of large successful corporations.

Shannon was somewhat inner directed and shy, but very easy to talk to after the initial connection had been made. It was relatively rare for him to be the actual supervisor of a thesis, but in many cases, when he talked to a student, he would find an interesting and novel new direction for the student's work. Although these interactions were not frequent, they were extremely important, since students learned to focus on the formulation and approach to a problem rather than getting immediately involved in technical details.

Photograph courtesy of the M.I.T. Museum.

Claude Shannon, early years at MIT.

Shannon did not teach ordinary courses, but would give relatively frequent seminars, and once gave an entire seminar course with new results at each lecture. He did not like to replow old ground, and was so creative that, if he started to think of something old, he would look at it in a different way and create something entirely novel. He also disliked writing papers, although he recognized the need for doing this. Fortunately, he could work out all the needed details for a paper in his head, and then dictate the paper in virtually finished form.

While at M.I.T., he fleshed out quite a few results from his masterwork, *The Mathematical Theory of Communication.* Several of these papers developed bounds on achievable error probability for coding on noisy channels. These bounds were important, both to give an indication of whether noisy channel coding could be practical, and also to provide guidance on what types of schemes would work. Essentially he showed that almost any code would work very well and that the problem was in finding implementable decoders. It was during this period also that he looked at problems of feedback, side information, and interference from other channels.

Shannon always tended to work (or play) with many different types of problems at the same time. Along with writing his information theory papers at M.I.T., he was developing optimal portfolio theories for the stock market, continuing his interest in chess-playing machines, and investigating many other topics. Increasingly in his later years at M.I.T., he worked at home, and eventually retired in 1978.

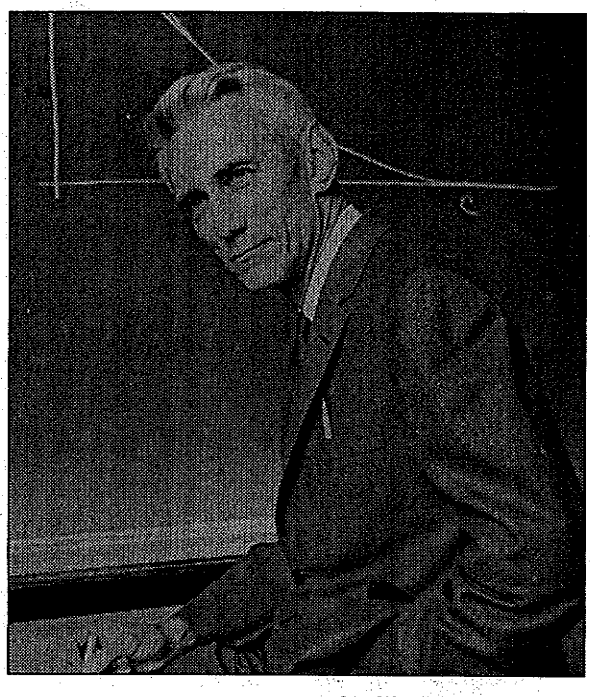
Robert G. Gallager is professor emeritus of electrical engineering and computer science at M.I.T. His e-mail address is gallager@lids.mit.edu.

James L. Massey

Shannon and the Development of Cryptography

Shannon's published work on cryptography is limited to the single paper "Communication theory of secrecy systems" [6], which appeared in October 1949. The first footnote in this paper indicates

James L. Massey is professor emeritus at Eidgenössische Technische Hochschule, Zürich. His e-mail address is JamesMassey@compuserve.com.



Photograph courtesy of the M.I.T. Museum.

At the blackboard, M.I.T.

that its contents had appeared as a September 1945 confidential Bell Laboratories memorandum that was now declassified. There has been speculation that Shannon's work on cryptography during the war (he wrote another Bell Laboratories memorandum on the subject in May 1943) led him to his formulation of information theory, but this seems not to be true. His October 1949 paper begins: "The problems of cryptography and secrecy systems furnish an interesting application of communication theory", and he later confirmed that this was indeed the motivation for his interest in cryptography. Whatever its source, there is no doubt that this paper is one of Shannon's "blockbusters" and that it has had an enormous influence on the subsequent development of cryptography. It is also vintage Shannon: insightful definitions, elegant proofs, and grand scope.

The title of Shannon's October 1949 paper is itself significant. It is now generally understood that cryptographic techniques have two quite independent goals, secrecy and authenticity. Shannon makes it clear that he is dealing only with secrecy. It would take another thirty-five years before a theory of authenticity, roughly on a par with that for secrecy provided by Shannon, was published by G. J. Simmons.

Shannon makes it very clear that there are two basic types of secrecy systems: those designed to protect against an attacker with unlimited computational resources and those designed to protect against an attacker with a given finite computational capability. Shannon called the kind of secrecy achieved by the former "theoretical secrecy" and that furnished by the latter "practical secrecy"—these terms have been replaced by "unconditional security" (or sometimes "information-theoretic security") and "computational security" in modern usage, but their meaning is unchanged.

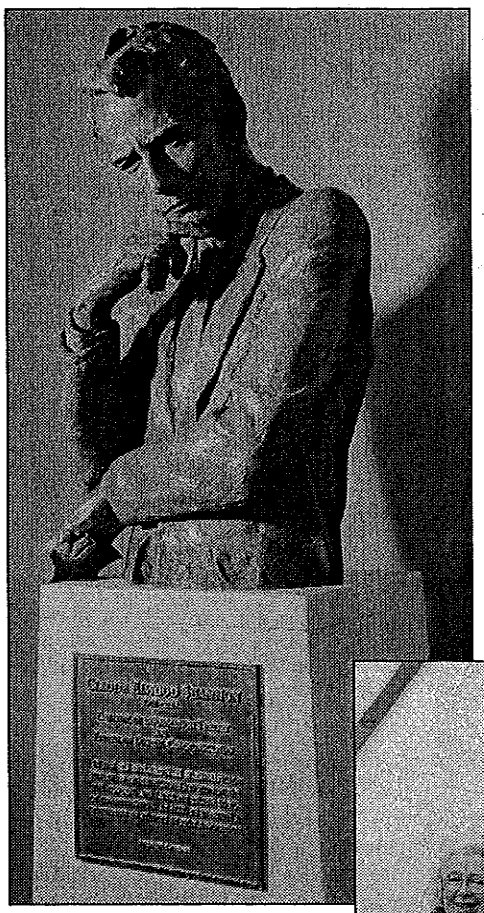
Shannon's treatment of theoretical secrecy is conceptually rich. He gave the first precise definition of the "unbreakability" of a cipher, restricting himself to a ciphertext-only attack, as meaning that the cryptogram and the message it represents are statistically independent. He showed that the cipher proposed by G. S. Vernam in 1926, now often called the "one-time pad", achieves "perfect secrecy"—which was Shannon's term for such unbreakability. [Vernam had claimed that the "unbreakability" of his cipher was confirmed by field trials with the U.S. Army Signal Corps.] More significantly, Shannon showed that perfect secrecy requires a secret key whose length in binary digits is at least as great as the number of bits of information in the message encrypted. This made clear that practical secrecy is the best that one can hope to achieve in most realistic situations where the secret key is relatively short, and it led Shannon to define the "unicity distance" of a cipher as the amount of plaintext that essentially determines the secret key. His formula for estimating unicity distance is still widely used today.

Perhaps the most important aspect of Shannon's October 1949 paper is its thoroughly scientific nature. The second section of his paper begins: "As a first step in the mathematical analysis of cryptography, it is necessary to idealize the situation suitably, and to define in a mathematically acceptable way what we shall mean by a secrecy system." This was a radical departure from previous papers in cryptography where conjecture and imprecision reigned. It is no exaggeration to say that Shannon's paper marked the transition of cryptography from art to science.

Even where Shannon argues on intuitive grounds, he was apparently right on the mark. His principles of "confusion" and "diffusion" for practical cipher design, for which he provided broad semimathematical definitions, were cited in the design of the enciphering algorithm of the 1977 U.S. Data Encryption Standard and are the principles most widely used today in the design of secret-key ciphers.

It must be said that Shannon appears to have missed the most important development of the past fifty years in cryptography, namely that a secret key shared between the communicating parties is not necessary for secrecy. M. E. Hellman, who together with W. Diffie announced this startling development in the 1976 paper that founded "public-key cryptography", has nonetheless credited these words from Shannon's 1949 paper as the inspiration for this discovery: "The problem of good cipher design is essentially one of finding difficult problems ... We may construct

our cipher in such a way that breaking it is equivalent to ... the solution of some problem known to be laborious." Whether Shannon truly missed something is not yet certain. After twenty-five years of public-key cryptography, there is still no proof that trapdoor one-way functions, which are the fundament of the theory, exist.



Left: Statue of Claude Shannon at the Laboratory for Information and Decision Systems (LIDS) at M.I.T. ***(Photograph by Michael Lewy.)*** **Below: Mrs. Betty Shannon at the dedication (October 2000) of another statue, this one commissioned by the IEEE Information Theory Society and located in Gaylord, Michigan, Shannon's hometown.** ***(Photograph courtesy of ITS website.)***

Andrew J. Viterbi

Attaining Maximum Achievable Channel Transmission Rates: Fulfilling Claude Shannon's Prophesy

In his uniquely remarkable 1948 papers [7], Claude Shannon established all the key parameters and limits for the optimal compression and transmission of digital information. The most unexpected was the hard limit on the maximum rate of reliable transmission over noisy, or error-prone, channels. Its comprehension by the communication engineering community was severely limited, partly because of the unconventional nature of the result, whose proof was based on statistical averaging over ensembles of codes rather than on the construction of specific good codes.

Over the next two decades there emerged two disparate approaches to attaining the promise of Shannon's limit. The first involved the construction of code classes and their respective decoding algorithms based on algebraic theory. Though producing elegant results and some very useful encoder and decoder implementations, this fell far short of fulfilling the original purpose. The second more profitable approach built on Shannon's ensemble concept to establish reasonably tight upper and lower bounds on the achievable error probabilities for important code classes as a function of code length and channel parameters. Out of this came the first tangible progress, yielding results which promised the attainment of reliable transmission at rates above one-half of channel capacity.

The codes were drawn from an ensemble of very long convolutional codes, originally proposed by Elias. But the major contribution was that of an implementable decoding algorithm which performed a sequential search for the most likely path along the tree structure generated by the code. Progressively more refined decoding algorithms were developed by Wozencraft, Reiffen, and Fano. Sequential decoding suffered, however, from a drawback which limited its achievable rate to lie below a so-called computational cutoff rate, always less than capacity and approaching half capacity for very noisy channels. This was due to the fact that the computational load of sequential decoding is a random variable with a Pareto distribution whose exponent is less than negative unity (and hence has bounded mean) only below the cutoff rate.

Partly as an exercise to better understand the potential of convolutional codes, Viterbi in the late 1960s proposed a maximum-likelihood algorithm that recognized the reconvergence of tree paths to reduce the search to the best path traversed within a stationary Markov graph,[1] whose complexity is proportional to its number of states, which in turn is exponential in the length of the convolutional encoder. Good performance was demonstrated even for short codes, guaranteeing a manageable number of states, and this led to their overwhelming acceptance as the codes of choice for most digital wireless transmission, whether over satellites or

Andrew J. Viterbi is professor emeritus at the University of California at San Diego and retired vice chairman of Qualcomm Incorporated. He is currently president of the Viterbi Group, LLC. His e-mail address is andrew.viterbi@viterbigroup.com.

[1] *This model subsequently gave rise to numerous applications of the algorithm unrelated to channel coding, ranging from voice recognition to DNA sequence alignment.*

terrestrial base stations, as well as for many wireline data channels.

To achieve very low error rates and to approach closer to capacity, the effective code length was greatly enlarged in the 1970s by employing concatenation, a technique first proposed by Forney, which involves passing the digital message through two encoders serially, with an interleaver in between, and at the receiver decoding the two codes in the inverse order to that in which they were encoded. This approach, however, still failed to attain capacity, largely because information was lost in the process of passing (hard decision) information between decoders.

In the 1980s, Hagenauer proposed means for preserving information by passing soft decisions (essentially likelihood ratios) between decoders, but it was not until the 1990s that Berrou and Glavieux recognized the need to iteratively refine these soft decisions by repeating the decoding process by each decoder with ever improved channel symbol estimates provided to and from the other decoder. The resulting overall performance of this so-called "turbo decoding" algorithm was so close to Shannon's capacity as to launch an intensified worldwide effort to scale the final peak.

Work of dozens of researchers validated and refined the turbo decoding concept and related it to Bayesian statistical concepts. The final and most promising word, however, came not from the convolutional and turbo decoding results but from a much earlier concept initially developed by Gallager in 1963, known as low-density parity check (LDPC) block codes, whose decoding employed both soft decisions and the iterative process, alternating between rows and columns of the parity check matrix. Modern refinements of Gallager's technique by numerous authors at the turn of the century have led to implementable codes for reliable transmission within epsilon of channel capacity.

So ends successfully the half-century saga, leaving though the question of why it took so long when we were so close almost forty years ago. I propose a dual answer. First, the enabling technology for implementation of the (then) seemingly complex algorithms was nonexistent, Moore's Law not yet even having been pronounced. Secondly, the research style of the time was first to prove theorems and then to attempt applications. Today, with practically unlimited computing power, we can simulate and computationally validate algorithms without first needing rigorous proofs of their performance.

References

[1] THOMAS M. COVER, PÉTER GÁCS, and ROBERT M. GRAY, Kolmogorov's contributions to information theory and algorithmic complexity, *Ann. Probab.* **17** (1989), no. 3, 840–865.

[2] THOMAS M. COVER and ERIK ORDENTLICH, Universal portfolios with side information, *IEEE Trans. Inform. Theory* 42 (1996), no. 2, 348–363.

[3] J. L. DOOB, Review of [7], *Mathematical Reviews* 10:133e (1949).

[4] J. L. KELLY JR., A new interpretation of the information rate, *Bell System Tech. J.* **35** (1956) 917-926.

[5] CLAUDE ELWOOD SHANNON, *Collected papers*, edited by N. J. A. Sloane and Aaron D. Wyner, IEEE Press, New York, 1993.

[6] C. E. SHANNON, Communication theory of secrecy systems, *Bell System Tech. J.* **28** (1949) 656–715.

[7] ____ , A mathematical theory of communication, *Bell System Tech. J.* **27** (1948) 379-423, 623-656.

[8] ____ , A symbolic analysis of relay and switching circuits, *Trans. American Institute of Electrical Engineers* **57** (1938) 713–723.

[9] CLAUDE E. SHANNON and WARREN WEAVER, *The Mathematical Theory of Communication*, The University of Illinois Press, 1949.

[10] BORIS TSYBAKOV, contribution to the website "Remembering Claude Shannon", `http://echo.gmu.edu/shannon/survey/memories.php`.