The Early Days of Information Theory

Invited Paper

J. R. PIERCE

***Abstract*—Shannon's communication (information) theory cast about as much light on the problem of the communication engineer as can be shed. It reflected or encompassed earlier work, but it was so new that it was not really understood by many of those who first talked or wrote about it. Most of the papers published on information theory through 1950 are irrelevant to Shannon's work. Later work has given us useful information and encoding schemes as well as greater rigor, but the wisdom of Shannon's way of looking at things and his original theorems are of primary importance.**

INFORMATION theory came into a confusing and confused world. Communication engineers had a problem, but none had successfully formulated the problem or provided a measure of the commodity with which communication engineers deal. There were on the one hand interesting and useful special cases, and on the other hand inconclusive attempts to provide a general theory that would successfully include and elucidate such special cases. In 1948 Claude Elwood Shannon published in the *Bell System Technical Journal* a two-part paper [1], which cast about as much light on the problem of the communication engineer as can be shed.

My approach will be to say something about the pre-Shannon world, to say something about the circumstances in which he worked, to say something about what he accomplished, and to describe the immediate aftermath through, say, 1950. Finally, I will give my own view of the promises (then) and the accomplishments (now) of information theory.

I. The Pre-Shannon World

In 1951, E. Colin Cherry published "A history of information theory" [2], which is worth consulting in itself and which provides an invaluable bibliography, to which I am indebted, as I am to the bibliography in his book, *On Human Communication* [3]. I recommend Cherry's paper and book, though what I say now is not in entire accord with what he said then.

With 20–20 hindsight, it is easy to pick out the earlier work most contributive to Shannon's synthesis.

Among these is the invention in 1832 of Morse's method of telegraphy with its highly efficient code [4]. Another was the elucidation of the problem of sidebands by Carson in 1922 [5]. In retrospect, Morse's work showed a way of speeding transmission; Carson's an apparently inviolable preservation or widening of the frequency spectrum of a signal in the process of modulation or, as we would now say, encoding.

Manuscript received August 12, 1972.
The author is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, Calif.

In 1924 Nyquist [6] in the United States and Küpfmüller [7] in Germany showed that in order to transmit telegraph signals at a definite rate a certain definite bandwidth was required. This result, like Carson's, is as central and valuable now as it was then—but it was only a part of the story.

In 1928 Hartley [8] tried to formulate a theory of the transmission of information. In retrospect, his idea of defining the quantity of information as the logarithm of the number of symbols seems very modern, but its success in elucidating knotty problems was not great. It is Shannon's feeling, and mine, that Nyquist's work was more fruitful.

One knotty problem was posed by Armstrong's realization in 1936 [9] that a positive advantage in combatting noise and interference could be realized by increasing the bandwidth used for transmission. Armstrong's wide-deviation frequency modulation was and remains an epochal step in processing signals for transmission.

On the other hand, Homer Dudley (as he told me a number of years ago) set out to and did overthrow any idea that communication requires a bandwidth at least as wide as the bandwidth of the message to be communicated. He did this by inventing the vocoder, which he described in 1939 [10]. At the transmitting end, the vocoder analyzes speech and produces a number of slowly varying parameters or control signals. These are transmitted and used to control a distant speaking machine that mimics the utterance of the talker. Together with Morse and the men who devised shorthand and commercial code books [2], Dudley was a pioneer in source encoding. Gabor's "frequency compression" [11] came much later.

In 1945, Potter published "Visible patterns of sound" [12]. This was elaborated in 1947, with George H. Kopp and Harriet Green Kopp, in a wonderful book, *Visible Speech* [13]. These publications present a variety of sound spectrograms in which intensity is represented by degree of darkness; frequency is the ordinate and time is the abscissa. The spectrograms show clearly that one can increase the frequency resolution (by using a narrow filter) at the expense of time resolution, or one can increase the time resolution (by using a broader filter) at the expense of the frequency resolution, but one cannot increase the frequency and time resolution simultaneously.

In 1946 Dennis Gabor, in "Theory of communication" [14], proposed a time-frequency uncertainty of the form $\Delta t \cdot \Delta f \doteq 1$. This uncertainty, which is a direct consequence of Fourier transform relationships, is not unrelated to Heisenberg's uncertainty principle of wave mechanics. His diagrams look much like Potter's sound spectrograms. From his considerations, Gabor arrived at the idea of a *logon* or unit of information. This idea is really related to Nyquist's work on signaling speed. It is nonstatistical in nature.

The idea of a statistical message source is central to Shannon's work. The study of random processes had entered into communication before his communication theory. There was a growing understanding of and ability to deal with problems of random noise. Rice was at work in this field [15]. Wiener had dealt extensively with the extrapolation, interpolation, and smoothing of time series [16]. Although Wiener's book was published in 1949, it had been available earlier in a wartime version known as the Yellow Peril (the cover was yellow). Shannon and Bode took considerable pains to put Wiener's work in a form more directly useful to them (and to many others) [17].

Thornton Fry headed a group of mathematicians at the Bell Laboratories with whom Shannon was associated, both during the summer of 1940 and as a regular employee beginning in 1941. Fry told Brockway McMillan that he had once called Shannon into his office and told him that he (Fry) felt that the work of Nyquist and Hartley was deficient in that it ignored noise. Shannon does not recall such an incident.

This constitutes, I believe, the work that in retrospect seems most germane to information theory. Some of the work is substantial and rather general, like that of Carson and Nyquist, Rice and Wiener; some can best be viewed as special cases of the highest importance, like that of Morse, Armstrong, and Dudley, and some like that of Hartley, Potter, and Gabor is suggestive.

It is hard to picture the world before Shannon as it seemed to those who lived in it. In the face of publications now known and what we now read into them, it is difficult to recover innocence, ignorance, and lack of understanding. It is easy to read into earlier work a generality that came only later. It is also easy to overlook important, if not directly pertinent, features of the environment in which Shannon worked.

The postwar world was filled with novel modulation schemes, many of which had been proposed or used for military communication. Besides frequency modulation, these included pulse-length modulation, pulse-rate or pulse-position modulation, and complicated variants of these. Finally, pulse-code modulation, invented earlier [18], had intriguing features. Notably, the signal-to-noise ratio in decibels of the recovered message was proportional to the transmission bandwidth, rather than to the logarithm of that bandwidth, as in wide-deviation frequency modulation.

Pulse-code modulation (PCM) was in an exciting stage at the Bell Telephone Laboratories where Shannon worked. The AN/TRC-6, a military communication system using PCM, had been demonstrated publicly on October 31, 1945. Details of this work were later published in 1947 [19]. Goodall was working toward the exploitation of PCM [20]. C. B. H. Feldman, together with Meacham and Peterson, were (vainly, it turned out) pushing PCM for the first Bell System long-haul microwave radio-relay system [21]. Some of us felt that PCM was the wave of the future and that the future was already upon us [22]. (It was not.)

There were other heady and inspirational activities. In 1940, George Stibitz had demonstrated the "Complex Computer" at the fall meeting of the American Mathematical Society [23]. The complex computer was a relay device which added, subtracted, multiplied, and divided complex numbers. Its input terminal was a teletypewriter; in the 1940 demonstration the computer in New York was operated remotely from Hanover, N.H., thus anticipating by a quarter century the essentials of today's time-shared computers. A number of other relay computers were built at Bell Laboratories up to 1950 [23].

Shannon's master's thesis had been on the use of Boolean algebra in the analysis of relay (or other) logical circuits. At the Bell Laboratories he found that Stibitz had had similar ideas. Shannon's interest in computers overlapped his interest in the problems of communication and probably provided an added stimulus for his emphasis on the digital aspects of communication.

While pulse-code modulation and early work on computers may seem less directly related to information theory than the work of Nyquist, Hartley, Gabor, or Wiener, these were the activities that Shannon's colleagues were pursuing while he was trying to formulate a general theory of communication.

II. Information Theory

Shannon told me recently that he had been working toward information theory as early as 1940, when he was a National Research Fellow at Princeton. Although we worked in quite different fields, I saw Shannon frequently after the war, during the gestation of information theory at Bell Laboratories. When he visited my office I asked him if he had proved any new theorems; if he had, I asked him to write them down in a notebook (alas, the notebook has been lost, but there were few theorems in it). I cannot really give any details of the progress of communication theory in Shannon's mind. I do remember lots of talk about pulse-code modulation and digital transmission, and criteria—sometimes wrong—for optimizing these. In the end, "The Mathematical Theory of Communication," [1] and the book based on it [25] came as a bomb, and something of a delayed-action bomb.

It is presumptuous, but necessary, to say something about the content of Shannon's paper. This I shall do.

In Shannon's words, "the fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

Shannon's model of the message source is a stochastic process that chooses messages from among possible messages on the basis of known (or in some sense knowable) probabilities. This is a fair but by no means exact fit to the behavior of many message sources. We should observe that the probabilities of a source of English text are not knowable and do not exist in any precise sense. What we can say about Shannon's model is that it is a good phenomenological description of a message source, and that, using it, we can either make life simple by assuming an ergodic source, or tractable by assuming a source that is stationary, or much more complicated by considering various nonstationary sources.

Shannon gives mathematical definitions of the information rate or entropy of a source and the capacity of a communication channel, noiseless or noisy. In principle, and within his assumptions, both are measurable quantities. These quantities are important because they are related by a theorem. If the information rate of a source is greater than the capacity of a communication channel, messages from the source cannot be transmitted over the channel without error; there will necessarily be errors of a degree that can be measured by a quantity called the equivocation. But, if the capacity of the channel is greater than the rate of the source, it is possible to send information through the channel "with as small a frequency of errors or equivocation as desired." The nigh-perfect performance possible when the channel capacity exceeds the information rate of the source can be realized only at the cost of complicated encoding involving a long delay in transmission.

After some examples of sources and a discussion of the discrete noiseless channel, Shannon analyzes the discrete noisy channel. Shannon's proof that messages can be transmitted with vanishingly small error if the source rate is less than the channel capacity is indirect rather than constructive. But in discussing encoding for transmission he gives a variable-length encoding in which short codes are assigned to more probable messages and long codes to less probable messages. He also cites an error-correcting code from work, not then published, of Hamming.

After discussing the discrete case without and with noise, Shannon turns to the more difficult continuous case. Here one clear and sometimes misapplied result caught people's attention: the formula for the channel capacity C, for a signal of average power P in additive white Gaussian noise of power N. For a bandwidth W the channel capacity C is

$$C = W \log_2 (1 + (P/N)) \text{ bits/s.} \qquad (1)$$

The channel capacity is the signaling rate for the best possible encoding, which in this case leads to a Gaussian signal. Shannon distinguished and made some calculations concerning channels with a *peak* power limitation, a distinction many carelessly disregard.

Other matters concerning continuous information go perhaps a little less smoothly. The entropy measure suffers from a problem of entropy encountered in classical statistical mechanics; if we don't do something the entropy of a continuous distribution is infinite. The continuous case is saved by introducing a fidelity criterion.

It is useless, wasteful, and impossible to provide the recipient of a continuous message, such as voice or a scene, with a perfect reproduction of the message. He will be satisfied if the message he receives meets some appropriate fidelity criterion. This fidelity criterion is *not* necessarily mean square error. It can be a kind of error specified *a priori*. In the case of the discrete source the rate or entropy can be determined only by an exhaustive and exhausting examination of the message source statistics. In the case of the continuous source, the proper fidelity criterion can be obtained only by an exhaustive and exhausting examination of the requirements of the message destination.

It is interesting to examine the immediate reaction to Shannon's work [1]. We have in print two reviews. The first is a review of the paper in the *Bell System Technical Journal* [1] by J. L. Doob [26]. After a careful summary of the content of the paper, Doob says,

> The discussion is suggestive throughout, rather than mathematical, and it is not always clear that the author's mathematical intentions are honorable. The point of view is that stressed by Wiener in his NDRC report (soon to be published as a book) "The Interpolation, Extrapolation, and Smoothing of Stationary Time Series," in which communication is considered as a statistical problem, specifically in its mathematical formulation as the study of stationary stochastic processes, and of the results of various operations performed on them.

It is clear that Doob, a mathematician, did not appreciate the engineering importance of Shannon's work. He appears not to have seen, either, that a brand new field of mathematics had been opened up.

The book with Weaver [25] was not widely reviewed, presumably because of the earlier appearance of Shannon's part in the *Bell System Technical Journal*. We do have a review by Wiener [27].

Wiener's head was full of his own work and an independent derivation of (1), the capacity of a channel with a limited average power, and a Gaussian noise. After a few introductory remarks, Wiener took off on the trail of Maxwell's demon and the human brain. Competent people have told me that Wiener, under the misapprehension that he already knew what Shannon had done, never actually found out. This would be consistent with his review of Shannon's work.

We come next to Warren Weaver's part of his and Shannon's book [25]. A version of some of this material had also appeared in the *Scientific American* [28].

Weaver starts out by recapitulating and explaining what Shannon had done. Weaver recognized clearly the importance and nature of Shannon's work. Like many others later, he was not satisfied with the territory that Shannon claimed. Weaver plausibly divided the problem of communication into three levels.

Level A: How accurately can the symbols of communication be transmitted? (The technical problem.)

Level B: How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)

Level C: How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

Weaver recognized that Shannon's work pertained to level A, but he argued persuasively that it overlapped and had significance for levels B and C. Weaver was intrigued by the idea of using the powerful body of theory concerning Markov processes in connection with both languages and semantic studies. And, in pointing to the pertinence of information theory to cryptography, he argued that information theory "contributes to the problem of translation from one language to another—."

Such ideas were not foreign to Shannon. In a paper on programming a computer to play chess, published in 1950 [29], Shannon lists the following as some possibilities.

1) Machines for designing filters, equalizers, etc.

2) Machines for designing relay and switching circuits.

3) Machines to handle the routing of telephone calls based on the individual circumstances rather than by fixed patterns.

4) Machines for performing symbolic (nonnumerical) mathematical operations.

5) Machines capable of translating from one language to another.

6) Machines for making strategic decisions in simplified military operations.

7) Machines capable of orchestrating a melody.

8) Machines capable of logical deduction.

Such ideas are, however, foreign to information theory in the sense in which Shannon formulated it, a theory that provides models of message sources and communication channels and theorems relating entropy to channel capacity. In the same way, except for material concerning the capacity of a continuous channel, Wiener's book, *Cybernetics*, which appeared in 1948 [30], is also irrelevant to information theory in the sense in which Shannon proposed it. So is much else that was inspired by information theory.

One entirely legitimate diversion was to relate Shannon's entropy meaningfully to the entropy of statistical mechanics. Equation (1), for the capacity of a continuous channel, should apply to real physical channels plagued by classical Johnson noise, which is indeed Gaussian. Brillouin [31] and Mandelbrot [32] worked toward this end. How far they and others succeeded I will not judge. I later satisfied myself that (1) and the inoperability of Maxwell's demon are consistent in the classical case [33]. This early work *did* leave unsettled the capacity of a channel as limited by quantum effects, a problem not yet resolved in a general way.

While forays from information theory into statistical mechanics are legitimate, they miss the heart of Shannon's theory: the idea of source rate and deliberate, efficient encoding for transmission with negligible error over a channel of limited capacity. So did work by Jacobson concerning information and the eye and ear [34]. Each paper attempts to estimate something that may or may not be a rate for a part of the biological system. But, in such a system the encoding is given, and details of the message source and the message destination are unknown. I suppose that one might try to construe the calculations as bearing on fidelity criteria for hearing and seeing, but the rates arrived at appear unreasonably high (5×10^4 bits/s for the ear; 4.3×10^6 bits/s for the eye).

The examples that Shannon gave of stochastic text stimulated an interest in statistical models of language. Weaver [25] seems to have been convinced of the validity of such a model. Linguists later rejected the idea, but this lies well beyond the era I shall cover.

Shannon's fascinating examples of stochastic text, together with the compelling phrase *information theory*, created another diversion—information theory and art. Here I shall cite only two early papers which, unlike many later ones, are both informed and sensible [35].

Shannon's 1949 paper, "Communication theory of secrecy systems" [36] was not so much a diversion as an inspired application of the techniques he had developed to the determination of when a cryptogram should in principle be decipherable. The paper also contains a sensible discussion of what is desirable in cryptography—mixing transformations that are easy to do and undo and yet make the encrypted text seem random.

Thus, much of the early reaction to Shannon's work was either uninformed or a diversion from his aim and accomplishment. Some work, however, was germane and sound. Among this was, of course, Shannon's paper in the January 1949 issue of the PROCEEDINGS OF THE IRE [37]. In this paper Shannon 1) gave a proof of the sampling theorem, 2) pointed out that there must be a threshold effect in mapping a signal into a space of higher dimensionality (as in frequency modulation) because such mappings cannot be continuous (a result apparently arrived at by Kotelnikov earlier [38], but then inaccessible to Shannon), and 3) gave a clear and appealing geometrical derivation of (1).

In 1949, William G. Tuller published a paper giving his justification of (1) [39].

In 1950, S. O. Rice computed the probability of error for two encoding schemes [40]. Following Shannon, he chose codewords at random. He considered two cases: time-limited codes and band-limited codes. He showed that in each case channel capacity was approached as the codes became very long, and (a very important result) that the probability of error could be made to decrease exponentially with code length.

Choosing codewords at random is fine for proving theorems but impractical for transmitting messages. We have noted that Shannon quoted an error-correcting code due to Hamming. In 1950 Hamming published his work in a simple and clear paper that describes all perfect binary codes save one [41]. This had been given by Golay in 1949, in a paper that refers to the example of error correction cited by Shannon [42].

Error correction in binary signals has become strongly associated with information theory, yet Hamming makes it clear in his paper that he was led to error-correcting codes because of the errors made in computing and switching machines, and particularly in relay machines, rather than through information theory.

In January of 1951 Shannon published "Prediction and entropy of printed English" [43]. By an ingenious experiment in which a human subject guessed what letter would come next, Shannon reduced his estimate of the redundancy of English text from an earlier value of "around 50 percent" to an entropy of about 1 bit per character (lower bound 0.6, upper bound 1.3). This is of interest in connection with both source encoding and cryptography.

Shannon also sought to arrive at a general definition of information through a lattice theory of information. Information was taken as what is common to all encodings of a stochastic source. Apparently, the structure was not great enough to lead to anything very valuable. He published on this in at least two places [44], [45].

Strangely, we have now covered both a number of peripheral areas roughly to the end of 1950 and all central and substantial contributions to information theory in that period. One important matter had not emerged: a conviction on the part of mathematicians that Shannon had done something important and a motivation to search for proofs more satisfactory to them. Early steps in this direction came with the publication in 1953 of McMillan's paper, "The basic theorems of information theory" [46], and a paper by Khinchin [47].

III. Then and Now

We have seen the problem of communication worried but not resolved prior to 1948. We are now familiar with its resolution, but the first response to that resolution for many people was to ride off in their own directions, old and new, rather than to pay attention to what had happened and to the issues that Shannon's theory raised. This is particularly noticeable in the largely irrelevant material presented at the Burlington House Symposium in 1950 [44].

There were strong engineering reactions to information theory.

One of the earliest reactions was to try to reduce the bandwidth used for transmission by encoding a band-limited signal for transmission over a channel of narrower bandwidth. In the long run, this proved to be the wrong direction to go.

Another reaction was to speculate about and to try to measure something concerning the rate of sources (as, picture-element-to-picture-element probability). I think that no one obtained information of value comparable to Shannon's concerning English text.

Another reaction was to try to evaluate existing transmission systems in terms of information theory. One really cannot do this without knowing how well one can hope to encode as a function of delay. This led to work that others will presumably describe.

Another reaction was to seek better systems of modulation or encoding suited to various sorts of noise or distortion. This finally led to some very complex ideas and some complex systems.

Another reaction was to devise and exploit error-correcting codes. This is still a happy field of endeavor.

When I look back on the early days of information theory, I ask myself, what did we expect and what did we get? We expected knowledge—particular knowledge of how to do this or that simply by turning a mathematical crank. But most often, when the problem was real and serious, the crank refused to budge or the mill ground endlessly. Just imagine really trying to gather useful results by block encoding and taking statistics! When the crank spun most easily, there was no practical effect on the art of communication.

What some of us attained was perhaps wisdom rather than knowledge. Like the laws of thermodynamics, information theory divided a world into two parts—that which was possible and that which was not. Often these were separated by a gap between upper and lower bounds, but the general geography was clear. Ingenious people no longer invented coding or modulation schemes that were analogous to perpetual motion. But, they were offered the novel possibility of efficient error-free transmission over noisy channels.

Like the vocoder, color television thrives on human limitations—on slowness of eye and on our failure to resolve detail in color. And, ingenious frame-repeating experiments in Picturephone® transmission succeed because when a face moves the background does not change.

The way we look at these forms of source encoding comes from information theory; the schemes themselves from human ingenuity.

Occasionally the mathematical crank does turn to good advantage, as in error-correcting codes. These, however, are most useful in rare special cases—transmission back from Mariner spacecraft, or data transmission over noisy voice-grade circuits that may eventually be replaced by low-error-rate digital circuits.

In general, I am content with the wisdom that information theory has given us, but sometimes I wish that the mathematical machine could provide a few more useful details.

One cannot judge a scheme of modulation without knowing how good it is, yet many useful schemes, including frequency-modulation receivers with feedback and some other aspects of frequency modulation have proved to be mathematically rather intractable. Perhaps this lies outside of the range of information theory.

In data transmission, we use redundant encoding in order to shape the signal spectrum. Our knowledge of the relation between spectrum and efficiency, and as to what sort of spectra are attainable, is unsatisfactory.

The use of computers in connection with large data bases has raised the question of data compression. We usually feel that a code book comes free; in the computer memory it could occupy as much space as the message—or more. Attempts have been made to apply information theory to the problem—but, is information theory the right theory—or is there a right theory?

I think that I have never met a physicist who understood information theory. I wish that physicists would stop talking about reformulating information theory and would give us a general expression for the capacity of a channel with

® Registered service mark of the American Telephone and Telegraph Company.

quantum effects taken into account rather than a number of special cases.

Acknowledgment

The author wishes to thank P. Elias, S. W. Golomb, J. N. Franklin, B. McMillan, B. M. Oliver, and D. Slepian for reading and commenting on a draft of this paper; C. E. Shannon for replies to questions only he could answer; and H. C. Martel doubly, for reading the draft and for making necessary corrections and editorial changes while the author was away.

References

[1] C. E. Shannon, "The mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948; also *ibid.*, vol. 27, pp. 623–656, Oct. 1948.
[2] E. C. Cherry, "A history of information theory," in *Proc. Inst. Elec. Eng.* (London), vol. 98, pt. 3, pp. 383–393, Sept. 1951.
[3] ——, *On Human Communication*. Cambridge, Mass.: M.I.T. Press; New York: Wiley; London: Chapman and Hall, 1957.
[4] E. N. Gilbert, "How good is Morse code?" *Inform. Contr.*, vol. 14, pp. 559–565, June 1969.
[5] J. R. Carson, "Notes on the theory of modulation," in *Proc. Inst. Radio Eng.*, vol. 10, pp. 57–69, Feb. 1922.
[6] H. Nyquist, "Certain factors affecting telegraph speed," *Bell Syst. Tech. J.*, vol. 3, pp. 324–352, Apr. 1924.
[7] K. Küpfmüller, "Uber einschwingvorgange in Wellen filtern," *Elek. Nachrichtentech.*, vol. 1, pp. 141–152, Nov. 1924.
[8] R. V. L. Hartley, "Transmission of information," *Bell Syst. Tech. J.*, vol. 7, pp. 535–563, July 1928.
[9] E. H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation," in *Proc. Inst. Radio Eng.*, vol. 24, pp. 689–740, May 1936.
[10] H. Dudley, "The vocoder," *Bell Lab. Rec.*, vol. 18, pp. 122–126, Dec. 1939.
[11] D. Gabor, "New possibilities in speech transmission," *J. Inst. Elec. Eng.* (London), vol. 94, pt. 3, pp. 369–390, Nov. 1947; also *ibid.*, vol. 95, pt. 3, pp. 412–413, Sept. 1948.
[12] R. K. Potter, "Visible patterns of sound," *Science*, vol. 102, pp. 463–470, Nov. 9, 1945.
[13] R. K. Potter, G. A. Kopp, and H. G. Kopp, *Visible Speech*. New York: Van Nostrand, 1947.
[14] D. Gabor, "Theory of communication," *J. Inst. Elec. Eng.* (London), vol. 93, pt. 3, pp. 429–457, Sept. 1946.
[15] S. O. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282–332, July 1944; also *ibid.*, vol. 24, pp. 46–156, Jan. 1945.
[16] N. Weiner, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*. Cambridge, Mass.: M.I.T. Press; New York: Wiley, 1949.
[17] H. W. Bode and C. E. Shannon, "A simplified derivation of linear least square smoothing and prediction theory," *Proc. IRE*, vol. 38, pp. 417–425, Apr. 1950.
[18] A. H. Reeves, French Patent 852 183, Oct. 23, 1939; U.S. Patent 2 272 070, Feb. 3, 1942.
[19] H. S. Black and E. O. Edson, "PCM equipment," *Elec. Eng.*, vol. 66, pp. 1123–1125, Nov. 1947.
[20] W. M. Goodall, "Telephony by pulse code modulation," *Bell Syst. Tech. J.*, vol. 26, pp. 395–409, July 1947.
[21] L. A. Meacham and E. Peterson, "An experimental multichannel pulse code modulation system of toll quality," *Bell Syst. Tech. J.*, vol. 27, pp. 1–43, Jan. 1948.
[22] B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The philosophy of PCM," *Proc. IRE*, vol. 36, pp. 1324–1331, Nov. 1948.
[23] E. G. Andrews, "Telephone switching and the early Bell laboratories computers," *Bell Syst. Tech. J.*, vol. 42, pp. 341–353, Mar. 1963.
[24] C. E. Shannon, "A symbolic analysis of relay and switching currents," *AIEE Trans.*, vol. 57, pp. 713–723, Dec. 1938.
[25] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, Ill.: Illinois Press, 1949.
[26] J. L. Doob, Rev., in C. E. Shannon, "The mathematical theory of communication," *Math. Rev.*, vol. 10, p. 133, Feb. 1949.
[27] N. Wiener, "The mathematical theory of communication," *Phys. Today*, vol. 3, pp. 31–32, Sept. 1950.
[28] W. Weaver, "The mathematics of communication," *Sci. Amer.*, vol. 181, pp. 11–15, July 1949.
[29] C. E. Shannon, "Programming a computer for playing chess," *Phil. Mag.*, vol. 41, ser. 7, no. 314, pp. 256–275, Mar. 1950.
[30] N. Wiener, *Cybernetics; or, Control and Communication in Animal and Machine*. Cambridge, Mass.: M.I.T. Press, 1948.
[31] M. Brillouin, "Life, thermodynamics and cybernetics," *Amer. Sci.*, vol. 38, pp. 594–599, Oct. 1950.
[32] B. Mandelbrot, "Statistical mechanics and the theory of information," *C. R. Acad. Sci.* (Paris), vol. 232, pp. 1638 and 2003, Apr. 30, 1951.
[33] J. R. Pierce, *Signals, Symbols, and Noise*. New York: Harper, 1961.
[34] H. Jacobson, "The information capacity of the human eye," *Science*, vol. 113, pp. 292–293, Mar. 16, 1951.
——, "Information and the human ear," *J. Acoust. Soc. Amer.*, vol. 23, pp. 463–471, July 1951.
[35] J. J. Coupling, "Chance remarks," and "Science for art's sake," *Astounding Science Fiction*, Oct. 1949 and Nov. 1950. Available slightly edited in J. R. Pierce, *Science, Art and Communication*, New York: Potter, 1968.
[36] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.
[37] ——, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.
[38] V. A. Kotelnikov, *The Theory of Optimum Noise Immunity*. New York: McGraw-Hill, 1959.
[39] W. G. Tuller, "Theoretical limitations on the rate of transmission of information," *Proc. IRE*, vol. 37, pp. 468–478, May 1949.
[40] S. O. Rice, "Communication in the presence of noise—Probability of error for two encoding schemes," *Bell Syst. Tech. J.*, vol. 29, pp. 60–93, Jan. 1950.
[41] R. W. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.*, vol. 29, pp. 147–160, Apr. 1950.
[42] M. J. E. Golay, "Notes on digital coding," *Proc. IRE*, vol. 37, p. 657, June 1949.
[43] C. E. Shannon, "Prediction and entropy of printed English," *Bell Syst. Tech. J.*, vol. 30, pp. 50–64, Jan. 1951.
[44] Symp. Information Theory, Minist. Supply, Roy. Soc. Rep., Burlington House, London, England, Sept. 26–29, 1950; also C. E. Shannon, "Lattice theory of information," *ibid.*, pp. 105–107.
[45] C. E. Shannon, "Some topics on information theory," in *Proc. Int. Cong. Math.*, vol. 2, p. 262, 1950.
[46] B. McMillan, "The basic theorems of information theory," *Amer. Math. Statist.*, vol. 24, pp. 196–219, June 1953.
[47] A. I. Khinchin, *Uspekhi Matematicheskikh Nauk*, vol. 7, no. 3, pp. 3–20, 1953; also in A. I. Khinchin, *Mathematical Foundations of Information Theory*. New York: Dover, 1957.