NATIONAL ACADEMY OF SCIENCES
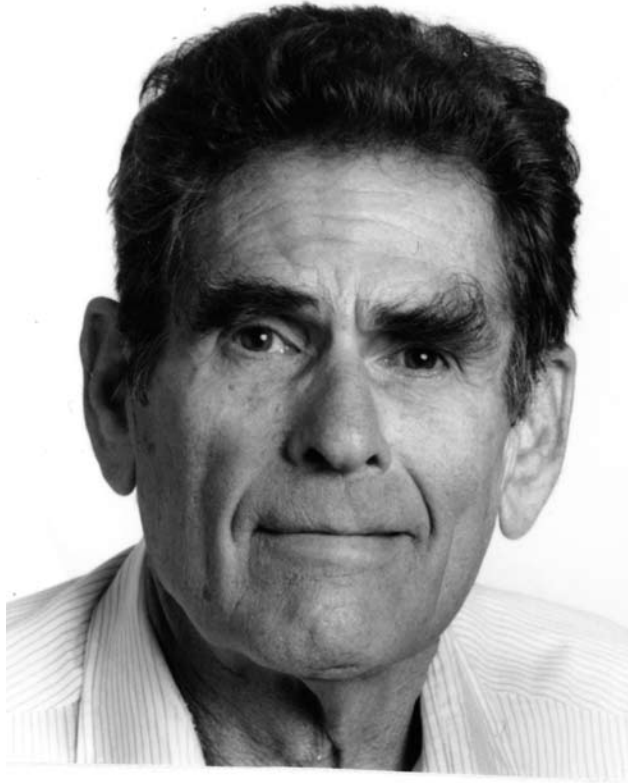
PETER ELIAS
1923—2001

A Biographical Memoir by
ROBERT G. GALLAGER

Biographical Memoir

Peter Elias

PETER ELIAS

November 26, 1923–December 7, 2001

BY ROBERT G. GALLAGER

PROFESSOR PETER ELIAS, PROBABLY THE most important early researcher in information theory after Claude Shannon, died from Creutzfeld-Jacob disease at his Cambridge, Massachusetts, home on December 7, 2001. His three children—Daniel Elias, Paul Elias, and Ellen Elias-Bursac—were with him. His wife, Marjorie (Forbes), predeceased him in 1993 after 43 years of marriage. Pete was distinguished not only for his research but also as the head of the Massachusetts Institute of Technology's Electrical Engineering Department during the crucial period from 1960 to 1966 when the department was changing its education from engineering practice to engineering science and when computer science was starting to be recognized as a major part of electrical engineering.

Among other honors, Pete was a fellow of the IEEE, a charter fellow of the Association for Computing Machinery, and a fellow of the American Academy of Arts and Sciences. He was elected to the National Academy of Sciences in 1975 and the National Academy of Engineering in 1979. He received the Claude E. Shannon Award, the highest honor of the IEEE Information Theory Society, in 1977. Somewhat belatedly he was the recipient of the Hamming Award, a major medal of the IEEE, immediately before his death.

3

EDUCATION (1923-1953)

Pete was born on November 23, 1923, in New Brunswick, New Jersey, where his father was an engineer at the Thomas Edison Laboratory. He completed his secondary education at the Walden School in New York City in 1940 and then enrolled in Swarthmore College. He transferred to MIT for his final two years and received an S.B. in management in 1944. After serving as an instructor for radio technicians in the U.S. Navy for the remainder of World War II, Pete entered Harvard University, where he received a master's degree in computation.

In 1948, immediately after the field of information theory was created by Claude Shannon's masterpiece, *A Mathematical Theory of Communication*, Pete started to search for a Ph.D. topic at Harvard. He soon came upon Shannon's work and was hooked for life. He was fascinated by the intellectual beauty of Shannon's theory, and immediately realized, first, that it provided the right conceptual basis for communication engineering and, second, that a great deal of work remained to be done before the practical benefits of the theory could be realized.

Norbert Wiener's contemporary work on cybernetics provided a complementary set of theories about communication, computation, and control. Pete's Ph.D. thesis, "Predictive Coding," used some of Wiener's results on prediction to attack the information-theoretic problem of determining the number of bits required to represent an analog data source. Pete's underlying idea here was simplicity itself; the predicted value of a symbol, based on previous symbols, is determined by those previous symbols, and thus carries no new information. This means that only the error in prediction (which is known at the encoder via the feedback) needs to be encoded. The full development of this idea, carried out in the thesis, led to a number of insights about data compression

and also to a better appreciation of the relationship between information theory and cybernetics.

After completing his Ph.D. thesis, Pete was appointed as a junior fellow in the Harvard Society of Fellows and spent the next three years doing research on a variety of subjects, including writing several pioneering papers on optical communication, continuing his core research on information theory, and working with Noam Chomsky on syntactic characterization in linguistic theory.

EARLY RESEARCH (1953-1960)

Information theory was a very popular topic among mathematicians, physicists, biologists, and social scientists during these years but most were looking in from the fringes. There was a relatively small community of researchers, concentrated at the Bell Telephone Laboratories and at MIT, who were focused on finding how the new theory could affect telecommunication systems. Robert Fano, the leader of the information theory group at MIT, persuaded Pete to accept an appointment at MIT in 1953 as an assistant professor in the Electrical Engineering Department.

The next seven years were extremely productive for Pete Elias, MIT, and information theory. The elegance and apparent importance of the theory attracted the very best graduate students, and the large number of accessible research problems created a heady and active research atmosphere. As will be seen, Pete's papers in this period opened up a large number of new approaches, forming the starting point for many Ph.D. theses then and later.

The cornerstone of information theory was Shannon's noisy-channel coding theorem, which showed that it is possible to send data over an essentially arbitrary noisy channel at any rate up to the capacity of that channel, and to do so with an arbitrarily small probability of error. Shannon showed

how to calculate that capacity, but his proof about small probability of error was an ingenious existence proof with almost no clues about implementation other than the need for considerable delay and computational complexity.

Although it would take another 40 years to learn how to reach capacity in practice, Pete's 1954 paper, "Error-Free Coding," developed the first algorithm for achieving zero error probability at a strictly positive rate. The algorithm was simple and elegant, but more important it introduced product codes (as they were called later) and iterative decoding to the field. Both of these ideas were sharp departures from the approaches being used in the coding research of that era. Generalizations of product codes and iterative decoding appear in most practical communication systems of today, including turbo codes and low-density parity-check codes.

Pete's next major paper, "Coding for Noisy Channels," in 1955, is perhaps the most influential early paper in information theory after Shannon's original papers. This was another attack on the central problem of finding coding and decoding techniques for reliable data transmission on noisy channels at rates close to channel capacity. It was a more fundamental approach than his "Error-Free Coding," and it provided three giant steps toward the search for effective coding strategies.

"Coding for Noisy Channels" is restricted to a particularly simple model of a noisy communication channel known as a binary symmetric channel, but this model contains all the significant conceptual problems of coding and decoding. Many mathematicians in that era were trying to generalize Shannon's theorems to the most general channels, but Elias saw that the more critical problem was to understand the simplest nontrivial cases first.

The usual approach to error-correcting codes (including those by Shannon and Hamming) was block coding (i.e.,

transmitting binary data in blocks of some given length n). Only some smaller number k of those bits would be independent data bits and the rest would check on those data bits. As a trivial example with $n = 3$ and $k = 1$, a 0 would be converted to 000 and a 1 to 111. If any one of these 3 bits were corrupted by noise, the other two would still permit correct decoding by majority rule. Shannon showed that by making n very large and keeping the rate $R = k/n$ in data bits per transmitted bit constant but less than capacity, the probability of correct decoding could be made to approach 0 with appropriately chosen coding and decoding. More surprisingly, Shannon proved his result by random choice of codewords, thus showing that the choice of code was not critical if the block length could be made arbitrarily large.

Elias realized that the potential for practical error-correcting codes would depend on the complexity of coding and decoding, which increase rapidly with n. Thus it was important to find how quickly error probability could be made to decrease with increasing n and also to discover whether good codes with special structure could simplify the implementation.

To answer the first question, Pete showed that the error probability (using optimal decoding), averaged over all codes of given n and $R < C$, decreases exponentially with n. He also found a lower bound on error probability for the best possible code of given n, R. By comparing the average result with the lower bound, he showed that the best code of given n, R is not substantially better than the average code. Thus error probability decreases rapidly with block length, and the choice of code is not critical. This showed that the essential problem of coding is simple implementation rather than optimum performance. Later researchers used techniques generalized from those of Elias to establish the same result for arbitrary noisy channels.

Now that Pete had established that the essential problem was complexity of implementation, he went on to show that parity-check codes, which had dominated the search for good codes of short block length, were also, on average, as good as the best codes for long block lengths. This meant that researchers could restrict their search to parity-check codes (which had enormous advantages in implementation) and simply avoid those few parity-check codes with very poor performance. This was not as simple as it appeared, since arbitrary parity-check codes were still quite difficult to decode, and the desire for simple implementation typically led to codes that were very much worse than average. This led to the popular saying of the time that "all codes are good except those we can find."

The third major result of this paper was the invention of convolutional codes. These were not block codes but instead resembled the digital equivalent of the linear filters that were so popular in circuit theory. Pete showed that these codes, on average, performed just as well as block codes, but were simpler to work with in a number of ways. The majority of practical coding systems used in practice ever since have used convolutional rather than block codes.

In summary, "Coding for Noisy Channels" provided the trade-off between block length, rate, and error probability on the binary symmetric channel. It showed that the best code is substantially no better than average, and that the average parity-check code and the average convolutional code are equally good. The result was to change the focus in error-correction theory from finding the optimum block codes of given length to finding classes of codes with simple decoding algorithms.

It is somewhat characteristic of Pete that this blockbuster paper appeared only in the convention record of a conference. Pete was never one to expand his publication record

by multiple versions of the same paper. The field was small enough at that time, however, that all the serious researchers were familiar with his work. This paper, as well as a number of his other papers, appeared later in the major anthologies of the most important papers in the field.

"Channel Capacity without Coding," in 1956, was another important paper. Shannon had just proven that feedback does not increase the capacity of a noisy channel, but it was clear that feedback could simplify coding and decoding. For analog sources, information theory showed that there is a minimum possible distortion when the source is encoded and sent over a given channel but that the necessary encoding could be very complex. For the special case of a band-limited Gaussian source sent over a band-limited Gaussian channel of the same bandwidth, however, no encoding at all was necessary to achieve the minimum mean-square distortion. If the bandwidths were different, however, all the complexity of the general case reappeared.

Pete's contribution here was to show that if the channel bandwidth were some integer multiple of the source bandwidth and if feedback were available, then, using the appropriate simple strategy, the need for complex coding disappears. As with a number of Pete's results, it was a very special case that was analyzed, but the idea was applicable in much greater generality. Elias's approach greatly simplifies many later results in the literature about coding with feedback.

"List Decoding for Noisy Channels," in 1957, was another paper that seemed quite specialized and impractical at first but which has had many later consequences. The idea here is that the decoder for a noisy channel, rather than choosing the most likely single transmitted codeword, might choose a small list of likely codewords. This could be envisioned as useful in multistage coding where the list of likely words at

one stage is reduced to a single choice at a later stage. Actually, however, the analysis was primarily intended to provide a better understanding of why randomly chosen codes are so good, and more generally to provide a better understanding of the geometry of binary n-space. The results in this paper provide an important link in finding the best-known bounds on error probability for arbitrary noisy channels, and even in algorithms for formal mathematical proof checking.

Another influential paper of a very different kind was an editorial entitled "Two Famous Papers" in the *IRE Transactions on Information Theory*, which was the premier journal for information theory. In an effort to improve the quality of papers in the journal, Pete described two imaginary extremes of bad technical papers. The first is facetiously entitled "Information Theory, Photosynthesis, and Religion" and the imaginary paper uses each term in the title to further obfuscate the others. The other paper, with a title three lines long, adds incremental detail to a problem previously worked to death. This editorial became well known for its amusement value but also helped give reviewers the needed backbone to reject papers of no merit.

This editorial was viewed by some as an attempt to emphasize practical engineering over more speculative efforts to explore new areas, but this was incorrect. Pete always had very broad interests, ranging across telecommunication, mathematics, the sciences, and liberal arts. He had also become a founding editor somewhat earlier of the journal *Information and Control*, which was intended to publish papers of somewhat broader scope than the *Transactions on Information Theory*.

Pete also always found the time to listen to and help undergraduates, graduates, other faculty members, and researchers in all sorts of areas from speculative to very detailed, and he often encouraged people to look at problems

from new angles. What his editorial was objecting to was not the honest effort of researchers to communicate but rather papers with no content or thought written solely to enhance a publication record.

DEPARTMENT HEAD (1960-1966)

In 1960 Pete was promoted to full professor and at the same time was appointed head of the Department of Electrical Engineering (later to become the Department of Electrical Engineering and Computer Science). He was 37 at the time, a remarkably tender age to be appointed department head of the largest department at MIT and the top-ranked electrical engineering department in the country. Two immediate questions come to mind: why was he appointed and why did he accept?

To understand why he was appointed, recall that 1960 was at the leading edge of the information age. It was clear at that time that communication, computation, and control were going to change the way we live, and the only uncertainty was the rapidity of the change and the exact nature of the change. Pete was a leader in one of these fields and highly knowledgeable about the others. He was also very knowledgeable about the physics-oriented part of electrical engineering, and recognized its necessity for developing the digital devices required for information technology. Perhaps most important, his integrity and goodwill were widely recognized. He could be counted on to recognize the many strengths in the department and to encourage each person to develop those strengths for the good of both the department and the individual.

It was a testament to the department and institute leaders at the time that the coming revolution was well anticipated, and that Pete's wisdom and capabilities to meet the challenge were recognized.

From Pete's standpoint his research career was in high gear, and he was in the right place with the right capabilities to solve even more fundamental problems in information theory, digital communication, and several new areas of computer science. He was being asked to put this research on hold and to lead a department of 72 faculty members, many older and more experienced than he, and to do this with almost no infrastructure of separate areas within the department. The departmental politics, given the size of the department, were relatively benign, but the rapid changes on the horizon could have led to ugly repercussions without remarkable tact and understanding from the department head.

Added to this, Pete was clearly an academic and an intellectual who fully enjoyed the pursuit of research problems. At the same time he had very general interests and thoroughly enjoyed interacting with the rest of the department to better understand what they were doing. He was also a humanist and enjoyed helping others in both technical and nontechnical ways. As department head, he would be in a position to interact with a large and outstanding group of people.

Pete's style of leadership was to listen to everyone carefully and help them develop ways of contributing, given the constraints on the department. He was not a person who enjoyed controlling others, but he did enjoy understanding and helping. I was one of Pete's Ph.D. students, and he was enormously helpful in discussing many issues with me, but he was never directive. I didn't realize until later what a great gift it was to be able to develop my own skills for formulating and doing research.

In the end Pete accepted the appointment, with some qualms but with an eagerness to meet the challenge. His style of leadership turned out to be just right, and the department changed and prospered enormously over the next

six years.

During this period, the department grew by more than 50 percent, and research topics changed even more. Many of the new hires were in the computer area, and many were in the communication network area. Many more were in system and control areas, which were in a state of rapid change and growth in the 1960s. Before 1960 the department viewed its research mission as being divided between processing and transmitting information on one side and processing and transmitting energy on the other. By 1966 the information side had dwarfed the energy side, and the information side had split into a very large number of new areas, with computer science being the most rapidly growing.

The department's mission while Pete was department head was not only education and research but also educating young faculty members in these new fields for careers in other universities and laboratories. The Ford Foundation was financing this effort, and it was highly successful as a means for rapid transfer of new research fields.

LATER CAREER (1966-2001)

When Pete completed his term as department head, he returned to a more academic life of research and teaching. He was by now viewed as a senior statesman and was in considerable demand for government, MIT, and professional committees requiring people of wisdom and tact. One particularly important MIT committee that he chaired was the Ad Hoc Committee on Family and Work. This committee was a response to the intense work pressure generally experienced at MIT, and explored how the institution could help people find a healthy balance between work and family. The report was issued in 1990 and is generally credited with a major increase in administrative sensibilities about these issues.

Pete's research after 1966 shifted somewhat toward the computer field, and he became affiliated with the Laboratory for Computer Science (now called the Computer Science and Artificial Intelligence Laboratory, or CSAIL). Much of his work in this area used information-theoretic arguments to approach questions about storage, organization, and retrieval for large files.

One of the interesting questions about file storage is how to achieve efficient storage in a universal manner. Information theory had solved the problem of efficient storage (encoding) for data with known probabilistic structure, but in practice the probabilistic structure is usually unknown. One would like to achieve the same efficiency without that added knowledge. This was a problem that had been attacked several times earlier, but Pete developed a number of simple theoretical and practical approaches that have found their way into many later universal coding schemes.

Pete became an Emeritus Professor in 1991. It is inappropriate to say he retired, since he still enjoyed coming to his office most days. He was still active advising students, organizing department colloquia, or simply playing an active role in the intellectual life of the community. He was a wonderful conversationalist, so well informed and well balanced that everyone just enjoyed being around him. The many colleagues who knew him miss him greatly.

SELECTED BIBLIOGRAPHY

1952

With D. S. Gray and D. Z. Robinson. Fourier treatment of optical processes. *J. Opt. Soc. Am.* 42:127-134.

1953

Optics and communication theory. *J. Opt. Soc. Am.* 43:229-232.

1954

Error free coding. *Trans. IRE (PGIT)* 4(4):29-37. (Reprinted in *Key Papers in the Development of Coding Theory*, ed. E. Berlekamp, pp. 39-47. New York: IEEE Press, 1974.)

1955

Predictive coding. *Trans. IRE (PGIT)* 1:16-33.

Coding for noisy channels. *IRE Convention Record* 3(4):37-46. (Reprinted in *Key Papers in the Development of Coding Theory*, ed. E. Berlekamp, pp. 48-55. New York: IEEE Press, 1974.)

1956

Coding for two noisy channels. In *Information Theory: Third London Symposium*, ed. C. Cherry, pp. 61-74. London: Butterworth Scientific.

Channel capacity without coding. *MIT/RLE Quarterly Progress Report*, Oct., pp. 90-93. (Reprinted in *Lectures in Communications System Theory*, ed. E. Baghdady, pp. 363-368, New York: McGraw-Hill.)

With A. Feinstein and C. E. Shannon. A note on the maximum flow through a network. *Trans. IRE (PGIT)* 2(4):117-119.

1957

List decoding for noisy channels. In *IRE WESCON Convention Record* 2:94-104.

1958

Two famous papers (editorial). *Trans. IRE (PGIT)* 4(3):99.

1967

Networks of Gaussian channels with applications to feedback systems.
IEEE Trans. Inform. Theory 13(3):493-501.

1970

Bounds on the performance of optimum quantizers. *IEEE Trans.
Inform. Theory* 16(2):172-184.

1972

The efficient construction of an unbiased random sequence. *Ann.
Math. Stat.* 43(3):853-862.

1974

Efficient storage and retrieval by content and address of static files.
J. Assoc. Comput. Mach. 21(2):246-260.
Minimum times and memories needed to compute the values of a
function. *J. Comput. Syst. Sci.* 9(2):196-212.

1975

Universal codeword sets and representations of the integers. *IEEE
Trans. Inform. Theory* 21(2):194-203.