Claude Shannon: His Work and Its Legacy[1]

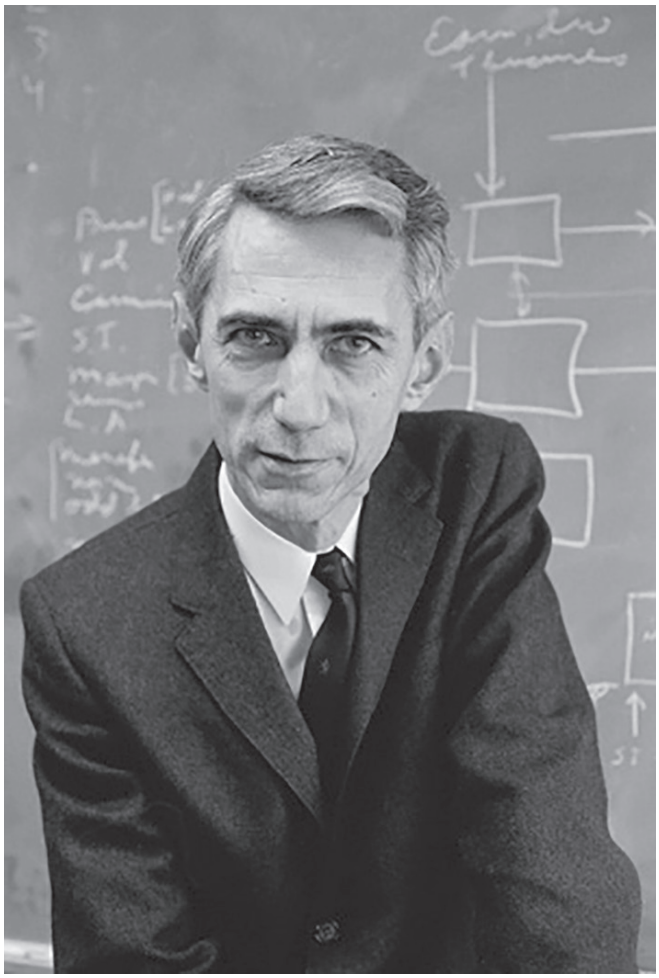
Michelle Effros (California Institute of Technology, USA) and H. Vincent Poor (Princeton University, USA)

The year 2016 marked the centennial of the birth of Claude Elwood Shannon, that singular genius whose fertile mind gave birth to the field of information theory. In addition to providing a source of elegant and intriguing mathematical problems, this field has also had a profound impact on other fields of science and engineering, notably communications and computing, among many others. While the life of this remarkable man has been recounted elsewhere, in this article we seek to provide an overview of his major scientific contributions and their legacy in today's world. This is both an enviable and an unenviable task. It is enviable, of course, because it is a wonderful story; it is unenviable because it would take volumes to give this subject its due. Nevertheless, in the hope of providing the reader with an appreciation of the extent and impact of Shannon's major works, we shall try.

To approach this task, we have divided Shannon's work into 10 topical areas:

- Channel capacity
- Channel coding
- Multiuser channels
- Network coding
- Source coding
- Detection and hypothesis testing
- Learning and big data
- Complexity and combinatorics
- Secrecy
- Applications

We will describe each one briefly, both in terms of Shannon's own contribution and in terms of how the concepts initiated by Shannon have influenced work in the intervening decades. By necessity, we will take a minimalist approach in this discussion. We offer apologies for the many topics and aspects of these problems that we must necessarily omit.



Channel capacity

By Shannon's own characterisation: "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." The channel is the medium – wire, cable, air, water, etc. – through which that communication occurs. Often, the channel transmits information in a way that is noisy or imperfect. The notion that truly reliable communication is possible even in the face of noise and the demonstration that a channel has an inherent maximal rate at which it can reliably deliver information are, arguably, Shannon's most important contributions to the field. These concepts were first expounded in his foundational 1948 paper. He developed these ideas further throughout the 1950s and even into the 1960s, examining the capacity of particular channels, looking at the effects of feedback and other features of existing communication networks and also, because capacity in his vision is an asymptotic quantity, looking at ways in which that asymptote is achieved.

Since Shannon's original work, the notion of capacity has evolved in several directions. For example, the traditional notion of capacity has been generalised to remove many of Shannon's original simplifying assumptions. In addition, the notion of capacity has been expanded to capture other notions of communication. For example, identification capacity was introduced to measure the

[1] This paper was adapted from a talk presented at The Bell Labs Shannon Conference on the Future of the Information Age, Murray Hill, NJ, 28–29 April 2016, celebrating the occasion of the Shannon centennial.

capacity when one is only interested in knowing when a message is present, not necessarily what is in the message, computation capacity was introduced to measure how many computations are possible in certain circumstances, and so on. Capacity has also been applied to many types of channels that have emerged since Shannon's day. Examples include quantum channels, which include both quantum as well as classical notions of transmission and noise, fading channels, which model signal attenuation in wireless transmissions, and, most famously, multiple-antenna channels, which form the basis of modern wireless broadband communications. Even more recently, Shannon's asymptotic concept of capacity, which relies on the ability to use a channel an unlimited number of times, has been examined in a finite-blocklength setting, where only a limited number of channel uses is considered; the finite-blocklength constraint is relevant to modern, delay-constrained applications such as multimedia communications.

Channel capacity has been an enduring concept. Even today, almost seven decades later, we are still using the notion of capacity to think about how communication channels behave. We have every expectation that it will continue to be an important concept well into the future.

Channel coding

In his 1948 paper, Shannon showed that, for any communication rate less than capacity, one can communicate with arbitrarily small error probabilities. In Shannon's paradigm, reliability is achieved through channel coding: transmitters protect signals against errors by introducing redundancy into each message before transmission, and receivers apply their knowledge of the type of redundancy employed to improve their probability of correctly determining the intended message from the channel output. The idea of adding redundancy to a signal was not new but, prior to Shannon, many communications engineers thought that achieving arbitrarily small error required more and more redundancy, therefore necessarily forcing the rate of transmission to zero. The idea that an arbitrarily small probability of error could be achieved with some constant rate of transmission therefore flew in the face of conventional wisdom at the time of its introduction.

Shannon's notion of channel coding initiated a tremendous amount of research and spawned entire subfields within the field of information theory. In particular, a significant amount of fundamental work went on in the 1950s through to the 1980s, when some of the very basic codes and decoding algorithms that we still use today were developed. Notable examples include algebraic codes, such as the Reed-Solomon family of channel codes that form the basis of codes used in modern storage media, and the Viterbi sequential decoding algorithm, which has found an astonishing array of applications, including its use in essentially every mobile phone in use today. The developments of more recent times have been no less impressive. In the 1990s, turbo codes were discovered, which, together with corresponding iterative decoding ideas, revolutionised the field of data transmission. This was followed quickly by another revolution, namely space-time coding. These ideas have driven a lot of what has happened in practice since that time, including the revival of the near-capacity-achieving low-density parity-check codes and the introduction of multiple-input multiple-output (MIMO) systems. These advances have enabled modern high-capacity data communication systems. And, of course, there have been many other key developments, including fountain and Raptor codes, polar codes, etc. In recent times, we have also seen a resurgence of some of the earlier ideas related to areas such as cloud storage and other distributed storage applications. So, channel coding provides a further example of a very early idea of Shannon's that has played a critical role in driving what is happening in technology today.

Multiuser channels

Shannon introduced the notions of channel coding and capacity in a very simple communication setting in which a single transmitter sends information to a single receiver. The techniques that he used to analyse channels in this setting are applicable well beyond this simple "point-to-point" communication model. Multiuser channel models generalise point-to-point channel models by incorporating multiple transmitters, multiple receivers, multi-directional flow of information or some combination of these features.

The generalisation from point-to-point channels to multiuser channels shows up in Shannon's own work as early as the 1950s. In his 1956 paper, Shannon generalised his network model from the point-to-point scenario to channels incorporating feedback; the goal in that work was to understand when feedback from the receiver to the transmitter increases the rate at which the transmitter can send to the receiver. That work employed two notions of capacity: the capacity achievable with an asymptotic notion of reliability and the capacity achievable with perfect reliability. In the former, information delivery is considered reliable if the probability of error can be made arbitrarily small. In the latter, information delivery is considered reliable only if the probability of error can be made to equal zero for a sufficiently large number of channel uses.

In 1960, Shannon generalised the network further by considering two-way channels. Two-way channels differ from point-to-point channels with feedback in that the point-to-point channel with feedback has only a single message travelling from the transmitter to the receiver while the two-way channel has messages travelling from each node to the other. The 1960 paper also mentions a channel in which a pair of transmitters sends information through a shared medium to a single receiver; that channel would today be called a "multiple access channel". The 1960 paper mentions future work to appear on this topic; while no such paper is found in the literature, it is clear that Shannon was thinking about generalisations beyond two-communicator models.

Starting in the late 1960s, multiuser channels became an important area for information theory research. Research

on feedback investigated the improved trade-offs between rate and error probability achievable through feedback. Research on two-way channels yielded improved upper and lower bounds on achievable rate regions. A wide array of new channel models were developed, including multiple access channels (in which multiple transmitters send information to a single receiver), broadcast channels (in which a single transmitter sends possibly distinct information to multiple receivers), relay channels (in which a single transmitter sends information to a single receiver with the aid of a relay that can both transmit and receive information but has no messages of its own to transmit) and interference channels (in which the transmissions of multiple transmitters interfere at the multiple receivers with which they are trying to communicate).

Generalisations of Shannon's channel model are not limited to increasing the number of transmitters or receivers in the networks. Other generalisations include compound channels, which capture channels with unknown or varying statistics, wiretap channels, which model channels with eavesdroppers, and arbitrarily varying channels, which capture channels under jamming. Joint source-channel coding has also been a major topic in the multiuser communication literature. While the optimality of separation between source and channel coding holds for the point-to-point scenario studied by Shannon, it does not hold in general and a good deal of work has gone into understanding when such separation is optimal and how to achieve optimal performance when it is not.

While interest in multiuser channels waxes and wanes over time due to the difficulty of the problems, the massive size and huge importance of modern communication networks makes multiuser information theory an important area for continuing research.

Network coding

The examples given above of multiuser channels are typically used to model wireless communication environments. But wireless networks are not the only multiuser communication networks. After all, Shannon's work was itself originally inspired by communication networks like the wireline phone and telegraph networks of his day, each of which connected vast numbers of users over massive networks of wires. The modern field of network coding studies such networks of point-to-point channels. Typically, the point-to-point channels in these models are assumed to be noiseless, capacitated links. The field of network coding began with questions about the capacity of network coding networks. In some scenarios, notably the case of multicast network coding, the capacity is known, and efficient algorithms are available for achieving those bounds in practice. However, for most networks, the network coding capacity remains incompletely understood.

Given the difficulty of solving the general network coding problem, a variety of special cases have been considered. One of these is the family of index coding networks. Unlike general network coding networks, index coding networks are networks in which only one node in the network has an opportunity to code. It has been shown that if one could solve all index coding networks then that would provide a means of solving all network coding networks as well. That is, any network coding instance can be represented by an index coding instance whose solution would give you a solution to the original network coding problem.

In addition to work on network coding capacity, there has also been quite a bit of work on network code design, as well as work on the relationship between networks of capacitated links and the corresponding networks of noisy channels that they are intended to model. Results in this domain demonstrate that the capacity of a network of noisy channels is exactly equal to the capacity of the network coding network achieved by replacing each channel by a noiseless, capacitated link of the same capacity. Thus, Shannon's channel capacity fully characterises the behaviour of noisy, memoryless channels at least insofar as they affect the capacity of the networks in which they are employed.

Other questions considered in the domain of network coding include network error correction, secure network coding, network coding in the presence of eavesdroppers, network coding techniques for distributed storage and network coding for wireless applications with unreliable packet reception.

Source coding

Source coding, also called data compression, refers to the efficient representation of information. Shannon's work introduces two types of source coding to the literature: lossless source coding, in which the data can be reconstructed from its description either perfectly or with a probability of error approaching zero, and lossy source coding, in which greater efficiency in data representation is obtained through the allowance of some level of inaccuracy or "distortion" in the resulting reproduction. While Shannon's 1948 paper famously discusses both source coding and channel coding and is often described as the origination point for both ideas, Shannon first posed the lossy source coding problem in an earlier communication.

Shannon's 1948 paper sets a lot of highly influential precedents for the field of lossless source coding. It introduces the now-classical approach to deriving upper bounds on the rates needed to reliably describe a source and gives a strong converse to prove that no better rates can be achieved. It also includes both the ideas of fixed- and variable-length codes, that is, codes that give the same description length to all symbols, and codes that give different description lengths to different symbols. Arithmetic codes, which remain ubiquitous to this day, have their roots in this paper. The notions of entropy, entropy rate, typical sequences and many others also come from the 1948 paper.

The 1948 paper also looks at lossy source coding, describing the optimal trade-off between rate and distortion in lossy source description and intuitively explaining its derivation. In a 1959 paper, Shannon revisits the trade-off between rate and distortion in lossy source coding, giving more details of the proof, coining the term "rate distortion function" to describe that bound and

presenting more examples of solutions of the rate distortion function for different sources.

Since then, there has been a lot of work in both lossless and lossy source coding. Much of the work in the 1950s through the 1970s looked at detailed proofs and extensions of the original ideas. Extensions include model generalisations to allow sources with memory, nonergodic sources, and so on. Advances were also made in practical code design for both lossless and lossy source coding. Huffman developed his famous source coding algorithm, which is still in use. Tunstall codes looked at coding from variable-length blocks of source symbols to fixed-length descriptions. Arithmetic codes were further developed for speed and performance. Algorithms for designing fixed and variable-rate vector quantizers were also introduced.

In the years that followed, a lot of work was done on universal source coding and multi-terminal source coding. Universal source codes are data compression algorithms that achieve the asymptotic limits promised by Shannon without requiring a priori knowledge of the distribution from which the source samples will be drawn. Results on universal source coding include code designs for both lossless and lossy universal source coding and analyses of code performance measures such as the rate at which a code's achievable rate (and, in the case of lossy coding, distortion) approaches the optimal bound. Like multiuser channel codes, multi-terminal source codes are data compression algorithms for networks with multiple transmitters of information, multiple receivers of information or both. Examples include the work of Ahlswede and Körner on source coding with coded side information and the work of Slepian and Wolf on distributed source coding networks, where source coded descriptions are sent by independent encoders to a single decoder.

In addition to advances in the theory of optimal source codes and their performance, there has been much research and development aimed at building and standardising lossless and lossy source codes for a variety of communication applications. These algorithms are critical parts of many of the data-rich applications that are becoming increasingly ubiquitous in our world.

Detection and hypothesis testing

Another field in which Shannon's influence has been felt has been that of signal detection and hypothesis testing. Although one might not normally think of Shannon in this context, he worked directly on signal detection in some of his very early work in 1944, in which he explored the problem of the best detection of pulses, deriving the optimal maximum a posteriori probability (MAP) procedure for signal detection; his work was one of the earliest expositions of the so-called "matched filter" principle. Also, by revealing the advantages of digital transmission in communications, he showed the importance of these fields to communication theory in general. And further, he expounded the idea that there is an optimal sampling rate for digitising signals through the famous Nyquist-Shannon sampling theorem.

These ideas have motivated quite a bit of work in subsequent years and to the present day. For example, channel decoding is, in essence, hypothesis testing with large numbers of hypotheses, and some famous results from this area have been developed within the context of sequence detection, including the Viterbi algorithm, noted above, and Forney's maximum likelihood sequence detector. Related to these developments is multiuser detection, which is also motivated by data detection in multiple-access communications, and the closely related problem of data detection in MIMO systems. Distributed detection, which is a problem motivated by wireless sensor networking, is also a successor to these ideas. And, returning to the sampling theorem, one of the major trends today in signal processing is compressed sensing, which exploits signal sparsity to go well beyond Nyquist-Shannon sampling to capture the essence of a signal with far fewer samples. So, again, although we might not think of Shannon as being a progenitor of this field, these connections show that his work has had a major influence either directly or as a motivator.

Machine learning

Another topic that is very much in evidence today is that of machine learning and its role in big data applications. Shannon was an early actor in the application of machine learning ideas – in 1950, he wrote one of the earliest chess-playing computer programs and, in 1952, he developed "Theseus", the famous maze-solving mouse. Of course, we have come a very long way in machine learning since those early contributions, driven by ever more powerful computers. For example, many games have been conquered: checkers in the 1950s, chess with Deep Blue in the 1990s, Jeopardy with Watson in 2011 and go with AlphaGo in 2016. And, of course, there have been many fundamental developments in learning and related tasks, such as neural networks and decision trees, and also graphical models, which have played a major role in channel decoding. These developments are behind contemporary developments such as deep learning and self-driving cars. So, again, Shannon was an early pioneer of a field that has turned out to be a very important part of modern technology and science.

Complexity and combinatorics

Quite a bit of Shannon's work related to and influenced the fields of complexity and combinatorics. Shannon's Master's thesis, perhaps his most famous work next to the 1948 paper, drew a relationship between switching circuits and Boolean algebra. His 1948 paper also introduced many tools that continue to be useful to combinatorial applications. Shannon's 1956 paper on zero-error capacity revisits the capacity problem – shifting the approach from a probabilistic perspective with asymptotic guarantees of reliability to a combinatoric perspective in which reliability requires the guaranteed accurate reproduction of every possible message that can be sent by the transmitter.

Over time, information theory has been and continues to be used for a variety of applications in the complexity and combinatorics literature. Results include generalisations of tools originally developed by Shan-

non in the probabilistic framework to their combinatoric alternatives. An example is Shannon's typical set, which captures a small set of sequences of approximately equal individual probability that together capture a fraction of the total probability approaching 1. This set generalises to more combinatoric alternatives such as that used in the method of types. Fano's inequality, entropy space characterisations, and a variety of other tools from information theory likewise play a role in the combinatorics and complexity literature.

The field of communication complexity also draws upon information theory tools. Concentration inequalities are another example of areas that sit at that boundary between combinatorics and information theory, bringing in tools from both of these communities to solve important problems. One can also find many examples in the literature involving bounding various counting arguments using information theory tools.

Cyber security

Cyber security is another extremely important aspect of modern technology that has its roots, at least in terms of its fundamentals, in Shannon's work. In particular, he established an information theoretic basis for this field in his 1949 paper (in turn based on earlier classified work), in which he addressed the question of when a cipher system is perfectly secure in an information theoretic sense. In this context, he showed the very fundamental result that cipher systems can only be secure if the key – that is, the secret key that is shared by sender and receiver and used to create an enciphered message – has at least the same entropy as the source message to be transmitted. Or, in other words, he showed that only one-time pads are perfectly secure in an information theoretic context.

Shannon's work was allegedly motivated by the SIGSALY system, which was used between Churchill and Roosevelt to communicate by radio telephone in World War II and which made use of one-time pads provided through physical transport of recordings of keys from Virginia to London. Most cyber security systems today, of course, do not use one-time pads. In fact, almost none do. Rather, they use smaller bits of randomness, expand that into a key and use computational difficulty to provide security. Nevertheless, the fundamental thinking comes from Shannon. Public key cryptosystems, of course, were not invented by Shannon but they are basically part of the legacy of looking at cyber security, or secret communications, from a fundamental point of view.

Another major advance in information theoretic characterisations of security was Wyner's introduction of the wire-tap channel in 1975, which gets away from a shared secret and uses the difference in the physical channels, from the transmitter to a legitimate receiver and to an eavesdropper, to provide data confidentiality. This setting introduces the notion of secrecy capacity, which is defined as the maximum rate at which a message can be transmitted reliably to the legitimate receiver while being kept perfectly secret from the eavesdropper. Wyner's work was extended by Csiszár and Körner to the broadcast channel with confidential messages, which is a model that has driven considerable research since, particularly in the recent development of wireless physical layer security, which makes use of radio physics to provide a degree of security in wireless transmission. The 1990s notion of common randomness as a source of distilling secret keys for use in cipher systems also has its roots in information theory and is another basis for wireless physical layer security.

So, again, we see another very important field of contemporary technology development influenced by Shannon's work.

Applications

While Shannon originally developed information theory as a means of studying the problems of information communication and storage, ideas from his work were very quickly taken up by other fields. In 1956, Shannon wrote about this phenomenon in an article titled "The Bandwagon," where he warned of "an element of danger" in the widespread adoption of information theory tools and terminology. In that article, he noted his personal belief that "information theory will prove useful in these other fields" but also argued that "establishing of such applications is not a trivial matter of translating words to a new domain, but rather the slow tedious process of hypothesis and experimental verification".

Today, information theory is used in a wide variety of fields. Biology and finance are two major examples of fields where people are starting to apply information theoretic tools: in one case to study how biological systems transmit and store information and in the other to model long-term behaviour of markets and strategies for maximising performance in such markets. Applications also exist in fields like linguistics, computer science, mathematics, probability, statistical inference and so on.

Concluding remarks

While one can barely skim the surface of Shannon's work and legacy in an article such as this, it should be clear that his genius has benefitted modern science and engineering, and thereby society, in countless ways. We hope that this very brief overview will inspire continuing interest in Shannon and his work and continuing interaction across the boundaries of the many distinct fields that share tools, philosophies, and interests with the field of information theory.

Selected bibliography

Shannon, C.E. (1938). A symbolic analysis of relay and switching circuits. *Transactions of the AIEE* 57(12): 713–723.

Shannon, C.E. (1944). The best detection of pulses. Bell Laboratories memorandum dated 22 June 1944. (In *Claude Elwood Shannon: Collected Papers*, N.J.A. Sloane and A.D. Wyner, Eds. Piscataway, NJ: IEEE Press, 1993, pp. 148–150.)

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27(3): 379–423.

Shannon, C.E. (1949). Communication theory of secrecy systems. *Bell System Technical Journal* 28(4): 656–715.

Shannon, C.E. (1949). Communications in the presence of noise. *Proceedings of the IRE* 37(1): 10–21.

"Claude Shannon Demonstrates Machine Learning", AT&T Archives. (Online at http://techchannel.att.com/play-video.cfm/2010/3/16/In-Their-Own-Words-Claude-Shannon-Demonstrates-Machine-Learning.)
Shannon, C.E. (1956). The bandwagon. *IRE Transactions on Information Theory* 2(1): 3.
Shannon, C.E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE Conv. Rec.* 7:142–163.
Berlekamp, E. (1974). *Key Papers in the Development of Coding Theory*. IEEE Press, New York.
Slepian, D. (1974). *Key Papers in the Development of Information Theory*. IEEE Press, New York.
Sloane, N.J.A., Wyner, A.D., Eds. (1993). *Claude Elwood Shannon: Collected Papers*. IEEE Press, Piscataway, NJ.
Verdú, S., McLaughlin, S., Eds. (2000). *Information Theory: 50 Years of Discovery*. IEEE Press, Piscataway, NJ.
MacKay, D.J.C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge, UK.
Cover, T.M., Thomas, J.A. (2006). *Elements of Information Theory – Second Edition*. John Wiley & Sons, New York.
Richardson, T., Urbanke, R. (2008). *Modern Coding Theory*. Cambridge University Press, Cambridge, UK.
El Gamal, A., Kim, Y.-H. (2011). *Network Information Theory.* Cambridge University Press, Cambridge, UK.
Elder, Y.C. (2015). *Sampling Theory: Beyond Bandlimited Systems*. Cambridge University Press, Cambridge, UK.
Poor, H.V., Schaefer, R. (2017). Wireless physical layer security. *Proceedings of the National Academy of Sciences of the USA* 114(1): 19–26.

Michelle Effros [effros@caltech.edu] is the George van Osdol Professor of Electrical Engineering at the California Institute of Technology. Her research focuses primarily on information theory, with a particular interest in information theoretic tools for the analysis and design of very large networks. She currently serves as Senior Past President of the IEEE Information Theory Society.

H. Vincent Poor [poor@princeton.edu] is the Michael Henry Strater University Professor of Electrical Engineering at Princeton University. His research interests include information theory and signal processing and their applications in various fields. He has served as President of the IEEE Information Theory Society and as Editor-in-Chief of the IEEE Transactions on Information Theory.

Visions for Mathematical Learning: The Inspirational Legacy of Seymour Papert (1928–2016)

Celia Hoyles and Richard Noss (UCL Knowledge Lab, University College, London, UK)

Seymour Papert, who died on 31 July, was a mathematician with two PhDs in pure mathematics, from the University of Witwatersrand, South Africa, and the University of Cambridge, UK. He was a founder of artificial intelligence with Marvin Minsky at MIT, a psychologist working alongside Jean Piaget, a political activist against apartheid and, on a personal level, a wonderful cook and loyal friend. Since his death, the web has been awash with reminiscences and detailed accounts of his intellectual contributions, not only to the fundamental subjects in which he was the undisputed leader but also to the field of education, to a scholar who believed and showed that the computer, or at least the very carefully crafted use of the computer, could introduce young and old alike to the joys and power of mathematics and mathematical thinking.

In this short article, we have selected four pieces of work that directly impacted on the mathematics education field and community. Significantly, these are among his less well-known lectures and papers and we hope that, by airing them, the realisation of Papert's vision of a new kind of learnable mathematics may be one step closer.

1980: Keynote in ICME Berkeley, USA

Seymour gave one of the four plenaries at ICME 1980. Sadly, as far as we can tell, there was no transcript produced of Seymour's remarks. We are, however, grateful to Jeremy Kilpatrick (who attended the talk) for pointing us to a 1980 book edited by Lynn Steen and Don Albers, which includes a 4-page synopsis of Seymour's talk.[1]

Apparently, Seymour was inspirational. From the abstract, we know that he began:

> *"We are at the beginning of what is the decade of mathematics education. Not just in how children learn, but what they learn: we will see dramatic changes in what children learn; we will see subject matters that formerly seemed inaccessible or difficult even at college level learned by young children; we will see changes in where learning takes place, and in the process of learning itself."*

[1] https://books.google.cz/books?id=zcq9BwAAQBAJ&pg=PA12&lpg=PA12&dq=%22.