THE GOLDEN YEARS OF INFORMATION THEORY

1955-1975

DEDICATED TO TOM KAILATH ON HIS 70TH BIRTHDAY.

Why talk about such a prehistoric era?

- Tom and I were both contributing to Information theory in that period.

- The research environment was very different then. Why?

- The effects of much of that research have only been apparent lately.

- Old people like to talk about the good old days.

Why focus on Information theory?

Information theory is a success story in that it has supplied both the architecture and the analytical tools that have governed modern digital communication systems.

The cleanness of Claude Shannon's thinking makes information theory an ideal model of how theories should contribute to engineering.

Shannon's genius lay in finding the "right way," the "simple way" to look at everyday technological problems.

Examples: communication systems, crypto systems, chess playing machines, solving mazes, controlling unicycles, gambling strategies, etc.

He built mathematical models to help understand these problems, but his focus was on the underlying problem, not in mathematics per se nor in problem details.

Fresh out of U. Penn in 1953, I joined the switching department at Bell Labs.

They had a job training program about information theory, switching, mathematics, physics, etc.

There were intellectual puzzles, basic concepts, and not much concern for “practical engineering.”

It was the beginning of a life-long question for me: is it enough to just have fun doing research, or should we work seriously on real problems?

Shannon (and most of the best researchers at Bell Labs) were driven by intellectual curiosity rather than applications.

The curiosity generally addressed basic engineering issues rather than purely mathematical abstraction.

-but the interest was in general principles rather than immediate design.

The interest was in developing an intellectual framework - an architecture - within which to view applications.

Shannon's puzzle-solving research style was in full swing when I was an MIT graduate student (1956-60).

Intellectualism was in the air. Everyone wanted to understand mathematics and physics as well as communication.

Starting companies, making millions, developing real applications was secondary.

There was interest in bringing the theory closer to reality, but it was theory-based.

Our role models were relaxed, curious, and had time to reflect.

There was no shortage of simple research problems that had never been explored.

The underlying mathematical disciplines, however, were stochastic processes, discrete mathematics, and algorithms, all quite new to engineering.

The combination of new mathematics, simply stated new problems, and intellectual culture was irresistible to the best students.

Tom Kailath, Jacob Ziv, Dave Forney, Jim Massey, Elwyn Berlekamp, Irwin Jacobs, Len Kleinrock, and many others were there at the time.

IT was built on a probabilistic model of sources and of noisy channels.

Shannon used the law of large numbers in a highly creative way to determine the number of typical source sequences.

The same ideas determined the number of typical noise sequences on channels.

The theory was like a Beethoven symphony with recurring themes of increasing intensity and depth.

Shannon started with text compression.

Text was modeled as a random, independent identically distributed (iid) letter sequence.

Why random? It makes sense for designing a telecomm system.

Why iid? It explains the basic idea behind compression; it builds the framework for 'better' models.

Let $p(i)$ be the probability of the letter i; the probability of a letter sequence $\mathrm{x} = x_1, \ldots, x_n$ is then

$$\mathrm{Pr}\{\mathrm{x}\} = p(x_1)p(x_2)\cdots p(x_n)$$

From the law of large numbers (LLN), long typical sequences x have about $np(i)$ appearances of letter i, and thus

$$\begin{aligned} \mathrm{Pr}\{\mathrm{x}\} &\approx \prod_i p(i)^{\,np(i)} \\ &= 2^{n \sum_i p(i) \log_2 p(i)} \\ &= 2^{-nH} \qquad \text{where} \\ H &= \sum_i -p(i) \log_2 p(i) \end{aligned}$$

All typical sequences have about the same probability.

Cumulatively, their probability is $\approx$ 1.

There are about 2^{nH} typical sequences.

Each can be represented by nH bits.

Hidden assumption: typicality is based on LLN. Long delays necessary for LLN behavior.

Shannon's entire theory was based on the LLN regime.

Was this an oversight?

No, it was a stroke of genius.

The theory fit together this way, and all major results depended on it.

Later research has made extensions for finite delay, feedback, and lack of LLN.

This theory was honed and polished for 30 years.

The mathematics became sharper and cleaner.

Elegant and semi-practical source codes and channel codes were developed.

There were applications for space probes and military applications.

Solid state technology was not ready for major commerical applications.

I invented “low density parity check codes” in my PhD thesis.

These codes approach capacity with increasing block lengths.

It generated enough theoretical interest to get me a faculty job at MIT, but it didn’t generate much practical interest.

Forty years later, the scheme is of major practical interest.

Information theory has prospered because of 4 major ingrediants:

1) There is a rich and elegant mathematical structure based on probability.

2) There are many toy problems that are fun and simple, but which can be extended to approach reality.

3) The application field is digital communication, which has rapidly grown in importance.

4) The culture is to attack new problems in a discipline oriented fashion.

The network area offers an interesting contrast.

Although there is no central intellectual framework for data networks, there is no shortage of theory.

Graph theory provides many insights about connectivity, path lengths, etc.

Routing theory is based on optimization.

Queueing networks is a well developed branch of probability.

Distributed algorithms are fascinating logical puzzles.

There are many varieties of network information theories.

Much of network design is ad hoc, and much is done by committee.

It is not clear that the theories above contribute greatly to this design, and it is not clear that the networks have enough structure to be helpful in enhancing the theory.

It is clear that the core problems of networks such as congestion remain unsolved.

Is a more cohesive structure possible?

Probably not without some fairly major changes in direction.

Theories develop slowly over time.

Shannon thought about communication for 8 years before writing his magnum opus.

Succeeding results appeared as evolutions of each other, not in order of interest to industry.

Many think the pace of research is accelerating, but it is probably getting slower.

When technology and architecture are all in place except for one missing link, rapid progress is usually made.

People know what to focus on.

Product cycle ‘research’ works fine.

Often, however, many links are missing.

Technology then stumbles along, year after year, with ad hoc solutions.

Multi-link problems take a very long time, even if they are looked at seriously.

Experience plus intuition helps locally, but not long term.

A complicated problem is really a problem whose structure is not understood.

Providing the appropriate structure makes the problem "simple."

Quote from Steven Weinberg: “In the study of anything outside human affairs, including the study of complexity, it is only simplicity that can be interesting.”

Unfortunately, simplicity is hard to define.

Quote from the sculptor Brancusi: “Simplicity is not an end in art, but one arrives at simplicity in spite of oneself, in approaching the real sense of things.”

Simplicity is the ‘A-HA’ that hits us after long contemplation of something.

Unfortunately, A-HA is not always easily communicated.

To me, A-HA means that I've placed the new thing in my own structure or organized it meaningfully for me.

To a college freshman, Riemann integration is simple and Lebesque integration complex. To a math graduate student, the reverse is true.

The search for simplicity is the search for a structure within which the complex becomes transparent.

The information (?) age(?)

Today it is clear that digital communication, digital networks, and computer systems are part of a major force changing the way we think and live.

This is called the information age (it should be called the data age).

It is astounding that we don't spend more time trying to understand the broad implications of these changes.

Consider the impact on basic research.

There is a basic dichotomy between science as structure and as a collection of facts.

The usual definition of scientific method:

- Observe and collect data
- Formulate hypothesis to explain data.
- Predict from the hypothesis.
- Experiment to test hypothesis.

This is fact oriented, but impossible without underlying structure.

The web is highly fact oriented.

It is equally important to constantly simplify the structure.

Detail must be abstracted away.

Simple but generalizable examples (and counter-examples) are critical.

Human minds do not evolve on technological time scales, and theories that are not accessible to human minds are not much use.

As data expands, the importance of simple structure becomes essential.

Information theory started with a simple structure - sources and channels.

There were very simple toy problems - iid sources, binary channels and Gaussian channels.

The structure has grown, but almost as a tree.

One can still start at the bottom and quickly get anywhere.

Fortunately, old branches die off as new ones start.

Information systems (communication, control, computers, networks, signal processing) are fairly new fields (compared with, e.g., physics).

They don't have the traditions that carry physical sciences through fads.

They also don't have physical reality to keep us honest.

That is, we can build more and more complex systems, and pretend that they simply need a little added debugging - witness microsoft word.

It is paradoxical that all the new tools of the web make enormous amounts of data available to us,

- but in our added focus on all this ‘stuff,’ it becomes harder and harder to think and to find simplicity and structure.

We need to change our research communities to cope with this surfeit of data.

Half serious suggestions

Universities and research organizations should hire new faculty/staff on the basis of their best 1 or 2 papers.

The research component of tenure should also be the best 1 or 2 papers.

Since everyone can put their papers on the web (and reference other such papers), journals should publish only papers of real archival interest.

Conferences should try for more interaction rather than more parallel sessions.

Universities, governments, and companies should encourage more Shannon style research.

This is very different from the scholar style, the programmatic style, the techno-jock style, and the multi-disciplinary style of research.

It is focussed on simplification.

Fortunately, the human spirit delights in simplicity.

We also have role models like Tom Kailath, who have brought simplications to many fields, so there is hope after all.