The early development of information theory, and what it means for today

Robert Gallager

What's so special about Information theory?

- **It's fun**
- **It's useful (or is it?)**
- **It pays room and board (but for how long?)**

Why focus on the theory side?

This is a school about information theory, dummy!

Information theory is a success story — it's supplied both the architecture and the analytical tools that govern modern digital communication systems.

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

‘The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.’

C. E. Shannon, 1948

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

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

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

Shannon was almost the opposite of an applied mathematician.

Applied mathematicians solve mathematical models formulated by others (perhaps with minor changes to suit the tools of their trade).

Shannon was a creator of models — his genius lay in determining the core of the problem, removing details that could be reinserted later.

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

Their curiosity was usually about mathematical abstractions of engineering issues rather than pure mathematical abstraction.

The interest was in general principles - an intellectual framework or architecture - rather than immediate design.

They would start with very simple and playful models before attempting to say anything general.

With a fresh BSEE from 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's puzzle-solving research style was in full swing when I was an MIT grauate student (1956-60).

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

Funding was easily available and the emphasis was on ideas, not writing papers, pleasing sponsors, starting companies, etc.

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 very bright young students.

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

Information theory 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 when viewed as the input to 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 an iid letter sequence $\mathbf{x} = x_1, \ldots, x_n$ is then

$$\begin{aligned}
\Pr\{\mathbf{x}\} &= p(x_1)p(x_2)\cdots p(x_n) \\
\Pr\{Shannon\} &= p(S)p(h)p(a)p(n)p(n)p(o)p(n) \\
&= p^3(n)p(S)p(h)p(a)p(o)
\end{aligned}$$

From the law of large numbers (LLN), long typical sequences $\mathbf{x}$ have about $np(i)$ appearances of letter i for each i, and thus

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

$$\Pr(\mathbf{x}) \approx 2^{-nH} \qquad \textbf{for } \mathbf{x} \textbf{ typical}$$

All typical sequences $\approx$ equiprobable.

Cumulatively, their probability is $\approx$ 1.

There are thus 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, but the IID assumption is unnecessary.

This highly oversimplified idea easily generalized to noise sequences, channel coding, and distortion.

The theory provides insight and structure into the range of telecommunication problems

Later research extended the theory to finite delay, feedback, and lack of LLN.

Example: Universal source codes

How does one design a source code without knowledge of the source statistics?

This is easy in principle if the source is unknown but ergodic.

Use no coding for a long period where sender and receiver measure statistics. Then use code for those statistics.

Doing this elegantly is a fun research problem (which others have already done).

Shannon's 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 occasional applications to space probes and military systems.

Solid state technology was too primitive 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.

My paper on LDPC to the major IT conference at the time was rejected.

Forty years later, the scheme started to have 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 possible that the complexity of modern heterogeneous networks make the problem impossible.

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

Perhaps there is hope for specialized networks (sensor, optical backbone, neural, automotive, etc.)

Perhaps we need another Claude Shannon.

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 the pace of basic system research is probably getting slower.

When a functioning theory and architecture is in place, rapid progress is usually made.

People know what to focus on.

Product cycle ‘research’ works fine.

Often, however, the basic theory is missing.

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

Providing a suitable theory for a complex area is difficult, even if the resources are present.

Evaluating the progress of research is even harder, even if the researchers themselves take the evaluation seriously.

Experience plus intuition helps in a small system, but not in a very large system - witness Microsoft Word.

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

Providing the appropriate structure makes the problem "simple."

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

Einstein: "Everything should be as simple as possible, but no simpler."

Whitehead: "Search for simplicity, but mistrust it."

Unfortunately, simplicity is hard to define.

To me, 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 topic in my own intellectual structure (*i.e.,* I can describe it in a sentence, a paragraph, or a chapter, whichever I please).

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 totally changing life, work, education, and society.

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 hypotheses to explain data.
- Predict from the hypotheses.
- Experiment to test hypotheses.

This is fact oriented, but impossible without underlying structure.

The web is highly fact oriented.

As known data expands, simplified structures, with details hidden until needed, are essential.

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.

Is it possible that the increasing complexity of technology and life is related to the growth of fundamentalism and fanaticism in many religions?

What does all this say about choosing good research problems?

The most useful theoretical research simplifies how we look at particular problem areas.

One approach is to find a new theoretical structure (like information theory), but this is very hard and very high risk.

My favorite approach is to get confused by something that appears to be simple.

I try to clarify the confusion by looking at the simplest examples I can think of.

Usually I find some minor carelessness in my thinking has caused the confusion - no harm is done other than relatively enjoyable time wasting.

Occasionally, resolving the confusion involves some small inconsistency in the overall theoretical structure of the problem area.

Clarifying such inconsistencies can lead to either minor or fairly major improvements in the overall theoretical structure.

Graduate students often mistake simplicity for triviality. They stumble on some simple and elegant result, and immediately try to complicate it as much as possible.

The misconception is that it takes the best students to solve the most complex problems.

Actually, it takes the best students to *find* **the simplest open problems.**

Shannon always had a knack for finding simple and instructive models for real problems.

He would look at these simple models from many angles, building up the intuition to see how different models related, and how they related to the real problem.

Mathematics only works on simple models - it is the intuition from studying many simple models that provides insight about real problems.

Mathematical models are preferable to the forms of argument we use in social settings because they keep us honest, not because they are 'more powerful.'

Modeling to Shannon was not a compromise between reality and tractibility.

There was also the need to provide insight into similar but more complex models - to suggest a general structure.

We must help students to develop the patience to look for structure, to look for the simple approach.

We only analyze simplified models of problems, and thus what is important is not the particular solution, but the insight.

Summary of my view: the goal of basic research is to create simple ways to view a field of interest.

This simplicity allows otherwise complex systems to be architected into understandable components.

Necessity can be the mother of invention, but only with the help of enough previous basic research.

Application needs should guide basic research, but cannot direct it.

The conventional wisdom
a.k.a. Really bad ideas

Basic research should look at applications 20 years in the future.

Massive computation, prototyping, and simulation can be the basis to design new large complex systems.

The goals and milestones of basic research should be carefully spelled out before starting.

More bad ideas

Basic research should be pure (uncorrupted by technology needs).

Basic research should be mathematical and/or theoretical.

Basic research should be experimental.

Basic research should have both theoretical and experimental components.

More bad ideas

Basic researchers should be encouraged to do whatever they wish (perhaps a good idea with a Claude Shannon).

The pay and advancement of researchers should be keyed to patents (or to number of publications, citations, etc.).

The truth: choosing good research problems is the hardest part of research; guiding and developing good researchers is the hardest part of management.

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.

Fortunately, the human spirit delights in simplicity, so there is hope after all.

We need to change our research communities somewhat though.