The impact of

INFORMATION THEORY (1948-1973)

on

INFORMATION TECHNOLOGY*

Robert Gallager

*These slides are an edited version of an MIT Theory of Computation colloquium given on Feb. 28, 2006. The editing was to allow the slides to stand on their own.

Why talk about such a prehistoric era?

- **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 on mathematics per se nor on 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.

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.

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 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 bright young graduate 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 $\mathrm{x} = x_1, \ldots, x_n$ is then

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

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

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

All typical sequences have about the same probability.*

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.

*A few ϵ's and δ's are needed to make this precise.

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

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

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

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

The mathematics was honed and polished for 30 years BUT:

- **Aside from space probes, early digital technology was too expensive for wide-spread applications**

- **Researchers (appropriately) focused on simple models**

- **Development engineers (appropriately) focused on practical detailed constraints**

Finally, the technology was ready and researchers and designers understood each other. Then the progress was very rapid.

Information theory is now a mature science.

It has conferences, workshops, journals, textbooks, and graduate courses in most major universities.

It has a coherent mathematical structure plus a well-developed set of simple examples.

It provides the architecture for digital communication systems, and provides insights at multiple levels of abstraction into the design of these systems.

Applications now stimulate research and vice-versa.

30 years were spent developing I.T. into a coherent theoretical structure before solid state technology permitted major commercial applications.

By that time, many theoreticians were actively looking for commercial applications.

Similarly, physicists and mathematicians were highly successful designing weapons in World War 2 after a lifetime of theoretical work

Hypothesis: Theories are most effective for complex* systems if the theories have time to develop and evolve first.

*Complex throughout here means conceptually complex, not computationally complex.

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

I invented "low density parity check codes" in my PhD thesis, but it was forgotten for 35 years until it became economically feasible a few years ago.

Theories develop slowly over time, particularly when they change the way we think about systems as complex as communication systems.

Inventions based on a known theory can be applied rapidly, but entire theoretical structures evolve slowly.

The field of data networks forms an interesting contrast to Information Theory. There are many underlying mathematical theories.

Graph theory, queueing theory, routing theory, switching theory, and distributed algorithms provide the structure for sub-problems in networks.

There is no cohesive theory that contributes much to overall network architecture.

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

In the rush to design larger and faster networks, there was never time for a cohesive theory to be developed.

COMPLEX INFORMATION SYSTEMS

Communication systems and networks are examples of very large complex information systems.

Communication systems are success stories partly because the theory provides a comprehensive understanding of system issues.

Networks have mixed success; the data rates are a small fraction of what is available. Theory provides only limited understanding.

Are theoretical disciplines really worth the time and effort in complex information technologies?

Successful theories about complex systems contain many simplified models (e.g., binary symmetric or Gaussian channels) that provide insight into possible applications.

They also provide a comprehensive mathematical theory that allows precise questions to be asked and answered.

This allows disagreements to be resolved and allows the theoretical structure (and then commercial structure) to evolve.

The need for mathematics here is much the same as the need for mathematics in any engineering discipline.

Devil's advocate position: What's the big deal about theory? Networks are enormously successful and are rapidly changing the way we live.

So what if better theory could increase network capacity a hundredfold? Moore's law will achieve that before theory can get started!

Incremental design change, Rube Goldberg software, and massive simulation are working fine.

If a team can build a system and it works, who cares whether they understand it?

Aside from being boorish, this often works for several incremental changes if the original design is understandable.

After several changes, the system becomes obscure, and further changes have unexpected consequences.

Witness microsoft word and many complex defense systems.

The problem with these systems is not lack of efficiency (although they are certainly inefficient). It is that they are so complex and unstructured that they cannot be made to work correctly.

Large complex information systems are created by and for humans.

The human mind has not evolved appreciably in the last thousand years.

There are very powerful computer aids for designing and using complex systems BUT:

I believe that if no one has a gut-level understanding of the major components and interactions of a system, then that system is a disaster waiting to happen.

I don't believe that computers can design systems in the absence of human gut-level understanding.

Summary: Large complex systems are increasingly limited by understanding rather than Moore's law.

That is, smaller, faster, and cheaper components will not help a great deal for many current systems.

We need to create simpler ways of thinking about complex systems, and that is what theories are for.

But simplicity is not a simple concept.

SIMPLICITY

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."

Quote from Albert North Whitehead: "Seek simplicity and distrust it."

Quote from Albert Einstein: "Everything should be made as simple as possible, but no simpler."

To me, simplicity is the 'A-HA' that hits me after long contemplation of something, when it becomes consistent within the structure of my knowledge.

What is simple to one person is often not simple to another (since they do not have the same intellectual structure within which to understand).

Between people in any given theoretical field, however, there is enough shared structure that simplicity can also be shared, and thus the shared understanding of the field can grow.

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

For complex systems in information technology, the primary goal of research is creating this simplicity.

CHOOSING RESEARCH PROBLEMS

The most useful theoretical research simplifies or clarifies the existing theoretical structure.

A useful approach is to take a simple intuitive view of what the theory is saying, and if there is no precise statement of this in the theory, find out why.

Similarly, look at a major theorem and use intuition to see what it seems to say about some new application area.

In both cases, start by looking at trivially simple examples.

If you find something paradoxical or confusing, this is very promising.

In resolving the confusion, you learn something, and in some cases it is something new.

If the confusion is interesting, explore multiple aspects of it.

Research is a little like doing homework problems, except that you think about the problem rather than doing it. It is even more like making up a good and instructive homework problem.

Be like Shannon and play with simple problems - if you find it interesting, it probably is interesting.

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

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

> **We know neither what those applications will be nor what the difficulties will be.**

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

> **If we don't understand the structure, it won't work!**

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

> **Basic research is the search for basic understanding and structure. We can't say what this is a priori, and thus clearly can't establish milestones.**
>
> **We can certainly spell out the underlying questions that motivate us, and can and should report on progress as it occurs.**

Should research guide the direction of information technology?

The impact of the information age on our lives is obvious to all of us.

We have virtually instant communication (voice and data) anywhere in the world.

The data/knowledge of our civilization is available on the web and indexed by google.

We can easily calculate or simulate virtually any problem we can formulate.

Productivity advances eliminate the need for many boring, repetitive, demeaning jobs.

Yet in this apparent Utopia, we have huge problems:

Those who have jobs work longer hours and under high stress, but many people cannot find jobs.

The inequities between haves and have-nots is rapidly growing.

Education is turning into narrow job training, which is too delayed to meet market needs anyway.

The web's focus on facts, and its inability to treat concepts, has encouraged education to focus on isolated facts rather than concepts.

It is now possible for production to meet basic human needs worldwide, but we cannot organize ourselves to accomplish this.

The need for better education is universally agreed upon, but we cannot organize ourselves to provide essential student/teacher ratios.

Technology creates video games that appeal to our basest instincts, but lags in supporting education.

Global warming and other pending eco-disasters are well documented, but we can't organize ourselves to act.

Our lives are increasingly complex and frustrating.

I gave up trying to understand my computer years ago, and am now challenged by the TV set, the digital camera, and the microwave.

When people can't understand and control their lives, they become fearful and angry. They look for simplistic (not simple) beliefs that once accepted, are never again questioned.

Perhaps this partly explains the recent growth of fundamental religious sects (in all religions) focused on hatred and intolerance.

These problems are partly caused by incompetent government.

But we in the U.S.A. sort of elected and re-elected that government.

Even the secret spying, imprisonment without trial, and torture have led to much talk but no action.

Should we, as researchers, take a more active role in the uses to which our research is put?

I don't know, but I believe it is an important subject to consider, both individually and collectively.