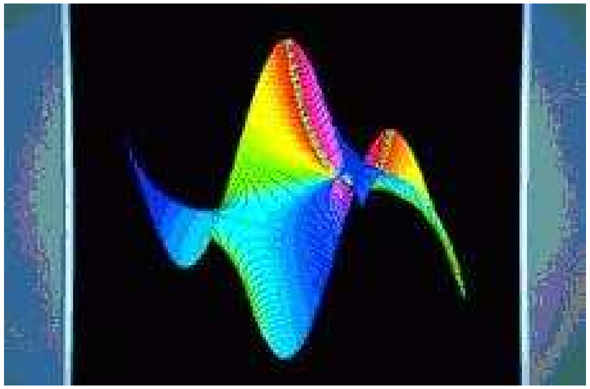
What is Information?*

W. Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907

November 27, 2006

AofA and **IT** logos



*Participants of Information Beyond Shannon, Orlando, 2005, and J. Konorski, Gdansk, Poland.

Outline

1. Standing on the Shoulders of Giants . . .
2. What is Information?
3. Shannon Information
4. Physics of Information
   - Shannon vs Boltzmann
   - Maxwell's Demon, Szilard's Engine, and Landauer's Principle
5. Ubiquitous Information (Biology, Chemistry, Economics, Physics)
6. Computer Science and Information Theory Interface
7. Today's Challenges

Standing on the Shoulders of Giants . . .

C. F. Von Weizsäcker:

"Information is only that which produces information" (relativity).
"Information is only that which is understood" (rationality)
"Information has no absolute meaning."

R. Feynman:

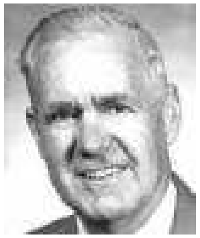
". . . Information is not simply a physical property of a message: it is a property of the message and your knowledge about it."

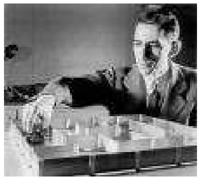
J. Wheeler:

"It from Bit". (Information is physical.)

C. Shannon:

"These semantic aspects of communication are irrelevant . . ."



Structural and Darwin Information

F. Brooks, jr. (JACM, 50, 2003, "Three Great Challenges for . . . CS "):

"Shannon and Weaver performed an inestimable service by giving us a definition of Information and a metric for Information as communicated from place to place. We have **no theory** however that gives us a metric for the Information embodied in **structure** . . . ...this is the most fundamental gap in the theoretical underpinning of Information and computer science. . . . A young information theory scholar willing to spend years on a deeply fundamental problem need look no further."

M. Eigen (Biology):

"The differentiable characteristic of the living systems is Information. Information assures the controlled reproduction of all constituents, thereby ensuring conservation of viability . . . . Information theory, pioneered by **Claude Shannon**, cannot answer this question . . . in principle, the answer was formulated 130 years ago by **Charles Darwin**."

What is then Information?

Information has flavor of:

relativity (depends on the activity undertaken),
rationality (depends on the recipient's knowledge),
timeliness (temporal structure),
space (spatial structure).

Informally Speaking:
A piece of data carries information if it can impact a recipient's ability to achieve the objective of some activity within a given context.

Event-Driven Paradigm

- **System**: A universe is populated by systems performing activities in order to achieve objectives (e.g., living organisms, software agents, communication networks).
- **State**: A system is characterized by its state (e.g., memory content, traffic conditions in networks).
- **Event**: Any change is caused by an event (e.g., clock tick, execution of a specific operation, reception of a message).
- **Context**: Partial order on the set of events: the set $C(E)$ of events preceding event E is called the context of event E.
- **Attributes**: Events may have attributes (e.g., time of occurrence, type of task being executed).
- **Objective**: Objective functional $objective(R, C)$: maps system's rules of conduct R and the present context C into any space with a well-defined order of points (e.g., number of correctly decoded bits).

Examples

Example 1. (Decimal representation) Our objective is to learn digits of π. Each computed digit is then an event and objective(R,C) is a real-valued function monotonically increasing in C (e.g., number of computed digits).

Example 2. (Synergy of data) In a distributed secret sharing scheme, the event corresponding to the reception of a part of a key does not improve the ability to decrypt a given cipher text unless all the other parts of the key are already in C, and objective(C,R) is to decode the message.

Example 3. (Wireless and ad-hoc networks) In a wireless network, the closer the users are in time and space, the more reliable information can be delivered in time allowed. This reflects upon the objective of sending a maximum data flow in **space-time** to the destination.

Example 4. (Quantum information and communication) In some communication environments an attempt to interpret data leads to the distortion of the data itself.

. . . any attempt to measure (that) property
destroys (at least partially) the influence
of earlier knowledge of the system . . . (W. Pauli)

Furthermore, the order in which experiments are performed may change information (e.g., spin experiment, Brukner and Zeilinger, 2001).

Formal Definition

Definition 1. *The amount of information (in a faultless scenario)* info*(E) carried by the event* E *in the context* C *as measured for a system with the rules of conduct* R *is*

$$\mathrm{info}_{R,C}(E) = \mathbf{cost}[\mathrm{objective}_R(C(E)), \mathrm{objective}_R(C(E) + E)]$$

where the **cost** *(weight, distance) is taken according to the ordering of points in the space of objectives.*

Remark: Thus an event only carries nonzero information if it changes objective(R,C) (intuitive flavor of relativity, rationality, and timeliness).

Definition 2. *The capacity of a channel between the event sources and the recipient is (in a faultless scenario)*

$$\mathrm{capacity} = \max_{R,C}\{\mathrm{info}_R(C(E) + E)\}$$

where the maximum is taken subject to some constraints and $C(E)$ *is a prefix of* C *preceding* E.

Outline Update

1. Standing on the Shoulders of Giants . . .
2. What is Information?
3. Shannon Information
4. Physics of Information
5. Ubiquitous Information
6. Today's Challenges

Shannon Information . . .

In 1948 C. Shannon created a powerful and beautiful **theory of information** that served as the backbone to a now classical paradigm of digital communication.

In our setting, Shannon defined:

objective: statistical ignorance of the recipient;
statistical uncertainty of the recipient.

cost: # binary decisions to describe E;
= $-\log P(E)$; $P(E)$ being the probability of E.

Context: the semantics of data is irrelevant . . .

Self-information for E_i: $\text{info}(E_i) = -\log P(E_i)$.
Average information: $H(P) = -\sum_i P(E_i) \log P(E_i)$
Entropy of $X = \{E_1, \ldots\}$: $H(X) = -\sum_i P(E_i) \log P(E_i)$
Mutual Information: $I(X;Y) = H(Y) - H(Y|X)$, (faulty channel).

Shannon's statistical information tells us how much a recipient of data can reduce their statistical uncertainty by observing data.

Shannon's information is not absolute information since $P(E_i)$ (prior knowledge) is a subjective property of the recipient.

Shortest Description, Complexity

Example: X can take eight values with probabilities:

$$\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right).$$

Assign to them the following code:

$$0, 10, 110, 1110, 111100, 111101, 111110, 111111,$$

The entropy X is

$$H(X) = 2 \text{ bits}.$$

The shortest description (on average) is 2 bits.

In general, if X is a (random) sequence with entropy $H(X)$ and average code length $L(X)$, then

$$H(X) \leq L(X) \leq H(X) + 1.$$

Complexity vs Description vs Entropy
The more complex X is, the longer its description is, and the bigger the entropy is.

Three Jewels of Shannon

Theorem 1. (**Shannon 1948; Lossless Data Compression**).

compression bit rate $\geq$ source entropy $H(X)$.

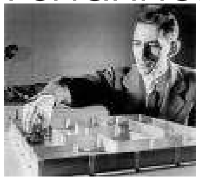
(There exists a codebook of size 2^{nR} of universal codes of length n with

$$R > H(X)$$

and probability of error smaller than any $\varepsilon > 0$.)

Theorem 2. (**Shannon 1948; Channel Coding**)

In Shannon's words:



It is possible to send information at the capacity through the channel with as small a frequency of errors as desired by proper (**long**) encoding. This statement is not true for any rate greater than the capacity.

(The maximum codebook size $N(n, \varepsilon)$ for codelength n and error probability ε is asymptotically equal to: $N(n, \varepsilon) \sim 2^{nC}$.)

Theorem 3. (**Shannon 1948; Lossy Data Compression**).

For distortion level D:

lossy bit rate $\geq$ rate distortion function $R(D)$.

Beyond Shannon

Participants of the **2005 Information Beyond Shannon** workshop realize:

Delay: In computer networks, delay incurred is a nontrivial issue not yet addressed in information theory (e.g., complete information arriving late maybe useless).

This is not a question of understanding the classical delay-rate trade-off, but a complex issue involving our choice of how and what to transmit, as well as the actually utility of the information being transmitted.

Space: In networks the spatially distributed components raise fundamental issues of limitations in information exchange since the available resources must be shared, allocated and re-used.

Information and Control: Again in networks our objective is to reliably send data with high bit rate and small delay (control), that is,
information is exchanged in space and time for decision making, thus timeliness of information delivery along with reliability and complexity constitute the basic objective.

Beyond Shannon

Utility: The utility of what is transmitted depends on different factors (e.g., time, space, content of a message,recipient's activities). How can such utility considerations be incorporated into the classical coding problem?

Semantics: In many scientific contexts experimenters are interested in signals, without knowing precisely what these signals represent (e.g., DNA sequences, spike trains between neurons, are certainly used to convey information, but little more than that can be assumed a priori).
Estimating the entropy is typically not appropriate (indeed it offers a measure of the structural complexity of the signal, but it does not measure its actual information content by ignoring noise present in the signal).

Dynamic information: In a complex network in a space-time-control environment (e.g., human brain information is not simply communicated but also processed) how can the consideration of such dynamical sources be incorporated into the Shannon-theoretic model?

Outline Update

1. Standing on the Shoulders of Giants . . .
2. What is Information?
3. Shannon Information
4. Physics of Information
   - Shannon vs Boltzmann
   - Maxwell's Demon, Szilard's Engine, and Landauer's Principle
5. Ubiquitous Information
6. Today's Challenges

Shannon vs Boltzmann

R. Clausius in 1850, and then Boltzmann in 1877 introduced entropy in statistical mechanics. How are Shannon and Boltzmann's entropies related (cf. Brillouin, Jaynes, Tribus)?

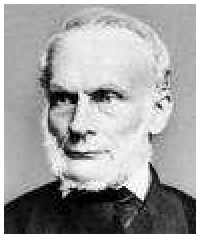
Boltzmann's entropy S is defined as

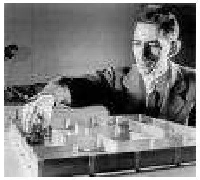
$$S = k \log W$$

where W is the number of molecule macrostates, and k is Boltzmann's constant (to get the correct units).

Boltzmann wanted to find out:

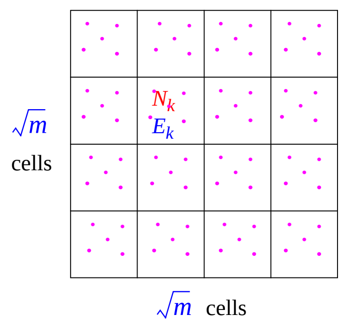
How are molecules distributed?





Boltzmann's Solution

Divide space into m cells each containing N_k molecules with energy E_k.
How many configurations, W, are there?



$$W = \frac{N!}{N_1! N_2! \cdots N_m!}$$

subject to

$$N = \sum_{i=1}^{m} N_i, \qquad E = \sum_{i=1}^{m} N_i E_i$$

We assume now that $E_1 = E_2 = \cdots = E_m = E/N$.

Boltzmann asked: Which distribution is the most likely to occur?

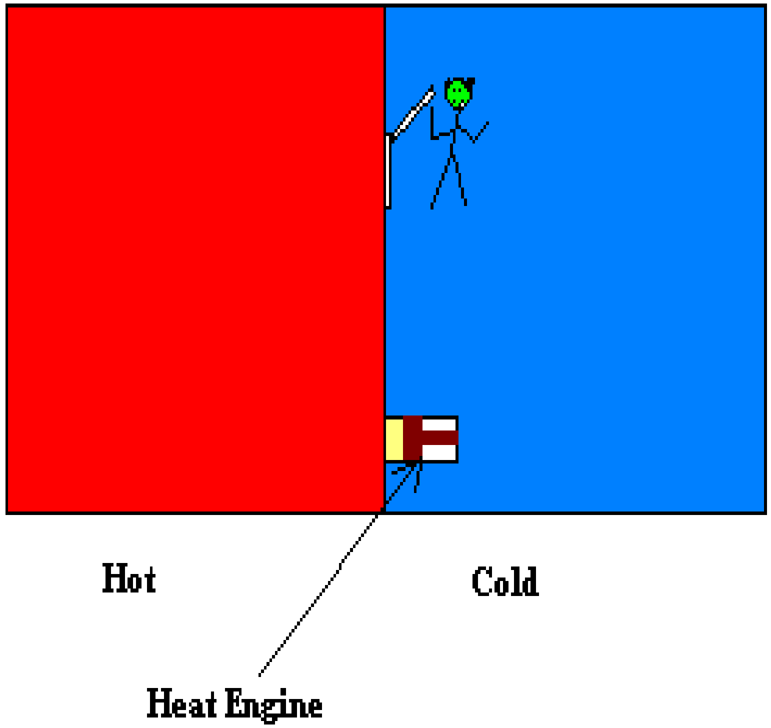
Boltzmann's answer: the most probable distribution is the one that occurs in the greatest number of ways!

Solving the optimization constrained problem, we find (cf. E. Jaynes)

$$\begin{aligned} \log W &= -N \sum_{i=1}^{m} \left(\frac{N_i}{N}\right) \log\left(\frac{N_i}{N}\right) \\ &= H(N_i/N) \qquad \text{(Shannon entropy)} \end{aligned}$$

where all N_i are equal ($N_i = \alpha N$ for some constant α).

Maxwell's Demon and Szilard's Engine



Is the Second Law of Thermodynamics violated by Maxwell's Demon?

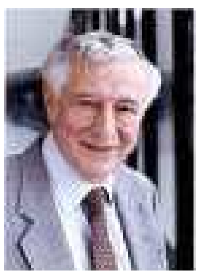
Szilard's Engine: (Acquiring) Information $\Rightarrow$ Energy.

1 bit of information $= k \ln 2$ (joules/kelvin) of energy.

Landauer's Principle: Limits of Computations

Landauer's Principle (1961):

Any logically irreversible manipulation of information (e.g., erasure of a bit), must be accompanied by a corresponding entropy increase (of $kT \log 2$ joules of heat) in an isolated system.
If no information is erased, computation may in principle be achieved that is thermodynamically reversible, and require no release of heat (reversible computing).

$$1 \text{ bit} = k \ln 2 \text{ (joules per Kelvin)}$$

Maxwell's demon explained: C.H. Bennett observed that to determine what side of the gate a molecule must be on, the demon must store information about the state of the molecule. **Eventually(!)** the demon will run out of information storage space and must begin to erase the information and by Landauer's principle this will increase the entropy of a system.

Outline Update

1. Standing on the Shoulders of Giants . . .
2. What is Information?
3. Shannon Information
4. Physics of Information
5. Ubiquitous Information (Biology, Chemistry, Economics, Physics)
6. Today's Challenges

Ubiquitous Information (Biology)

1. Life is an interplay of energy, entropy, and information (i.e., it's consumption, processing, preservation, and duplication of information).

2. Cells process information from in to out in a rather faulty environment.

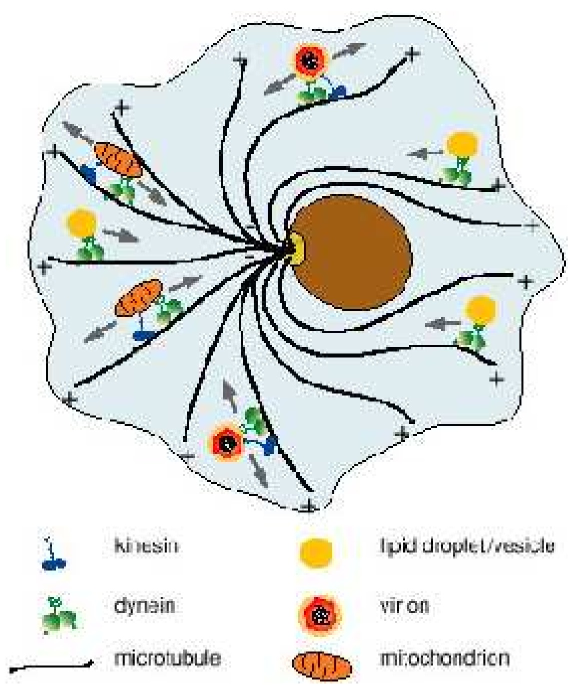
Why does it work so well?

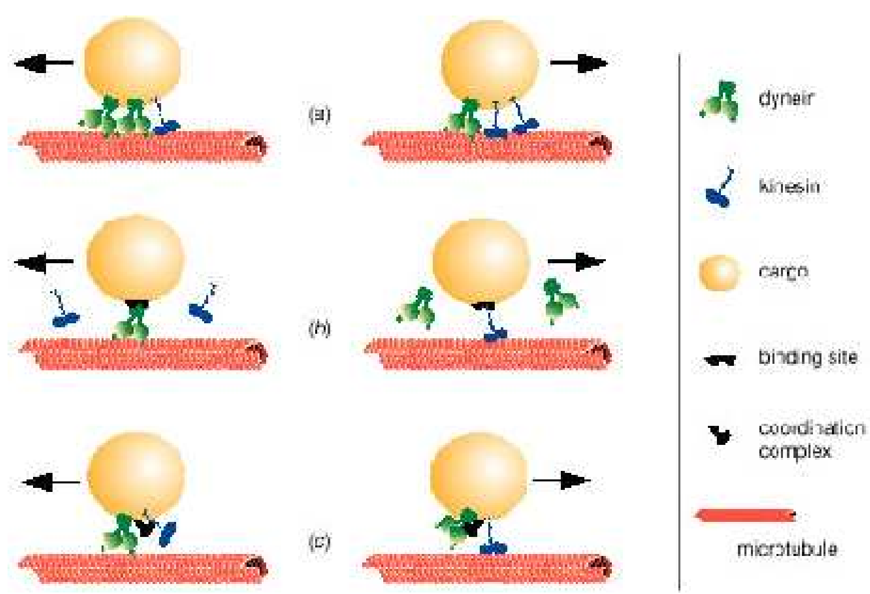
3. Some enzymes correct errors in DNA, but some pass through, and then perhaps Darwin natural selection takes care of them.

Do living organisms possess any error correction capability?

4. How much information does a protein need in order to recognize a binding site amidst all the noise and jostling of a busy cell?

5 Why **bi-directional** microtubule-based transport in cells?

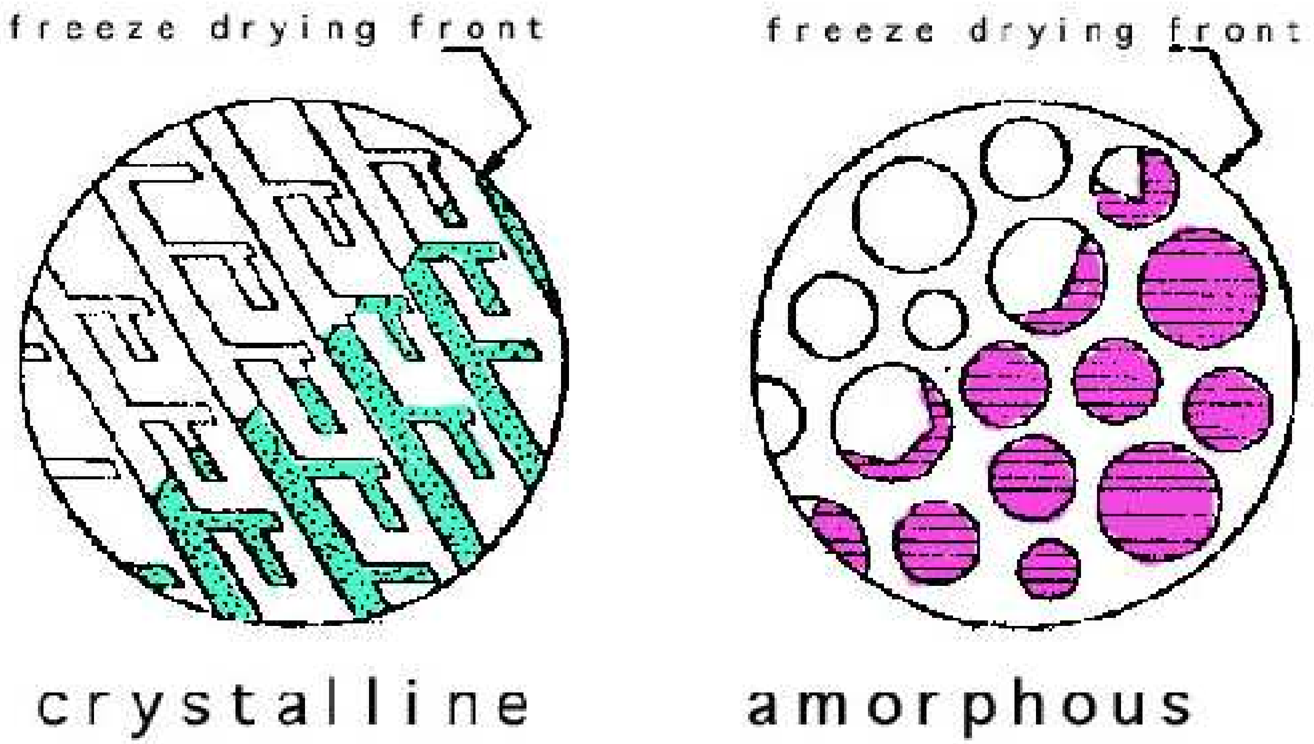




Ubiquitous Information (Chemistry)

In chemistry information may be manifested in **shapes** and **structures**.

Example: Amorphous solid vs crystalline solid



An amorphous solid is a solid in which there is **no long-range order** of the atoms while in crystalline solids there is long-range atomic order. How to assess that amorphous solids are more **complex** than crystalline solids?

Can Körner's graph entropy be used to describe the complexity of shapes and structures (of molecules)? **NO!**

Ubiquitous Information (Economics)

In economics the value of information may be measured by the difference in economic rewards of an **informed action** over an **uninformed** one.

Amount vs Value of Information:

Amount of information: $H(A) = -\sum_i p(a_i) \log p(a_i)$ (bits)

Value of information: $V(A) = \sum_i p(a_i)\nu(a_i, d_i)$ (dollars)

where $d_i \in \mathcal{D}$ is a decision and $\nu(a, d)$ is a payoff function.

Example (Marschak, 1960). Stock X can change in the interval $[-6, 6]$. We assume a **faultless** environment.

Informant A: a_1 =stock rises; a_2 =stock drops.

Decision: d_1 =buy when a_1; d_2 =sell when a_2.

Informant B: b_1 =stock rises ≥ 2; b_2 =stock drops ≤ 2; b_3 =otherwise.

Decision: d_1 =buy when b_1; d_2 =sell when b_2, d_3 =do nothing when b_3.

Amount of Information: $H(A) = \log 2$, $H(B) = \log 3$ (bits)

Value of Information: $V(A) = 3$, $V(B) = 2\frac{2}{3} < V(A)$ (dollars)

More generally: Value of A (information before decision) is:

$$V^*(A) = \max_{d \in \mathcal{D}} \sum_i p(a_i)\nu(a_i, d_i).$$

Value of prefect information: $V_I^*(A) = \sum_i p(a_i) \max_{d \in \mathcal{D}} \nu(a_i, d) - V^*(A)$

i.e., **maximum amount a decision maker would pay**.

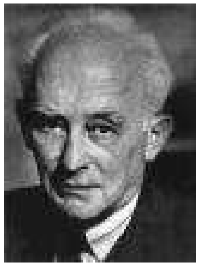
Ubiquitous Information (Quantum)

Information in Quantum Mechanics

Since Bohr and Copenhagen's interpretation, an **atom** is not a thing any more but rather a sum of total probabilities, hence total information.

Von Neumann entropy in quantum statistics describes randomness of a quantum state. It is a natural generalization of Shannon entropy. But

> . . . conceptual difficulties arise when we try to define the information gain in a quantum measurement using the notion of Shannon information. The reason is that, in contrast to classical measurement, quantum measurement, with very few exceptions, cannot be claimed to reveal a property of the individual quantum system existing before the measurement is performed. (cf. C. Brukner, and A. Zeilinger, Conceptual Inadequacy of the Shannon Information in Quantum Measurements, Phys. Rev. A 63, 2001).

Quantum Information

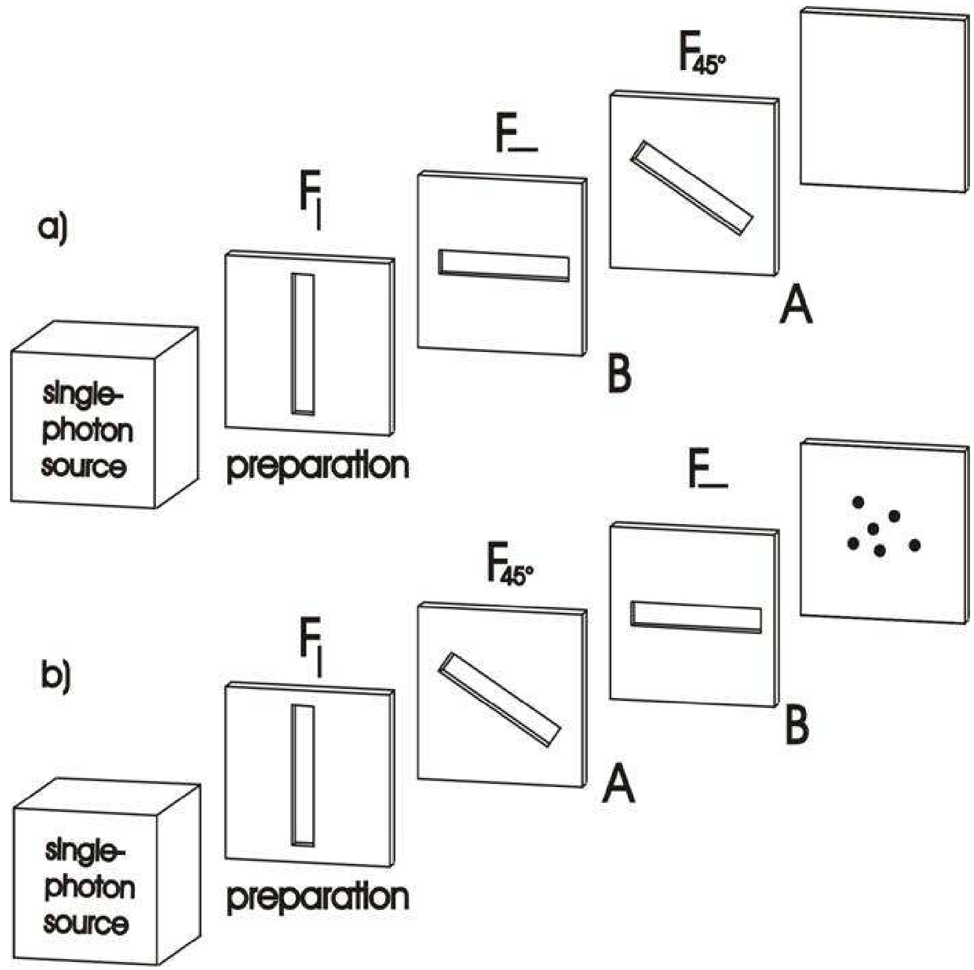


Figure 1: C. Brukner & A. Zeilinger, Phys. Rev. A 63, 022113 (2001).

In quantum mechanics events do not necessarily commute, therefore, joint entropy cannot even be defined since

$$H(A,B) \neq H(B,A)!$$

Law of Information?

The flow of information about an object into its surrounding is called **decoherence** (increases entanglement with its environment) (H. Zeh, W. Zurek).

Decoherence occurs very, very, very fast, in $10^{-10} - 10^{-20}$ seconds.

The essential difference between microscopic world (quantum) and macroscopic world is **decoherence**.

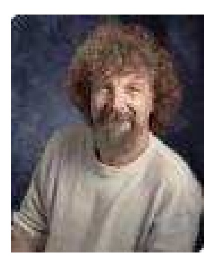
Entropy and decoherence are related, but while entropy operates on a time scale of microseconds, decoherence works a billion times faster.

A new law of Information(?):

Information can be neither created nor destroyed.

or perhaps

stored information of any "isolated system" tends to dissipate.



Information Theory and Computer Science Interface

The interplay between IT and CS dates back to the founding father of information theory, Claude E. Shannon. In 2003 was the first Workshop on Information Theory and Computer Science Interface held in Chicago.

Examples of IT and CS Interplay:

Lempel-Ziv schemes and data compression (Ziv, Lempel, Louchard, Jacquet, Seroussi, Weinberger, Szpankowski)

LDPC coding, Tornado and Raptor codes (Gallager, Luby, Mitzenmacher, Shokrollahi, Urbanke)

List-decoding algorithms for error-correcting codes (Gallager, Sudan, Guruswami, Koetter, Vardy);

Kolmogorov complexity (Kolmogorov, Cover, Li, Vitanyi, Lempel, Ziv);

Analytic information theory (Jacquet, Flajolet, Drmota, Savari, Szpankowski);

Quantum computing and information (Shor, Grover, Schumacher, Bennett, Deutsch, Calderbank);

Network coding and wireless computing (Kumar, Yang, Effros, Bruck, Hajek, Ephremides, Shroff, Verdu).

Today's Challenges

- We still lack measures and meters to define and appraise the amount of structure and organization embodied in artifacts and natural objects.
- Information accumulates at a rate faster than it can be sifted through, so that the bottleneck, traditionally represented by the medium, is drifting towards the receiving end of the channel.
- Timeliness, space and control are important dimensions of Information. Time and space varying situations are rarely studied in Shannon Information Theory.
- In a growing number of situations, the overhead in accessing Information makes information itself practically unattainable or obsolete.
- Microscopic systems do not seem to obey Shannon's postulates of Information. In the quantum world and on the level of living cells, traditional Information often fails to accurately describe reality.
- What is the impact of rational/noncooperative behavior on information? What is the relation between value of information and information?

Vision

Perhaps it is time to initiate an

Information Science Institute

integrating research and teaching activities aimed at investigating the role of **information** from various viewpoints: from the fundamental theoretical underpinnings of information to the science and engineering of novel information substrates, biological pathways, communication networks, economics, and complex social systems.

The specific means and goals for the Center are:

- initiate the Science Lecture Series on Information to collectively ponder short and long term goals;
- study dynamic information theory that extends information theory to time–space–varying situations;
- advance information algorithmics that develop new algorithms and data structures for the application of information;
- encourage and facilitate interdisciplinary collaborations;
- provide scholarships and fellowships for the best students, and support the development of new interdisciplinary courses.