# McCoy Presentation for:

W. Szpankowski
Department of Computer Science
Purdue University
USA

March 22, 2011

# Outline

1. Szpankowski's McCoy Nomination Overview
2. Shannon Information Theory
3. Szpankowski's Technical Contributions
   - Entropy of Hidden Markov Process
   - Constrained Channel Capacity
4. Szpankowski's Vision: Science of Information
   - What is Information?
   - Post Shannon Information Theory
   - NSF Science and Technology Center
5. Quotes from Letter Writers

# Szpankowsk's McCoy Nomination Overview

McCoy Nomination of Szpankowski for:

- Solving long-standing open problems: the entropy of hidden Markov processes and the noisy constrained capacity.
- For developing innovative analytic methods for Shannon information theory leading to solutions of several open problems (e.g., Ziv's conjecture, Wyner-Ziv conjecture, Steinberg-Gutman conjecture, Huffman's code redundancy, Csiszár-Shields renewal process redundancy )
- Visionary ideas that led first to the creation of the field of "analytic information theory," and subsequently to broadening of Shannon Information Theory to a new science of information, leading to the establishment of Indiana's first NSF Science of Technology Center for Science of Information (CSoI) and one of only two centers ever awarded in computing disciplines.

# Szpankowsk's McCoy Nomination Overview

This nomination is based on the following recent technical results:

(1) P. Jacquet, G. Seroussi and W. Szpankowski, On the Entropy of a Hidden Markov Process, *Theoretical Computer Science*, 395, 203-219, 2008.

(2) J. Konorski and W. Szpankowski, What is Information? *Festschrift in Honor of Jorma Rissanen*, 154-172, 2008.

(3) P. Jacquet and W. Szpankowski, Noisy Constrained Capacity for BSC, *IEEE Transaction on Information Theory*, 56, 5412- 5423, 2010.

# Outstanding Challenges in Computing

The most pressing challenge of our times is the data deluge and the transformation from data to information, and subsequently to knowledge.

1. 25.21 billion web pages (2009), over 1 trillion distinct URLs (2008).

2. The amount of data in the deep web far exceeds this.

3. About 56% of the text data is in english.

4. Easy Questions: How much unique data? How much information in text? Translating this information into actionable form?

5. Increasingly data is not in the form of text – social networks, tweets, scientific data (interactions, geometries, time series), economic transactions, etc.

6. Harder Questions: How do we quantify this data, how do we extract information from these datasets? How do we act on this information?

7. Really Hard Questions: Information has cause and consequence – How do we reach beyond information?

# Outstanding Challenges in Computing

These are profound questions and Wojtek is an acknowledged world leader in quantitative methods addressing these problems.

- The best researchers from the premier institutions, worldwide, have rallied around him to define and promote the area **(visibility)**.
- Wojtek has solved some of the longest standing problems in the area **(Depth)**.
- Wojtek's unique contributions transcend computing – reaching out to scientific disciplines such as life sciences and physics **(Breadth and Impact)**.
- He has driven the research agenda of the community at large, through his tireless contributions in the form of conferences, journals, and workshops **(Service)**.

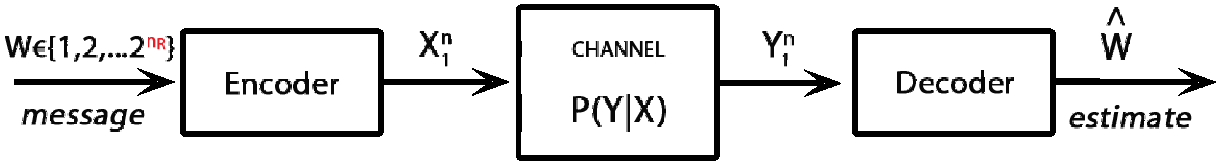
# Three Theorems of Shannon

**Theorem 1 & 3**. (**Shannon 1948; Lossless & Lossy Data Compression**)

compression bit rate $\geq$ source entropy $H(X)$;
for distortion level $D$:
lossy bit rate $\geq$ rate distortion function $R(D)$.

**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.



# Theorem 1: AEP and Typical Sequences

**Shannon-McMilan-Breiman**:

$-\frac{1}{n}\log P(X_1^n) \to H(X) \quad (\text{pr.})$

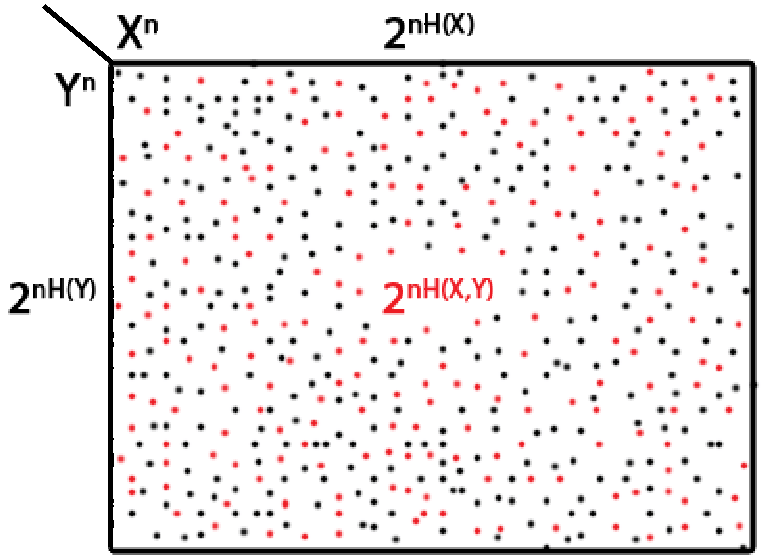
$H(X)$ is the entropy rate.

**Asymptotic Equipartition Property**: Sequences of length $n$ can be partitioned into

| | | |
|---|---|---|
| good set | $G_n^\varepsilon$ | $P(w) \sim 2^{-nH(X)}, \quad w \in G_n^\varepsilon$ |
| bad set | $B_n^\varepsilon$ | $P(B_n^\varepsilon) < \varepsilon.$ |

Also, $|G_n^\varepsilon| \sim 2^{nH(X)}$.

# Theorem 2: Shannon Random Coding




There are $2^{nH(X)}$ X-typical sequences

There are $2^{nH(Y)}$ Y-typical sequences

There are $2^{nH(X,Y)}$ jointly X,Y-typical pair of sequences

**Decoding Rule**: Declare that sequence sent $X$ is the one that is jointly typical with the received sequence $Y$ provided there is unique $X$.

The probability of error (more than one typical pair is):

$$\frac{2^{nH(X,Y)}}{2^{n(H(X)+H(Y))}} = 2^{n(H(X)+H(Y)-H(X,Y))} = 2^{-nI(X,Y)}.$$

Since there are $2^{nR}$ messages sent, the total error probability is approximately

$$\min P(\text{error}) \sim 2^{-n(\sup_{P(X)} I(X,Y)-R)} = 2^{-n(C-R)}, \quad C = \sup_{P(X)} I(X,Y).$$

# (Szpankowsky 2010) Noisy Constrained Channel

Let $\mathcal{S}$ denote the set of binary constrained sequences of length $n$. Here:

$$\mathcal{S}_{d,k} = \{\text{(d,k) sequences}\},$$

i.e., no sequence contains a run of zeros shorter than $d$ or longer than $k$ (applications: DVD, CD, blue-rays, biology).

**Sequence $X \in \mathcal{S}_{(d,k)}$ can be represented as a MARKOV PROCESS.**

$C(\mathcal{S}, \varepsilon)$ – noisy constrained capacity defined as

$$C(\mathcal{S}, \varepsilon) = \sup_{X \in \mathcal{S}} I(X;Y) = \lim_{n\to\infty} \frac{1}{n} \sup_{X_1^n \in \mathcal{S}_n} I(X_1^n, Y_1^n).$$

**This is/was an open problem since 1948 Shannon work.**

# Entropy of Hidden Markov Process

**Hidden Markov Process**: Since

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(\varepsilon)$$

($H(\varepsilon) = -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$) we need to find $H(Y)$.
But $Y$ is a Hidden Markov Process (HMP) since it is a noisy version of the Markov Process $X$.

Entropy of HMP was first investigated by Blackwell in 1956 but no significant progress since then. Why?

**Theorem 1** (Jacquet, Seroussi, & Szpankowski, 2004). *Consider the HMP $Y$ as defined above. The entropy rate*

$$H(Y) = \lim_{n\to\infty} \frac{1}{n} \mathbf{E}[-\log\left(\mathbf{p}_1 \mathbf{M}(Y_1, Y_2) \cdots \mathbf{M}(Y_{n-1}, Y_n) \mathbf{1}^t\right)] = \mu(P)$$

*where $\mu(P)$ is a top Lyapunov exponent of random matrices $\mathbf{M}(Y_1, Y_2) \cdots \mathbf{M}(Y_{n-1}, Y_n)$ defined as*

$$\mathbf{M}(Y_{n-1}, Y_n) = \begin{bmatrix} (1-\varepsilon) P_X(Y_n|Y_{n-1}) & \varepsilon P_X(\bar{Y}_n|Y_{n-1}) \\ (1-\varepsilon) P_X(Y_n|\bar{Y}_{n-1}) & \varepsilon P_X(\bar{Y}_n|\bar{Y}_{n-1}) \end{bmatrix}.$$

# Asymptotic Expansion

We now assume that $P(E_i = 1) = \varepsilon \to 0$ is small (never studied before).

**Theorem 2** (Jacquet and Szpankowski, 2004, 2007). *Assume $r$th order Markov. Then the entropy rate of $Y$ for small $\varepsilon$ is*

$$H(Y) = H(X) - f_0(P)\varepsilon \log \varepsilon + f_1(P)\varepsilon + o(\varepsilon)$$

*for explicitly computable $f_0(P)$ and $f_1(P)$; e.g.,*

$$f_1(P) = \sum_{z_1^{2r+1}} P_X(z_1^{2r+1}) \log \frac{P_X(z_1^{2r+1})}{P_X(\bar{z}_1^{2r+1})} = \mathbb{D}\left(P_X(z_1^{2r+1}) || P_X(\bar{z}_1^{2r+1})\right) ,$$

*where $\bar{z}^{2r+1} = z_1 \dots z_r \bar{z}_{r+1} z_{r+2} \dots z_{2r+1}$. In the above, $\mathbb{D}$ denotes the Kullback-Liebler divergence.*

# Noisy Constrained Capacity

In 2004 Marcus at al. stated:
"*. . . while calculation of the noise-free capacity of constrained sequences is well known, the computation of the capacity of a constraint in the presence of noise . . . has been an unsolved problem in the half-century since Shannon's landmark paper . . .*"

We just showed that

$$H(Y) = H(P) - f_0(P)\varepsilon \log \varepsilon + f_1(P)\varepsilon + o(\varepsilon).$$

Let $P^{\max}$ be the maxentropic maximizing $H(P)$. Then we prove

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) - (1 - f_0(P^{\max}))\varepsilon \log \varepsilon + (f_1(P^{\max}) - 1)\varepsilon + o(\varepsilon)$$

where $C(\mathcal{S})$ is the capacity of noiseless system ($\varepsilon = 0$)

**Theorem 3** (Jacquet & Szpankowski, 2007). *For $k \leq 2d$, we have*

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) + A \cdot \varepsilon + O(\varepsilon^2 \log \varepsilon).$$

*For $k > 2d$, we shall prove that*

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) + B \cdot \varepsilon \log \varepsilon + O(\varepsilon),$$

*where $A$ and $B$ are explicitly computable constants.*

# Broader Vision: What is Information?

**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".

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

Event-Driven Paradigm: Systems, State, Event, Context, Attributes, Objective: Objective function objective$(R, C)$ maps systems' rule $R$ and context $C$ in to an objective space.

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

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

*where the* **cost** *(weight, distance) is a cost function.*

# Post-Shannon Challenges

Classical Information Theory needs a recharge to meet new challenges of emerging applications in biology, modern communication, knowledge extraction, economics and physics, . . . .

We need to extend traditional formalisms for information to include ("meaning"):

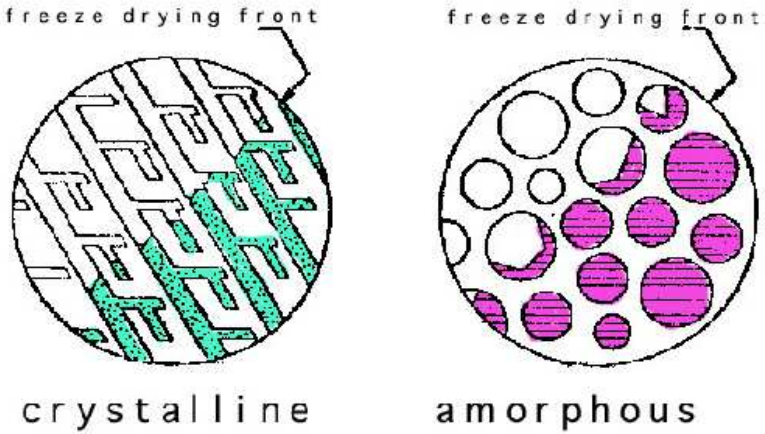
structure, time, space, and semantics ,

and others such as:

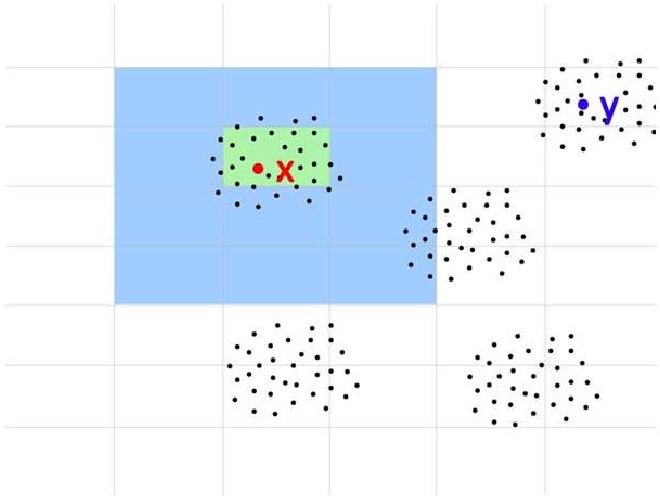
dynamic information, limited resources, complexity, physical information, representation-invariant information, and cooperation & dependency.

# Structure, Time & Space, and Semantics

**Structure**:
Measures are needed for quantifying information embodied in structures (e.g., material structures, nanostructures, biomolecules, gene regulatory networks protein interaction networks, social networks, financial transactions).
(Y. Choi & W.S., ISIT, 2009.)



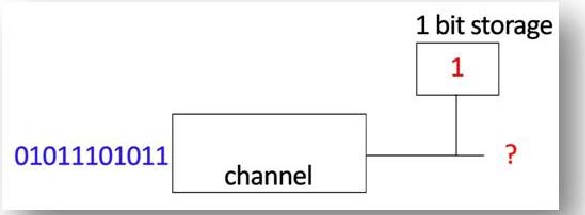

**Time & Space**:
Classical Information Theory is at its weakest in dealing with problems of delay (e.g., information arriving late maybe useless or has less value).
(P. Jacquet et al., IT 2010.)



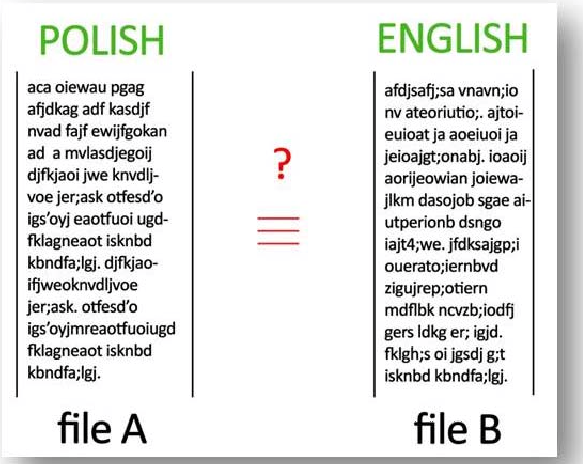

**Semantics & Learnable information**:
Data driven science focuses on extracting information from data. How much information can actually be extracted from a given data repository? How much knowledge is in Google's database?
(M. Sudan et al., 2010.)

# Limited Resources, Representation, and Cooperation

**Limited Computational Resources**:
In many scenarios, information
is limited by available
computational resources
(e.g., cell phone, living cell).
(Helman & Cover, 1970, "Learning with Limited Memory".)




**Representation-invariant of information**:
How to know whether two representations
of the same information
are information equivalent?




**Cooperation**. Often subsystems may be in conflict (e.g., denial of service) or in collusion (e.g., price fixing). How does cooperation impact information? (In wireless networks nodes should cooperate in their own self-interest.)
(Cuff, et al. IT, 2010).

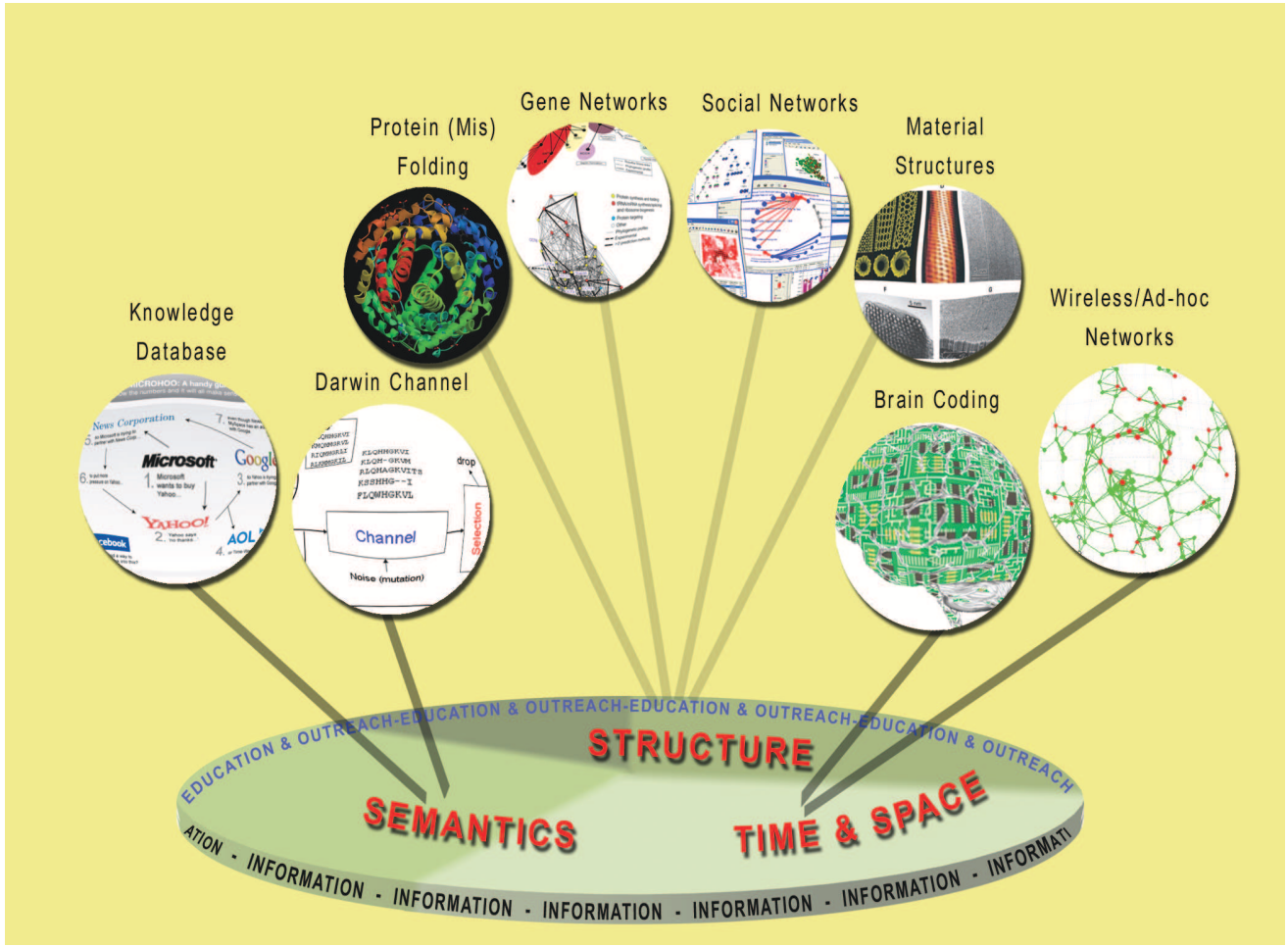
# Selected Quotes from Letter-Writers

- "Many of Szpankowski's research works are jewels of discrete mathematics." (Flajolet)
- "I believe that Purdue University, and indeed the worldwide scientific community in the fields of information theory, communication theory, computation theory, and systems biology, are fortunate in having a scientist and a leader of his accomplishments and vision.." (Kumar)
- "I consider his work on the analysis of Lempel-Ziv compression schemes in information theory his best. Just for this alone, I would not be surprised to see him capture one day the Shannon award." (Devroye)
- ".. he has already created a remarkable intellectual atmosphere where people from different disciplines are coming together drawn by their interest in various aspects of information science, bringing their own distinct perspective to the field." (Datta on Wojtek's STC Center)

Thank You.

# Science of Information

The overarching vision of **Science of Information** is to develop rigorous principles guiding the extraction, manipulation, and exchange of information, integrating elements of space, time, structure, and semantics.



# Institute for Science of Information

In 2008 Szpankowski launched at Purdue the

**Institute for Science of Information**

and in 2010 National Science Foundation established $25M

**Science and Technology Center**

at Purdue to do ccollaborative work with Berkeley, MIT, Princeton, Stanford, UIUC and Bryn Mawr & Howard U. integrating research and teaching activities aimed at investigating the role of **information** from various viewpoints: from the fundamental theoretical underpinnings of greeninformation 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:

- develope post-Shannon Information Theory,
- Prestige Science Lecture Series on Information to collectively ponder short and long term goals;
- organize meetings and workshops (e.g., Information Beyond Shannon, Orlando 2005, and Venice 2008).
- initiate similar world-wide centers supporting research on information.

# That's It