Complexity

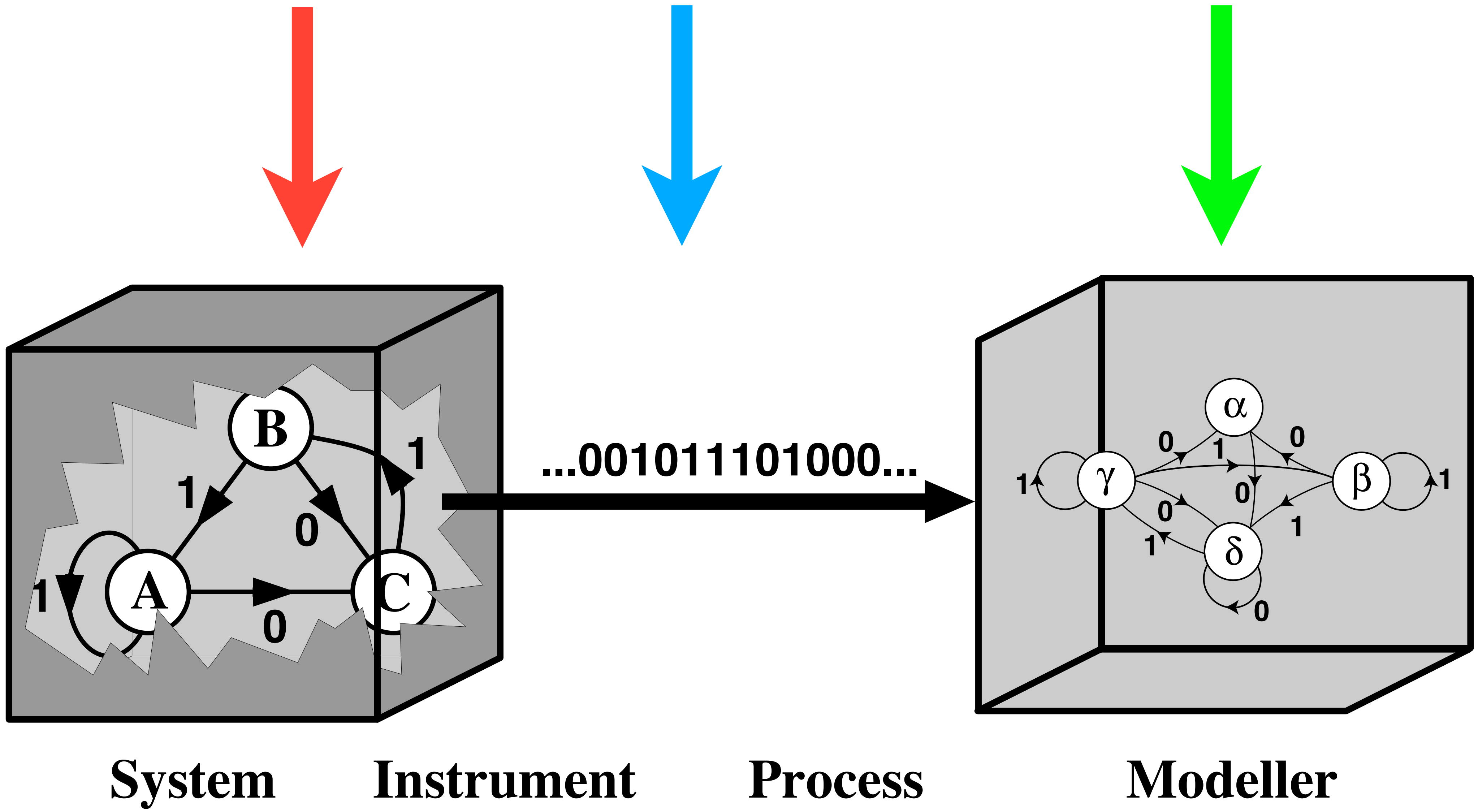
Jim Crutchfield
Complexity Sciences Center
Physics Department
University of California at Davis

Complex Systems Summer School
Santa Fe Institute
St. John's College, Santa Fe, NM
19 June 2017

Main Question

Randomness versus Structure?

Now Next Tomorrow



The Learning Channel

Complexity

Today:

Information Theory for Complex Systems

I. Information Theory

- Algorithmic Basis of Probability
- Information Theory
- Information Measures

Tomorrow:

II. Information & Memory in Processes

Intrinsic Computation

- Measuring Structure
- Intrinsic Computation
- Optimal Models
- Physics of Information

Complexity

References? For example:

Stanislaw Lem, *Chance and Order*, New Yorker **59** (1984) 88-98.
T. Cover and J. Thomas, *Elements of Information Theory*,
Wiley, Second Edition (2006) Chapters 1 - 7.
M. Li and P.M.B. Vitanyi, *An Introduction to Kolmogorov Complexity and its Applications*,
Springer, New York (1993).
J. P. Crutchfield and D. P. Feldman,
"Regularities Unseen, Randomness Observed: Levels of Entropy Convergence",CHAOS **13**:1 (2003) 25-54.
J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney,
"Time's Barbed Arrow: Irreversibility, Crypticity, and Stored Information",
Physical Review Letters **103**:9 (2009) 094101.
R. G. James, C. J. Ellison, and J. P. Crutchfield,
"Anatomy of a Bit: Information in a Time Series Observation",CHAOS **21**:1 (2011) 037109.
J. P. Crutchfield,
"Between Order and Chaos", Nature Physics **8** (January 2012) 17-24.

See http://csc.ucdavis.edu/~cmg/
See online course: http://csc.ucdavis.edu/~chaos/courses/ncaso/

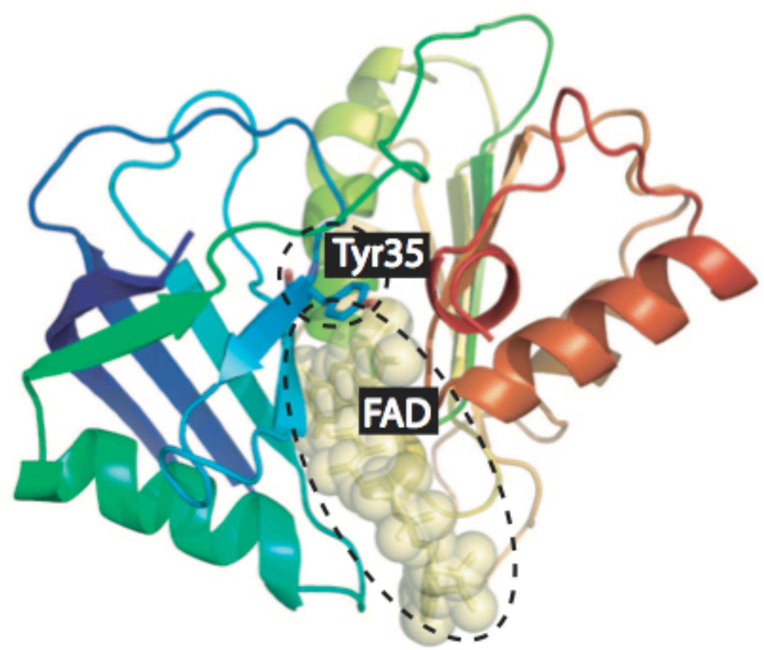
Applications?

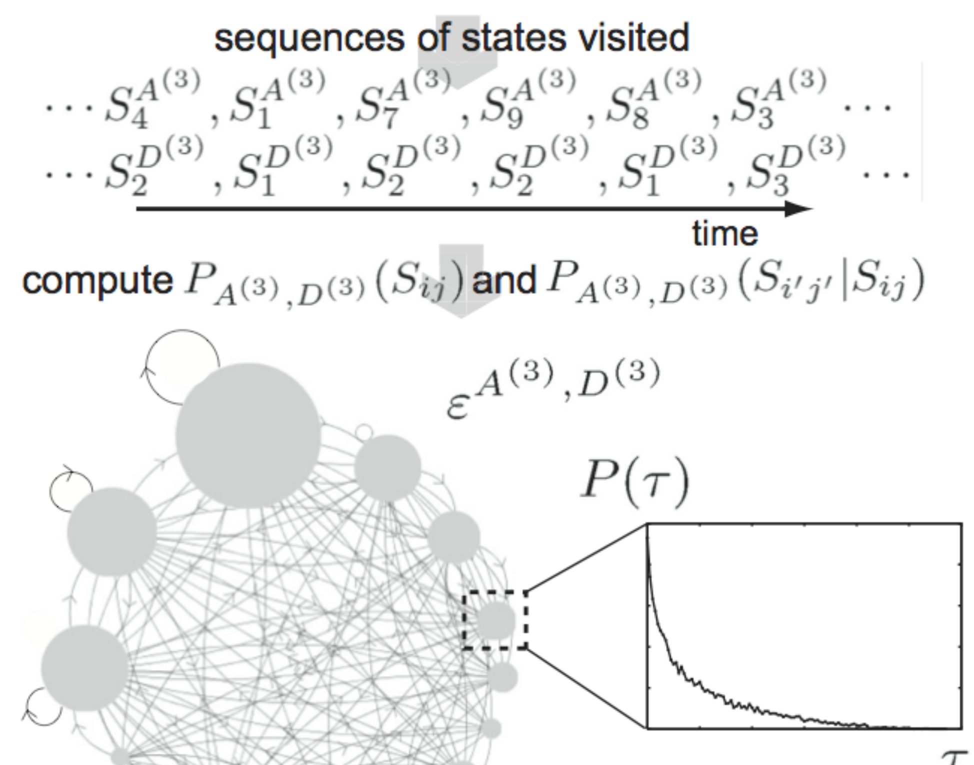
Computational Mechanics:
Application to Experimental
Molecular Dynamics Spectroscopy

Multiscale complex network of protein conformational fluctuations in single-molecule time series

Chun-Biu Li*†‡, Haw Yang§¶, and Tamiki Komatsuzaki*†‡‖

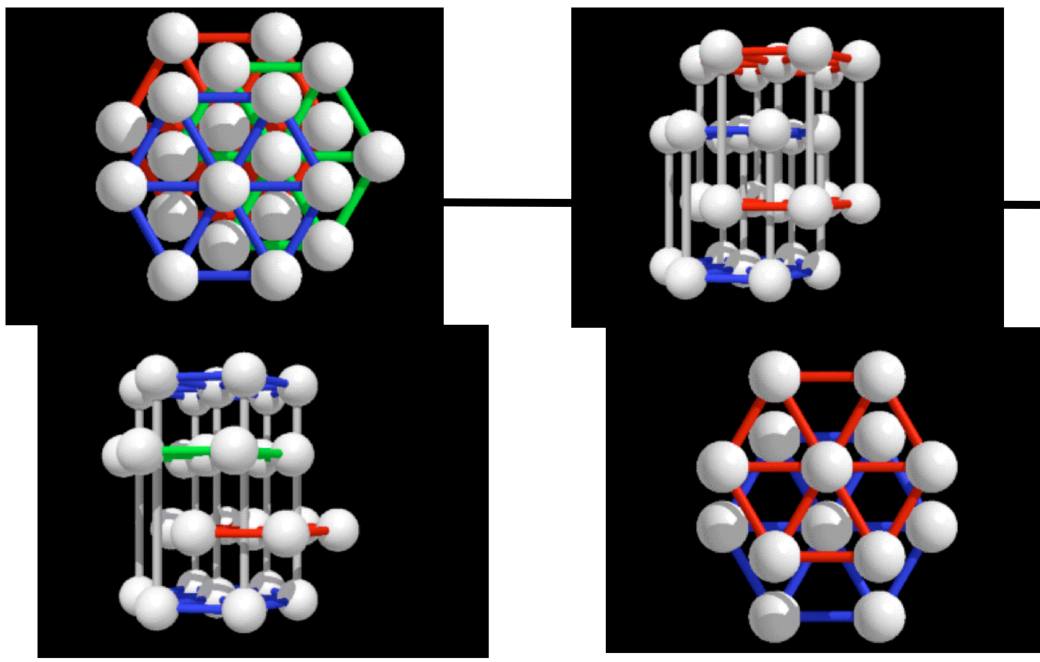
*Nonlinear Sciences Laboratory, Department of Earth and Planetary Sciences, Faculty of Science, Kobe University, Nada, Kobe 657-8501, Japan; †Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan; §Department of Chemistry, University of California, Berkeley, CA 94720; and ¶Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

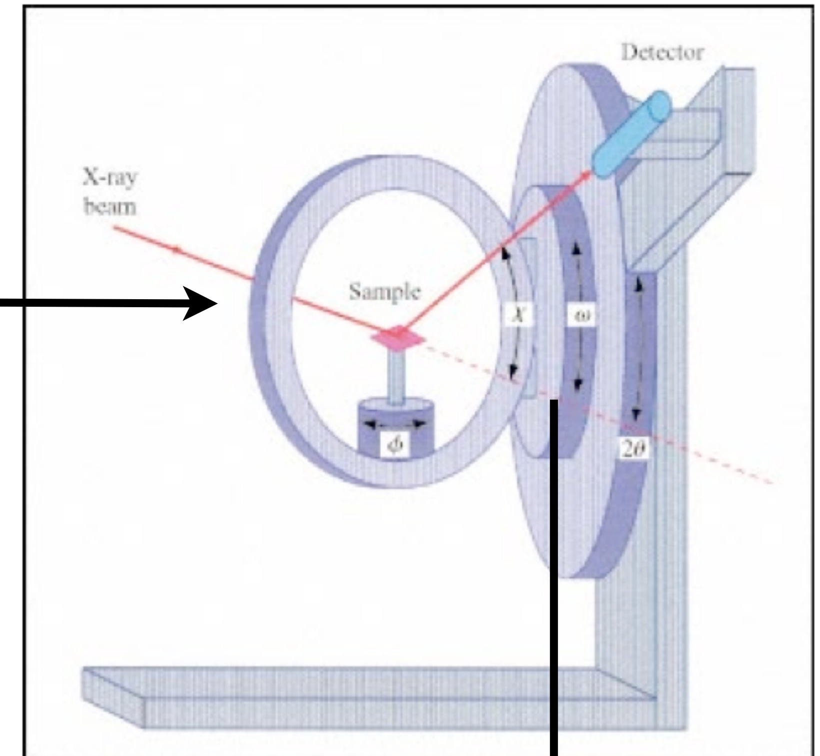


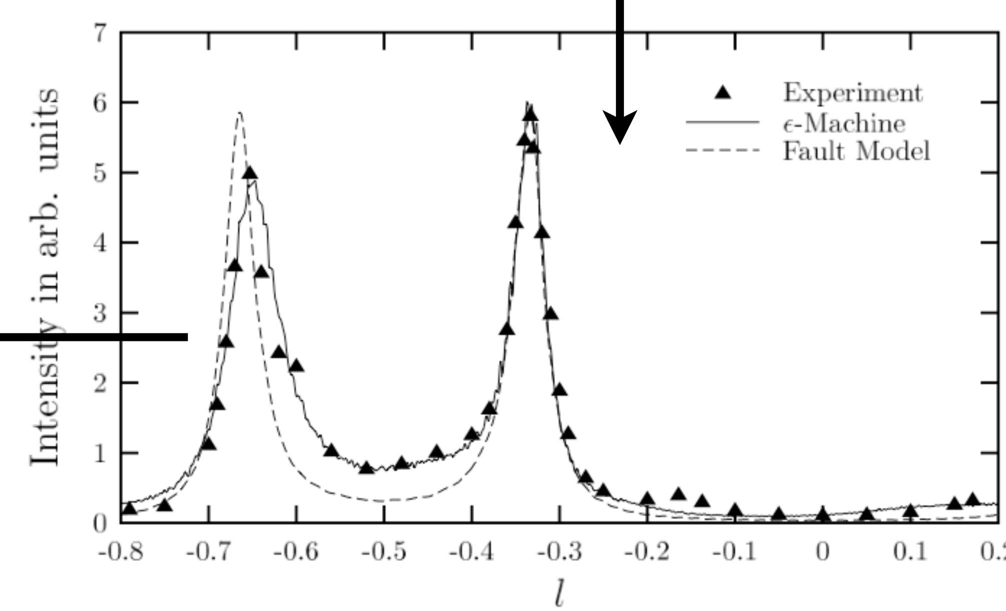


C.-B. Li, H. Yang, & T. Komatsuzaki, Proc. Natl. Acad. Sci USA **105**:2 (2008) 536–541.

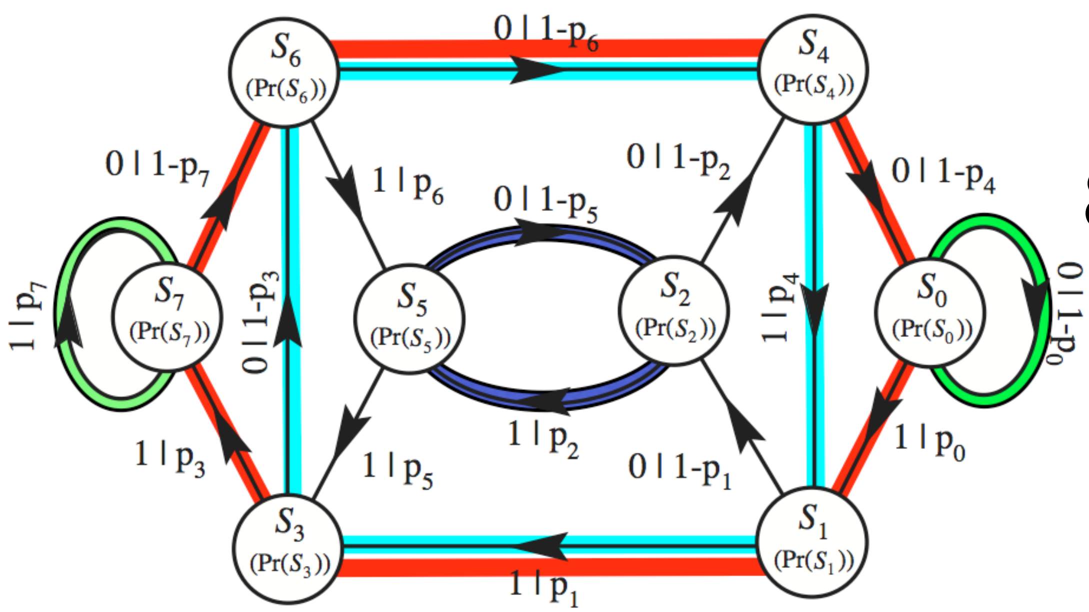
Computational Mechanics: Application to Experimental X-Ray Diffraction





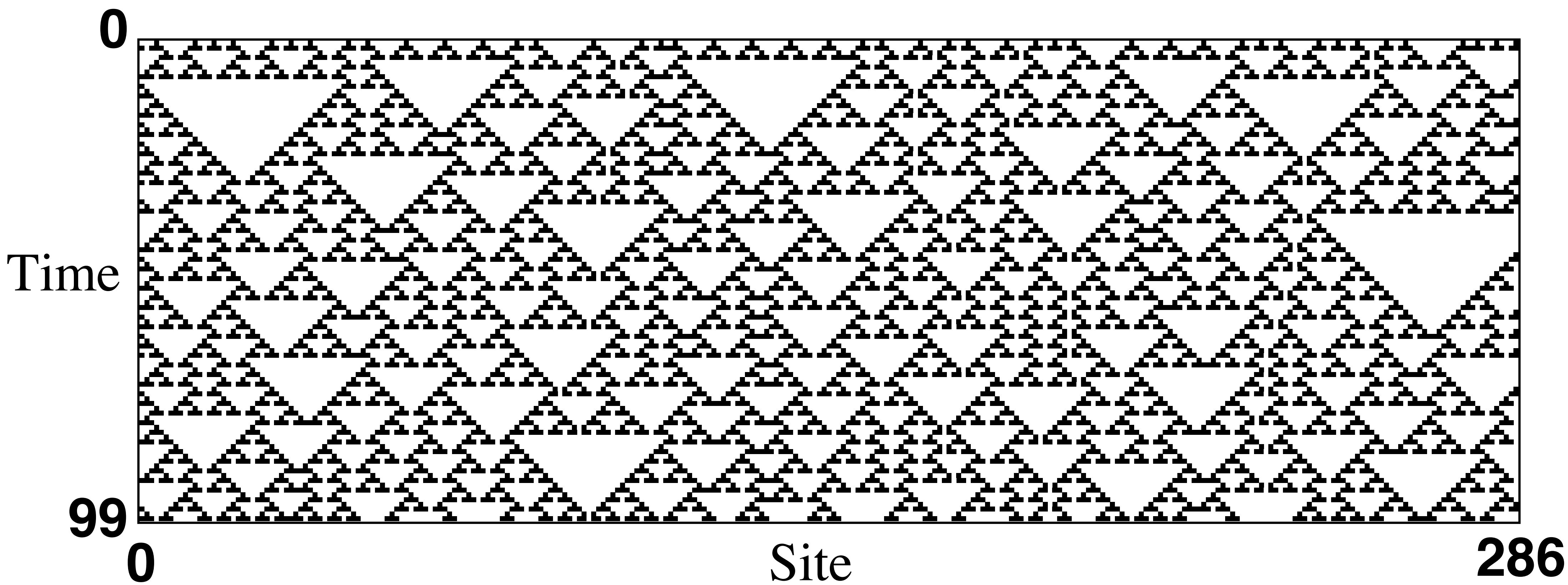


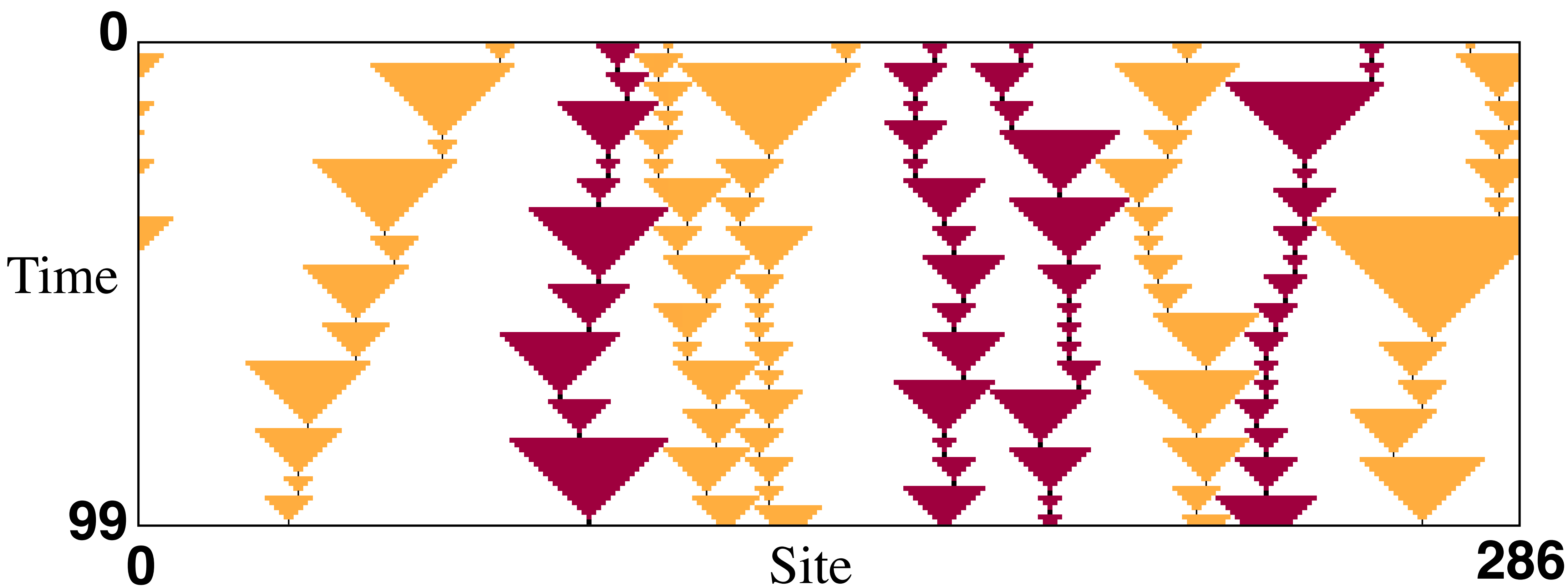
ε-MSR



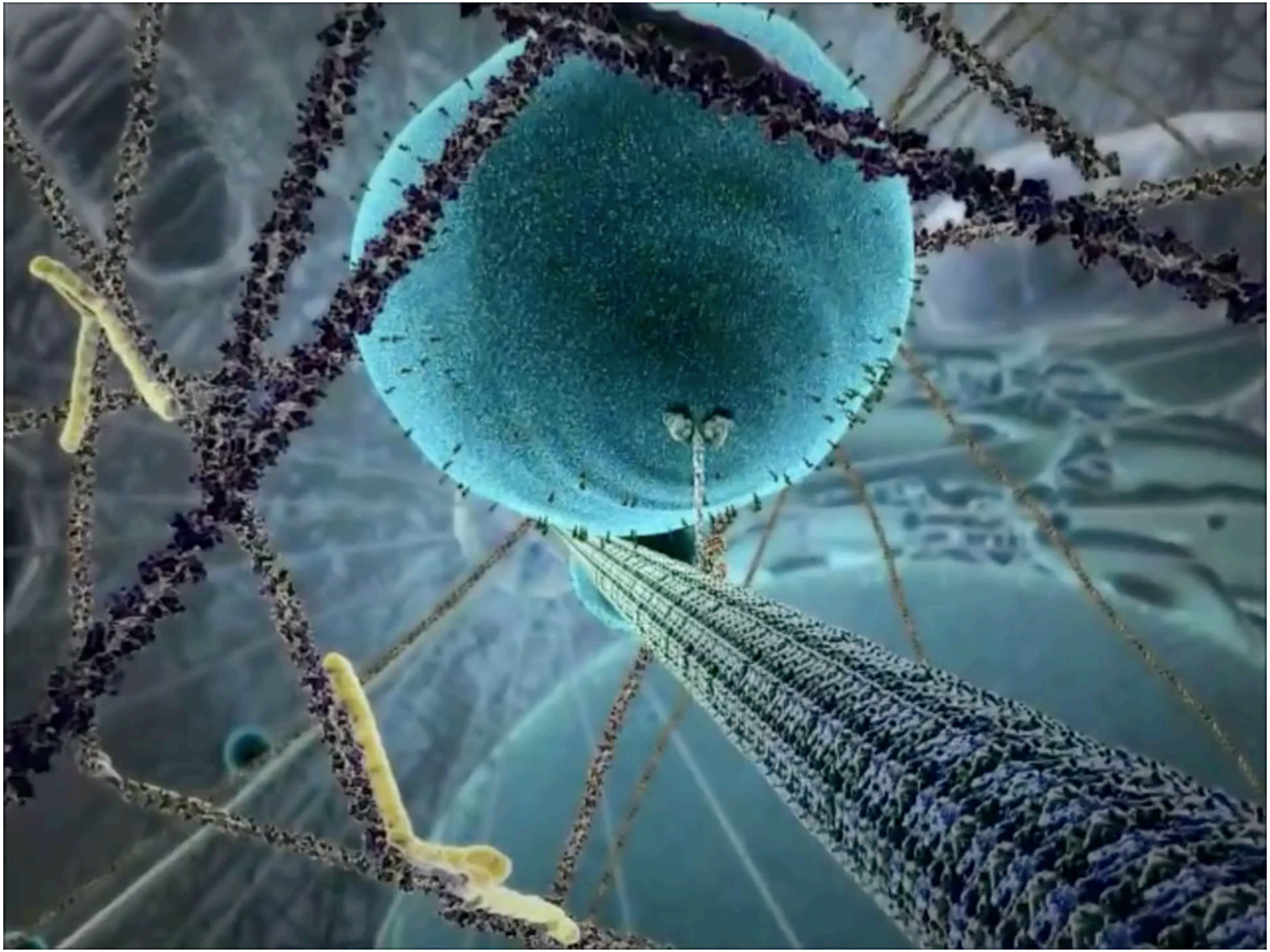
D. P. Varn, G. S. Canright, J. P. Crutchfield, "Discovering Planar Disorder in Close-Packed Structures from X-Ray Diffraction: Beyond the Fault Model", Phys. Rev. B **66**: 17 (2002) 174110-2.

Cellular Automata Computational Mechanics



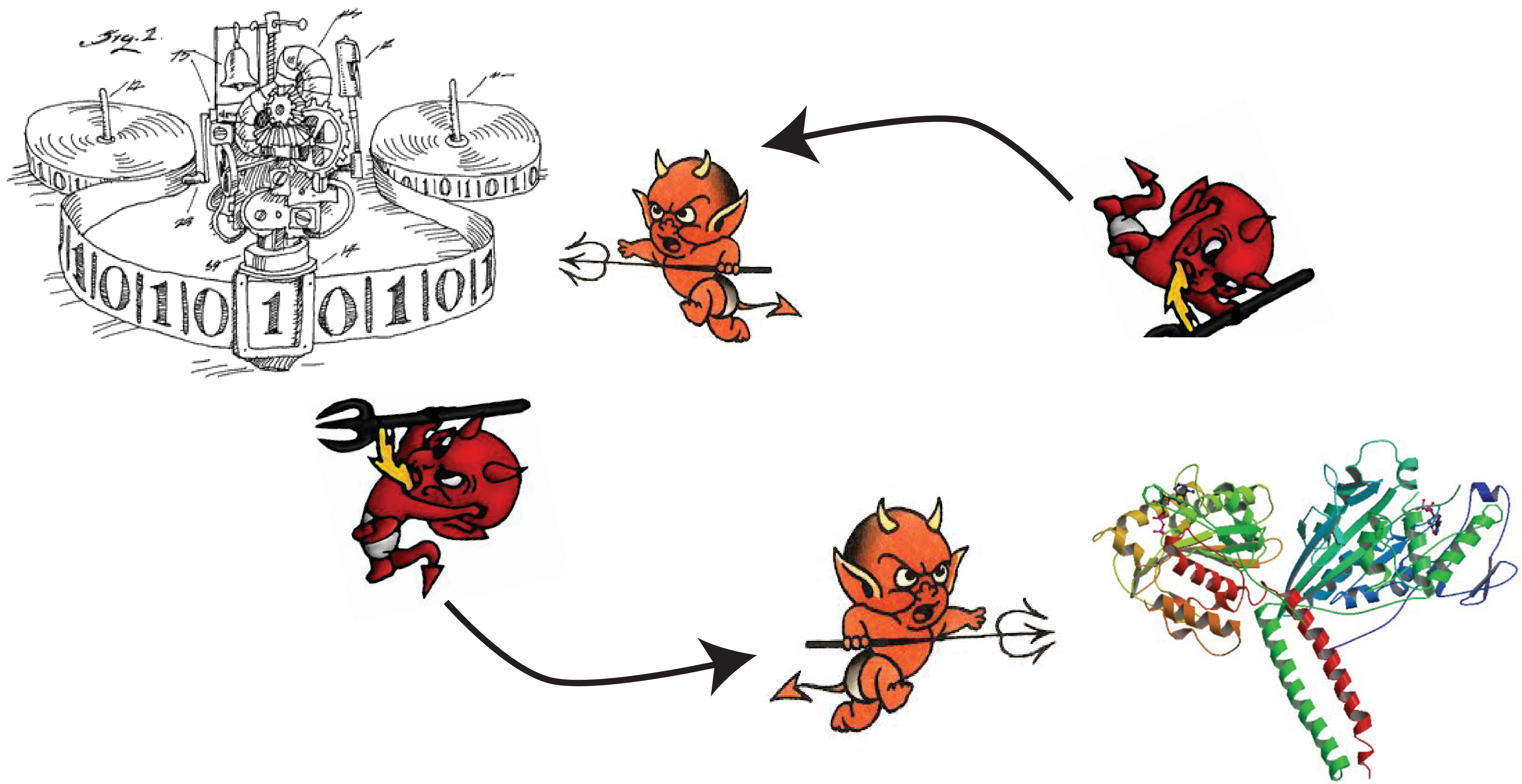


Thermodynamics of Adaptive Complex Systems

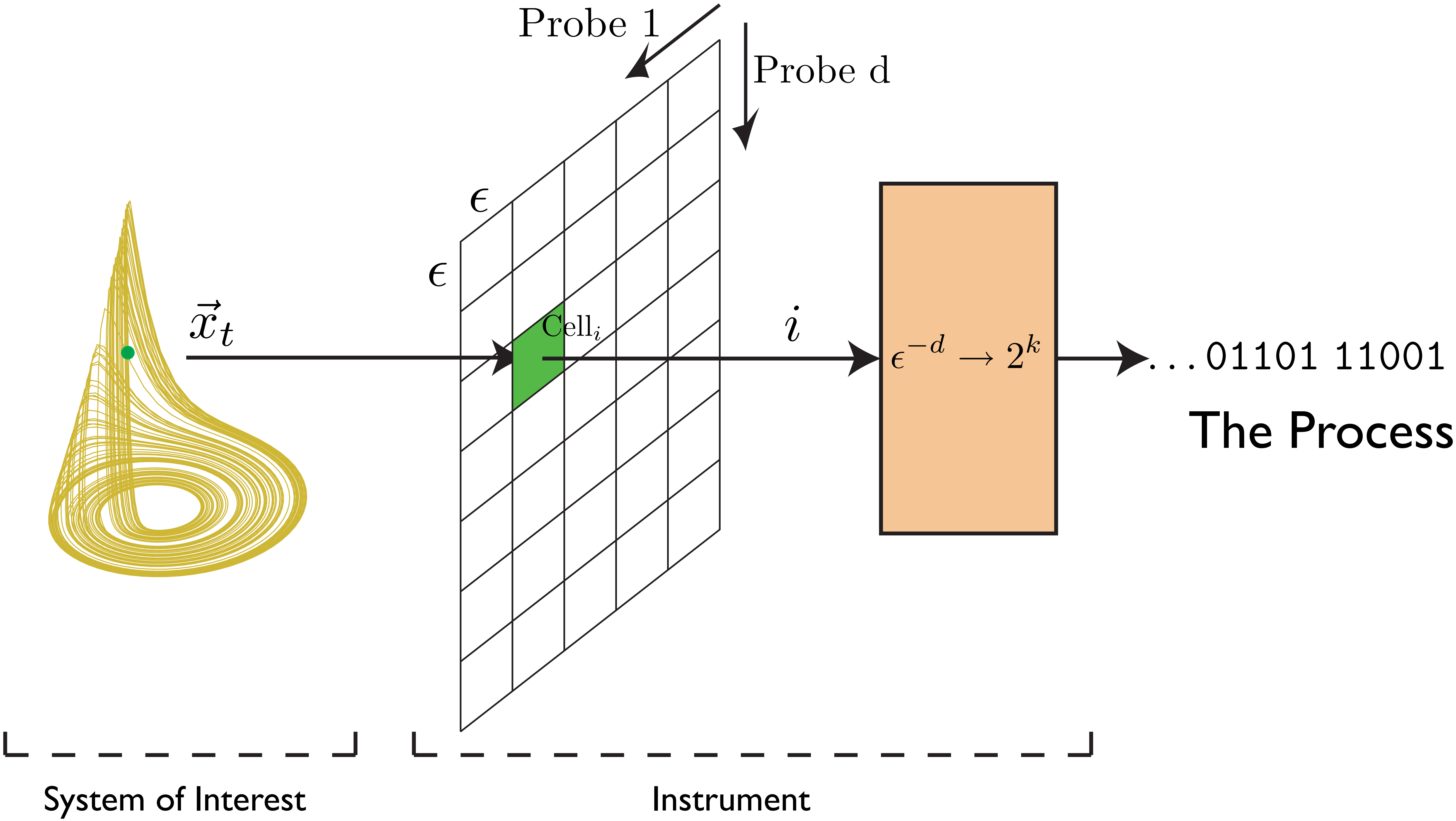


Thermodynamics of Adaptive Complex Systems

Role of “intelligence”
in functioning?
in overcoming fluctuations?



Processes and Their Models



Measurement Channel

Main questions now:

- How do we characterize the resulting process?
  - Measure degrees of unpredictability & randomness.
  - Use probabilities?
  - What correlational structure is there?
- How do we build a model from the process itself?
- How much can we reconstruct about the hidden internal dynamics?

What to do with all of this complicatedness?

1. Algorithmic basis

2. Information theory for complex processes

3. Measures of complexity

4. Optimal models and how to build them

Algorithmic Basis of Probability

Kolmogorov-Chaitin Complexity Theory

The question:

Algorithmic foundation for probability?

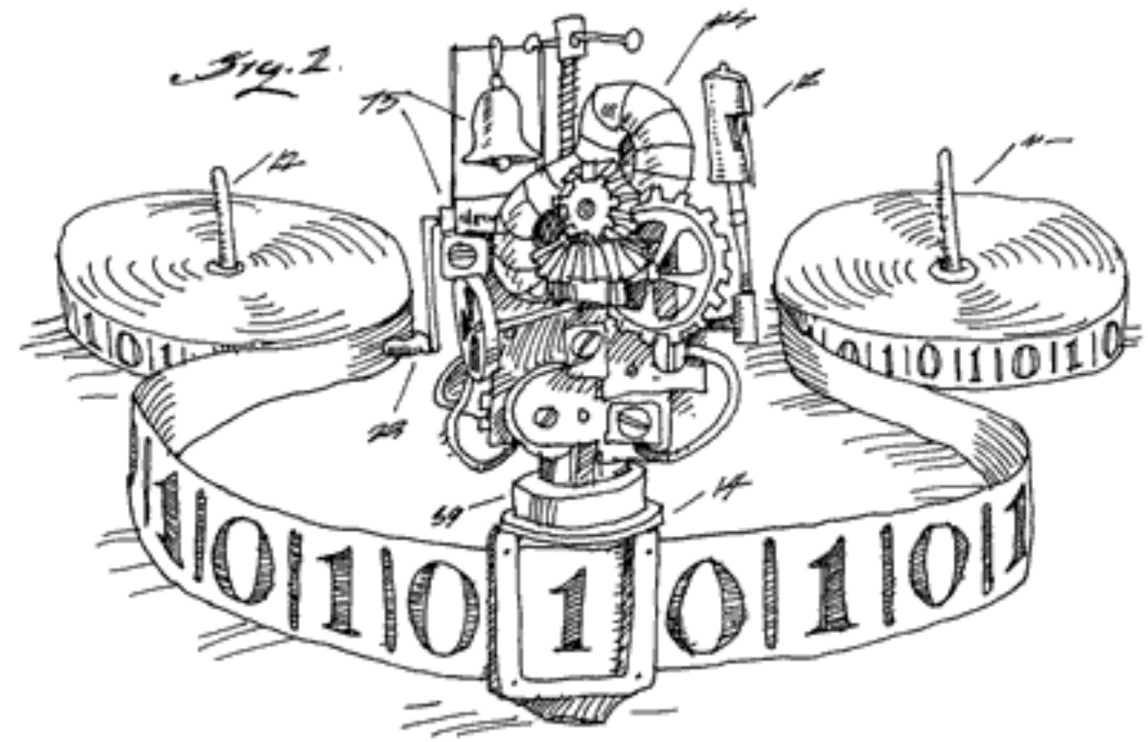
History:

1776: Treatise on probability theory (Laplace)
1920s: Frequency stability (von Mises)
1930s: Foundations of probability theory (Kolmogorov)
1940s: Information theory (Shannon … Szilard 1920s!)
1940s: Automata & computing theory (Turing)
1960s: KC Complexity Theory
(Kolomogorov, Chaitin, Solomonoff, ...)

Kolmogorov-Chaitin Complexity

Turing's machine (1937):

Finite-state controller +
Infinite read-write tape



Machine M :

Device to generate output $x = 1010111\ldots$ from program p :

$$M(p) = x$$

Kolmogorov-Chaitin Complexity

Universal Turing Machine: U

Sufficient states, control logic, and tape alphabet

$\Rightarrow$ Calculate any input-output function

UTM programs generate output: $U(p) = x$

(Python interpreter w/ infinite memory.)

Kolmogorov-Chaitin Complexity:

Size of smallest program p that generates object x

$$K(x) = \min\{|p| : \ U(p) = x\}$$

Kolmogorov-Chaitin Complexity

Consider Python program:

```
def generate_x():
      print x
```

And so:

$$K(x) \leq |x| + \text{constant}$$

For most objects:

$$K(x) \approx |x|$$

Kolmogorov-Chaitin Complexity is not computable.

(Theorem: No program can calculate $K(x)$.)

Kolmogorov-Chaitin Complexity

Exercise! Which has high, which low $K(x)$?

00100100001111110110101010001000
10000101101000110000100011010011
00010011000110011000101000101110
00000011011100000111001101000100

π

Algorithm $\Rightarrow$ low $K(x)$

(Bailey–Borwein–Plouffe 1997)

10000010100011011111110111001100
01101101001100010110010001010100
00101100011011000110001110111000
10110100010000111000111001110011

Random

High $K(x)$

Kolmogorov-Chaitin Complexity

Lessons:

A random object is its own shortest description.

$K(x)$ maximized by random objects.

Probability of objects:

$$\Pr(x) \approx 2^{-K(x)}$$

Alternatives?

Computable?
Scientifically applicable?

Information!

Information ...

Information as uncertainty and surprise:

Observe something unexpected:
Gain information

Bateson: "A difference that makes a difference"

Information ...

Sources of Information?

Apparent randomness:
- Uncontrolled initial conditions
- Actively generated: Deterministic chaos

Hidden regularity:
- Ignorance of forces
- Limited capacity to model structure

Information ...

Information as uncertainty and surprise ...

How to formalize?

Shannon's approach:

A measure of surprise.

Connection with Boltzmann's thermodynamic entropy

Self-information of an event $\propto -\log \Pr(\text{event})$.

Predictable: No surprise $-\log 1 = 0$

Completely unpredictable: Maximally surprised

$$-\log \frac{1}{\text{Number of Events}} = \log(\text{Number of Events})$$

Information ...

Shannon Entropy: $X \sim P$

$$x \in \mathcal{X} = \{1, 2, \ldots, k\}$$
$$P = \{\Pr(x = 1), \Pr(x = 2), \ldots\}$$

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$$

Note: $0 \log 0 = 0$

Units:

Log base 2: $H(X) = [\text{bits}]$

Natural log: $H(X) = [\text{nats}]$

Properties:

1. Positivity: $H(X) \geq 0$
2. Predictive: $H(X) = 0 \;\Leftrightarrow\; p(x) = 1$ for one and only one x
3. Random: $H(X) = \log_2 k \;\Leftrightarrow\; p(x) = U(x) = 1/k$

Information ...

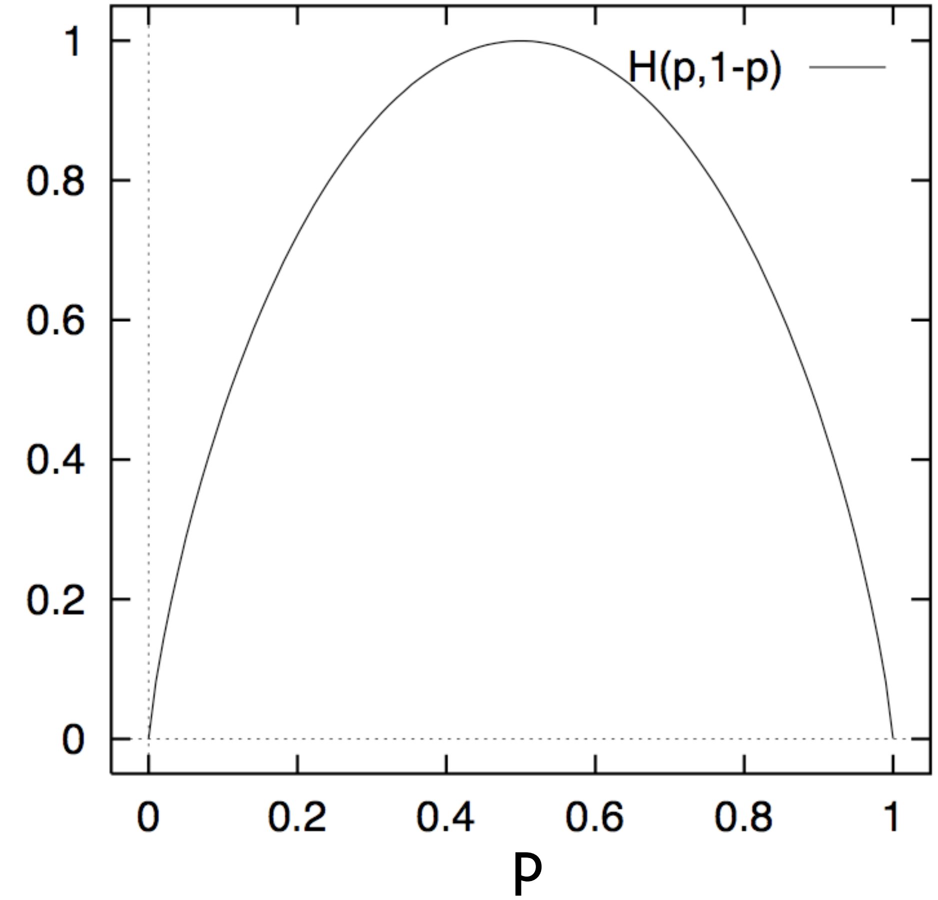
Example: Binary random variable X (Biased Coin)

$$\mathcal{X} = \{0, 1\} \qquad \Pr(1) = p \;\&\; \Pr(0) = 1 - p$$

$H(X)$?

Binary entropy function:

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$



Fair coin: $p = \frac{1}{2}$

$$H(p) = 1 \text{ bit}$$

Completely biased coin: $p = 0 \text{ (or } 1)$

$$H(p) = 0 \text{ bits}$$

Recall: $0 \cdot \log 0 = 0$

Information ...

Example: Independent, Identically Distributed (IID) Process over four events

$$\mathcal{X} = \{a, b, c, d\} \qquad \Pr(X) = (\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8})$$

Entropy: $H(X) = \frac{7}{4}$ bits

Number of questions to identify the event?

x = a? (must always ask at least one question)
x = b? (this is necessary only half the time)
x = c? (only get this far a quarter of the time)

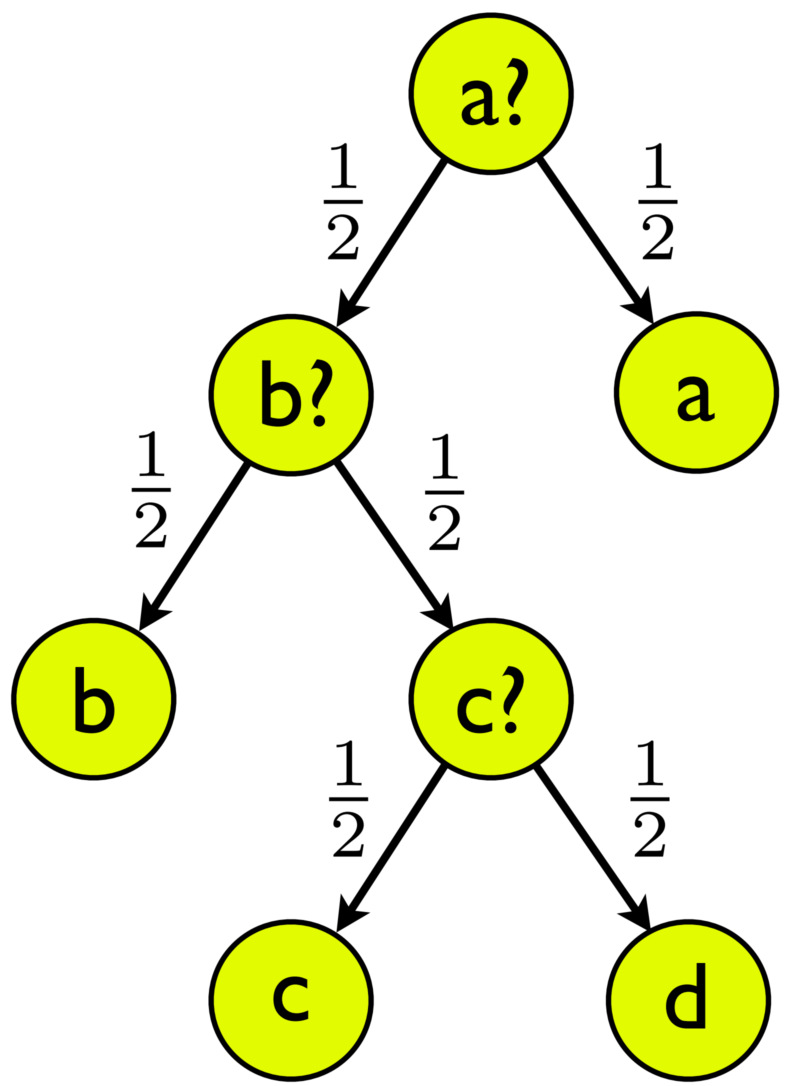
Average number: $1 \cdot 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 1.75$ questions

Interpretation? Optimal way to ask questions.

Information ...

Example: IID Process over four events ...

Average number: $1 \cdot 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 1.75$ questions



$$\Pr(X) = (\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8})$$

Information ...

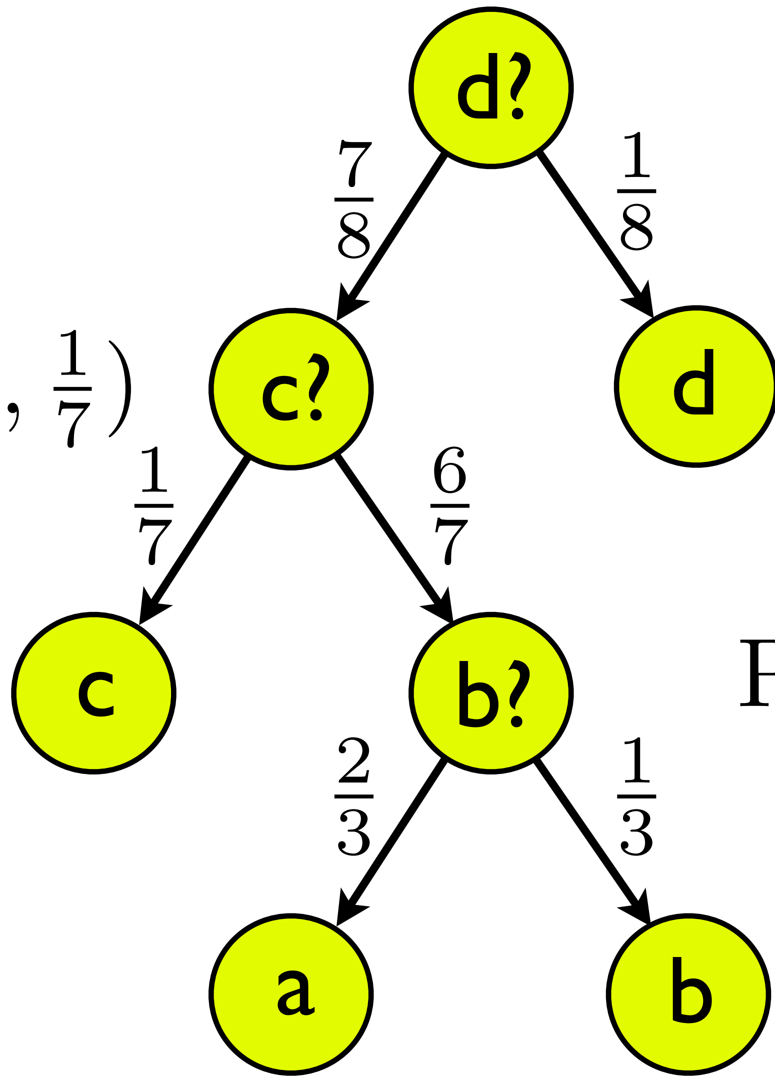
Example: IID Process over four events ...

Query in a different order:

Average number: $1 \cdot 1 + 1 \cdot \frac{7}{8} + 1 \cdot \frac{6}{7} \approx 2.7$ questions

$\Pr(X) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$

$\Pr(a, b, c) = (\frac{4}{7}, \frac{2}{7}, \frac{1}{7})$



$\Pr(a, b) = (\frac{2}{3}, \frac{1}{3})$

Information ...

Example: IID Process over four events

Entropy: $H(X) = \frac{7}{4}$ bits

At each stage, ask questions that are most informative.

Choose partitions of event space that give "most random" measurements.

Theorem:
Entropy gives the smallest number of questions to identify an event, on average.

Information ...

Interpretations of Shannon Entropy:

Observer's *degree of surprise* in outcome of a random variable

Uncertainty *in* random variable

Information required to *describe* random variable

A measure of *flatness* of a distribution

Information ...

Two random variables: $(X, Y) \sim p(x, y)$

Joint Entropy: Average uncertainty in X and Y occurring

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 p(x, y)$$

Independent:

$$X \perp Y \Rightarrow H(X, Y) = H(X) + H(Y)$$

Conditional Entropy: Average uncertainty in X, knowing Y

$$H(X|Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log_2 p(x|y)$$

$$H(X|Y) = H(X,Y) - H(Y)$$

Not symmetric: $H(X|Y) \neq H(Y|X)$

Information ...

Common Information Between Two Random Variables:

$$X \sim p(x) \ \& \ Y \sim p(y)$$

$$(X, Y) \sim p(x, y)$$

Mutual Information:

$$I(X;Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$

Information ...

Mutual Information ...

Properties:

(1) $I(X;Y) \geq 0$

(2) $I(X;Y) = I(Y;X)$

(3) $I(X;Y) = H(X) - H(X|Y)$

(4) $I(X;Y) = H(X) + H(Y) - H(X,Y)$

(5) $I(X;X) = H(X)$

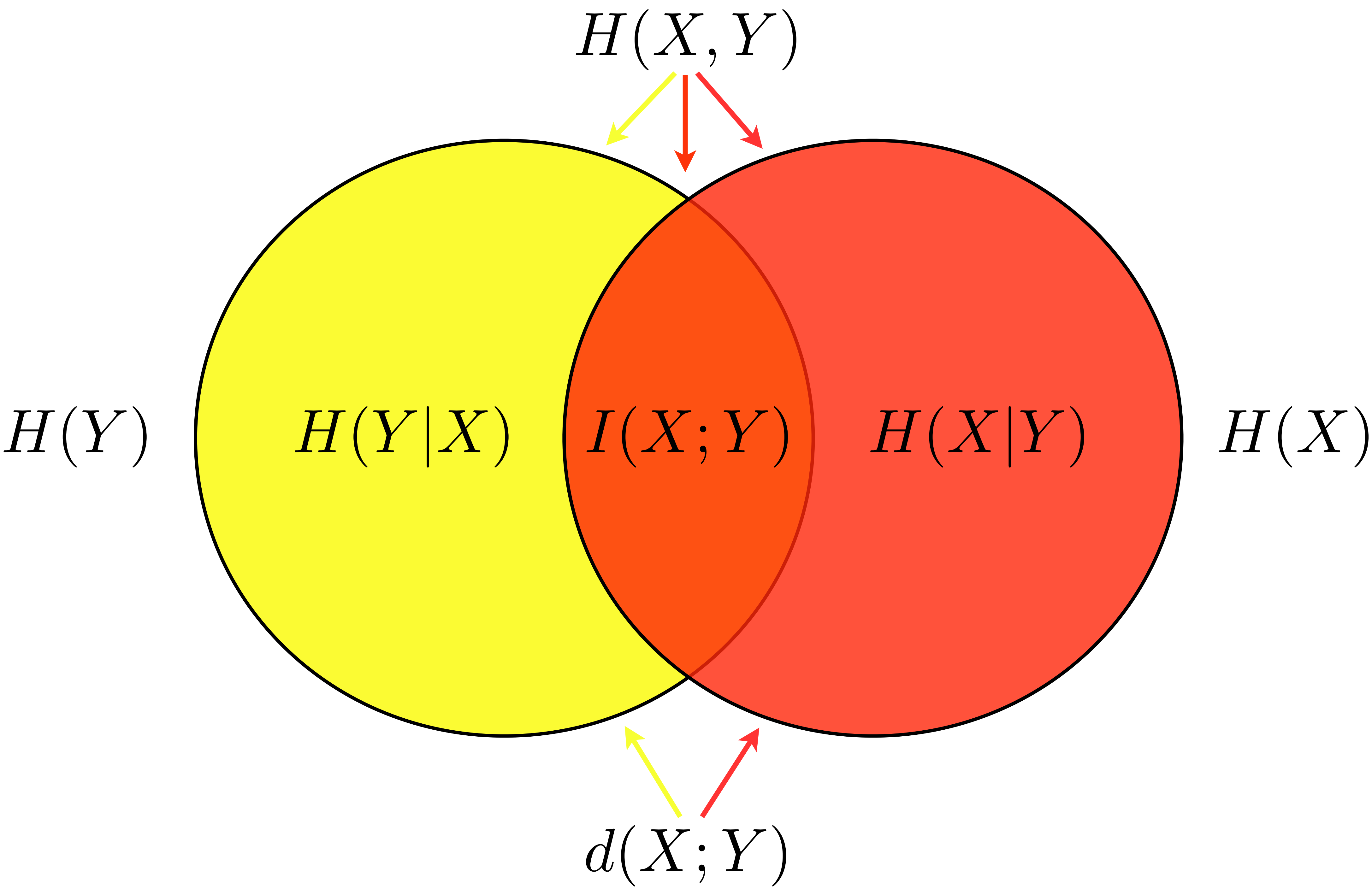
(6) $X \perp Y \Rightarrow I(X;Y) = 0$

Interpretations:

Information one variable has about another
Information shared between two variables
Measure of dependence between two variables

Information ...

Event Space Relationships of Information Quantifiers:

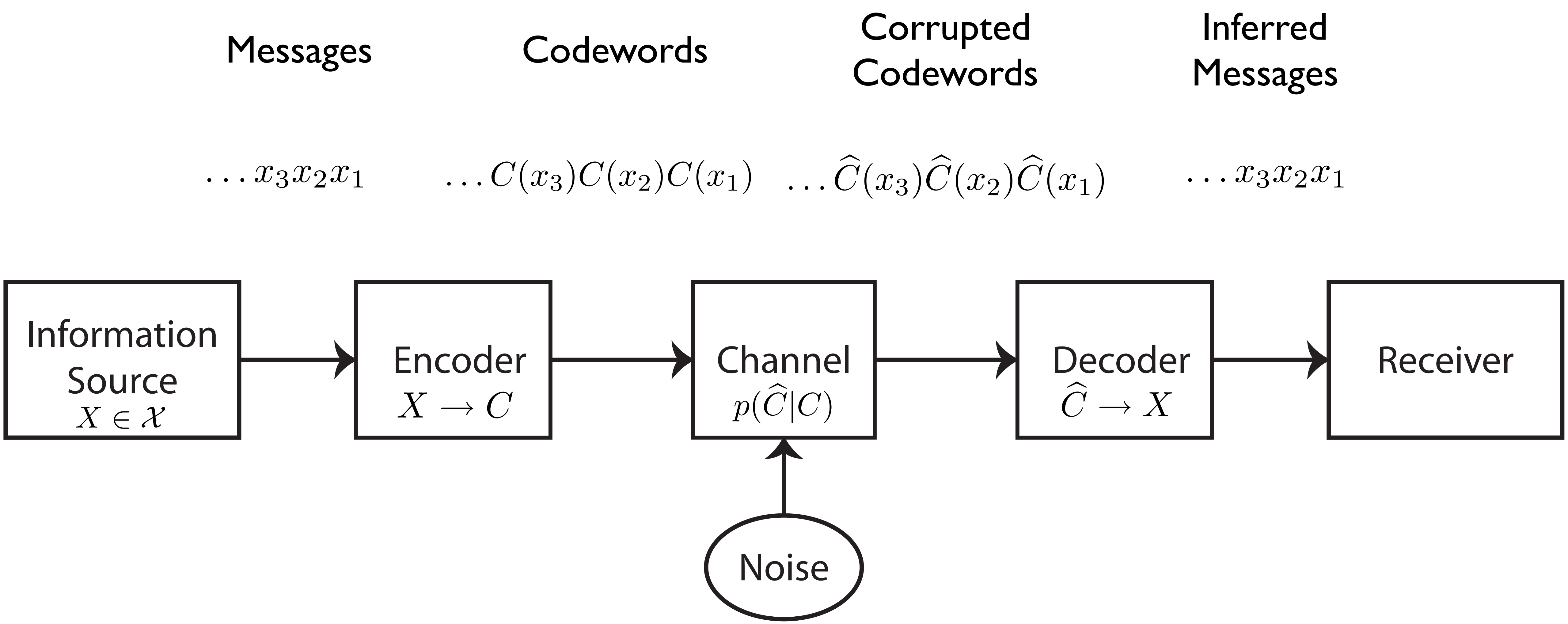


Why information?

1. Accounts for any type of co-relation
   - Statistical correlation ~ linear only
   - Information measures nonlinear correlation
2. Broadly applicable:
   - Many systems don't have "energy", physical modeling precluded
   - Information defined: social, biological, engineering, ... systems
3. Comparable units across different systems:
   - Correlation: Meters v. volts v. dollars v. ergs v. ...
   - Information: bits.
4. Probability theory ~ Statistics ~ Information
5. Complex systems:
   - Emergent patterns!
   - We don't know these ahead of time

Information in Processes ...

Communication channel:



Real Information Theory:

How to compress a process:

Can't do better than $H(X)$

(Shannon's First Theorem)

How to communicate a process's data: $H(X) \leq \mathcal{C}$

Can transmit error-free at rates up to channel capacity

(Shannon's Second Theorem)

Both results give operational meaning to entropy.

Previously, entropy motivated as a measure of surprise.

Complexity

Information Theory for Complex Systems

Today:

- Complex Processes
- Information Measures

Tomorrow:

- Information & Memory in Processes

Intrinsic Computation

- Measuring Structure
- Optimal Models
- Structure = Computation

See online course:

http://csc.ucdavis.edu/~chaos/courses/ncaso/