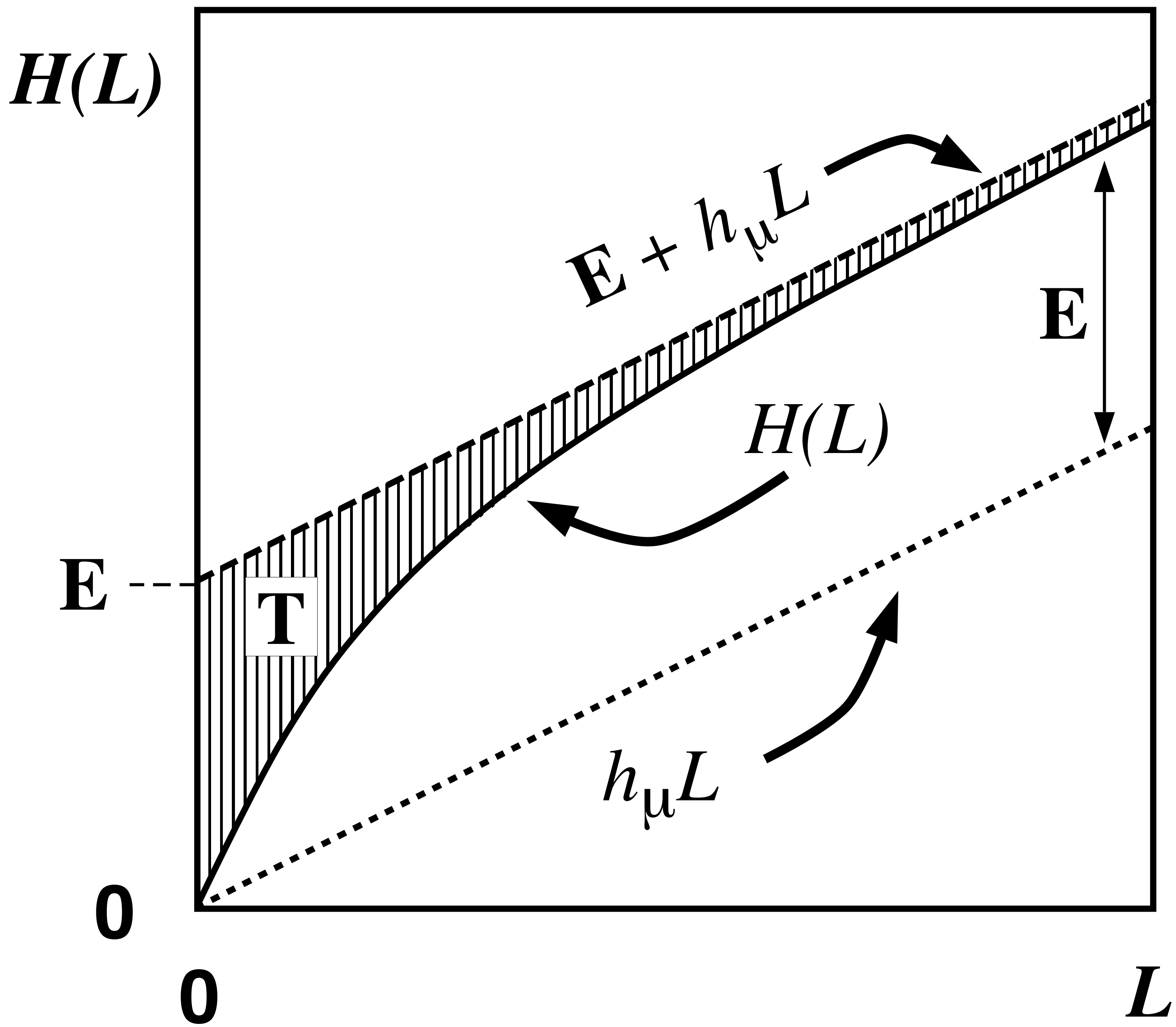
Information in Complex Systems

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

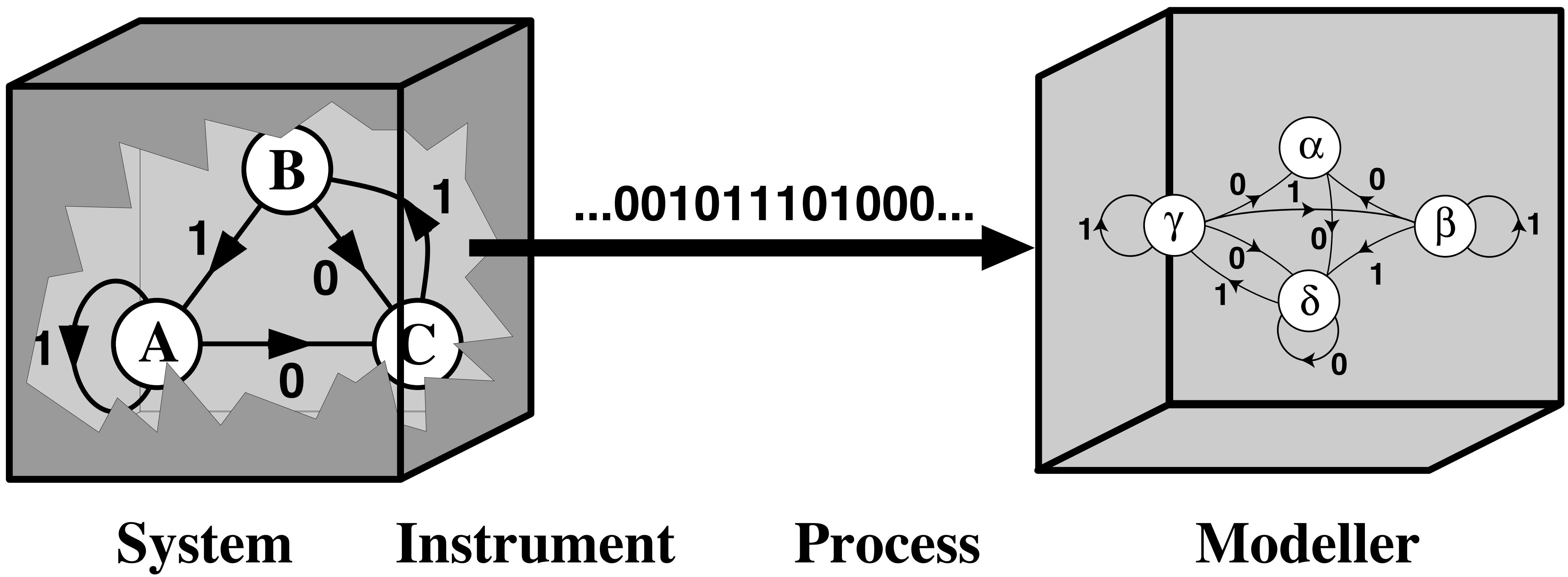
Complex Systems Summer School
Institute of American Indian Arts
Santa Fe, New Mexico
20 June 2019

Information Roadmap for a Complex Process



What's wrong with information theory?

The Learning Channel:



Central questions:

What are the states?

What is the dynamic?

The Learning Channel ...
The Prediction Game

Rules:

1. I give you a data stream (an observed past sequence).

2. You predict its future.

3. You give a model (states & transitions) describing the process.

The Learning Channel ...

The Prediction Game ...

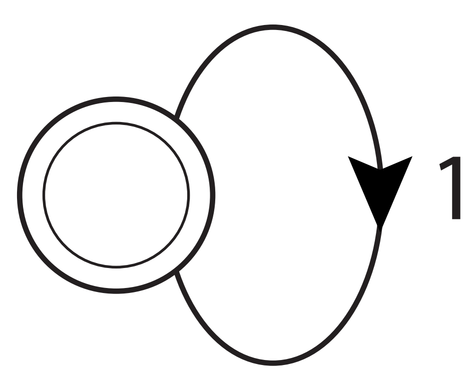
Process I:

Past: $\ldots 111111111111$

Your prediction is?

Future: $111111111111\ldots$

Your model (states & dynamic) is?



The Learning Channel ...

The Prediction Game ...

Process II:

Past: $\ldots 10110010001101110$

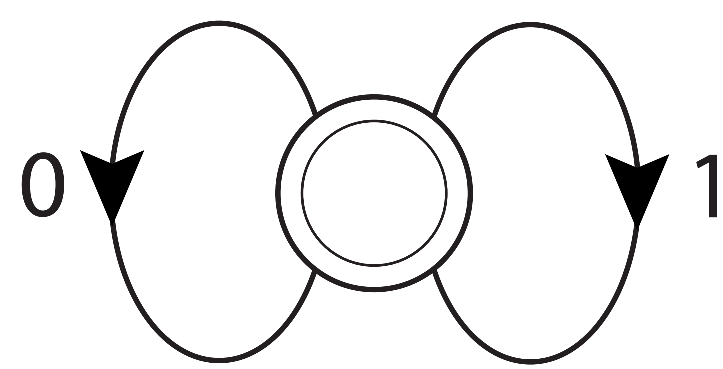
Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

$$0101011101001101\ldots$$

Your model is?



The Learning Channel ...

The Prediction Game ...

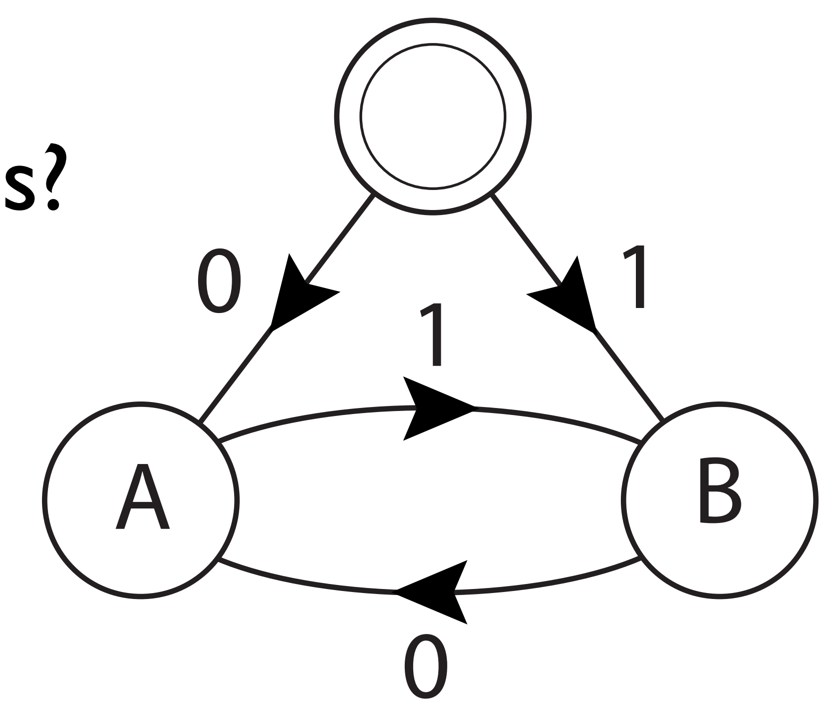
Process III:

Past: $\ldots 1010101010101010$

Your prediction is?

Future: $1010101010101010\ldots$

Your model is?



The Learning Channel ...

Theory? Algorithms?

Computational Mechanics

The Learning Channel ...

Goal:

Predict the future $\overrightarrow{S}$

using information from the past $\overleftarrow{S}$

But what "information" to use?

We want to find the effective "states"

and the dynamic (state-to-state mapping)

How to define "states", if they are hidden?

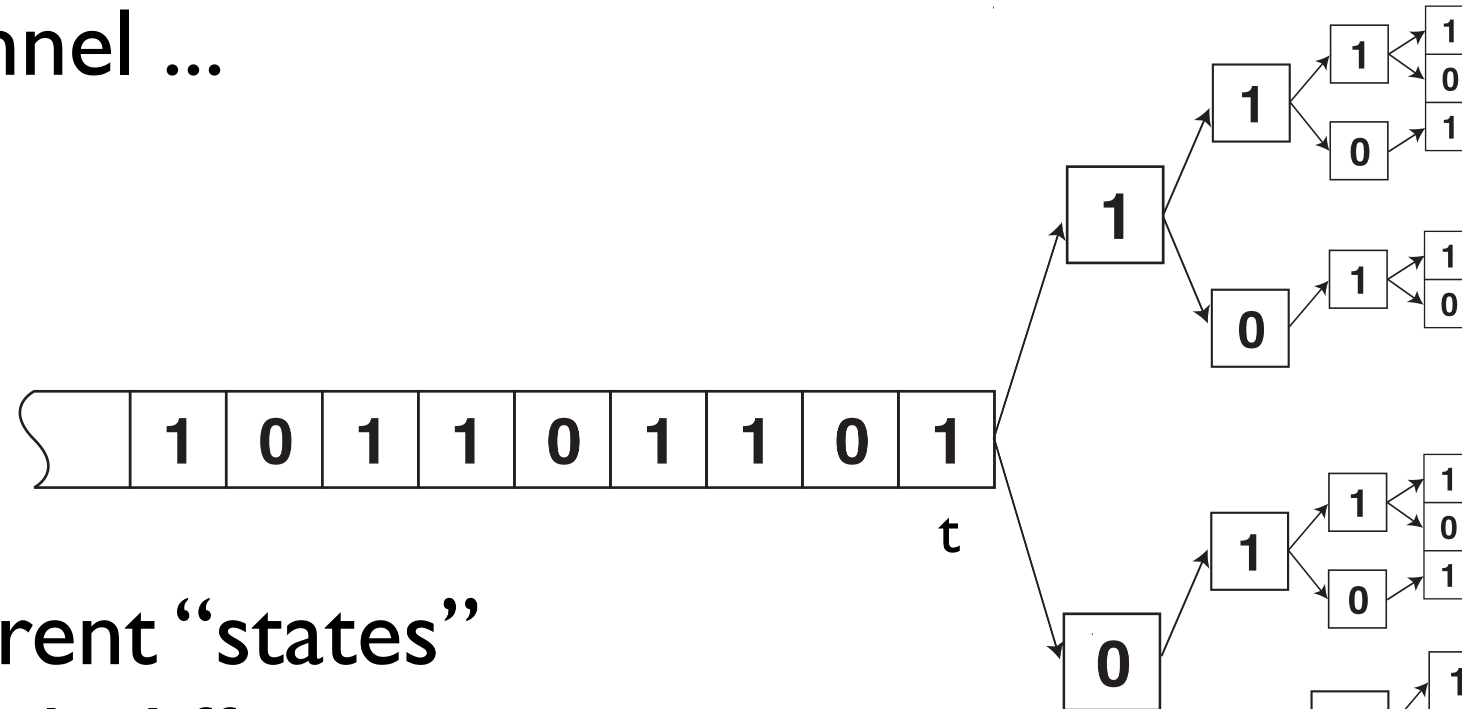
All we have are sequences of observations

Over some measurement alphabet $\mathcal{A}$

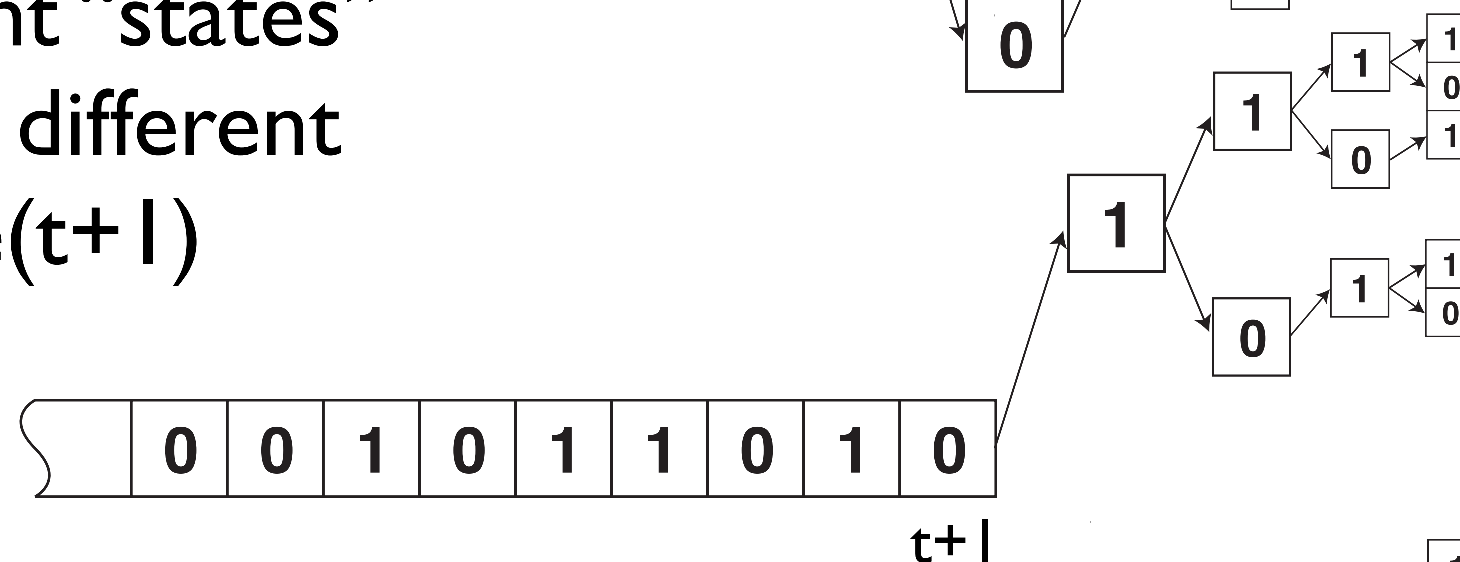
These symbols only indirectly reflect the hidden states

The Learning Channel ...

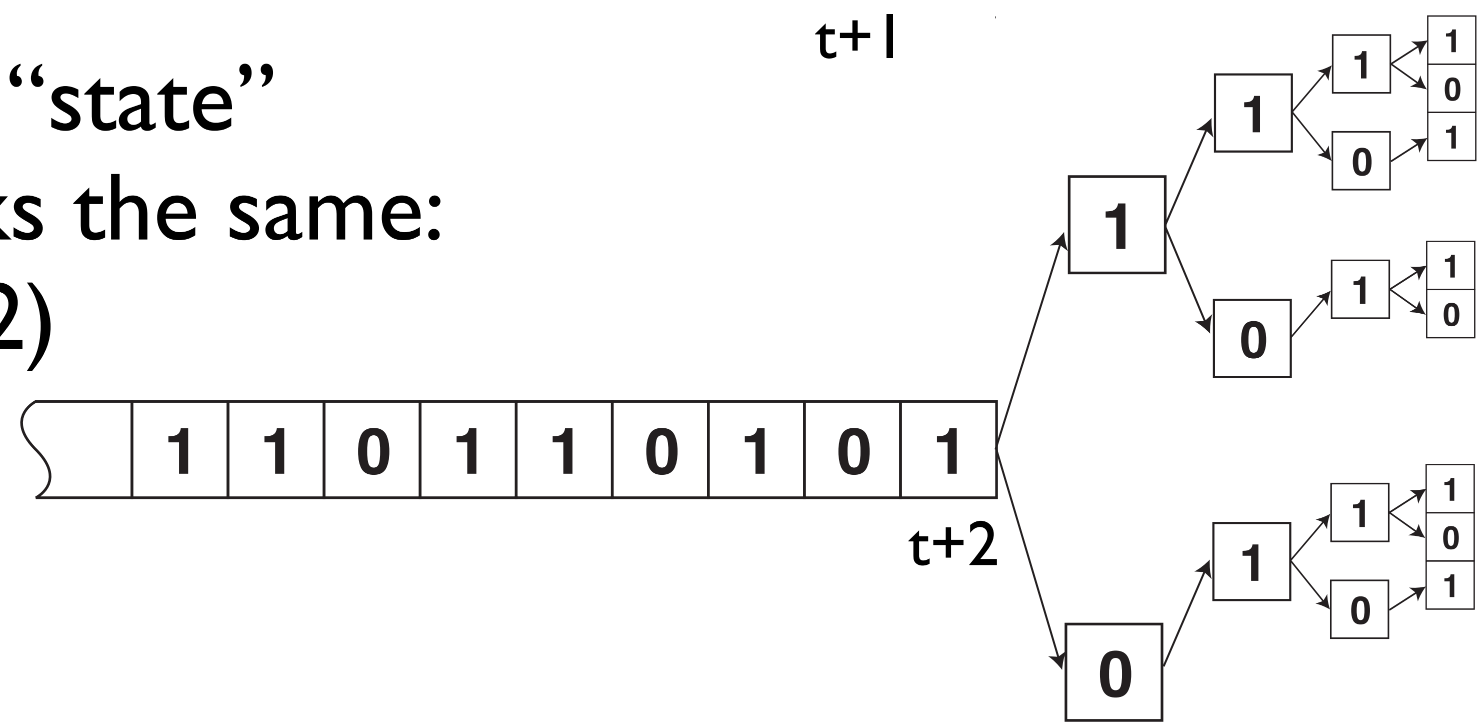
Effective States:



Process is in different "states"
when futures look different
State(t) $\nsim$ State(t+1)



Process is in the same "state"
when the future looks the same:
State(t) ~ State(t+2)



Causal States:

Causal State:

Set of pasts with same morph $\Pr(\overrightarrow{S} \mid \overleftarrow{s})$.

Set of histories that lead to same predictions.

Predictive equivalence relation:

$$\overleftarrow{s}' \sim \overleftarrow{s}'' \iff \Pr(\overrightarrow{S} \mid \overleftarrow{S} = \overleftarrow{s}') = \Pr(\overrightarrow{S} \mid \overleftarrow{S} = \overleftarrow{s}'')$$

$$\overleftarrow{s}', \overleftarrow{s}'' \in \overleftarrow{\mathbf{S}}$$

The Learning Channel ...

Causal States ...

We've answered the first part of the modeling goal:

We have the effective states!

Now,

What is the dynamic?

Causal State Dynamic ...

Conditional transition probability:

$$
\begin{aligned}
T_{ij}^{(s)} &= \Pr(\mathcal{S}_j, s | \mathcal{S}_i) \\
&= \Pr\left(\mathcal{S} = \epsilon(\overleftarrow{s}\, s) | \mathcal{S} = \epsilon(\overleftarrow{s})\right)
\end{aligned}
$$

State-to-State Transitions:

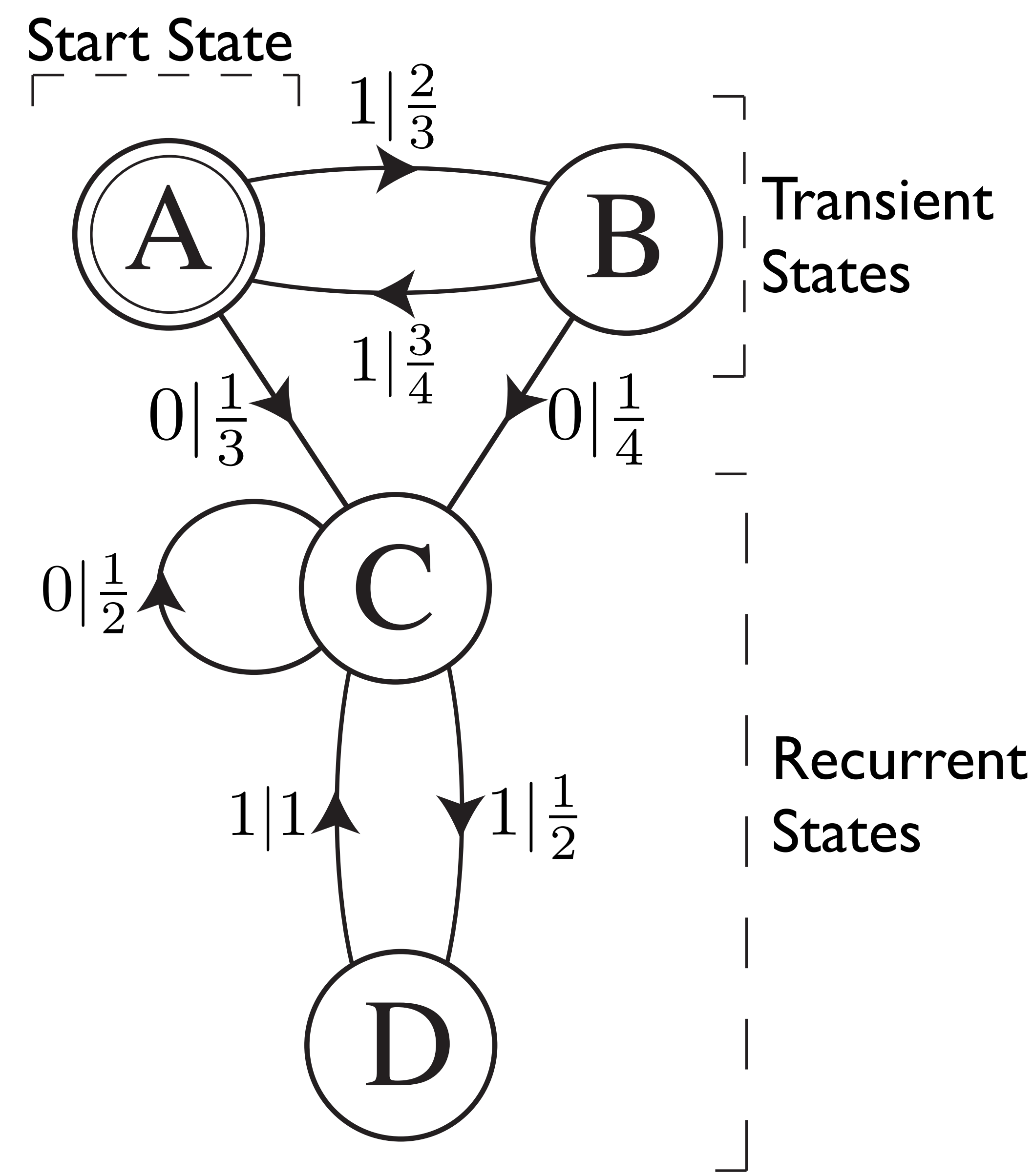
$$
\{T_{ij}^{(s)} : s \in \mathcal{A},\ i, j = 0, 1, \ldots, |\boldsymbol{\mathcal{S}}|\}
$$

The ϵ-Machine ...

$$\text{Process} \;\Rightarrow\; \text{Predictive equivalence} \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\Pr(\overleftrightarrow{S}) \;\Rightarrow\; \overleftarrow{\mathbf{S}} \,/\sim \;\Rightarrow\; \epsilon - \text{Machine}$$

$$\mathcal{M} = \left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$



Unique Start State:

$$\mathcal{S}_0 = [\lambda]$$

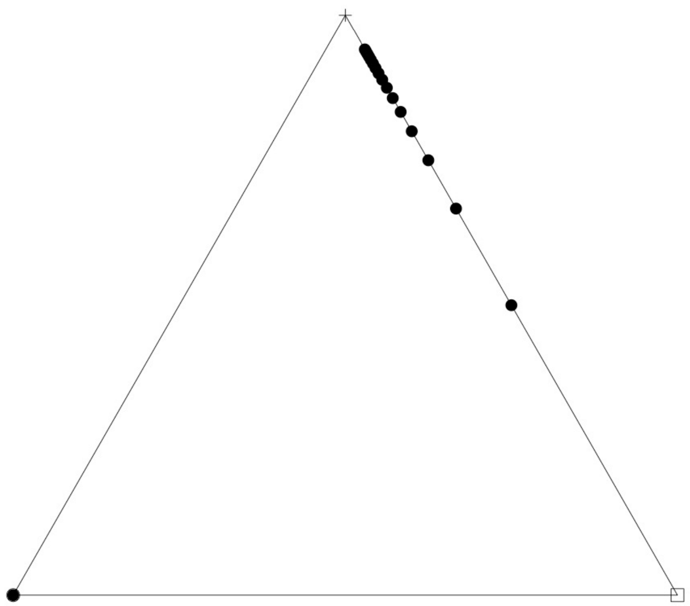
$$\Pr(\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots) = (1, 0, 0, \ldots)$$

Transient States

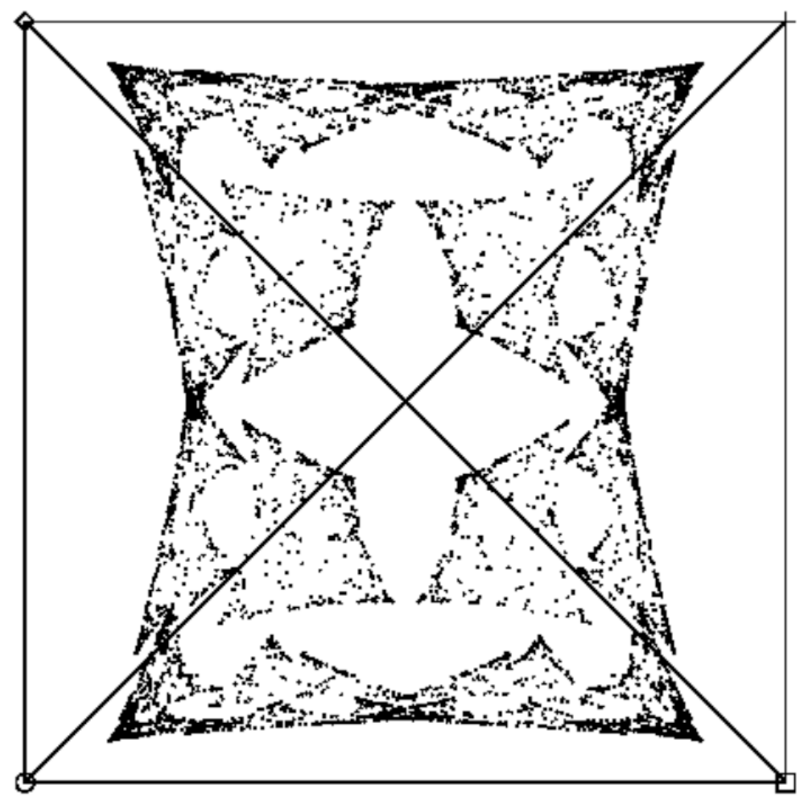
Recurrent States

The Learning Channel ...

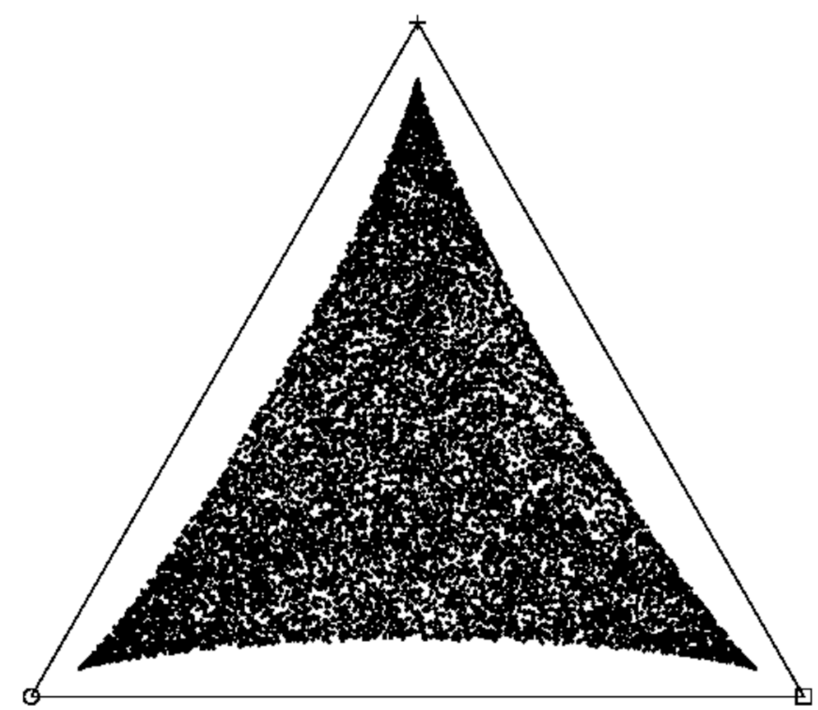
The ϵ-Machine of a Process ...



Denumerable
Causal States



Fractal



Continuous

The ϵ-Machine ...

Summary:

$\epsilon\mathrm{M}$:

(1) Optimal predictor: Lower prediction error than any rival.

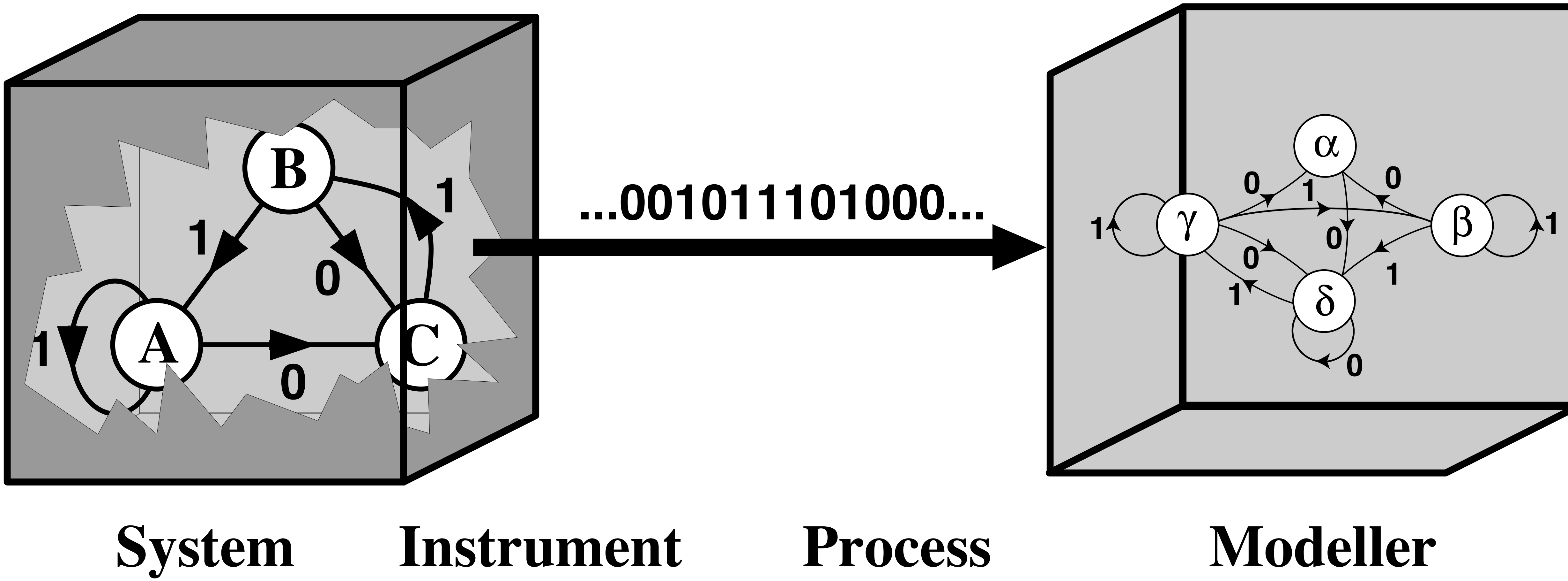
(2) Minimal size: Smallest of the prescient rivals.

(3) Unique: Smallest, optimal, unifilar predictor is equivalent.

(4) Model of the process: Reproduces all of process's statistics.

(5) Causal shielding: Renders process's future independent of past.

The Learning Channel:



Central questions:

What are the states? Causal States

What is the dynamic? The ϵ-Machine

Measures of Intrinsic Computation

Measures of Intrinsic Computation ...

A complex process's intrinsic computation:

(1) How much of past does process store?

$$C_\mu$$

(2) In what architecture is that information stored?

$$\left\{ \boldsymbol{\mathcal{S}}, \{T^{(s)}, s \in \mathcal{A}\} \right\}$$

(3) How is stored information used to produce future behavior?

$$h_\mu$$

Measures of Intrinsic Computation ...

Measures of Structural Complexity:

Information Measures		Interpretation
Entropy Rate	h_μ	Intrinsic Randomness
Excess Entropy	$\mathbf{E}$	Info: Past to Future
Total Predictability	$\mathbf{G}$	Redundancy
Transient Information	$\mathbf{T}$	Synchronization

How related to statistical complexity C_μ?

How to get from ϵM?

Measures of Intrinsic Computation ...

Measures from the $\epsilon\mathrm{M}$:

Entropy Rate of a Process:

$$h_\mu(\Pr(\overleftrightarrow{S})) = \lim_{L \to \infty} \frac{H(L)}{L}$$

Entropy Rate given $\epsilon\mathrm{M}$:

$$h_\mu(\boldsymbol{\mathcal{S}}) = - \sum_{\mathcal{S} \in \boldsymbol{\mathcal{S}}} \Pr(\mathcal{S}) \sum_{s \in \mathcal{A}, \mathcal{S}' \in \boldsymbol{\mathcal{S}}} T^{(s)}_{\mathcal{S}\mathcal{S}'} \log_2 T^{(s)}_{\mathcal{S}\mathcal{S}'}$$

where $\Pr(\mathcal{S})$ is casual-state asymptotic probability.

Possible only due to $\epsilon\mathrm{M}$ unifilarity!

1-1 mapping between measurement sequences & internal paths.

Measures of Intrinsic Computation ...

Measures from the $\epsilon\mathrm{M}$...

Statistical Complexity of a Process:

$$C_\mu(\boldsymbol{\mathcal{S}}) = -\sum_{\mathcal{S} \in \boldsymbol{\mathcal{S}}} \Pr(\mathcal{S}) \log_2 \Pr(\mathcal{S})$$

where $\Pr(\mathcal{S})$ is casual-state asymptotic probability.

Meaning:

Shannon information in the causal states.

The amount of historical information a process stores.

The amount of structure in a process.

Measures of Intrinsic Computation ...

Measures from the $\epsilon\mathrm{M}$...

Excess Entropy: Three versions, all equivalent for 1D processes

$$\mathbf{E} = \lim_{L \to \infty} [H(L) - h_\mu L]$$

$$\mathbf{E} = \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu]$$

$$\mathbf{E} = I[\overleftarrow{S}; \overrightarrow{S}]$$

How to get, given $\epsilon\mathrm{M}$?

Special cases: When $\epsilon\mathrm{M}$ is IID, periodic, or spin chain.

General case: Need a new framework.

Measures of Intrinsic Computation ...

Measures from the $\epsilon\mathrm{M}$ ...

Bound on Excess Entropy: $\mathbf{E} \leq C_\mu$

Executive Summary:

C_μ is the amount of information the process uses

to *communicate*

$\mathbf{E}$ bits of information from the past to the future.

Measures of Intrinsic Computation ...

Measures from the $\epsilon\mathrm{M}$ ...

Bound on Excess Entropy: $\mathbf{E} \leq C_\mu$

Consequence:

The inequality is Why We Must Model.

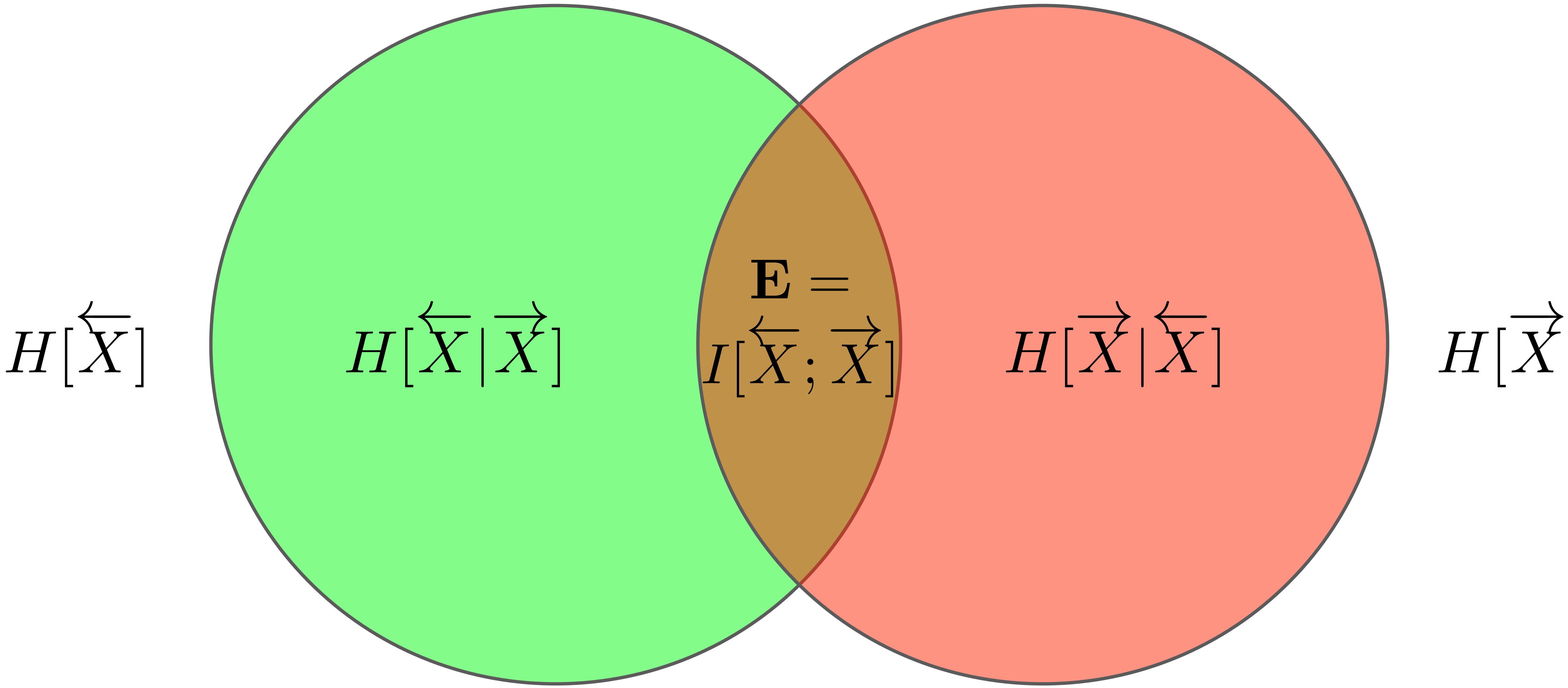
Cannot simply use sequences as states.

There is internal structure not expressed by this.

Information Diagrams for Processes

Information Diagrams for Processes

Process I-diagram:



Information Diagrams for Processes

Process I-diagram using ε-machine:

Start with 3-variable I-diagram and whittle down:

- Past as composite random variable: $\overleftarrow{X}$
- Future as composite random variable: $\overrightarrow{X}$
- Causal states: $\mathcal{S} \in \boldsymbol{\mathcal{S}}$

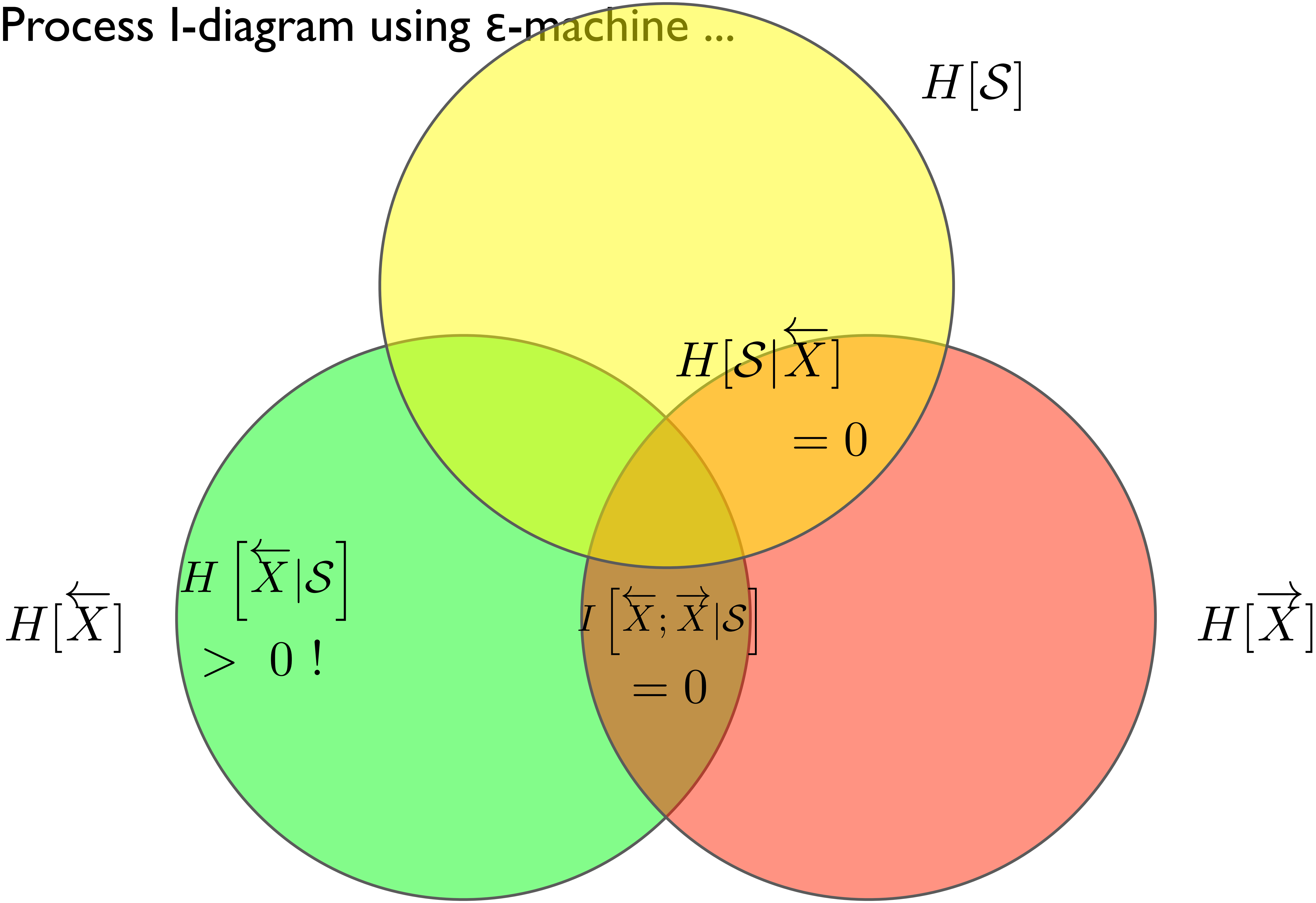
Information measures:

$$H[\overleftarrow{X}] \quad H[\overrightarrow{X}] \quad H[\mathcal{S}] \quad \cdots \quad I[\overrightarrow{X};\overleftarrow{X};\mathcal{S}] \quad \cdots \quad H[\overrightarrow{X},\overleftarrow{X},\mathcal{S}]$$

There are $8 = 2^3$ atomic information measures.

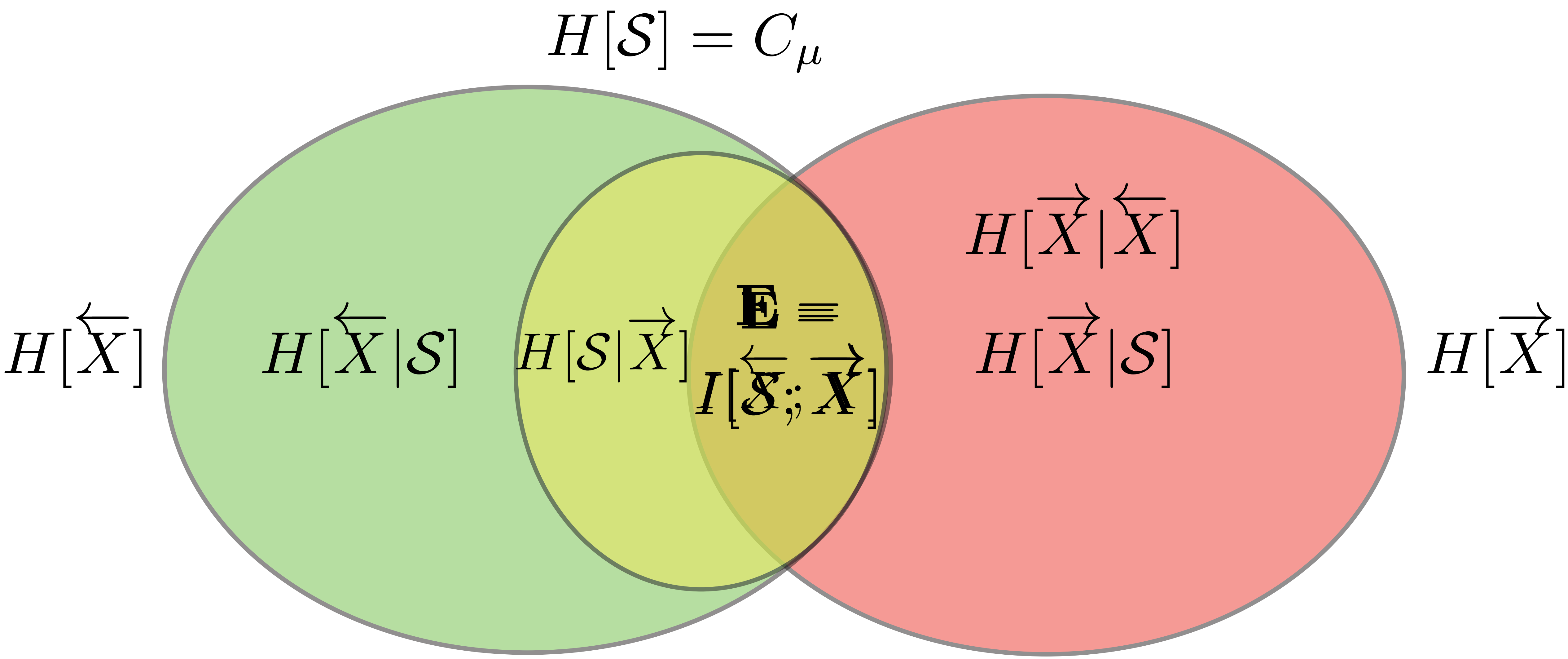
Information Diagrams for Processes

Process I-diagram using ε-machine ...



Information Diagrams for Processes

ε-Machine I-diagram:

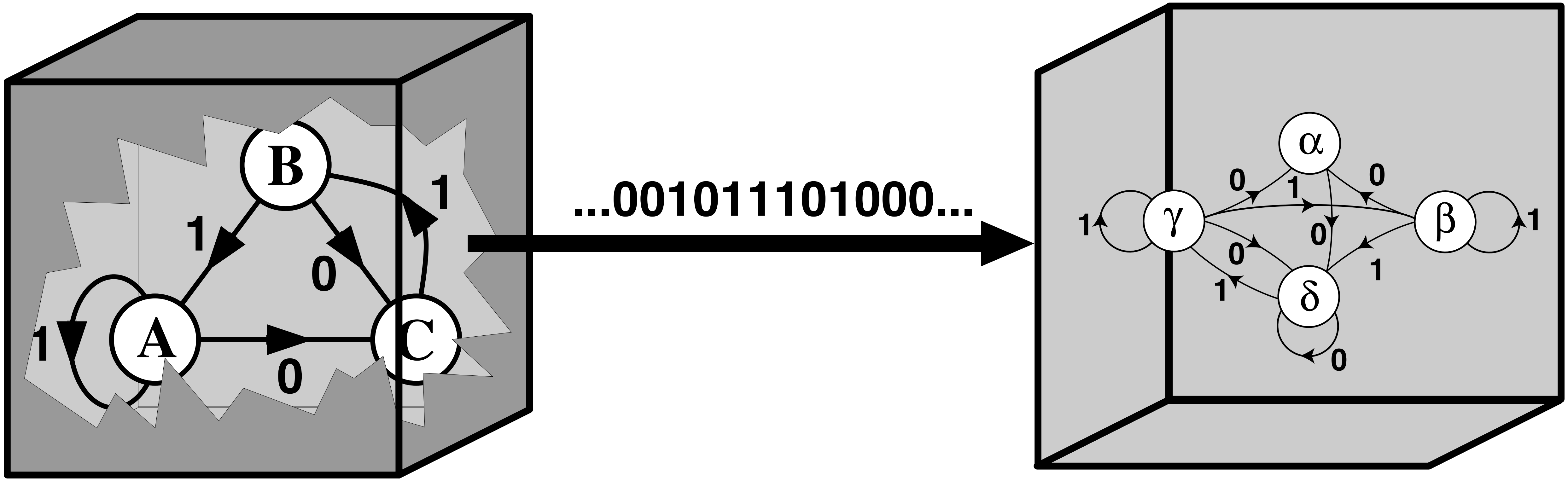


What is this Mystery Wedge?

What do we know about this?

Intrinsic Computation ...

Analysis narrative:



System **Instrument** **Process** **Modeller**

Forms of Chaos:
- Deterministic sources of novelty
- Mechanisms that produce unpredictability
- Sensitive dependence on initial condition
- Sensitive dependence on parameter

Measurement Theory:
- Partitions
- Optimal Instrument:

$$\max_{\{\mathcal{P}\}} h_\mu$$

$$\min_{\{\mathcal{P}\}} C_\mu$$

How random?

$$\lambda, H(L), h_\mu$$

How structured?

$$C_\mu, \mathbf{E}, \mathbf{T}, \mathbf{G}, \mathcal{R}$$

Universal model:
- $\epsilon - \mathrm{Machine}$
- Pattern defined
- Causal Architecture
- Intrinsic Computation

Intrinsic Computation ...

A system is unpredictable
if it has positive entropy rate: $h_\mu > 0$

A system is complex
if it has positive structural complexity measures: $C_\mu > 0$

A system is emergent
if its structural complexity measures increase over time:
$$C_\mu(t') > C_\mu(t), \text{ if } t' > t$$

A system is hidden
if its crypticity is positive: $\chi > 0$

Algorithmic Basis of Information ...

Kolmogorov-Chaitin Complexity versus Statistical Complexity

KC Complexity versus Statistical Complexity

We saw that:
KC complexity of typical realizations from an information source grows proportional to the Shannon entropy rate:

$$K(x) \propto h_\mu |x|$$

Thus, KC complexity is a measure of randomness.

KC Complexity versus Statistical Complexity

What's the relationship to Statistical Complexity?

Since randomness drives Kolmogorov-Chaitin complexity, let's discount for generating randomness:

Programs consist of model m and data d (random part unexplained by m).

Sophistication of object:

$$S_k(x) = \min\{|m| : p = m + d \text{ and } |p| - K(x) \leq k\}$$

Also, uncomputable.

KC Complexity versus Statistical Complexity

Consider the average sophistication:

$$S(\ell) = \langle S_0(x_{0:\ell}) \rangle$$

It is statistical complexity:

$$C_\mu \propto_{\ell \gg 1} S(\ell)$$

Since program = model + data:

$$K(\ell) = S(\ell) + \langle |d| \rangle_{x_{0:\ell}}$$

We have:

$$K(\ell) \approx C_\mu + h_\mu \ell$$

Since a process has a structure, as ℓ gets large,
with probability 1 each possible $x_{0:\ell}$ has the same model.

KC Complexity versus Statistical Complexity

Recall the Block Entropy

$$H(\ell) \approx C_\mu + h_\mu \ell$$

Similar scaling.

$K(\ell)$ versus $H(\ell)$:

$\mathbf{E}$ quantifies the amount of information observed as ℓ gets large, whereas C_μ quantifies how much information it takes to predict as ℓ gets large.

KC Complexity versus Statistical Complexity

Kolmogorov-Chaitin Theory versus Computational Mechanics

First, ε−machine describes distribution over a system's behaviors, including individual realizations.

Second, one can exactly calculate the Shannon entropy rate for a system's behaviors.

Third, the computational model is a probabilistic UTM: a Bernoulli-Turing Machine.

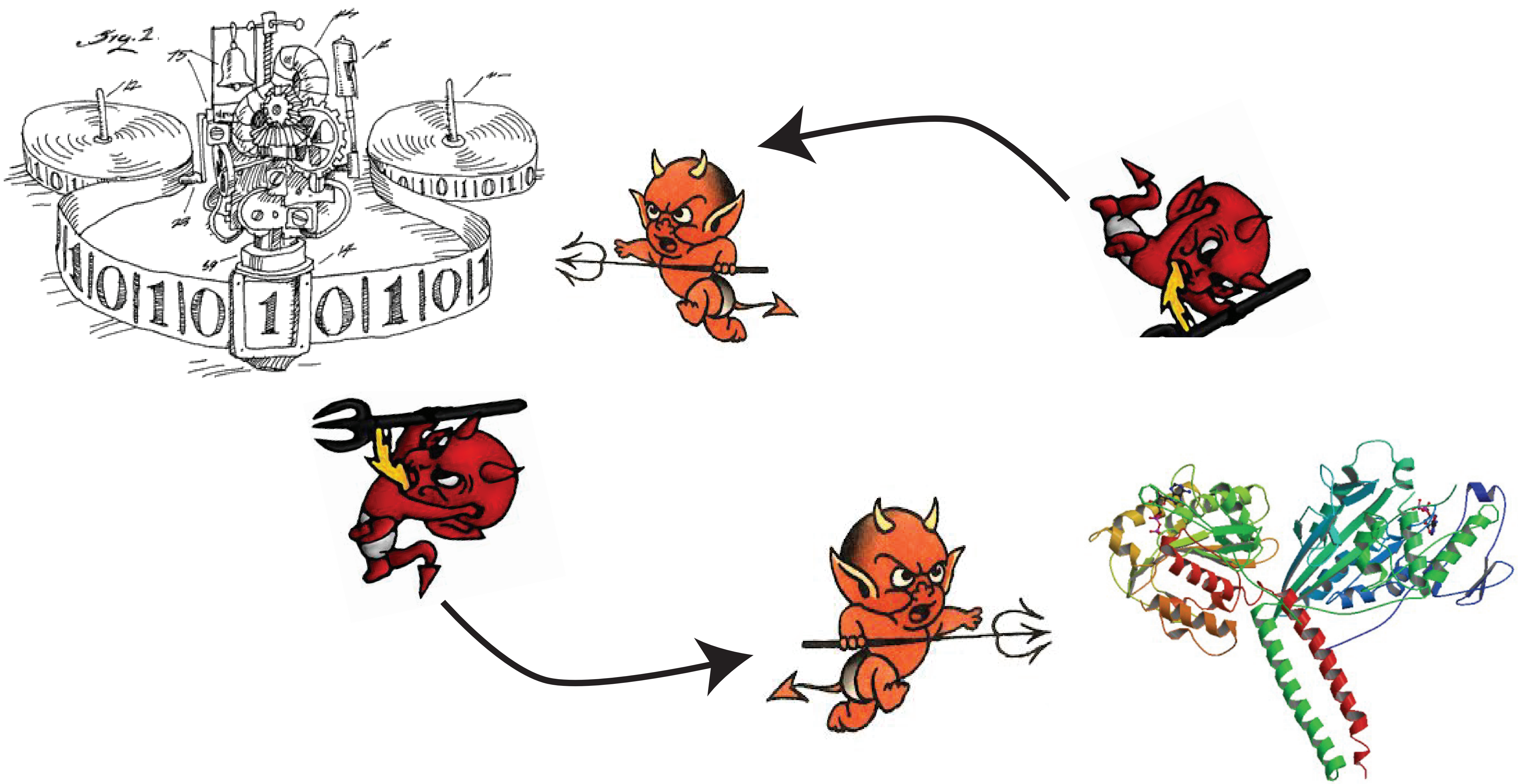
KC Complexity versus Statistical Complexity

Computational Mechanics was introduced to be a calculable, quantitative version of KC Complexity Theory.

Constructive! For finite eMs, all complexity/information measures
- can be calculated in closed form.
- $O(1)$ computational complexity.

So, much computational complexity in KC Theory and in Information Theory obviated.

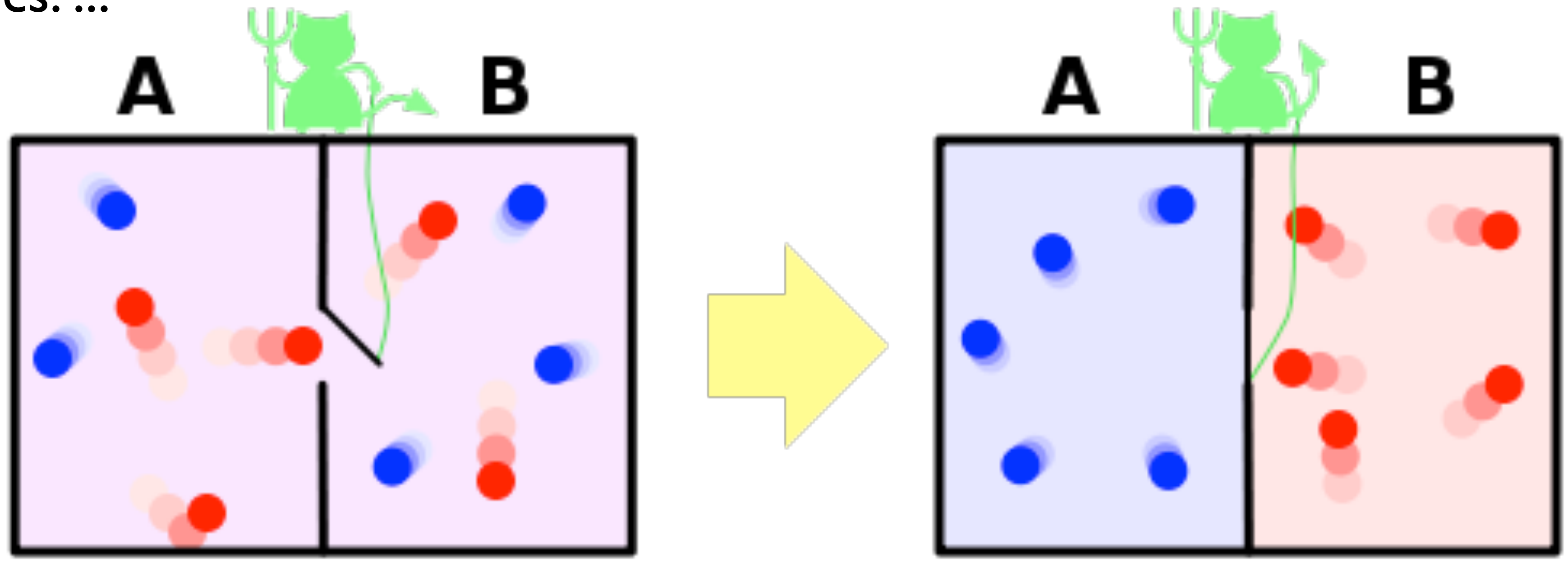
Thermodynamics of Adaptive Complex Systems: An Application



Maxwell's Demon

James Clerk &
Katherine Maxwell
(1865)

... if we conceive of a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are as essentially finite as our own, would be able to do what is impossible to us. ... Now let us suppose that ... a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower molecules to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics. ...



Maxwell's Demon

- Demon creates order out of chaos.

- Uses molecular information to convert heat to temperature difference & so to useful work.

Szilard's Engine:

"ON THE DECREASE OF ENTROPY IN A THERMODYNAMIC SYSTEM BY THE INTERVENTION OF INTELLIGENT BEINGS",
Leo Szilard, Zeitschrift fur Physik **65** (1929) 840-866.

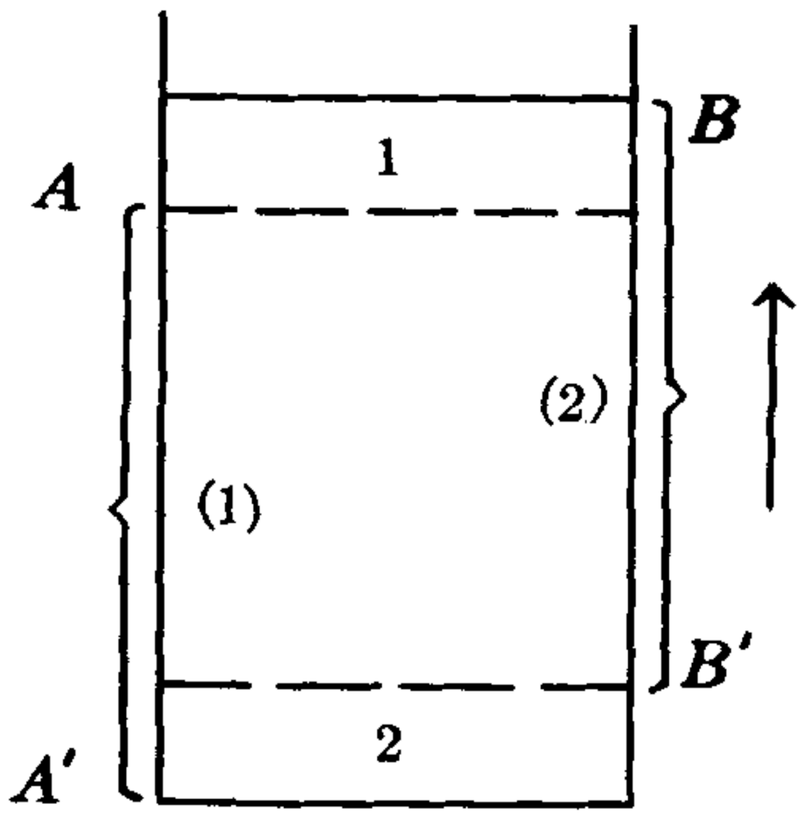


FIG. 1

... we must conclude *that the intervention which establishes the coupling between y and x, the measurement of x by y, must be accompanied by a production of entropy.*

... a simple inanimate device can achieve the same essential result as would be achieved by the intervention of intelligent beings. We have examined the "biological phenomena" of a nonliving device and have seen that it generates exactly that quantity of entropy which is required by thermodynamics.

A Technological Interlude or Why your Lap(top) is Hot!

Landauer Principle: $Q_{\text{erase}} \geq k_{\text{Boltzmann}} T \ln 2$

Erasing a bit dissipates energy to the heat bath.

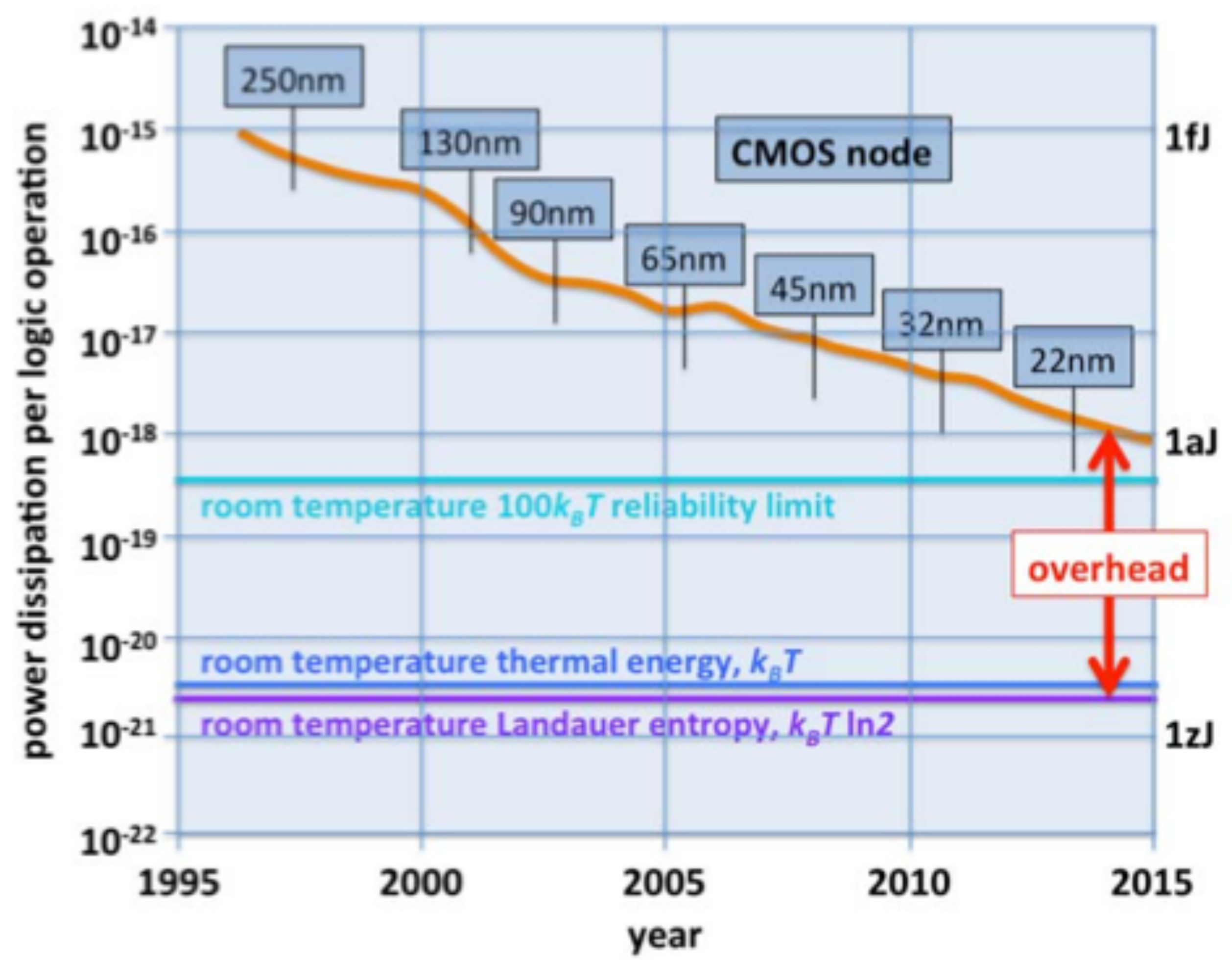
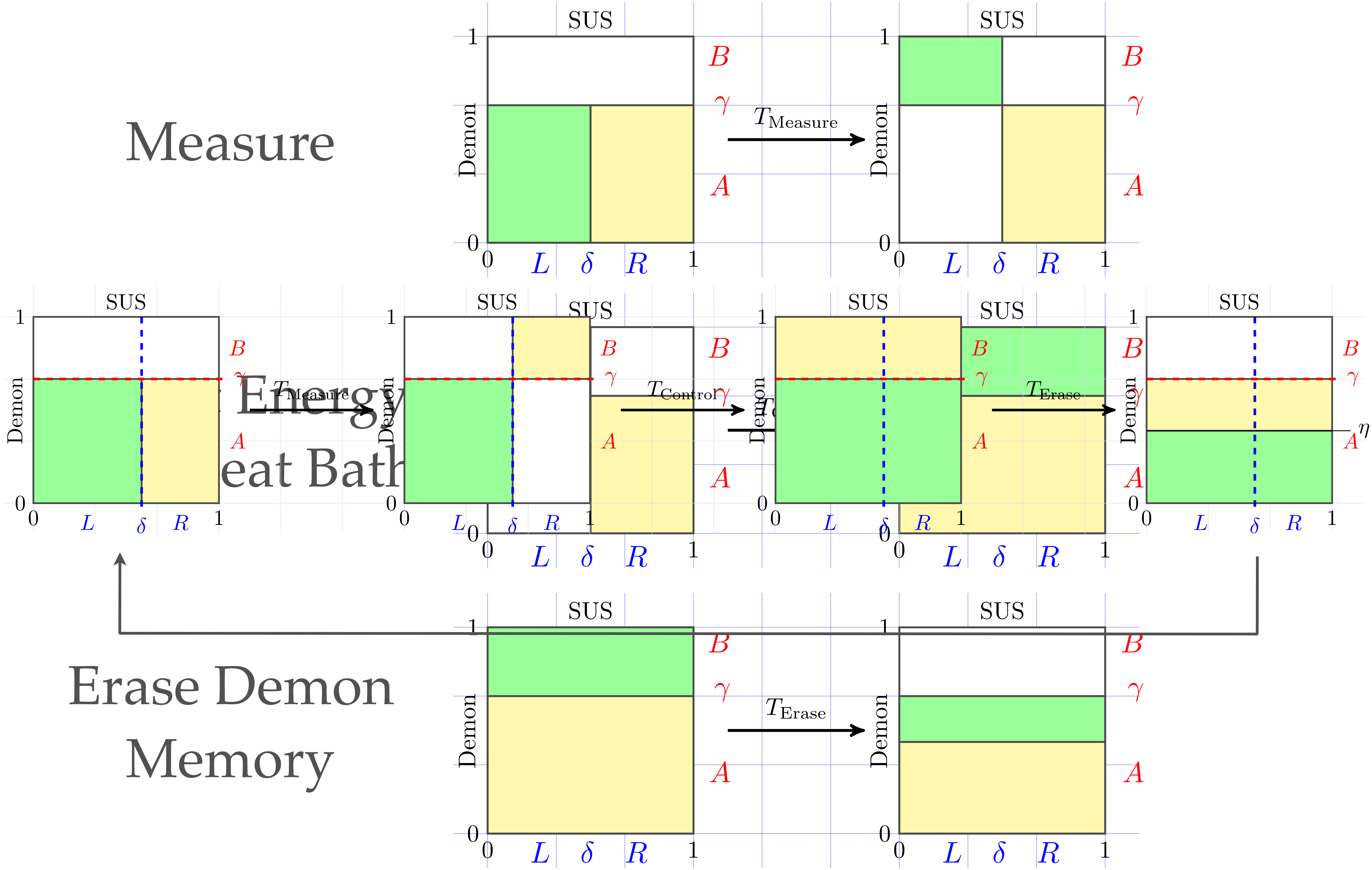
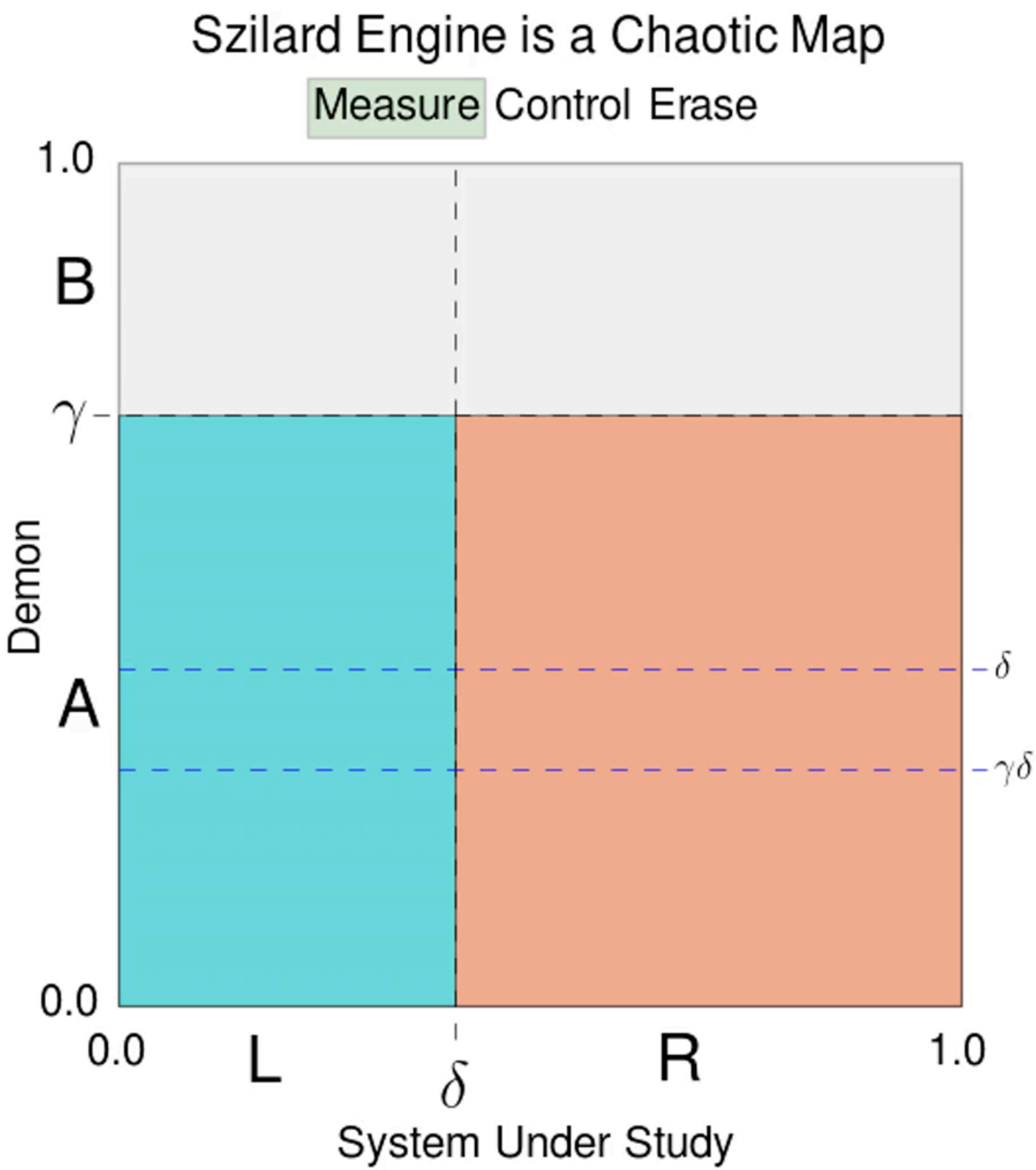


Figure 1. Evolution over time of the energy per CMOS logic operation. Energy overhead (see text) for the latest node is shown by the red arrow.

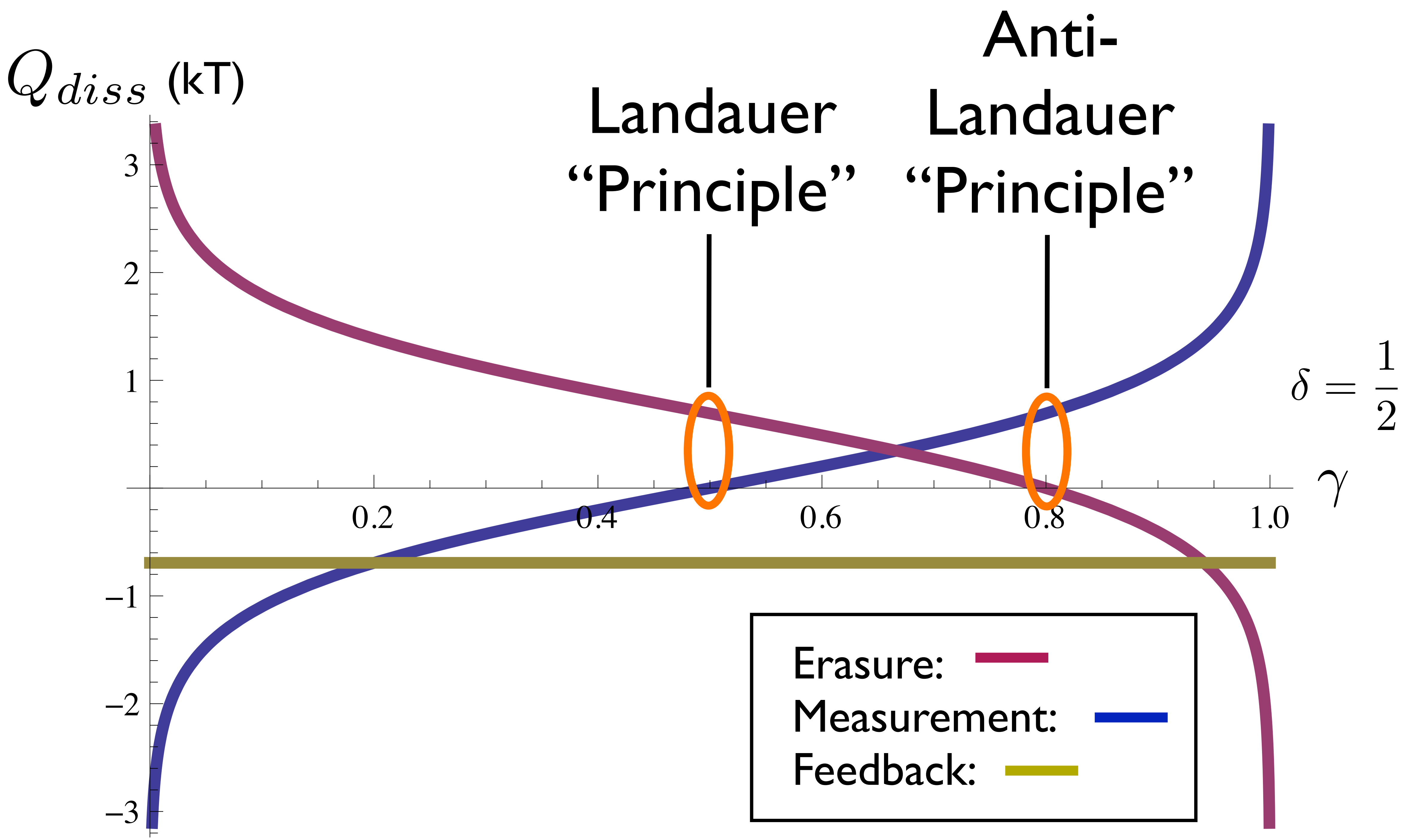
Szilard's Engine is a Chaotic Dynamical System



A. B. Boyd and J. P. Crutchfield, *Demon Dynamics: Deterministic Chaos, the Szilard Map, and the Intelligence of Thermodynamic Systems*, **Physical Review Letters** 116 (2016) 190601.


Szilard Engine is a Chaotic Map
Measure Control Erase
1.0
B
γ
Demon
A
δ
$\gamma\delta$
0.0
0.0
L
δ
R
1.0
System Under Study

Beyond Landauer:
Energy Dissipation during Erasure and Measurement!



Information in Complex Systems

Done:

Algorithmic Basis of Probability

Information Theory

Information Measures

Done:

Measuring Structure

Intrinsic Computation

Optimal Models

Physics of Information

See online course:

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