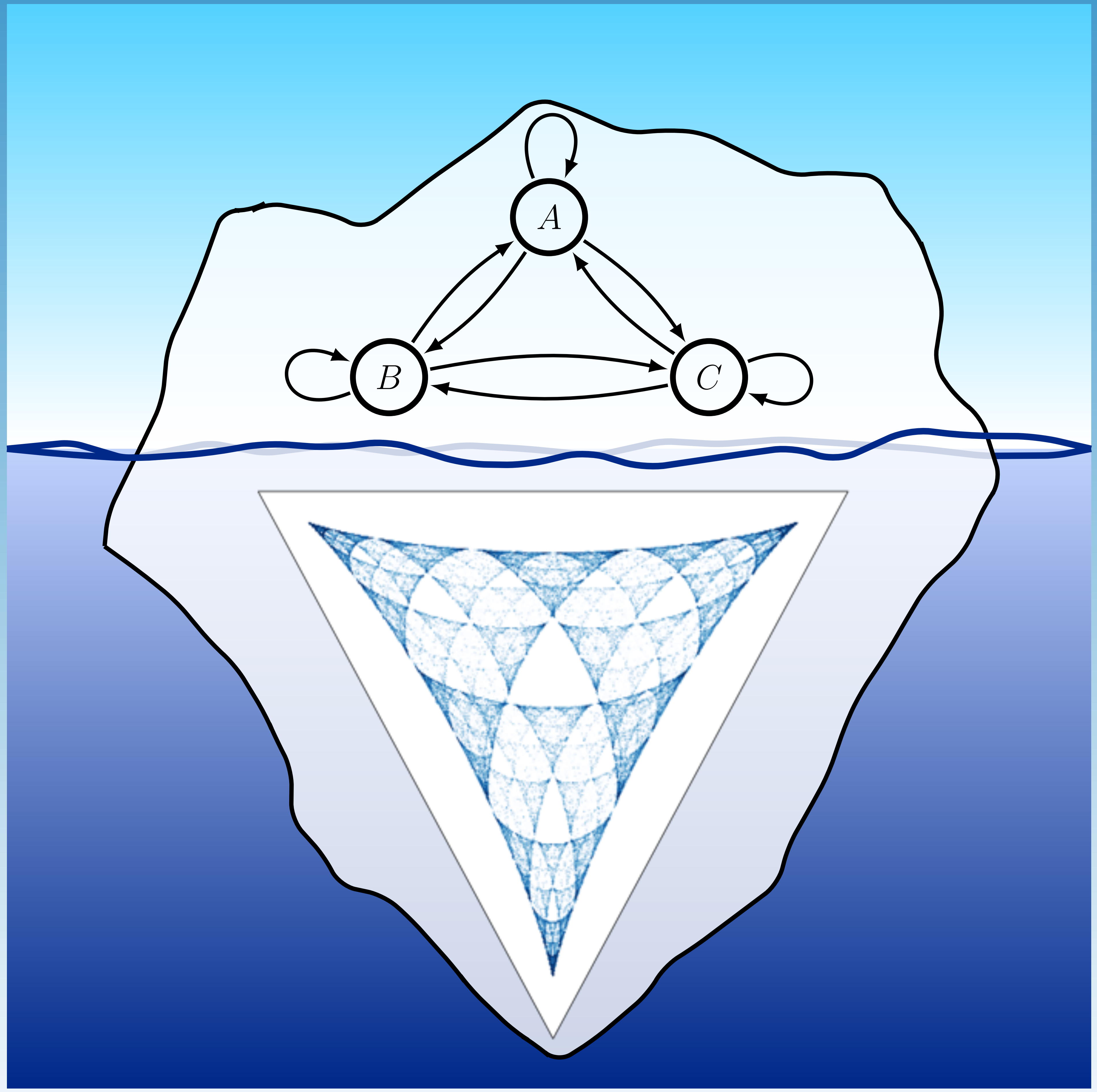
PHYS 256: POCI

Physics of Information & Computation

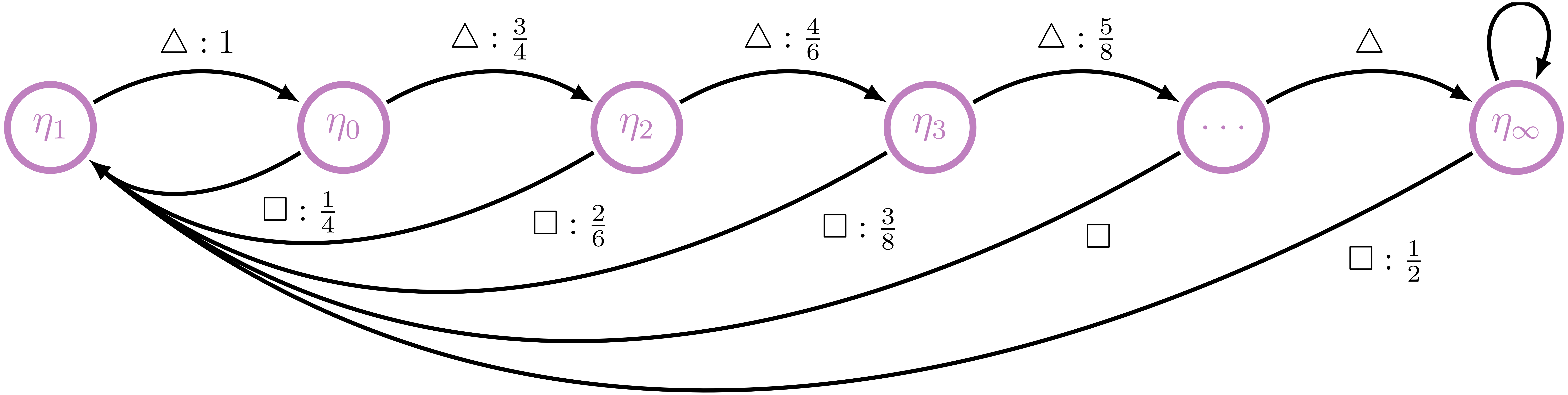
Alexandra Jurgens
Inria Centre at the University of Bordeaux
05/20/2025

Infinite State Processes II



Alexandra Jurgens
Inria Centre at the University of Bordeaux
05/20/2025

Blackwell Entropy Formula



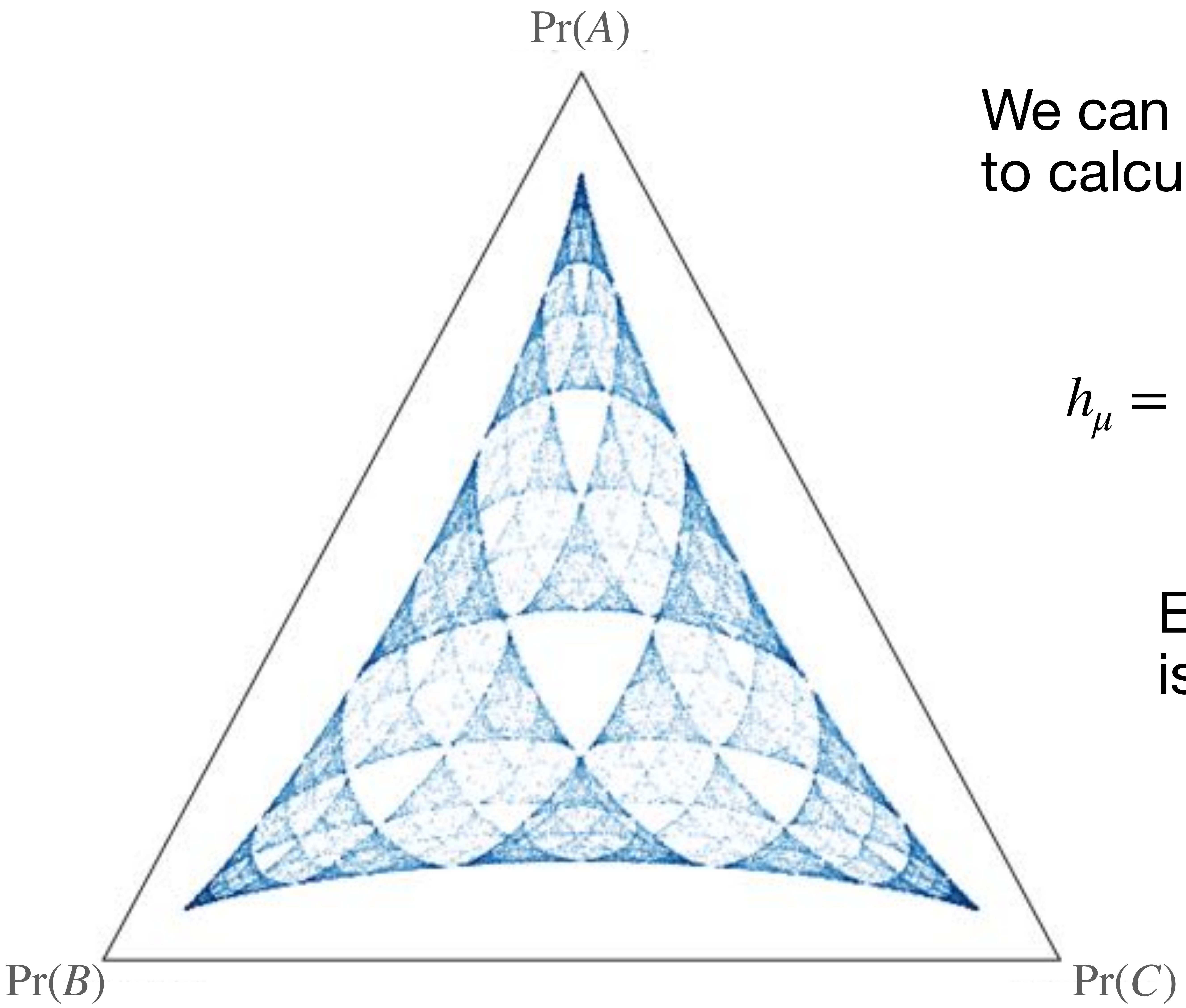
$$h_\mu = \int_R d\mu\left(\eta\right)\ H[x\,|\,\eta] = \lim_{N\to\infty} - \sum_n^N \Pr\left(\eta_n\right) \sum_{x\in A} \Pr\left(x\,|\,\eta_n\right) \log_2 \Pr\left(x\,|\,\eta_n\right)$$

$$\longrightarrow\ h_\mu = 0.67$$

Blackwell, D. (1957) The entropy of functions of finite-state Markov chains.

Jurgens, A. M., & Crutchfield, J. P. (2021) Shannon entropy rate of hidden Markov processes. *Journal of Statistical Physics*, *183*(2), 32.

Mixed State Sets are Fractals

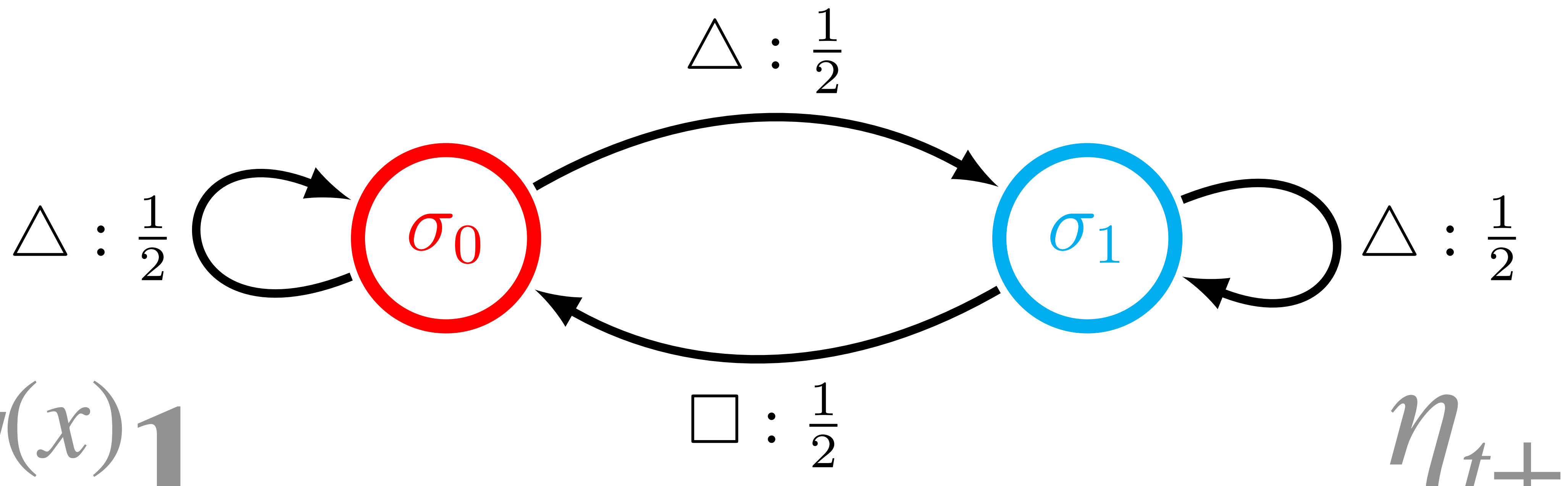


We can still use the Blackwell formula to calculate entropy:

$$h_\mu = \int_R d\mu(\eta)\, H[x \,|\, \eta]$$

Except finding the Blackwell measure is much more difficult.

Mixed State Generation → IFS

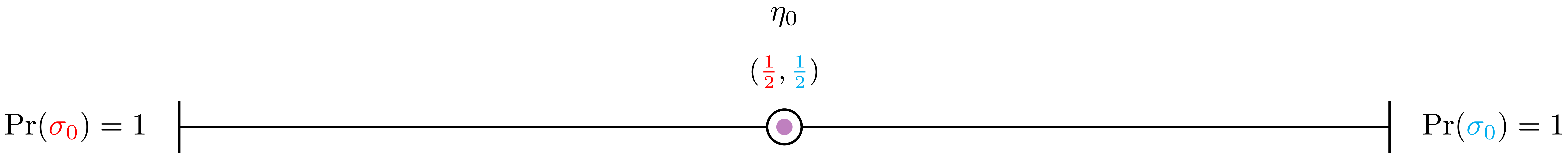


$$\Pr(x \mid \eta_t) = \eta_t T^{(x)} \mathbf{1}$$

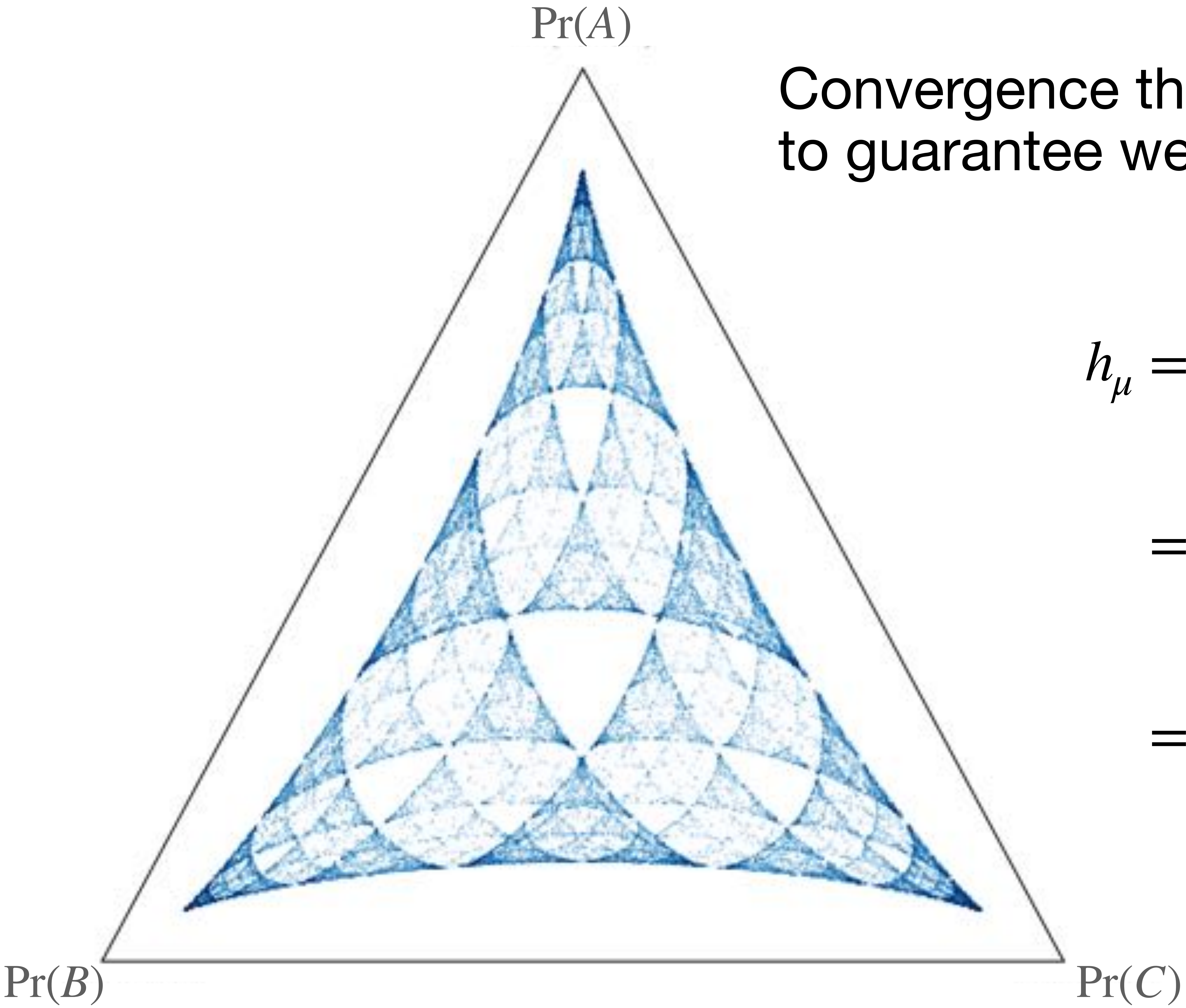
$$\eta_{t+1} = \frac{\eta_t T^{(x)}}{\Pr(x \mid \eta_t)}$$

$$p^x(\eta) = \eta T^{(x)} \mathbf{1}$$

$$f^x(\eta) = \frac{\eta T^{(x)}}{p_x(\eta)}$$



Entropy Rate of Fractal HMMs



Convergence theorems about IFSs can now be applied to guarantee we can apply the ergodic theorem:

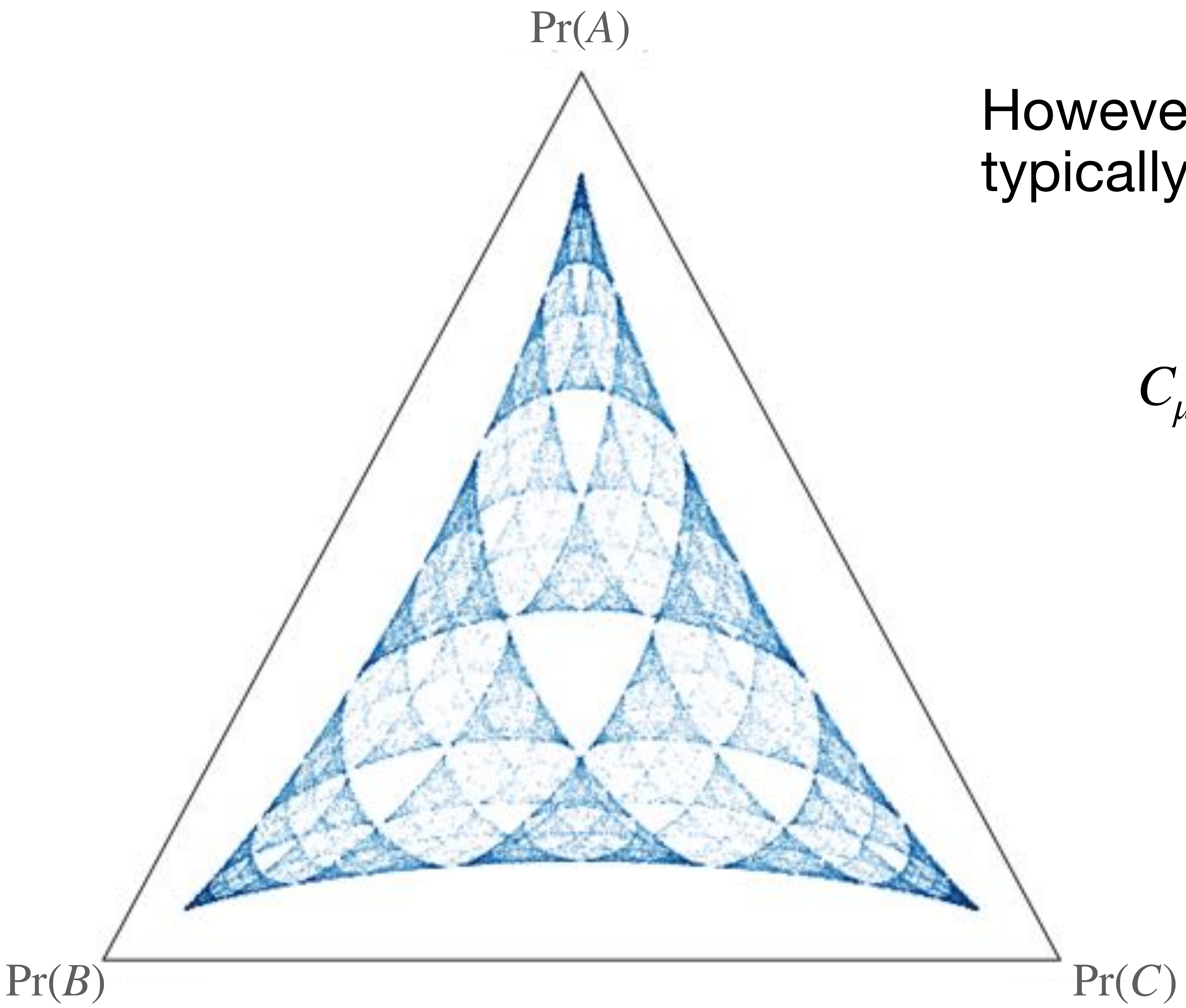
$$h_\mu = \int_R d\mu(\eta)\, H[x \mid \eta]$$

$$= \lim_{T\to\infty} -\frac{1}{T}\sum_{t}^{T} H\left[x \mid \eta_t\right]$$

$$= \lim_{T\to\infty} -\frac{1}{T}\sum_{t}^{T}\sum_{x} p^x\left(\eta_t\right)\log_2 p^x\left(\eta_t\right)$$

Jurgens, A. M., & Crutchfield, J. P. (2021) Shannon entropy rate of hidden Markov processes. *Journal of Statistical Physics*, *183*(2), 32.

Complexity of Fractal HMMs



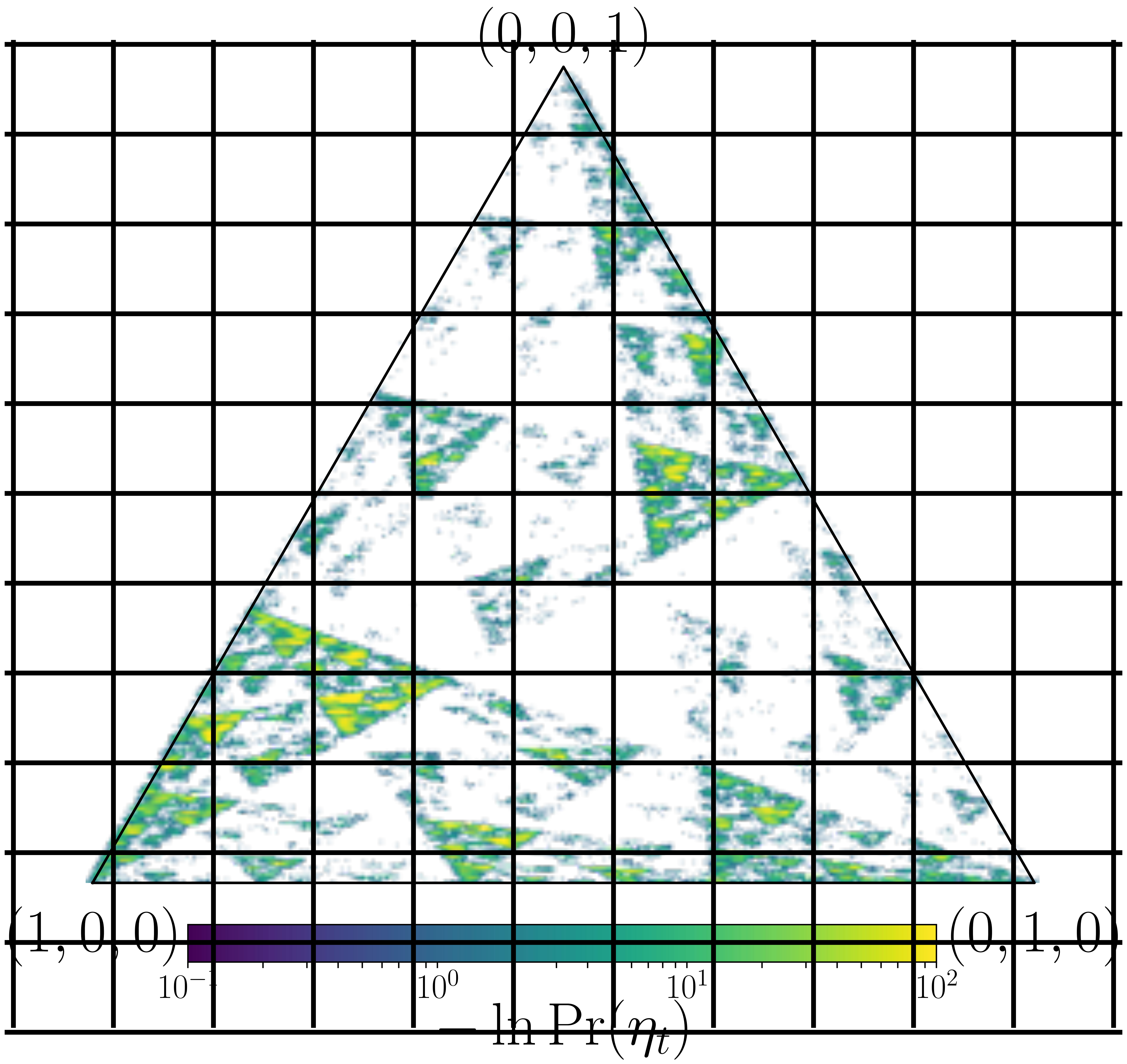
However, the statistical complexity still typically diverges.

$$C_\mu = -\int_R \Pr\left(\eta\right) \log_2 \Pr\left(\eta\right) d\mu\left(\eta\right)$$

Infinite complexity?

Jurgens, A. M., & Crutchfield, J. P. (2021). Divergent predictive states: The statistical complexity dimension of stationary, ergodic hidden Markov processes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *31*(8).

Statistical Complexity Dimension

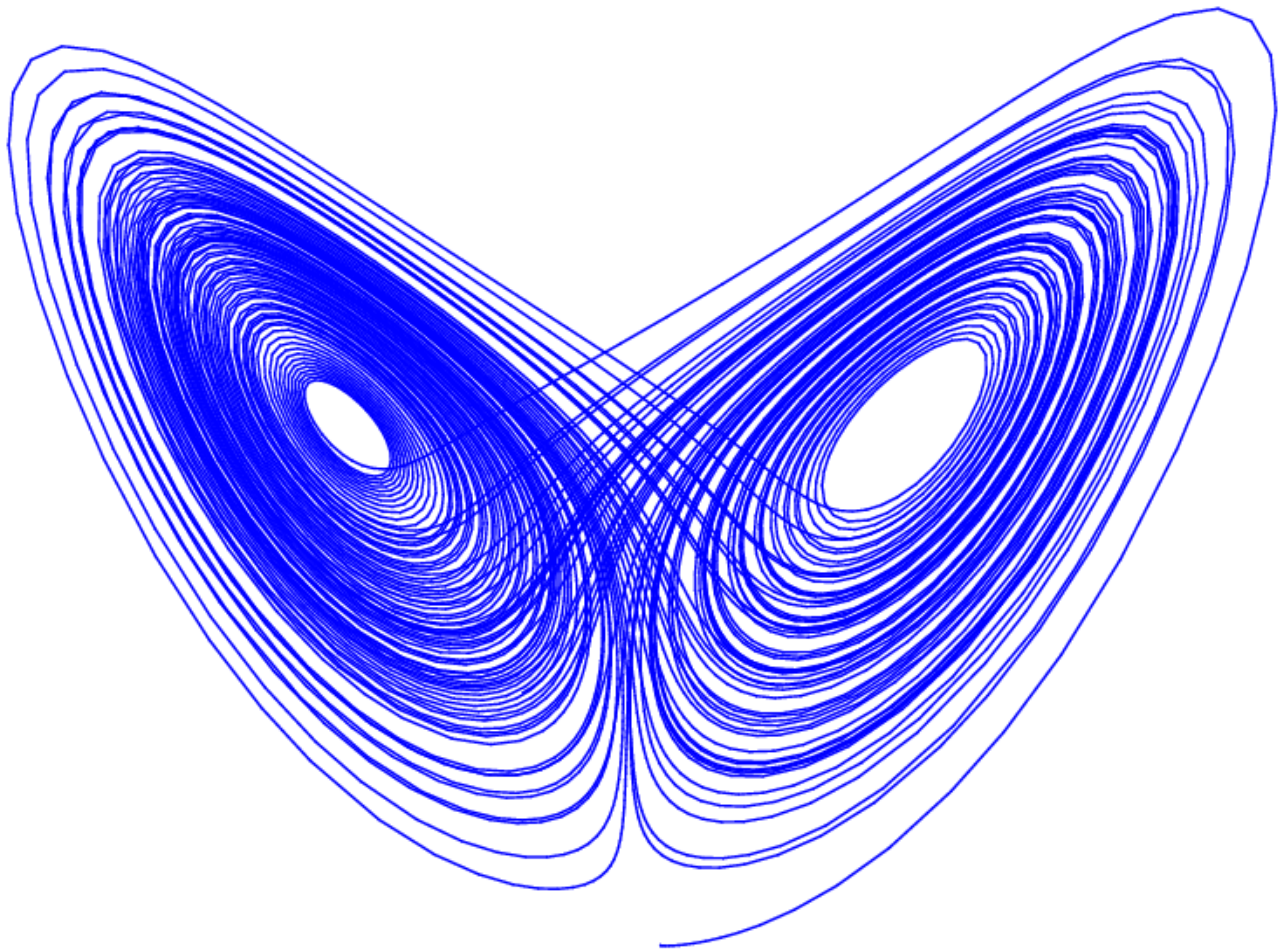


So, we define the *statistical complexity dimension* as the information dimension of the mixed state set:

$$d_\mu = \lim_{\epsilon \to 0} - \frac{H[R]}{\log(\epsilon)}$$

It gives the rate at which memory resources diverge as we increase predictive power of a finitized ϵ-machine.

Kaplan-Yorke Conjecture



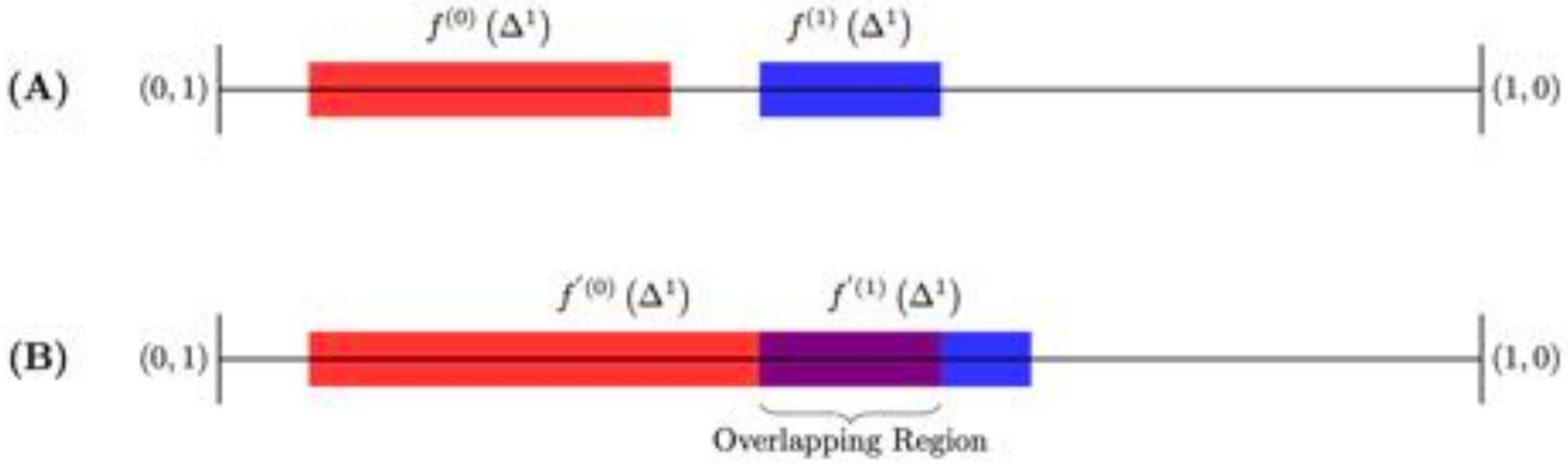
Kaplan—Yorke conjecture:

$$\dim_{KY} = k + \frac{\sum_i^k \lambda_i}{|\lambda_{k+1}|} \stackrel{?}{=} \dim_I$$

Where k is the largest index for which the sum $\sum_i^k \lambda_i$ is positive.

P. Frederickson, J. Kaplan, E. Yorke, J. Yorke. (1983). *The Lyapunov Dimension of Strange Attractors. J. Diff. Eqs.* 49 (2): 185–207.

The Overlapping Problem



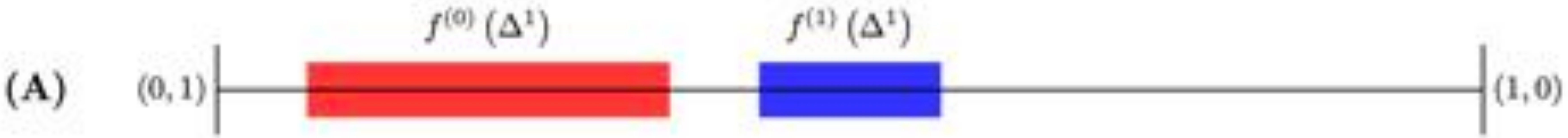
Problem: this solution does not work for "overlapping" IFSs.

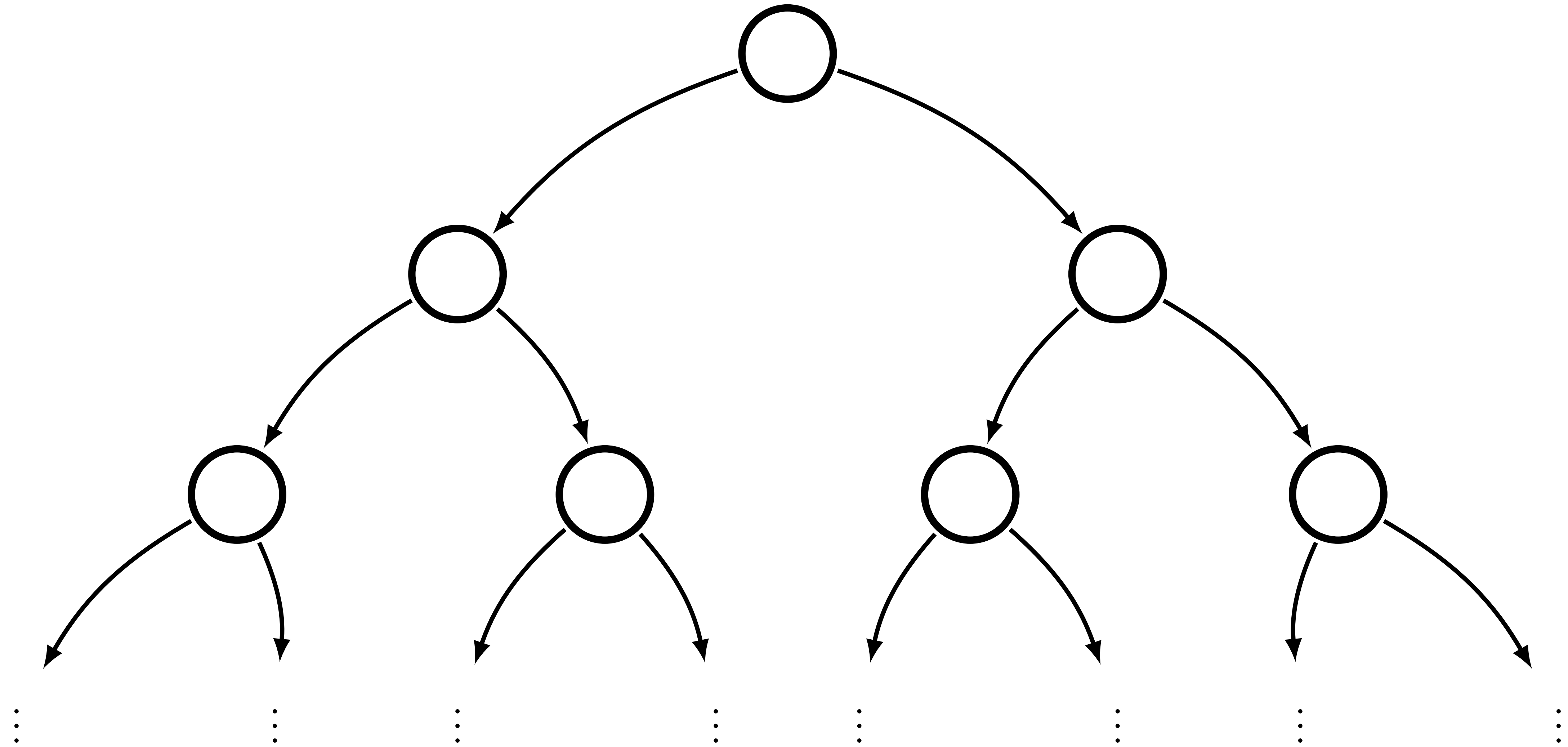
We end up "double counting" some of the states.

$$\dim_\mu(R) \leq k + \frac{h_\mu + \sum_i^k \lambda_i}{|\lambda_{k+1}|}$$

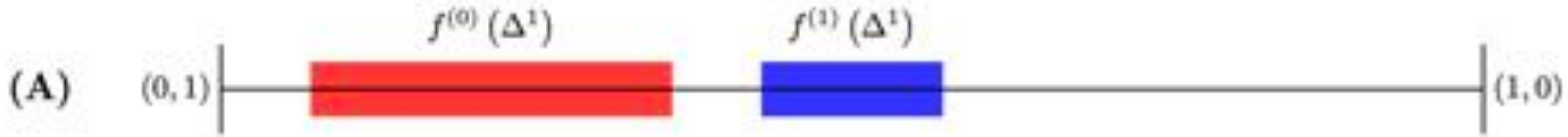
Alexandra M. Jurgens, James P. Crutchfield. *Divergent Predictive Memory: The Statistical Complexity Dimension of Stationary, Ergodic Finite-State Hidden Markov Processes*. Chaos 31, 083114, 2021.
Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

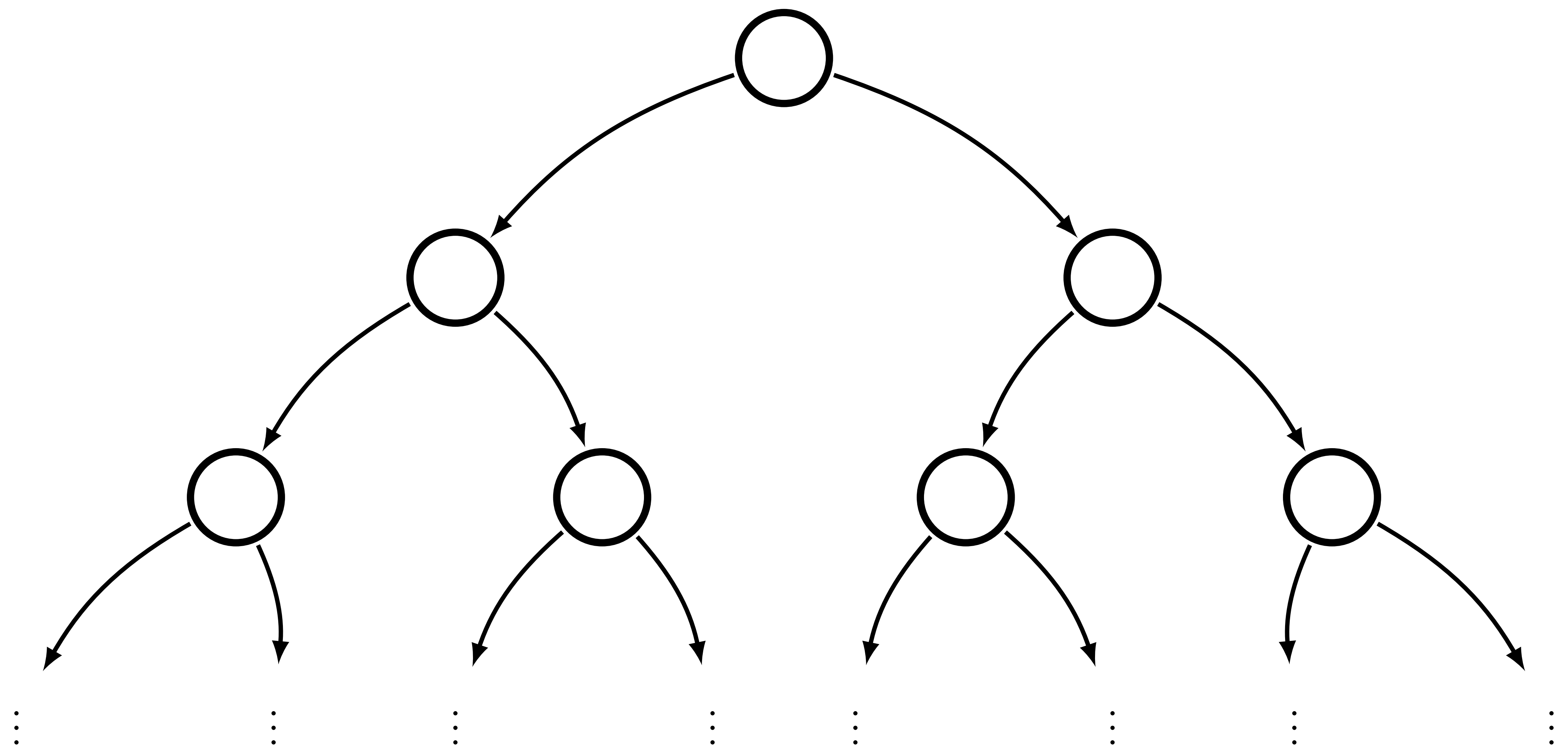
When KY Conjecture Works





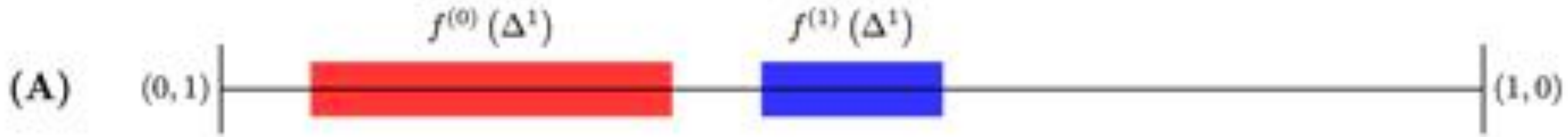
When KY Conjecture Works

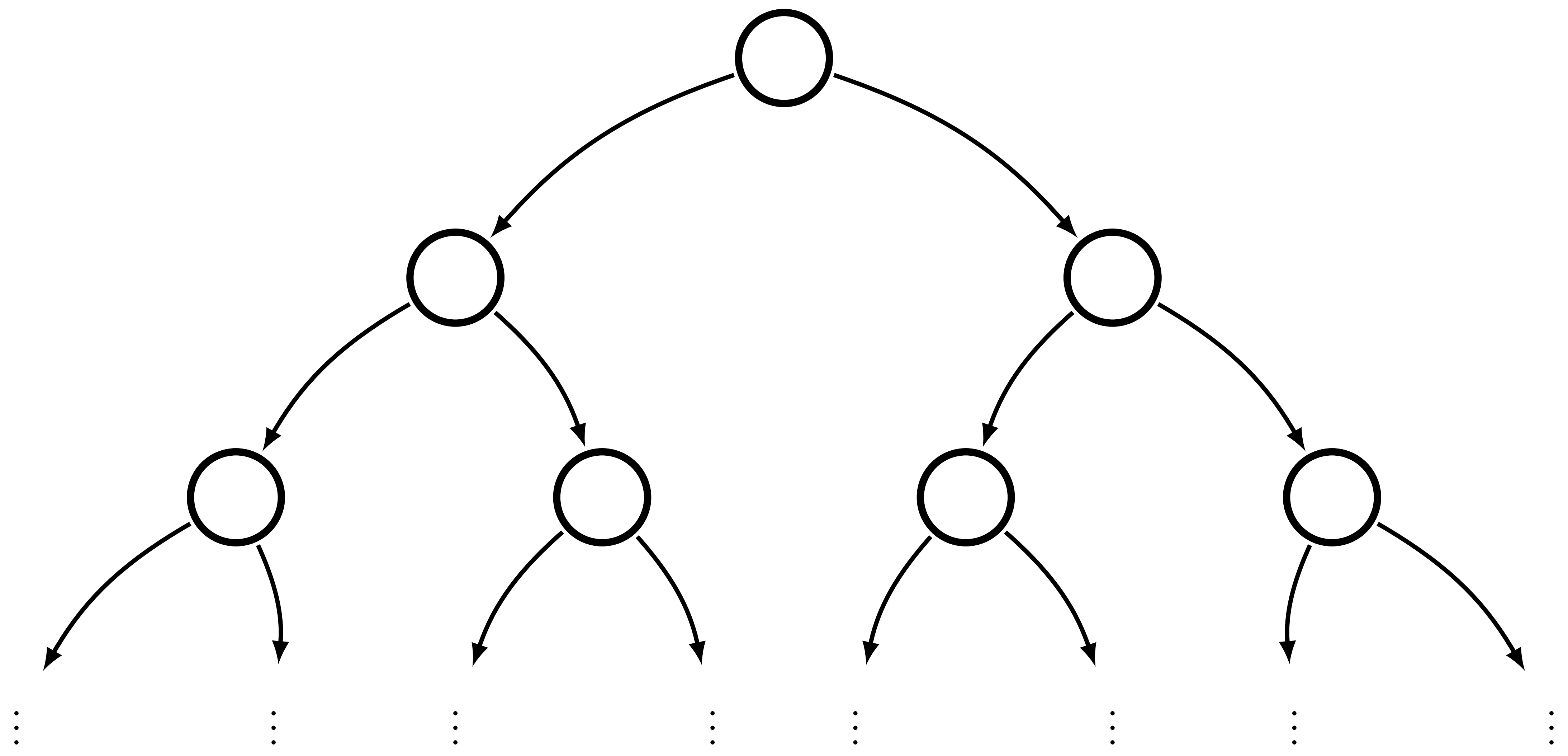




$$\dim_\mu(R) = k + \frac{h_\mu + \sum_i^k \lambda_i}{|\lambda_{k+1}|}$$

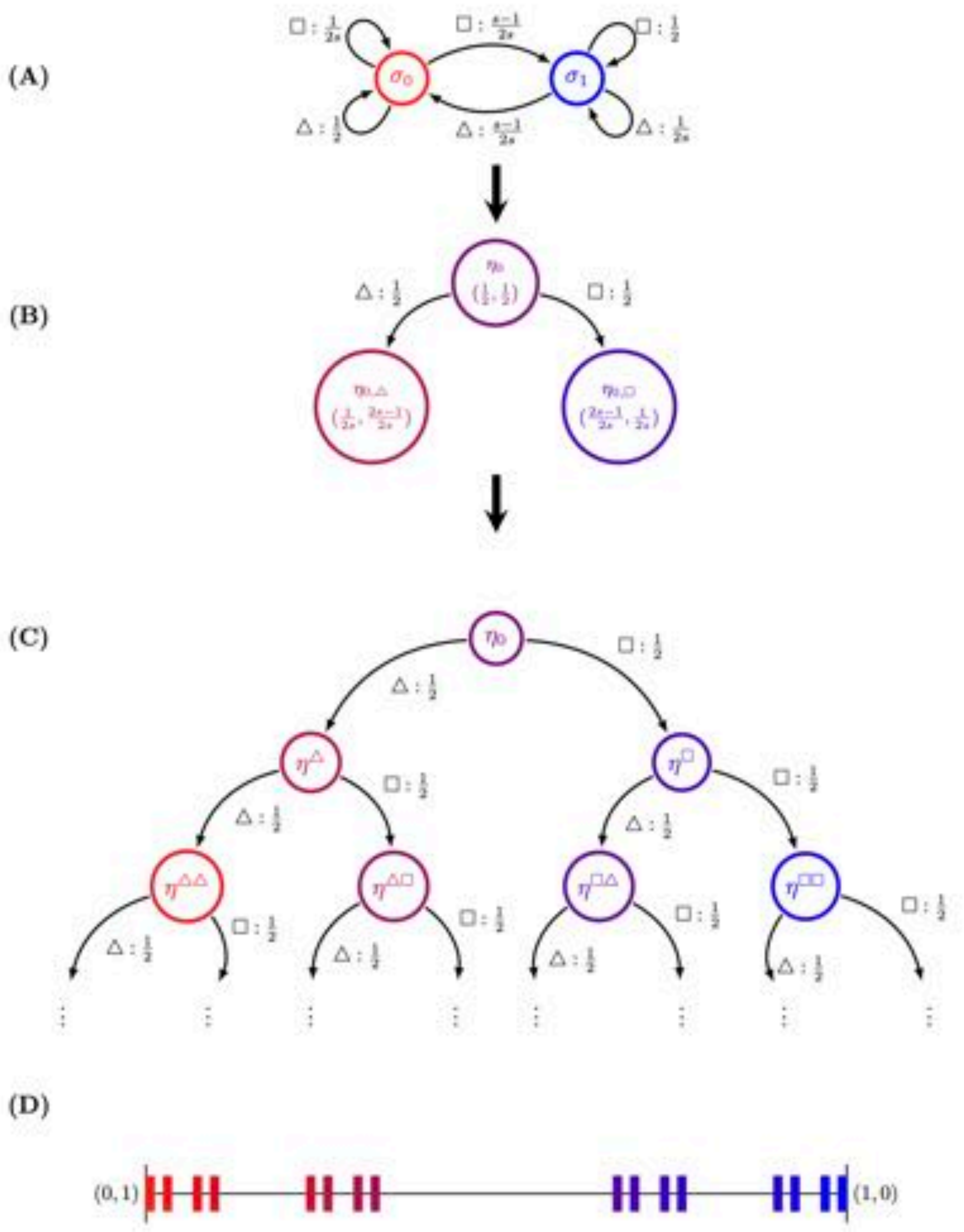
When KY Conjecture Works





$$\dim_\mu(R) = \frac{h_\mu}{|\lambda_1|}$$

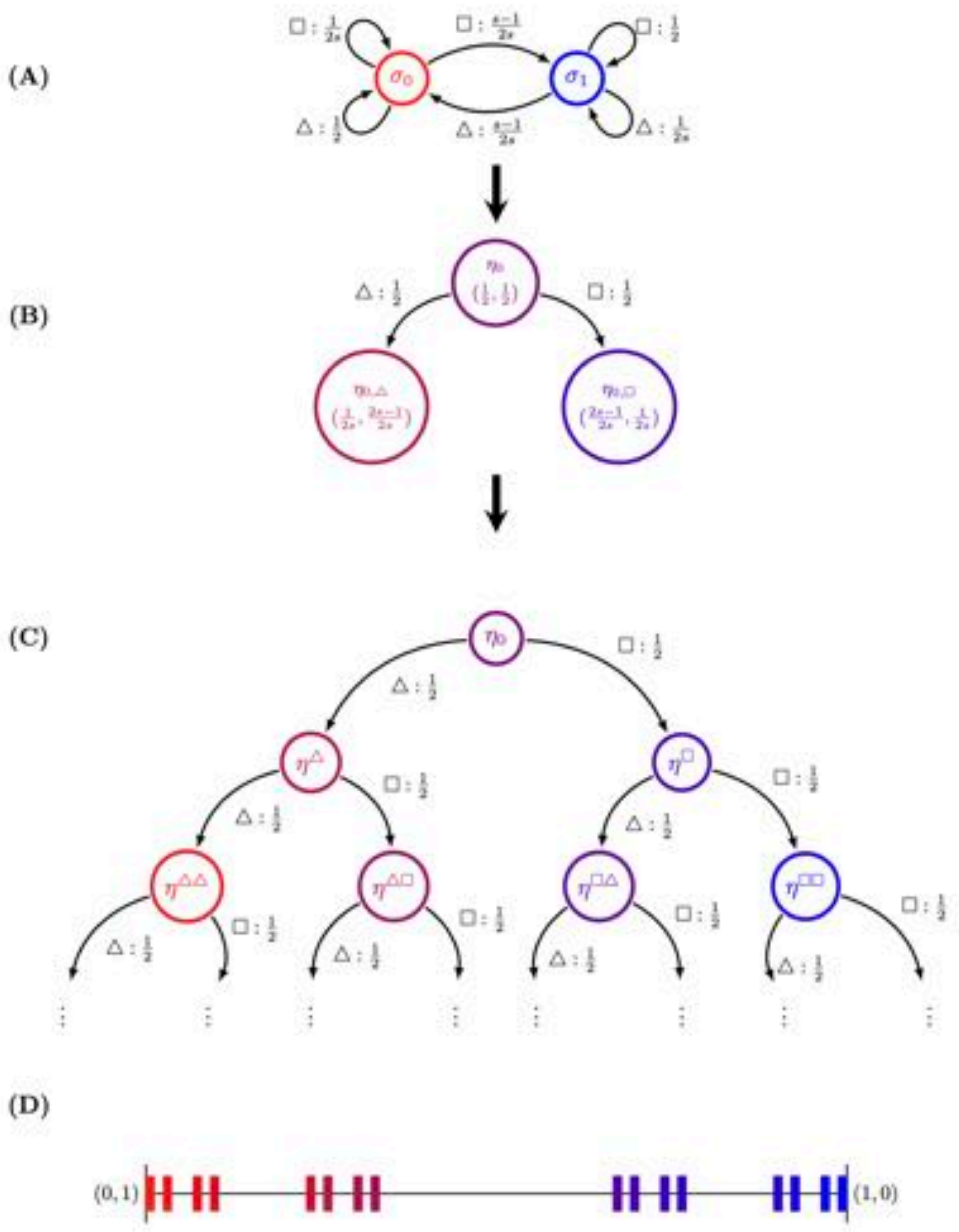
When KY Conjecture Works



$$\dim_\mu(R) = \frac{h_\mu}{|\lambda_1|}$$

Where λ_1 is the time-averaged Lyapunov exponent of each map.

When KY Conjecture Works

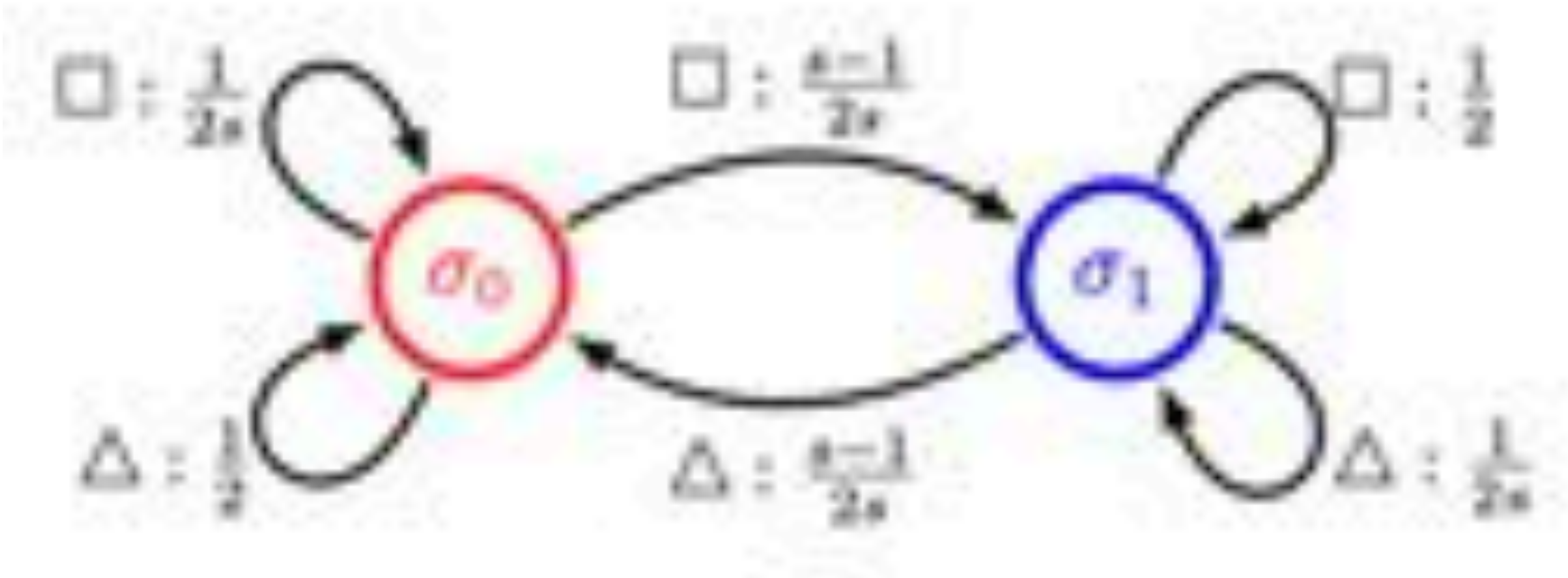


$$\dim_\mu(R) = \frac{h_\mu}{|\lambda_1|}$$

Let w be an L length word.
Let η_t be the mixed state at time t, which emits observation w_t.
Let $\eta_t = (x_t, y_t)$.

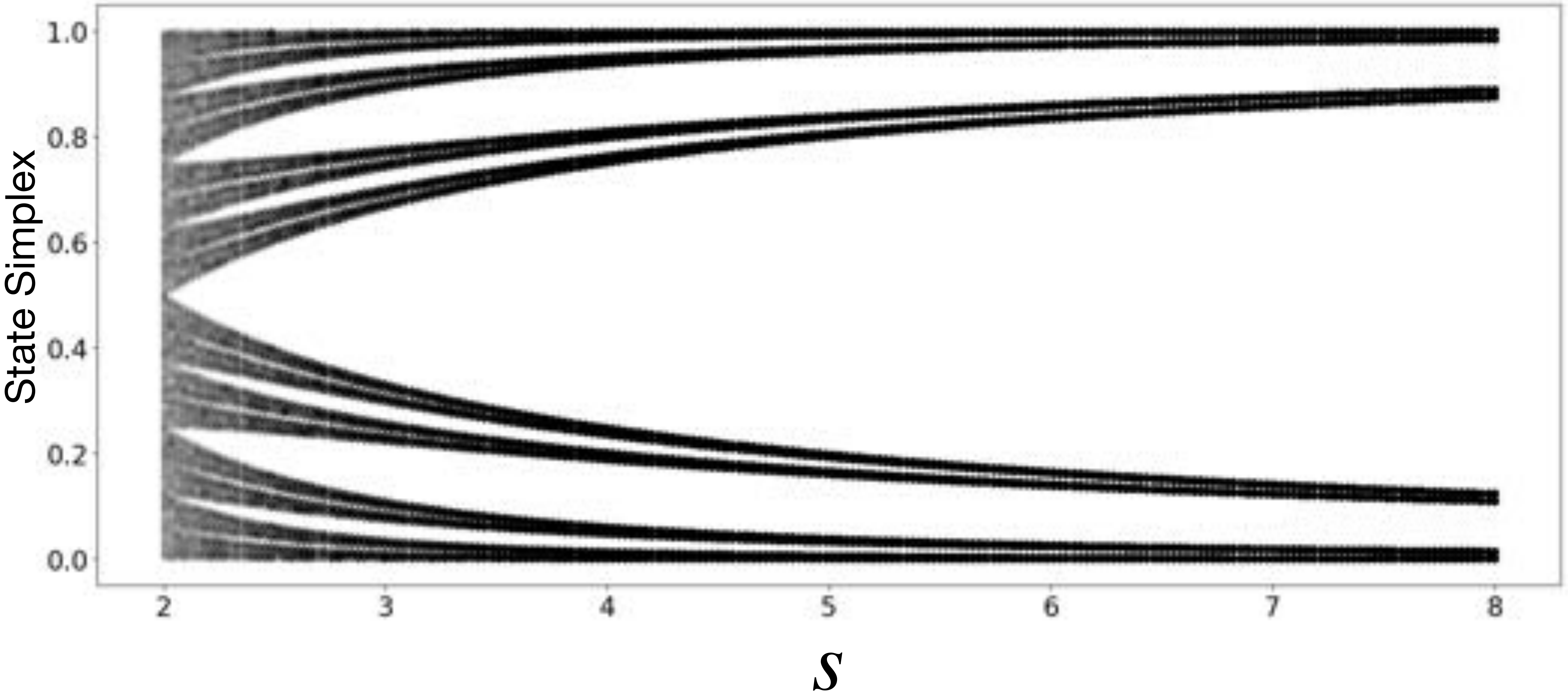
$$\lambda(w) = \lim_{L\to\infty} \frac{1}{L} \sum_{i=0}^{n-1} \ln \left| \frac{df^{(w_i)}(x)}{dx} \Big|_{x = x_t} \right|$$

Cantor Set Machine

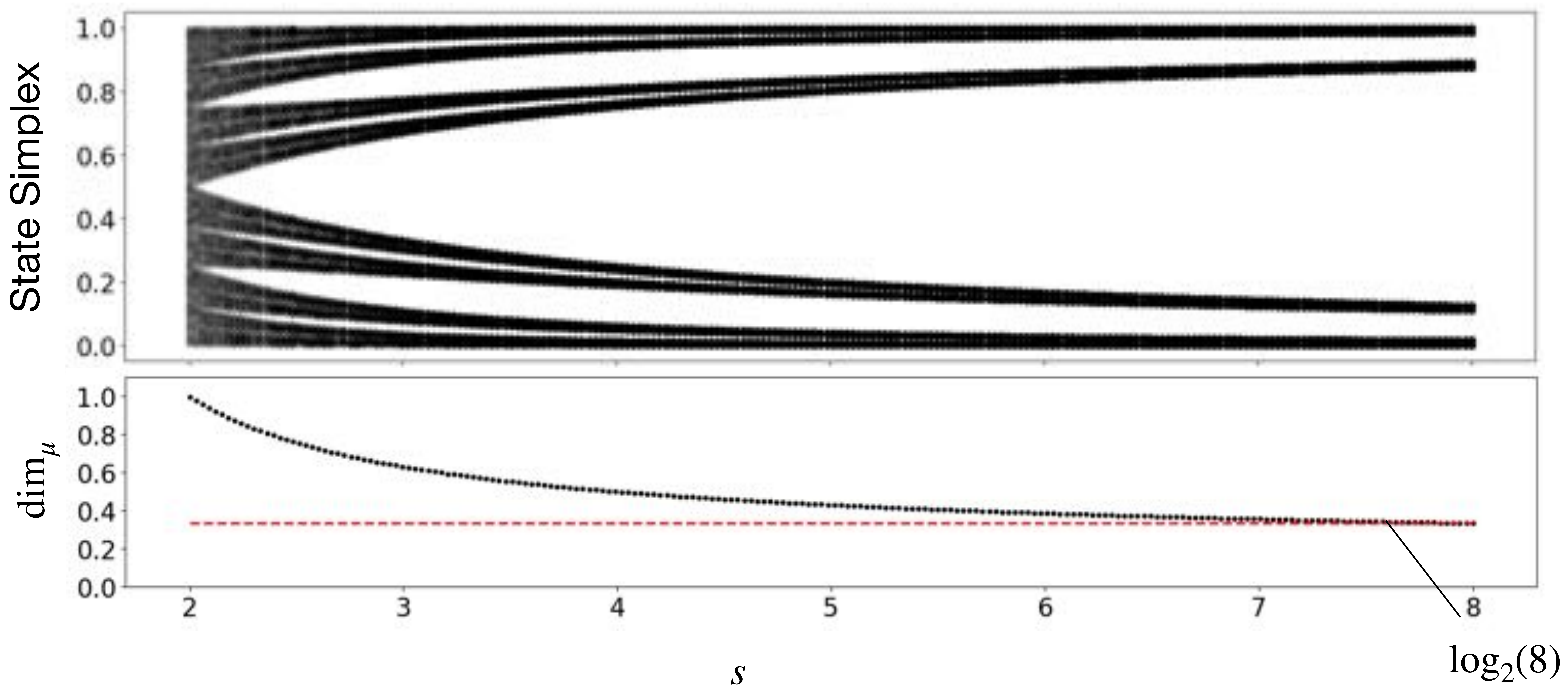


This machine produces a missing s Cantor set.

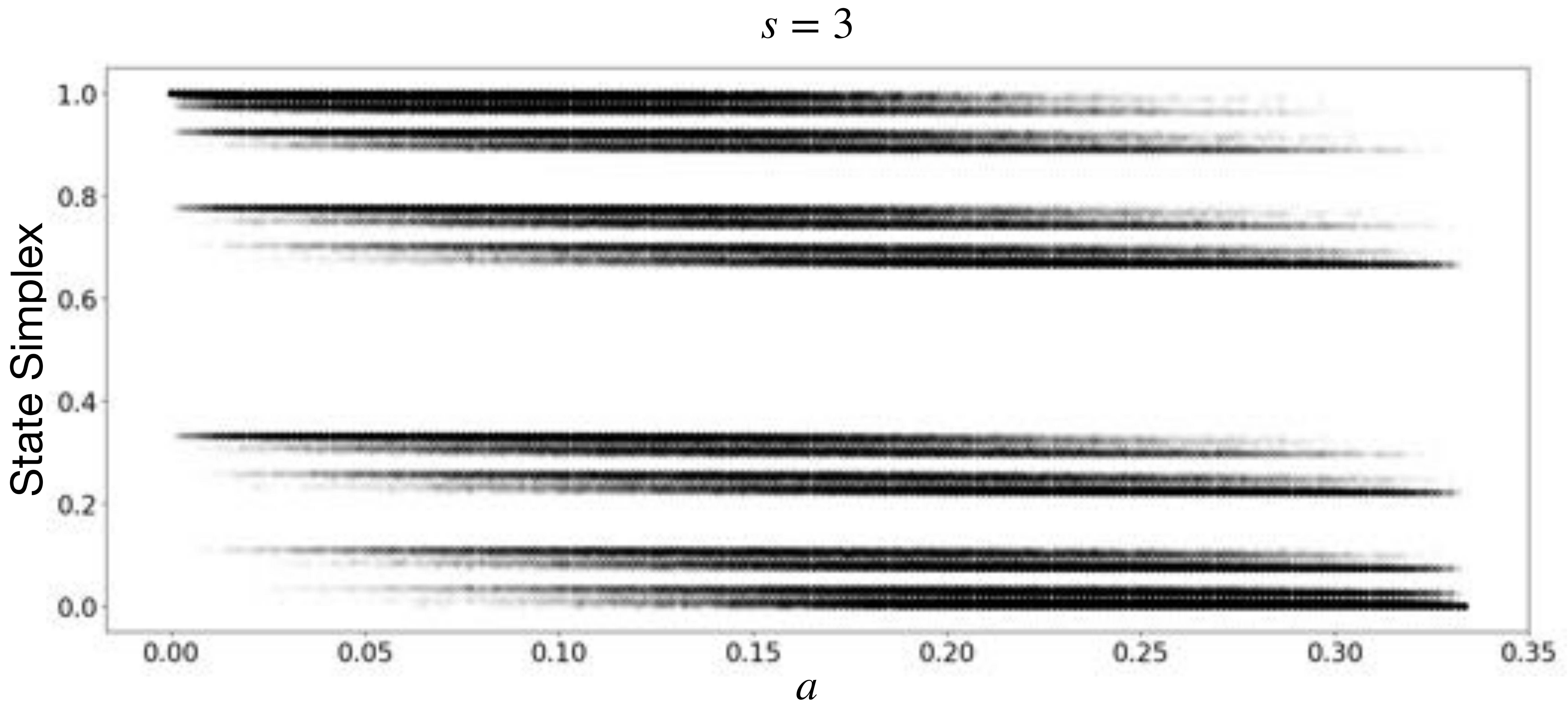
Cantor Set Machine



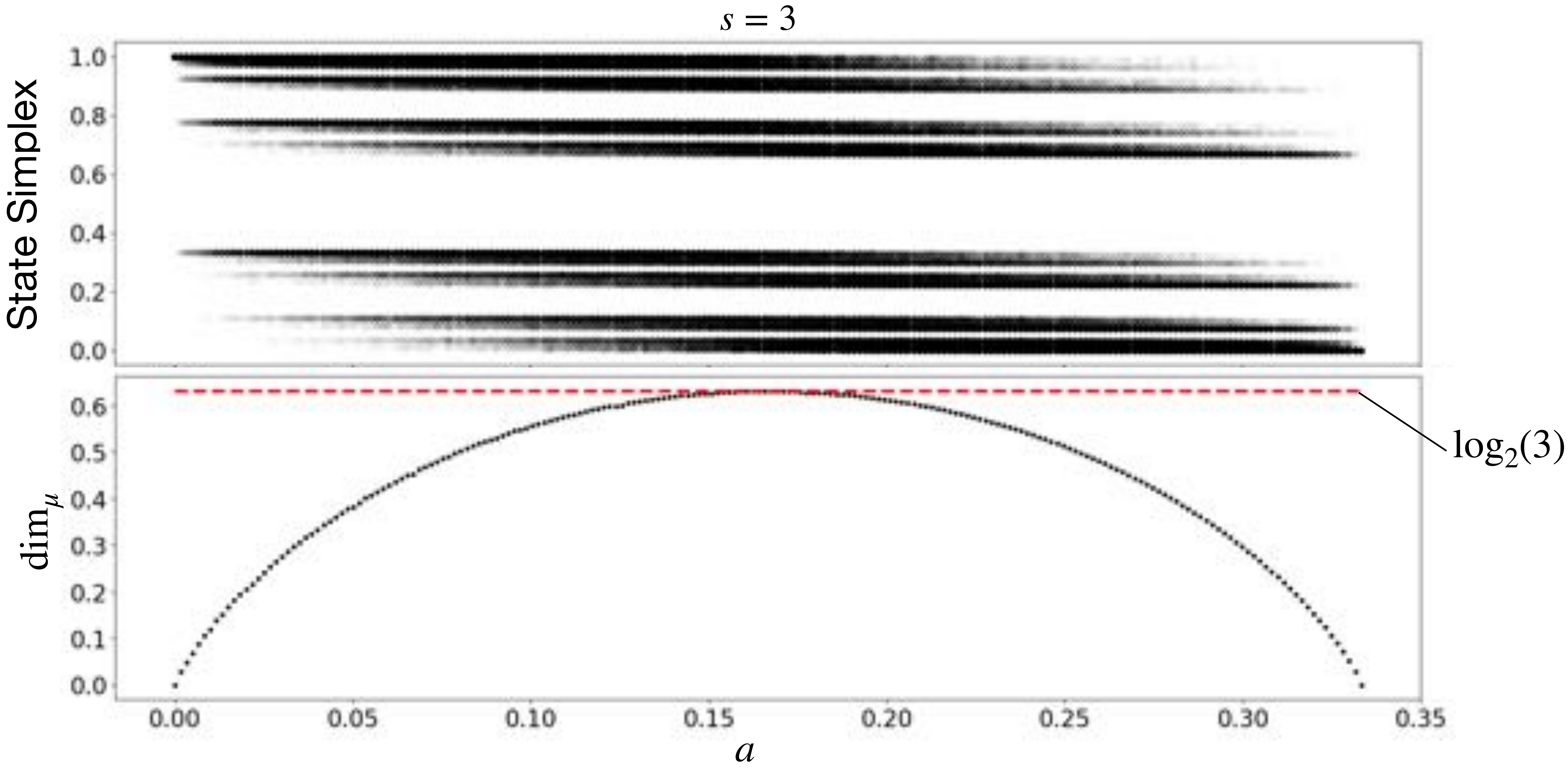
Cantor Set Machine



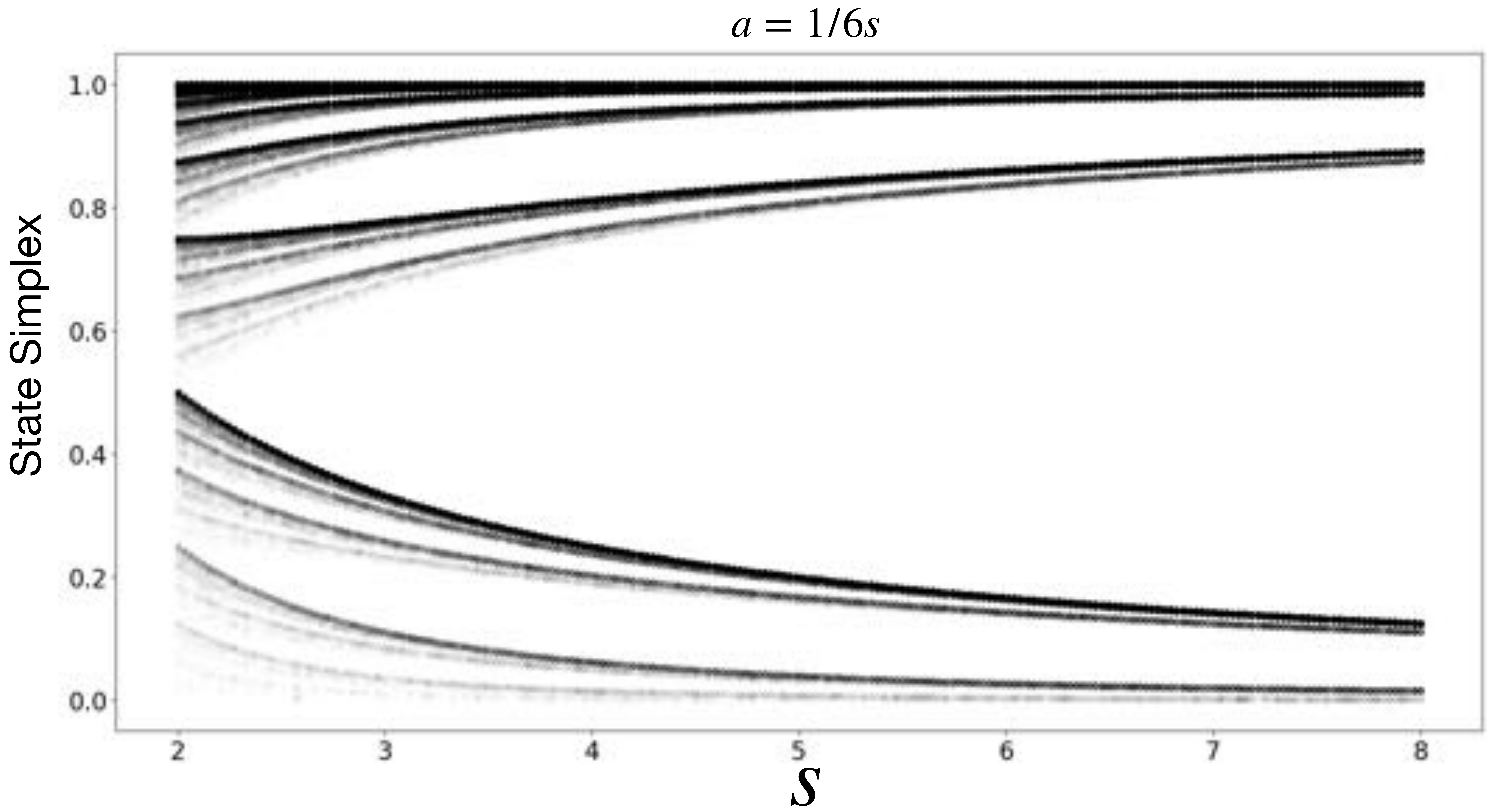
Cantor Set Machine



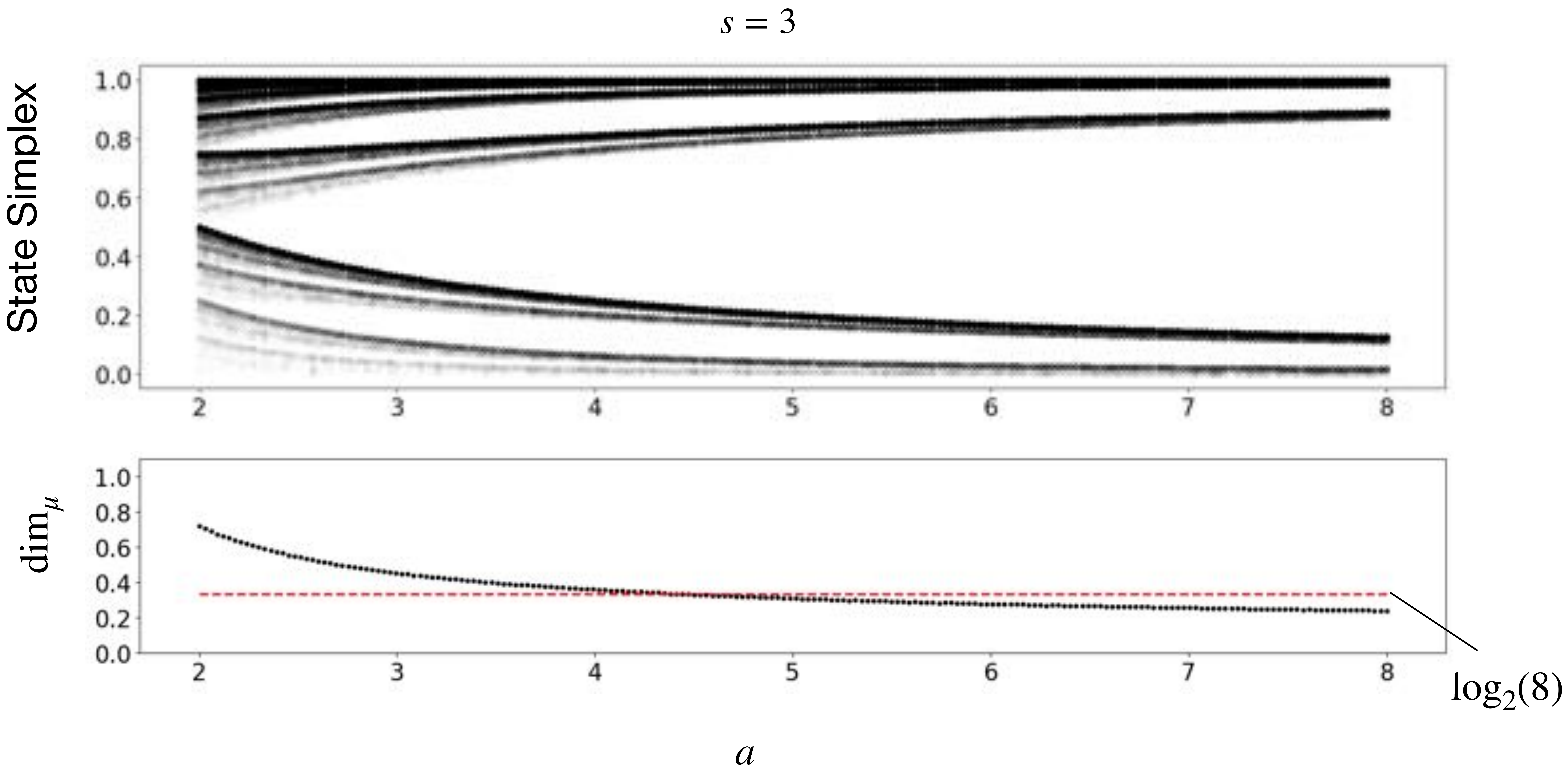
Cantor Set Machine



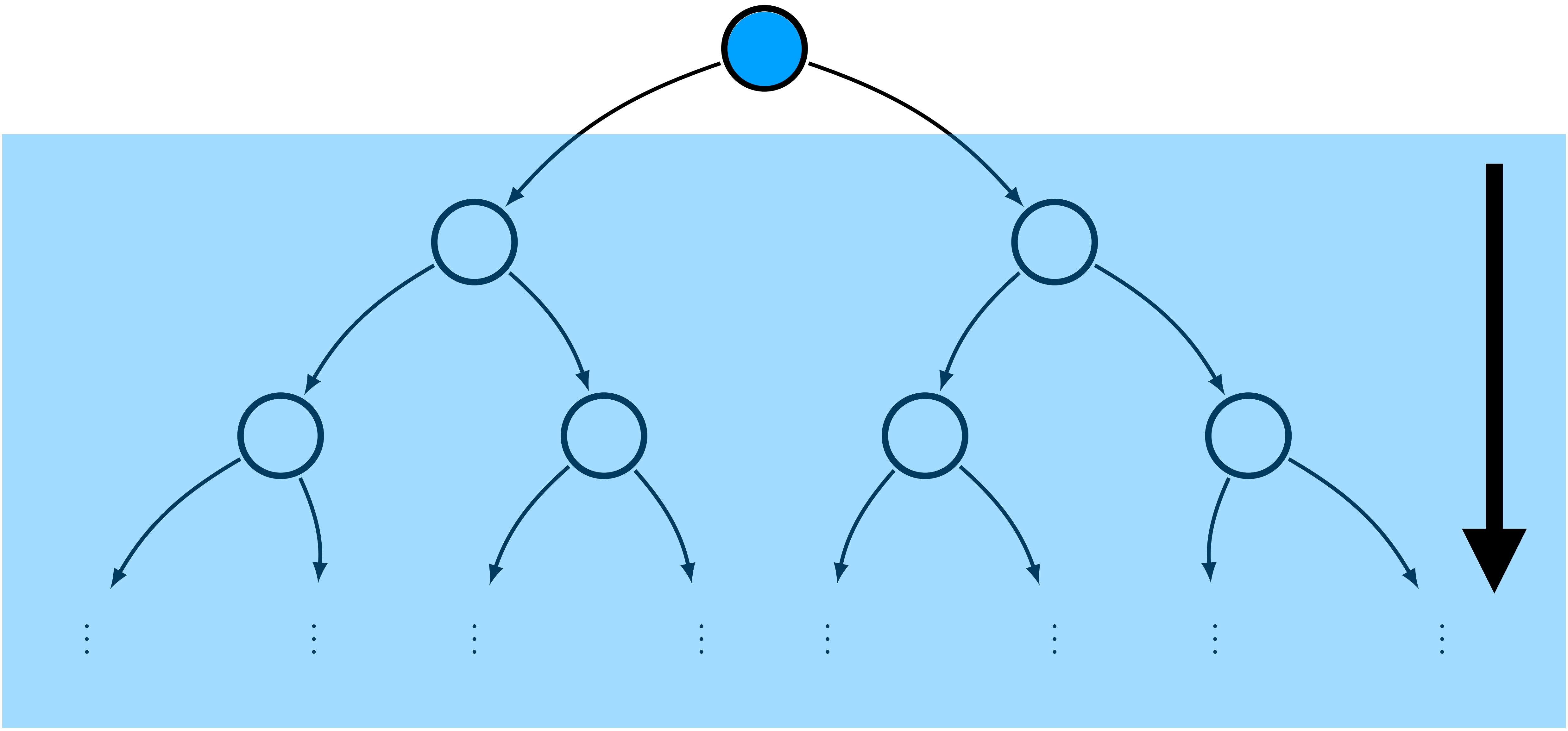
Cantor Set Machine



Cantor Set Machine



Overlap Problem



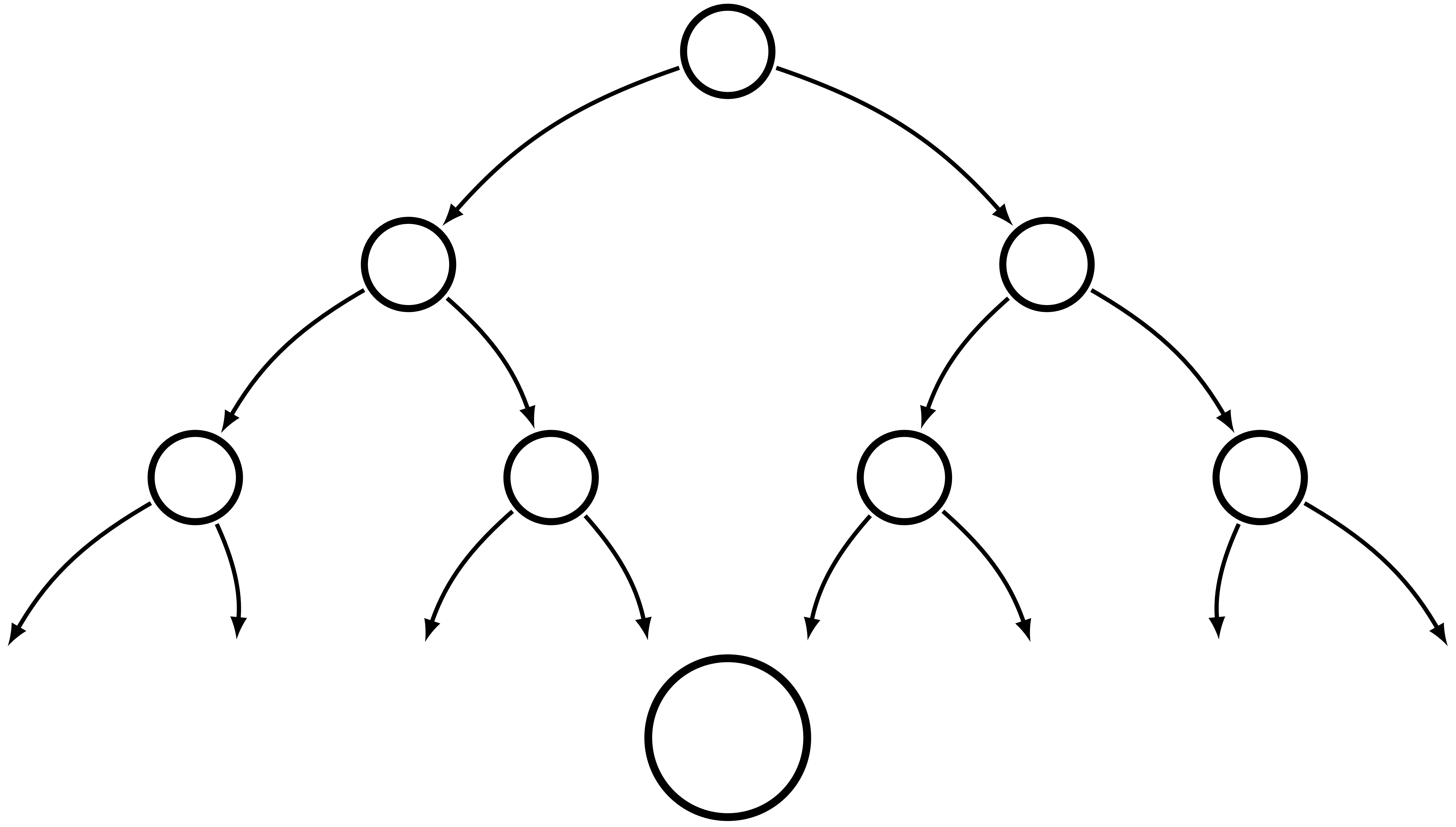
Entropy rate measures uncertainty in the next symbol given the present.

$$h_\mu = H\left[X_0, S_1 \mid S_0\right]$$

Rate of new symbols, on average…

Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

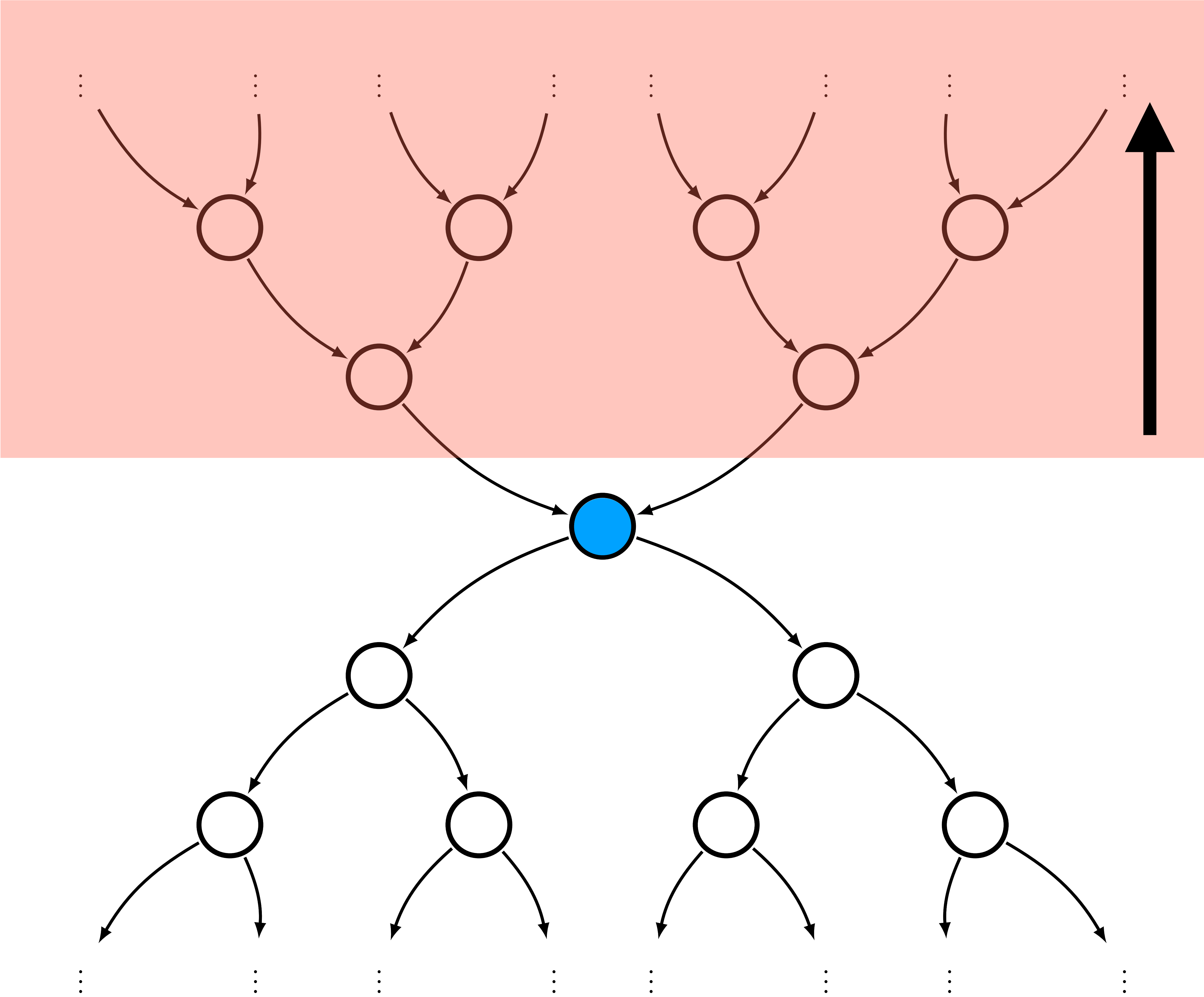
Overlap Problem



But, the branching entropy *overcounts* the average number of states generated when it is possible for two states to map into the same state.

We need a correction to count the average number of new states properly.

New Quantity: Ambiguity Rate



Entropy rate measures uncertainty in the next symbol given the present.

$$h_\mu = H\left[X_0, S_1 \mid S_0\right]$$

Ambiguity rate measures uncertainty *in prior symbol given the present*.

$$h_a = H\left[X_{-1}, S_{-1} \mid S_0\right]$$

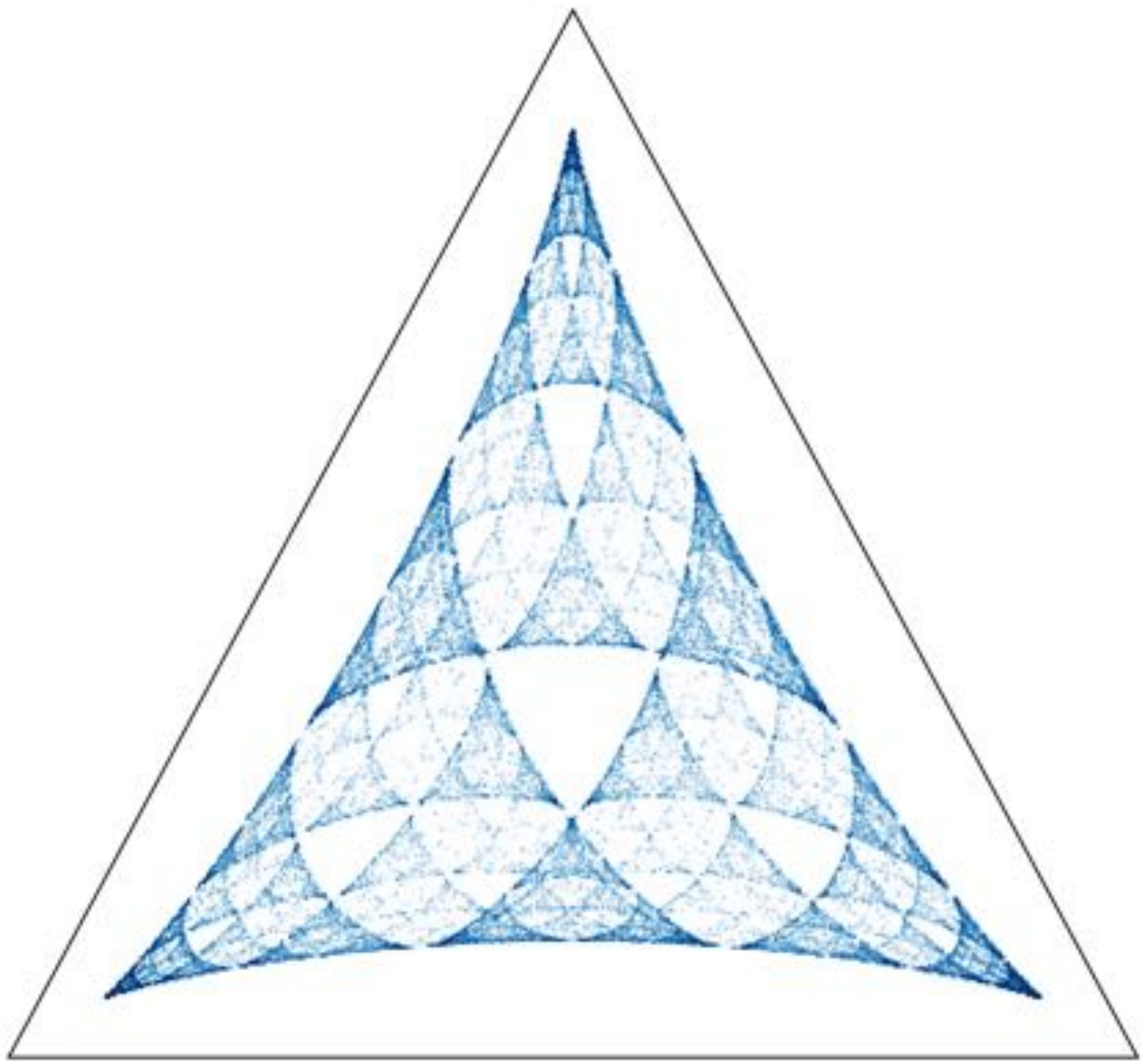
Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

Proposed Correction to Kaplan Yorke Conjecture

$$\dim_\mu(R) = k + \frac{h_\mu - h_a + \sum_i^k \lambda_i}{|\lambda_{k+1}|}$$

Alexandra M. Jurgens, James P. Crutchfield. *Divergent Predictive Memory: The Statistical Complexity Dimension of Stationary, Ergodic Finite-State Hidden Markov Processes*. Chaos 31, 083114, 2021.
Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

Proposed Correction to Kaplan Yorke Conjecture



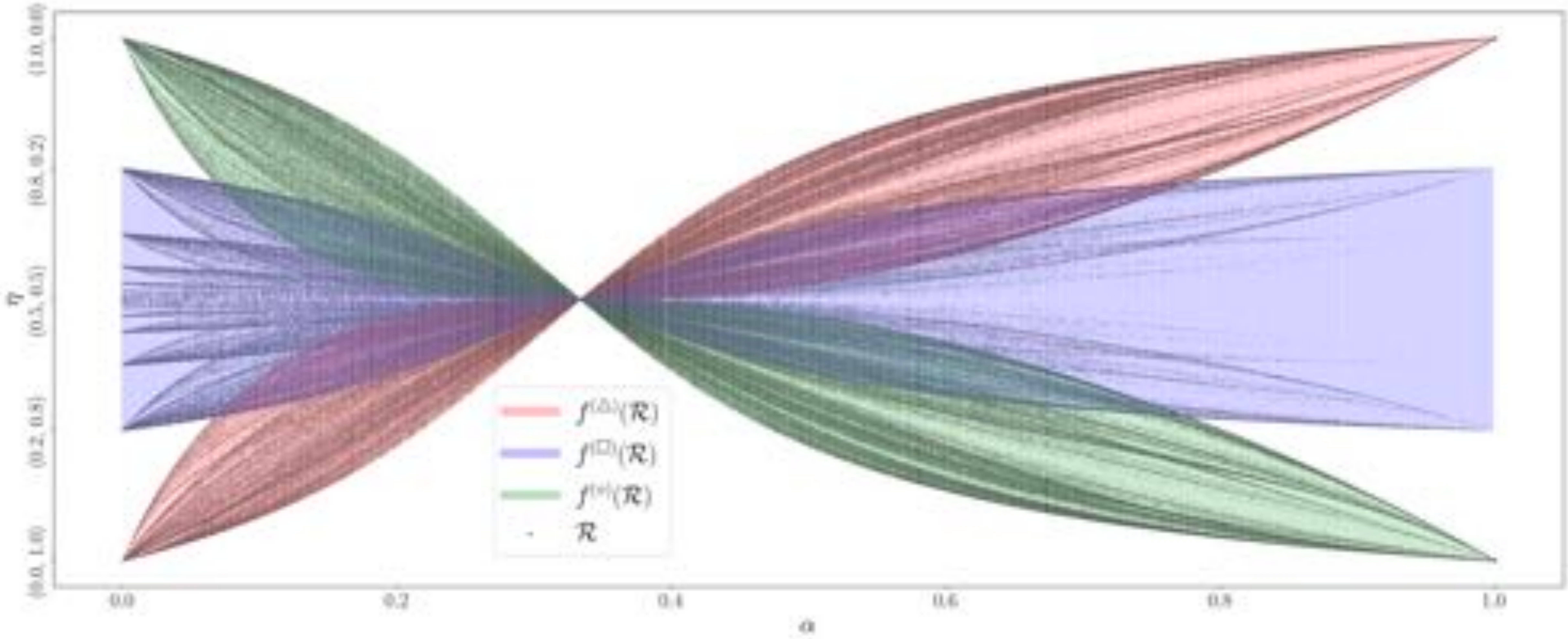
Entropy rate

Ambiguity rate

$$\dim_\mu(R) = k + \frac{h_\mu - h_a + \sum_i^k \lambda_i}{|\lambda_{k+1}|}$$

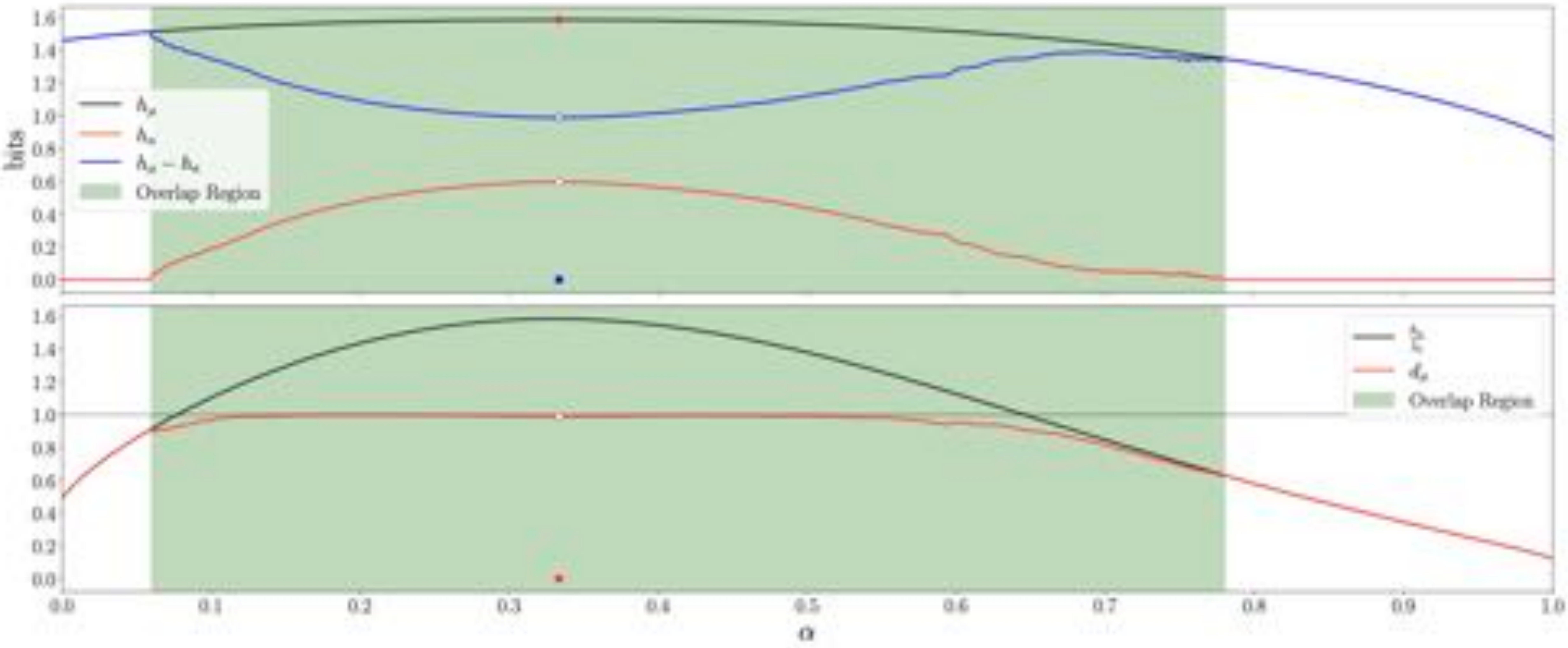
Alexandra M. Jurgens, James P. Crutchfield. *Divergent Predictive Memory: The Statistical Complexity Dimension of Stationary, Ergodic Finite-State Hidden Markov Processes*. Chaos 31, 083114, 2021.
Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

Proposed Correction to Kaplan Yorke Conjecture



Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

Proposed Correction to Kaplan Yorke Conjecture

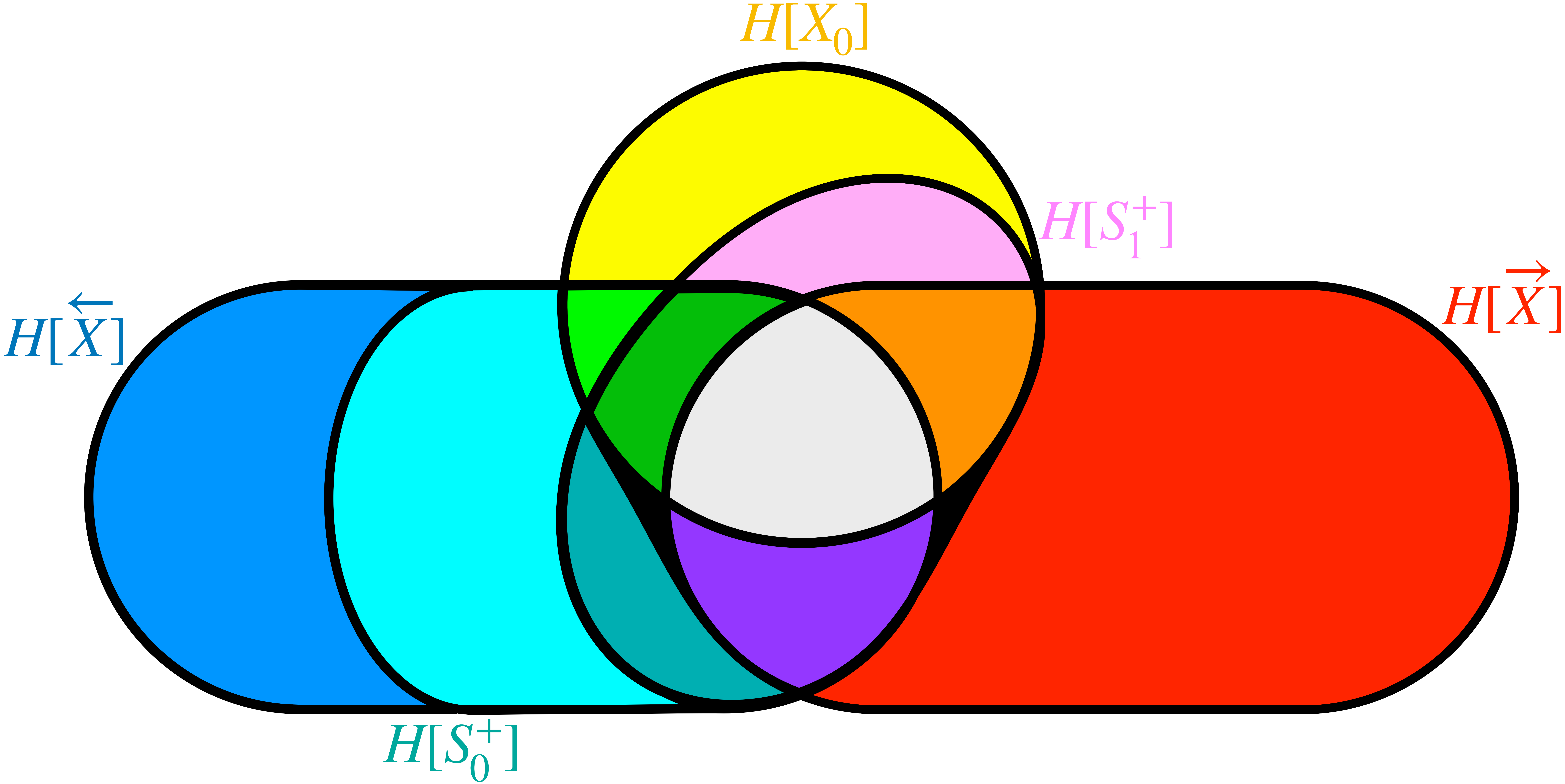


Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

What is h_a?

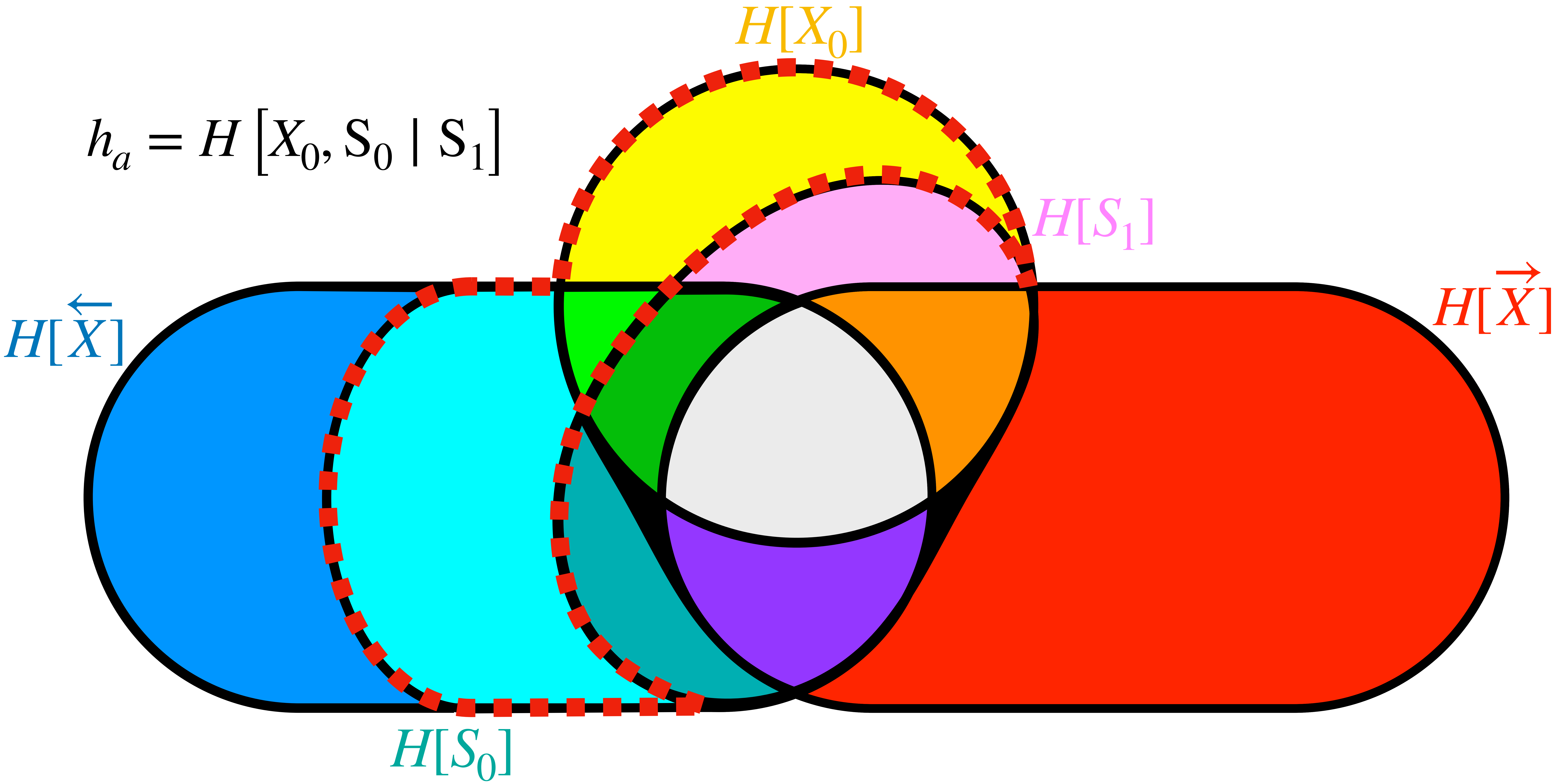
Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

What is h_a?

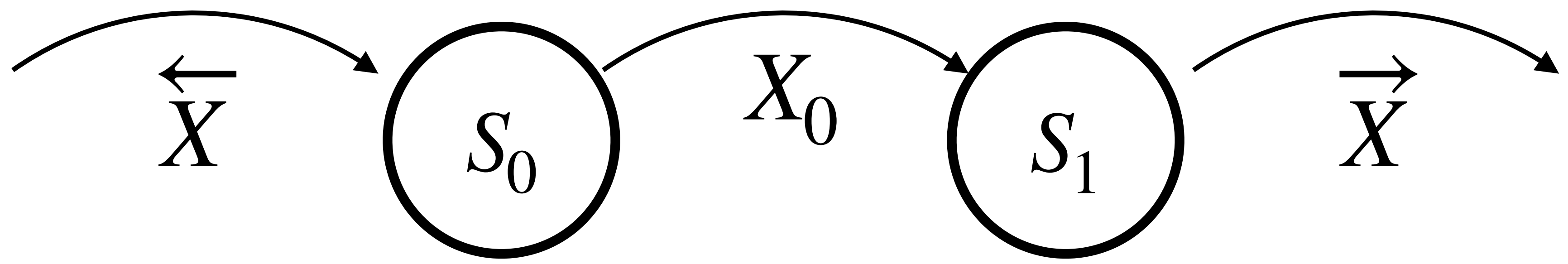


What is h_a?

$$h_a = H\left[X_0, S_0 \mid S_1\right]$$



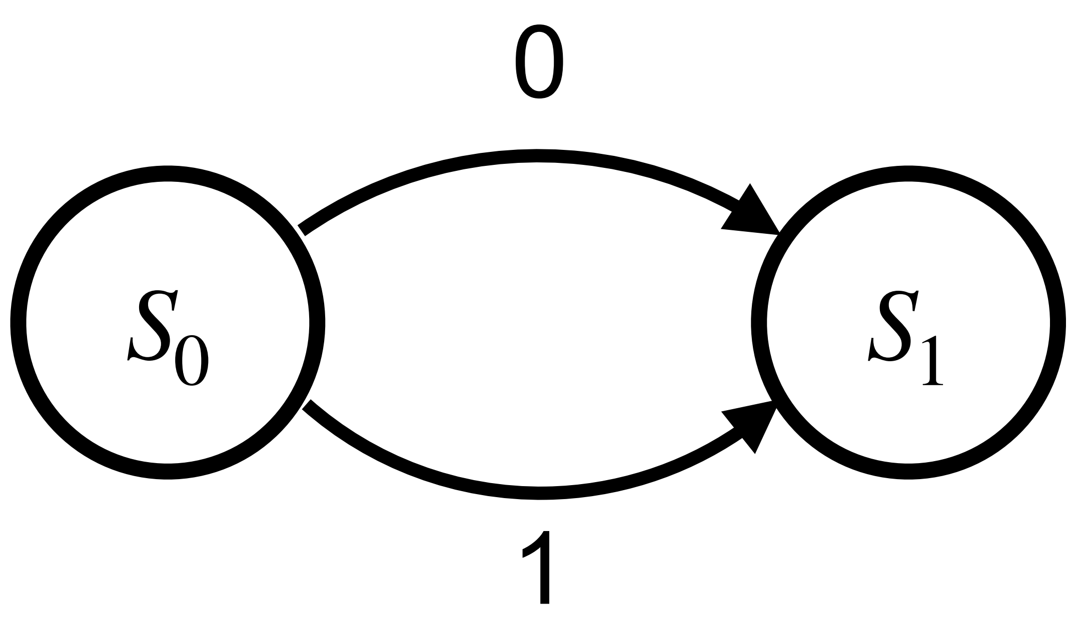
Components of Ambiguity Rate



Ephemeral information

$$H[X_0 \mid S_0, S_1]$$

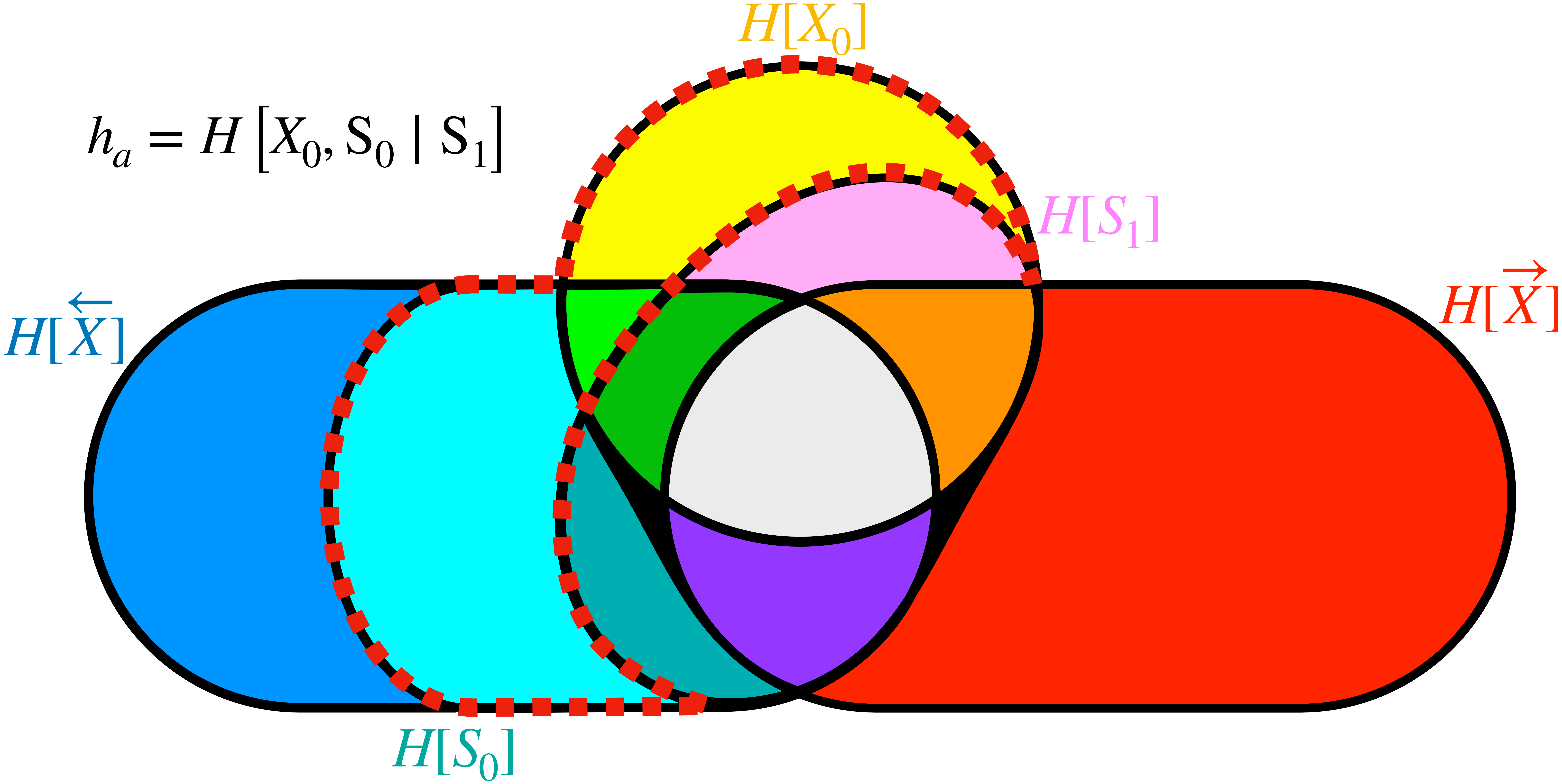
Information generated in
present, not used for
prediction.



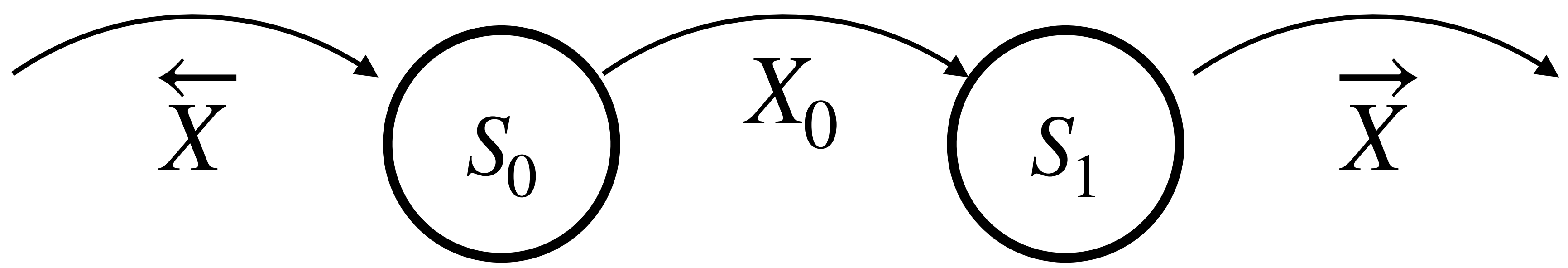
A Jurgens, JP Crutchfield. Taxonomy of Prediction. arXiv preprint arXiv:2504.11371, 2025
R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *CHAOS*. 17. 21(3):037109, 2011. doi:10.1063/1.3637494

Components of Ambiguity Rate

$$h_a = H\left[X_0, S_0 \mid S_1\right]$$



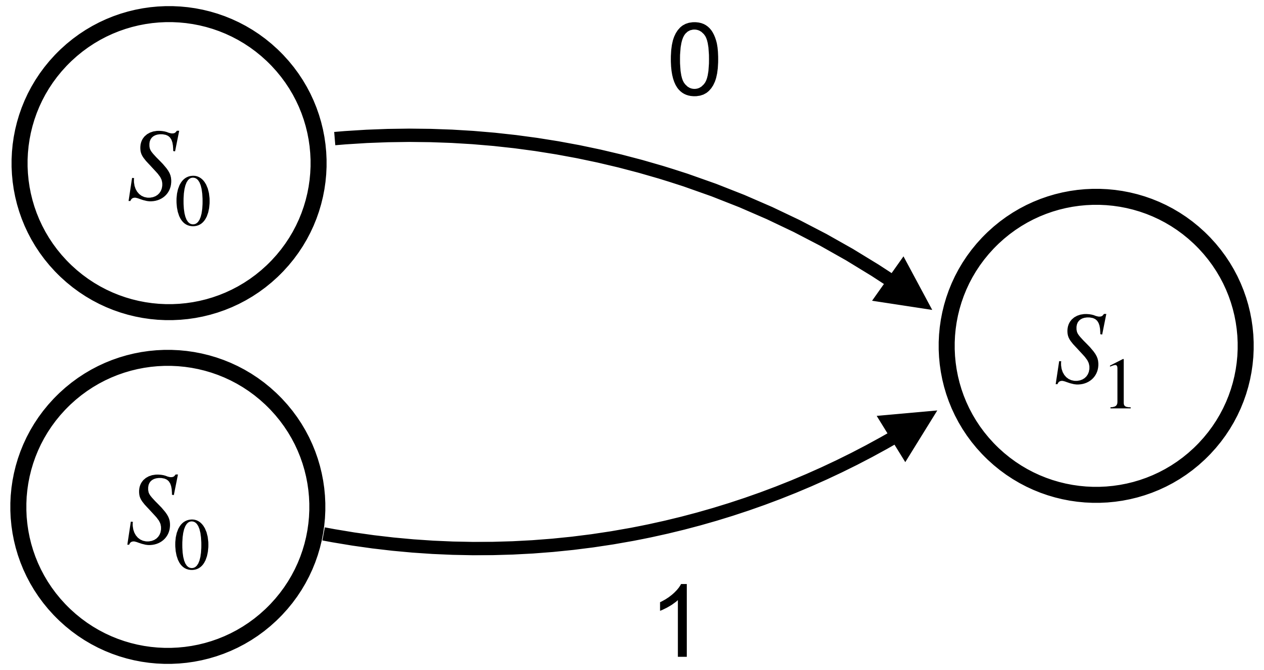
Components of Ambiguity Rate



Binding information

$$I[X_0; S_0 \mid S_1]$$

Information in the past and
present that doesn't get
used for future prediction.

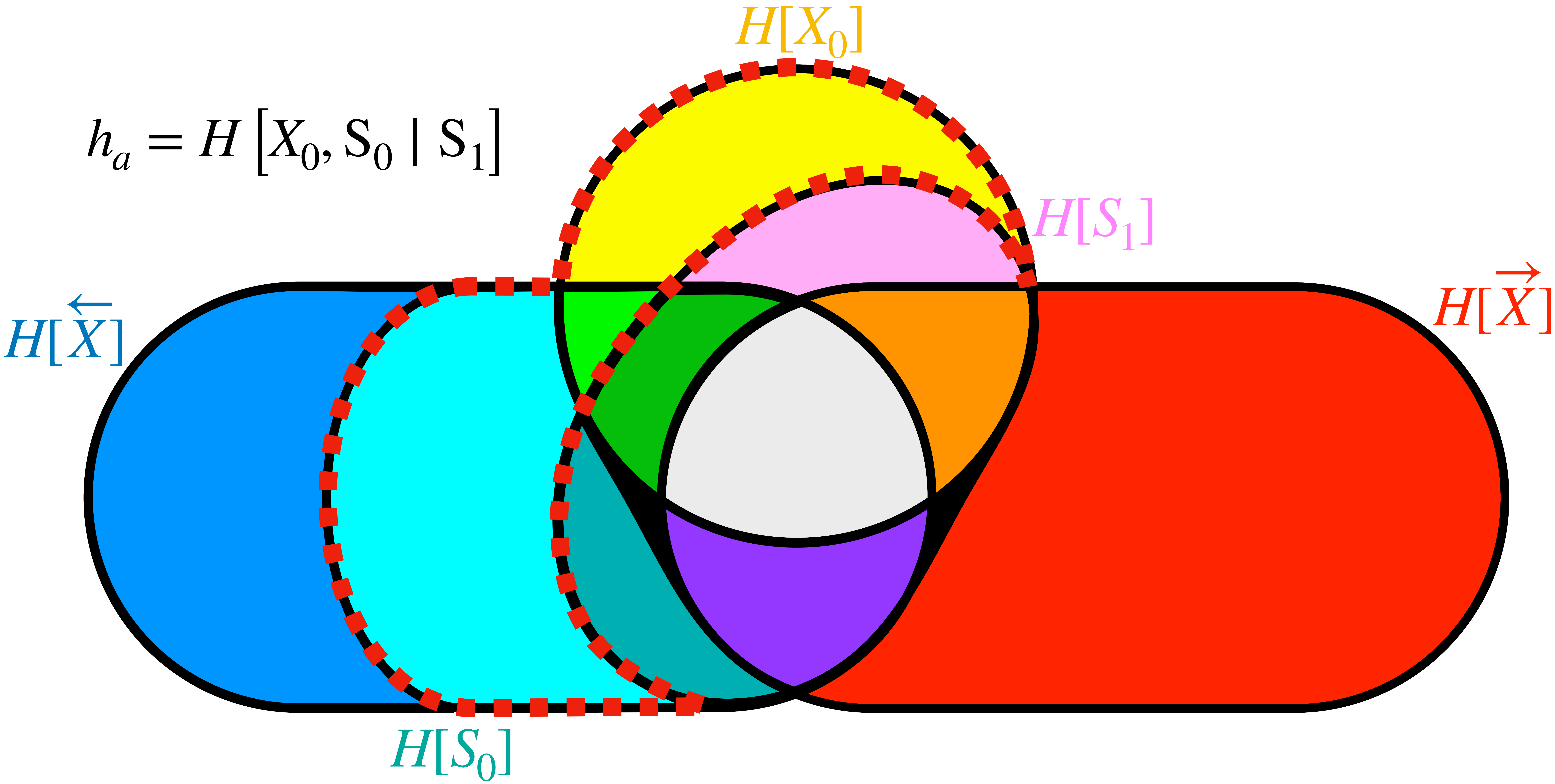


A Jurgens, JP Crutchfield. Taxonomy of Prediction. arXiv preprint arXiv:2504.11371, 2025

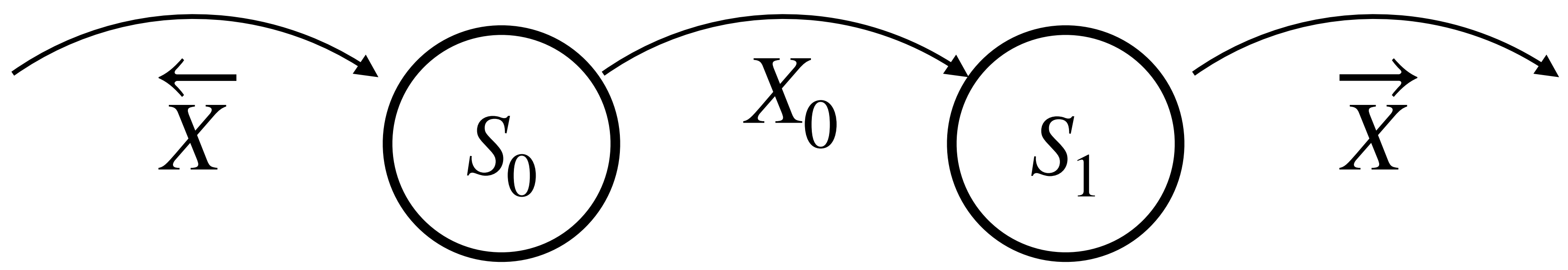
R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *CHAOS*. 17. 21(3):037109, 2011. doi:10.1063/1.3637494

Components of Ambiguity Rate

$$h_a = H\left[X_0, S_0 \mid S_1\right]$$



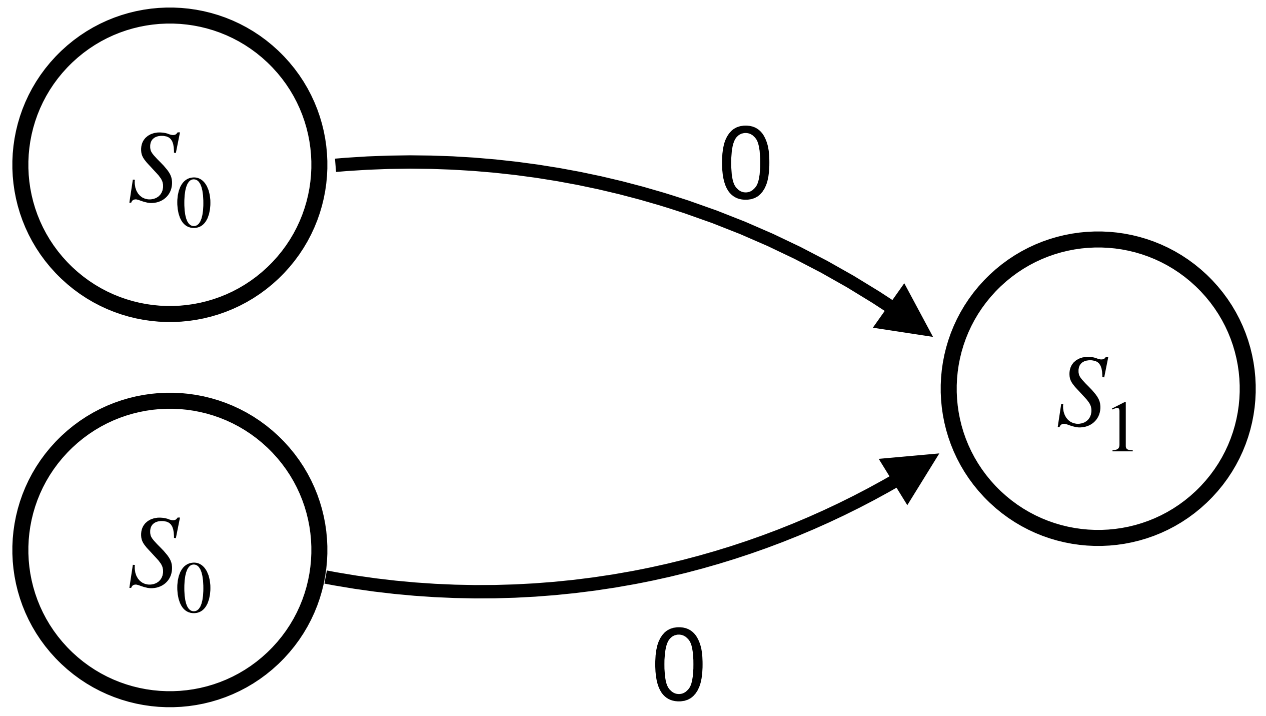
Components of Ambiguity Rate



Crypticity

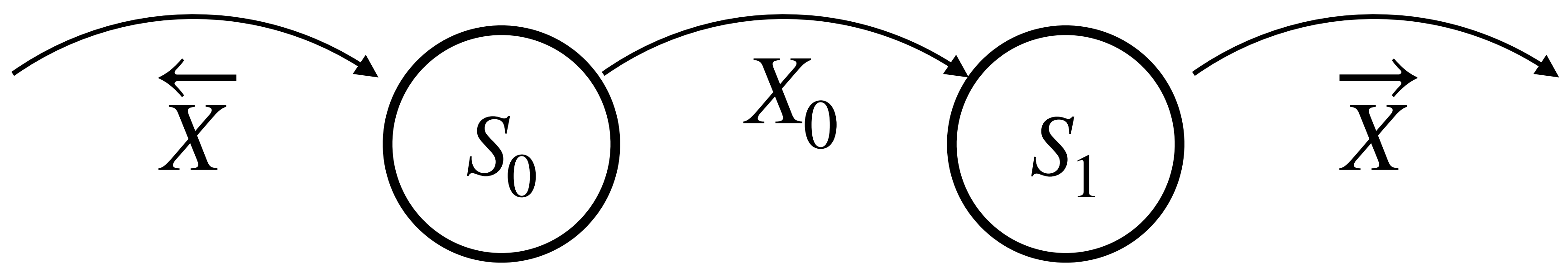
$$H[S_0 \,|\, X_0, S_1]$$

Information in past used for prediction but then forgotten for future prediction.



A Jurgens, JP Crutchfield. Taxonomy of Prediction. arXiv preprint arXiv:2504.11371, 2025
R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *CHAOS*. 17. 21(3):037109, 2011. doi:10.1063/1.3637494

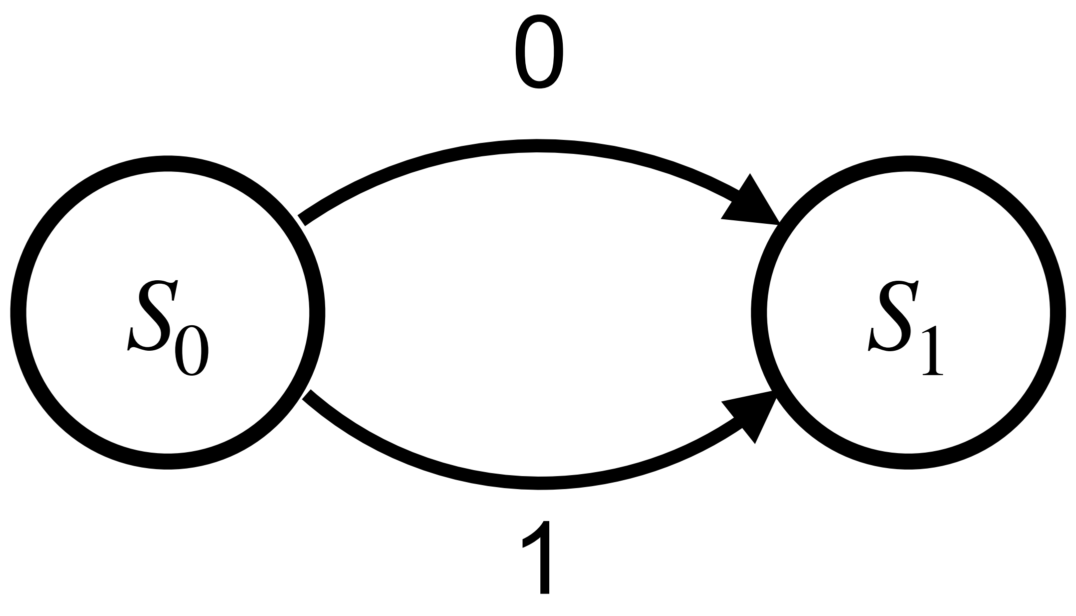
Components of Ambiguity Rate



Ephemeral information

$H[X_0 \,|\, S_0, S_1]$

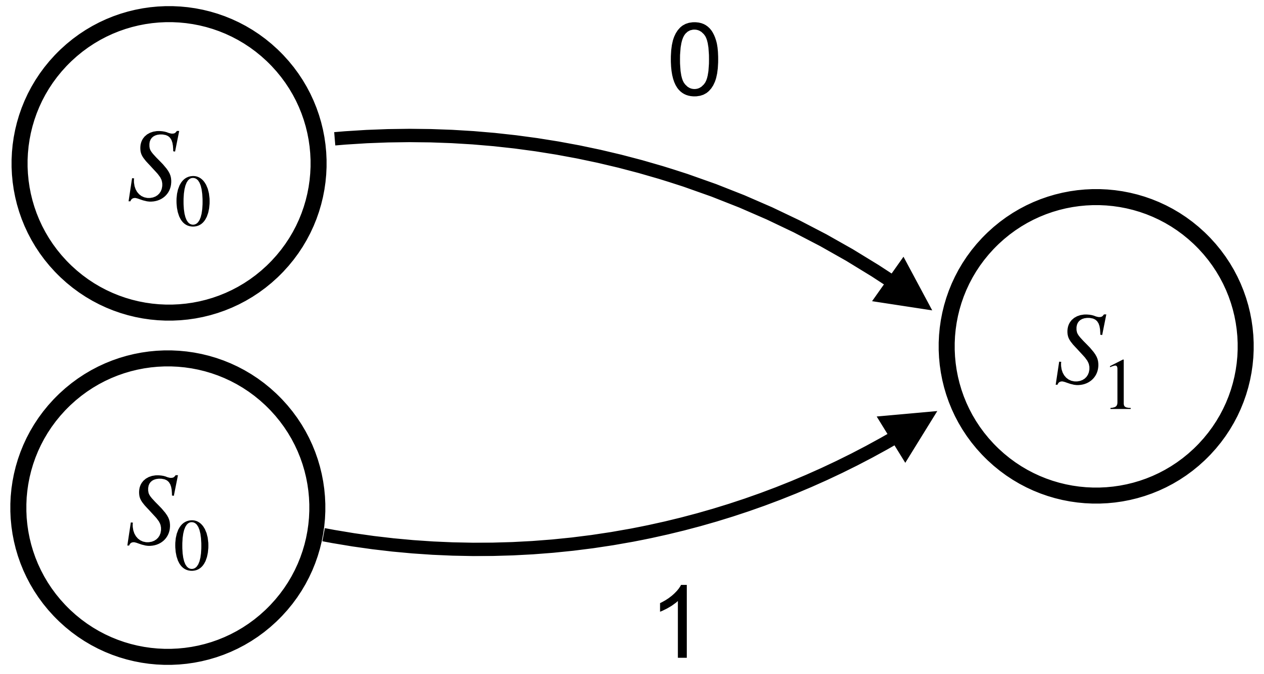
Information generated in present, not used for prediction.



Binding information

$I[X_0; S_0 \,|\, S_1]$

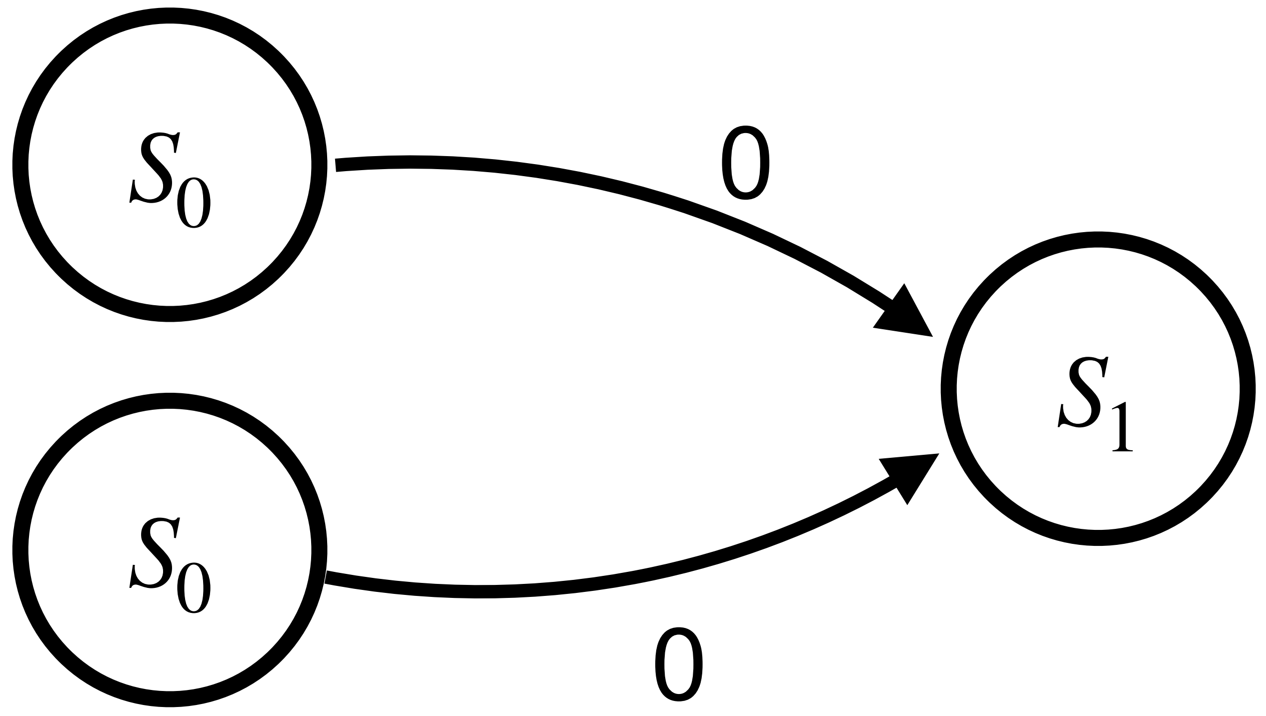
Information in the past and present that doesn't get used for future prediction.



Crypticity

$H[S_0 \,|\, X_0, S_1]$

Information in past used for prediction but then forgotten for future prediction.



A Jurgens, JP Crutchfield. Taxonomy of Prediction. arXiv preprint arXiv:2504.11371, 2025

R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *CHAOS*. 17. 21(3):037109, 2011. doi:10.1063/1.3637494

h_a as Model Growth Rate

$$\Delta H \left[\text{Predictive states}\right] \sim h_\mu - h_a$$

Entropy rate

Ambiguity rate

Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

h_a as Model Growth Rate

$$
\begin{aligned}
h_\mu - h_a &= H\left[X_t \mid S_t\right] - H\left[X_t, S_t \mid S_{t+1}\right] \\
&= H\left[S_{t+1} \mid S_t, X_t\right] + H\left[S_{t+1}\right] - H\left[S_t\right] \\
&= \Delta H\left[S\right]
\end{aligned}
$$

Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

h_a as Model Growth Rate

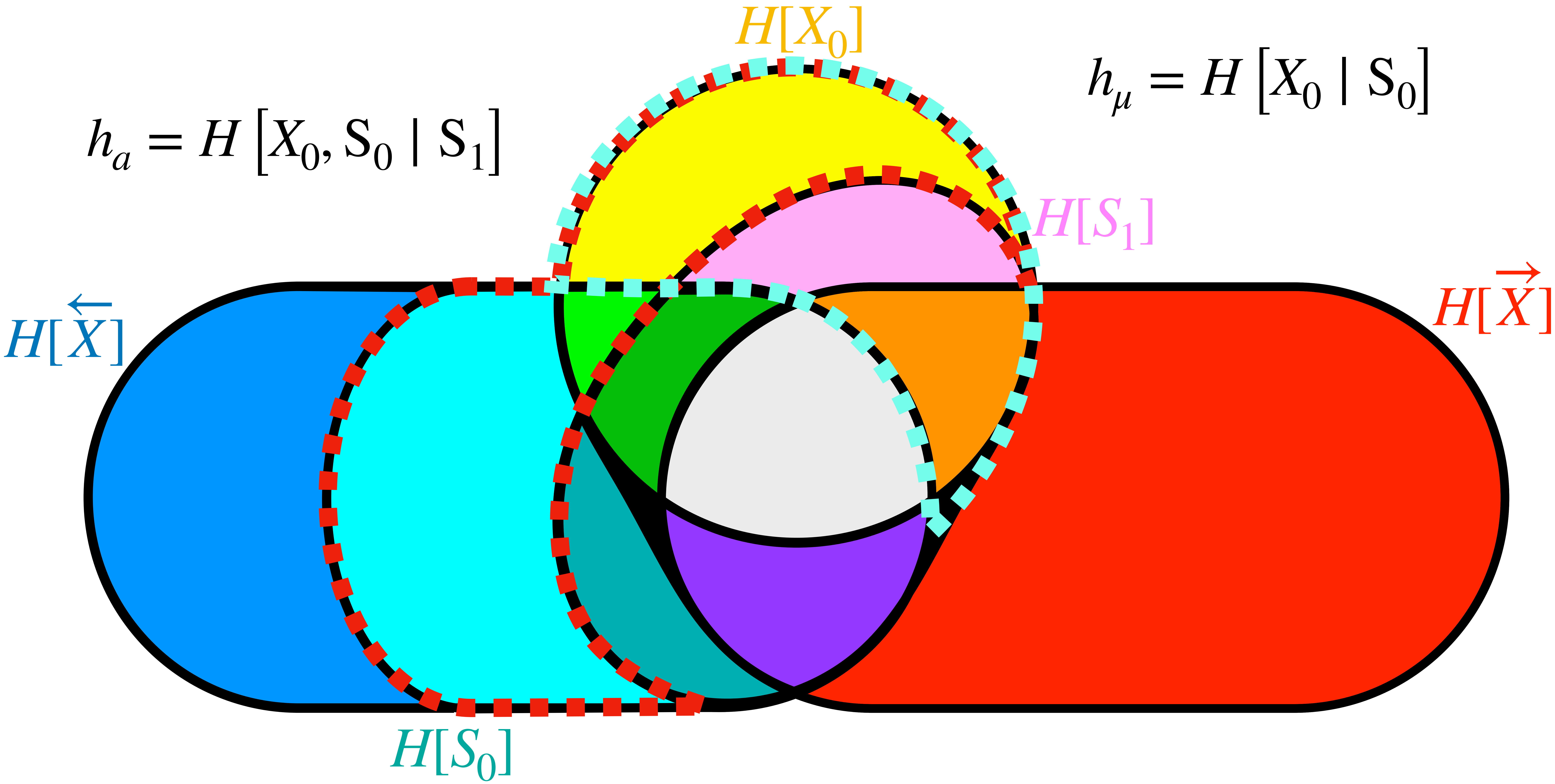
$$\begin{aligned} h_\mu - h_a &= H\left[X_t \mid S_t\right] - H\left[X_t, S_t \mid S_{t+1}\right] \\ &= H\left[S_{t+1} \mid S_t, X_t\right] + H\left[S_{t+1}\right] - H\left[S_t\right] \\ &= \Delta H\left[S\right] \end{aligned}$$

Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

What is h_a?

$$h_a = H\left[X_0, S_0 \mid S_1\right]$$

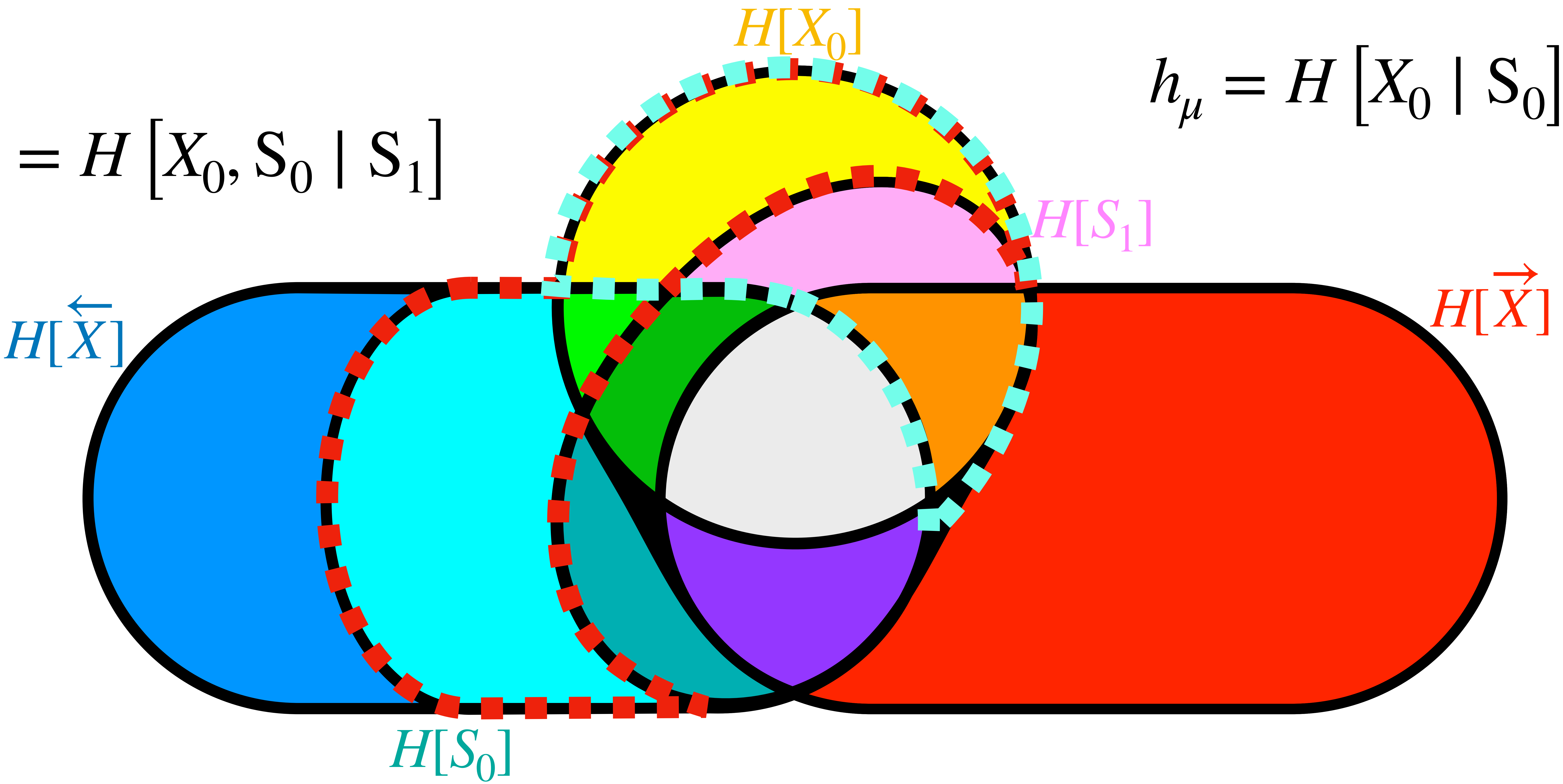
$$h_\mu = H\left[X_0 \mid S_0\right]$$



What is h_a?

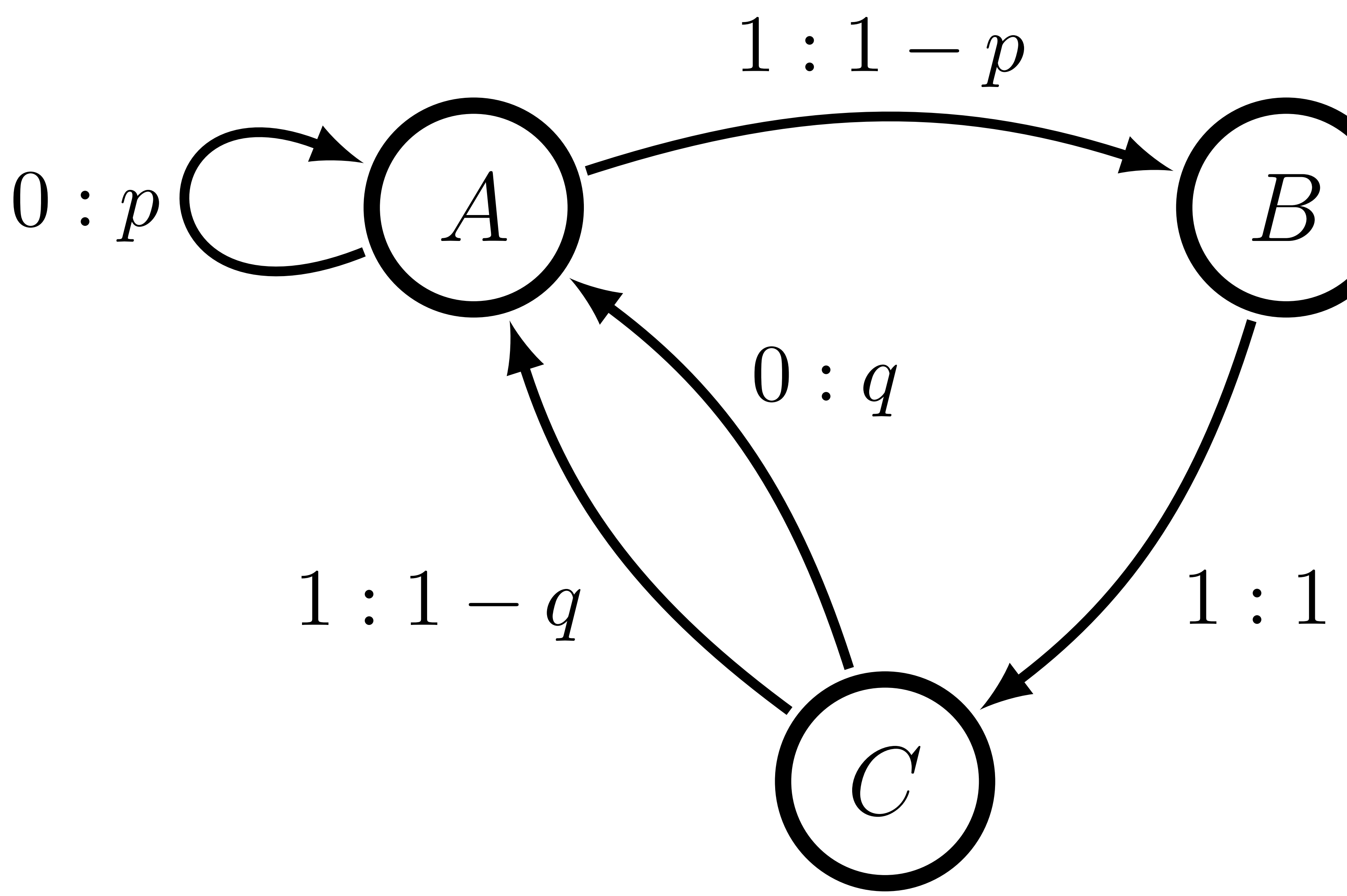
$$h_a = H\left[X_0, S_0 \mid S_1\right]$$

$$h_\mu = H\left[X_0 \mid S_0\right]$$



If $\Delta H\left[S\right] = 0$, then $h_\mu = h_a$

Growth Rate of Optimally Predictive Models

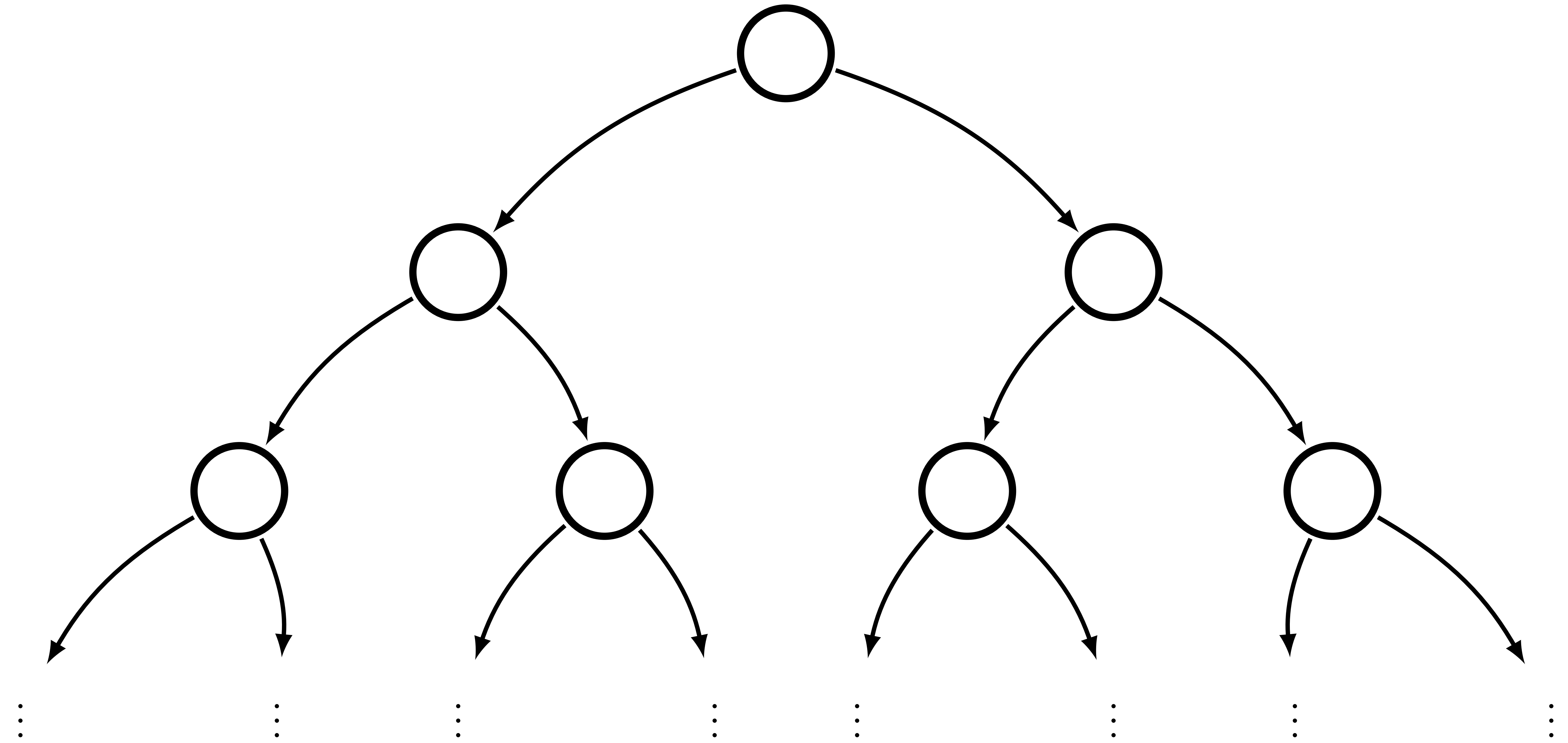


For a finite C_μ model:

$$h_\mu - h_a = 0$$

Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

Growth Rate of Optimally Predictive Models



No uncertainty in past:

$$h_\mu - h_a = h_\mu$$

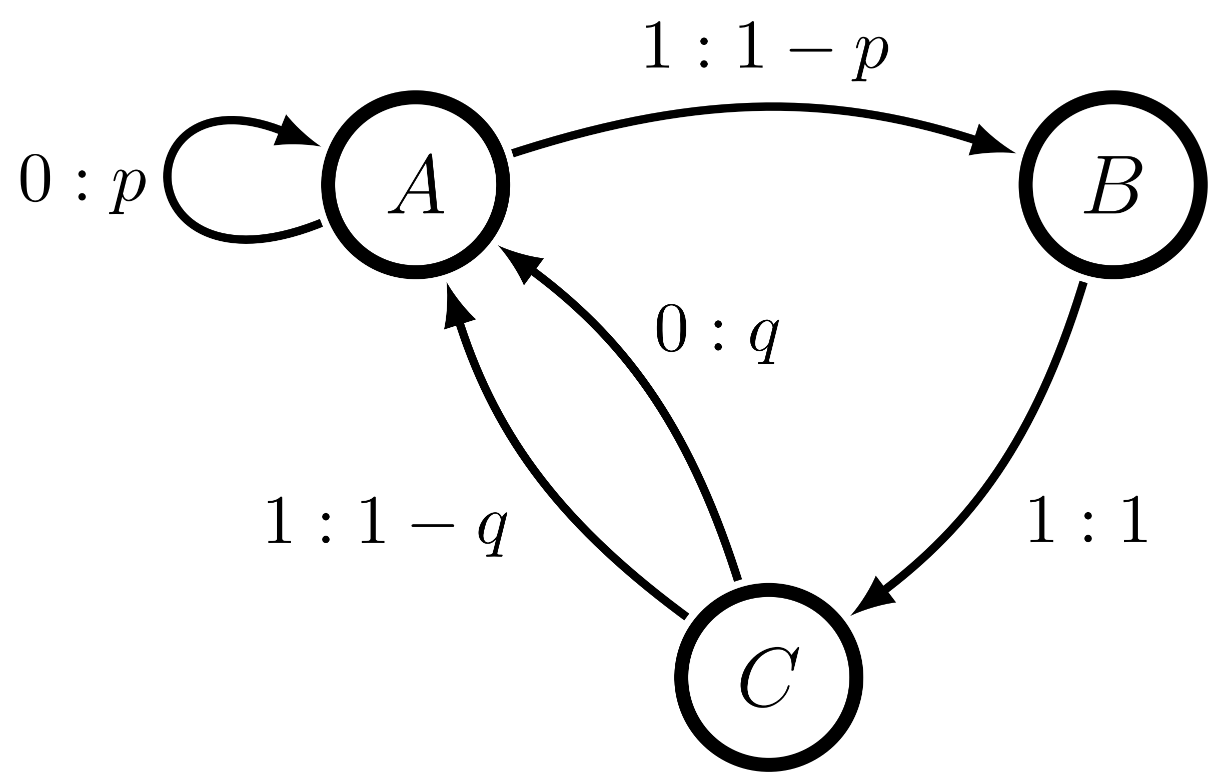
Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

Growth Rate of Optimally Predictive Models

$$\Delta H\left[\text{Predictive states}\right] = h_\mu - h_a$$

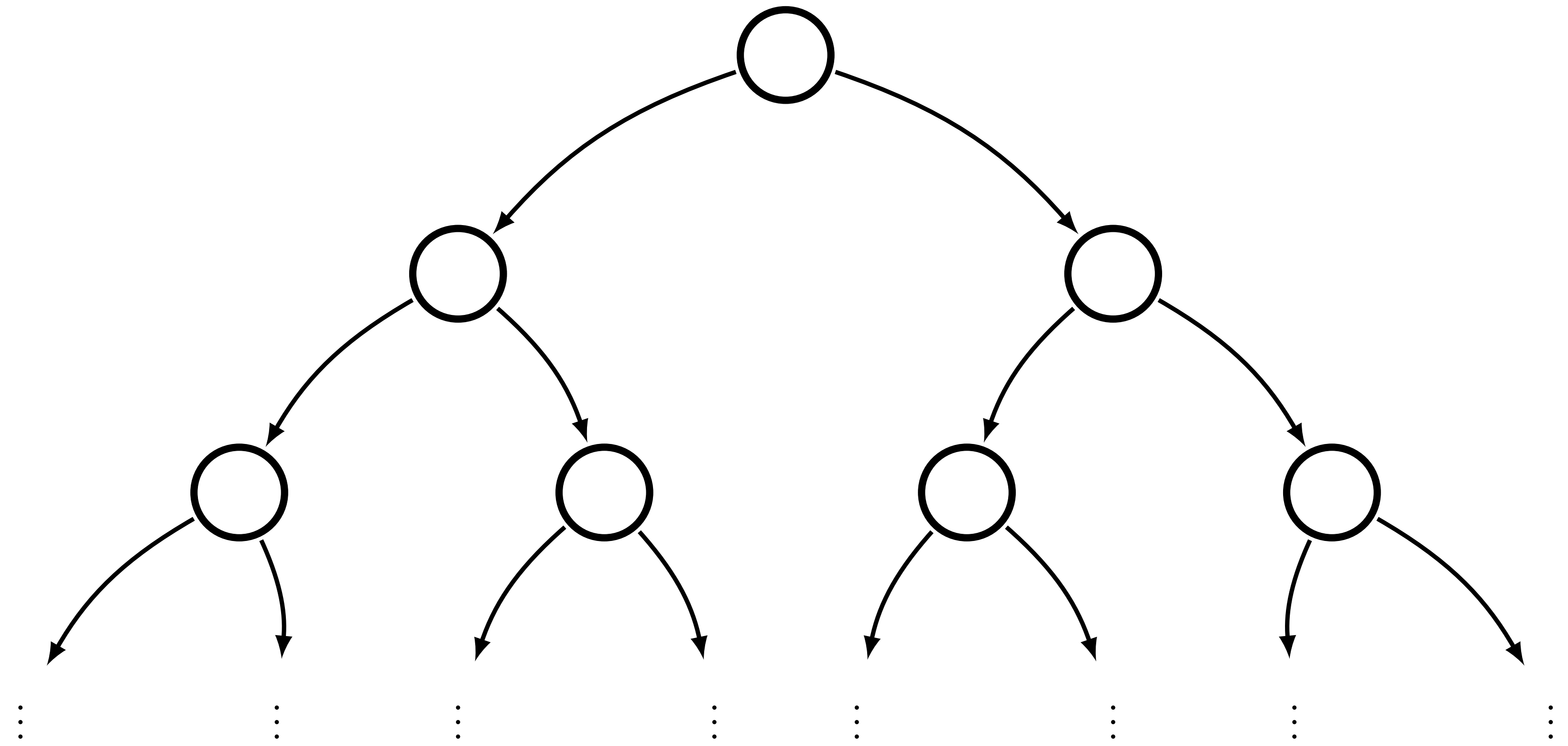
Finite states

$$h_\mu - h_a = 0$$



"Every history counts"

$$h_\mu - h_a = h_\mu$$



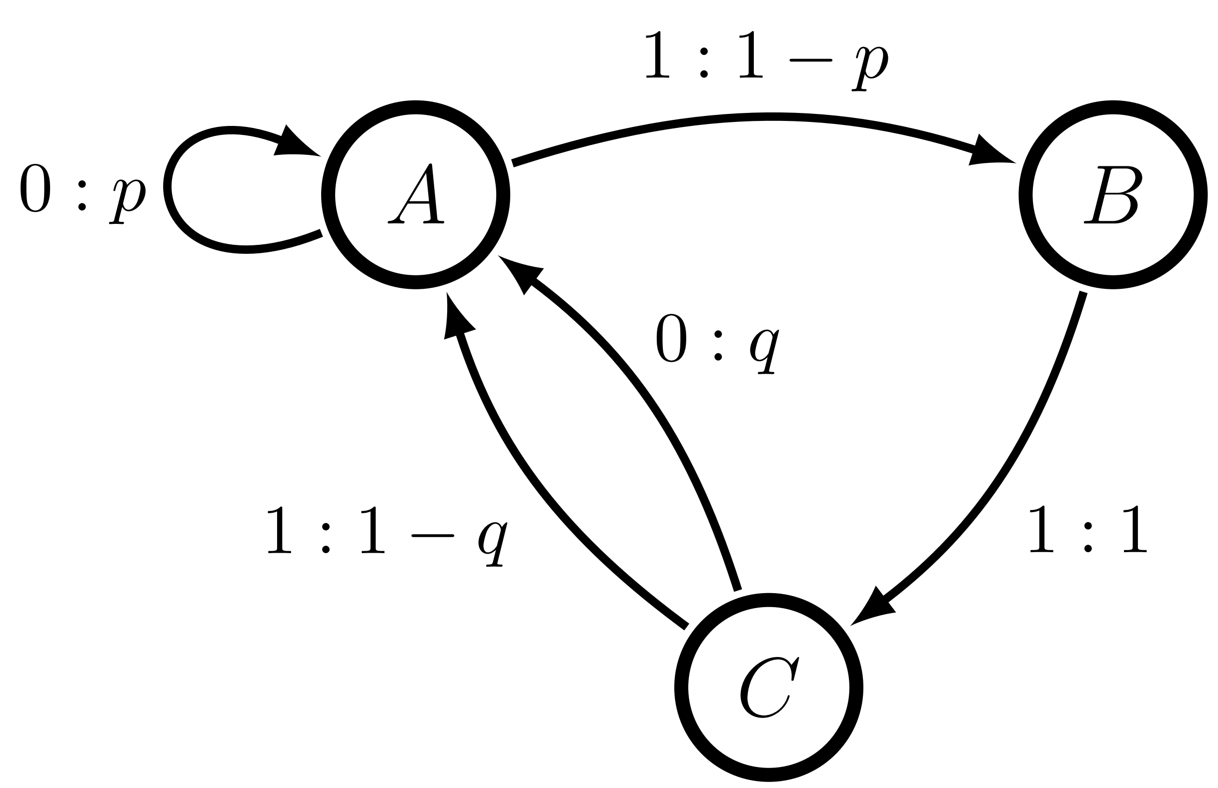
Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

Growth Rate of Optimally Predictive Models

$$\Delta H\left[\text{Predictive states}\right] = h_\mu - h_a$$

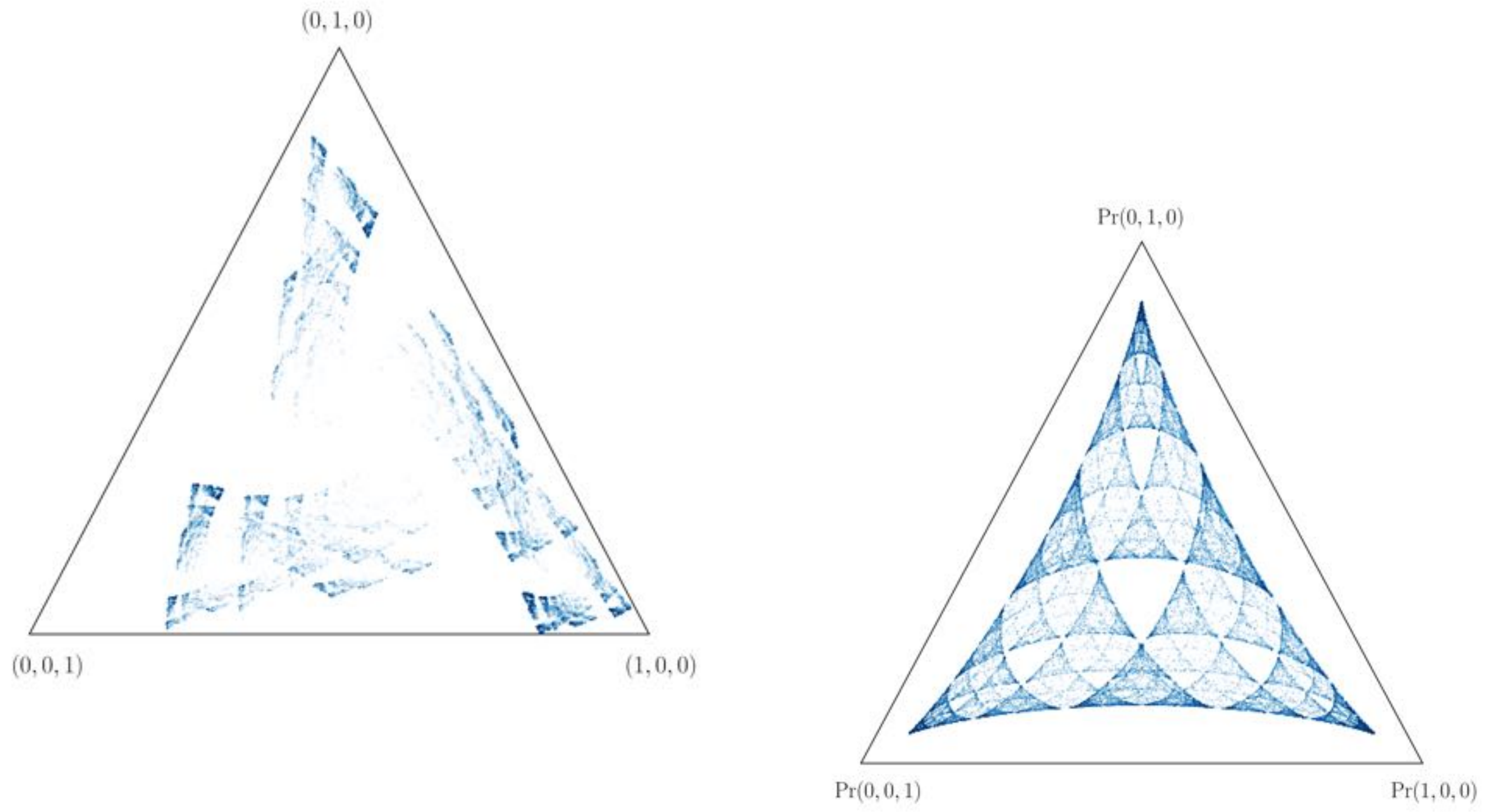
Finite states

$$h_\mu - h_a = 0$$



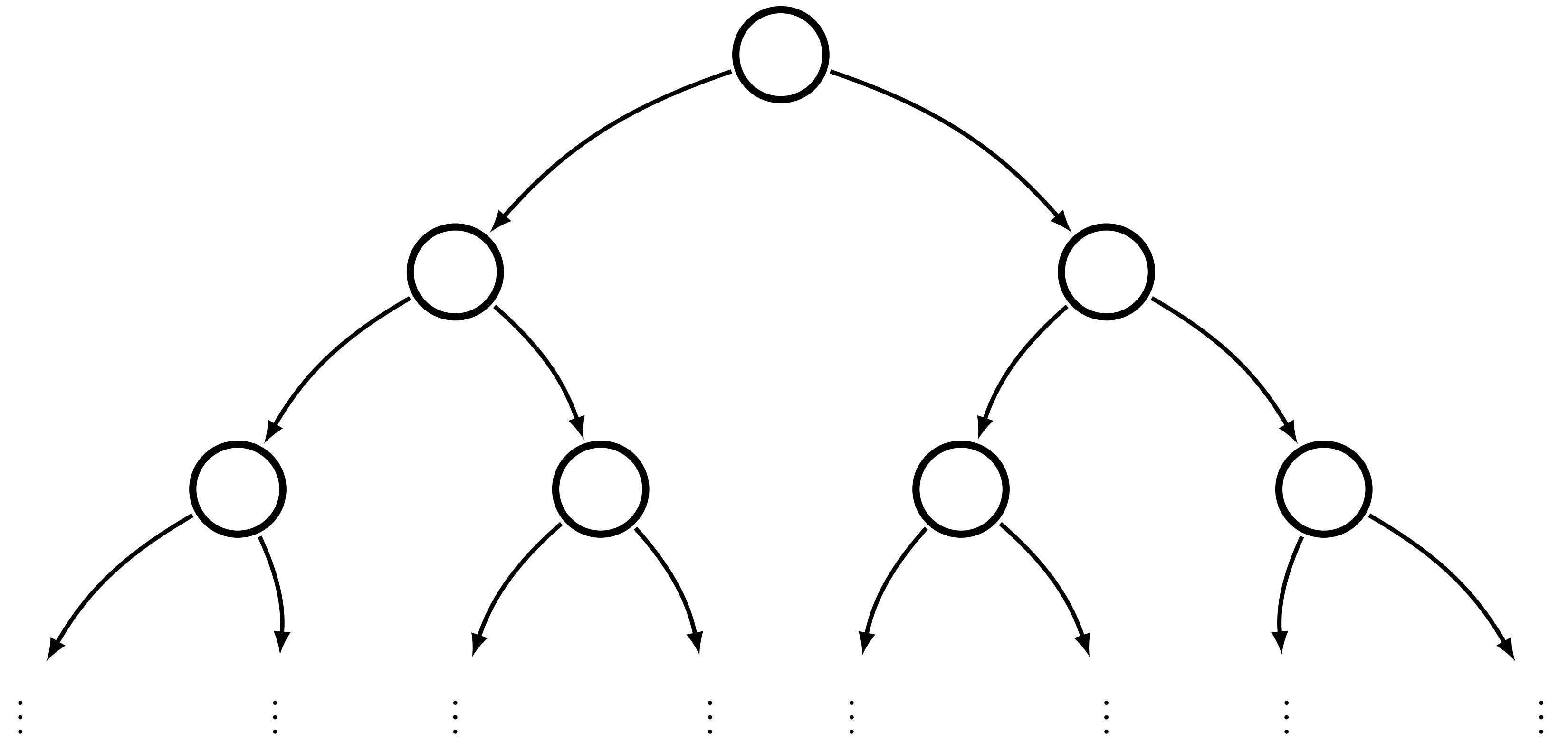
In general:

$$h_\mu > h_\mu - h_a > 0$$



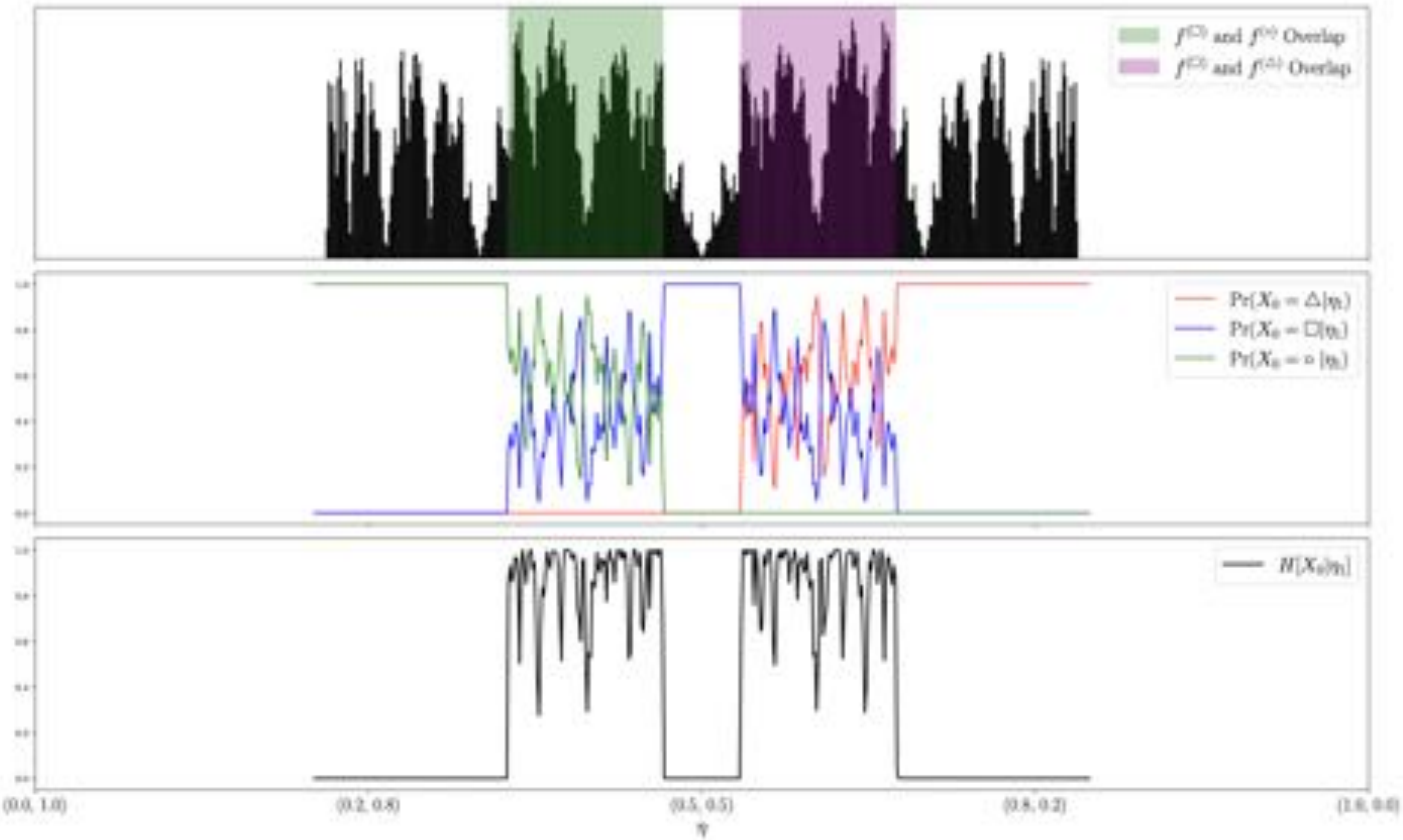
"Every history counts"

$$h_\mu - h_a = h_\mu$$



Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

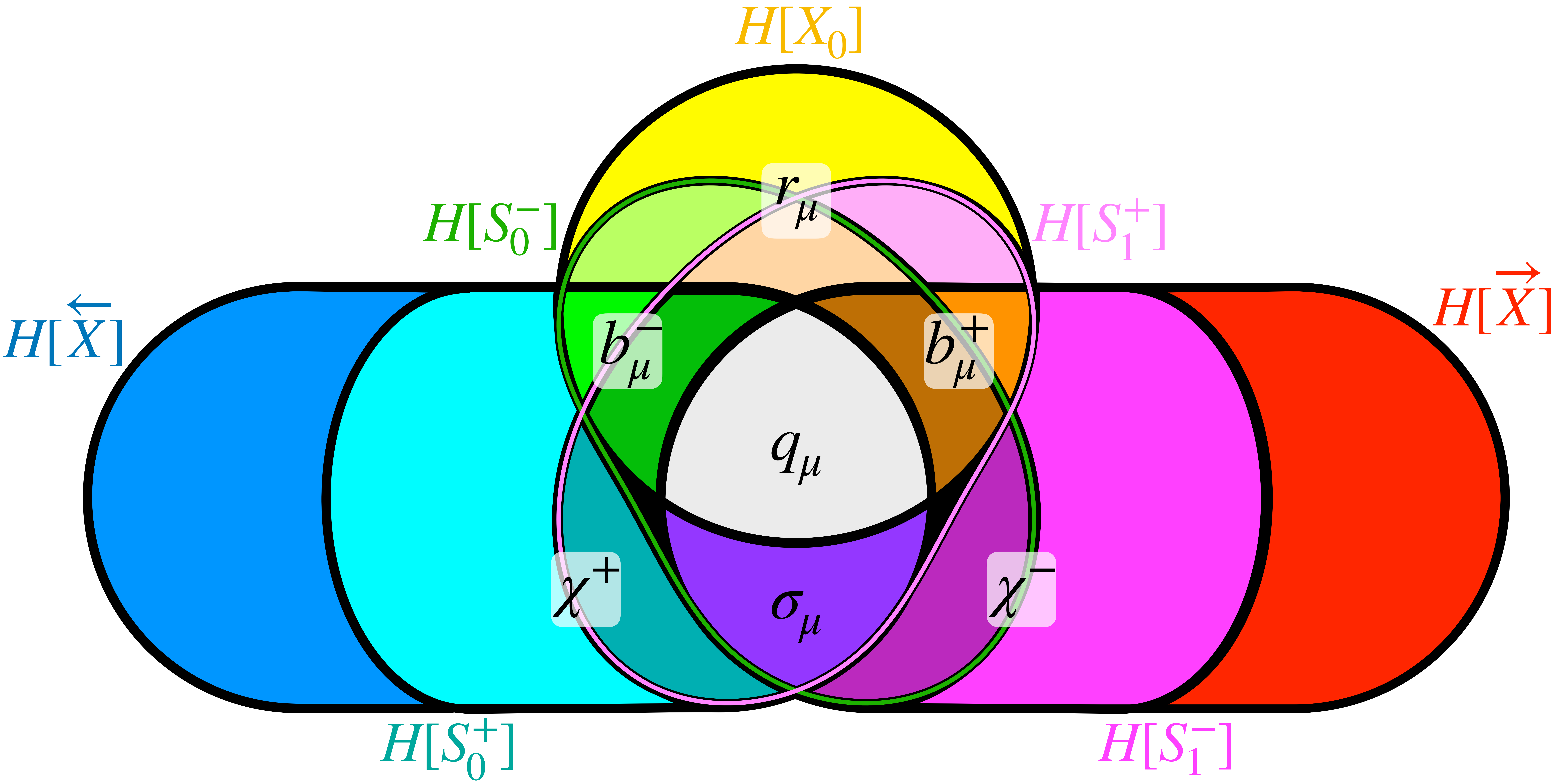
Calculation of Ambiguity Rate?



Ulam's method….

Alexandra M. Jurgens, James P. Crutchfield. *Ambiguity rate of hidden Markov processes*. Phys. Rev. E, 104 (2021)

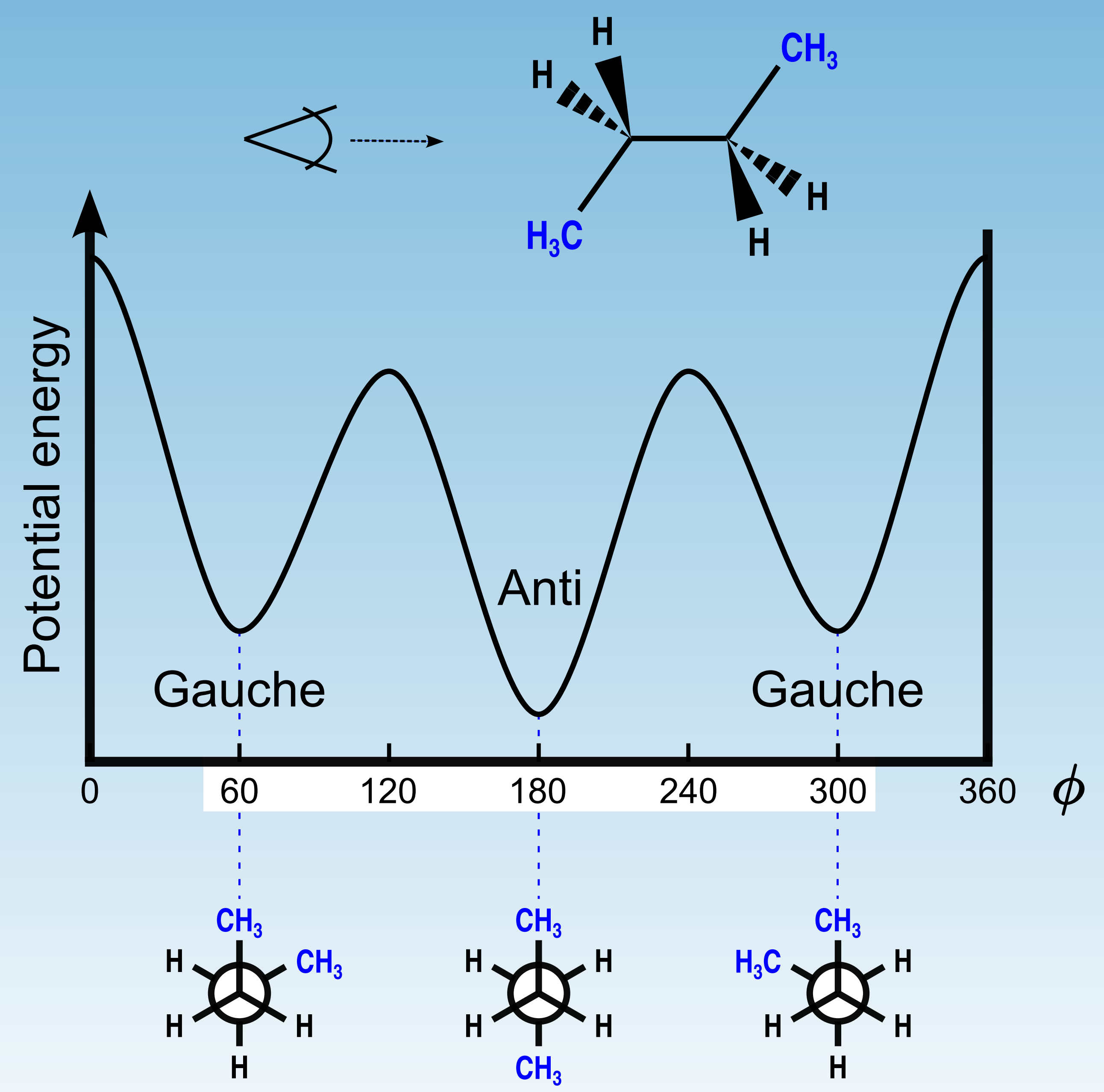
This work inspired…

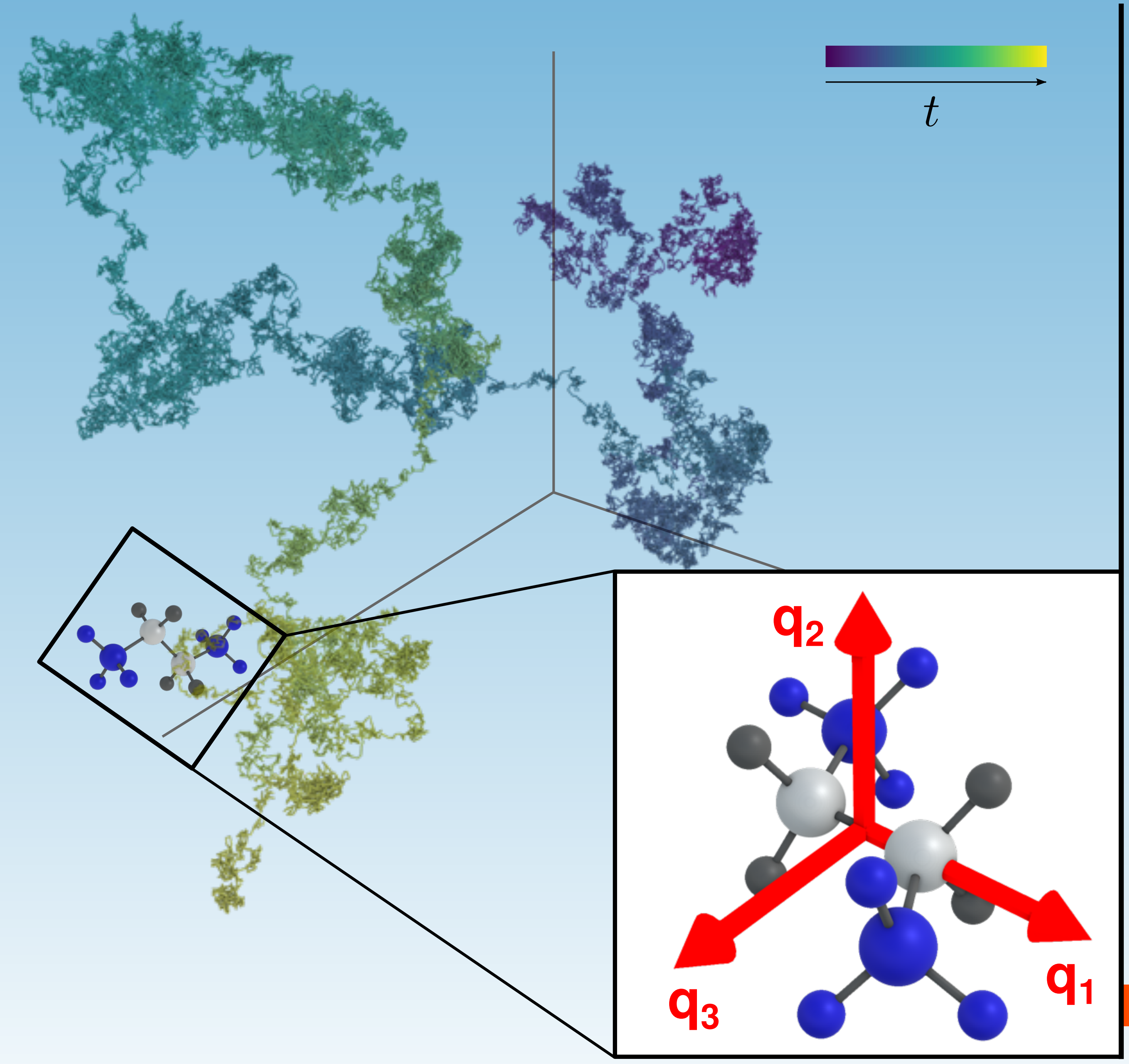


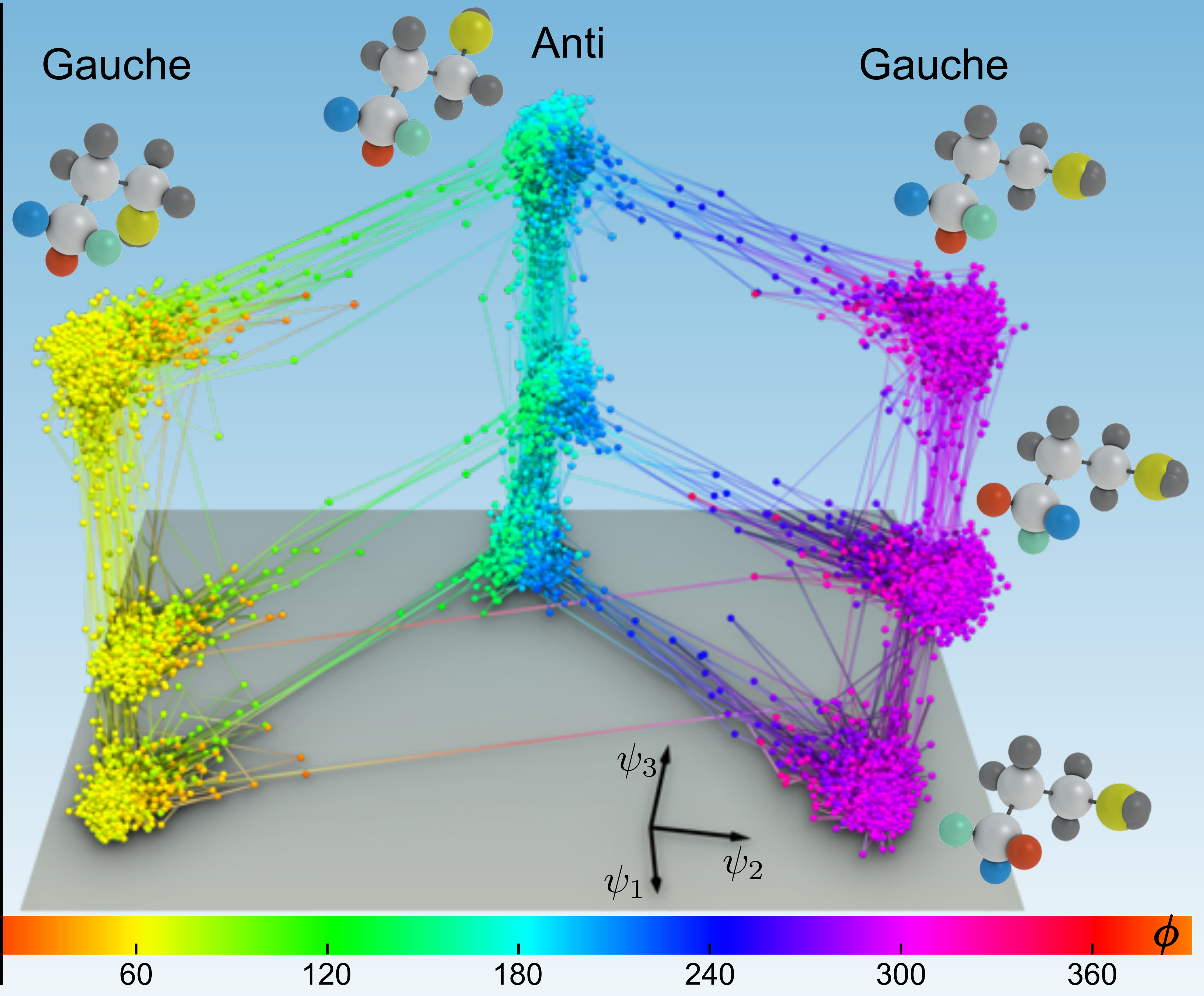
A Jurgens, JP Crutchfield. Taxonomy of Prediction. arXiv preprint arXiv:2504.11371, 2025

(Secret Bonus Talk)

Geometric Perspectives on ϵ-Machines







Context for this Talk

Geometry from a Time Series

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw
Dynamical Systems Collective, Physics Department, University of California, Santa Cruz, California 95064
(Received 13 November 1979)

It is shown how the existence of low-dimensional chaotic dynamical systems describing turbulent fluid flow might be determined experimentally. Techniques are outlined for reconstructing phase-space pictures from the observation of a single coordinate of any dissipative dynamical system, and for determining the dimensionality of the system's attractor. These techniques are applied to a well-known simple three-dimensional chaotic dynamical system.

PACS numbers: 47.25.-c

FIG. 1. (x,y) projection of Rossler (Ref. 7).

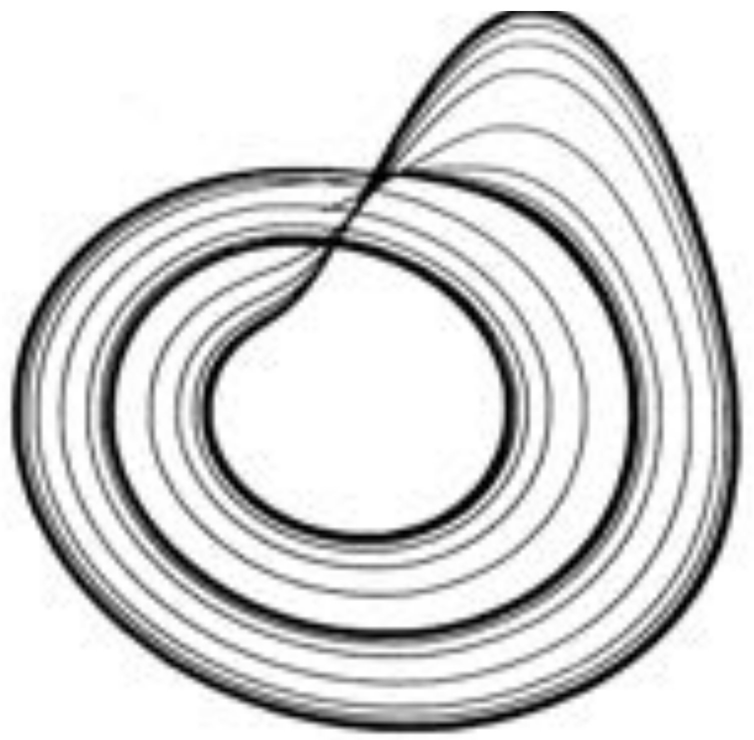


FIG. 2. $(x,\dot{x})$ reconstruction from the time series.

Continuous and deterministic dynamic → attractor reconstruction

Context for this Talk

Geometry from a Time Series

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw
Dynamical Systems Collective, Physics Department, University of California, Santa Cruz, California 9506
(Received 13 November 1979)

It is shown how the existence of low-dimensional chaotic dynamical systems describing turbulent fluid flow might be determined experimentally. Techniques are outlined for reconstructing phase-space pictures from the observation of a single coordinate of any dissipative dynamical system, and for determining the dimensionality of the system's attractor. These techniques are applied to a well-known simple three-dimensional chaotic dynamical system.

PACS numbers: 47.25.-c

FIG. 1. (x,y) projection of Rossler (Ref. 7).

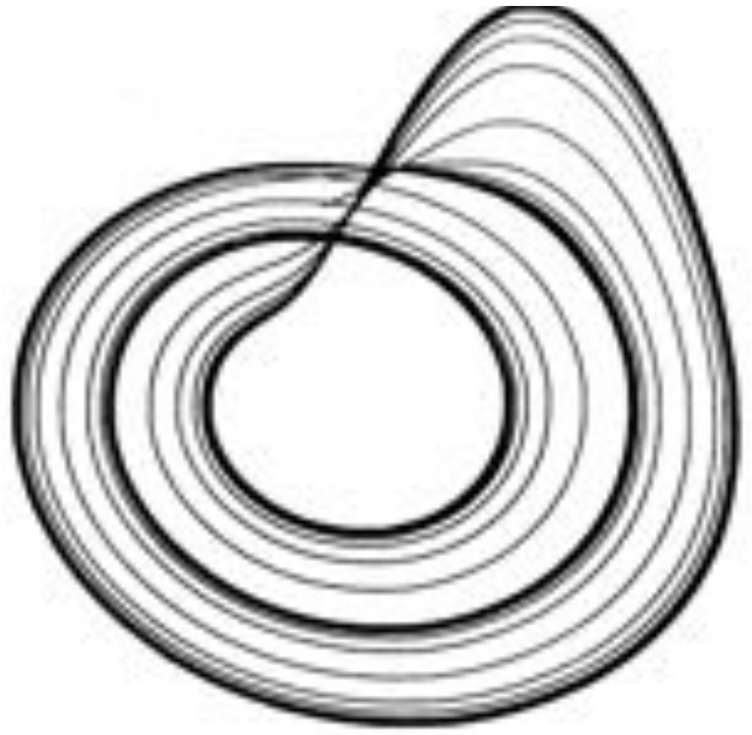


FIG. 2. $(x,\dot{x})$ reconstruction from the time series.

Continuous and deterministic dynamic → attractor reconstruction

Transformers Represent Belief State Geometry in their Residual Stream

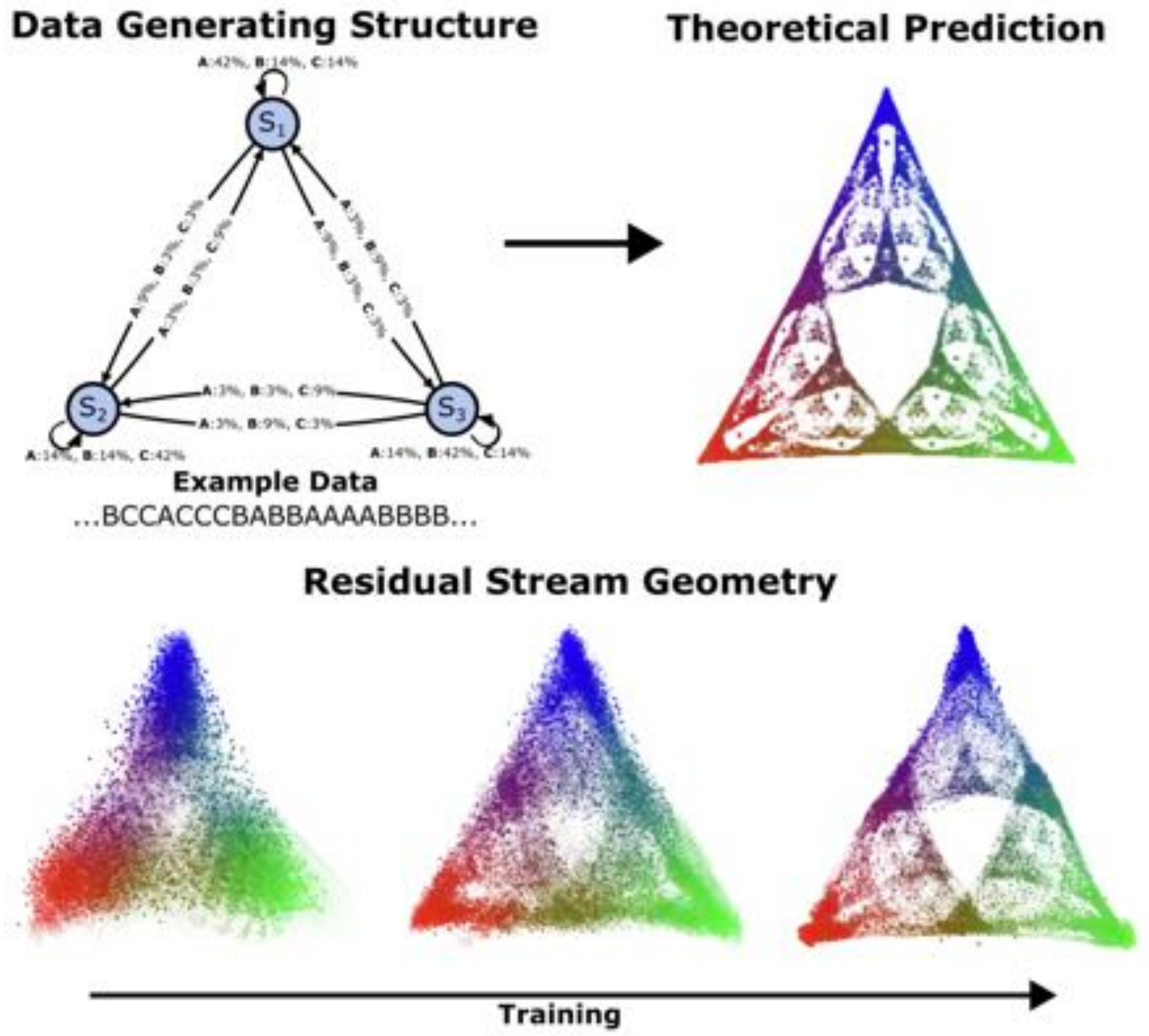
Adam S. Shai
Simplex
PIBBSS*

Sarah E. Marzen
W. M. Keck Science Department
Pitzer, Scripps, and Claremont McKenna College

Lucas Teixeira
PIBBSS

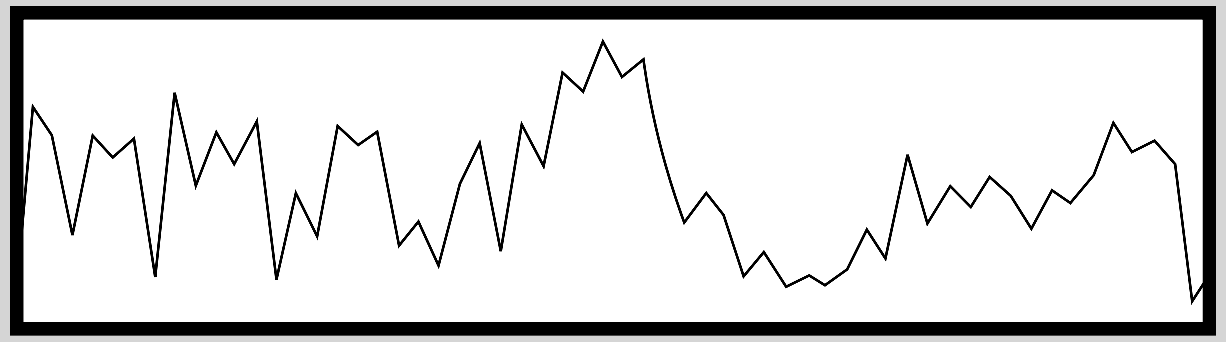
Alexander Gietelink Oldenziel
University College London
Timaeus

Paul M. Riechers
Simplex
BITS



Writing Down ϵ-Machines

Data



0100010000101110…

Lorem ipsum dolor sit amen…

⋮

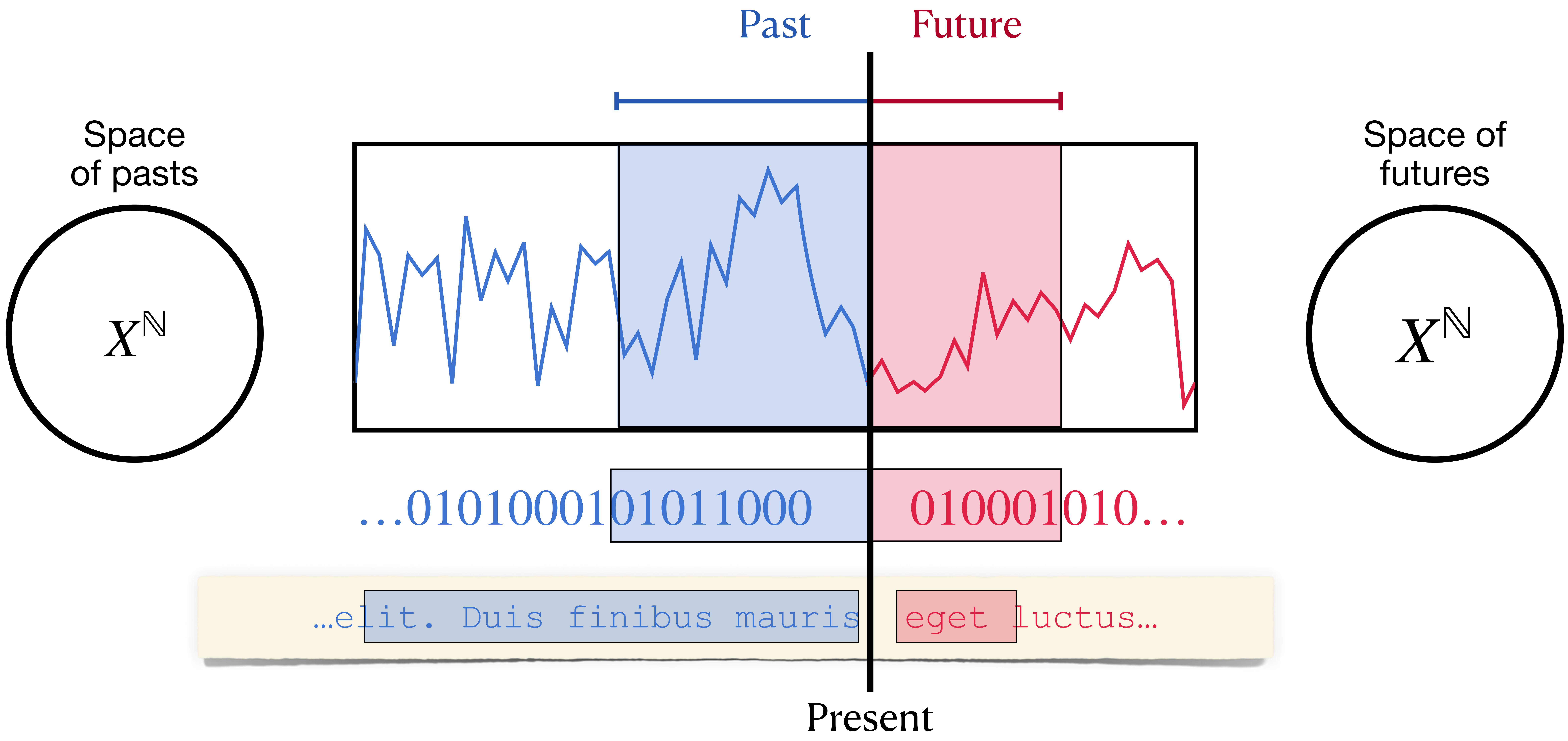
Causal States $\mathcal{S}$

- Via the predictive equivalence relation

ϵ-Machine

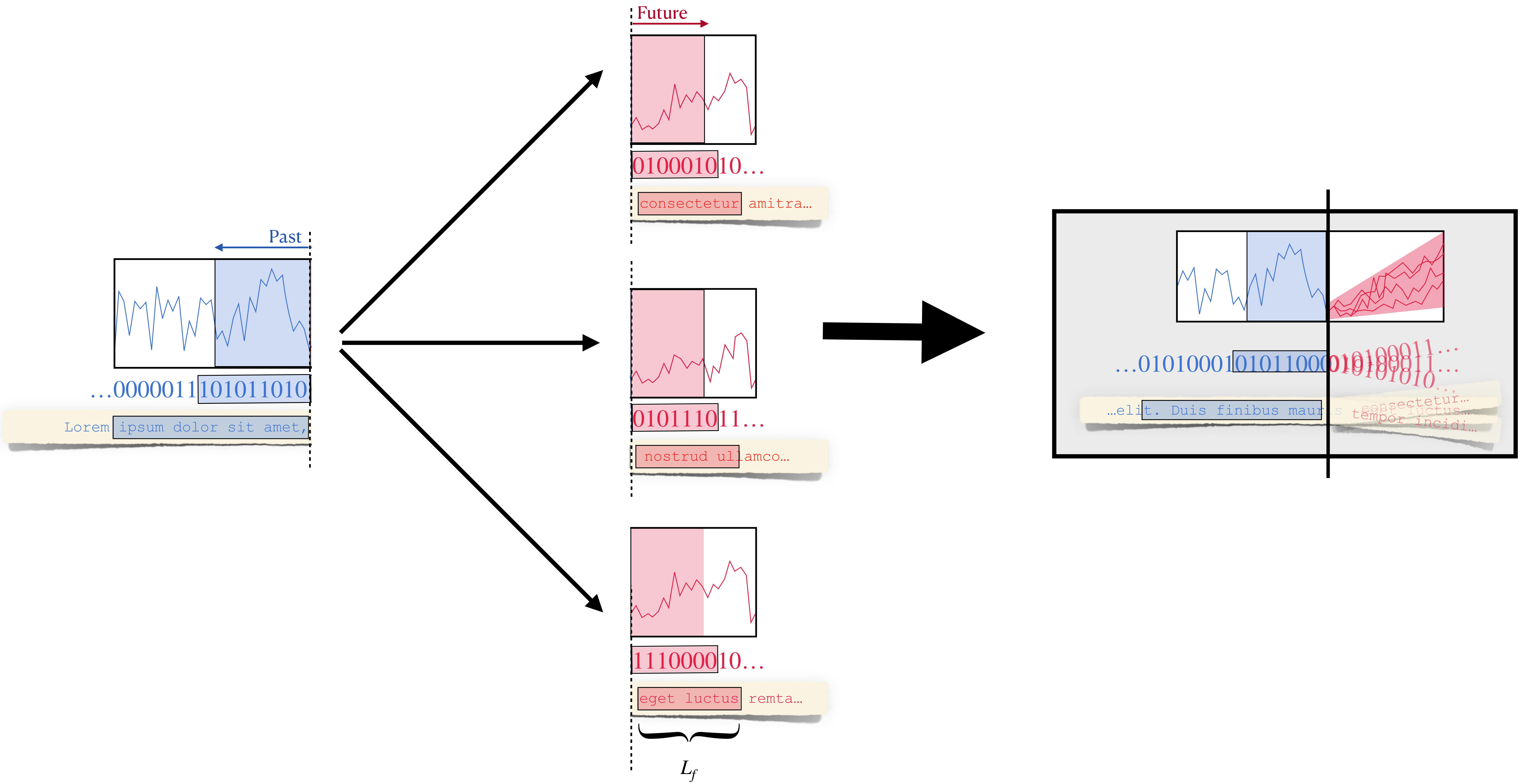
- Dynamic over the casual states

Pasts & Futures



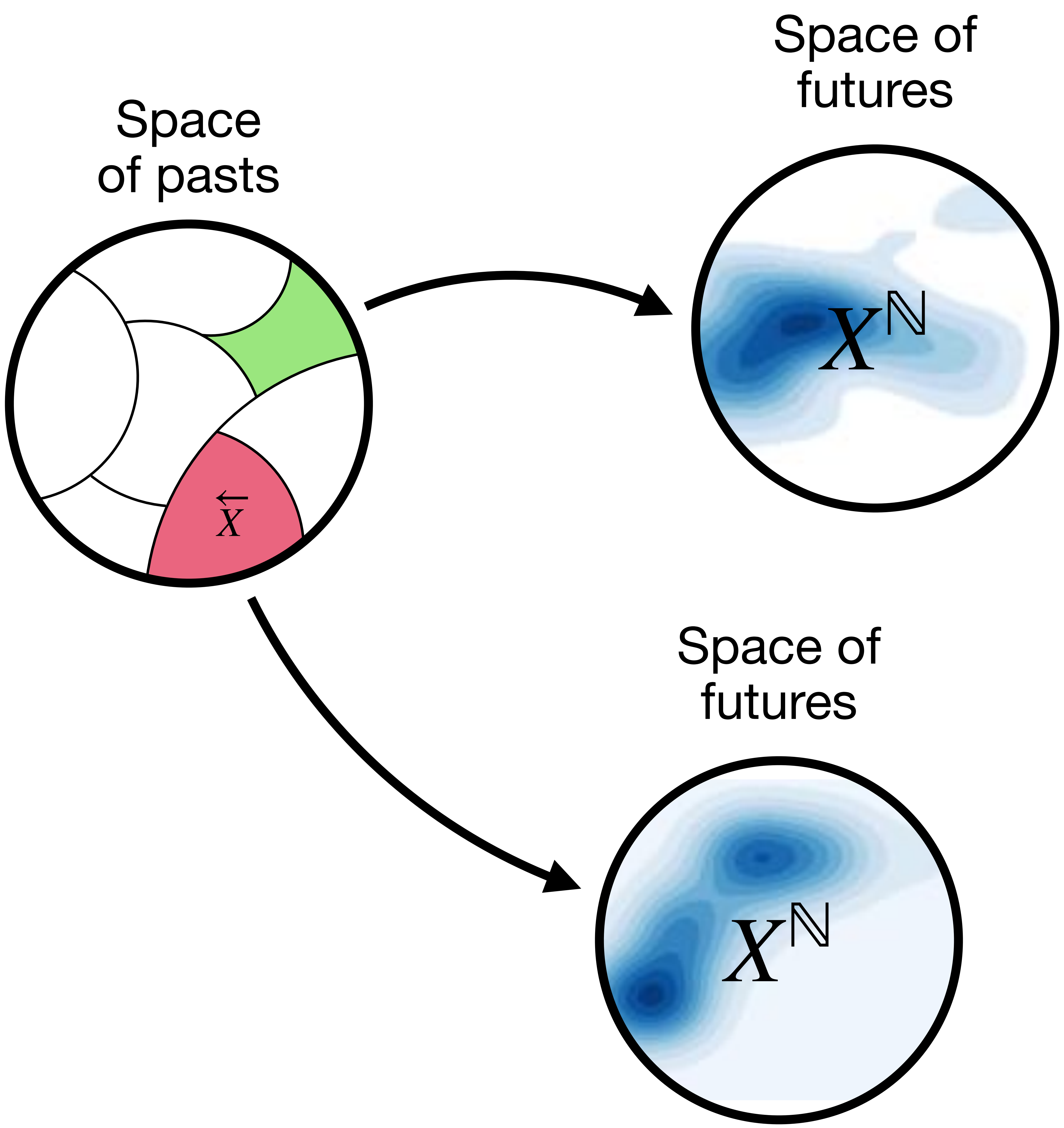
Shalizi, C.R., Crutchfield, J.P. Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *Journal of Statistical Physics* **104**, 817–879 (2001). https://doi.org/10.1023/A:1010388907793

Pasts Induce Distribution Over Futures



Past Partitioning

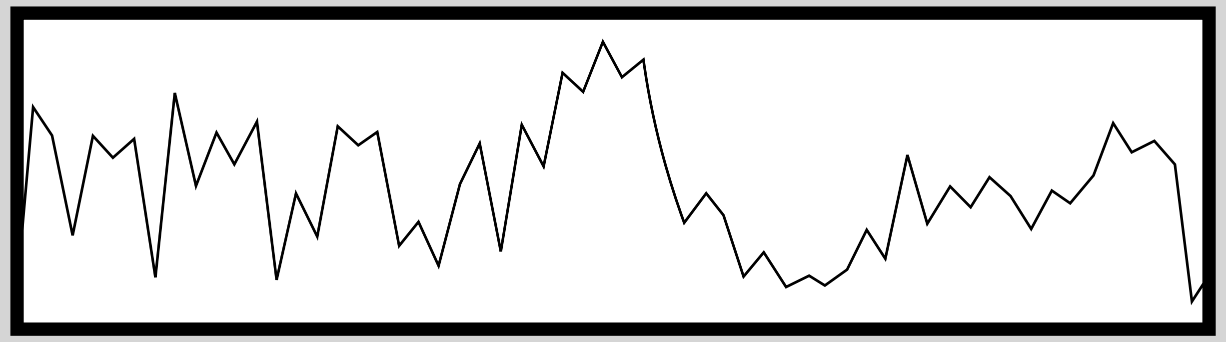
Predictive equivalence relation induces partition over pasts....



...where each element of the partition induces a unique distribution over futures.

Writing Down ϵ-Machines

Data



0100010000101110…

Lorem ipsum dolor sit amen…

⋮

Causal States S

- Via the predictive equivalence relation

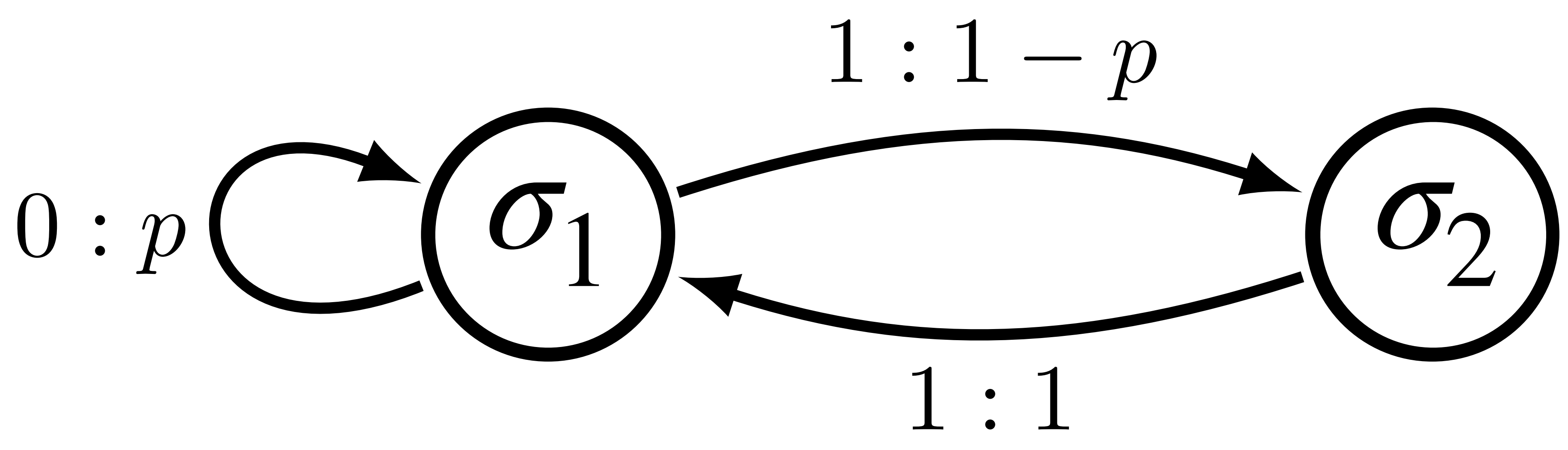
1. By inspection

ϵ-Machine

- Dynamic over the casual states

ϵ-Machine: Finite States

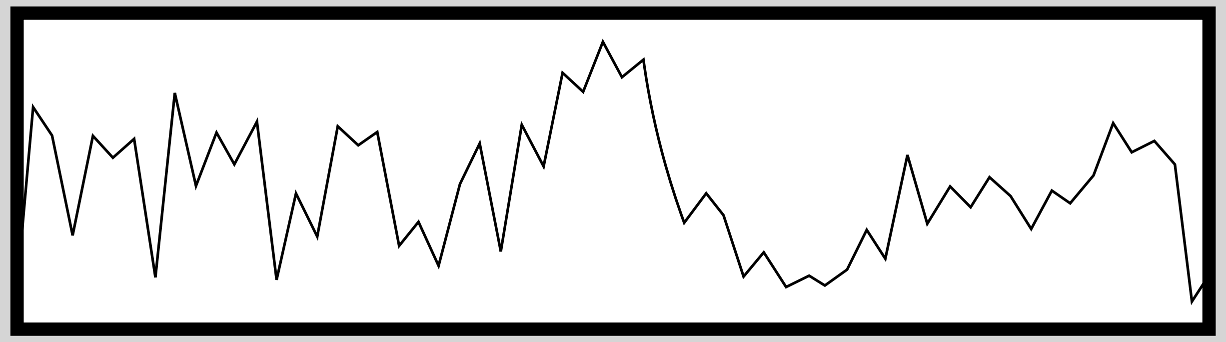
$011000001111011111000\ldots$ $\longrightarrow$



Causal states + dynamic over states: ϵ-machine

Writing Down ϵ-Machines

Data



0100010000101110…

Lorem ipsum dolor sit amen…

⋮

→

Causal States S

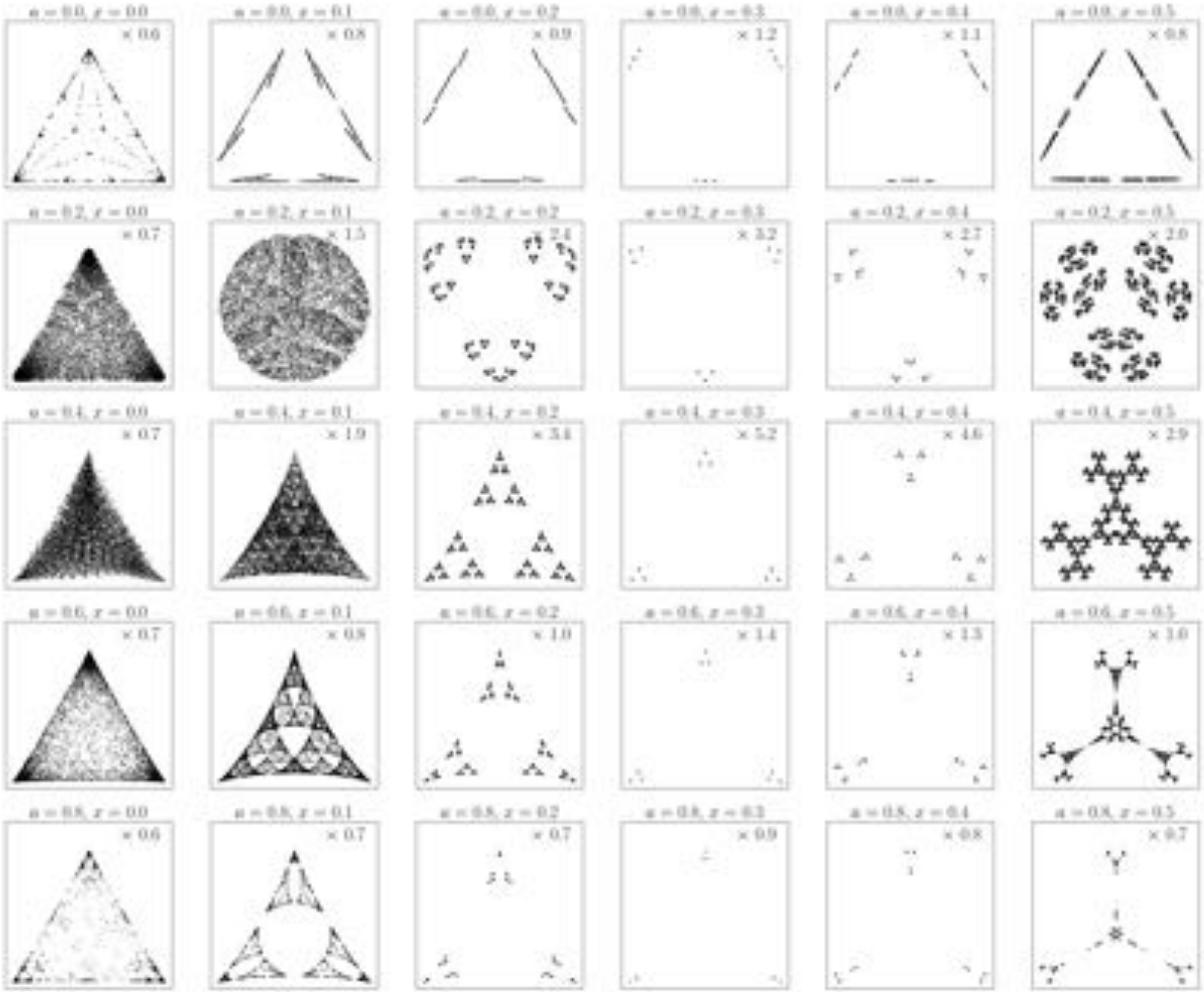
- Via the predictive equivalence relation

1. By inspection
2. By generation in the space spanned by a generating model

→

ϵ-Machine

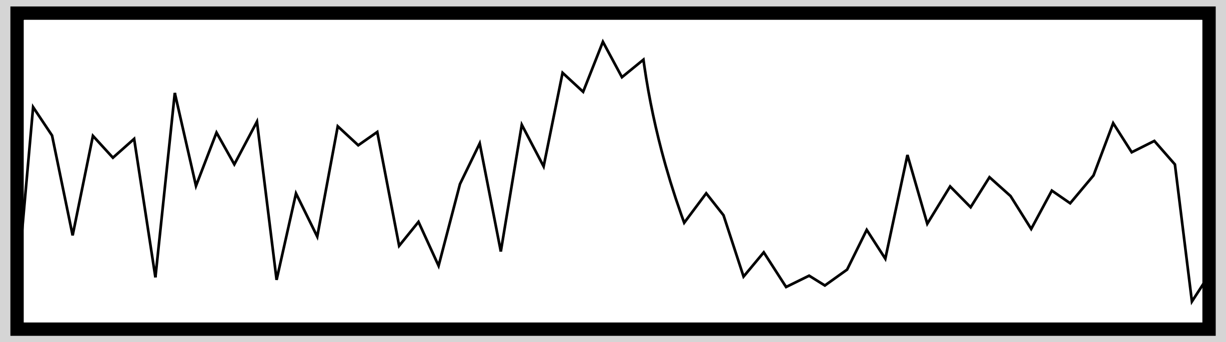
- Dynamic over the casual states



a = 0.0, r = 0.0
× 0.6
a = 0.0, r = 0.1
× 0.8
a = 0.0, r = 0.2
× 0.9
a = 0.0, r = 0.3
× 1.2
a = 0.0, r = 0.4
× 1.1
a = 0.0, r = 0.5
× 0.8
a = 0.2, r = 0.0
× 0.7
a = 0.2, r = 0.1
× 1.5
a = 0.2, r = 0.2
× 2.4
a = 0.2, r = 0.3
× 3.2
a = 0.2, r = 0.4
× 2.7
a = 0.2, r = 0.5
× 2.0
a = 0.4, r = 0.0
× 0.7
a = 0.4, r = 0.1
× 1.9
a = 0.4, r = 0.2
× 3.4
a = 0.4, r = 0.3
× 5.2
a = 0.4, r = 0.4
× 4.6
a = 0.4, r = 0.5
× 2.9
a = 0.6, r = 0.0
× 0.7
a = 0.6, r = 0.1
× 0.8
a = 0.6, r = 0.2
× 1.0
a = 0.6, r = 0.3
× 1.4
a = 0.6, r = 0.4
× 1.3
a = 0.6, r = 0.5
× 1.0
a = 0.8, r = 0.0
× 0.6
a = 0.8, r = 0.1
× 0.7
a = 0.8, r = 0.2
× 0.7
a = 0.8, r = 0.3
× 0.9
a = 0.8, r = 0.4
× 0.8
a = 0.8, r = 0.5
× 0.7

Writing Down ϵ-Machines

Data



0100010000101110…

```
Lorem ipsum dolor sit amen…
```

⋮

Causal States S

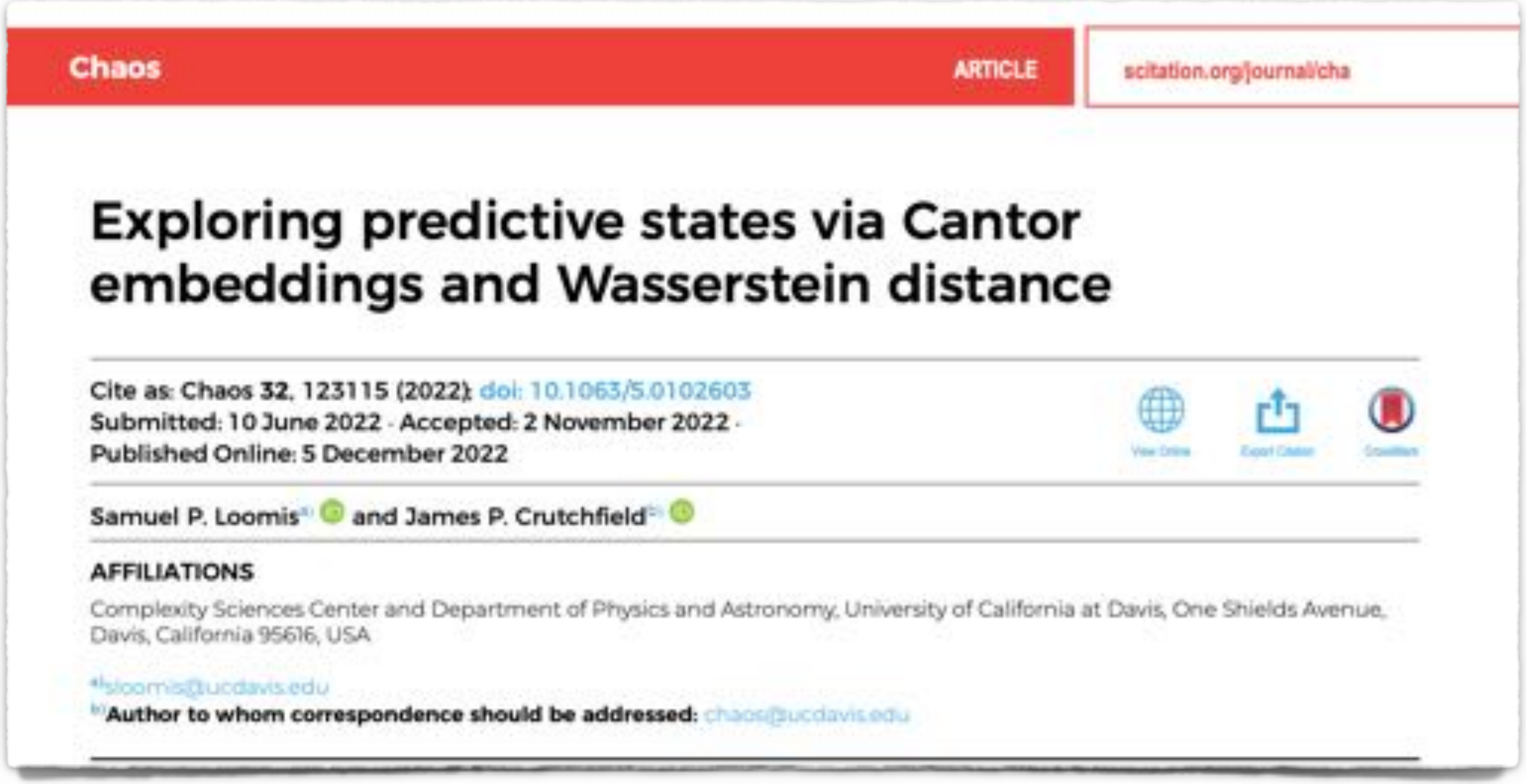
- Via the predictive equivalence relation

1. By inspection
2. By generation in the space spanned by a generating model
3. By embedding in a space

ϵ-Machine

- Dynamic over the casual states

Cantor Embedded ϵ-Machines



Chaos ARTICLE scitation.org/journal/cha

Exploring predictive states via Cantor embeddings and Wasserstein distance

Cite as: Chaos 32, 123115 (2022); doi: 10.1063/5.0102603
Submitted: 10 June 2022 · Accepted: 2 November 2022 ·
Published Online: 5 December 2022

Samuel P. Loomis[a] and James P. Crutchfield[b]

AFFILIATIONS

Complexity Sciences Center and Department of Physics and Astronomy, University of California at Davis, One Shields Avenue, Davis, California 95616, USA

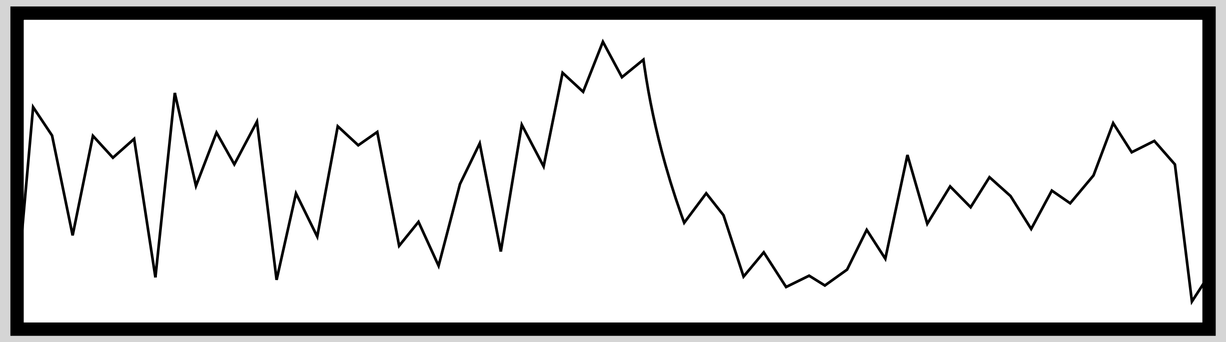
[a]sloomis@ucdavis.edu
[b]Author to whom correspondence should be addressed: chaos@ucdavis.edu

Given a sequence of symbols $x_1, x_2, x_3 \dots$ where each x_k is drawn from the alphabet $X = \{0,1,2,\dots,N\}$:

$$C\left(x_1, x_2, \dots\right) = \sum_{k=1}^{\infty} \frac{2x_k}{\left(2|X| - 1\right)^k}$$

Writing Down ϵ-Machines

Data



0100010000101110…

Lorem ipsum dolor sit amen…

⋮

Causal States S

- Via the predictive equivalence relation

1. By inspection
2. By generation in the space spanned by a generating model
3. By embedding in a space
   A. Which space? What metric properties do we need to recreate the predictive equivalence relation?
   B. Dimension reduction?

ϵ-Machine

- Dynamic over the casual states

ϵ-Machines in a Hilbert space

Chaos ARTICLE scitation.org/journal

Discovering causal structure with reproducing-kernel Hilbert space ϵ-machines

Cite as: Chaos 32, 023103 (2022); doi: 10.1063/5.0062829
Submitted: 8 July 2021 · Accepted: 10 January 2022 ·
Published Online: 1 February 2022 · Publisher error corrected: 28 March 2022

Nicolas Brodu[1,a] and James P. Crutchfield[2,b]

AFFILIATIONS

[1]Geostat Team—Geometry and Statistics in Acquisition Data, INRI
33405 Talence Cedex, France
[2]Complexity Sciences Center and Department of Physics and Astr
Davis, California 95616, USA

[a]Author to whom correspondence should be addressed: nicola
[b]Electronic mail: chaos@ucdavis.edu

Chaos ARTICLE pubs.aip.org/aip/cha

Inferring kernel ϵ-machines: Discovering structure in complex systems

Cite as: Chaos 35, 033162 (2025); doi: 10.1063/5.0242981
Submitted: 8 October 2024 · Accepted: 10 March 2025 ·
Published Online: 31 March 2025

Alexandra M. Jurgens[a] and Nicolas Brodu[b]

AFFILIATIONS

INRIA Bordeaux Sud Ouest, 33405 Talence Cedex, France

[a]alexandra.jurgens@inria.fr
[b]Author to whom correspondence should be addressed: nicolas.brodu@inria.fr

Contents
Physica D
journal homepage: www.elsevier.com/locate/physd
ELSEVIER

Topology, convergence, and reconstruction of predictive states

Samuel P. Loomis, James P. Crutchfield *

Complexity Sciences Center and Department of Physics and Astronomy, University of California at Davis, One Shields Avenue, Davis, CA 95616, United States of America

ARTICLE INFO

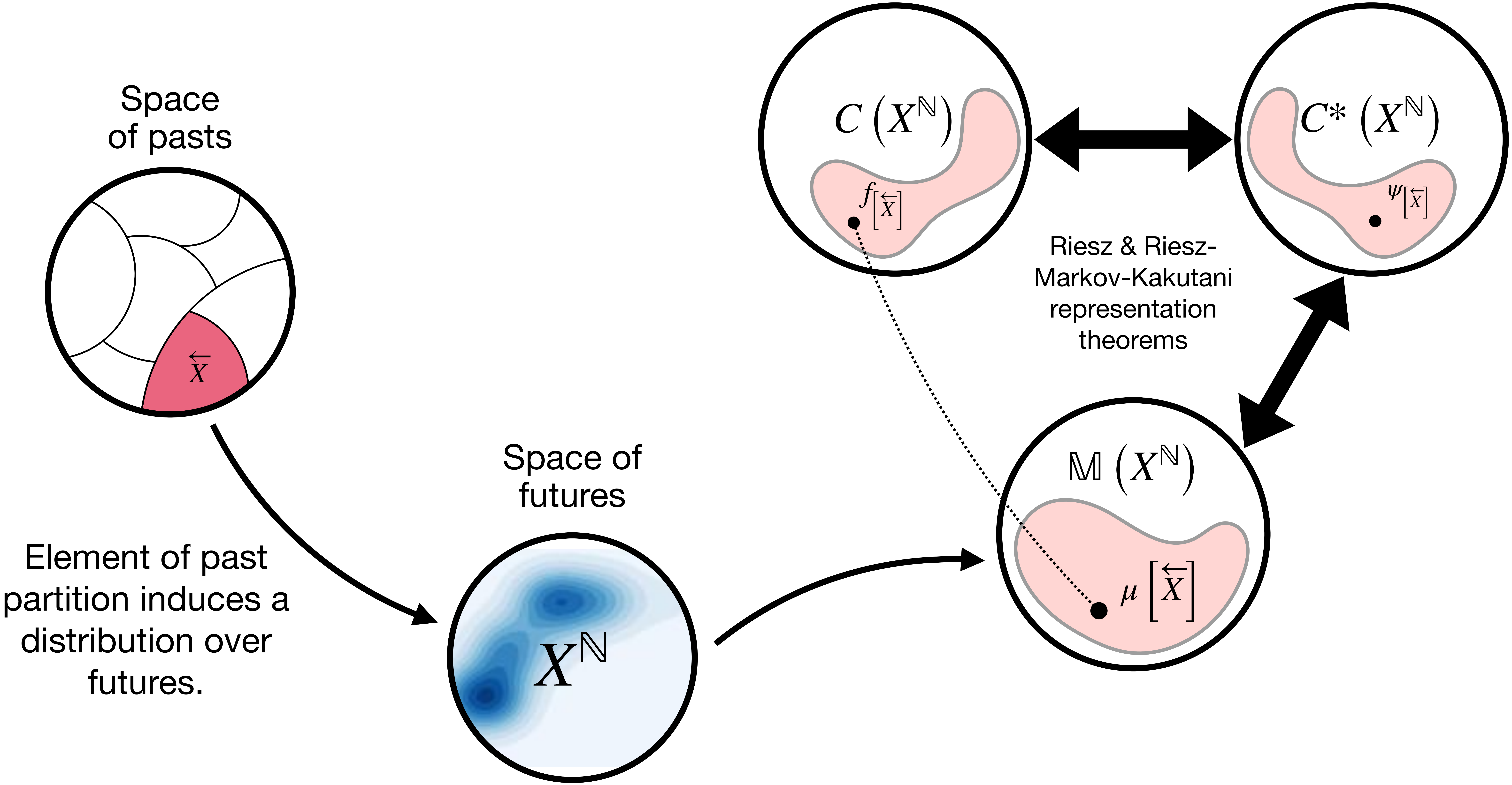
Article history:
Received 20 September 2021
Received in revised form 10 December 2022
Accepted 17 December 2022
Available online 3 January 2023
Communicated by G. Froyland

Keywords:
Stochastic process
Symbolic dynamics
Dynamical systems

ABSTRACT

Predictive equivalence in discrete stochastic processes has been applied with great success to identify randomness and structure in statistical physics and chaotic dynamical systems and to inferring hidden Markov models. We examine the conditions under which predictive states can be reliably reconstructed from time-series data, showing that convergence of predictive states can be achieved from empirical samples in the weak topology of measures. Moreover, predictive states may be represented in Hilbert spaces that replicate the weak topology. We mathematically explain how these representations are particularly beneficial when reconstructing high-memory processes and connect them to reproducing kernel Hilbert spaces.

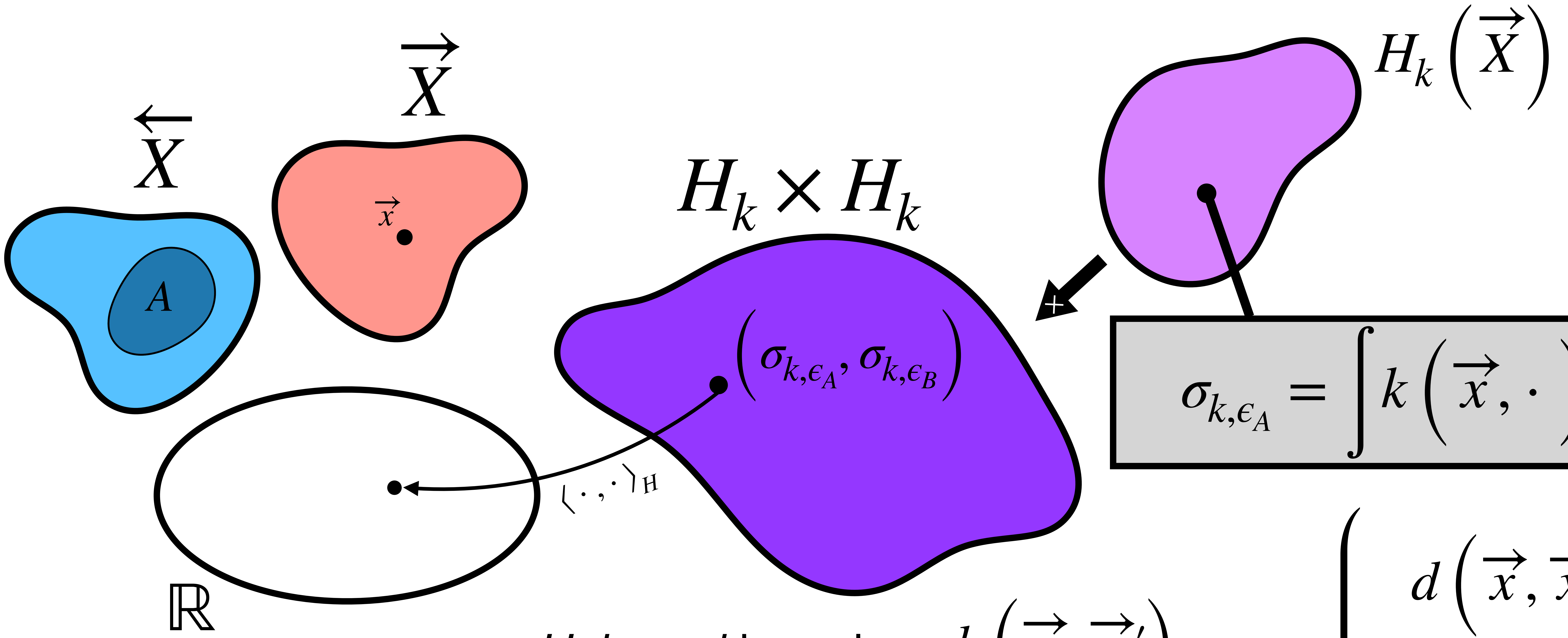
Causal States in a Hilbert space



Kernel Causal States

Kernel Causal States

Kernel Causal States

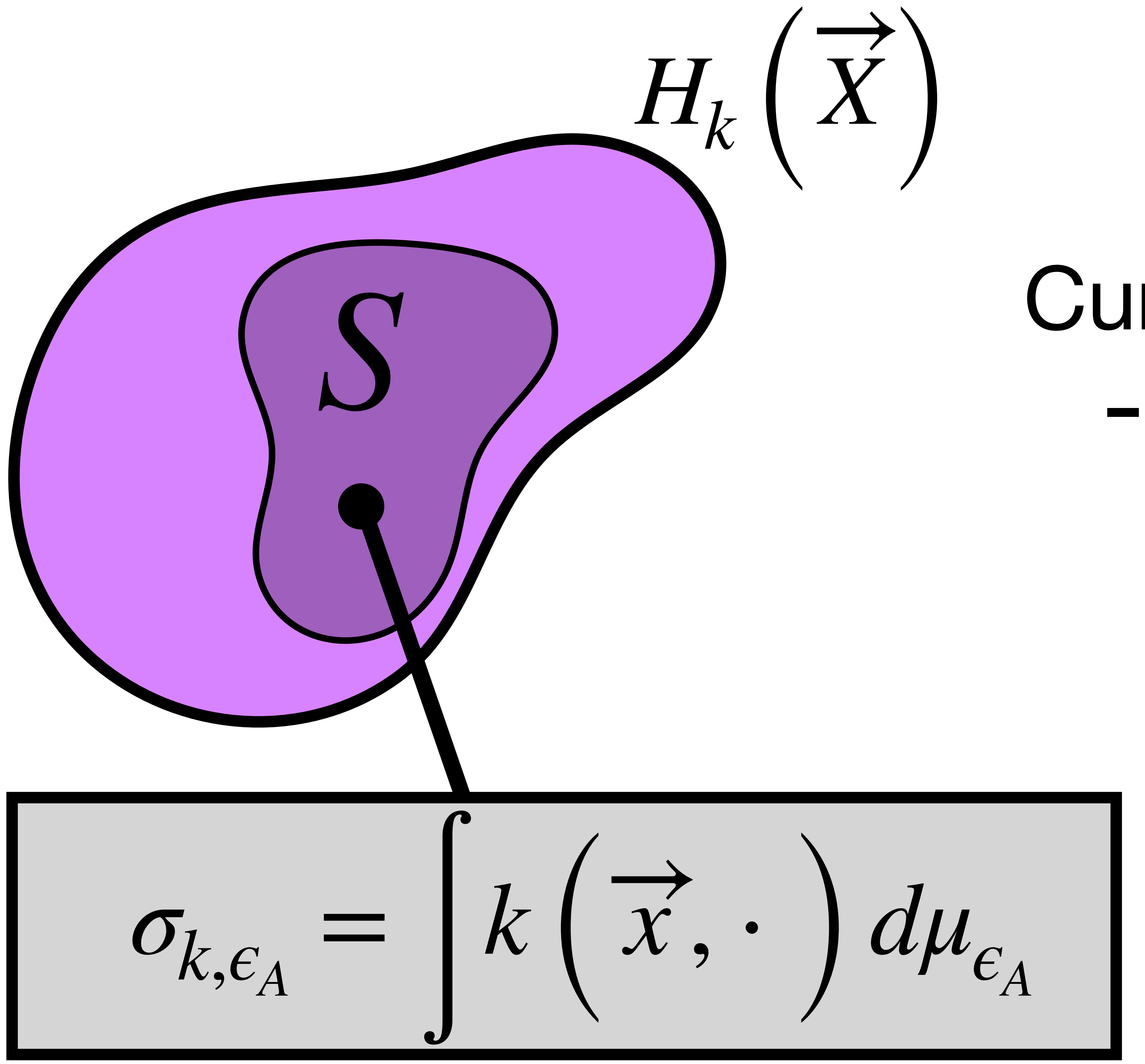


$$\sigma_{k,\epsilon_A} = \int k\left(\vec{x}, \cdot\right) d\mu_{\epsilon_A}$$

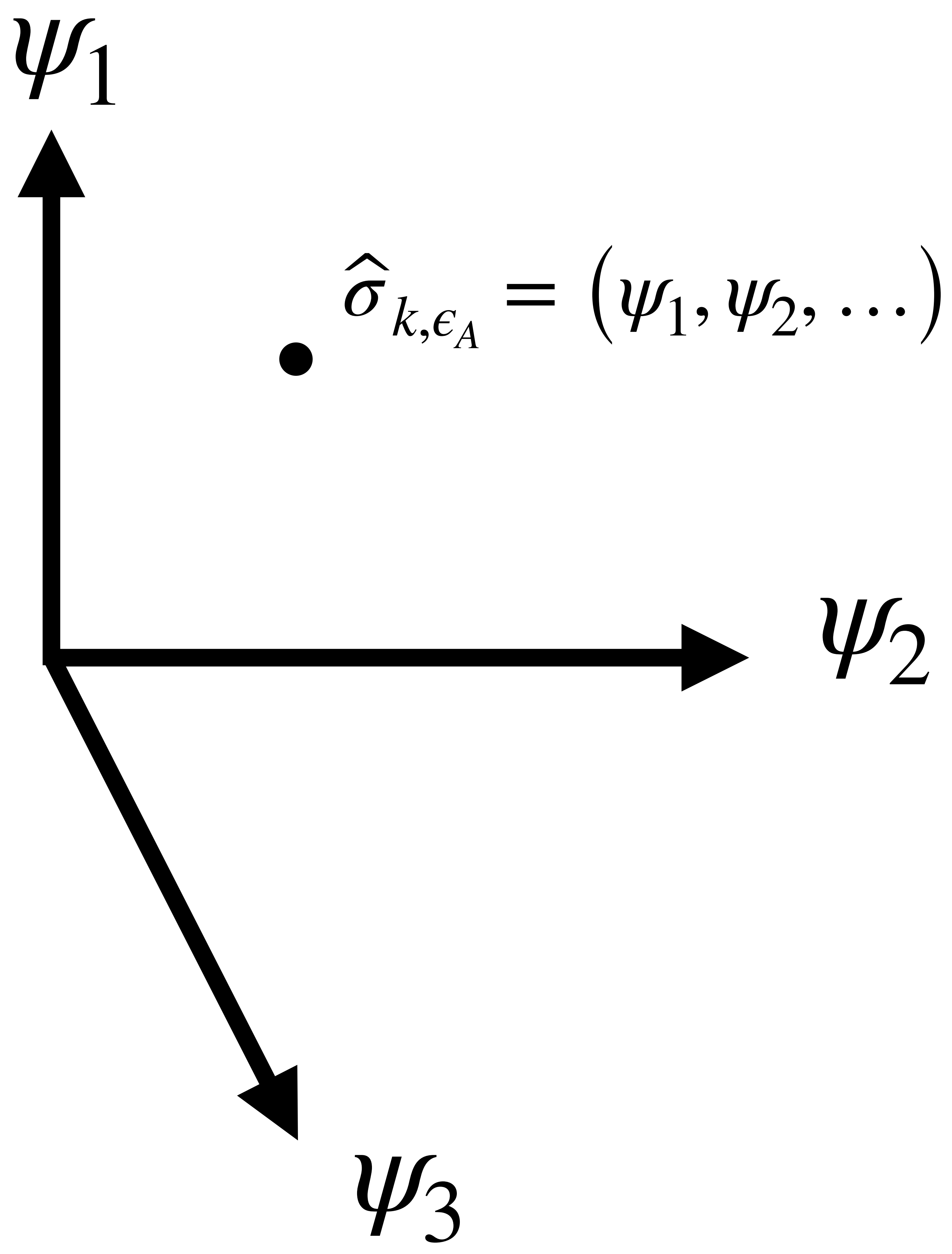
Universal kernels: $k\left(\vec{x}, \vec{x}'\right) = \exp\left(-\frac{d\left(\vec{x}, \vec{x}'\right)^2}{\beta^2}\right)$

→ Guarantees inner product space on measures is isometric to RKHS generated by k

Dimension Reduction…?



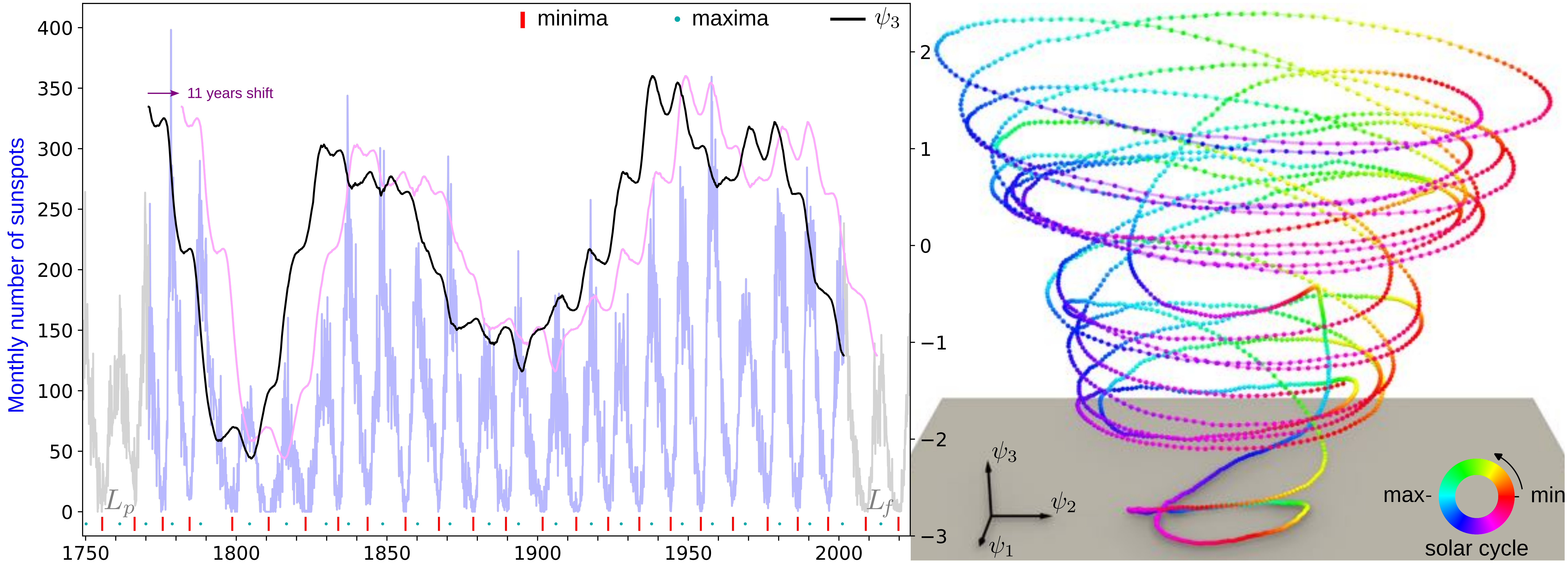
Current technique:
- diffusion mapping



Kernel Causal States: *n*-Butane

Alexandra M. Jurgens, Nicolas Brodu; Inferring kernel ϵ-machines: Discovering structure in complex systems. *Chaos* 1 March 2025; 35 (3): 033162. https://doi.org/10.1063/5.0242981

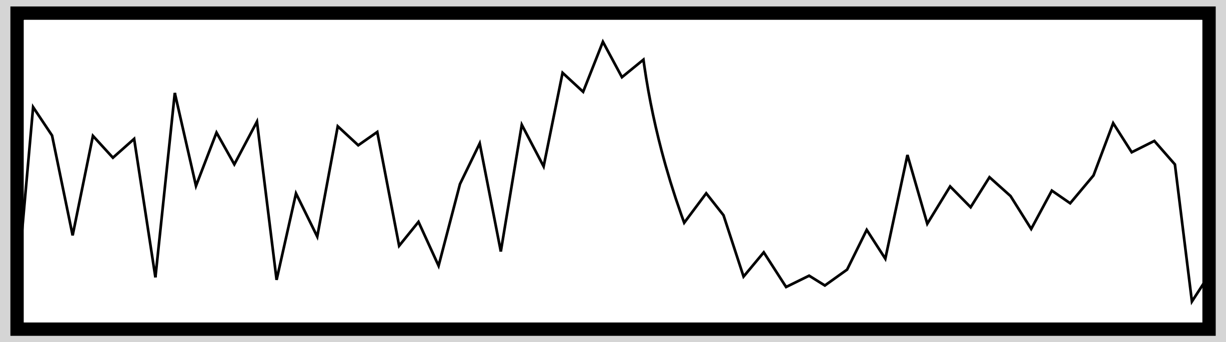
Kernel Causal States: Sunspot Sequence



Alexandra M. Jurgens, Nicolas Brodu; Inferring kernel ϵ-machines: Discovering structure in complex systems. *Chaos* 1 March 2025; 35 (3): 033162. https://doi.org/10.1063/5.0242981

Writing Down ϵ-Machines

Data



0100010000101110…

Lorem ipsum dolor sit amen…

⋮

Causal States S

- Via the predictive equivalence relation

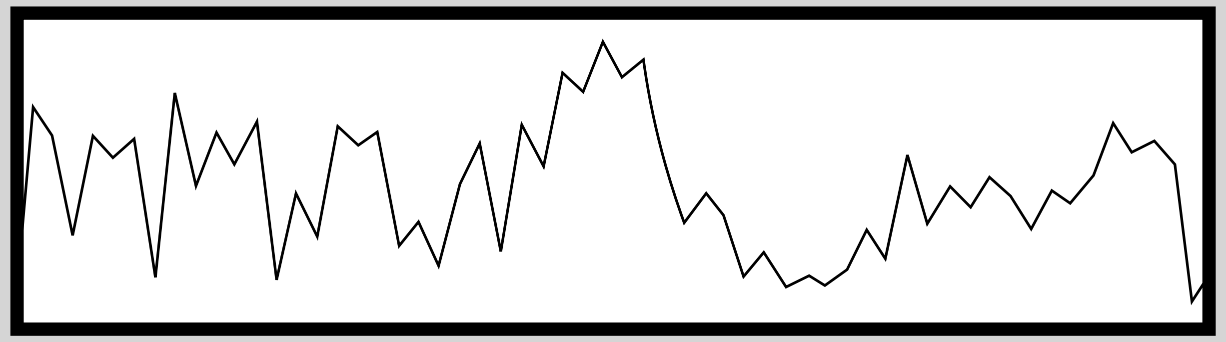
1. By inspection
2. By generation in the space spanned by a generating model
3. By embedding in a space
   A. Which space? What metric properties do we need to recreate the predictive equivalence relation?
   B. Dimension reduction?

ϵ-Machine

- Dynamic over the casual states?

Writing Down ϵ-Machines

Data



0100010000101110…

Lorem ipsum dolor sit amen…

⋮

Causal States S

- Via the predictive equivalence relation

1. By inspection
2. By generation in the space spanned by a generating model
3. By embedding in a space
   A. Which space? What metric properties do we need to recreate the predictive equivalence relation?
   B. Dimension reduction?

ϵ-Machine

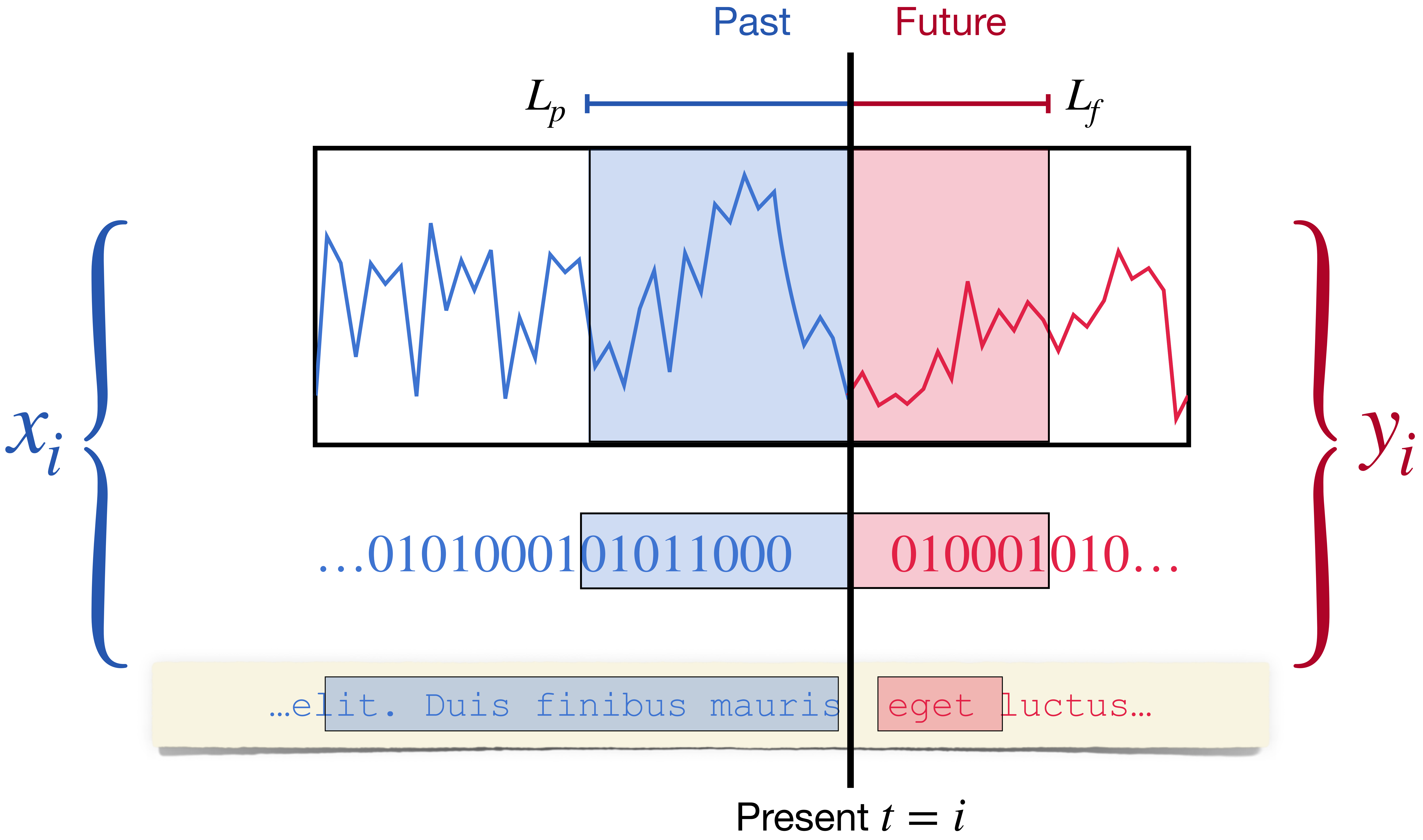
- Dynamic over the casual states?

Complexity & Information Measures

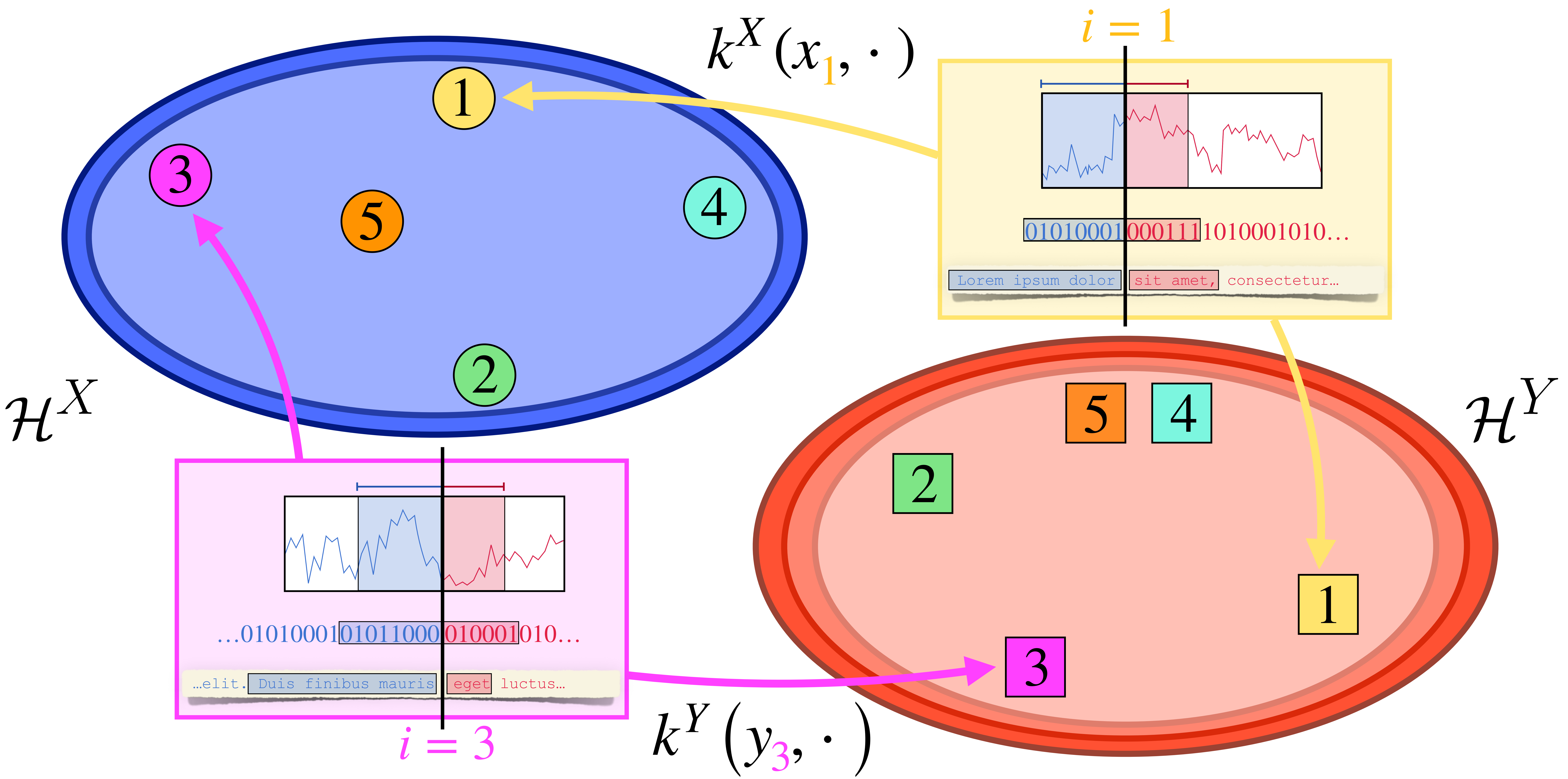
- Information measure zoo?

Thank you and Questions

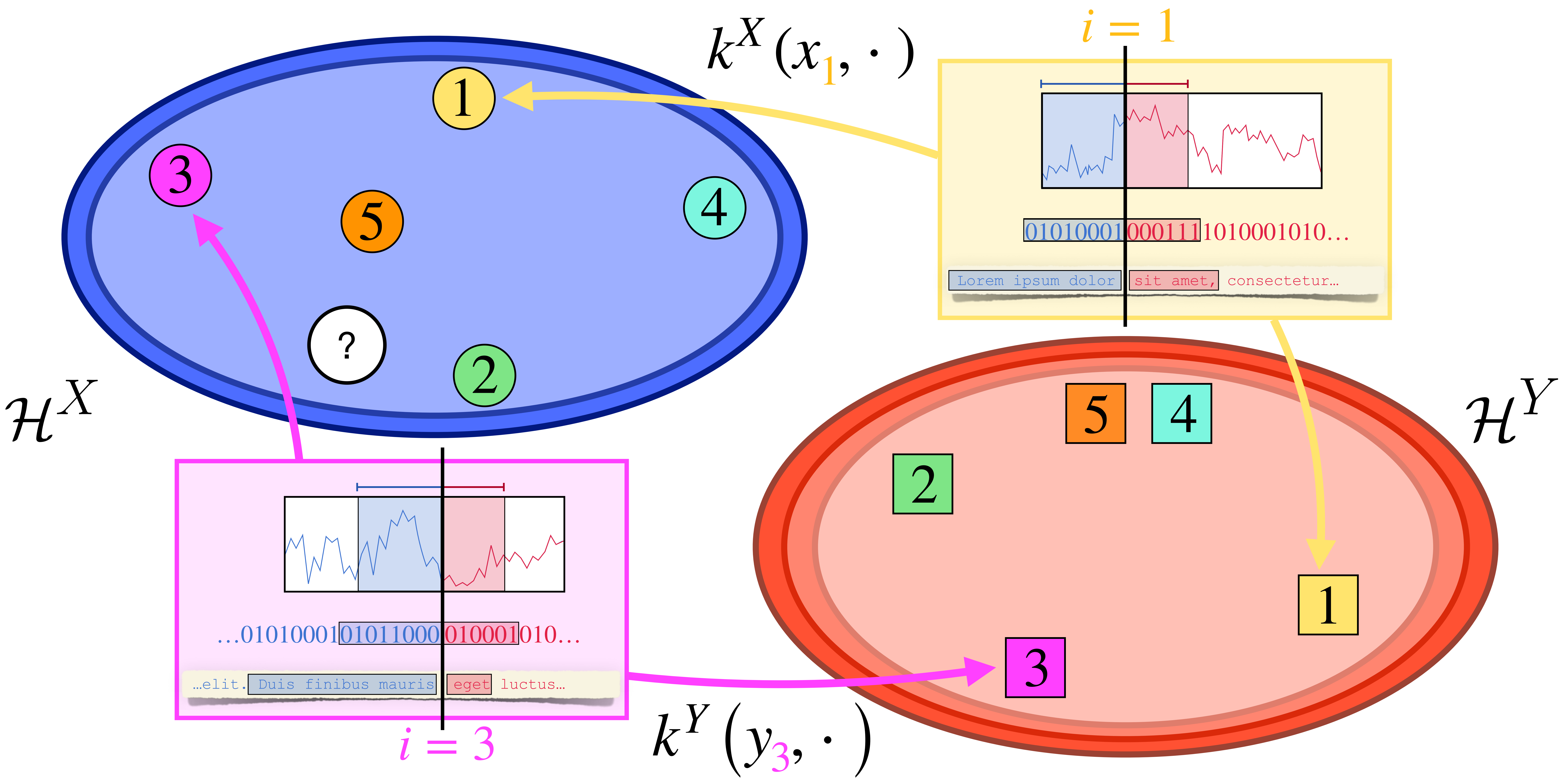
Kernel Causal States Algorithm



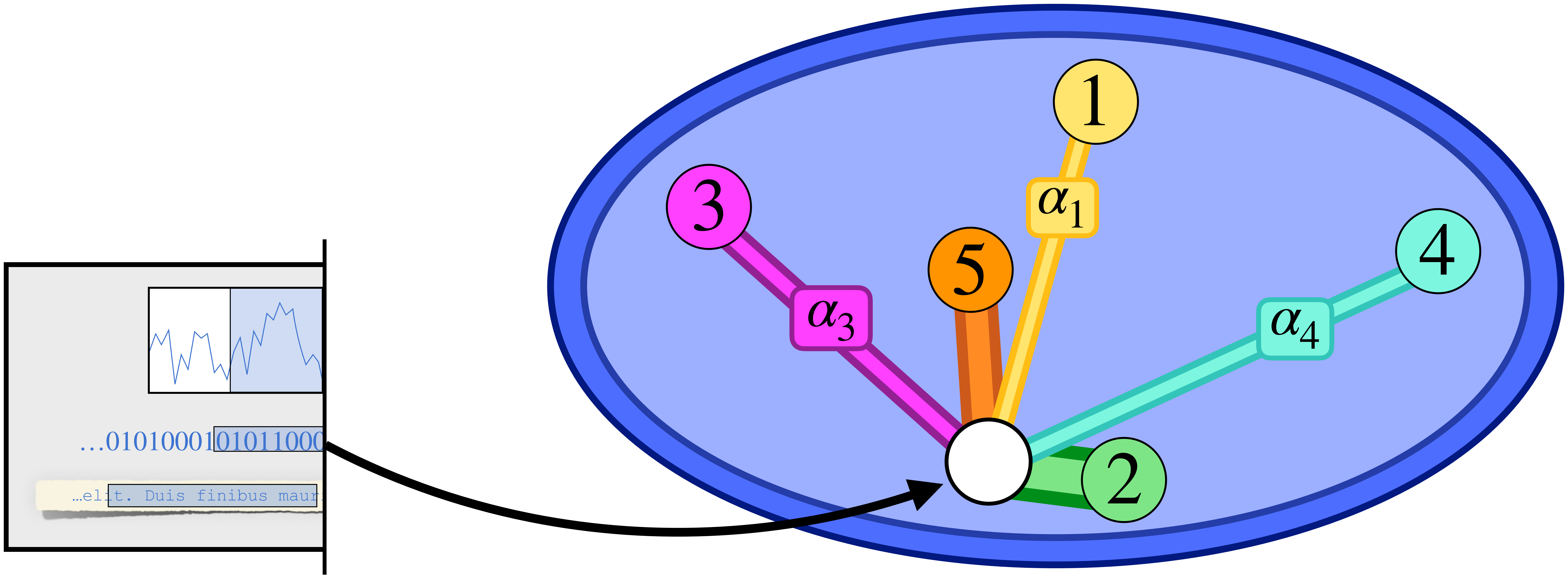
Embed Pasts *and* Futures



Embed Pasts *and* Futures

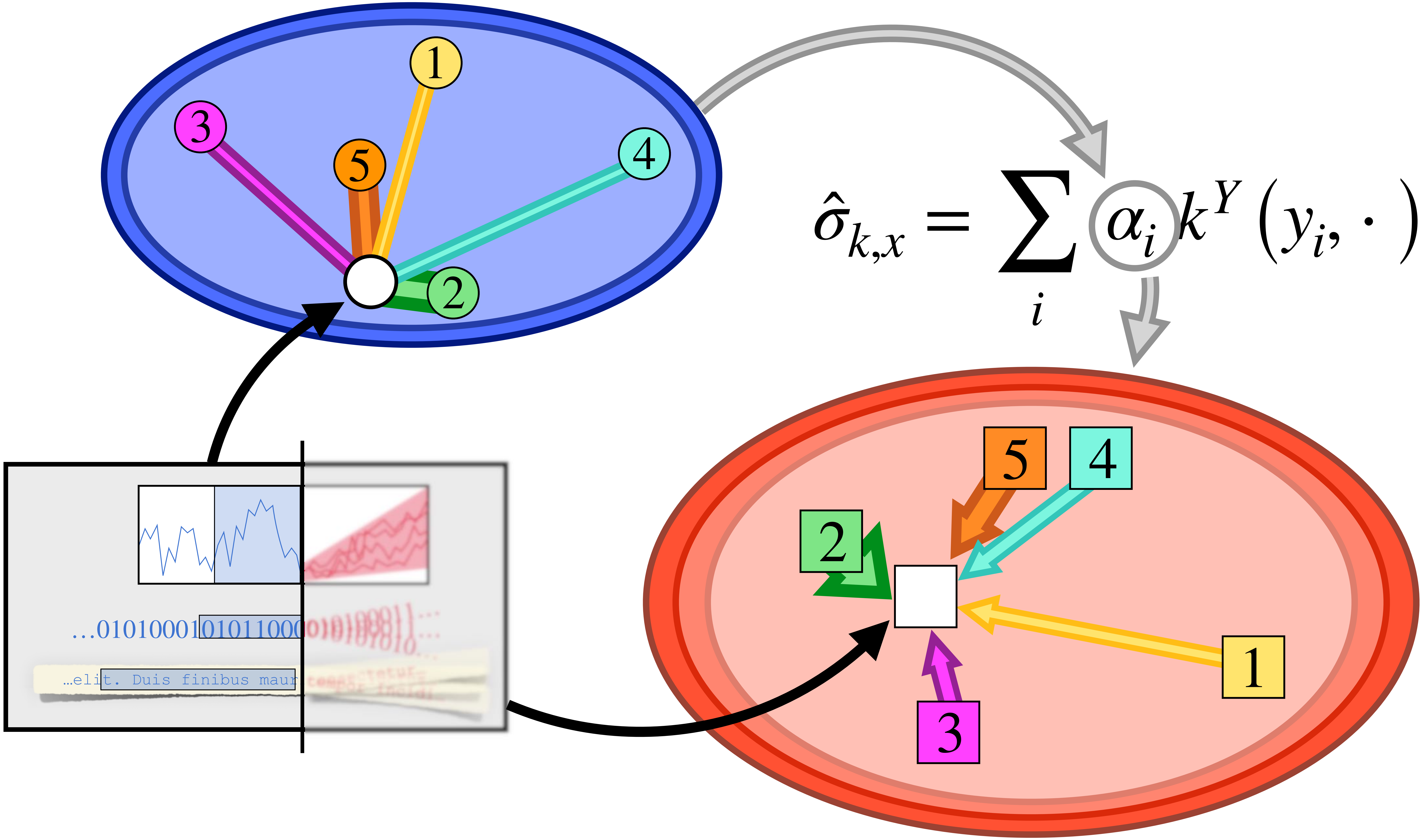


Calculate Proximity-Based Weights



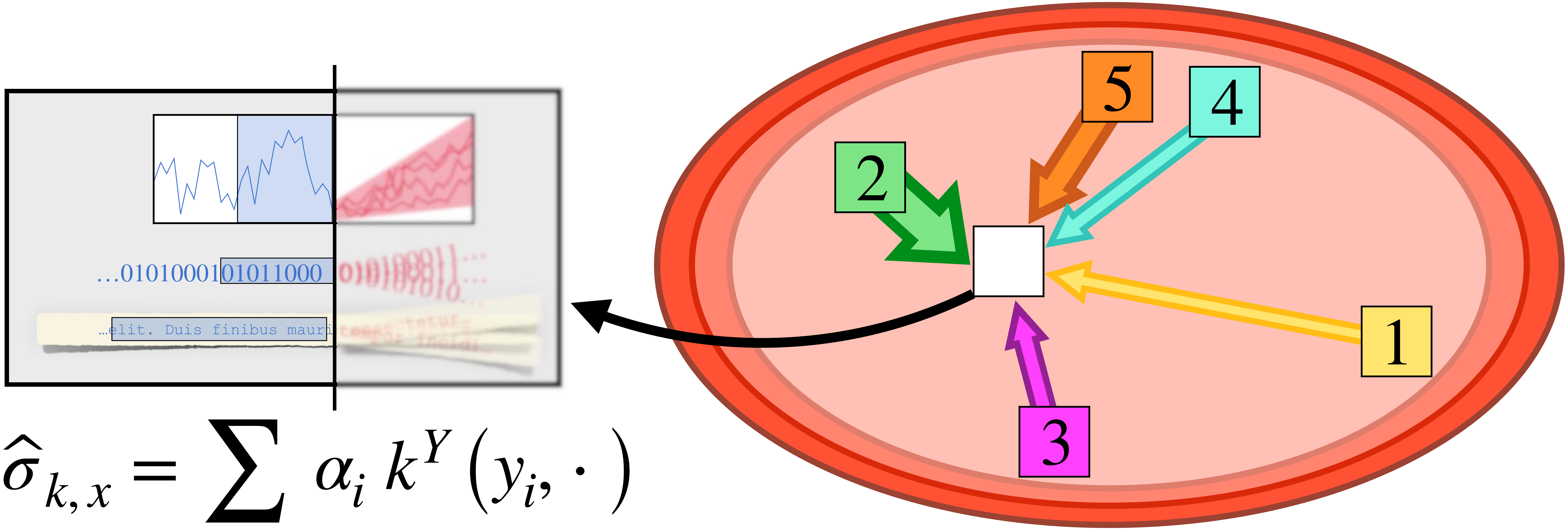
$$k^X\left(x, x_i\right) \uparrow \implies \boxed{\alpha_i} \uparrow$$

Construct Future Distribution



$$\hat{\sigma}_{k,x} = \sum_i \alpha_i k^Y \left(y_i, \cdot\right)$$

Embed Causal State



$$\hat{\sigma}_{k,x} = \sum_i \alpha_i k^Y(y_i, \cdot\,)$$