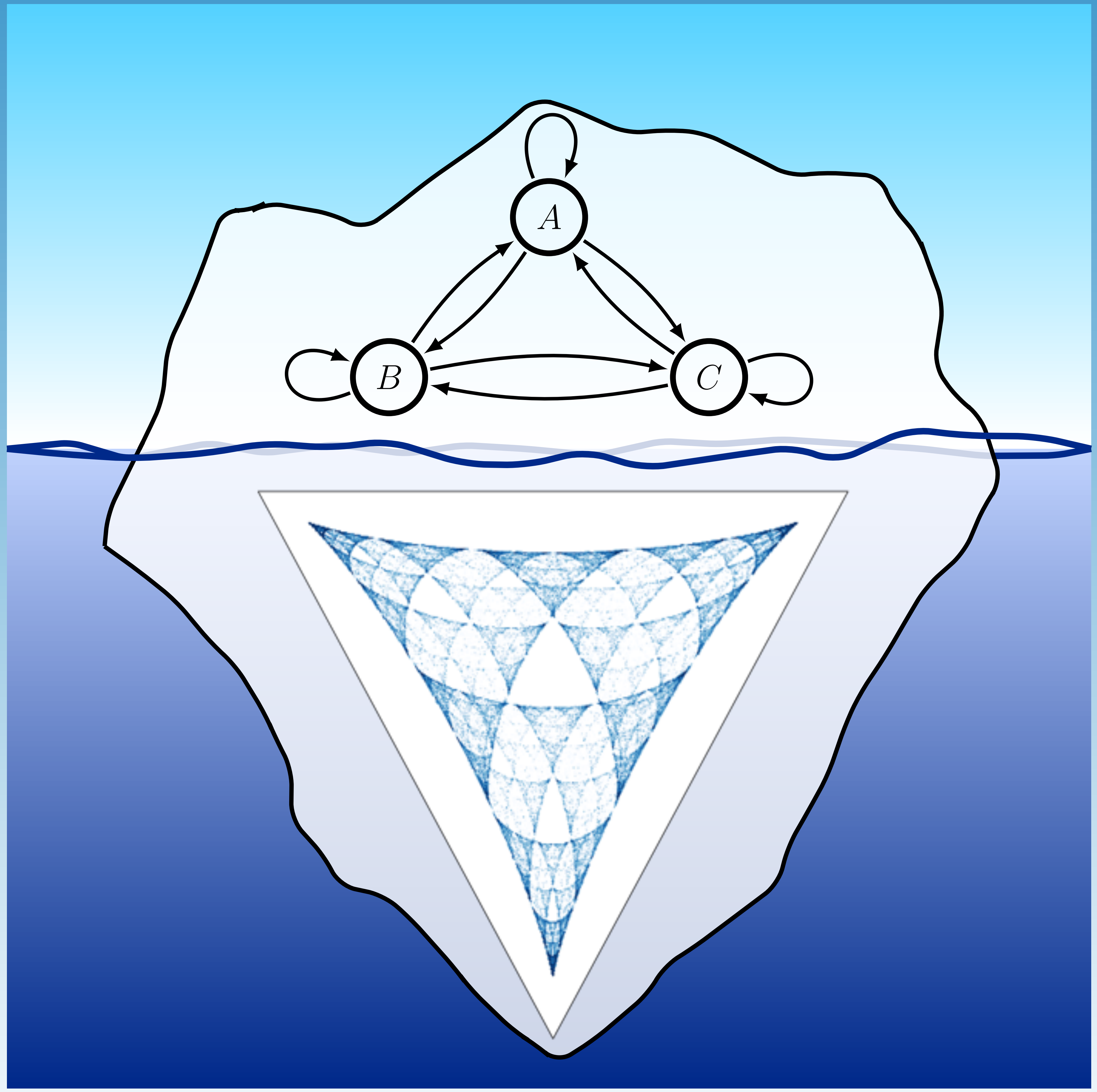
PHYS 256: POCI

Physics of Information & Computation

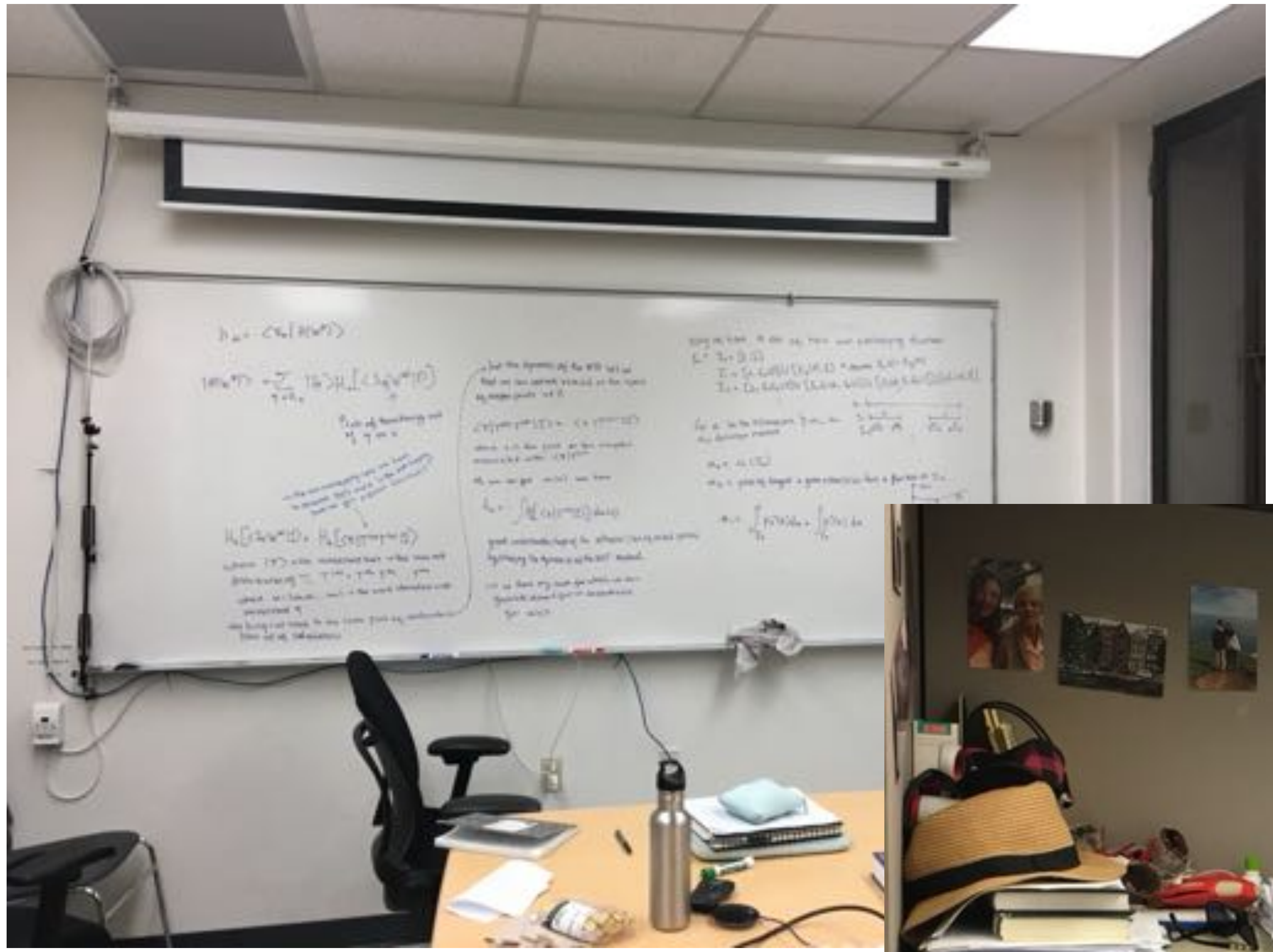
Alexandra Jurgens
Inria Centre at the University of Bordeaux
05/20/2025

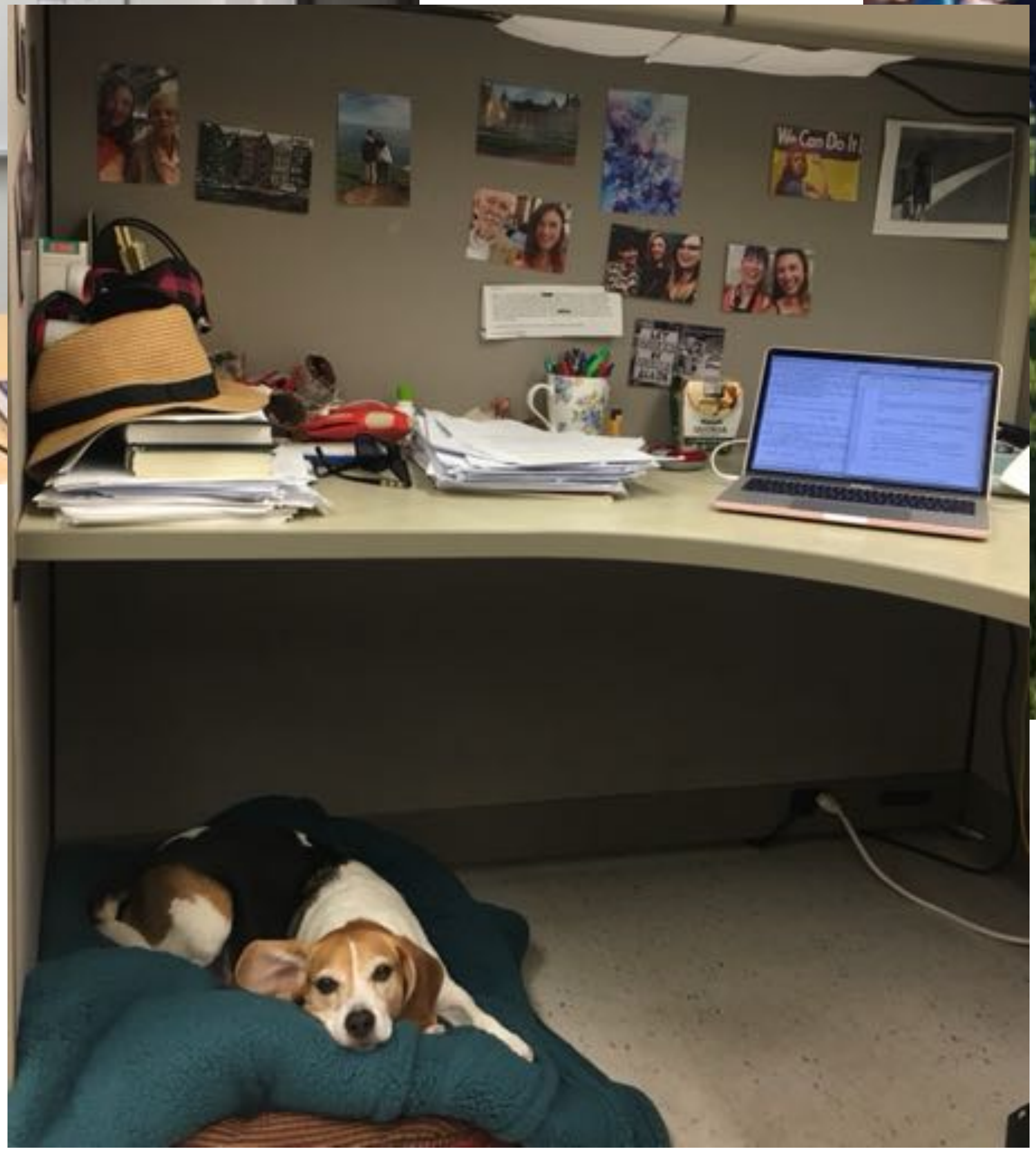
Infinite State Processes I



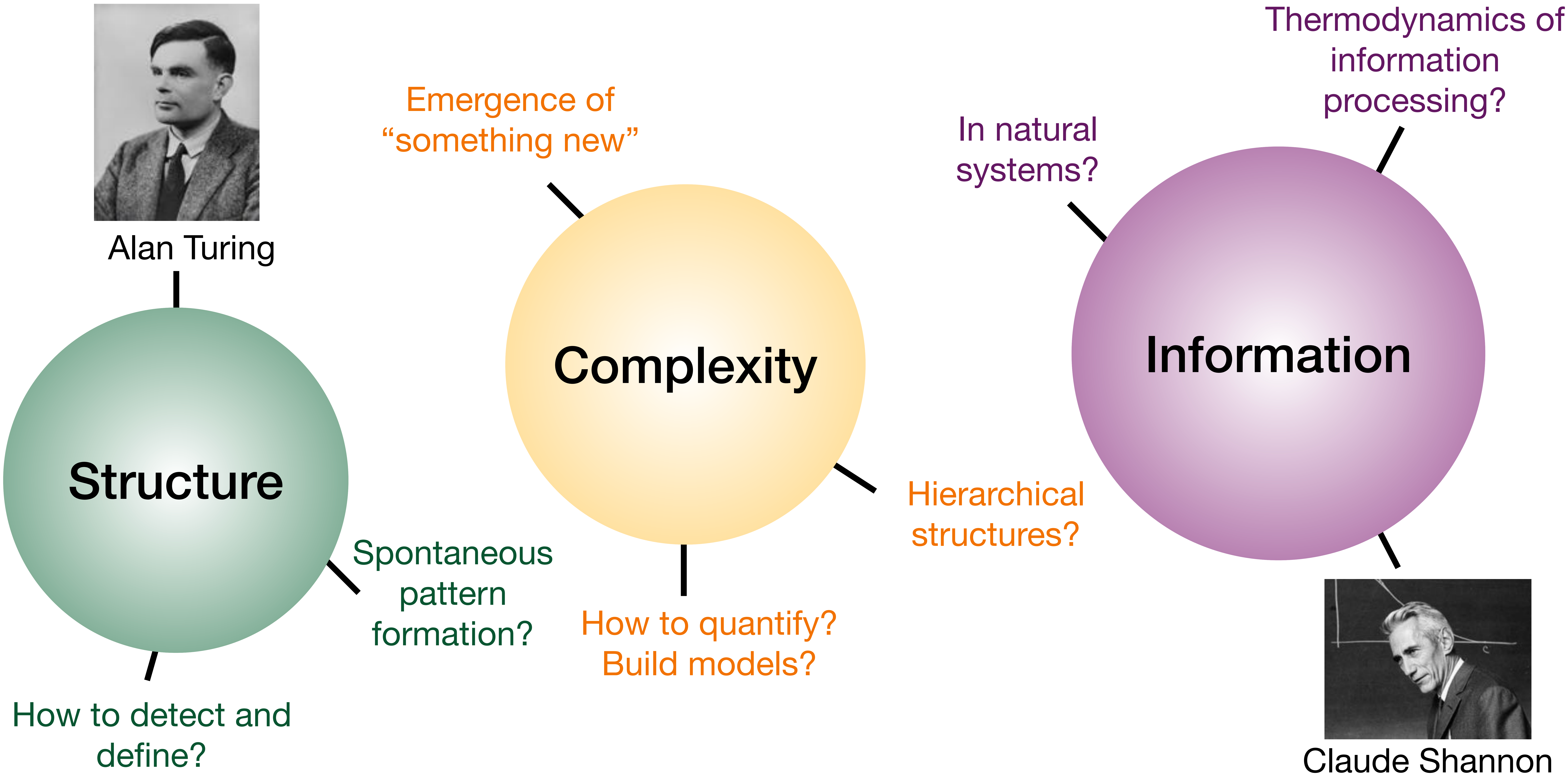
Alexandra Jurgens
Inria Centre at the University of Bordeaux
05/20/2025

UC Davis 2015 - 2021





Complex Systems Topics Overview



Relevant Papers

Journal of Statistical Physics (2021) 183:32
https://doi.org/10.1007/s10955-021-02769-3

Check for updates

Shannon Entropy Rate of Hidden Markov Processes

Alexandra M. Jurgens[1] · James P. Crutchfield[1]

Received: 6 October 2020 / Accepted: 24 April 2021 / Published online: 12 May 2021

PHYSICAL REVIEW E **104**, 064107 (2021)

Ambiguity rate of hidden Markov processes

Alexandra M. Jurgens * and James P. Crutchfield †
Complexity Sciences Center, Physics Department University of California at Davis Davis, California 95616, USA

(Received 6 August 2021; revised 17 November 2021; accepted 18 November 2021; published 6 December 2021)

The ϵ-machine is a stochastic process's optimal model—maximally predictive and minimal in size. It often happens that to optimally predict even simply defined processes, probabilistic models—including the ϵ-machine—must employ an uncountably infinite set of features. To constructively work with these infinite sets we map the ϵ-machine to a place-dependent iterated function system (IFS)—a stochastic dynamical system. We then introduce the ambiguity rate that, in conjunction with a process's Shannon entropy rate, determines the rate at which this set of predictive features must grow to maintain maximal predictive power over increasing horizons. We demonstrate, as an ancillary technical result that stands on its own, that the ambiguity rate is the (until now missing) correction to the Lyapunov dimension of an IFS's attracting invariant set. For a broad class of complex processes, this then allows calculating their statistical complexity dimension—the information dimension of the minimal set of predictive features.

DOI: 10.1103/PhysRevE.104.064107

Chaos ARTICLE scitation.org/journal/cha

Divergent predictive states: The statistical complexity dimension of stationary, ergodic hidden Markov processes

Cite as: Chaos 31, 083114 (2021); doi: 10.1063/5.0050460
Submitted: 15 March 2021 · Accepted: 4 July 2021 ·
Published Online: 30 August 2021

View Online Export Citation CrossMark

Alexandra M. Jurgens[a] and James P. Crutchfield[b]

AFFILIATIONS

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

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

Shannon Entropy Rate of HHMs

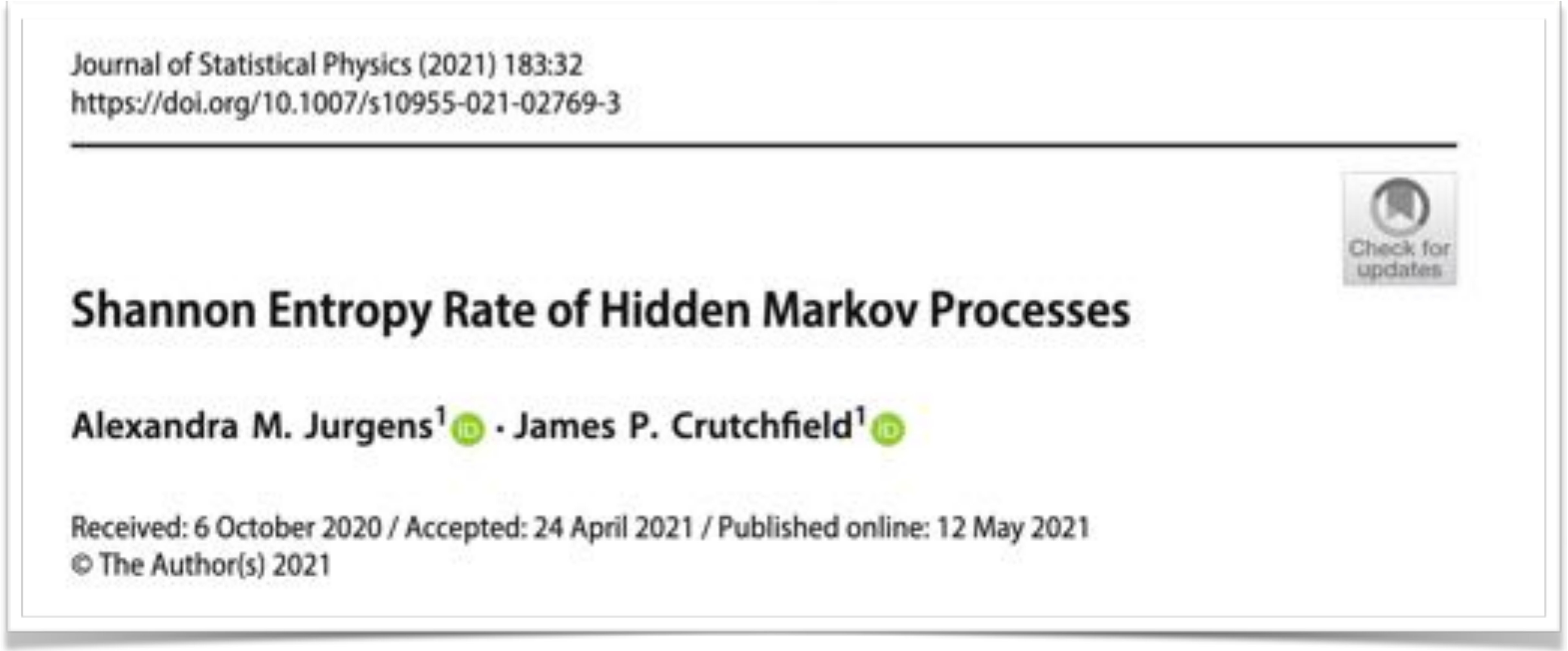
Journal of Statistical Physics (2021) 183:32
https://doi.org/10.1007/s10955-021-02769-3

Shannon Entropy Rate of Hidden Markov Processes

Alexandra M. Jurgens[1] · James P. Crutchfield[1]

Received: 6 October 2020 / Accepted: 24 April 2021 / Published online: 12 May 2021

Shannon Entropy Rate of HHMs


Journal of Statistical Physics (2021) 183:32
https://doi.org/10.1007/s10955-021-02769-3

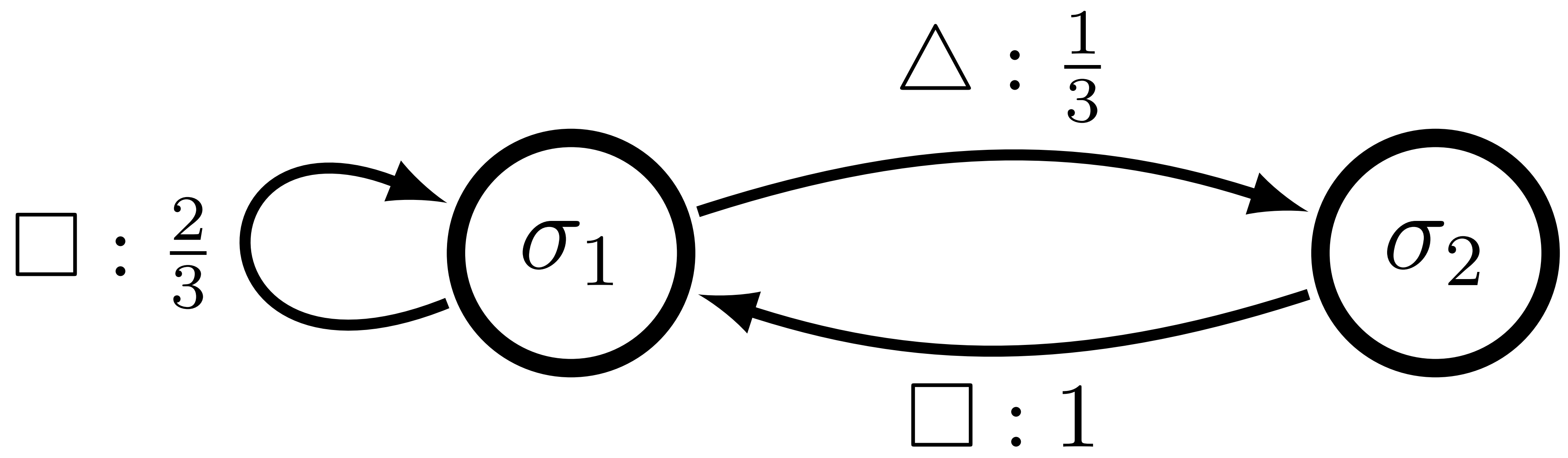
Check for updates

Shannon Entropy Rate of Hidden Markov Processes

Alexandra M. Jurgens[1] · James P. Crutchfield[1]

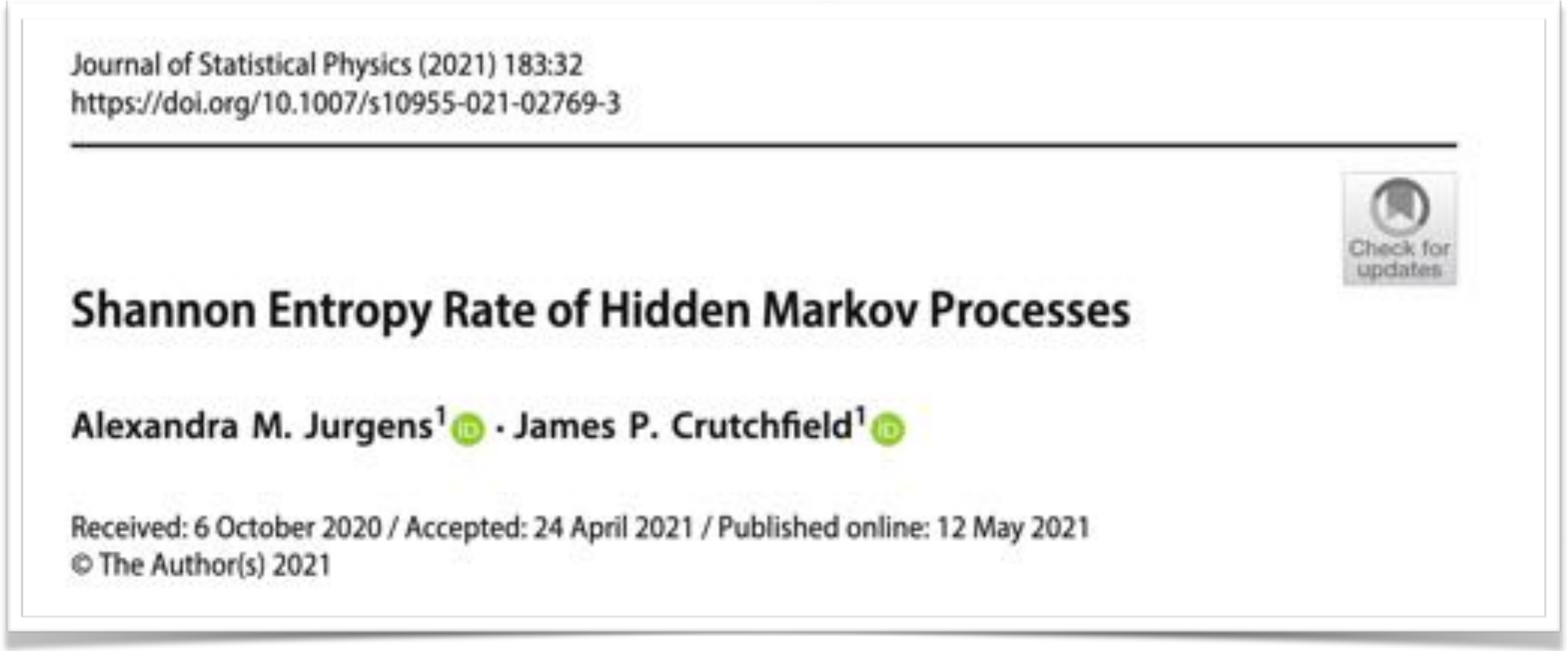
Received: 6 October 2020 / Accepted: 24 April 2021 / Published online: 12 May 2021

If you have a *unifilar* presentation:



Unifilar: At most *one* transition out of a state $s \in S$ per symbol $x \in A$

Shannon Entropy Rate of HHMs


Journal of Statistical Physics (2021) 183:32
https://doi.org/10.1007/s10955-021-02769-3

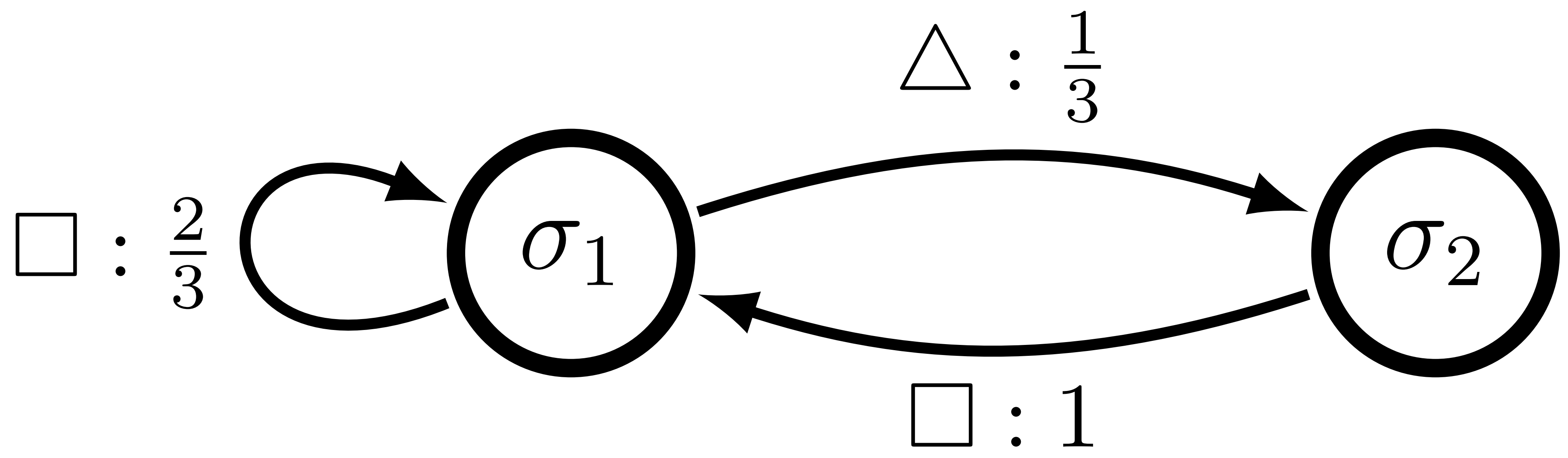
Check for updates

Shannon Entropy Rate of Hidden Markov Processes

Alexandra M. Jurgens[1] · James P. Crutchfield[1]

Received: 6 October 2020 / Accepted: 24 April 2021 / Published online: 12 May 2021
© The Author(s) 2021

If you have a *unifilar* presentation:



$$h_\mu = -\sum_{s \in S} \Pr(s) \sum_{x \in A} \Pr\left(x \mid s\right) \log_2 \Pr\left(x \mid s\right)$$

Shannon Entropy Rate of HHMs


Journal of Statistical Physics (2021) 183:32
https://doi.org/10.1007/s10955-021-02769-3

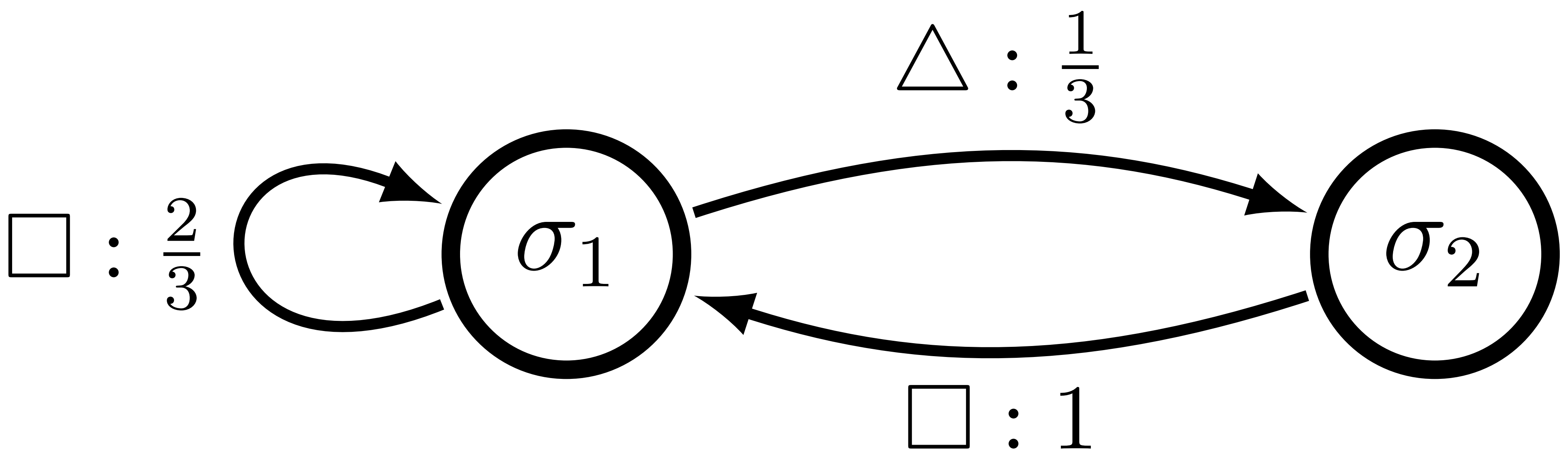
Check for updates

Shannon Entropy Rate of Hidden Markov Processes

Alexandra M. Jurgens[1] · James P. Crutchfield[1]

Received: 6 October 2020 / Accepted: 24 April 2021 / Published online: 12 May 2021

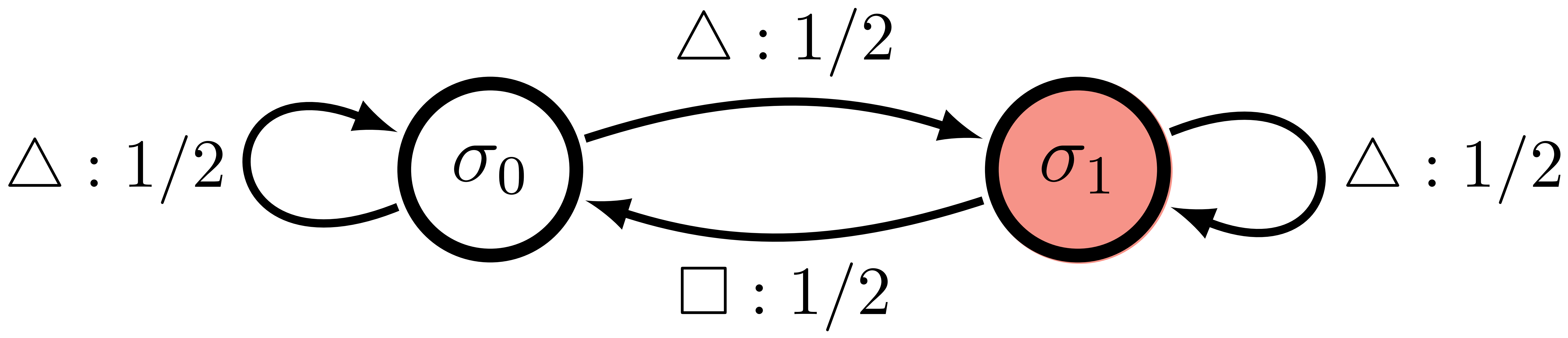
If you have a *unifilar* presentation:



$$h_\mu = -\sum_{s \in S} \Pr(s) \sum_{x \in A} \Pr\left(x \,|\, s\right) \log_2 \Pr\left(x \,|\, s\right)$$

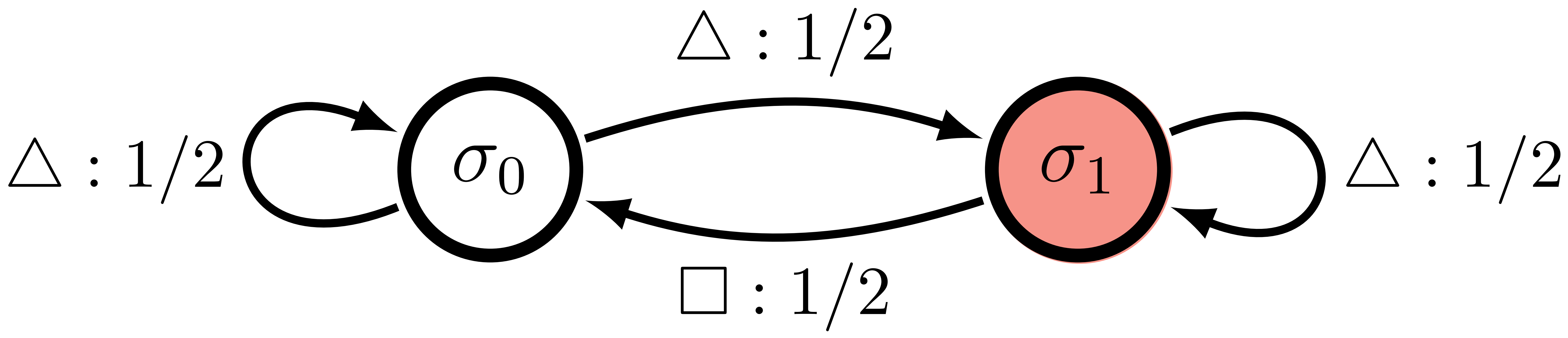
$h_\mu = 0.688$ bits

Non Unifilar Presentation?



State: σ_1

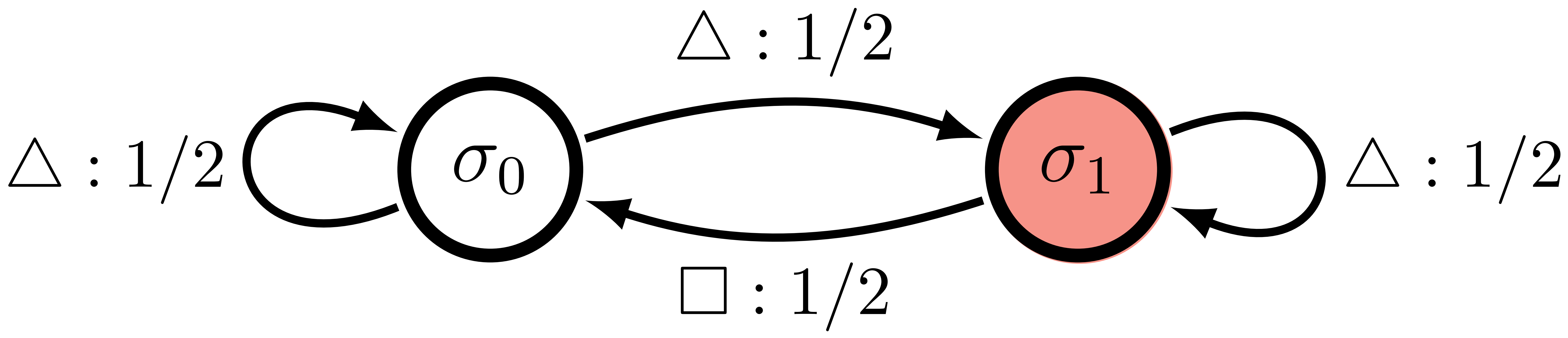
Non Unifilar Presentation?



State: σ_1

Symbol: △

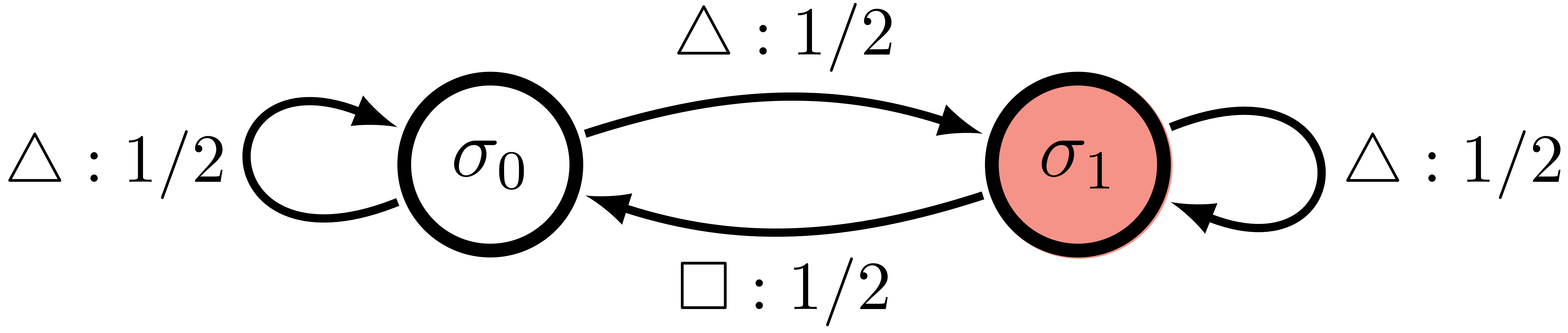
Non Unifilar Presentation?



State: σ_1 σ_1

Symbol: $\triangle$

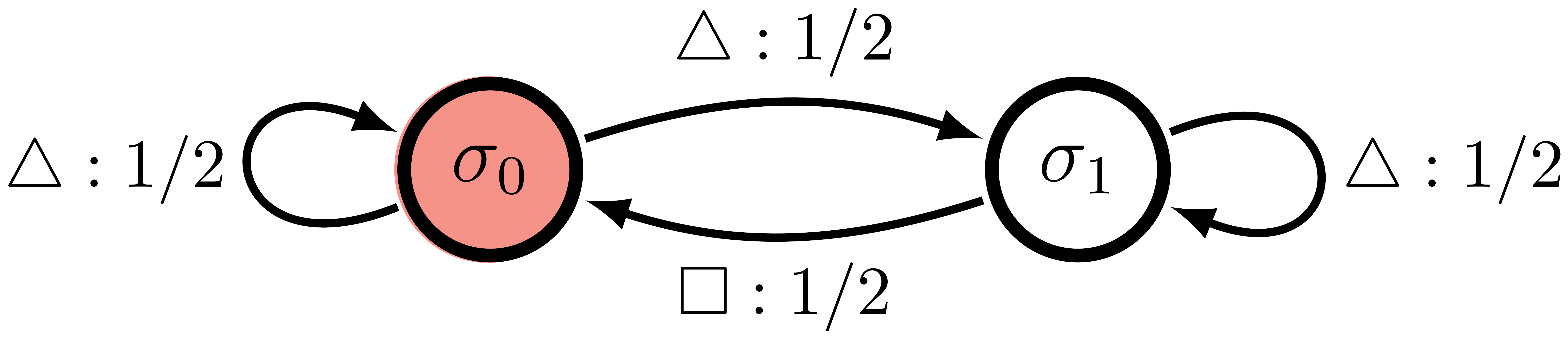
Non Unifilar Presentation?



State: σ_1 σ_1

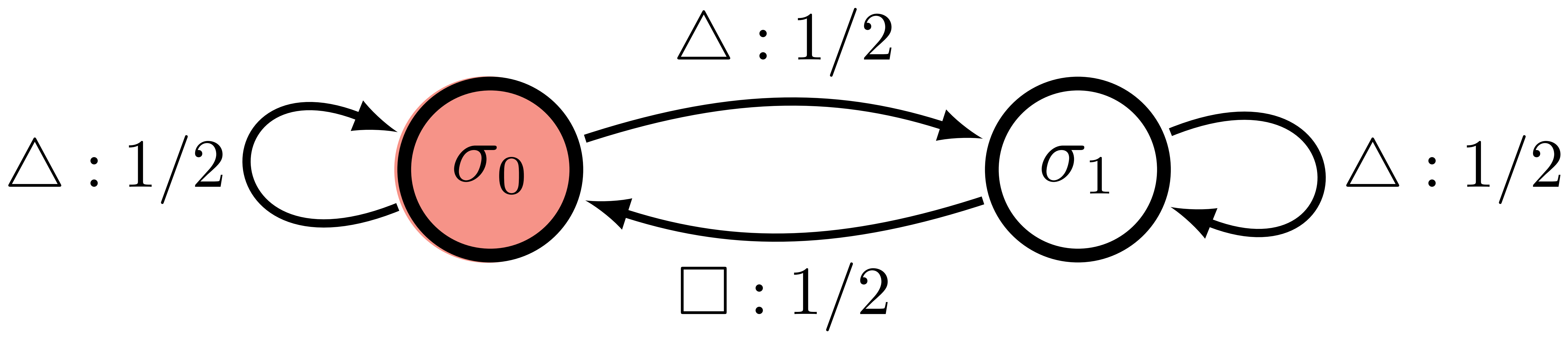
Symbol: $\triangle$ $\square$

Non Unifilar Presentation?



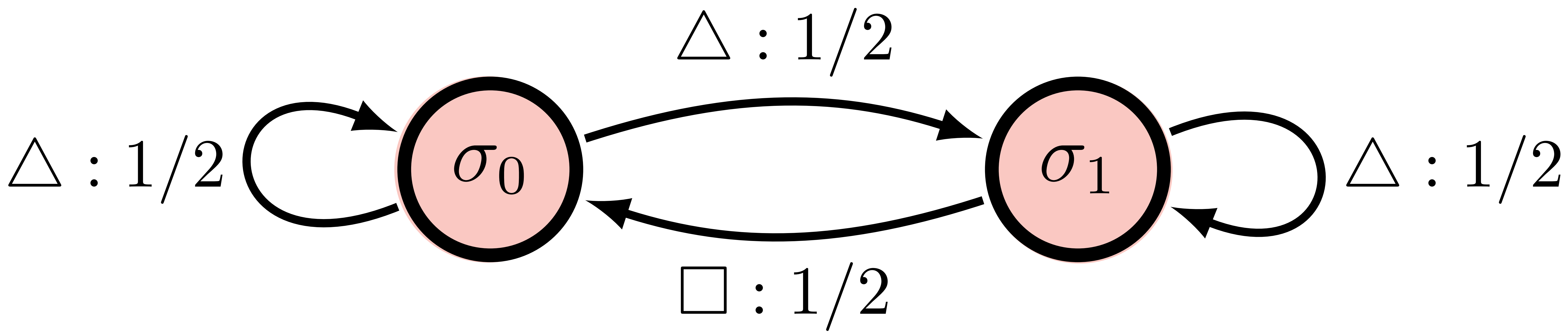
State:	σ_1	σ_1	σ_0
Symbol:	$\triangle$	$\square$	

Non Unifilar Presentation?



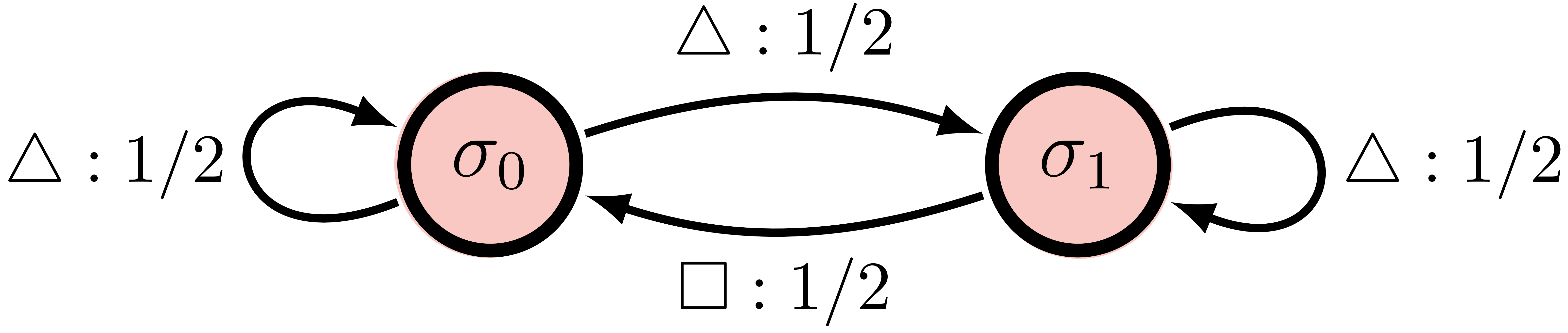
State:	σ_1	σ_1	σ_0
Symbol:	△	□	△

Non Unifilar Presentation?



State:	σ_1	σ_1	σ_0	**?**
Symbol:	$\triangle$	$\square$	$\triangle$	

Non-predictive States

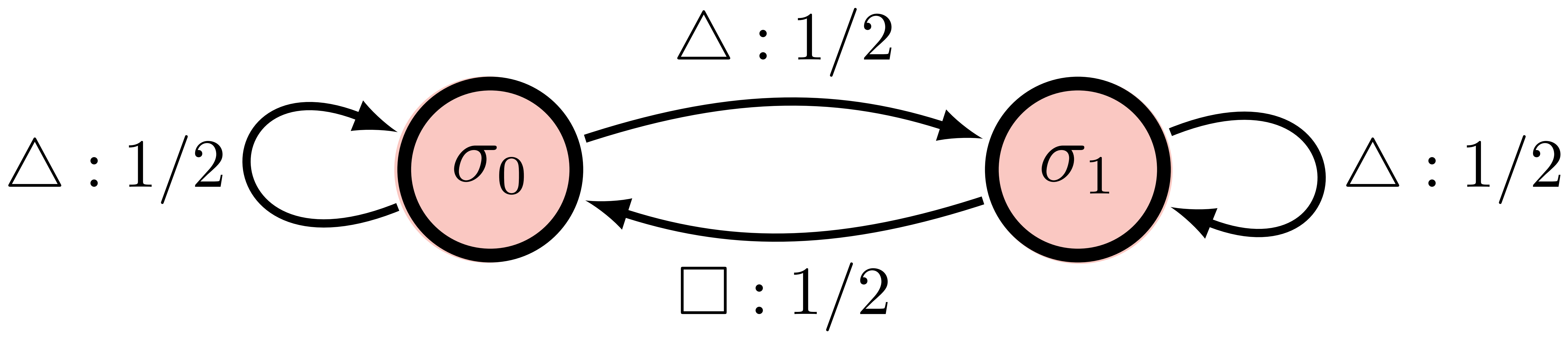


Can’t stay synched to internal state

↓

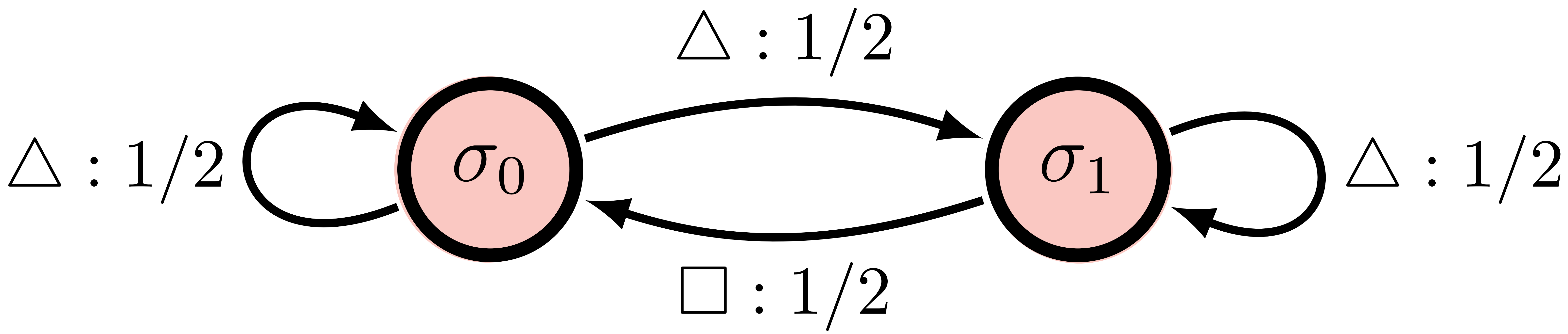
Can’t use states to compute h_μ

Non-predictive States



$$-\sum_{s\in S} \Pr(s) \sum_{x\in A} \Pr(x\,|\,s)\, \log_2 \Pr(x\,|\,s) \quad \leq \quad h_\mu \quad \leq \quad -\sum_{s\in S} Pr(s) \sum_{s'\in S} \Pr(s'\,|\,s)\, \log_2 \Pr(s'\,|\,s)$$

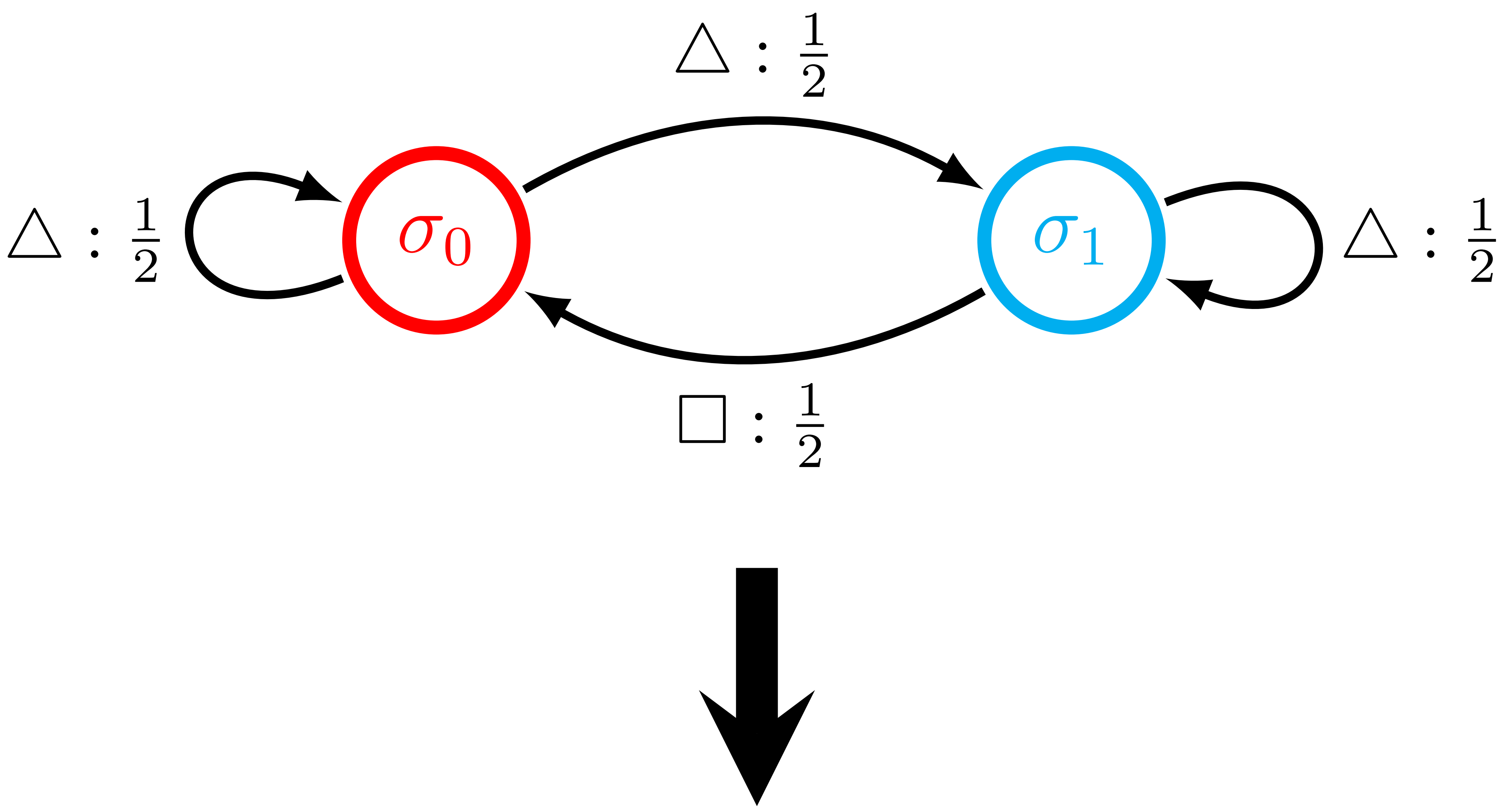
Non-predictive States



$$-\sum_{s\in S} \Pr(s) \sum_{x\in A} \Pr(x\,|\,s)\ \log_2 \Pr(x\,|\,s) \quad \leq \quad h_\mu \quad \leq \quad -\sum_{s\in S} Pr(s) \sum_{s'\in S} \Pr(s'\,|\,s)\ \log_2 \Pr(s'\,|\,s)$$

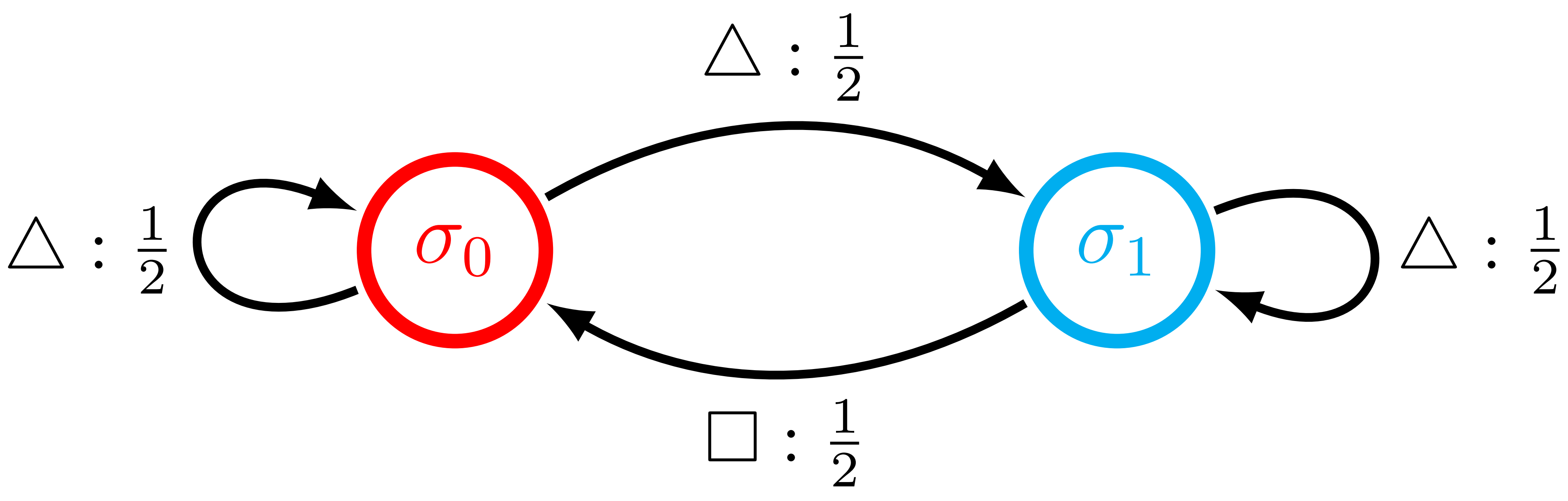
0.5 bits ? bits 2 bits

Mixed State Generation

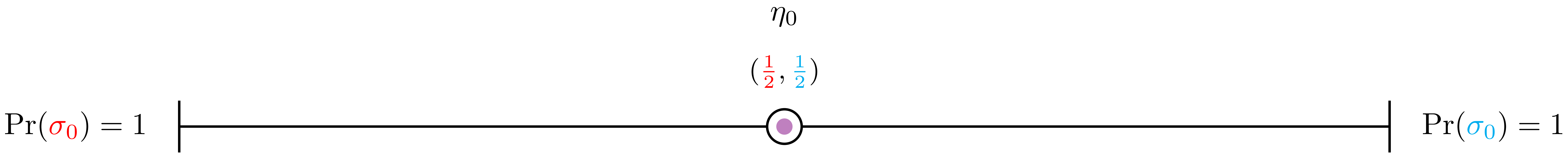


$\Pr(\sigma_0) = 1$ $\Pr(\sigma_0) = 1$

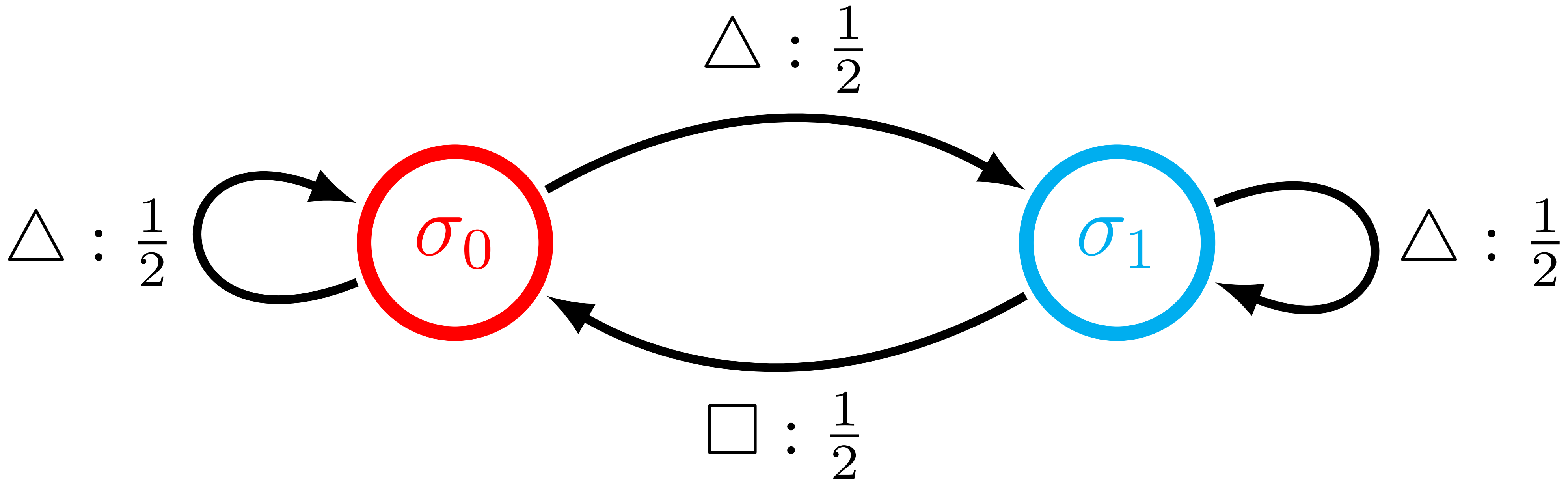
Mixed State Generation



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

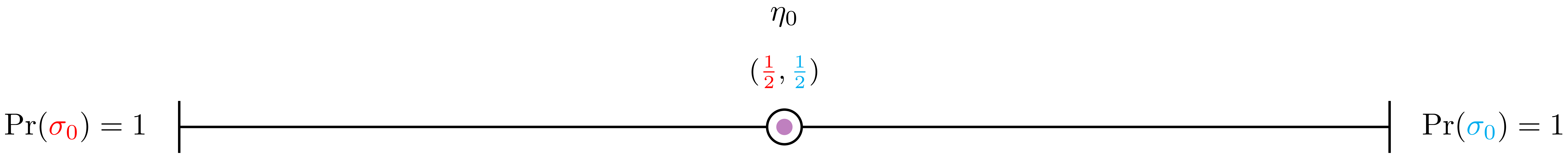


Mixed State Generation

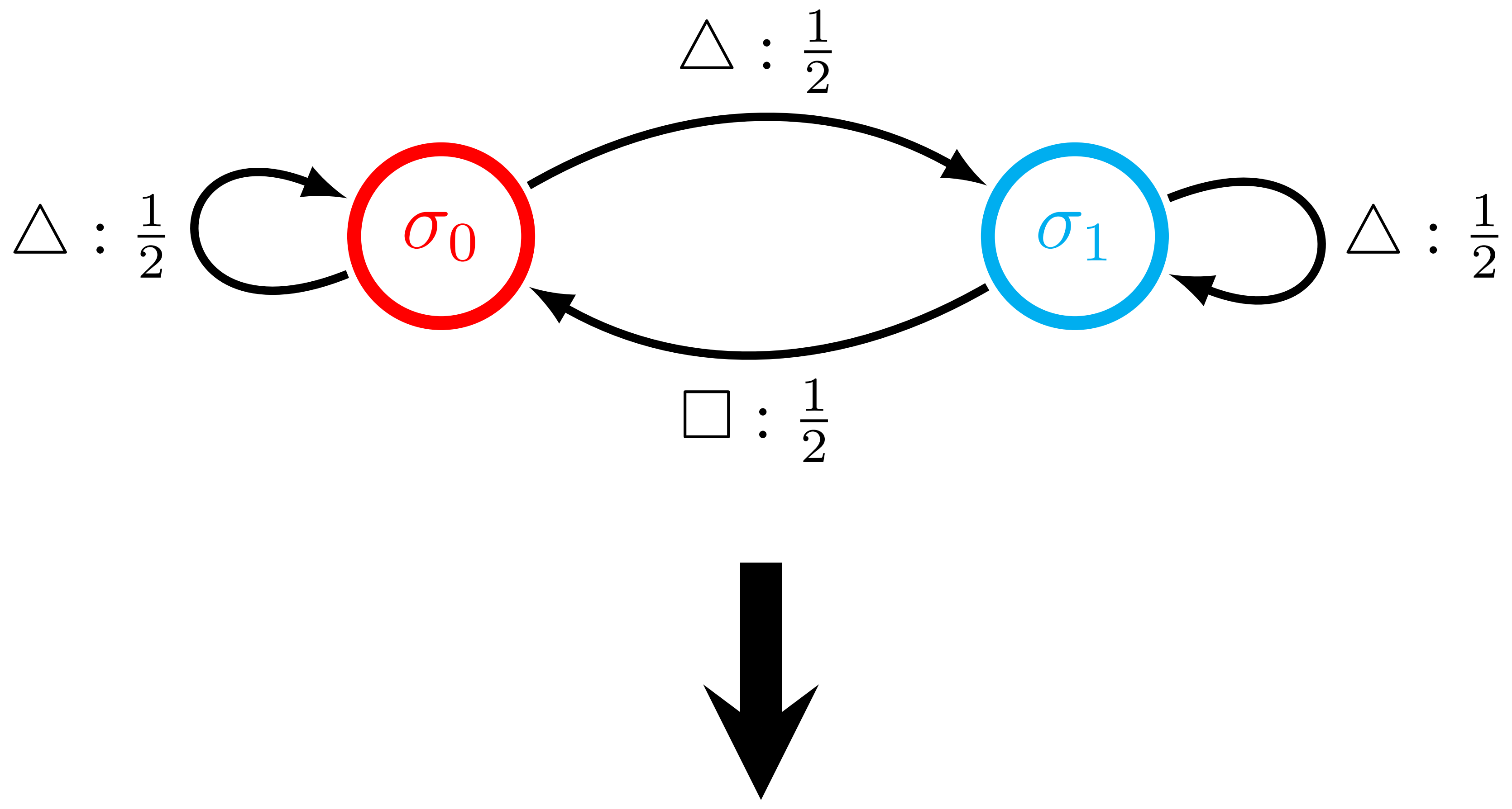


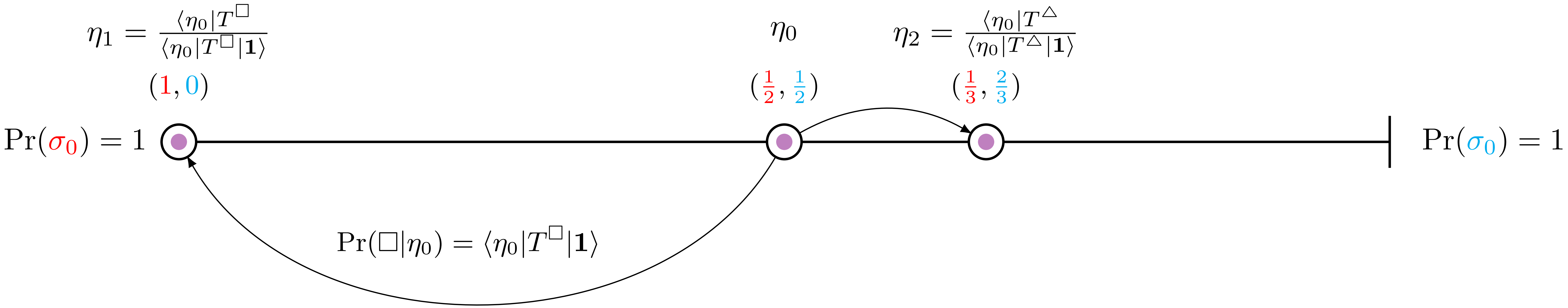
$$\Pr(x \mid \eta_t) = \eta_t T^{(x)} \mathbf{1}$$

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

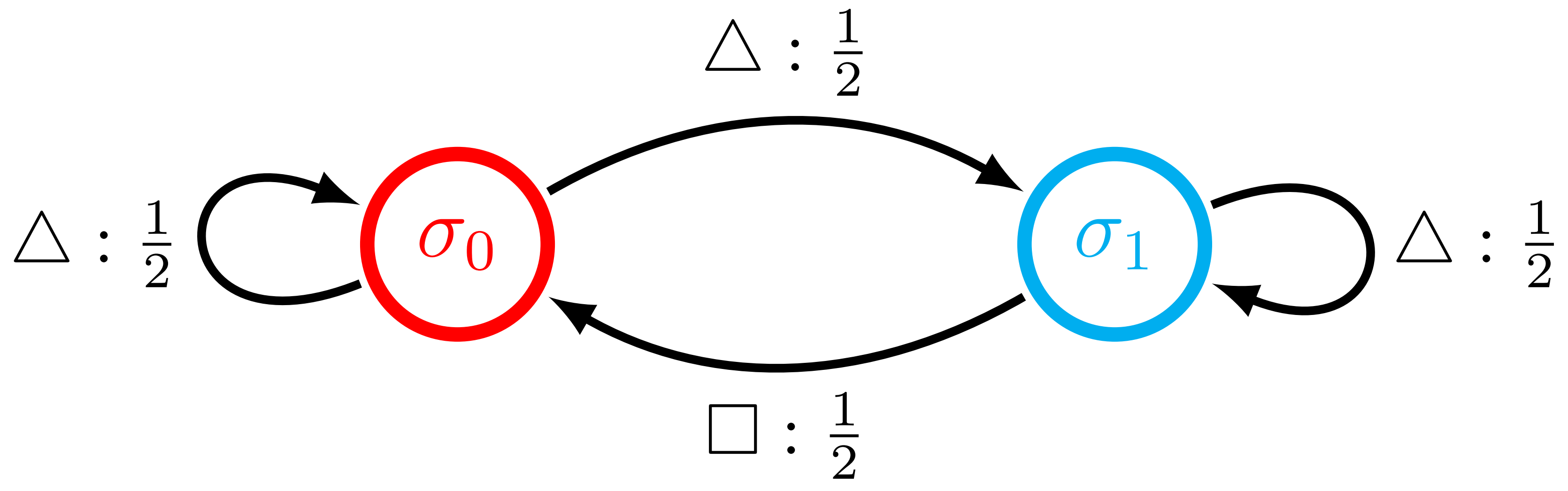


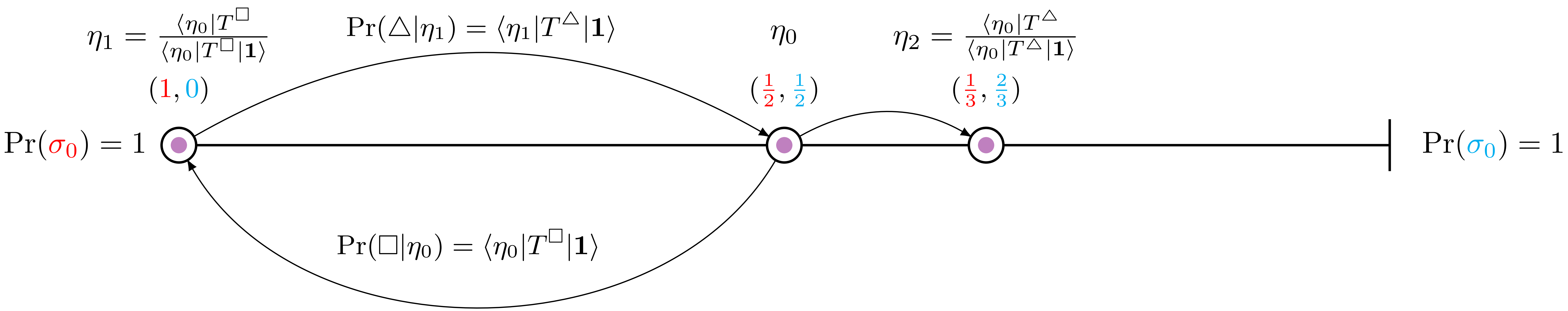
Mixed State Generation



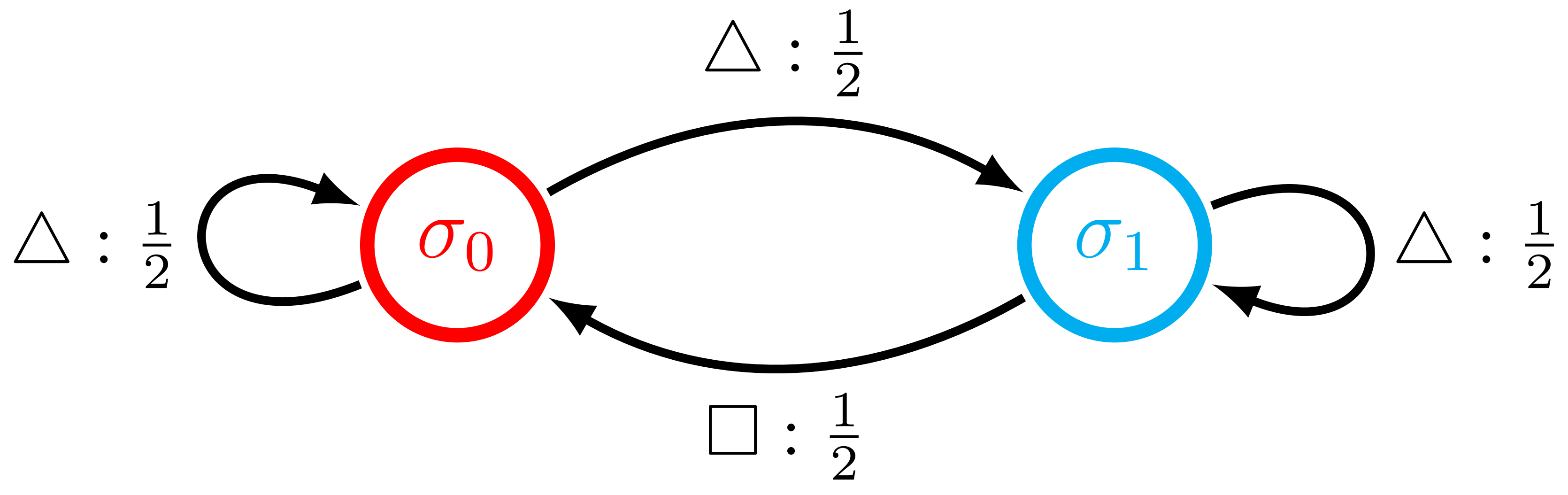


Mixed State Generation

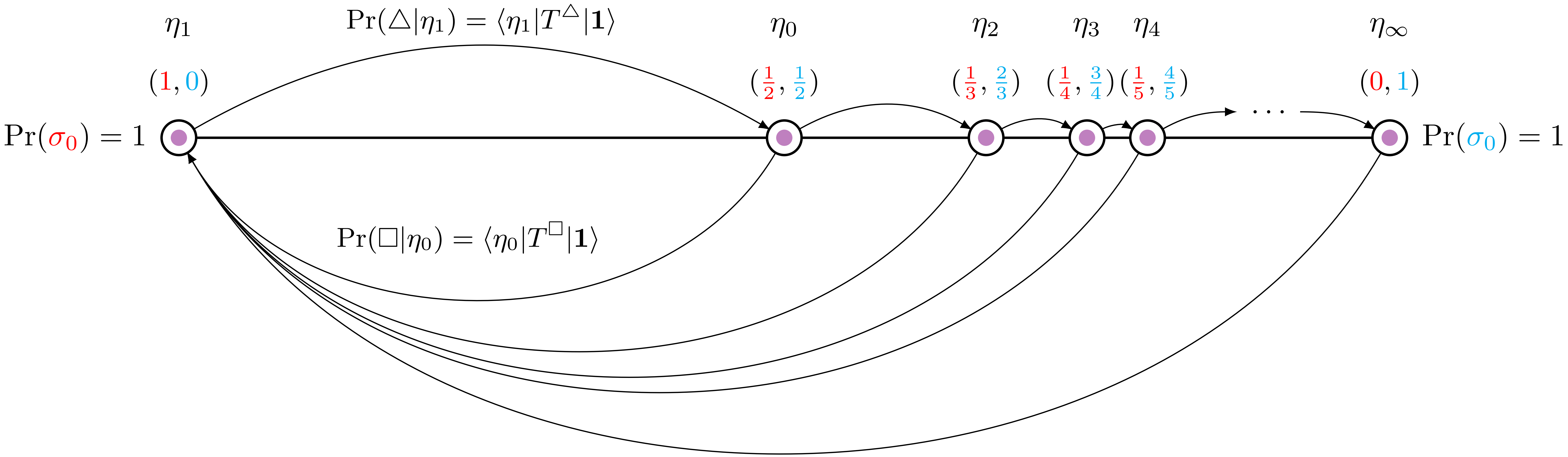




Mixed State Generation

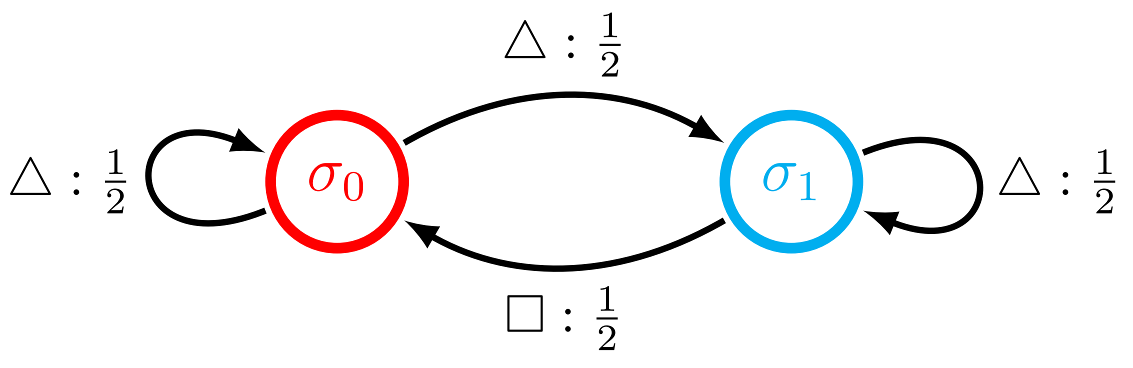


→ Build up the set $R = \left\{ \eta : \eta(w) = \Pr\left(S_l \,|\, X_{0:l} = w, S_0 = \pi\right) \right\}$ of mixed states.

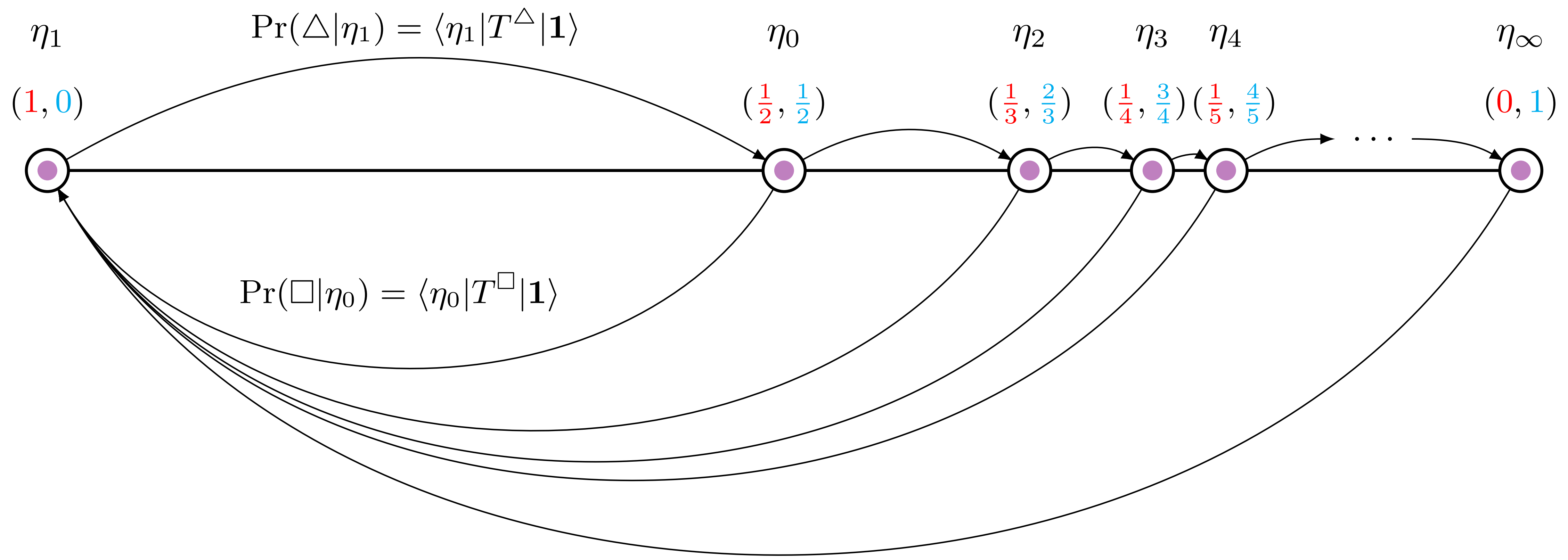


Constructing Predictive Model

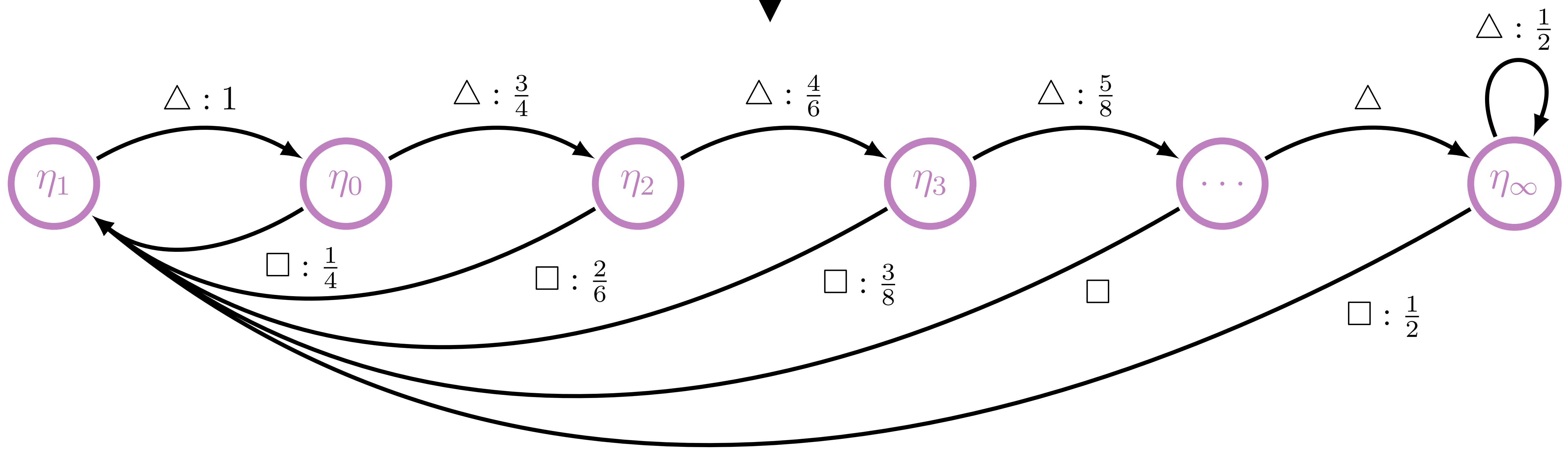
Non-Unifilar Model



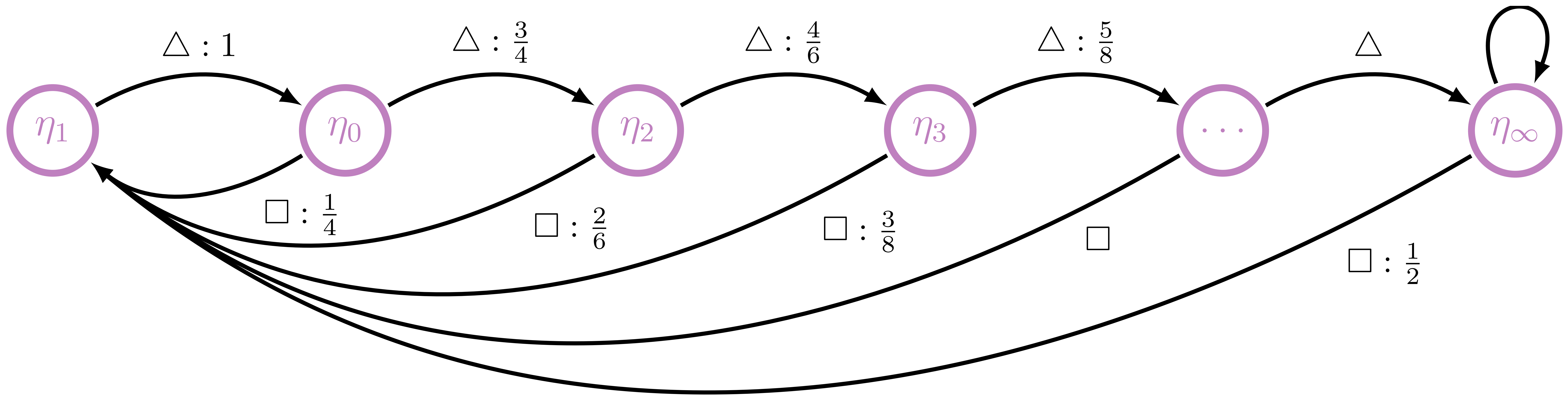
Simplex



Unifilar Model



Blackwell Entropy Formula

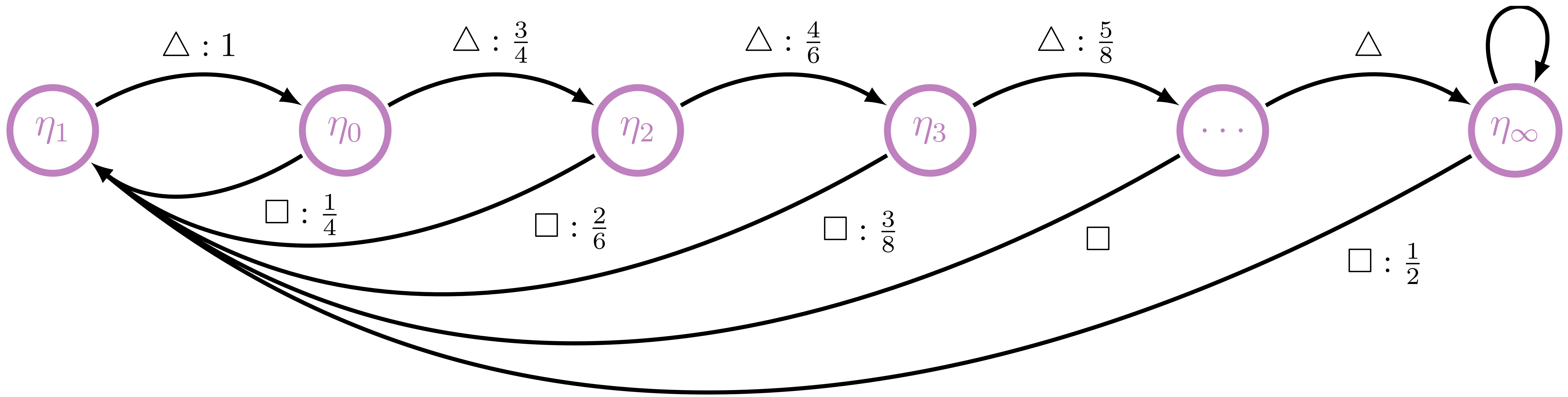


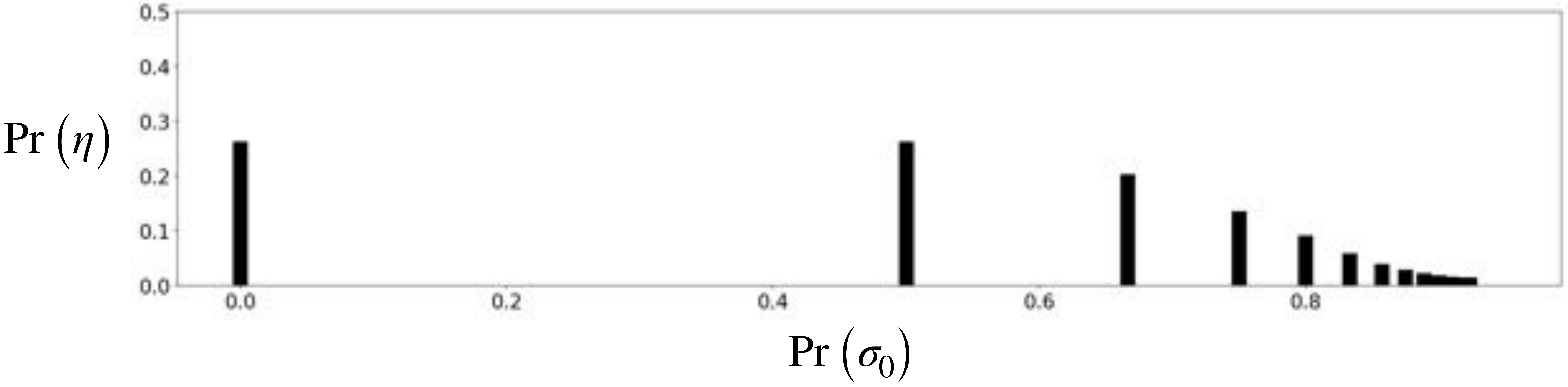
Can use mixed states to calculate entropy:

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

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

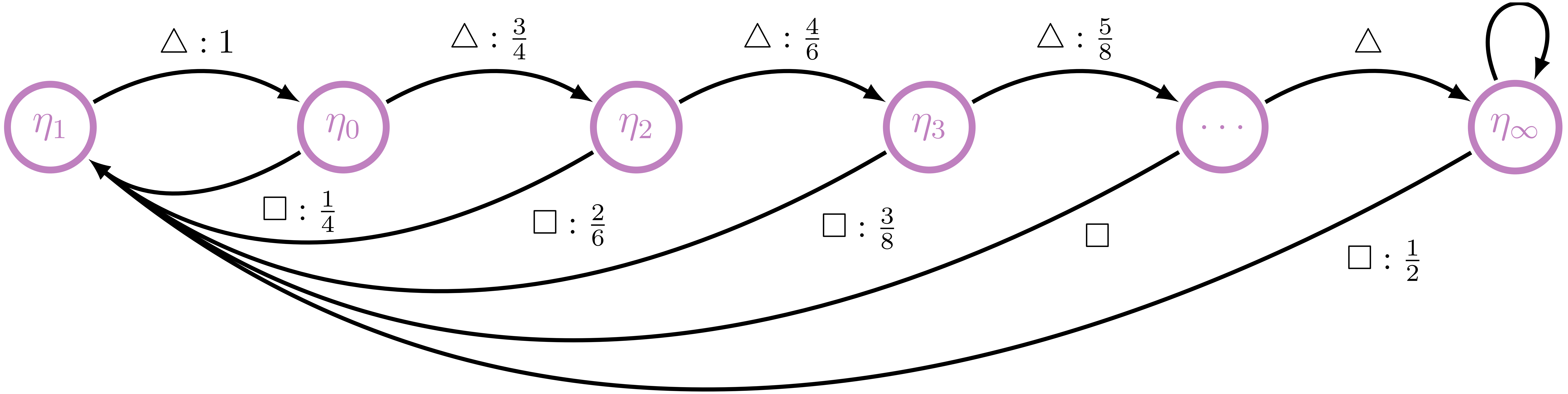
Blackwell Measure





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

Blackwell Entropy Formula

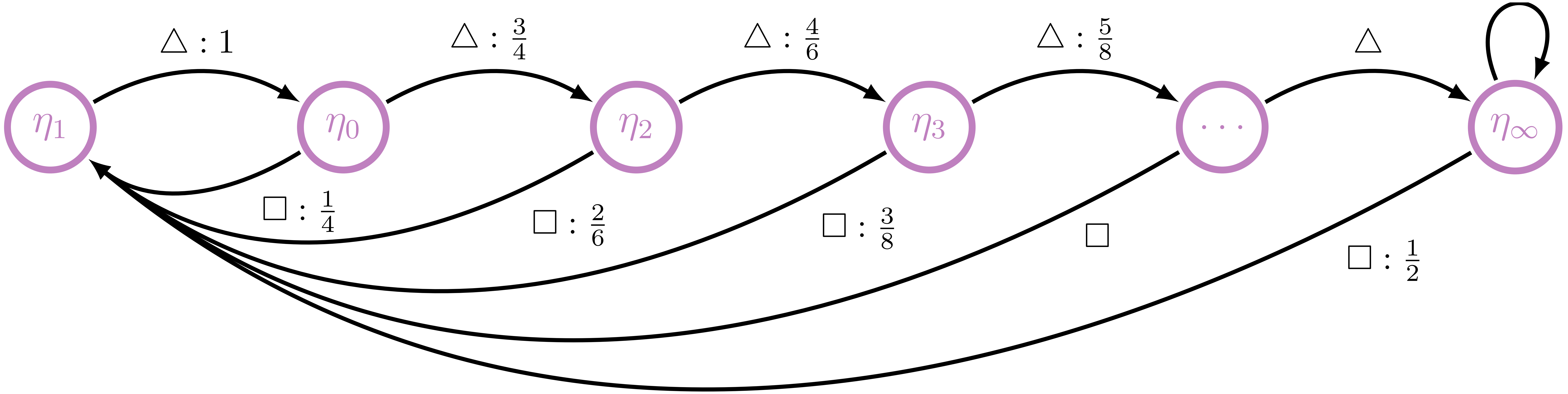


$$h_\mu = \int_R d\mu\left(\eta\right)\ H[x\,|\,\eta] = \lim_{N\to\infty} -\sum_n^N \Pr\left(\eta_n\right) \sum_{x\in A} \Pr\left(x\,|\,\eta_n\right) \log_2 \Pr\left(x\,|\,\eta_n\right)$$

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.

Blackwell Entropy Formula



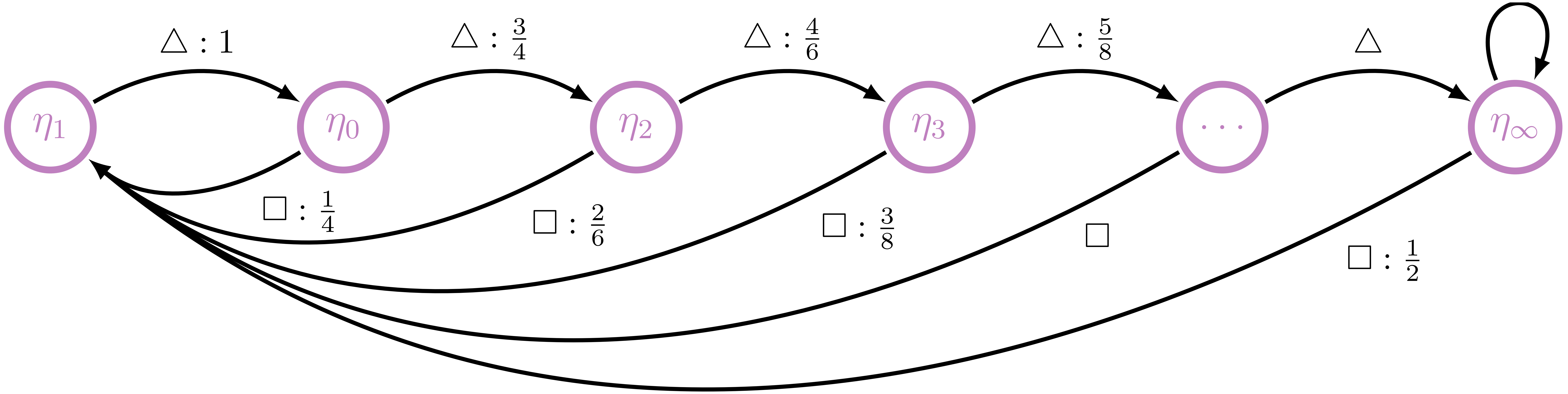
$$h_\mu = \int_R d\mu\left(\eta\right)\ H[x\,|\,\eta] = \lim_{N\to\infty} - \sum_n^N \Pr\left(\eta_n\right) \sum_{x\in A} \Pr\left(x\,|\,\eta_n\right) \log_2 \Pr\left(x\,|\,\eta_n\right)$$

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

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.

Statistical Complexity of Infinite Machines



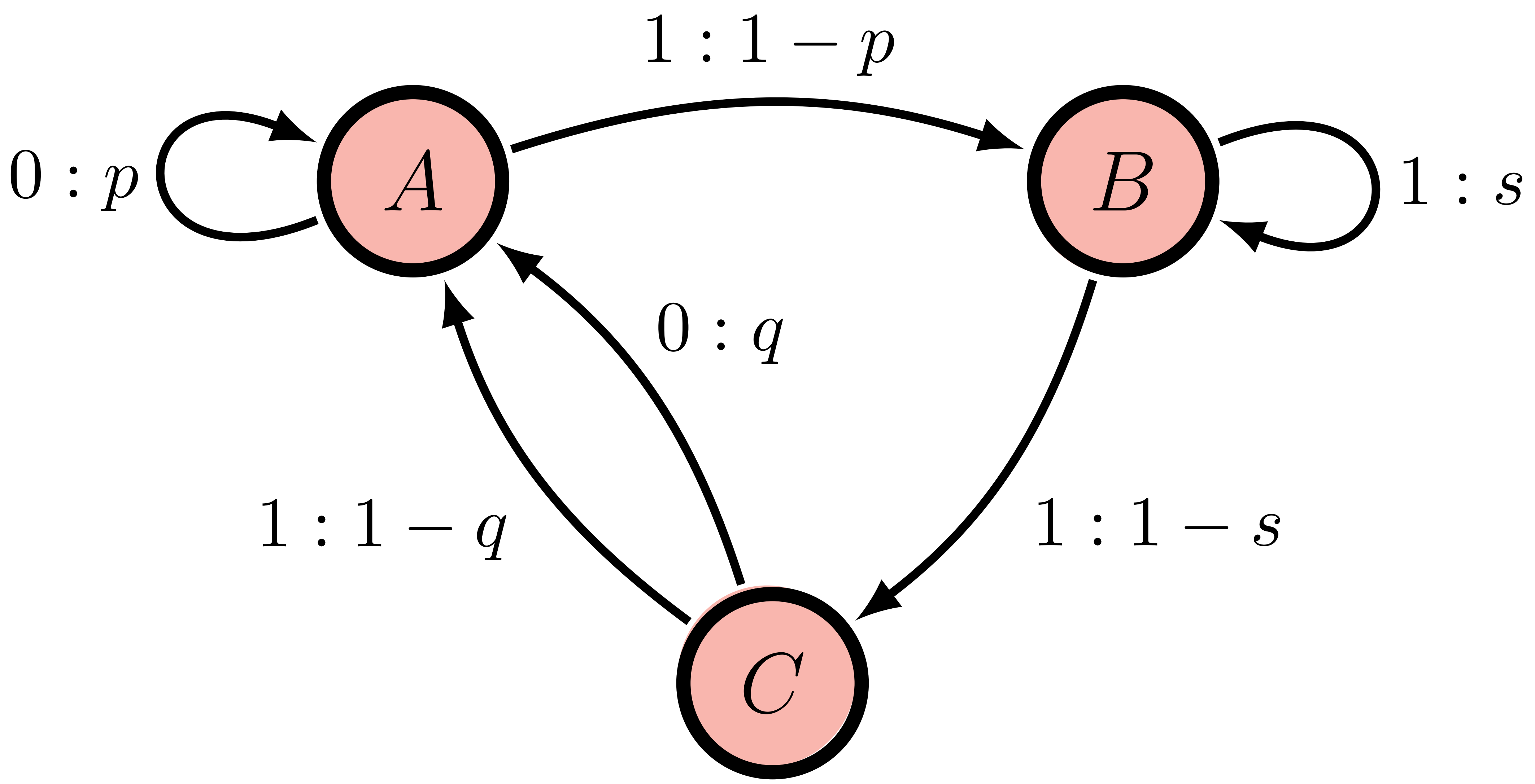
Can also calculate statistical complexity C_μ:

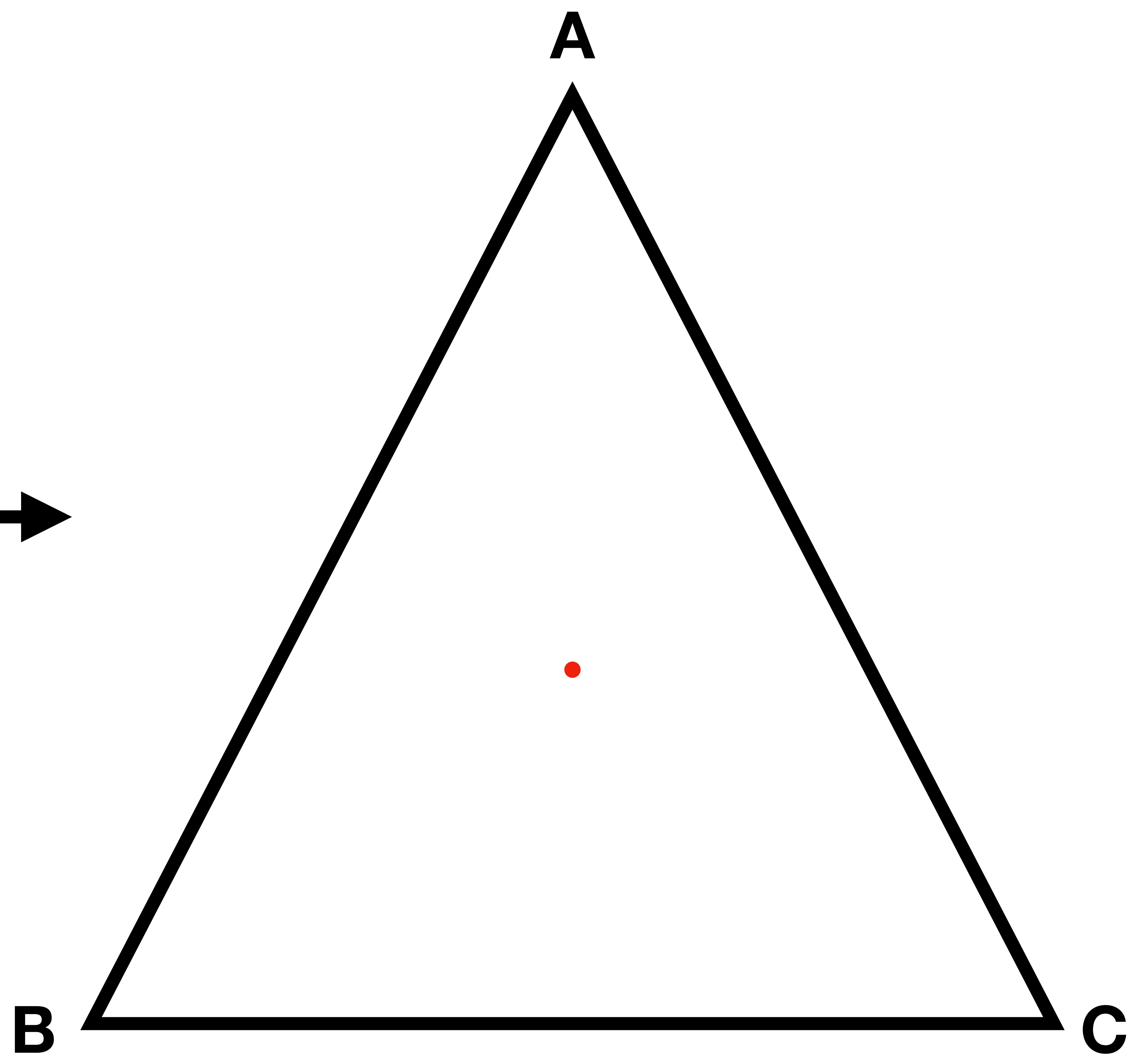
$$C_\mu = -\int_R \Pr(\eta) \log_2 \Pr(\eta)\, d\mu(\eta) = \lim_{N\to\infty} -\sum_n^N \Pr(\eta_n) \log_2 \Pr(\eta_n)$$

$$\longrightarrow \quad C_\mu \approx 2.71$$

Three State Mixed States

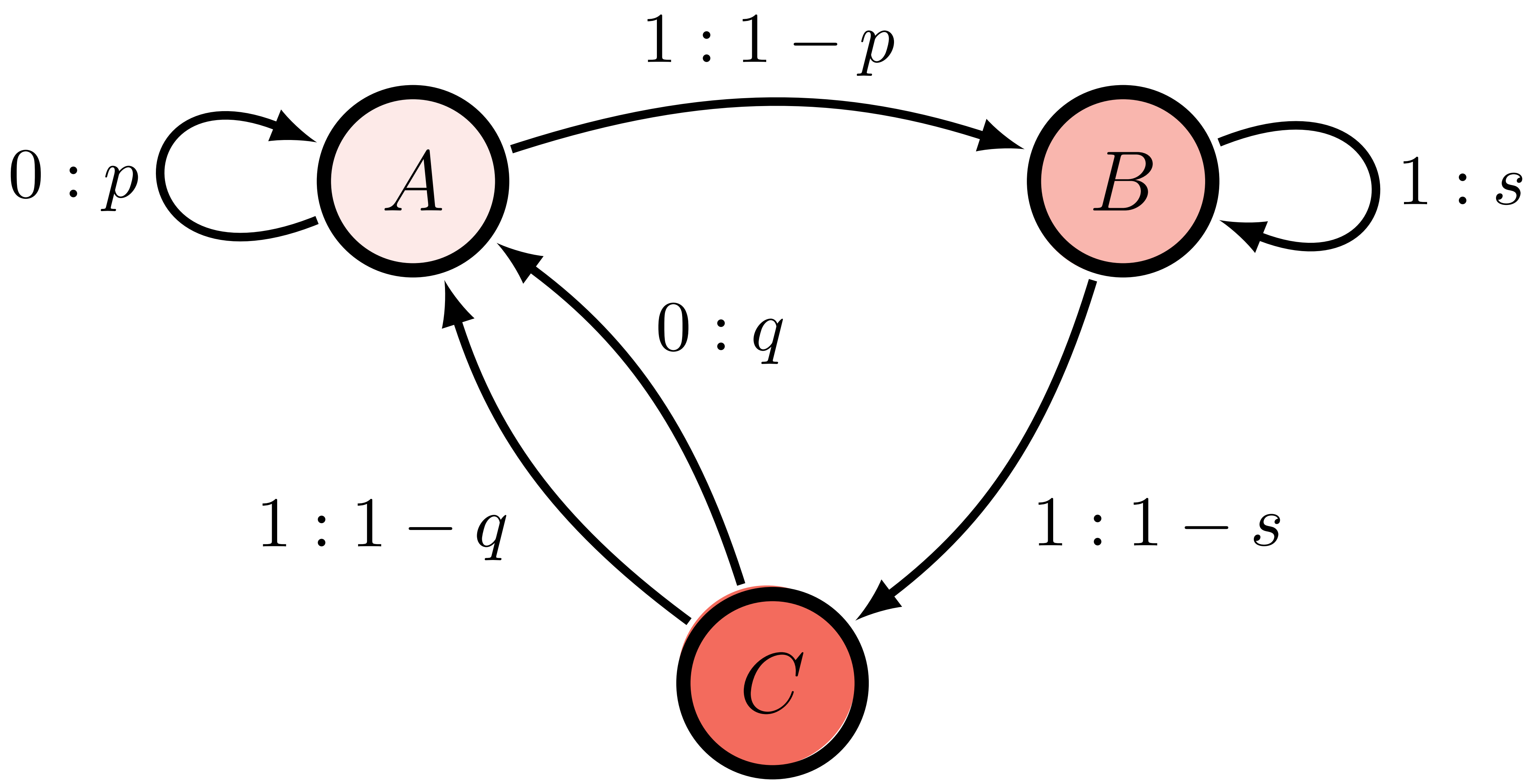
Find the mixed/belief states:

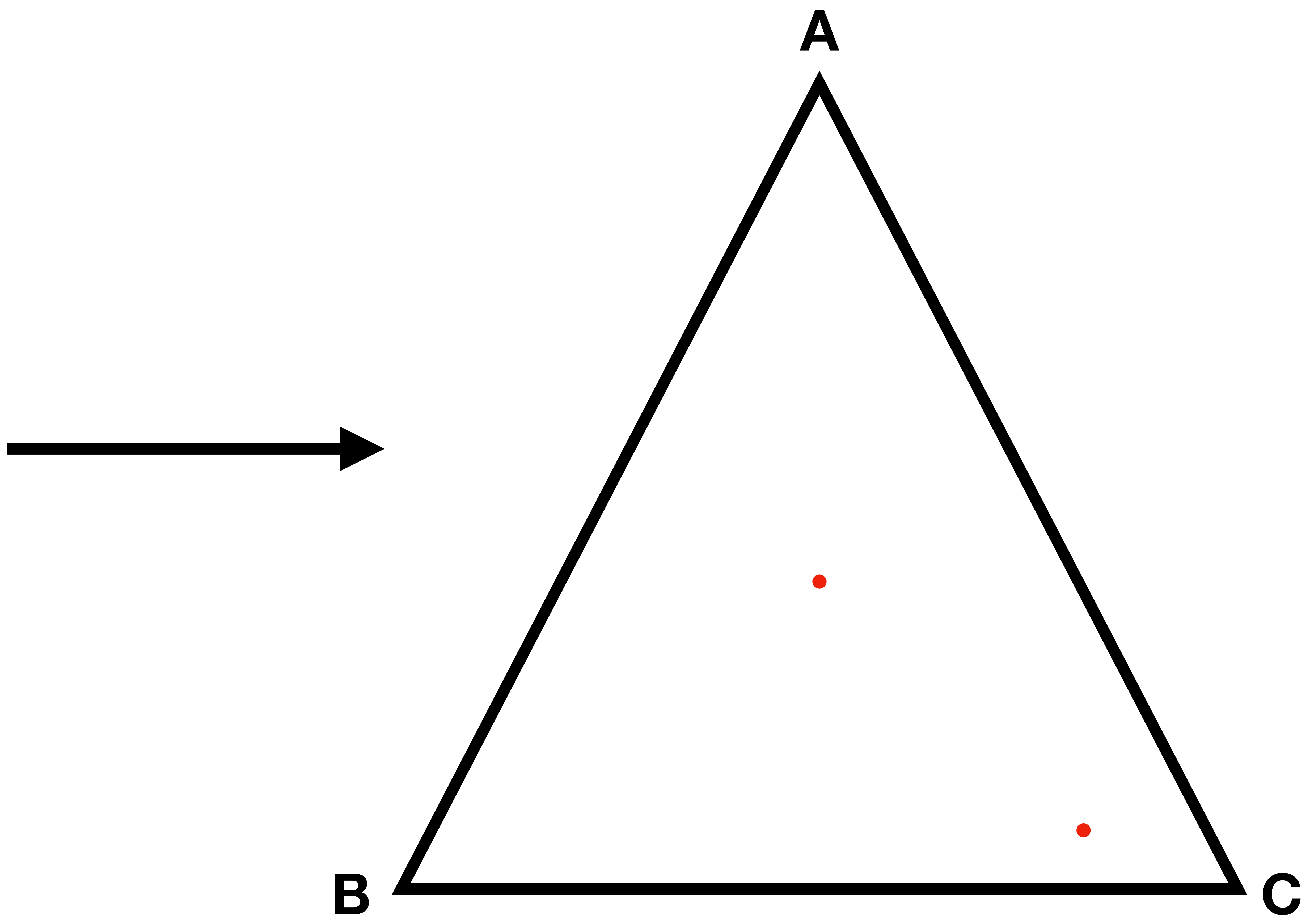




Observed sequence: λ

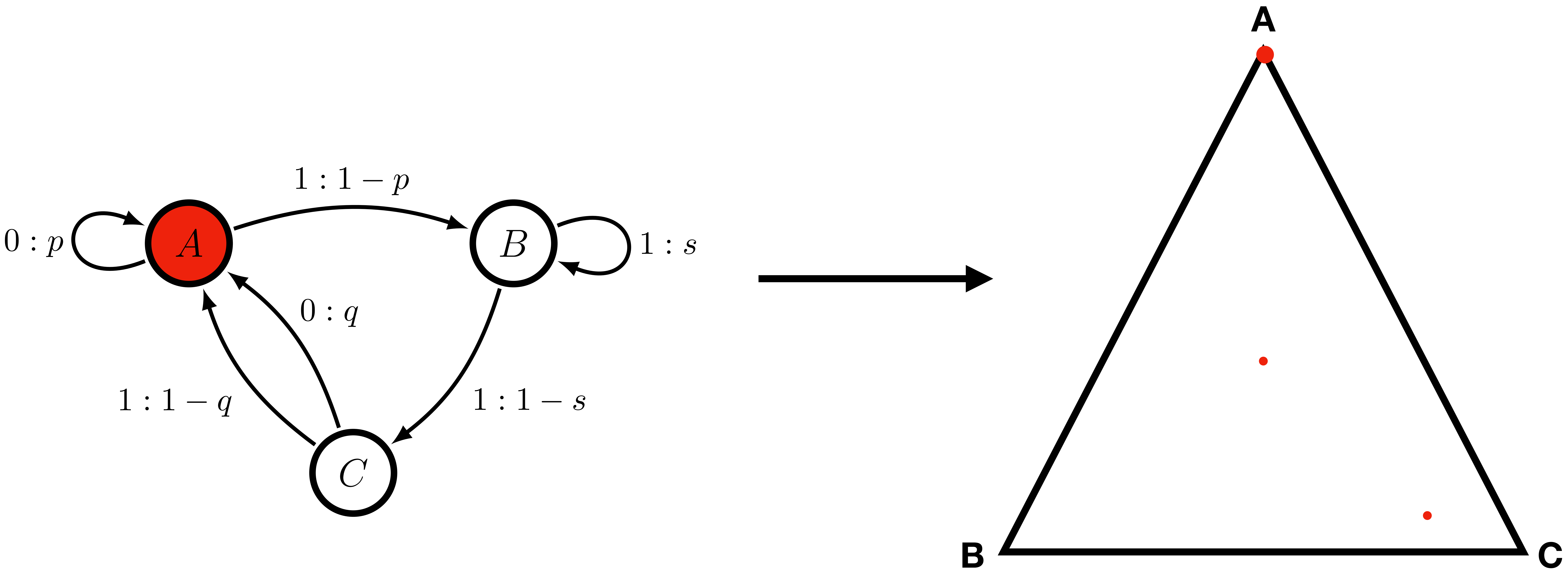
Three State Mixed States





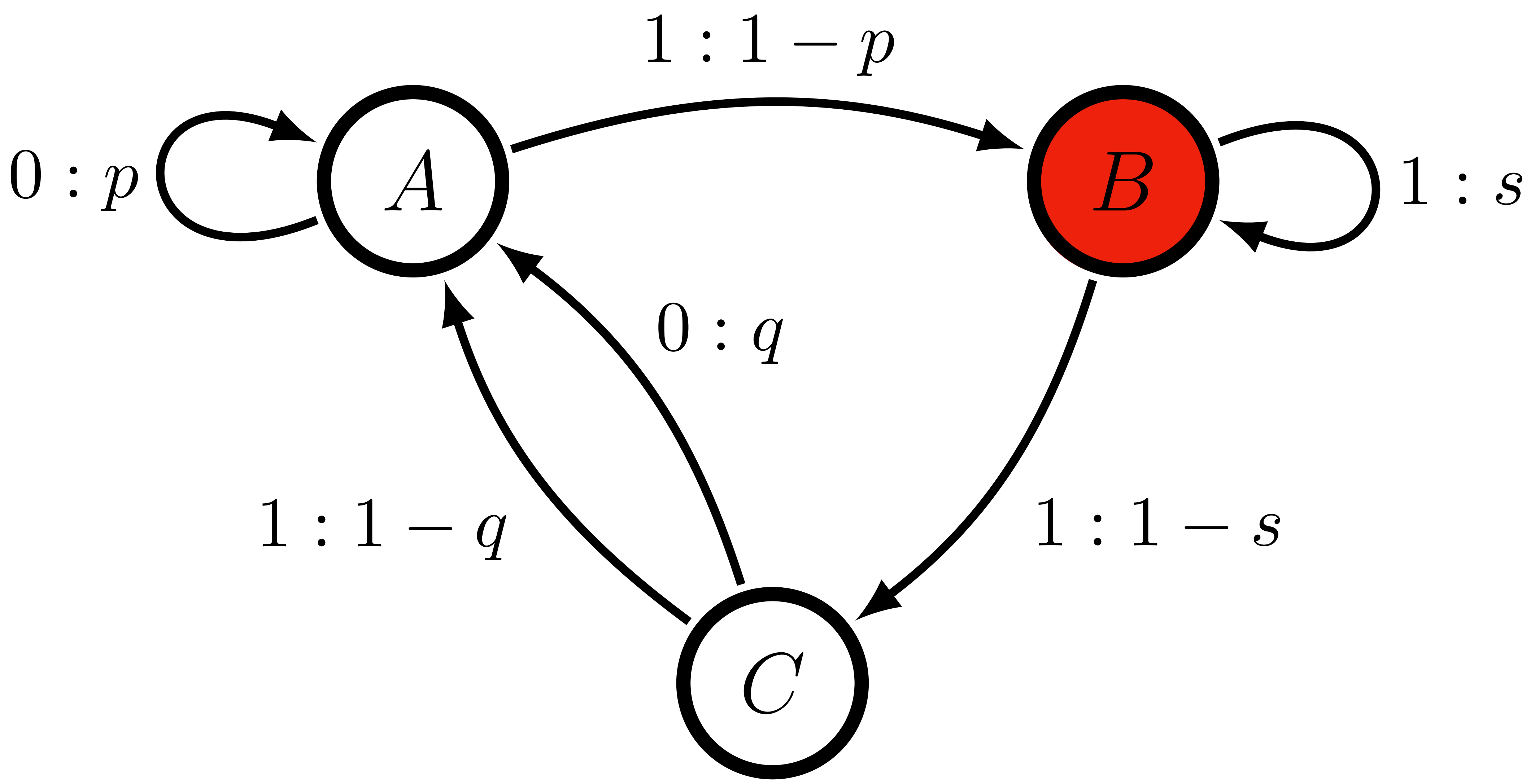
Observed sequence: $\lambda 1$

Three State Mixed States

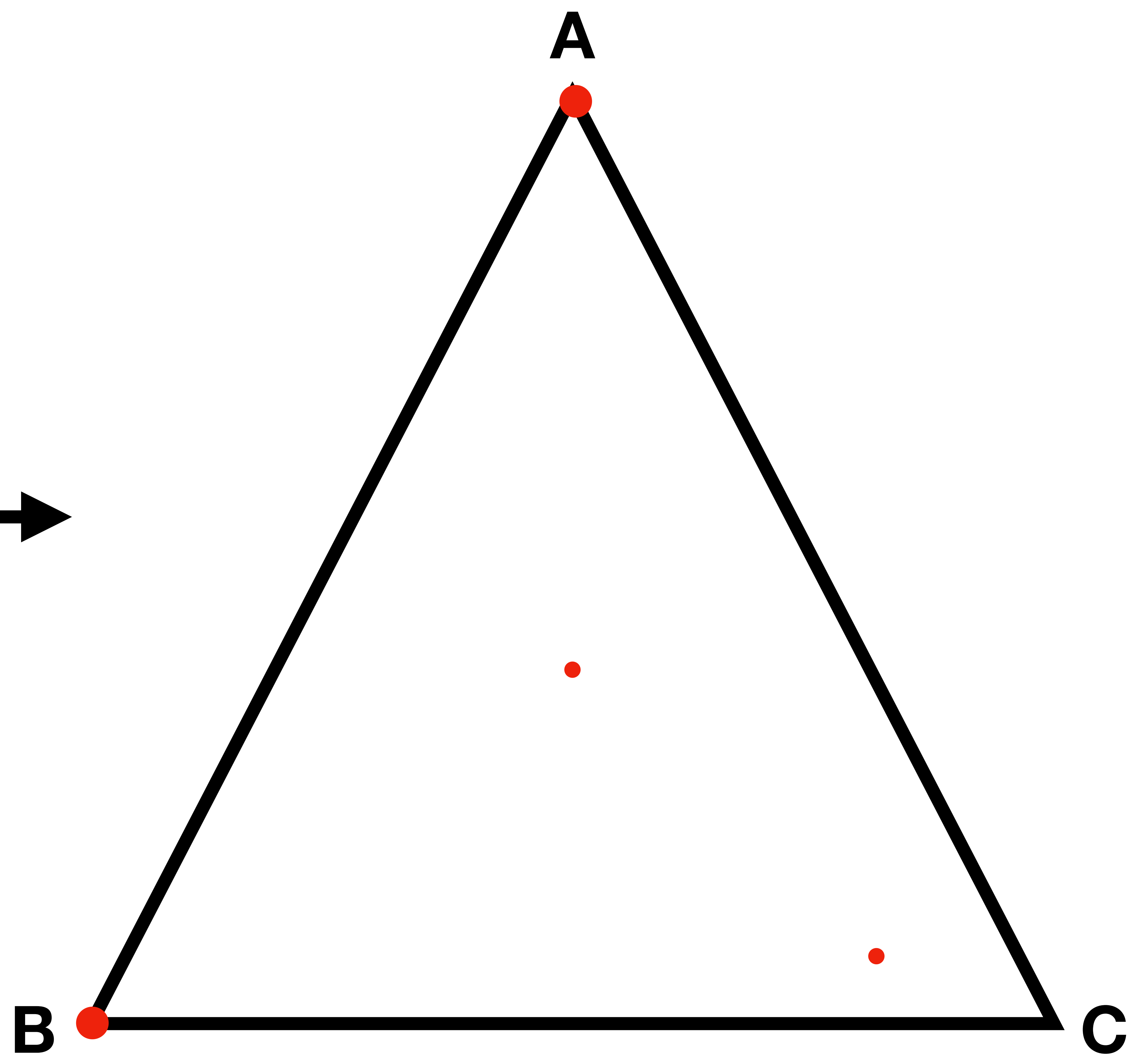


Observed sequence: $\lambda 10$

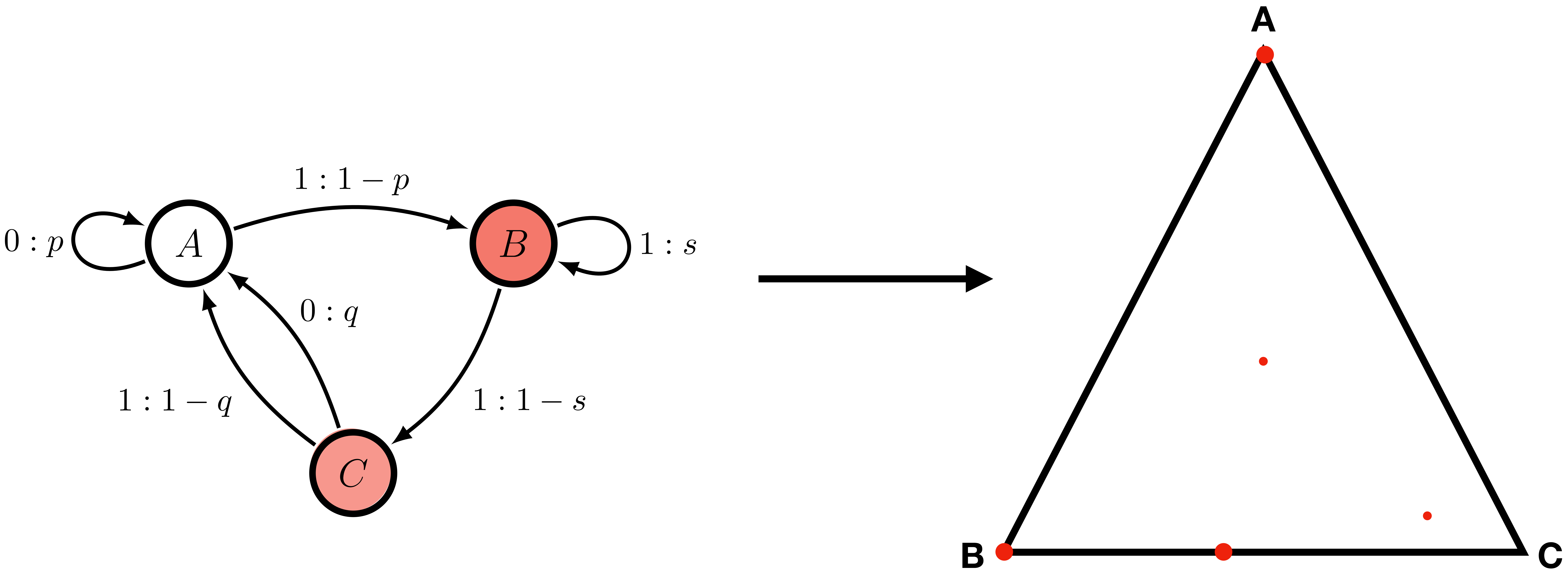
Three State Mixed States



Observed sequence: $\lambda 101$

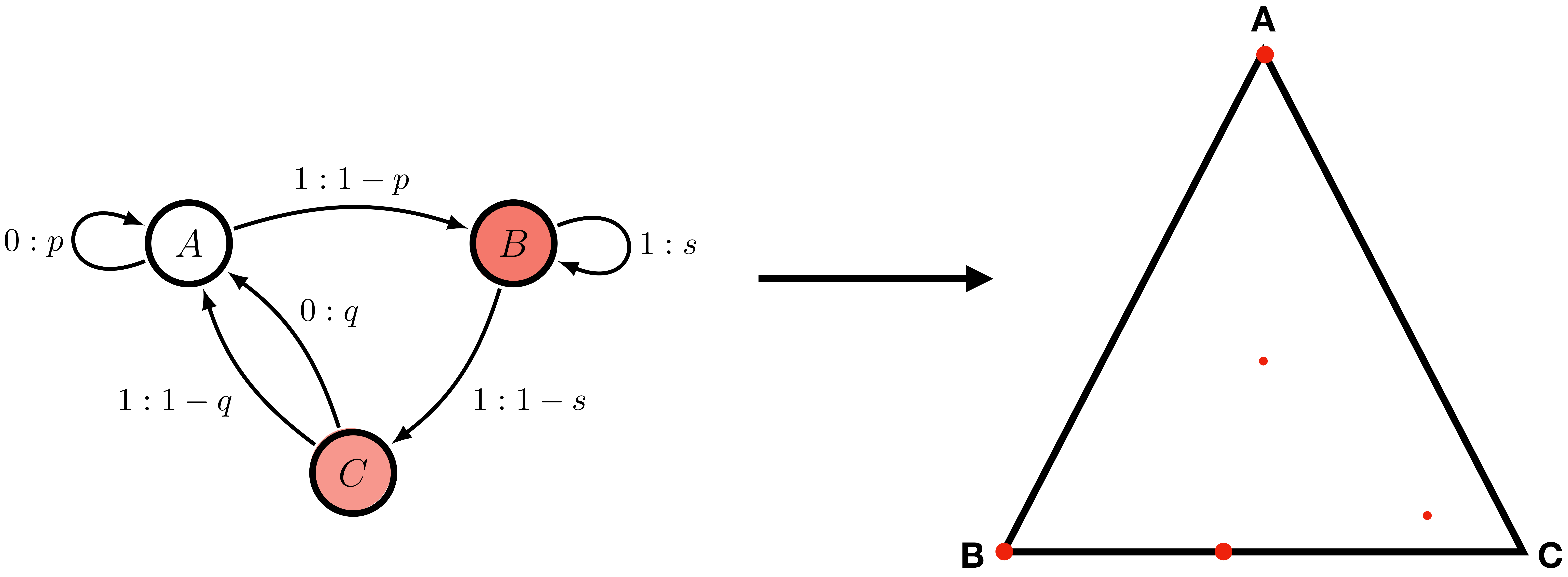


Three State Mixed States



Observed sequence: $\lambda 1011$

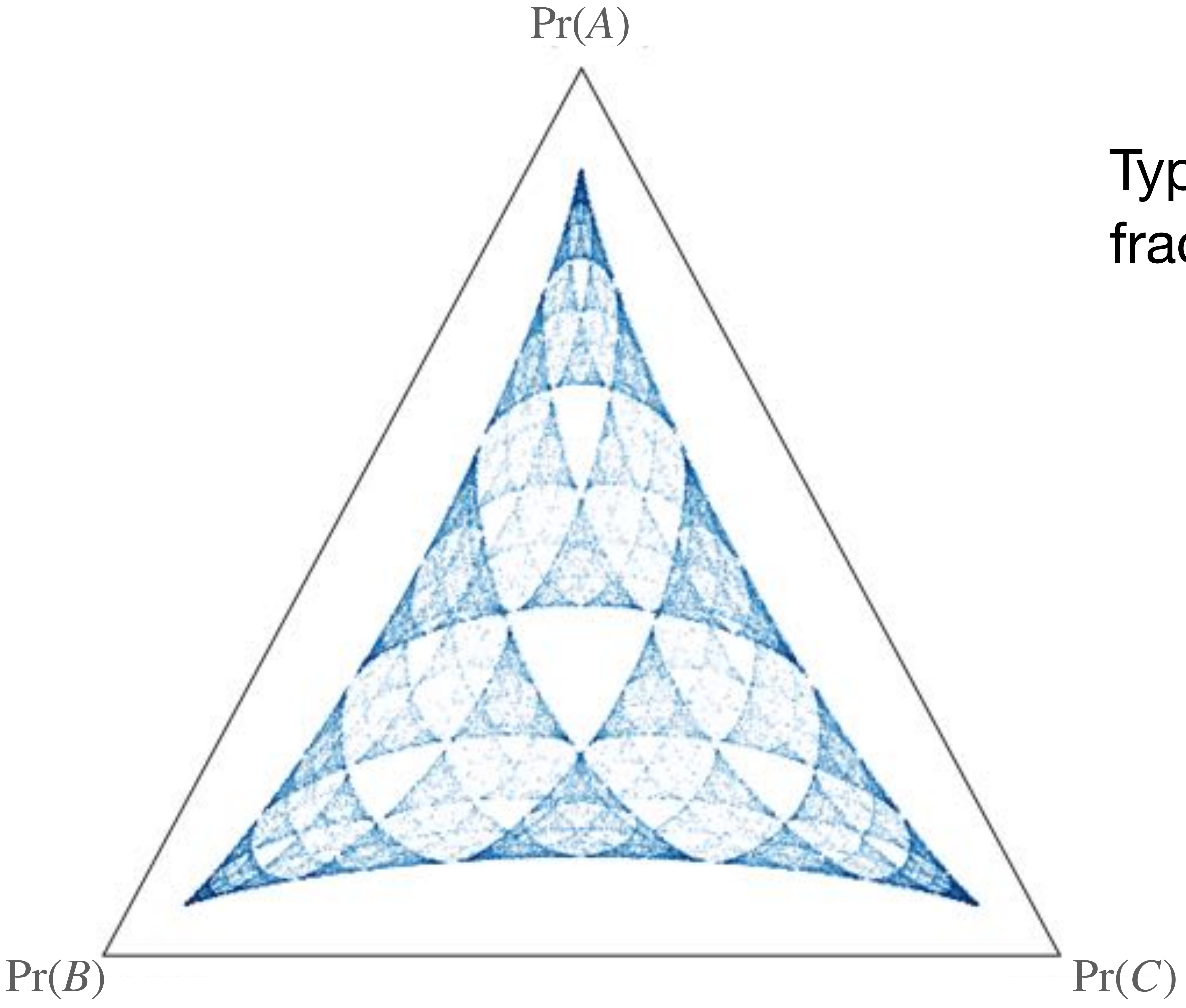
Three State Mixed States



→ Build up the set

$$R = \left\{ \eta : \eta(w) = \Pr\left(S_l | X_{0:l} = w, S_0 = \pi\right) \right\}$$ of belief states.

Mixed State Sets are Fractals

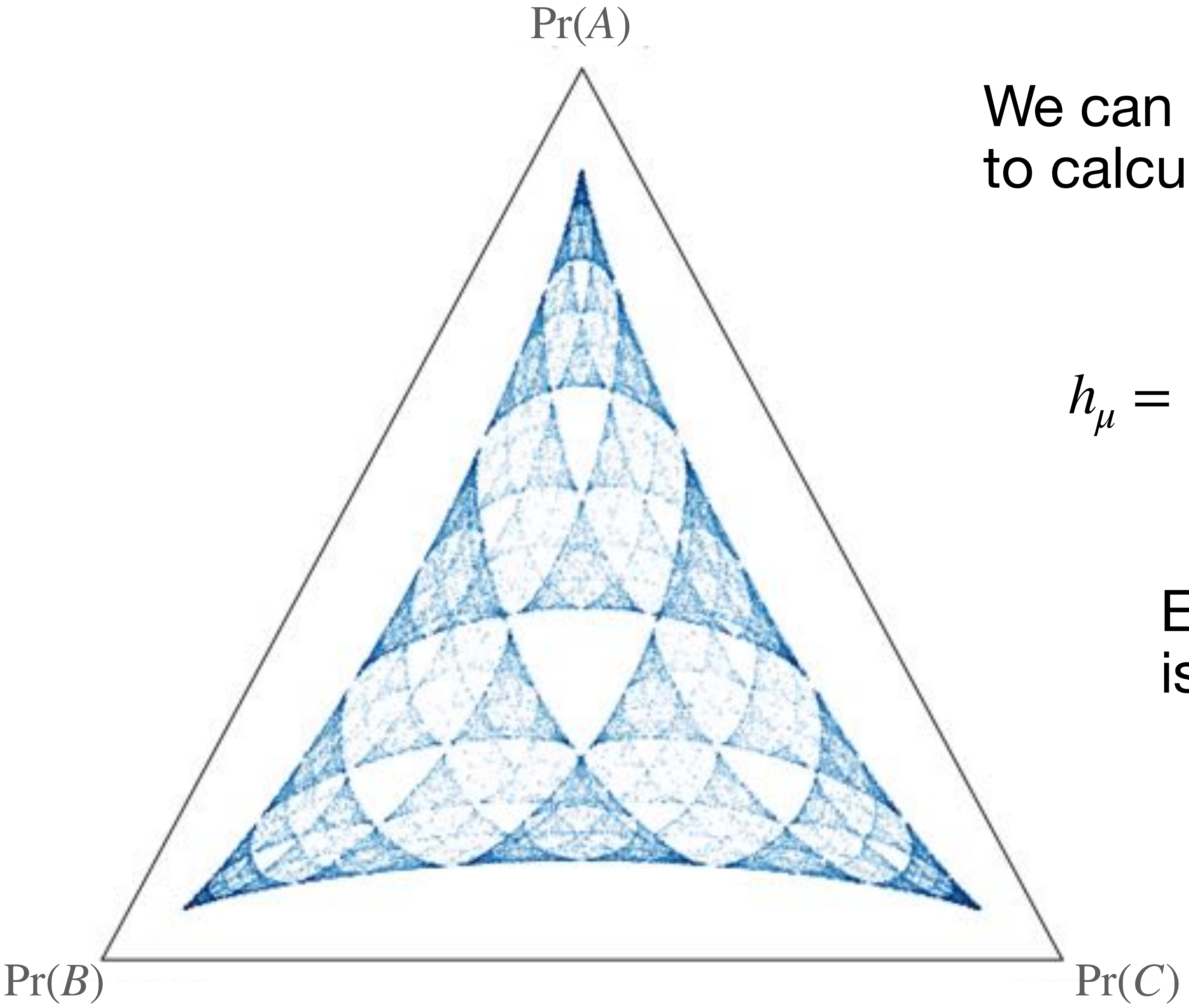


Typically, non-unifilar HMMs result in a fractal mixed state set R.

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).

Marzen, S.E.; Crutchfield, J.P. (2017) Nearly maximally predictive features and their dimensions. Phys. Rev. E 95, 051301(R)

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.

Iterated Function Systems

An iterated function system is a set of contractive mappings on a space X

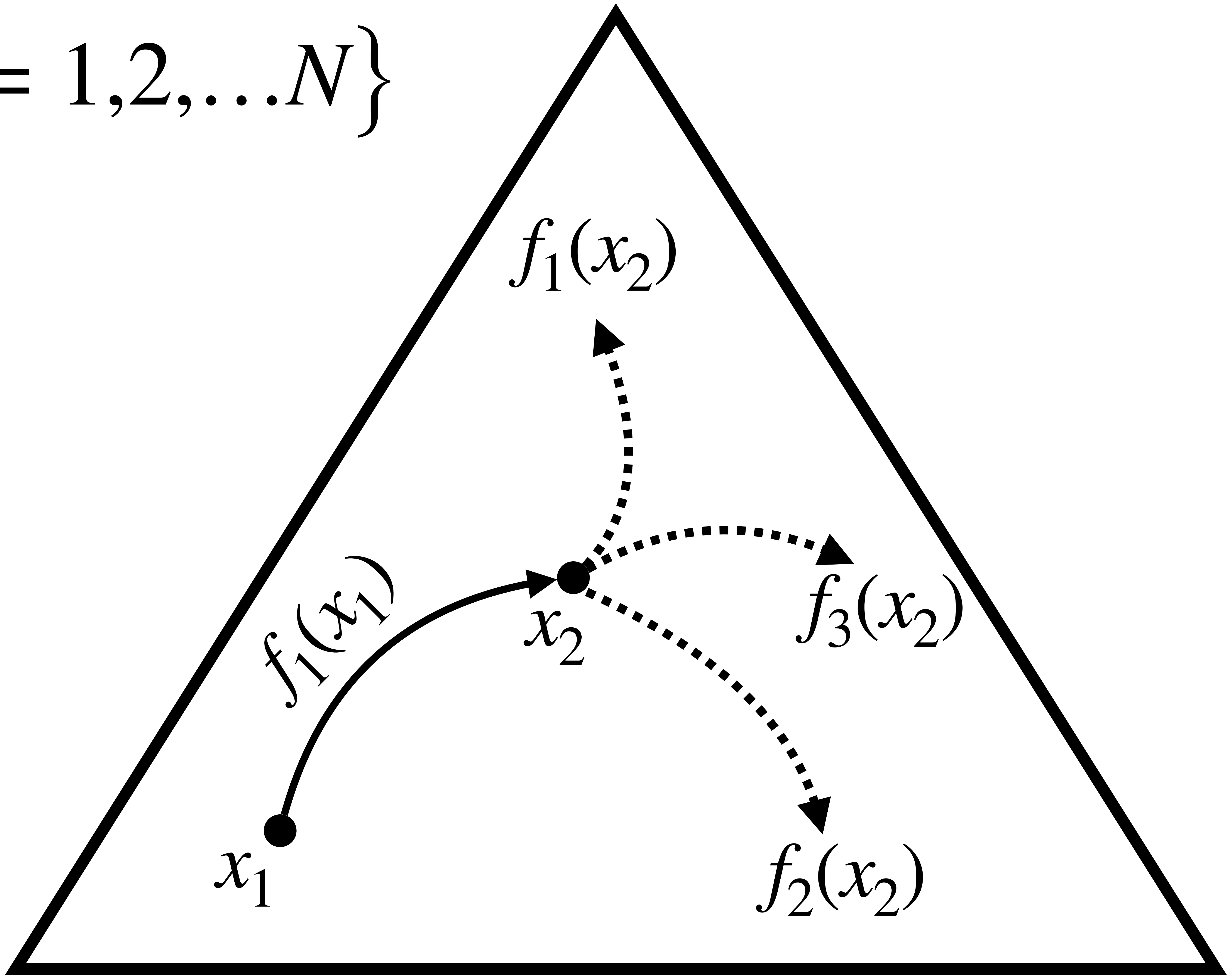
$$\left\{f_i : X \to X \,|\, i = 1,2,\ldots N\right\}$$

Iterated Function Systems

An iterated function system is a set of contractive mappings on a space X

$$\{f_i : X \to X \,|\, i = 1,2,\ldots N\}$$

The "chaos game" is to apply one mapping chosen at random, then another, then another, until the shape of the fractal is traced out.



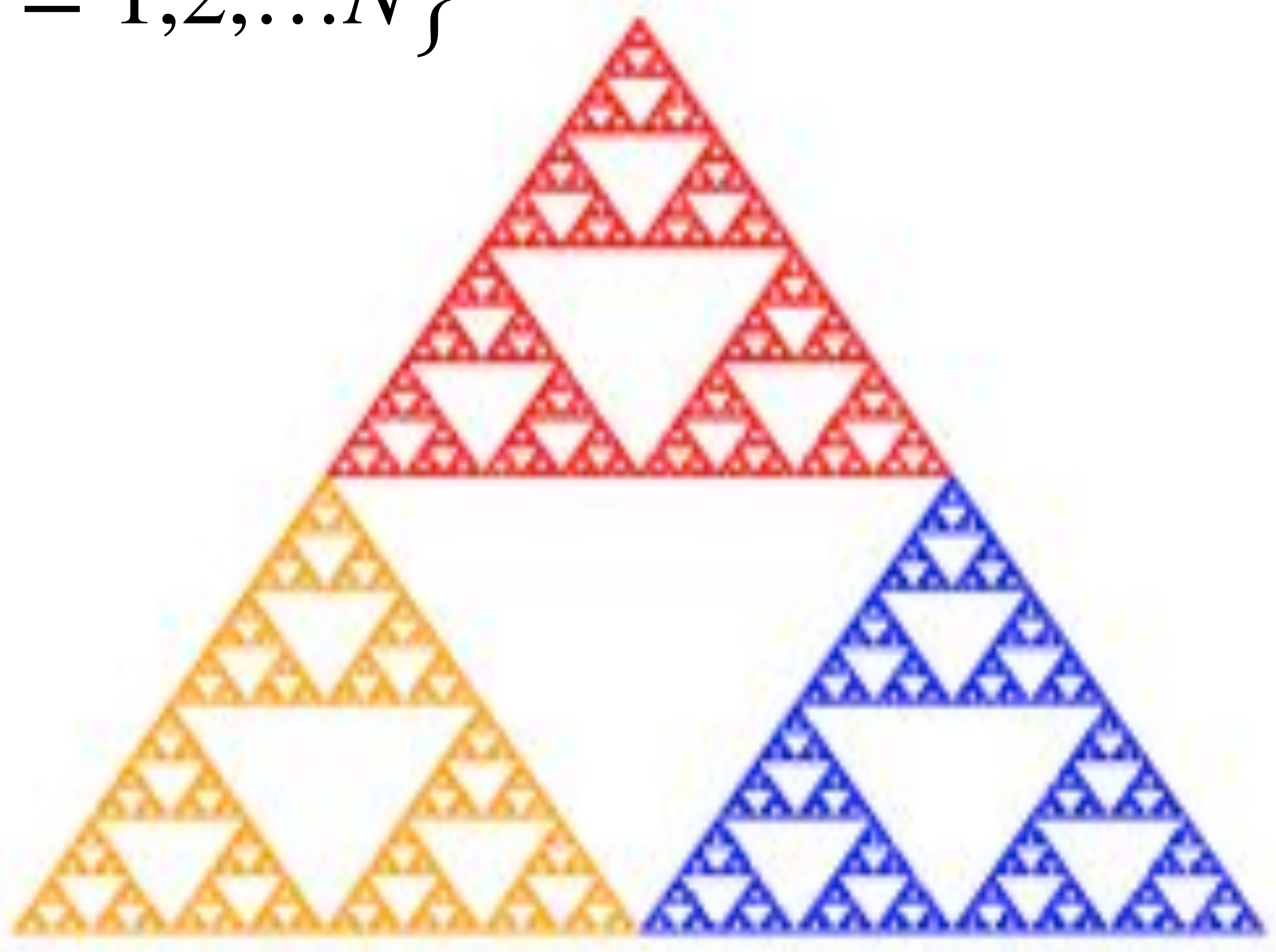
Iterated Function Systems

An iterated function system is a set of contractive mappings on a space X

$$\{f_i : X \to X \,|\, i = 1,2,\ldots N\}$$

The “chaos game” is to apply one mapping chosen at random, then another, then another, until the shape of the fractal is traced out.

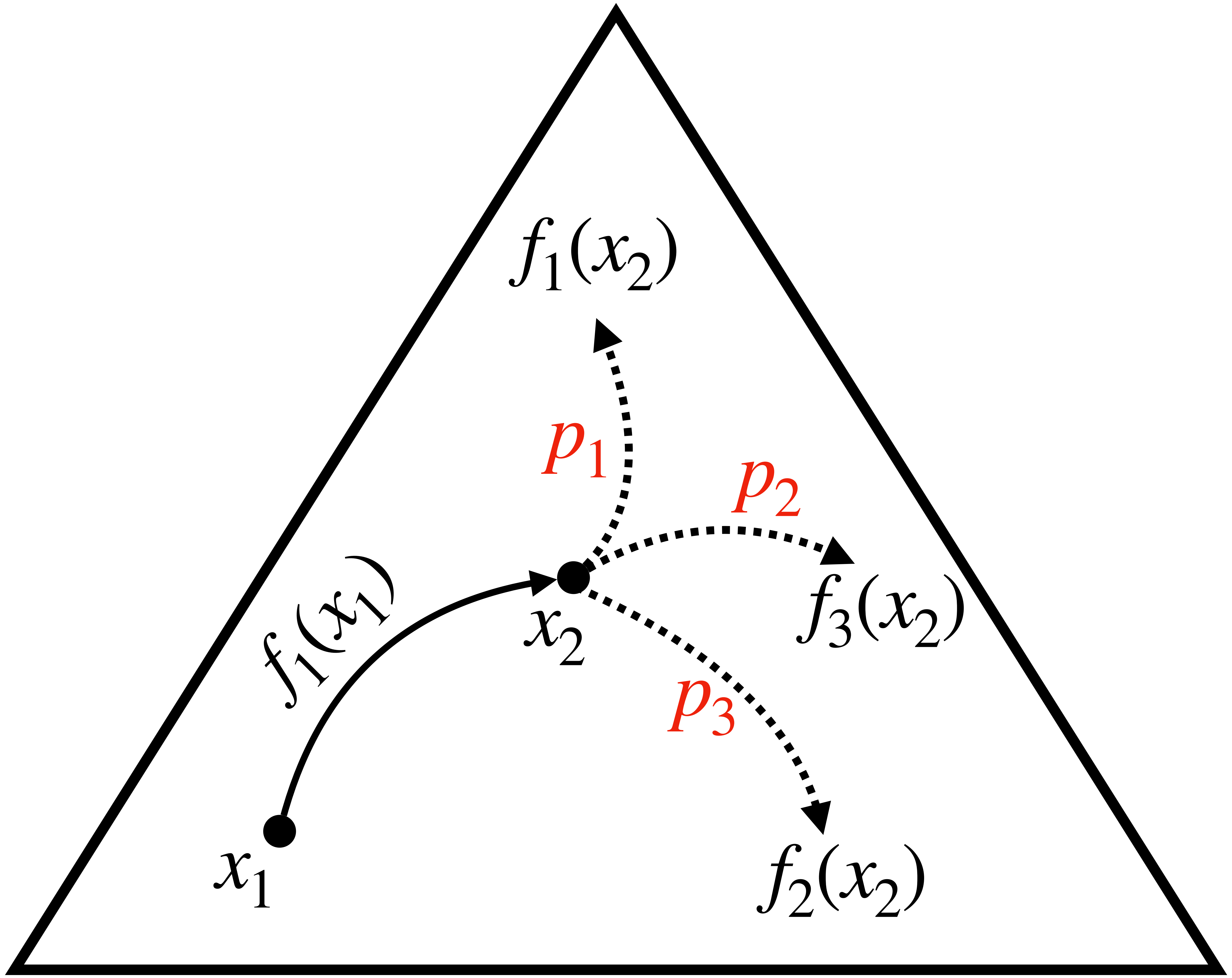
Sierpinski triangle can be understood as three contractive mappings on the triangle.



Weighted IFSs

A slightly more complicated version of an IFS pairs each function with a probability, so that the chaos game instead uses a weighted die

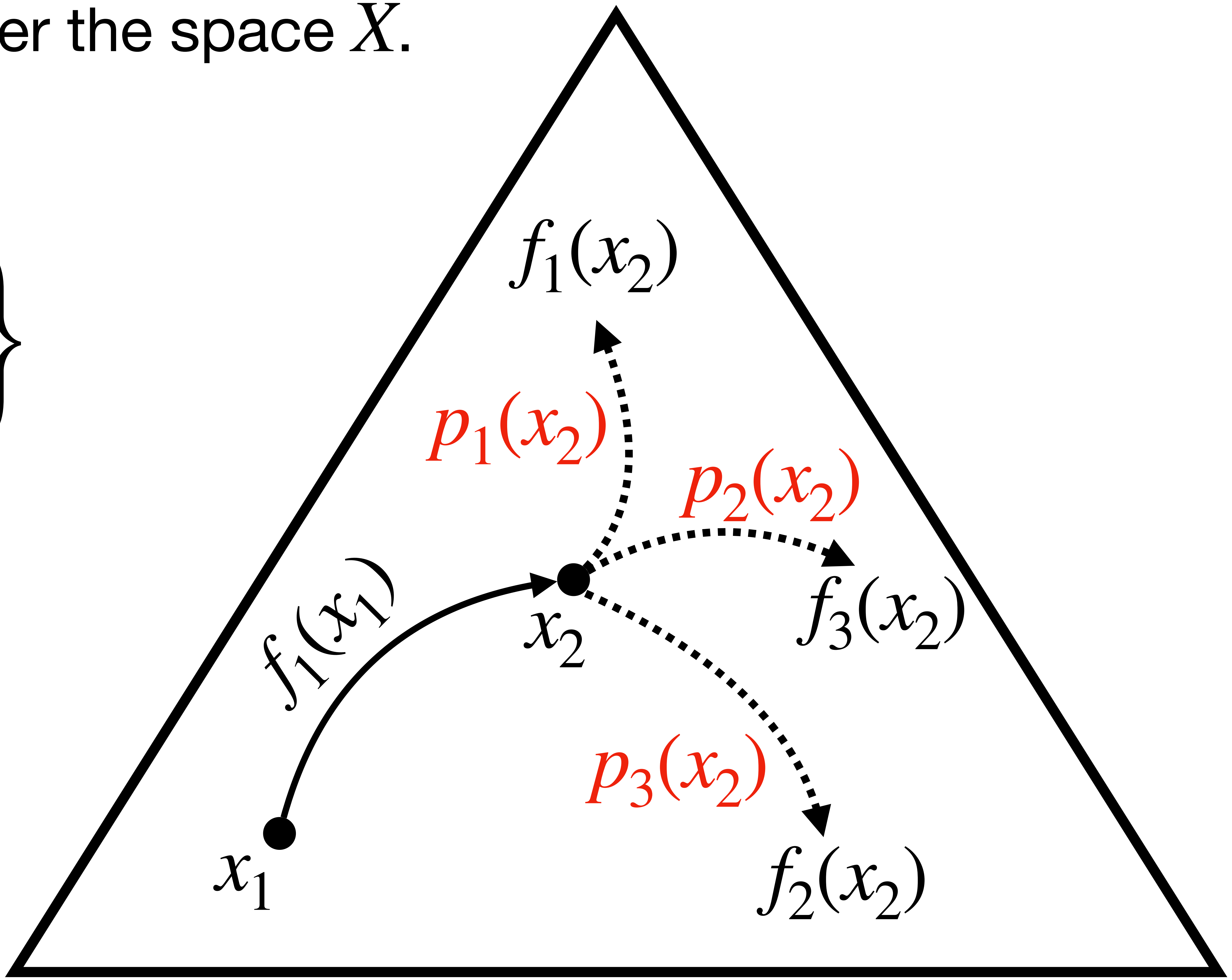
$$\left\{ \begin{matrix} f_i : X \to X \\ p_i \end{matrix} \;\middle|\; \begin{matrix} i = 1,2,\ldots N \\ \Sigma_i \, p_i = 1 \end{matrix} \right\}$$



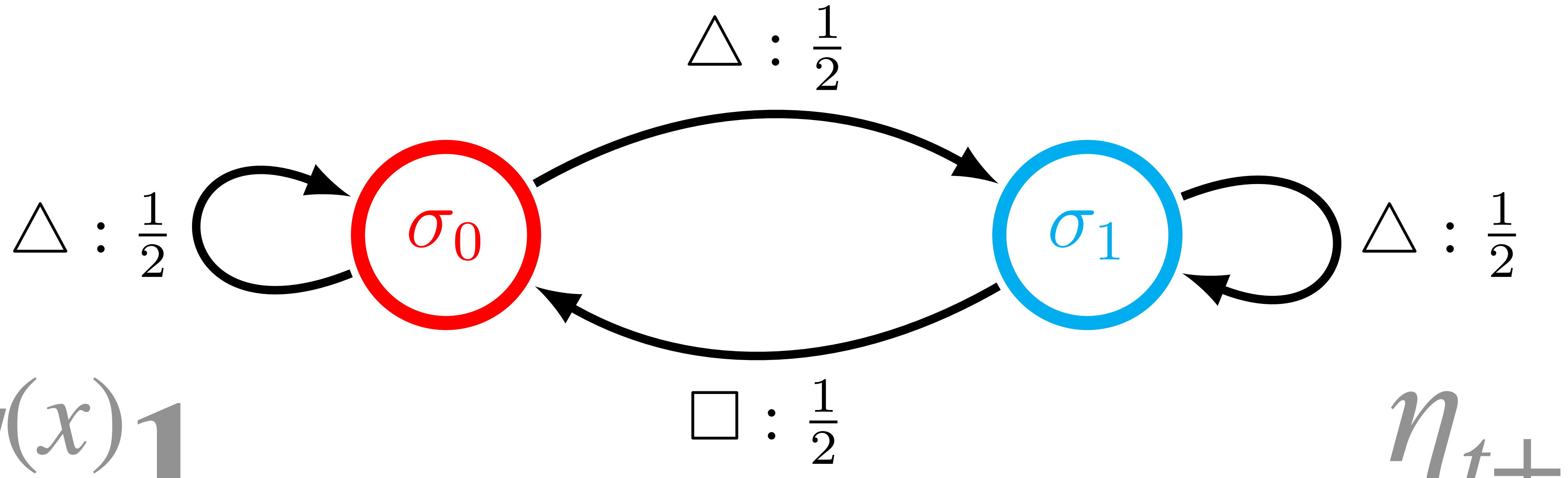
Place Dependent Weighted IFSs

Our IFSs are called “place dependent weighted IFSs” and replaces the static probability with a function over the space X.

$$\left\{ \begin{array}{r} f_i : X \to X \\ p_i : X \to [0,1] \end{array} \middle| \begin{array}{r} i = 1,2,\ldots N \\ \Sigma_i\, p_i = 1 \end{array} \right\}$$



Mixed State Generation → IFS



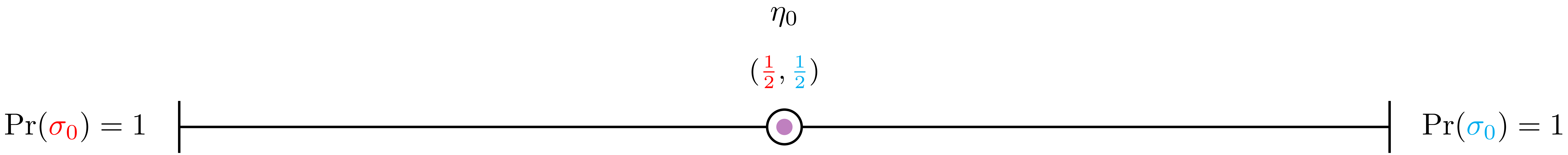
$$\Pr(x \mid \eta_t) = \eta_t T^{(x)} \mathbf{1}$$

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

↓

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

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

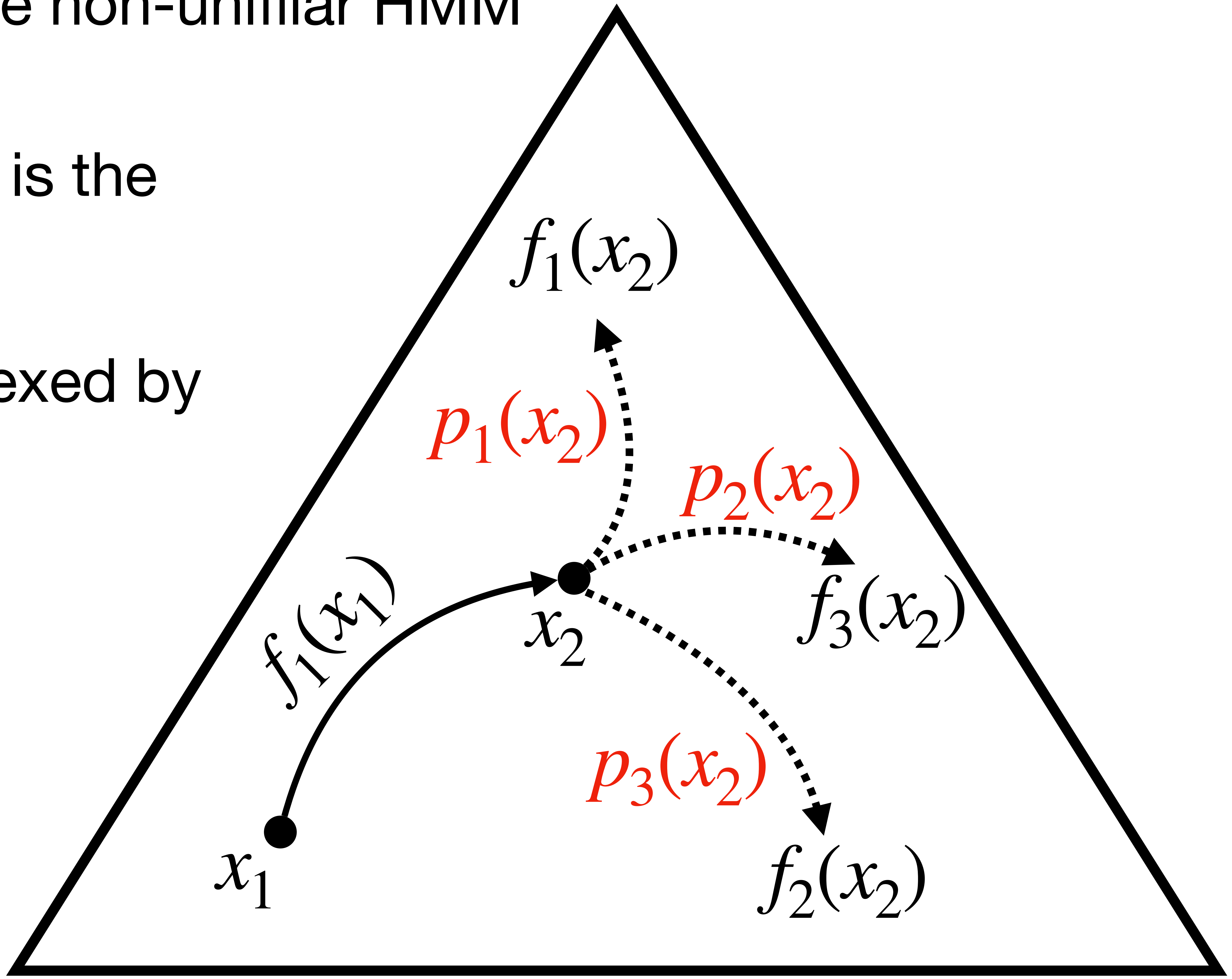


Markov-Driven IFS

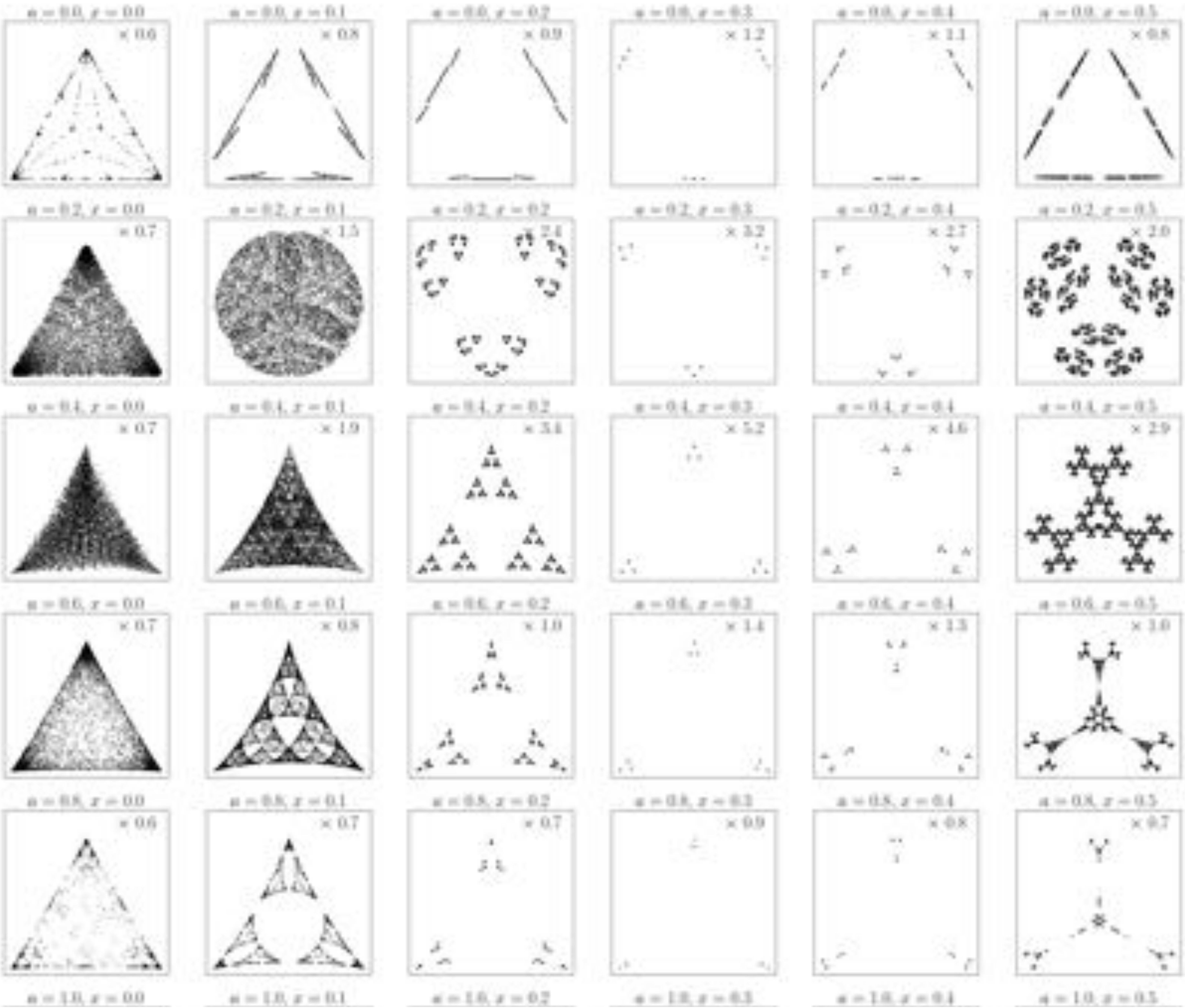
Concretely, we call our IFSs “Markov driven” (DIFSs) because the underlying map choice is derived from the non-unifilar HMM

The *space* is the N-Simplex Δ^N where N is the number of states minus one

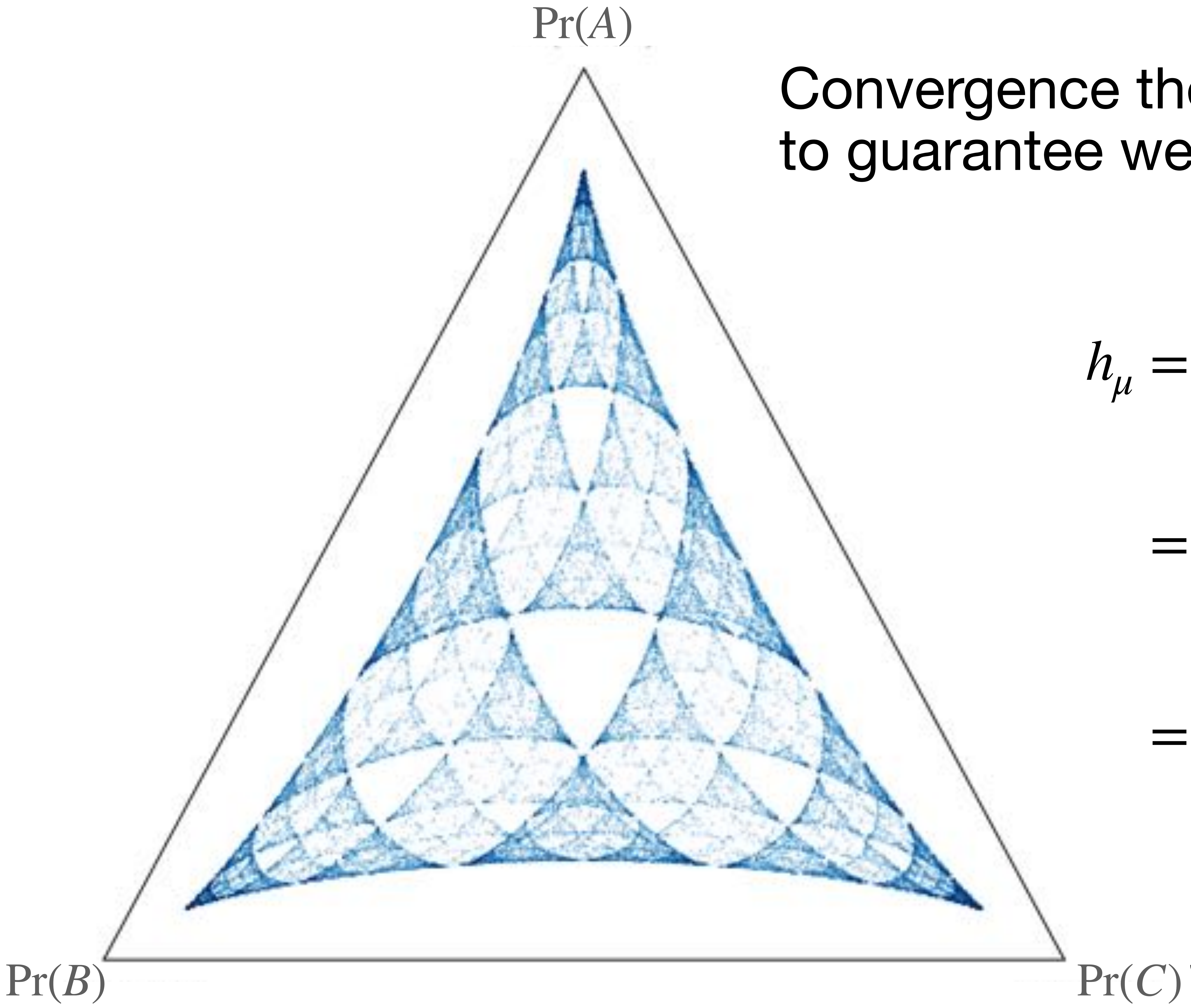
The mapping functions f^x and p^x are indexed by elements of the alphabet $x \in A$



The Fractal Zoo



Entropy Rate of Fractal HMMs



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

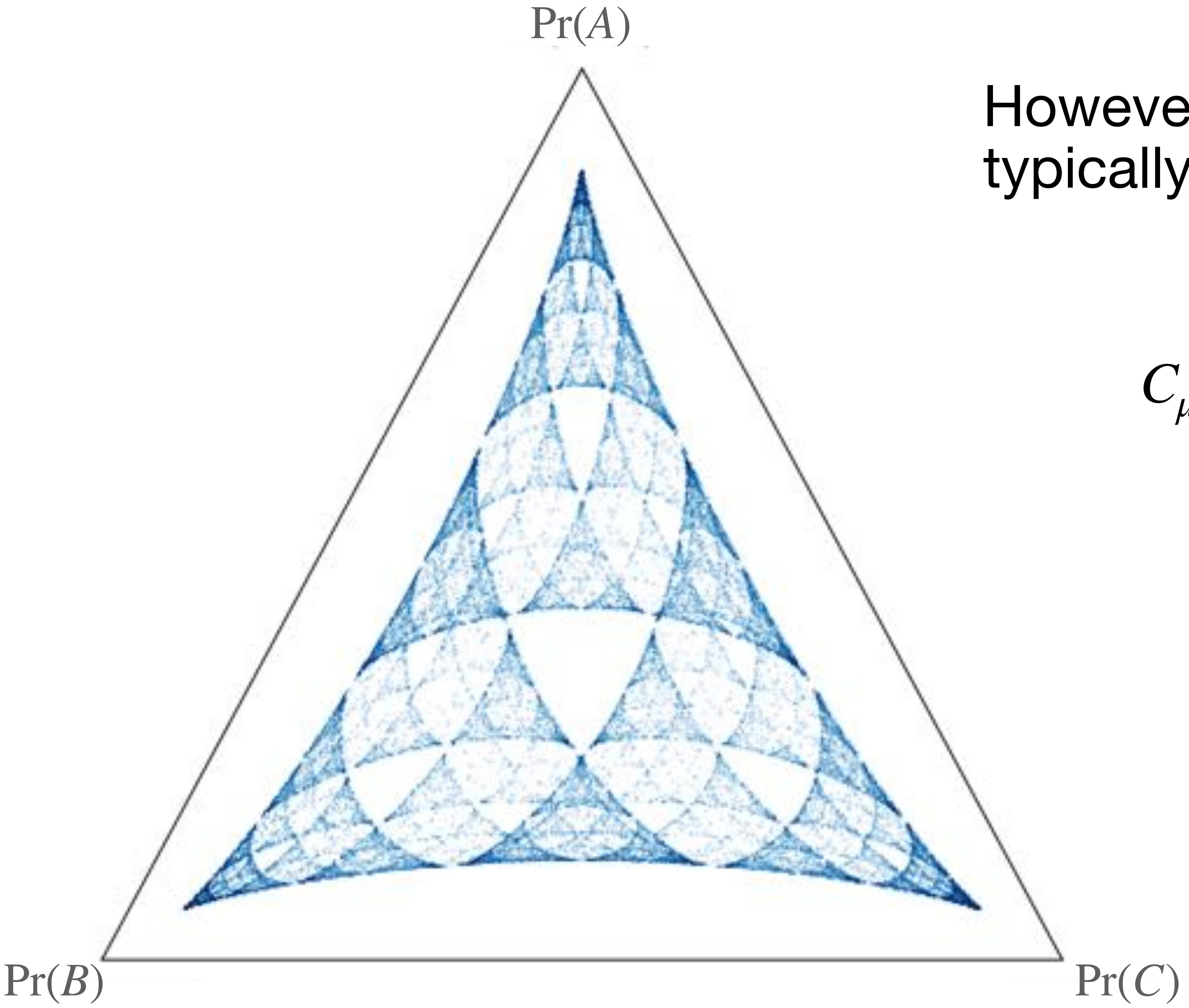
$$h_\mu = \int_R d\mu(\eta)\, H[x \mid \eta]$$

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

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

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



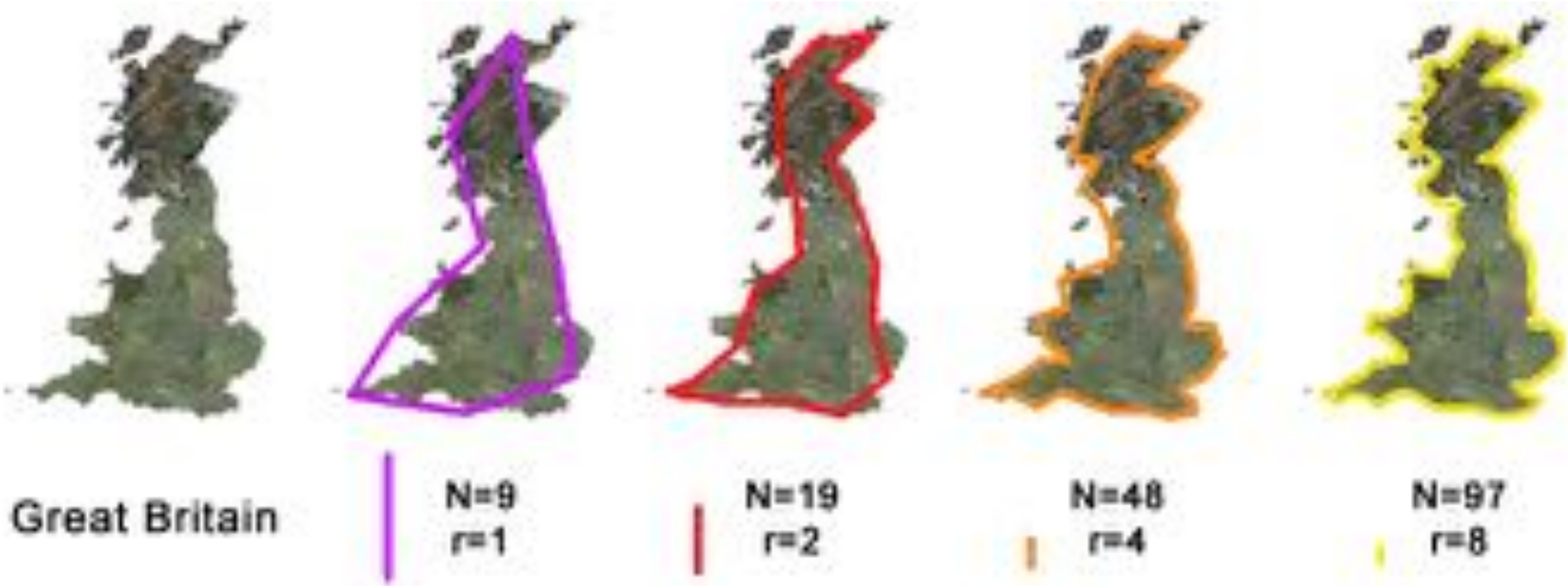
However, the statistical complexity still typically diverges.

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

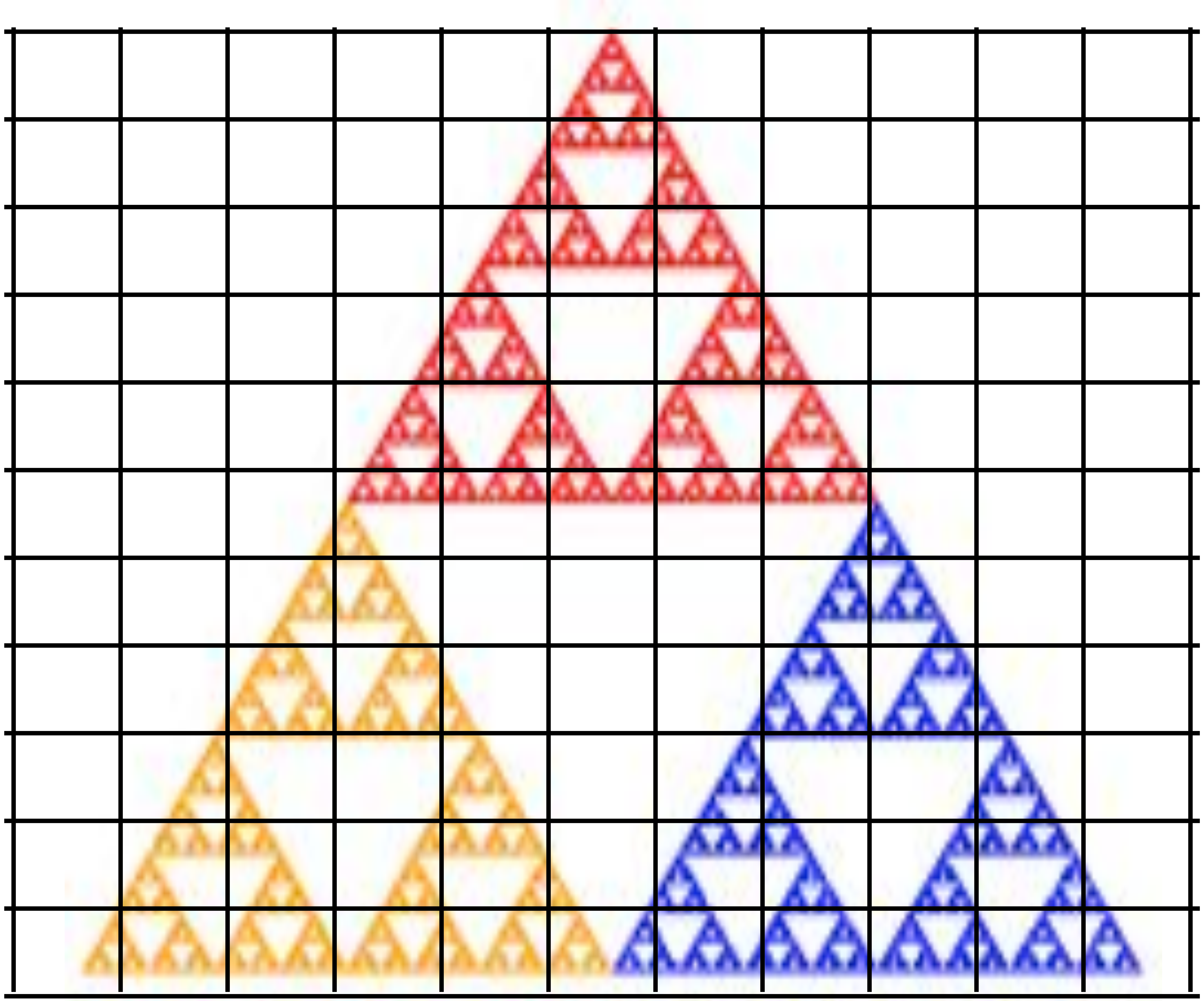
Infinite complexity?

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).

The Coastline Paradox



Fractal Dimension

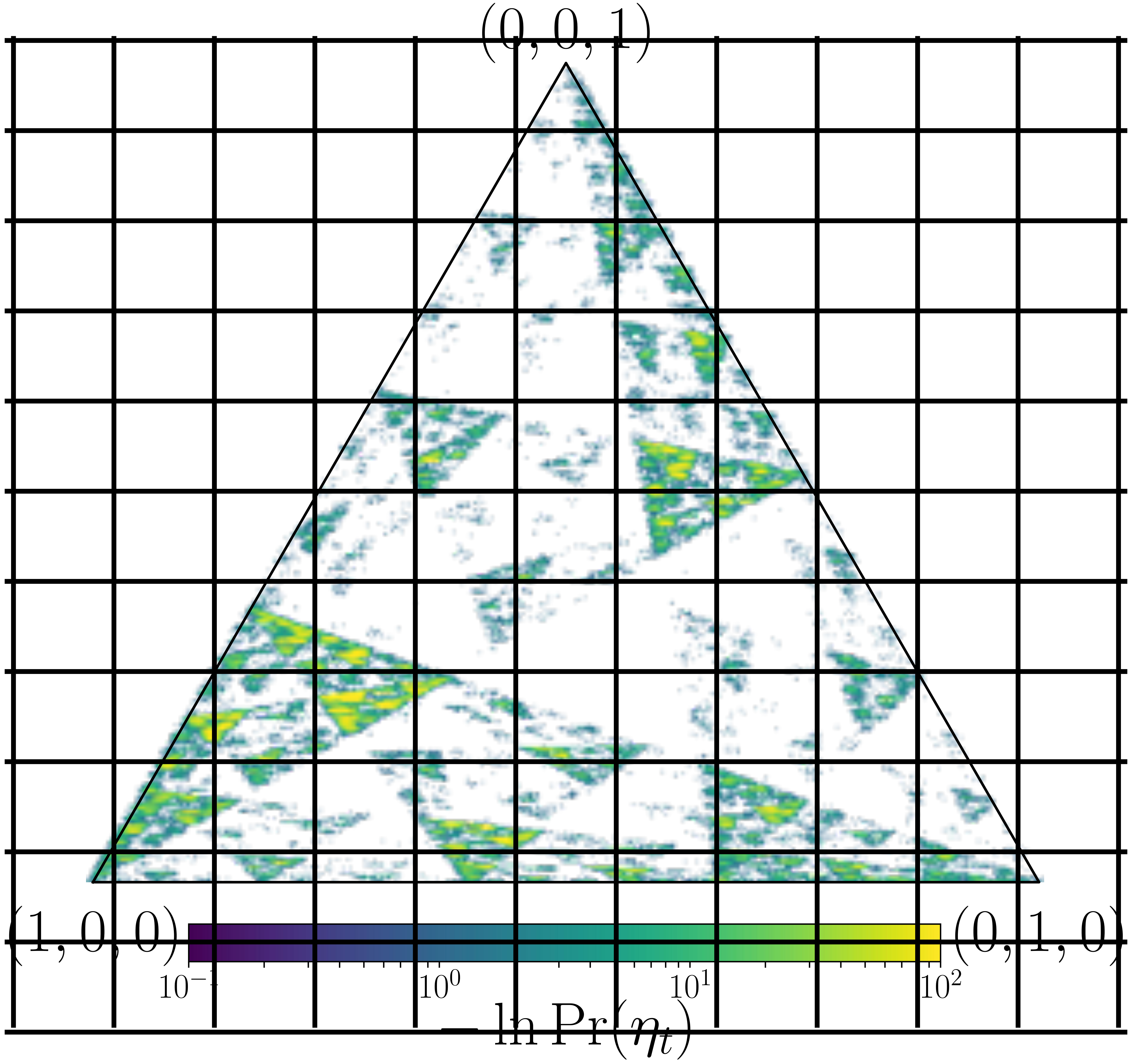


Box counting dimension:

$$d_0 = \lim_{\epsilon \to 0} - \frac{N_\epsilon}{\log(\epsilon)}$$

Where N_ϵ is the number of occupied boxes.

Information Dimension

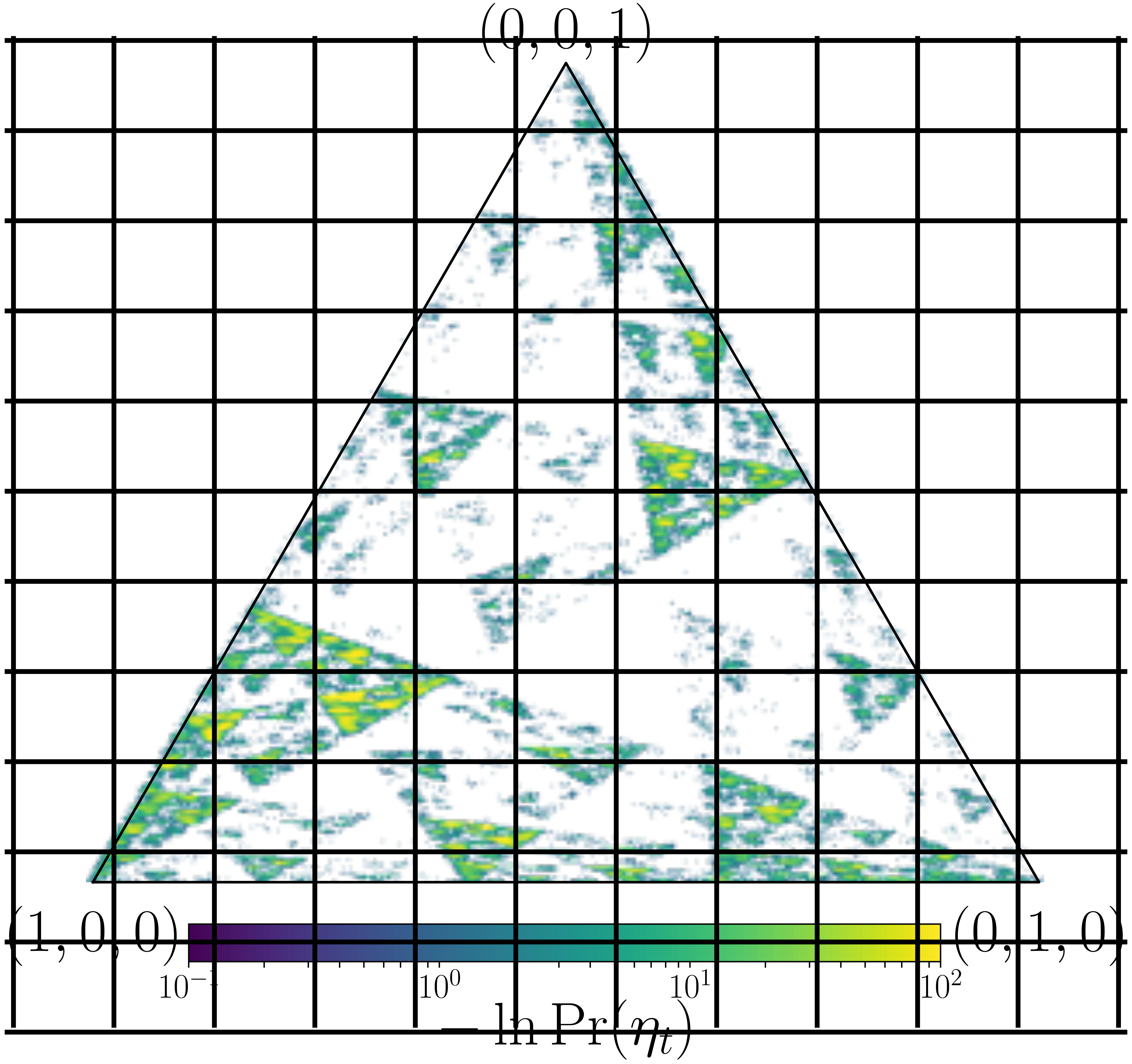


Information dimension:

$$d_1 = \lim_{\epsilon \to 0} -\frac{H\left[N_\epsilon\right]}{\log(\epsilon)}$$

Where H is taken according to the "natural measure"

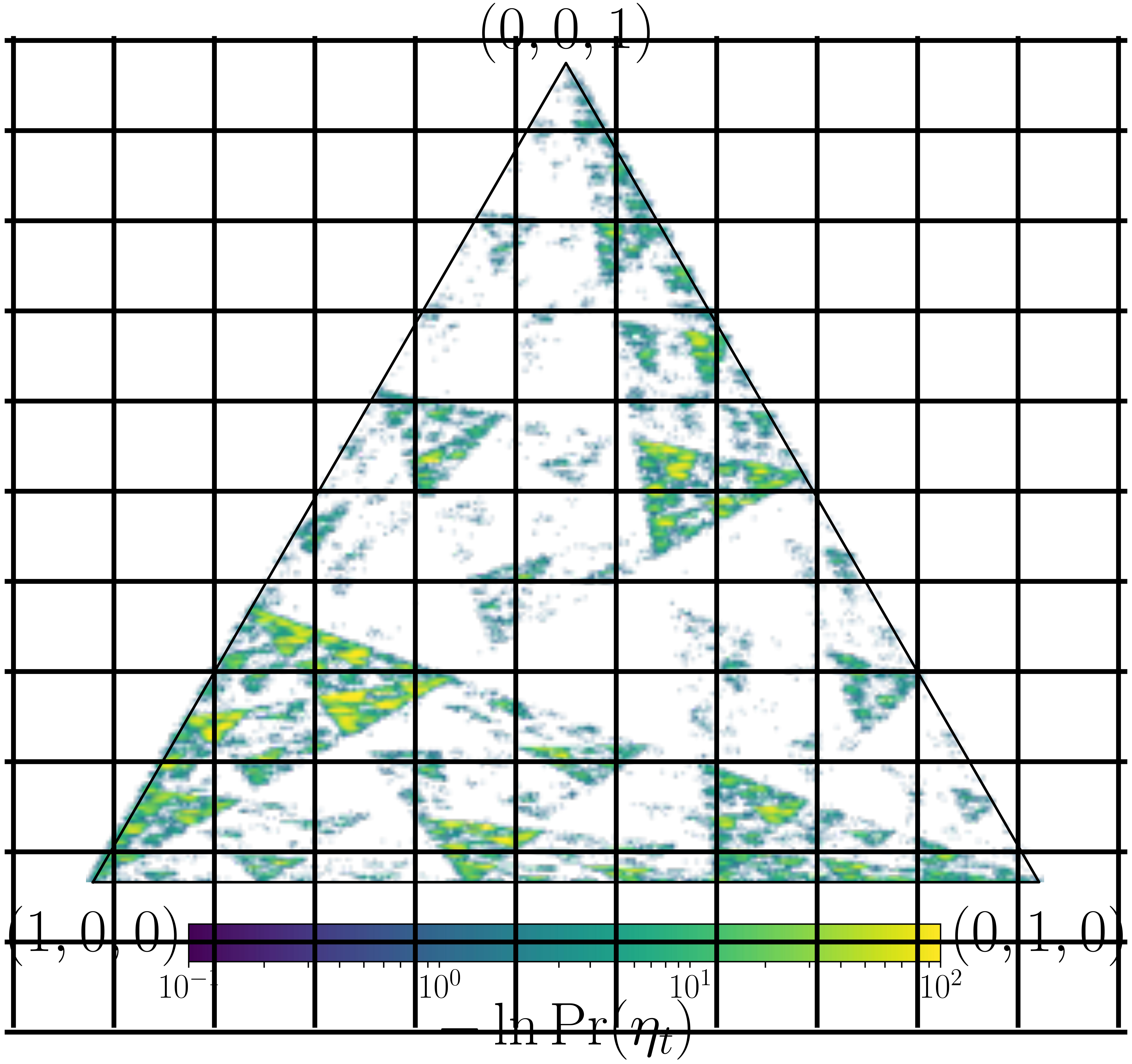
Information Dimension



Information dimension:

$$d_1 = \lim_{\epsilon \to 0} - \frac{H\left[N_\epsilon\right]}{\log(\epsilon)}$$

Information Dimension



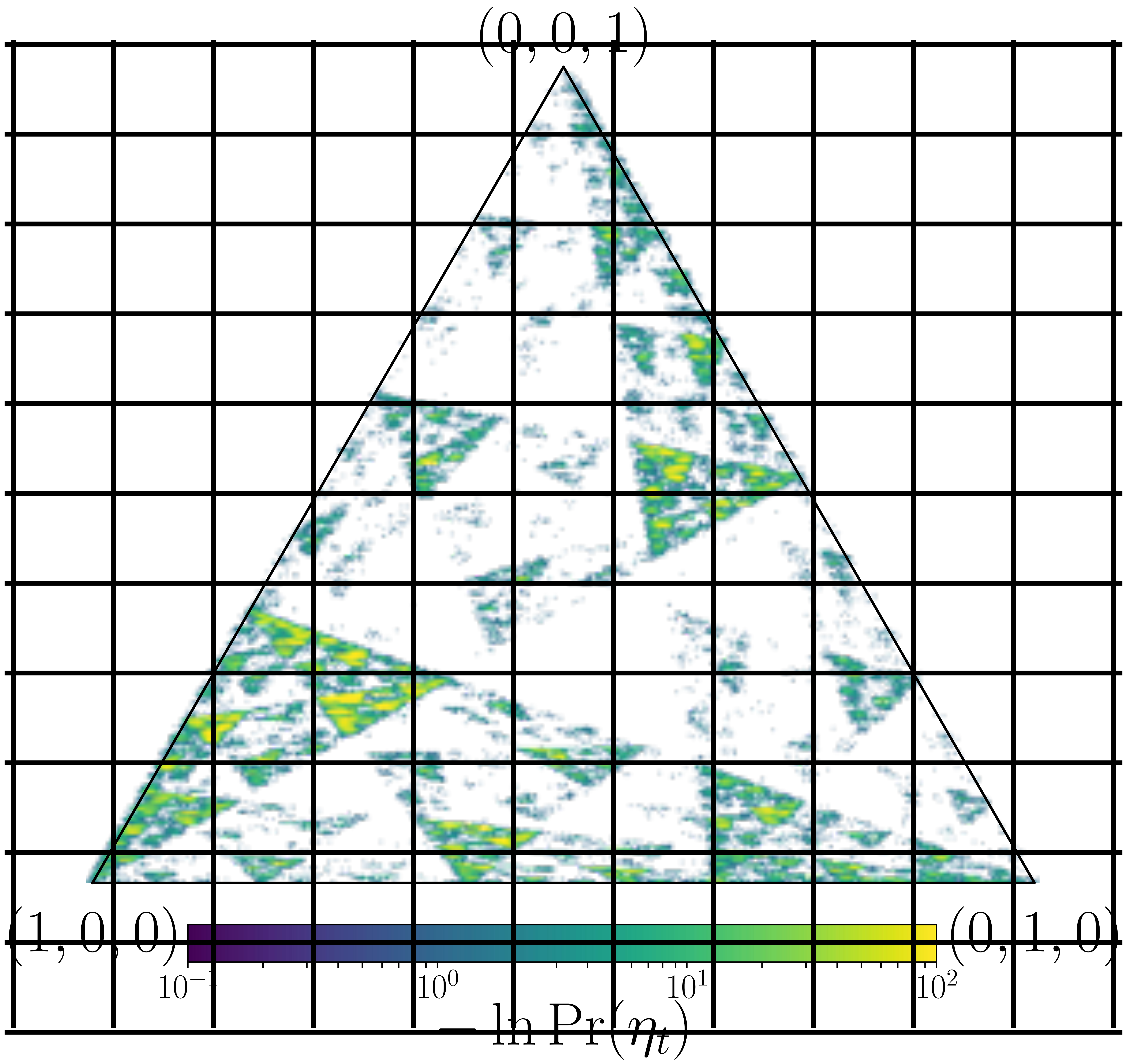
Information dimension:

$$d_1 = \lim_{\epsilon \to 0} - \frac{H\left[N_\epsilon\right]}{\log(\epsilon)}$$

Move things around:

$$H\left[N_\epsilon\right] \sim - d_1 \times \log(\epsilon)$$

Information Dimension



Information dimension:

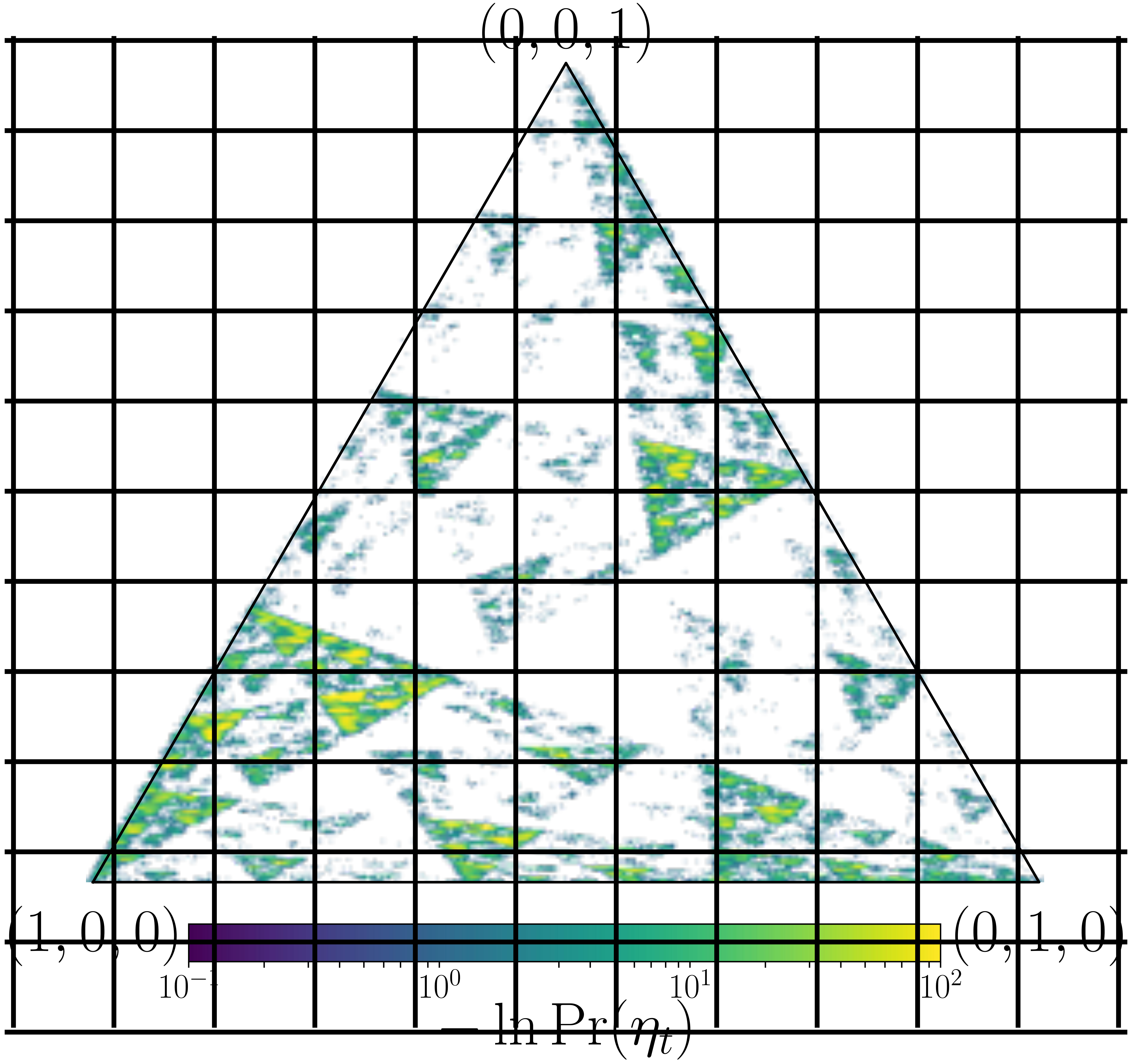
$$d_1 = \lim_{\epsilon \to 0} - \frac{H\left[N_\epsilon\right]}{\log(\epsilon)}$$

Move things around:

$$H\left[N_\epsilon\right] \sim - d_1 \times \log(\epsilon)$$

We can identify the left term as $C_{\mu,\epsilon}$ if the set is the causal states

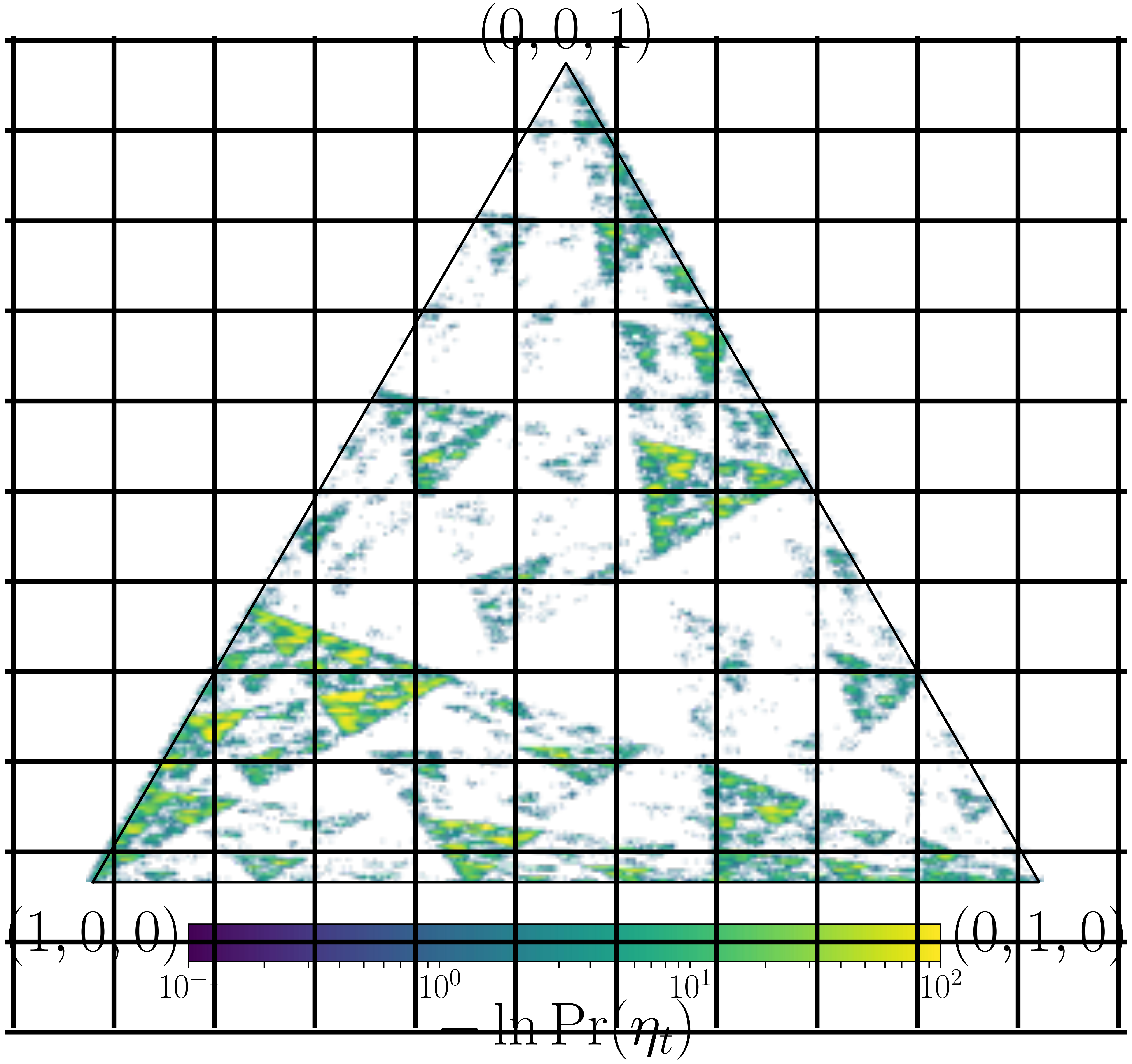
Statistical Complexity Dimension



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

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

Statistical Complexity Dimension

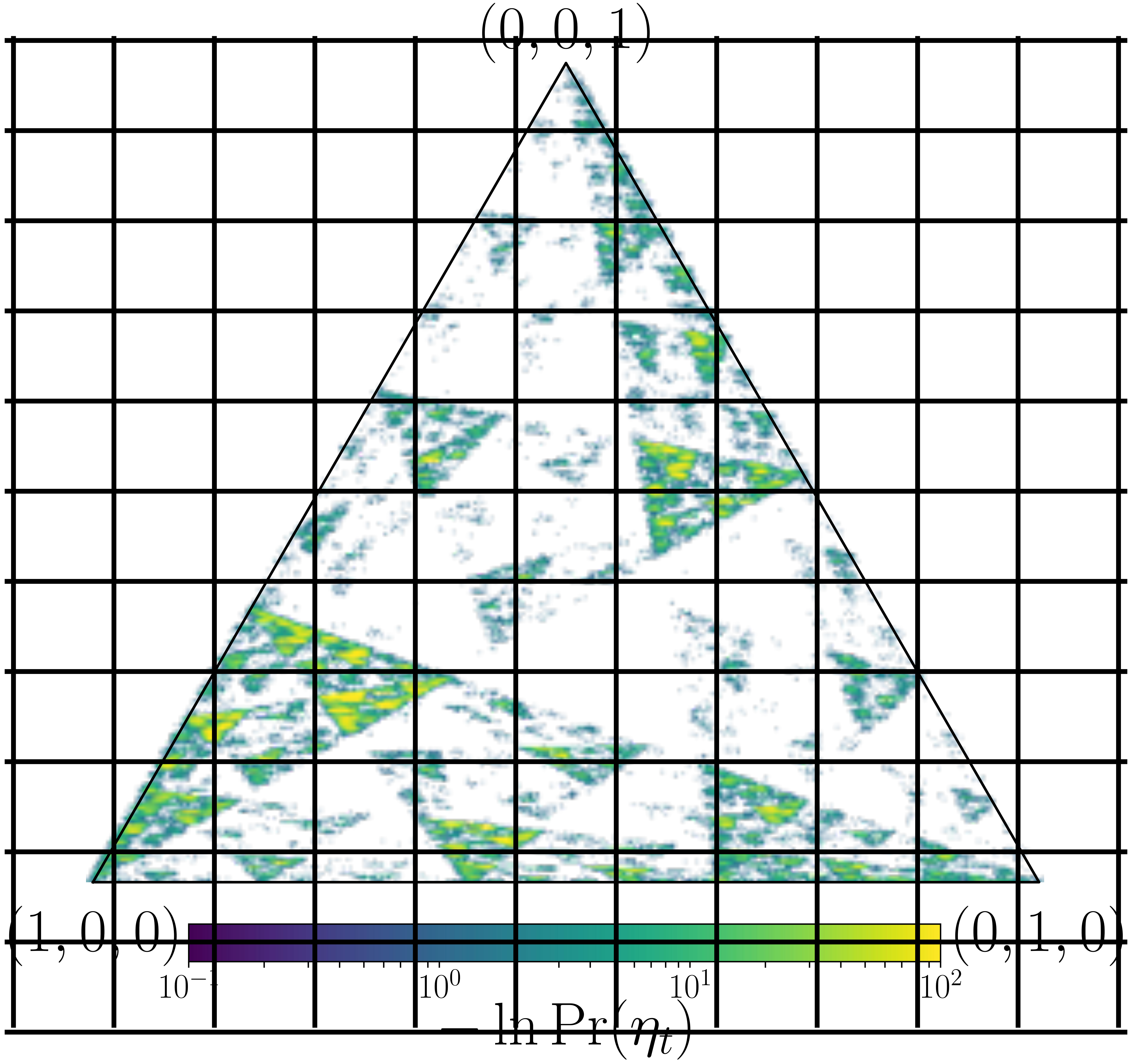


So, we define the *statistical complexity dimension* as the information dimension of the mixed state set:

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

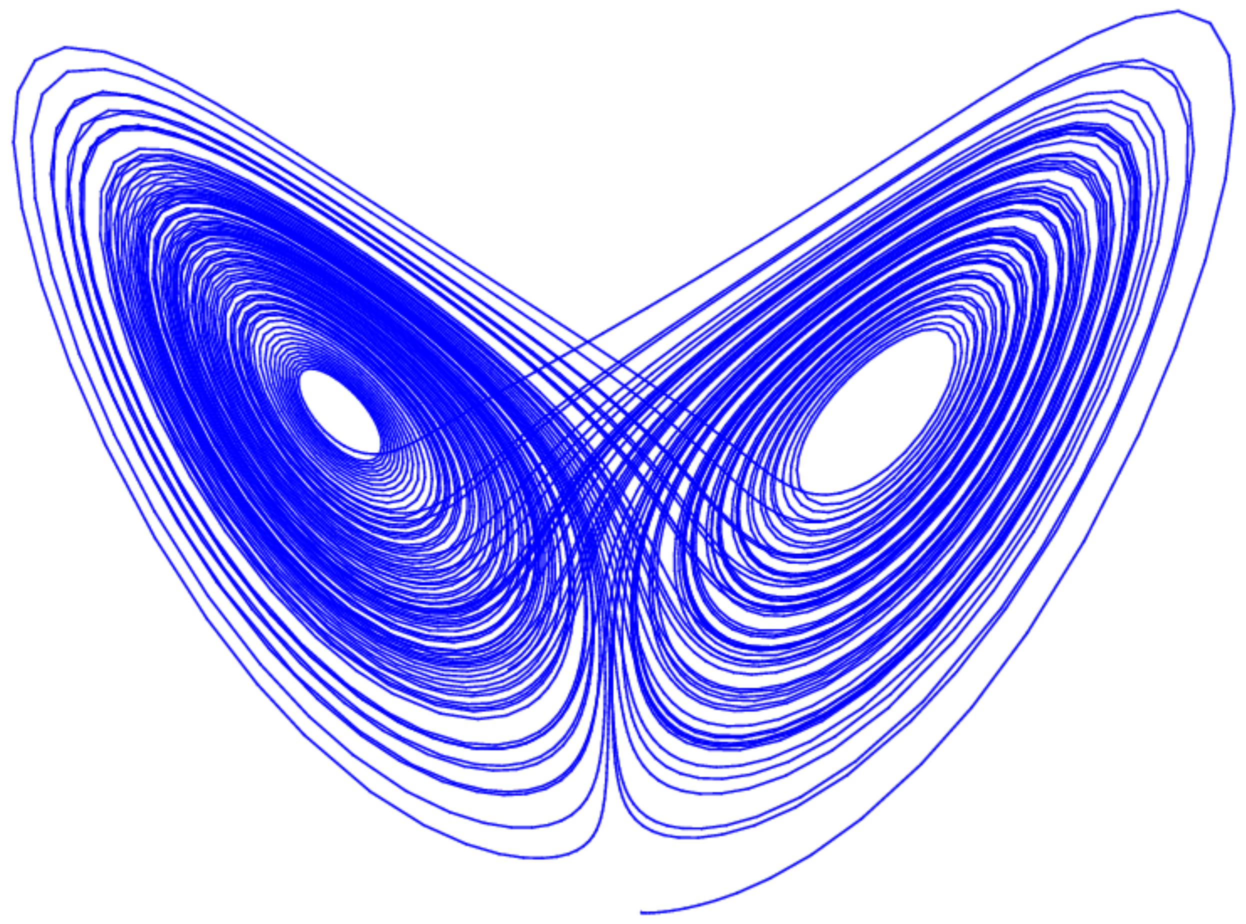
It gives the rate at which memory resources diverge as we increase predictive power of a finitized ϵ-machine.

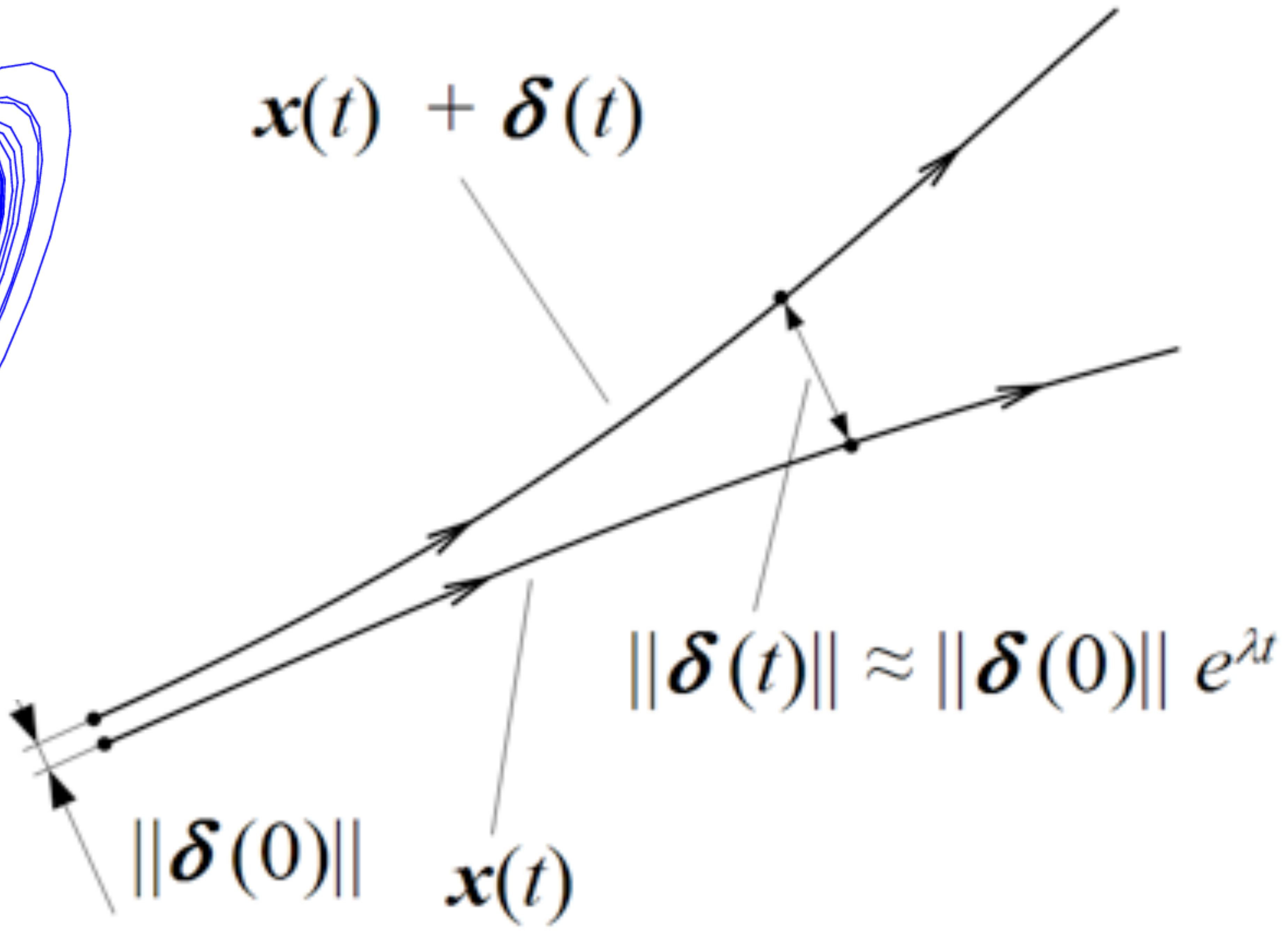
Calculating Statistical Complexity Dimension



We can approximate the measure using Ulam's method, but this method is fairly noisy.

Lyapunov Exponents



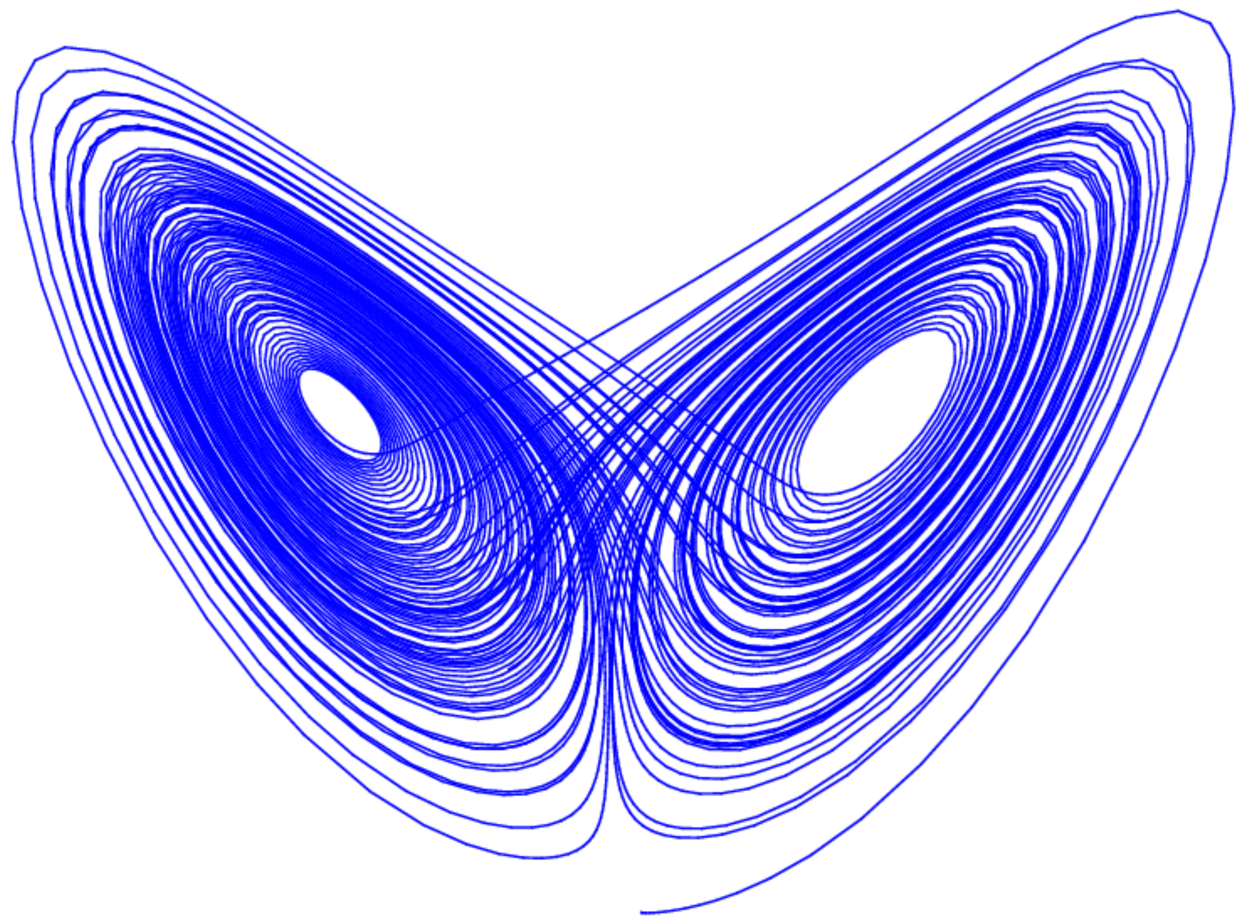


Calculate the Lyapunov spectrum: $\Gamma = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$

$$\text{s.t.} \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$$

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

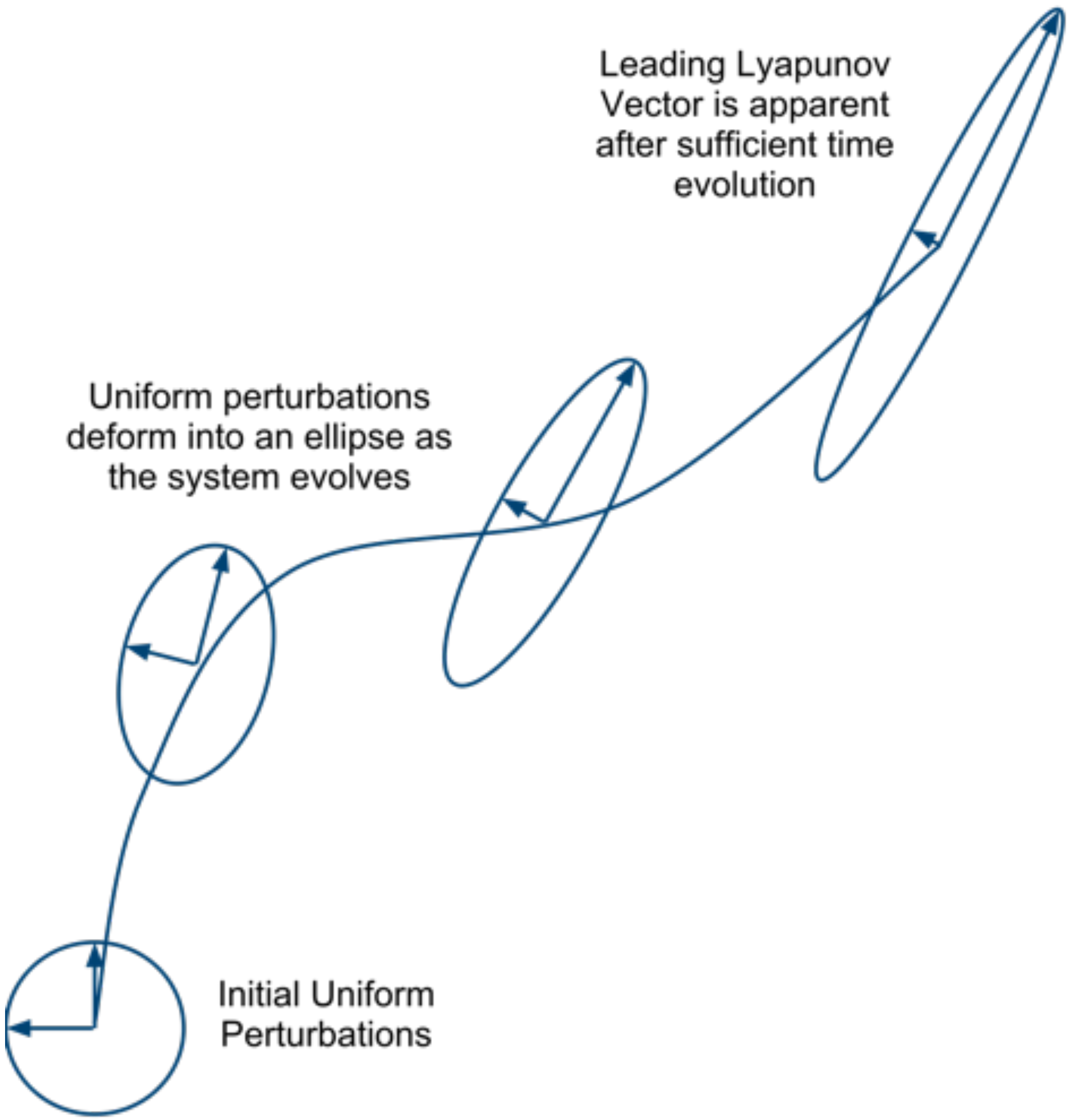
Lyapunov Exponents



Calculate the Lyapunov spectrum:

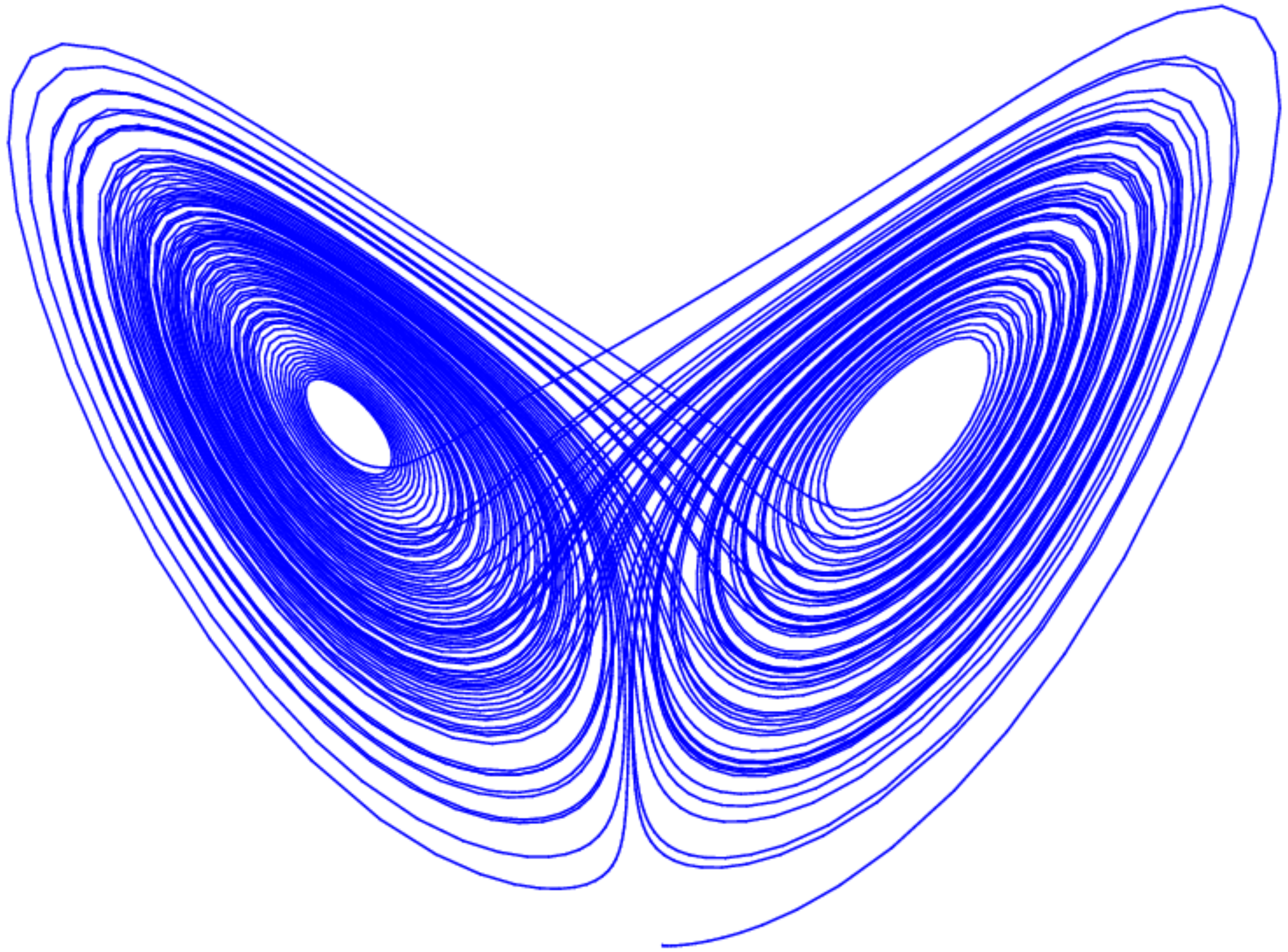
$$\Gamma = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$$

$$\text{s.t.} \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$$



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

Kaplan-Yorke Conjecture



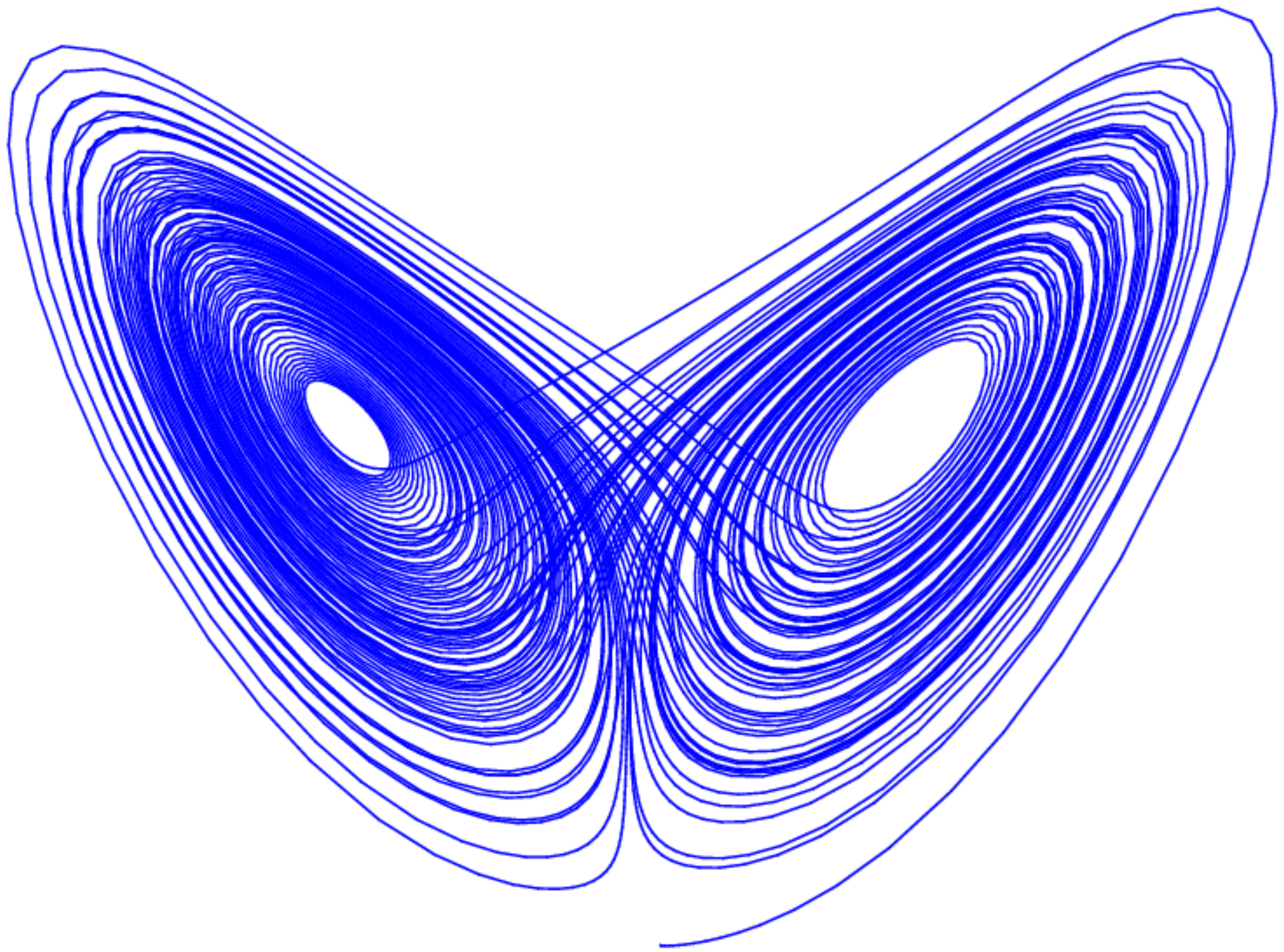
Kaplan—Yorke conjecture:

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

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.

Kaplan-Yorke Conjecture



For the Lorenz attractor with $\sigma = 10$, $r = 28$, $b = 8/3$:

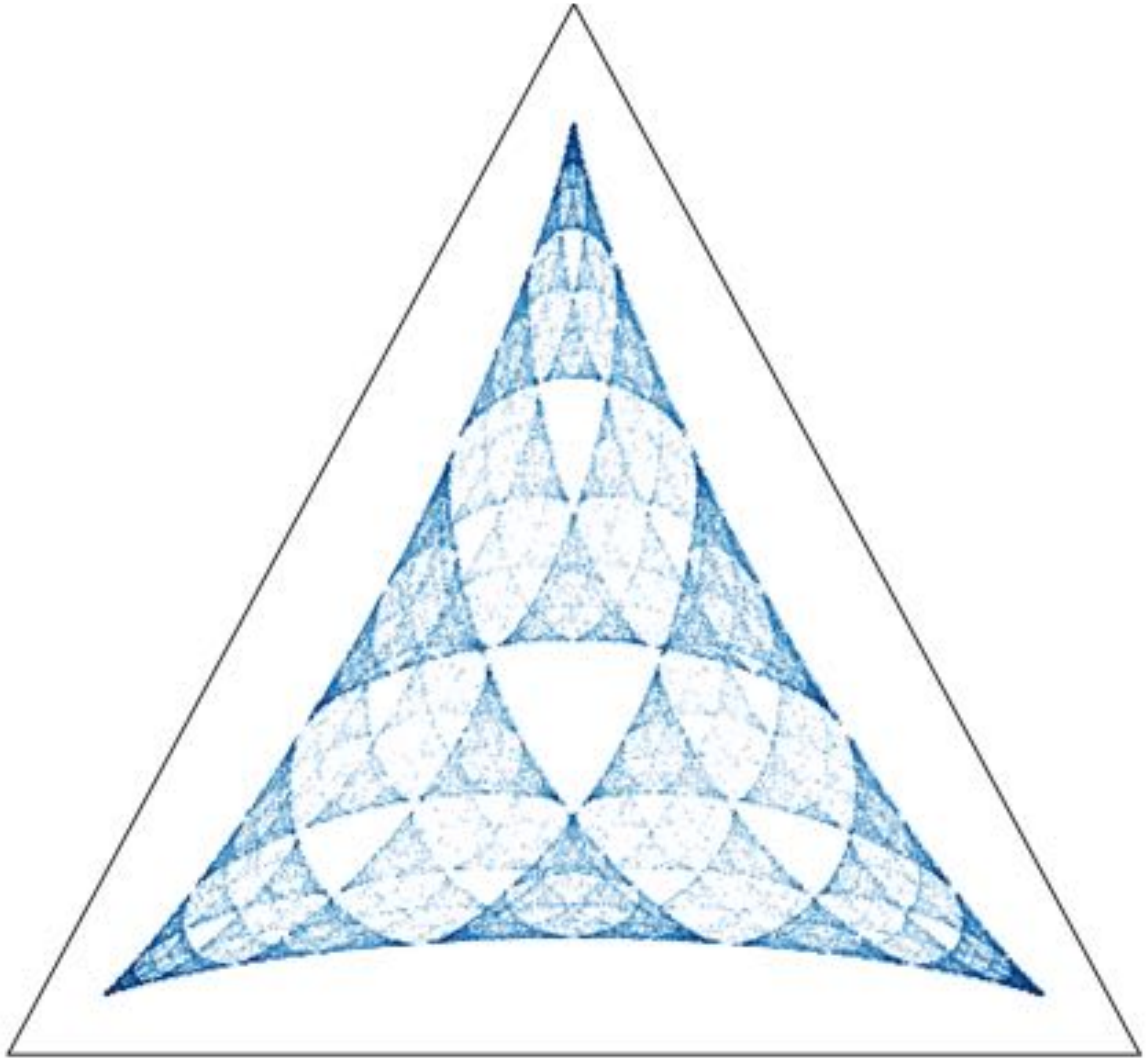
$$\Gamma = \{0.90563, 0, -14.57219\}$$

So the Lyapunov dimension is:

$$\dim_{KY} = 2.06215$$

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

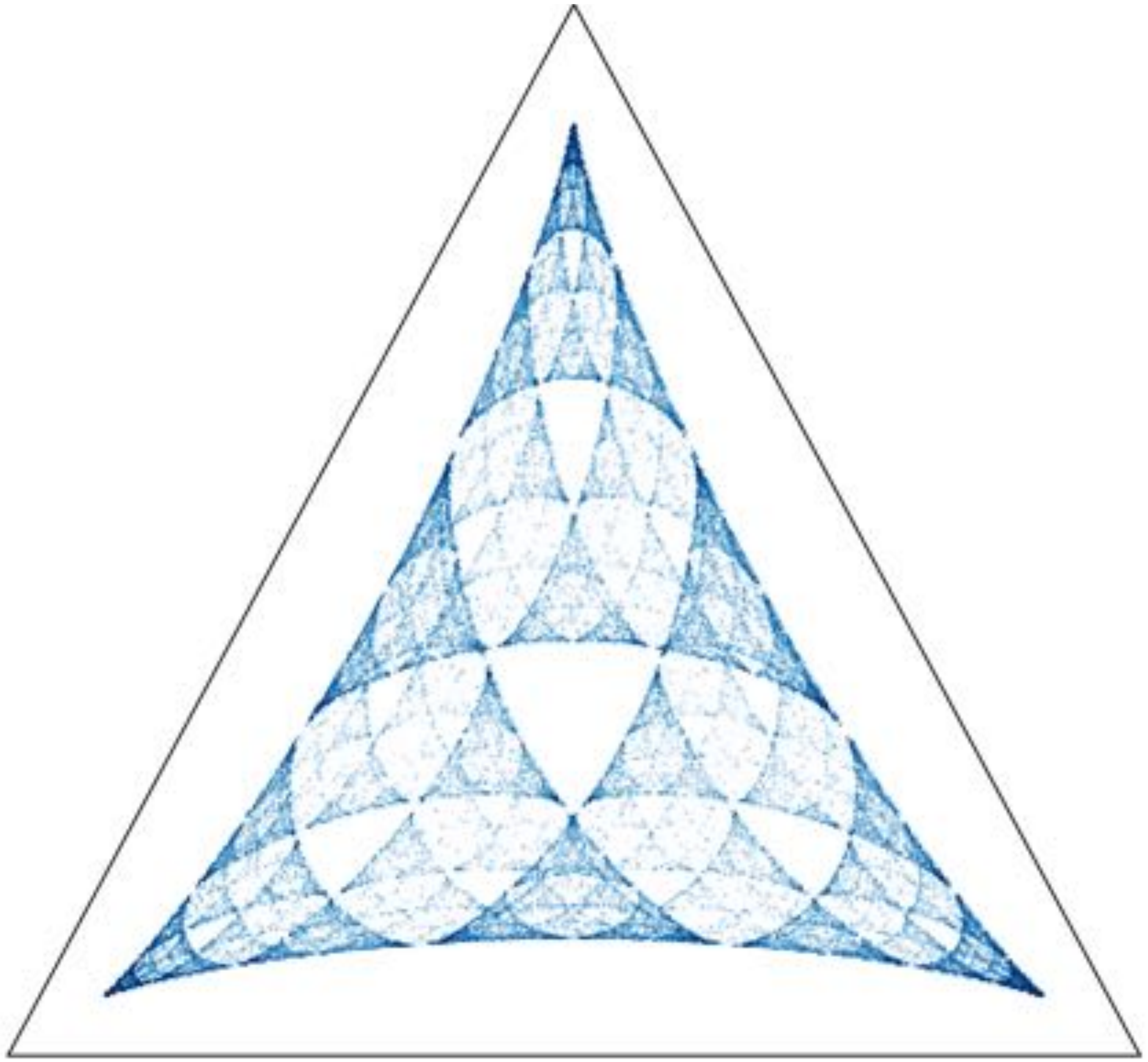
Kaplan-Yorke Conjecture for Statistical Complexity Dimension



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

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)

Kaplan-Yorke Conjecture for Statistical Complexity Dimension

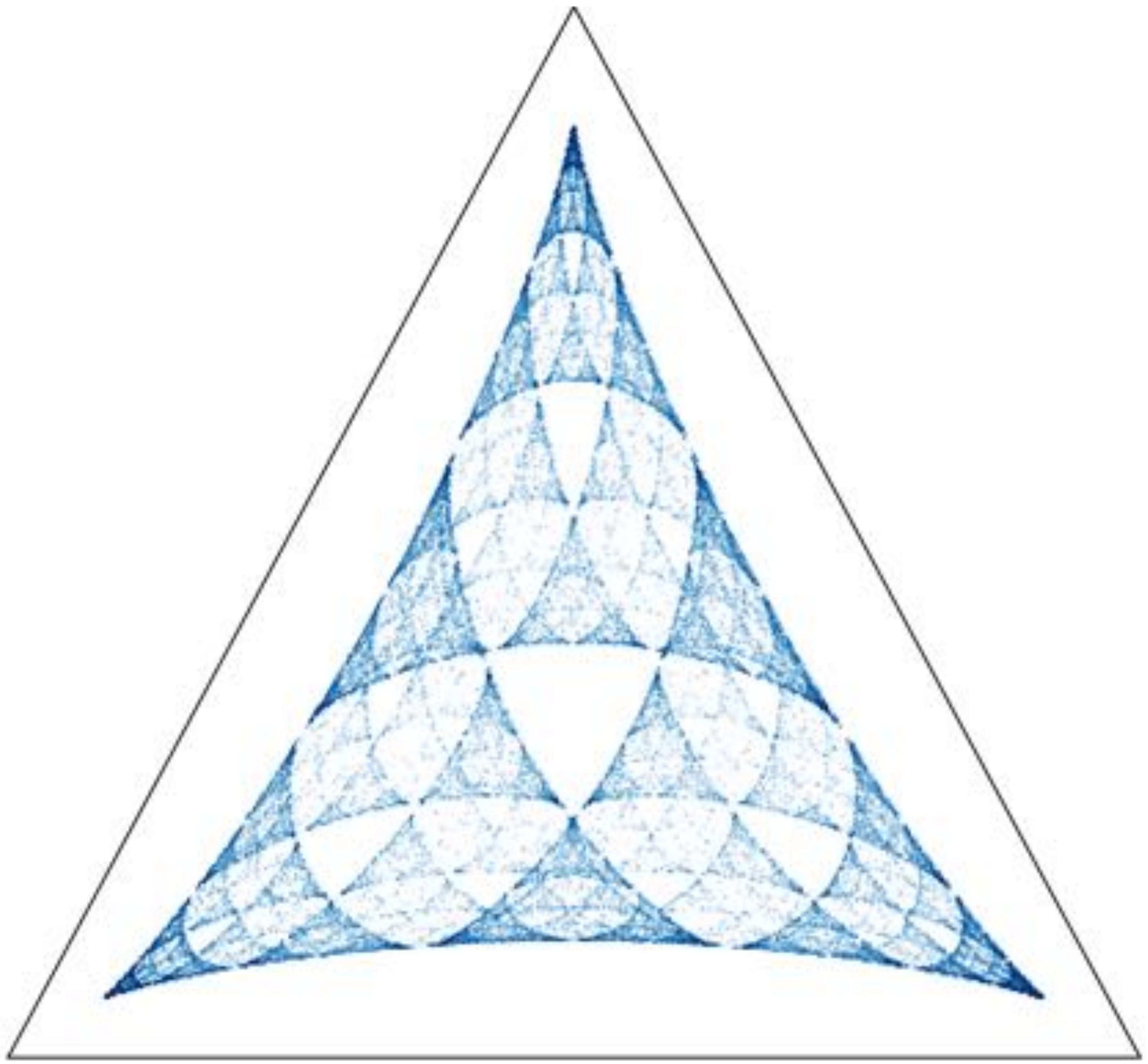


$$\dim_\mu(R) \stackrel{?}{=} k + \frac{\sum_i^k \lambda_i}{|\lambda_{k+1}|}$$

Problem: all maps are contractive, so all Lyapunov exponents are negative.

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)

Kaplan-Yorke Conjecture for Statistical Complexity Dimension



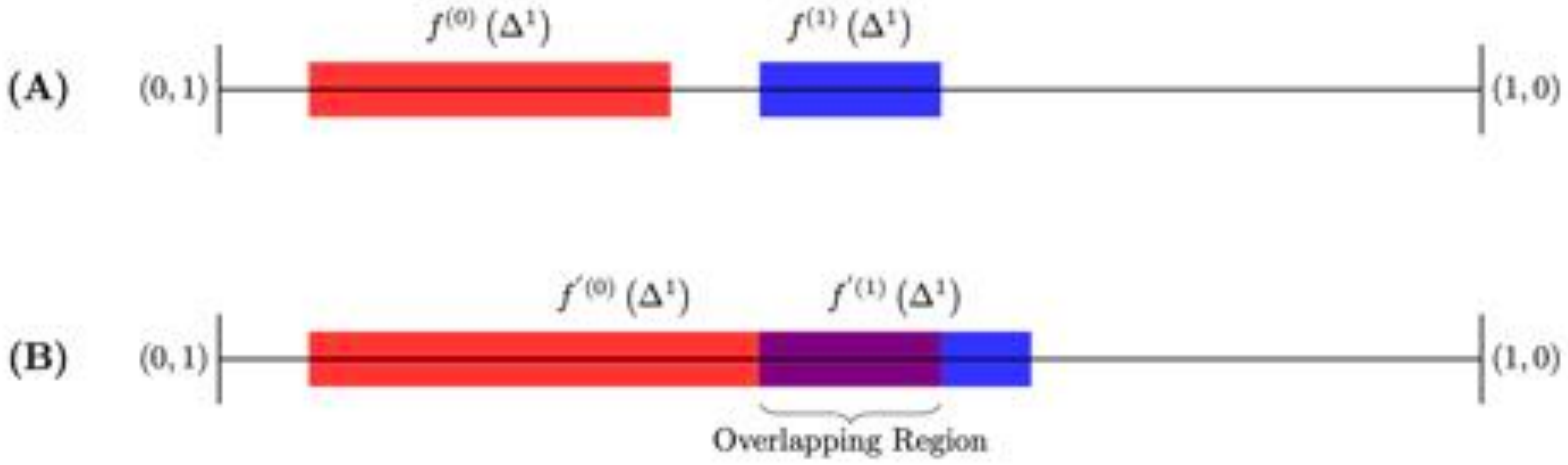
$$\dim_\mu(R) \stackrel{?}{=} k + \frac{h_\mu + \sum_i^k \lambda_i}{|\lambda_{k+1}|}$$

(Partial) solution: the entropy of the map choice plays the expansive role in an IFS, so we append it into the Lyapunov spectrum.

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)

The Overlapping Problem

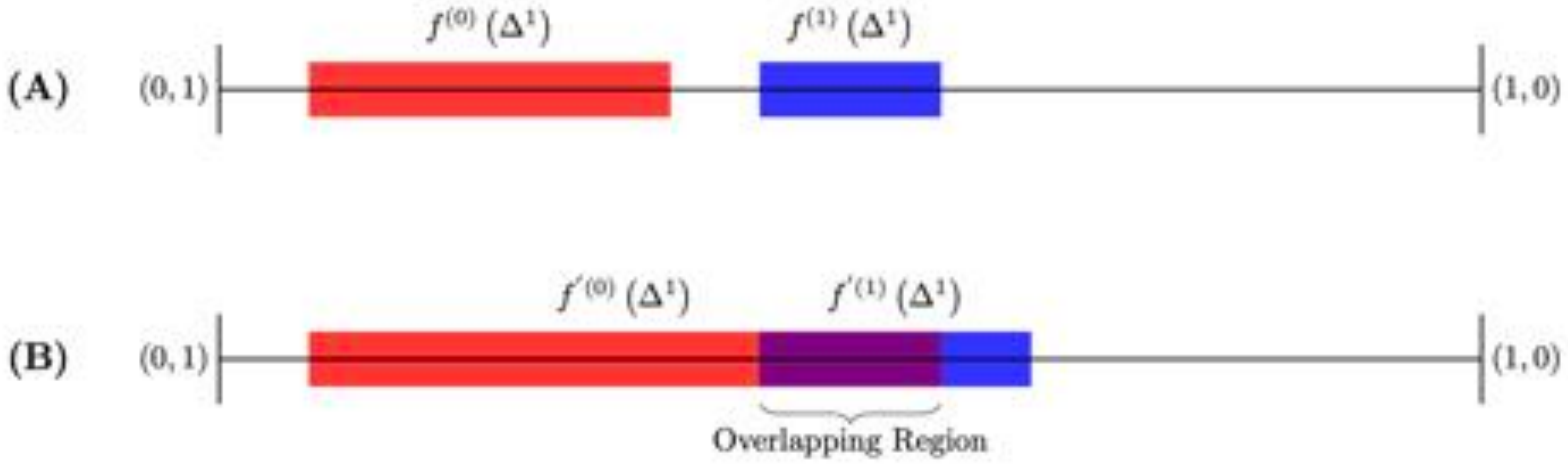


Problem: this solution does not work for “overlapping” IFSs.

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

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)

The Overlapping Problem



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

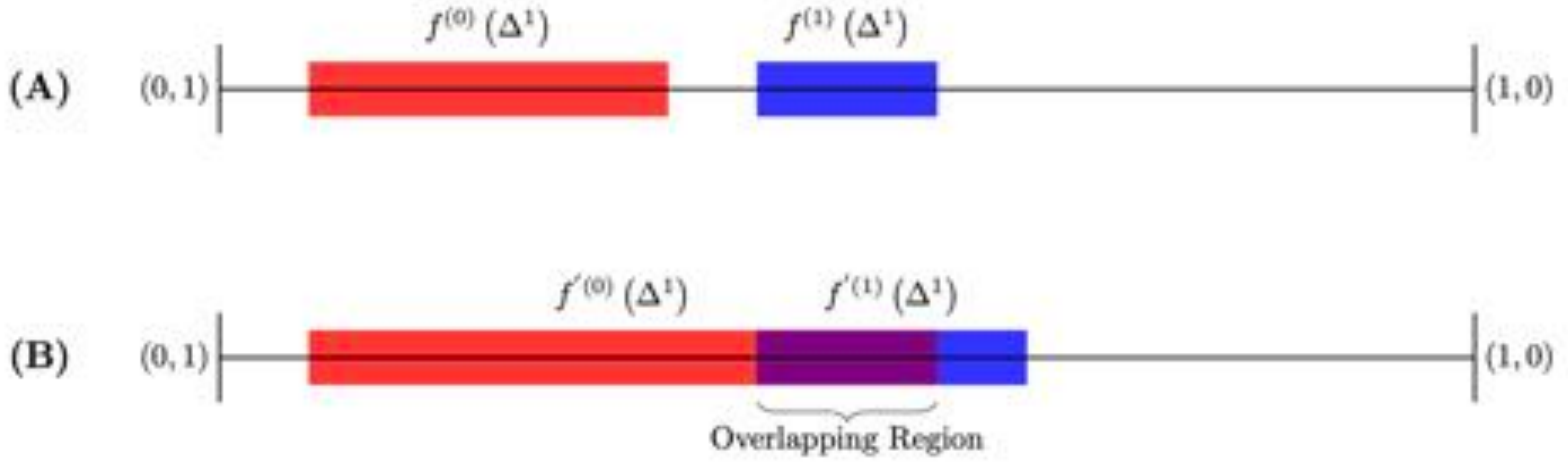
We end up "double counting" some of the states.

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

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)

Next time: Solving the Overlapping Problem



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

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)

Thank you and Questions