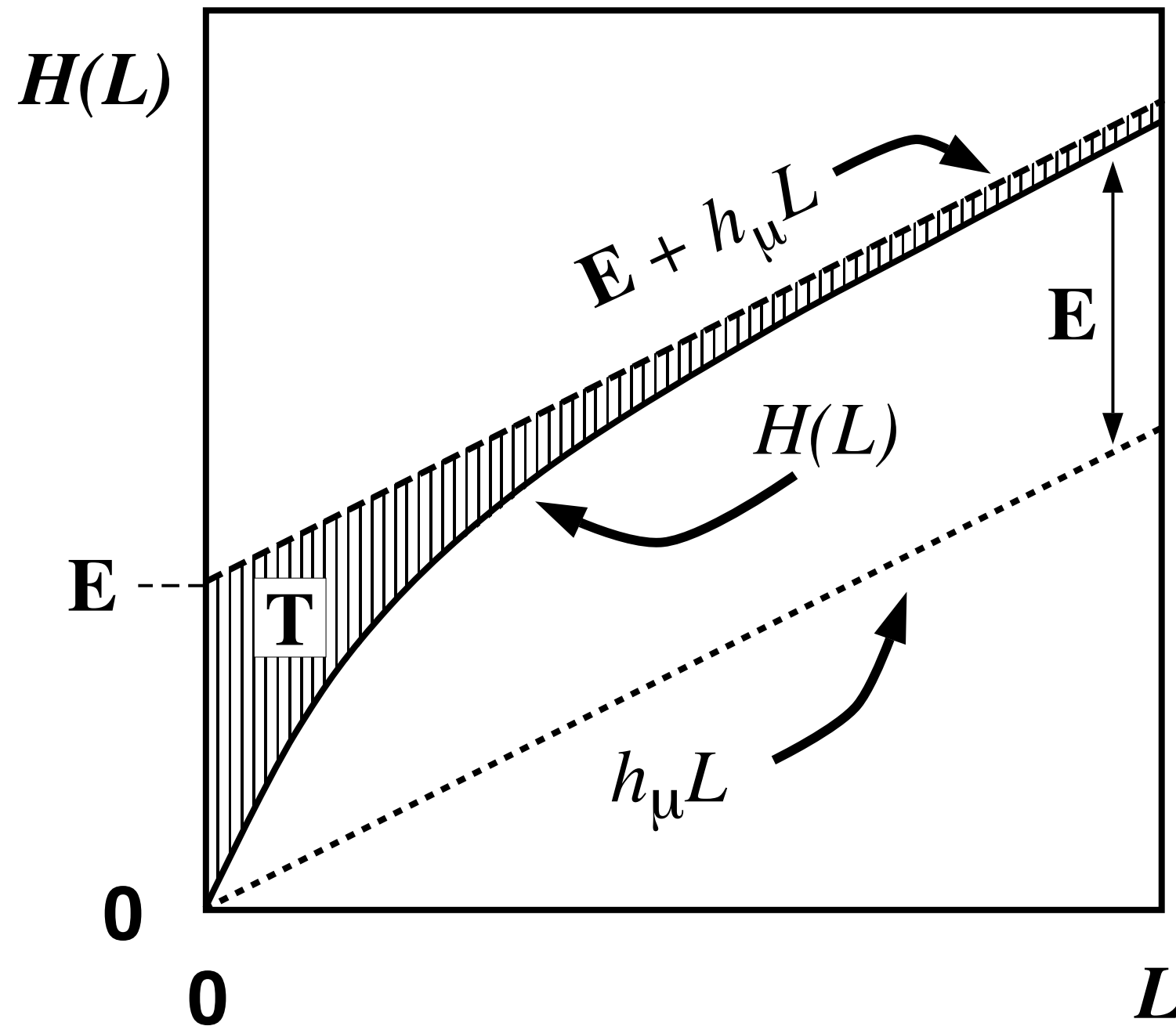


Intrinsic Computation

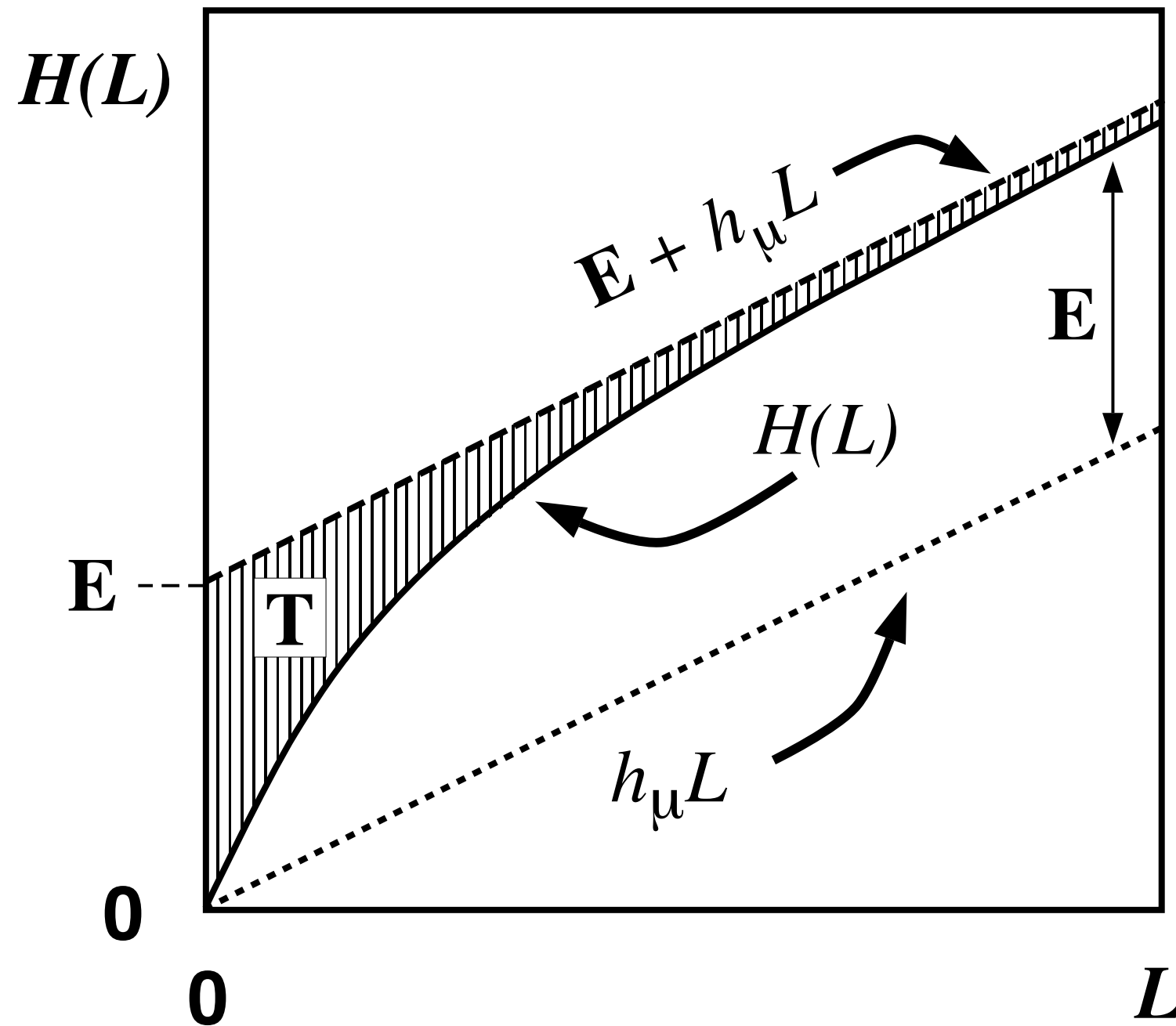
Jim Crutchfield
Complexity Sciences Center
Physics Department
University of California at Davis

Complex Systems Summer School
Santa Fe Institute
Santa Fe, NM
15 June 2016

Information Roadmap for a Complex Process

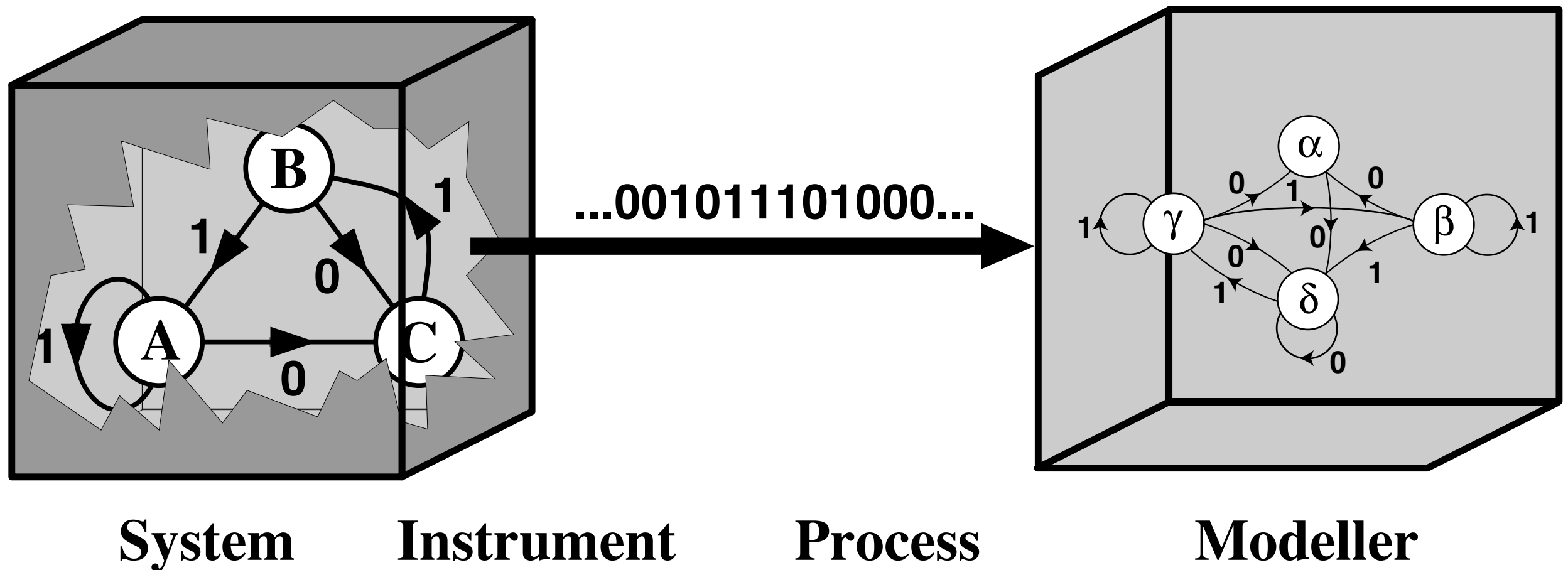


Information Roadmap for a Complex Process



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:

The Learning Channel ...

The Prediction Game ...

Process I:

Past: ... 111111111111

The Learning Channel ...

The Prediction Game ...

Process I:

Past: ... 111111111111

Your prediction is?

The Learning Channel ...

The Prediction Game ...

Process I:

Past: ... 111111111111

Your prediction is?

Future: 111111111111 ...

The Learning Channel ...

The Prediction Game ...

Process I:

Past: ... 111111111111

Your prediction is?

Future: 111111111111 ...

Your model (states & dynamic) is?

The Learning Channel ...

The Prediction Game ...

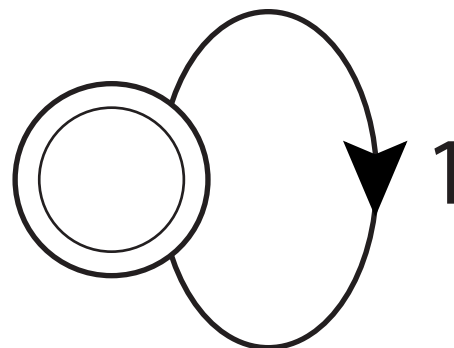
Process I:

Past: ... 111111111111

Your prediction is?

Future: 111111111111...

Your model (states & dynamic) is?



The Learning Channel ...
The Prediction Game ...

Process II:

The Learning Channel ...

The Prediction Game ...

Process II:

Past: ... 10110010001101110

The Learning Channel ...

The Prediction Game ...

Process II:

Past: ... 10110010001101110

Your prediction is?

The Learning Channel ...

The Prediction Game ...

Process II:

Past: ... 10110010001101110

Your prediction is?

Analysis: All words of length L occur & equally often

The Learning Channel ...

The Prediction Game ...

Process II:

Past: ... 10110010001101110

Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

The Learning Channel ...

The Prediction Game ...

Process II:

Past: ... 10110010001101110

Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

01010111010001101 ...

The Learning Channel ...

The Prediction Game ...

Process II:

Past: ... 10110010001101110

Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

01010111010001101 ...

Your model is?

The Learning Channel ...

The Prediction Game ...

Process II:

Past: ... 10110010001101110

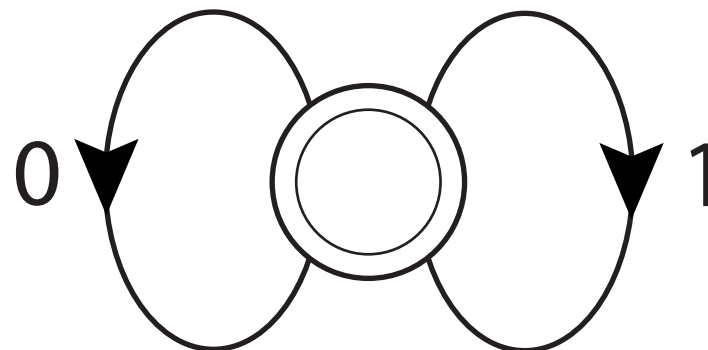
Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

01010111010001101 ...

Your model is?



The Learning Channel ...

The Prediction Game ...

Process III:

The Learning Channel ...

The Prediction Game ...

Process III:

Past: ... 1010101010101010

The Learning Channel ...

The Prediction Game ...

Process III:

Past: ... 1010101010101010

Your prediction is?

The Learning Channel ...

The Prediction Game ...

Process III:

Past: ... 1010101010101010

Your prediction is?

Future: 1010101010101 ...

The Learning Channel ...

The Prediction Game ...

Process III:

Past: ... 1010101010101010

Your prediction is?

Future: 1010101010101 ...

Your model is?

The Learning Channel ...

The Prediction Game ...

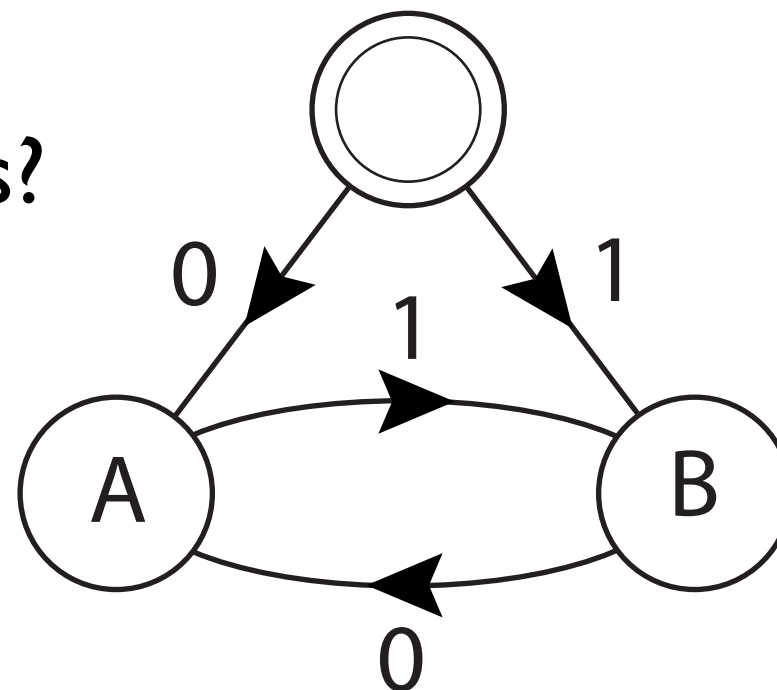
Process III:

Past: ... 10101010101010

Your prediction is?

Future: 1010101010101...

Your model is?



The Learning Channel ...

Theory? Algorithms?

Computational Mechanics

The Learning Channel ...

Goal:

Predict the future $\vec{S}$
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 $\mathcal{A}$

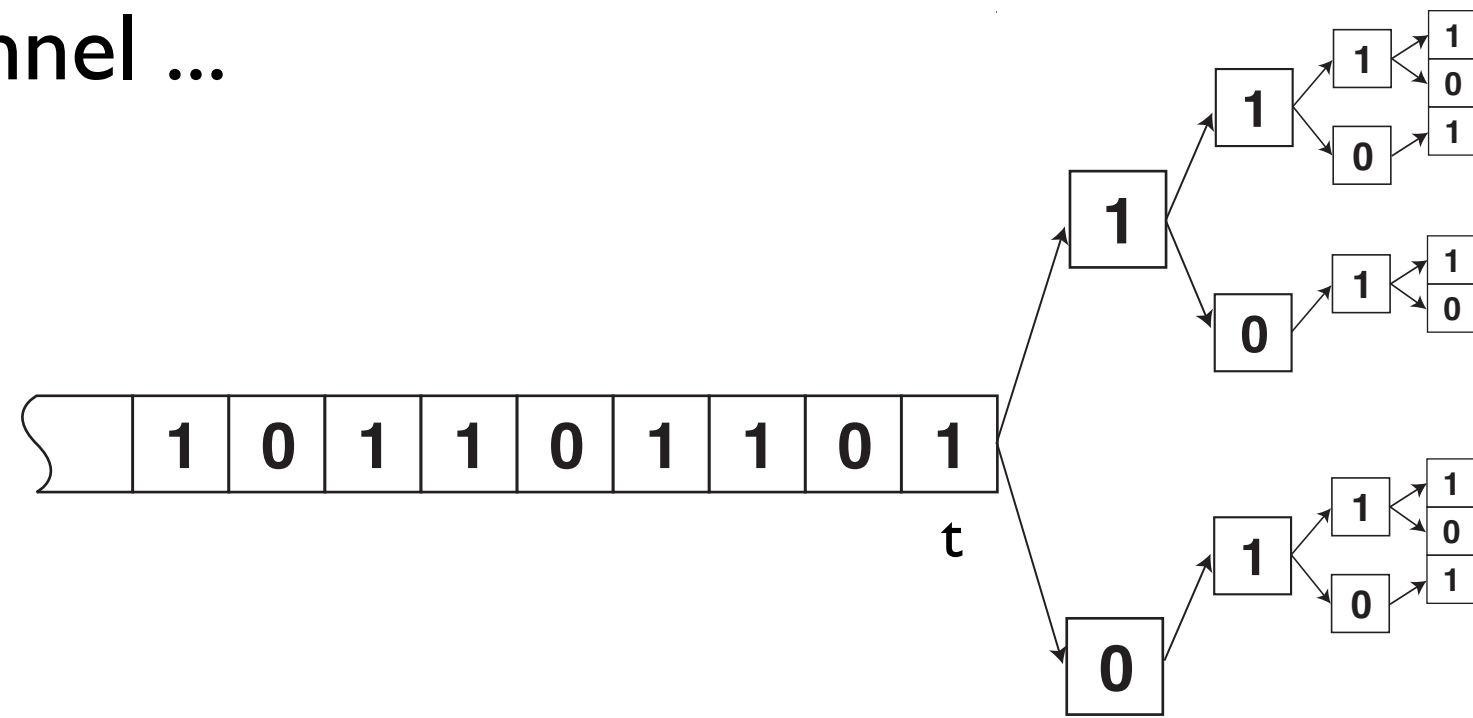
These symbols only indirectly reflect the hidden states

The Learning Channel ...

Effective States:

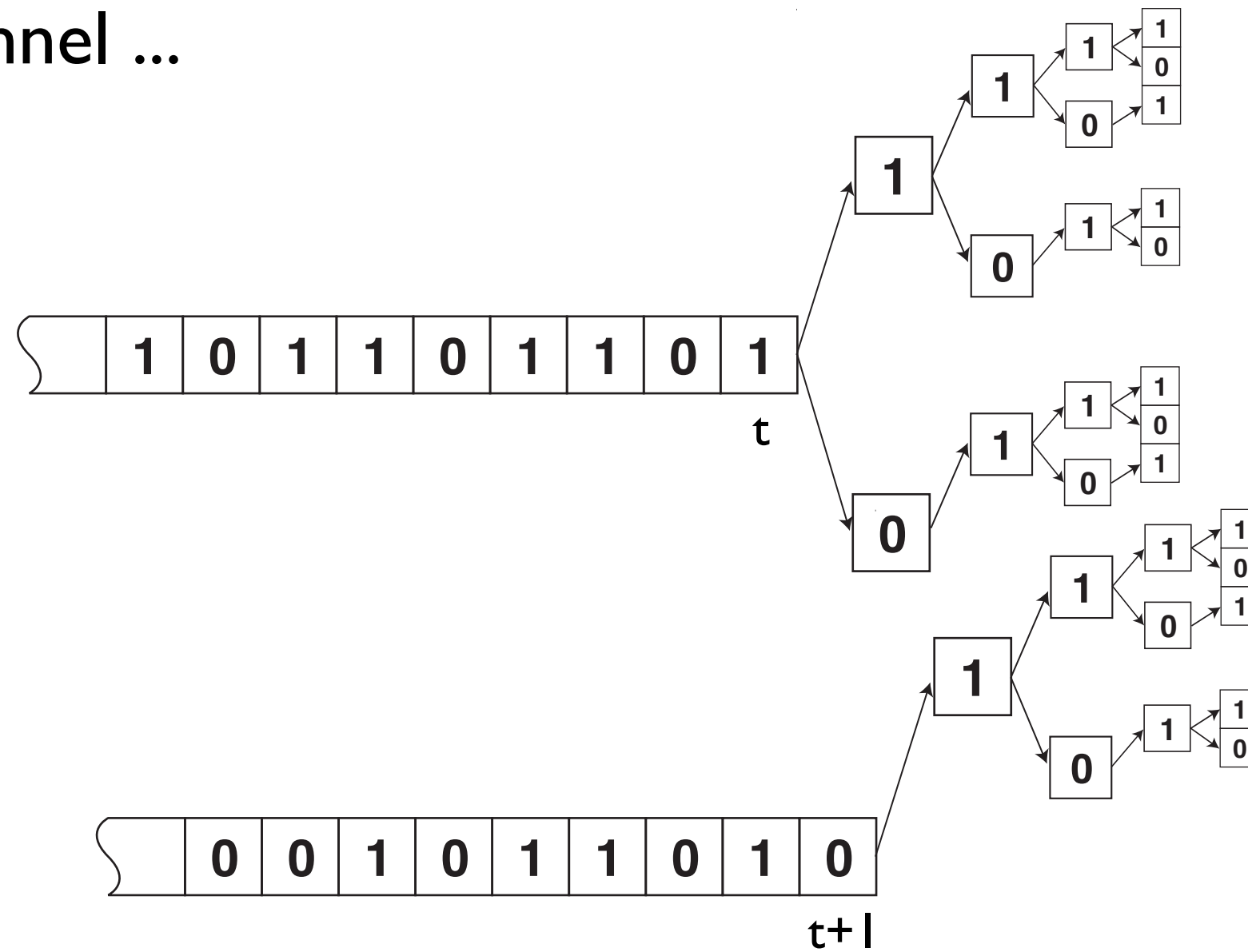
The Learning Channel ...

Effective States:



The Learning Channel ...

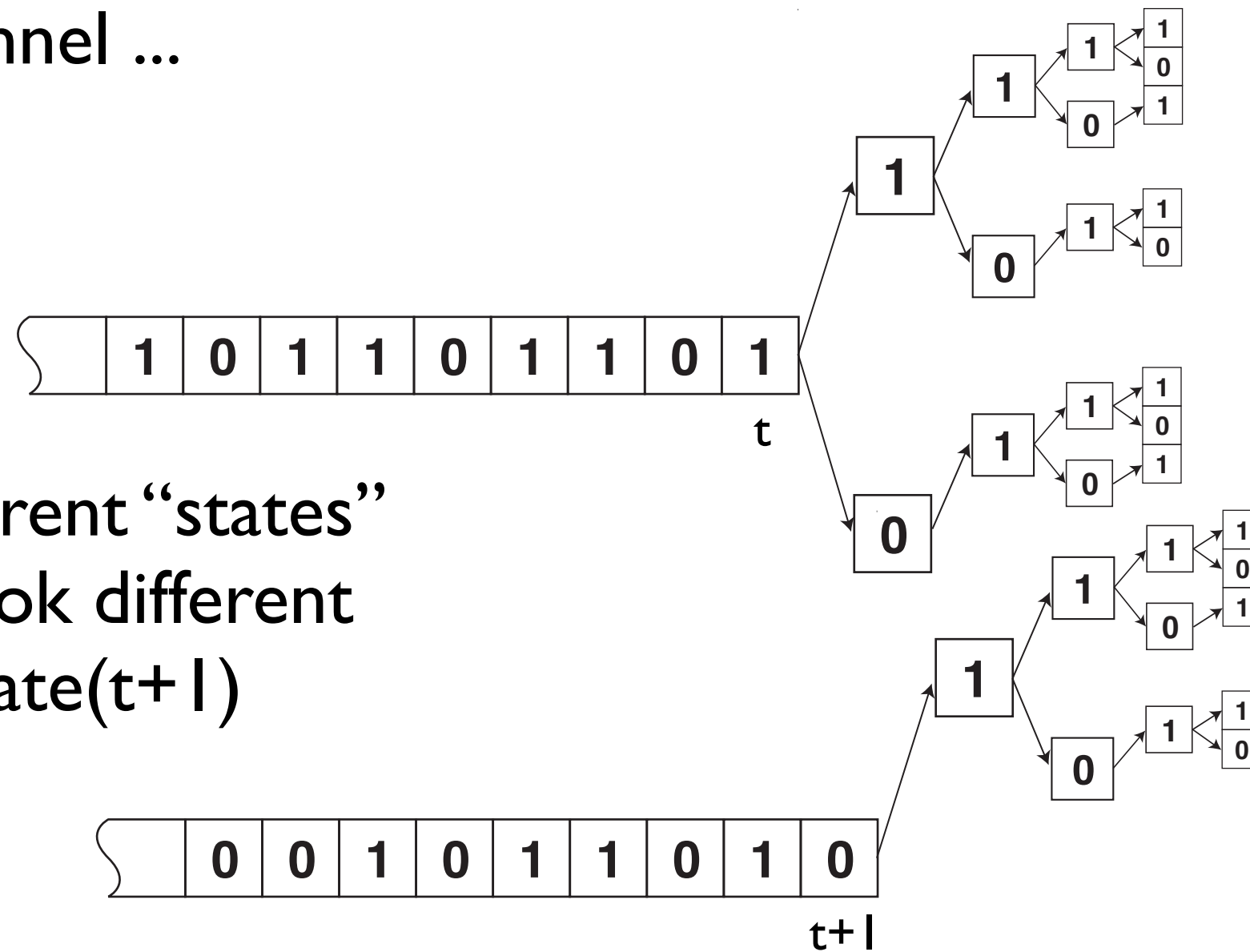
Effective States:



The Learning Channel ...

Effective States:

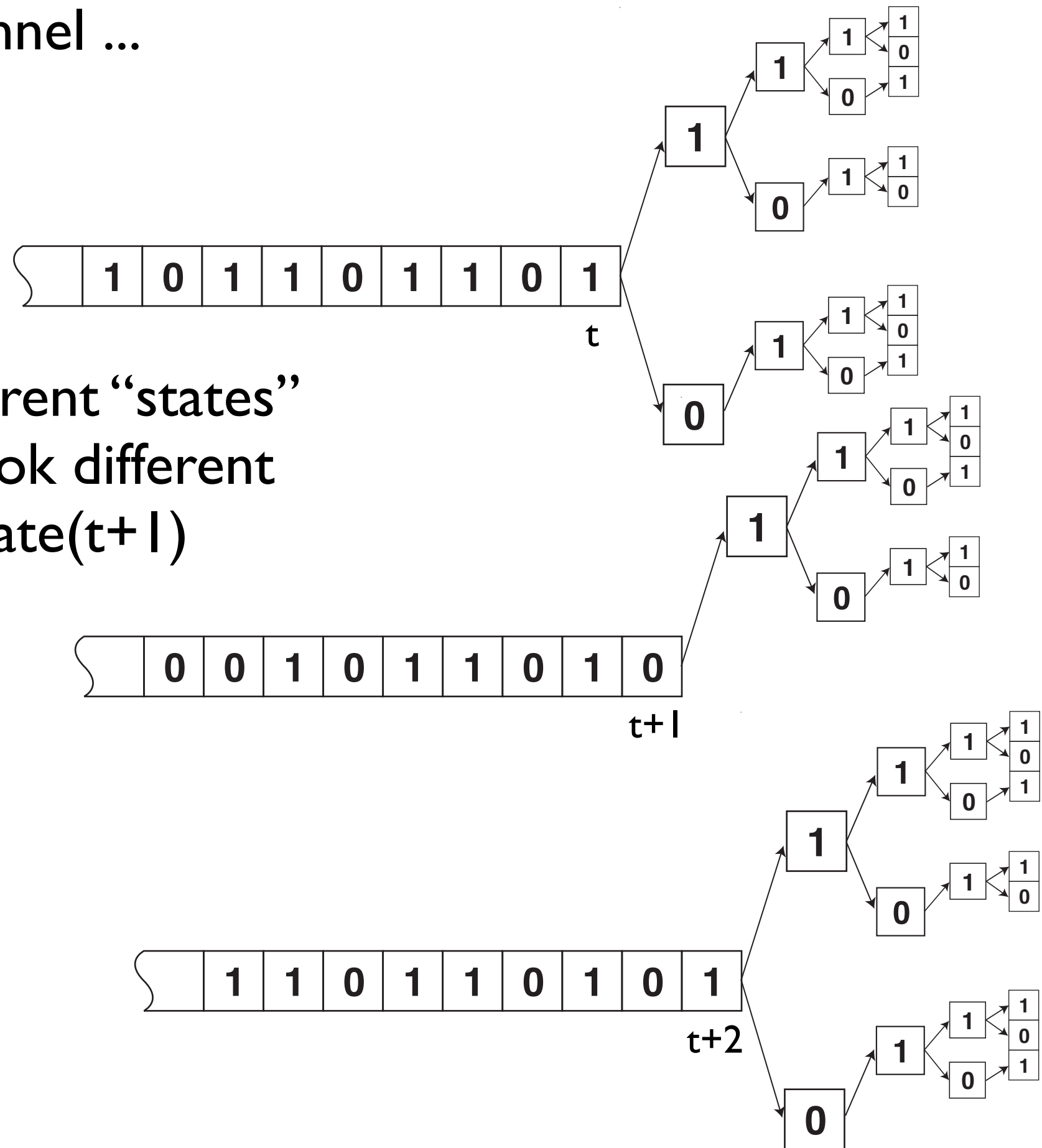
Process is in different “states”
when futures look different
 $\text{State}(t) \approx \text{State}(t+1)$



The Learning Channel ...

Effective States:

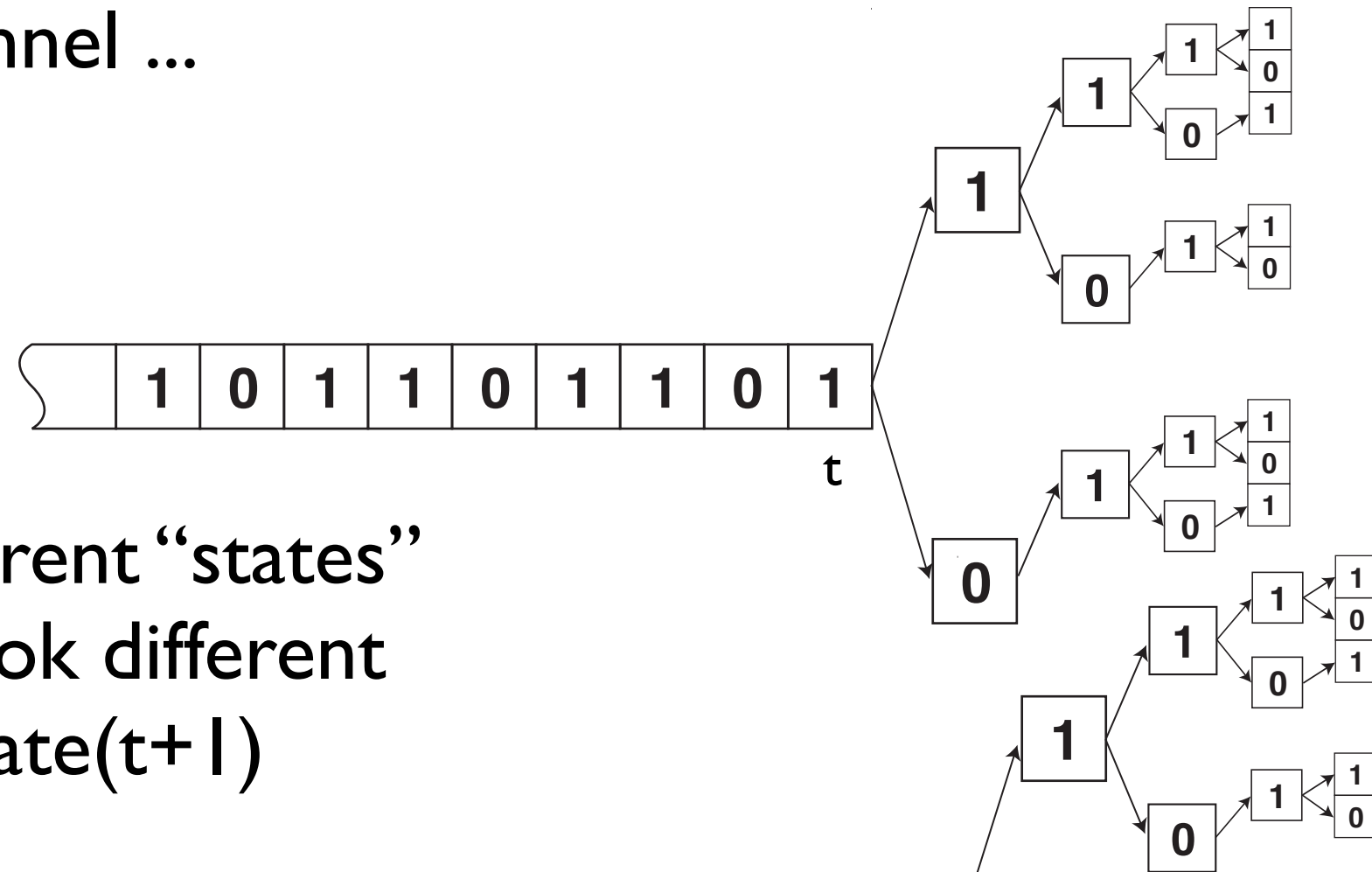
Process is in different “states”
when futures look different
 $\text{State}(t) \approx \text{State}(t+1)$



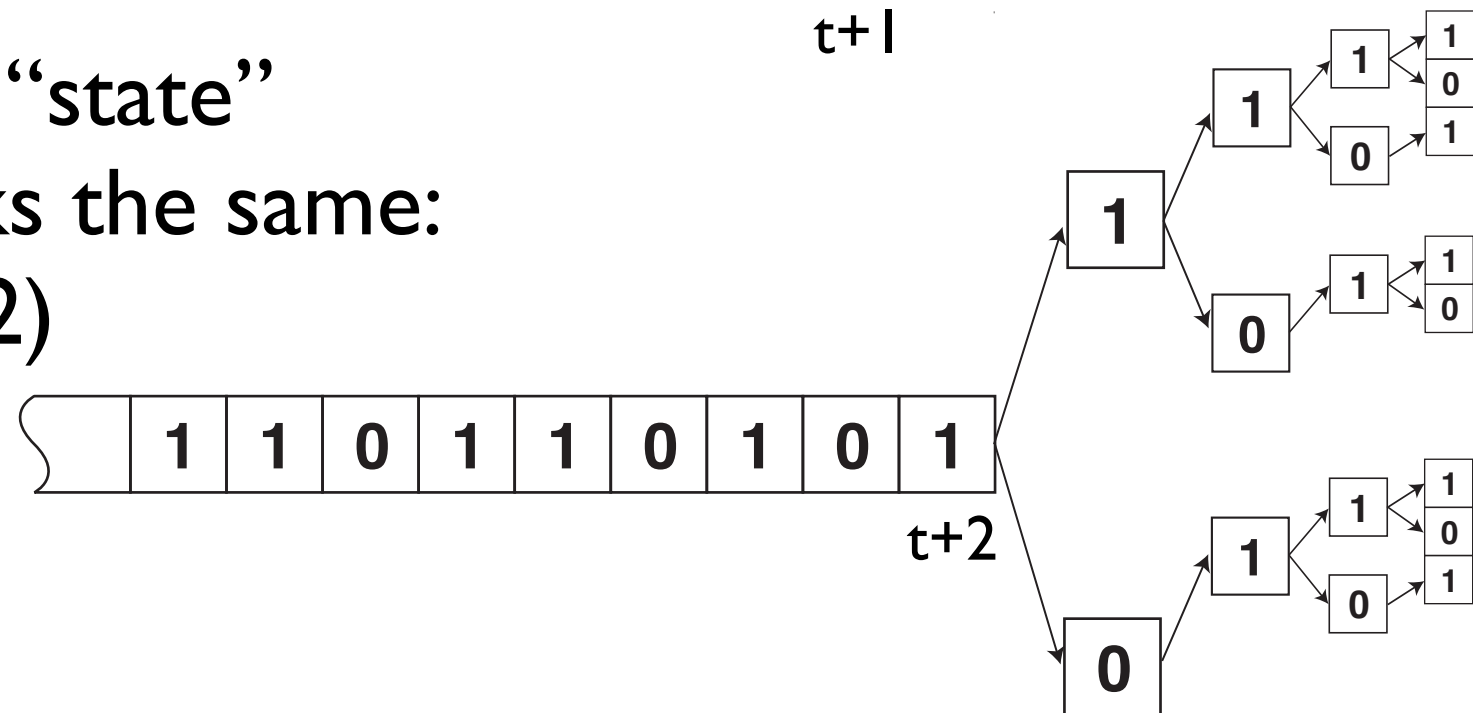
The Learning Channel ...

Effective States:

Process is in different “states”
when futures look different
 $\text{State}(t) \not\sim \text{State}(t+1)$



Process is in the same “state”
when the future looks the same:
 $\text{State}(t) \sim \text{State}(t+2)$



The Learning Channel ...

Effective for what?

What's a prediction?

A mapping from the past to the future.

Process $\Pr(\overleftrightarrow{S}) : \overleftrightarrow{S} = \overleftarrow{S} \overrightarrow{S}$

Future: $\overrightarrow{S}^L$

Particular past: $\overleftarrow{s}$

Future Morph: $\Pr(\overrightarrow{S}^L | \overleftarrow{s})$ (the most general mapping)

Refined goal:

Predict as much about the future $\overrightarrow{S}$,
using as little of the past $\overleftarrow{S}$ as possible.

The Learning Channel ...

How Effective are the Effective States?

Candidate “rival” model R

(Think some HMM)

A given mapping from pasts to future morphs

How to measure goodness?

Effective Prediction Error:

$$H[\vec{S}^L | R]$$

Uncertainty about future given effective states

Effective Prediction Error Rate:

$$h_{\mu}(R) = \lim_{L \rightarrow \infty} \frac{H[\vec{S}^L | R]}{L}$$

Entropy rate given effective states

The Learning Channel ...

How Effective are the Effective States?

Statistical Complexity of the Effective States:

$$C_{\mu}(R) = H[R] = H(\text{Pr}(R))$$

Interpretations:

Uncertainty in state.

Shannon information one gains when told effective state.

Model “size” $\propto \log_2(\text{number of states})$

Historical memory used by R .

The Learning Channel ...

Goals Restated:

Question 1:

Can we find effective states that give good predictions?

$$H[\vec{S}^L | R] = H[\vec{S}^L | \overset{\leftarrow}{S}]$$

or

$$h_{\mu}(R) = h_{\mu}$$

Question 2:

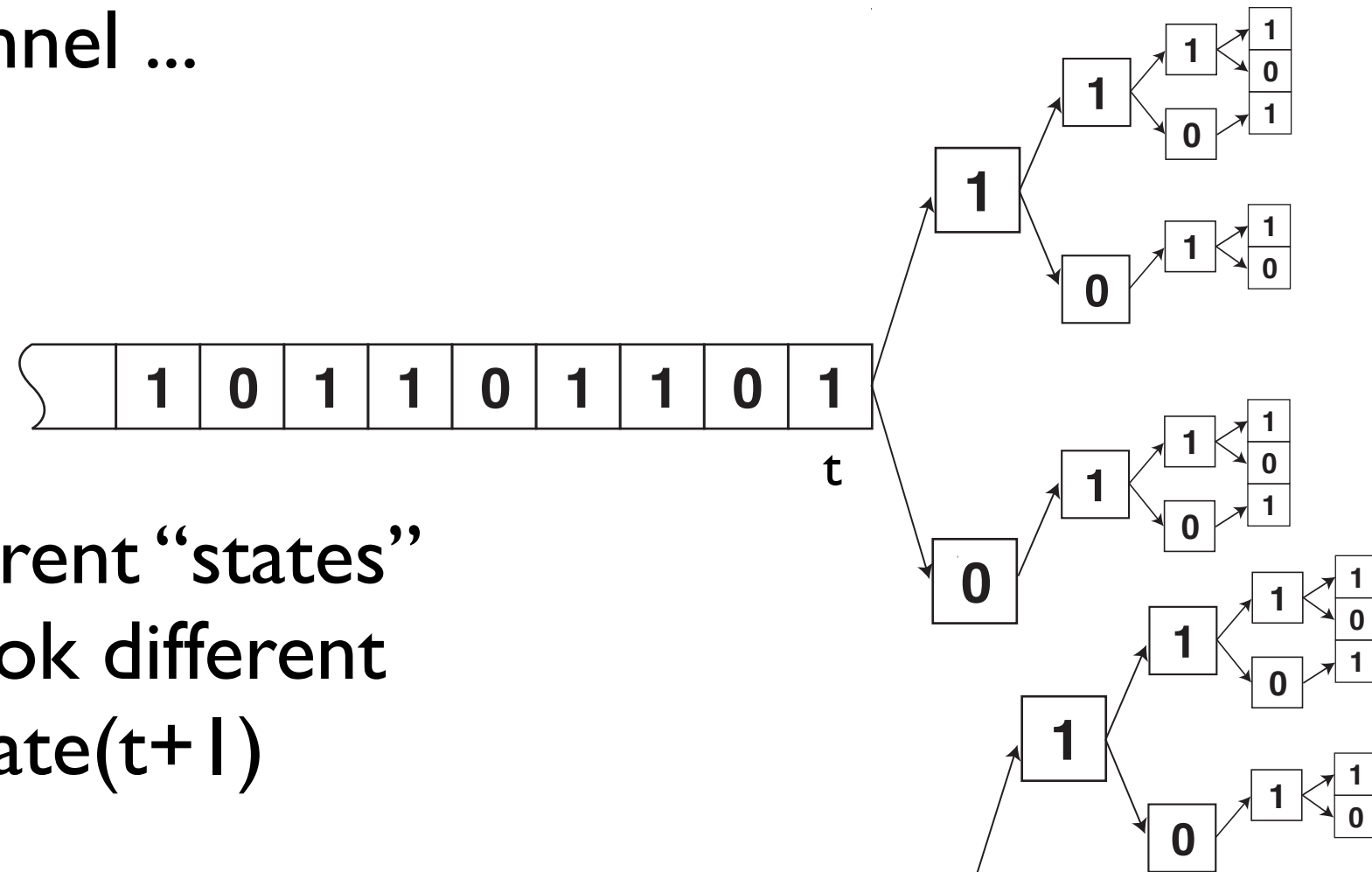
Can we find the smallest such set?

$$\min C_{\mu}(R)$$

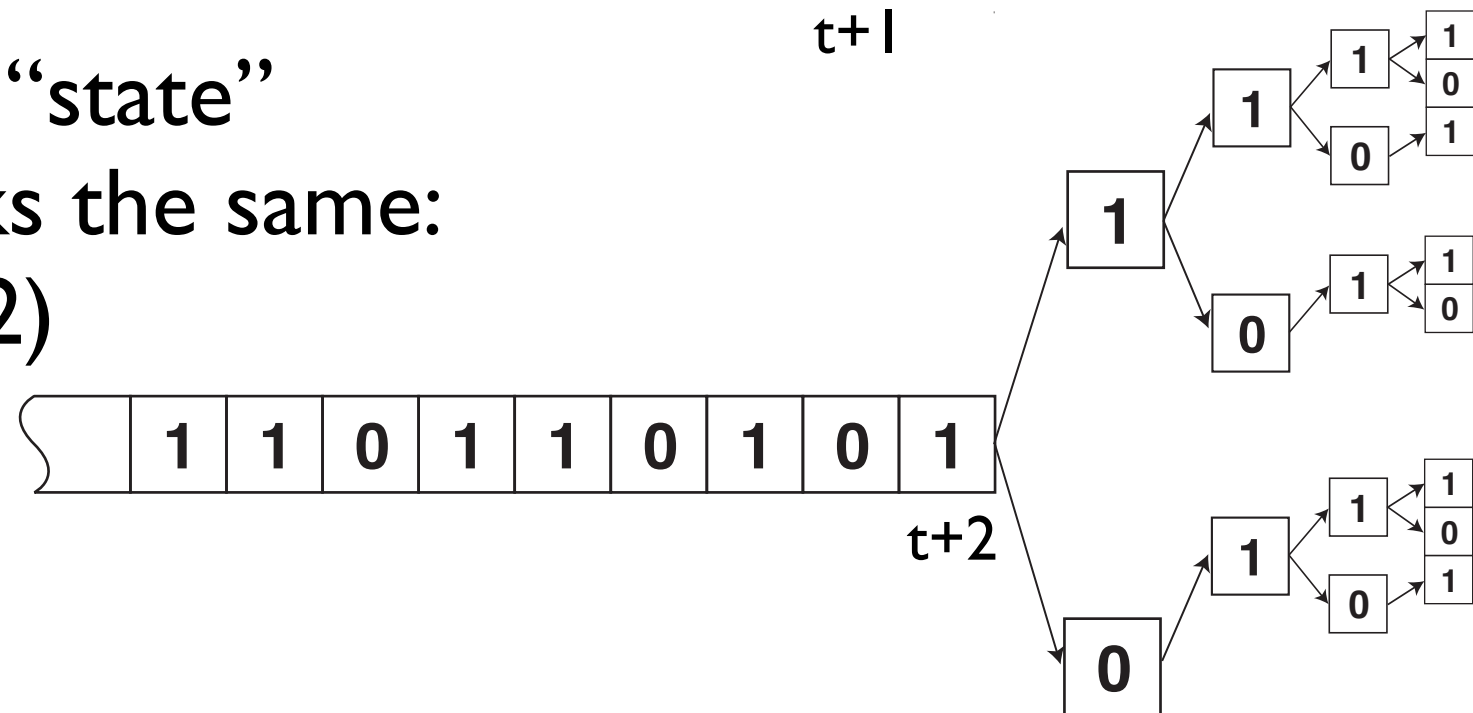
The Learning Channel ...

Effective States:

Process is in different “states”
when futures look different
 $\text{State}(t) \not\sim \text{State}(t+1)$



Process is in the same “state”
when the future looks the same:
 $\text{State}(t) \sim \text{State}(t+2)$



The Learning Channel ...

Causal States:

Causal State:

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

Set of histories that lead to same predictions.

Predictive equivalence relation:

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

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

The Learning Channel ...

Causal State Components

Causal State = Pasts with same morph: $\Pr(\vec{S} \mid \overleftarrow{s})$

$$\mathcal{S} = \{ \overleftarrow{s}' : \overleftarrow{s}' \sim \overleftarrow{s} \}$$

Set of causal states:

$$\mathcal{S} = \overleftarrow{\mathbf{S}} / \sim = \{ \mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \dots \}$$

Causal state map:

$$\epsilon : \overleftarrow{\mathbf{S}} \rightarrow \mathcal{S}$$

$$\epsilon(\overleftarrow{s}) = \{ \overleftarrow{s}' : \overleftarrow{s}' \sim \overleftarrow{s} \}$$

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?

The Learning Channel ...

Causal State Dynamic ...

Causal-state Filtering:

$$\begin{aligned}\overleftrightarrow{s} &= \dots s_{-3} \quad s_{-2} \quad s_{-1} \quad s_0 \quad s_1 \quad s_2 \quad s_3 \quad \dots \\ \epsilon(\overleftrightarrow{s}) &= \dots \epsilon(\overleftarrow{s}_{-3}) \epsilon(\overleftarrow{s}_{-2}) \epsilon(\overleftarrow{s}_{-1}) \epsilon(\overleftarrow{s}_0) \epsilon(\overleftarrow{s}_1) \epsilon(\overleftarrow{s}_2) \epsilon(\overleftarrow{s}_3) \dots \\ \overleftrightarrow{\mathcal{S}} &= \dots \mathcal{S}_{t=-3} \mathcal{S}_{t=-2} \mathcal{S}_{t=-1} \mathcal{S}_{t=0} \mathcal{S}_{t=1} \mathcal{S}_{t=2} \mathcal{S}_{t=3} \dots\end{aligned}$$

Causal-state process:

$$\Pr(\overleftrightarrow{\mathcal{S}})$$

The Learning Channel ...

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:

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

The ϵ -Machine ...

Process $\Rightarrow$ Predictive equivalence $\Rightarrow \epsilon$ - Machine

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

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

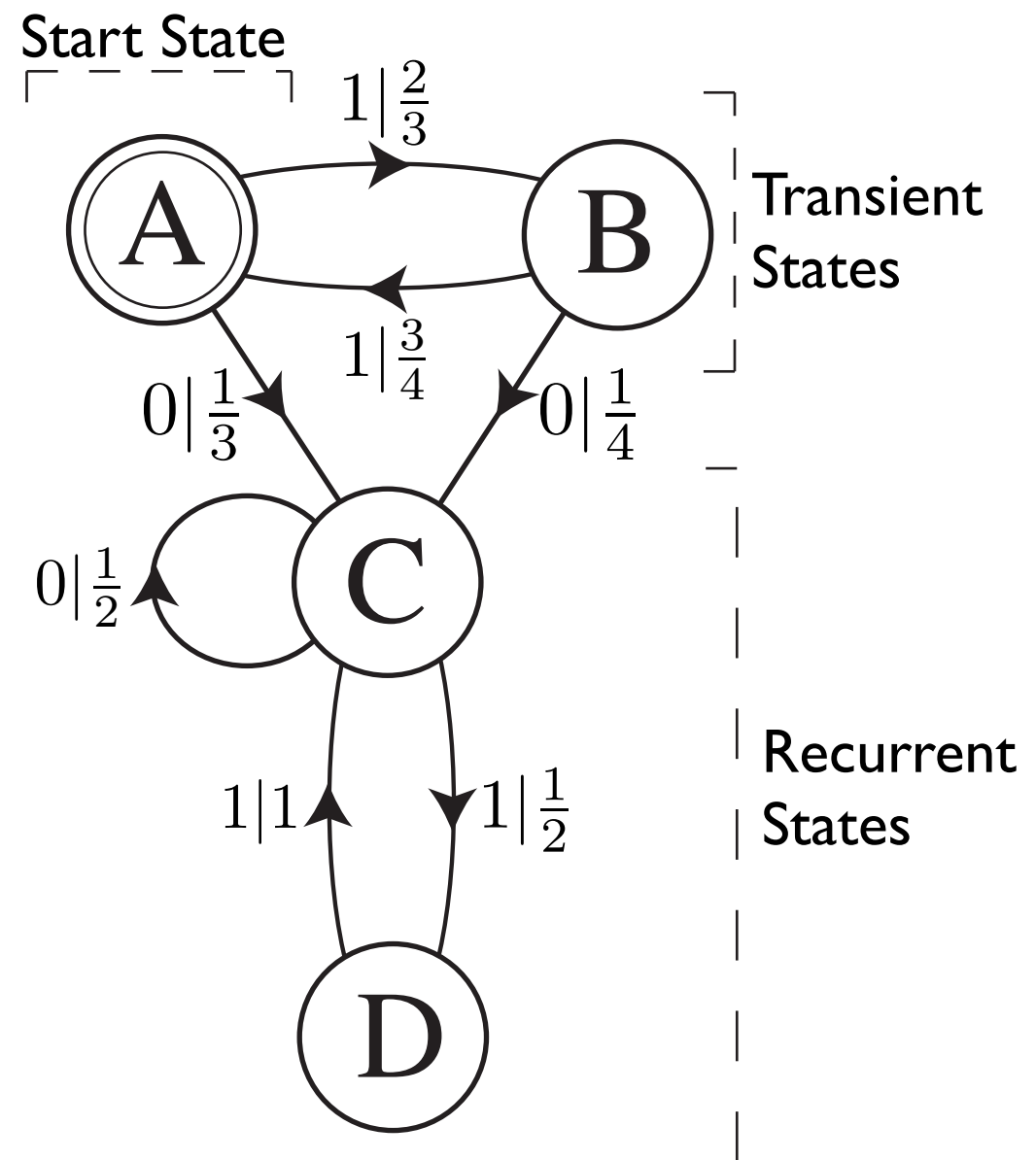
Unique Start State:

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

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

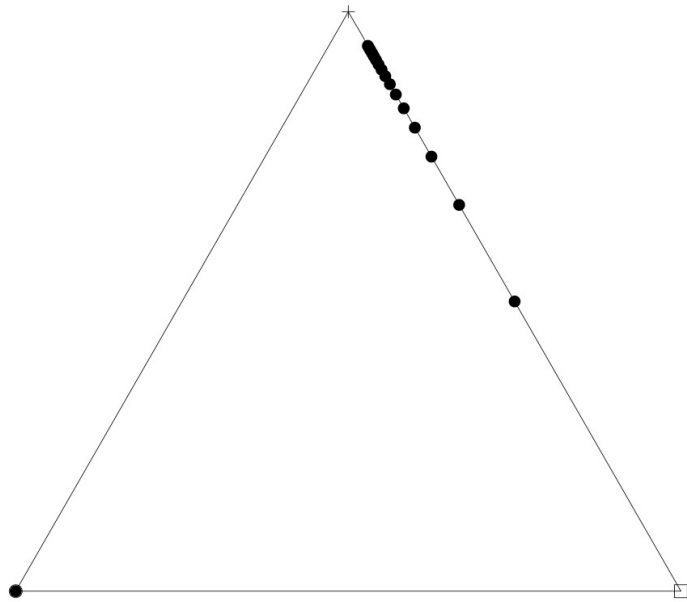
Transient States

Recurrent States

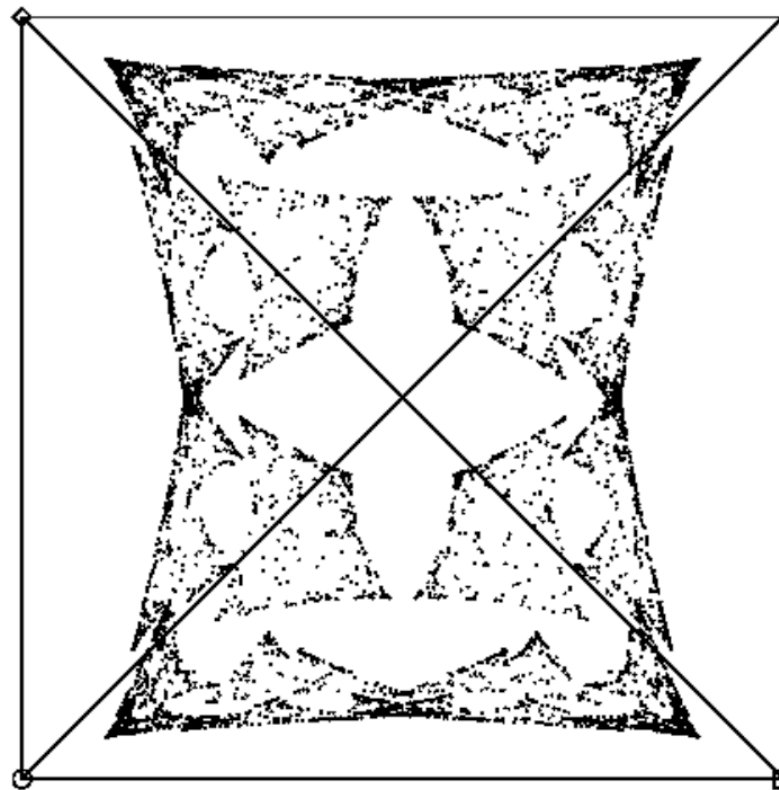


The Learning Channel ...

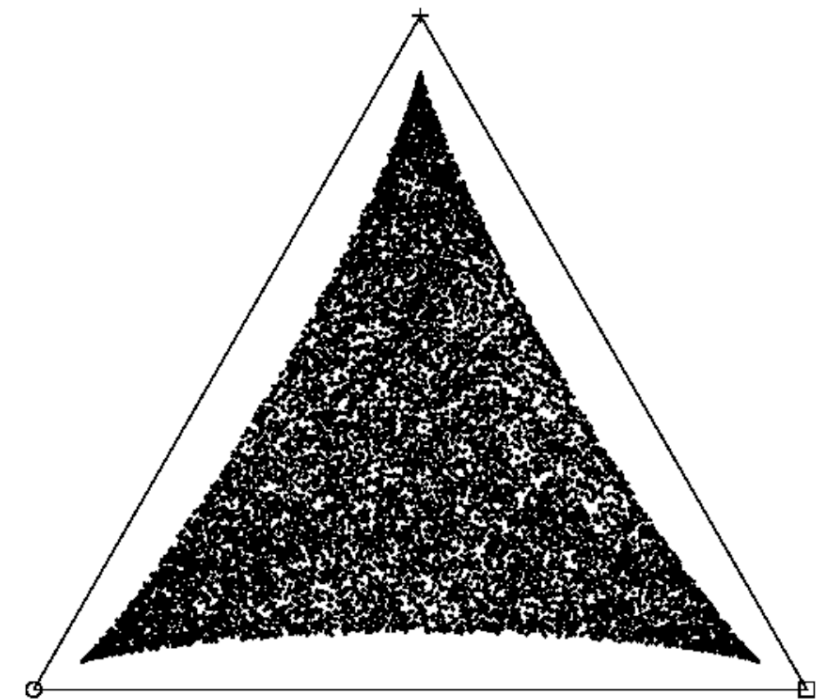
The ϵ -Machine of a Process ...



**Denumerable
Causal States**

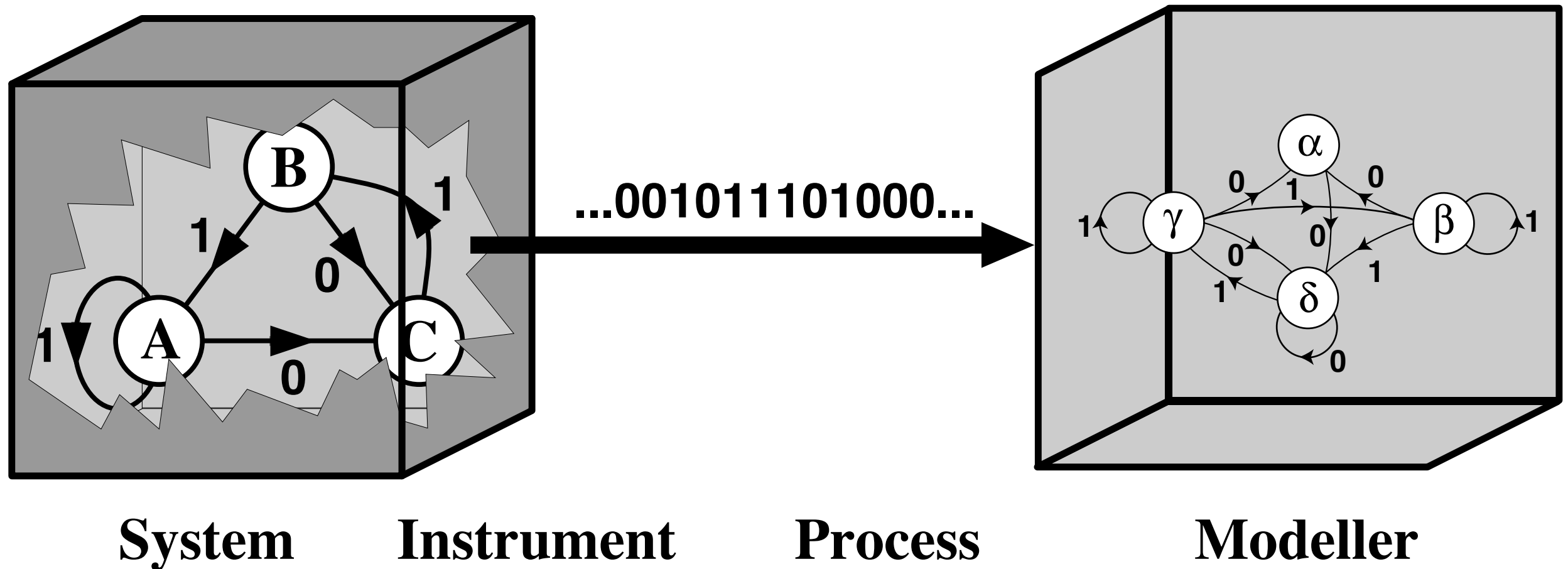


Fractal



Continuous

The Learning Channel:



Central questions:

What are the states? Causal States

What is the dynamic? The ϵ -Machine

The ϵ -Machine

The ϵ -Machine ...

A **Model** of a Process $\Pr(\vec{S})$:

ϵ -Machine reproduces the process's word distribution:

$$\Pr(s^1), \Pr(s^2), \Pr(s^3), \dots$$

$$s^L = s_1 s_2 \dots s_L \quad \mathcal{S}(t=0) = \mathcal{S}_0$$

$$\begin{aligned} \Pr(s^L) = & \Pr(\mathcal{S}_0) \Pr(\mathcal{S}_0 \rightarrow_{s=s_1} \mathcal{S}(1)) \Pr(\mathcal{S}(1) \rightarrow_{s=s_2} \mathcal{S}(2)) \\ & \dots \Pr(\mathcal{S}(L-1) \rightarrow_{s=s_L} \mathcal{S}(L)) \end{aligned}$$

Initially, $\Pr(\mathcal{S}_0) = 1$.

$$\Pr(s^L) = \prod_{l=1}^L T_{i=\epsilon(s^{l-1}), j=\epsilon(s^l)}^{(s^l)}$$

The ϵ -Machine ...

Causal shielding:

Past and future are independent given causal state

$$\text{Process: } \Pr(\vec{S}) = \Pr(\overleftarrow{S} \overrightarrow{S})$$

$$\Pr(\overleftarrow{S} \overrightarrow{S} | \mathcal{S}) = \Pr(\overleftarrow{S} | \mathcal{S}) \Pr(\overrightarrow{S} | \mathcal{S})$$

Causal states shield past & future from each other.

Similar to states of a Markov chain, but for hidden processes.

The ϵ -Machine ...

ϵ Ms are **Unifilar**: $(\mathcal{S}_t, s) \rightarrow \text{unique } \mathcal{S}_{t+1}$
(in automata theory, “deterministic”)

Consequence:

Unifilarity: 1-1 map between state-sequences & symbol-sequences.

Entropy rate expression requires this 1-1 mapping.

Can use ϵ M to calculate entropy rate h_μ .
(Any unifilar presentation will do.)

The ϵ -Machine ...

ϵ Ms are **Optimal Predictors**:

Compared to any rival effective states R :

$$H \left[\begin{array}{c|c} \vec{S}^L & R \end{array} \right] \geq H \left[\begin{array}{c|c} \vec{S}^L & \mathcal{S} \end{array} \right]$$

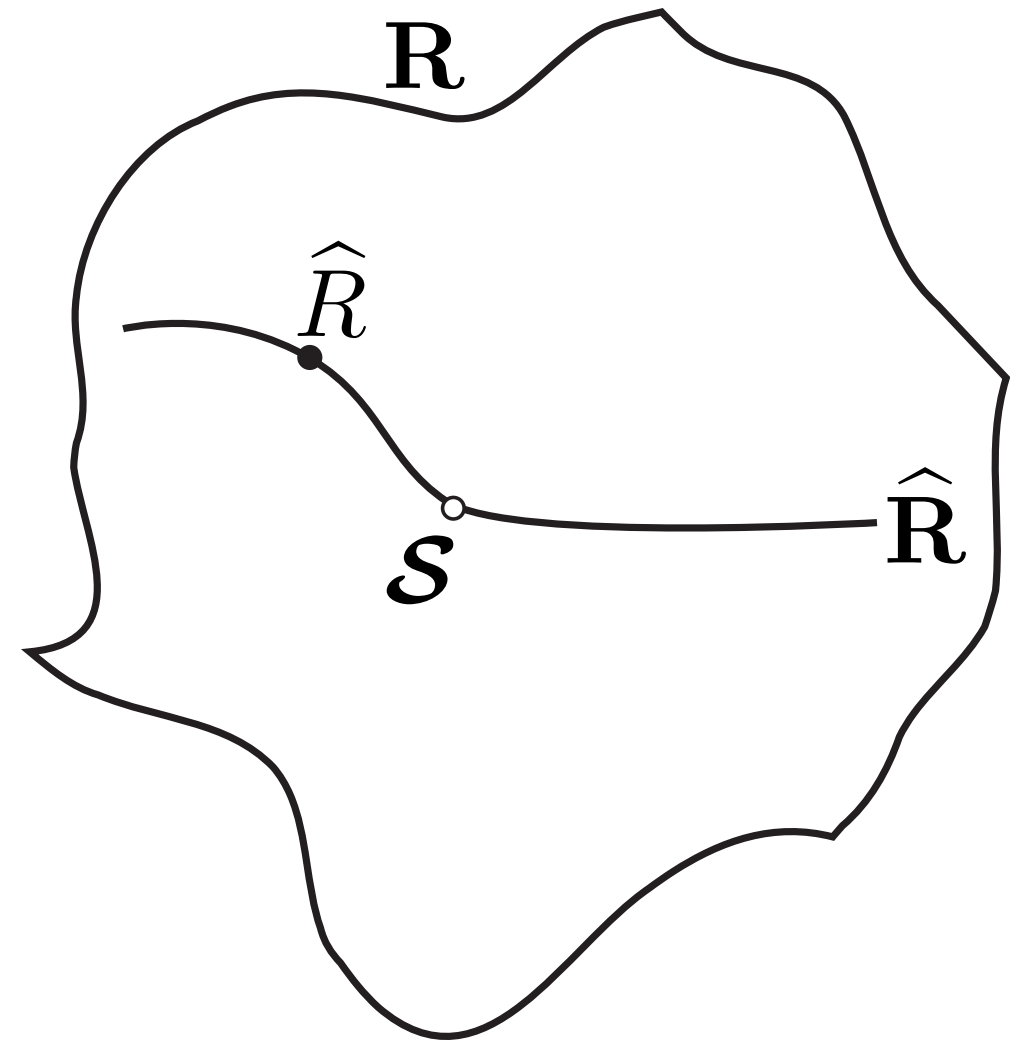
The ϵ -Machine ...

Prescient Rivals $\hat{\mathbf{R}}$:

Alternative models that are optimal predictors

$$\hat{R} \in \hat{\mathbf{R}}$$

$$H[\vec{S}^L | \hat{R}] = H[\vec{S}^L | \mathcal{S}]$$



(Prescient rivals are sufficient statistics for process's future.)

The ϵ -Machine ...

Minimal Statistical Complexity:

For all prescient rivals, ϵM is the smallest:

$$C_{\mu}(\hat{R}) \geq C_{\mu}(\mathcal{S})$$

Consequence:

- (1) C_{μ} measures historical information process stores.
- (2) This would not be true, if not minimal representation.

Remarks:

- (1) Causal states contain every difference (in past)
that makes a difference (to future) (Bateson “information”)
- (2) Causal states are sufficient statistics for the future.

The ϵ -Machine ...

Summary:

ϵM :

- (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.

Measures of Intrinsic Computation

Measures of Intrinsic Computation ...

A complex process's **intrinsic computation**:

(1) How much of past does process store?

(2) In what architecture is that information stored?

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

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?

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

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?

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

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

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?

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

(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	E	Info: Past to Future
Total Predictability	G	Redundancy
Transient Information	T	Synchronization

How related to statistical complexity C_μ ?

How to get from ϵM ?

Measures of Intrinsic Computation ...

Measures from the ϵM :

Entropy Rate of a Process:

$$h_{\mu}(\text{Pr}(\vec{S})) = \lim_{L \rightarrow \infty} \frac{H(L)}{L}$$

Entropy Rate given ϵM :

$$h_{\mu}(\mathcal{S}) = - \sum_{\mathcal{S} \in \mathcal{S}} \text{Pr}(\mathcal{S}) \sum_{s \in \mathcal{A}, \mathcal{S}' \in \mathcal{S}} T_{\mathcal{S}\mathcal{S}'}^{(s)} \log_2 T_{\mathcal{S}\mathcal{S}'}^{(s)},$$

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

Possible only due to ϵM unifilarity!

I-I mapping between measurement sequences & internal paths.

Measures of Intrinsic Computation ...

Measures from the ϵM ...

Statistical Complexity of a Process:

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

where $\text{Pr}(\mathcal{S})$ is causal-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 ϵM ...

Excess Entropy: Three versions, all equivalent for IID processes

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

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

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

How to get, given ϵM ?

Special cases: When ϵM is IID, periodic, or spin chain.

General case: Need a new framework.

Measures of Intrinsic Computation ...

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

Bound on Excess Entropy:

$$\mathbf{E} \leq C_\mu$$

Proof sketch:

$$(1) \mathbf{E} = I[\vec{S}; \overleftarrow{S}] = H[\vec{S}] - H[\vec{S} | \overleftarrow{S}]$$

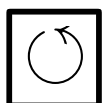
$$(2) \text{ Causal States: } H[\vec{S} | \overleftarrow{S}] = H[\vec{S} | \mathcal{S}]$$

$$(3) \mathbf{E} = H[\vec{S}] - H[\vec{S} | \mathcal{S}]$$

$$= I[\vec{S}; \mathcal{S}]$$

$$= H[\mathcal{S}] - H[\mathcal{S} | \vec{S}]$$

$$\leq H[\mathcal{S}] = C_\mu$$



Measures of Intrinsic Computation ...

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

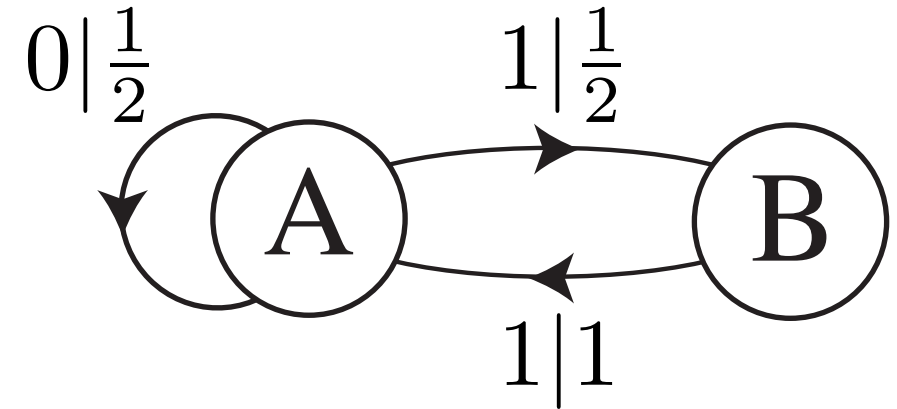
Bound on Excess Entropy ...

But, the bound is saturated!

Even Process:

$$C_\mu = H(2/3) \approx 0.9182$$

$$\mathbf{E} \approx 0.9182$$



$$T^{(0)} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}$$

$$T^{(1)} = \begin{pmatrix} 0 & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

$$\pi_V = (2/3, 1/3)$$

When does this occur?

In general, need a new framework for answering this question:
Directional Computational Mechanics.

Measures of Intrinsic Computation ...

Measures from the ϵM ...

Bound on Excess Entropy ...

Consequence: The Cryptographic Limit

Can have $\mathbf{E} \rightarrow 0$ when $C_\mu \gg 1$.

Excess entropy is *not* the process's stored information.

$\mathbf{E}$ is the *apparent* information,
as revealed in *measurement sequences*.

Statistical complexity *is* stored information.

Measures of Intrinsic Computation ...

Measures from the ϵM ...

Bound on Excess Entropy ...

Executive Summary:

C_μ is the amount of information the process uses

to communicate

E bits of information from the past to the future.

Measures of Intrinsic Computation ...

Measures from the ϵM ...

Bound on Excess Entropy: $E \leq C_\mu$

Consequence:

The inequality is Why We Must Model.

Cannot simply use sequences as states.

There is internal structure not expressed by this.

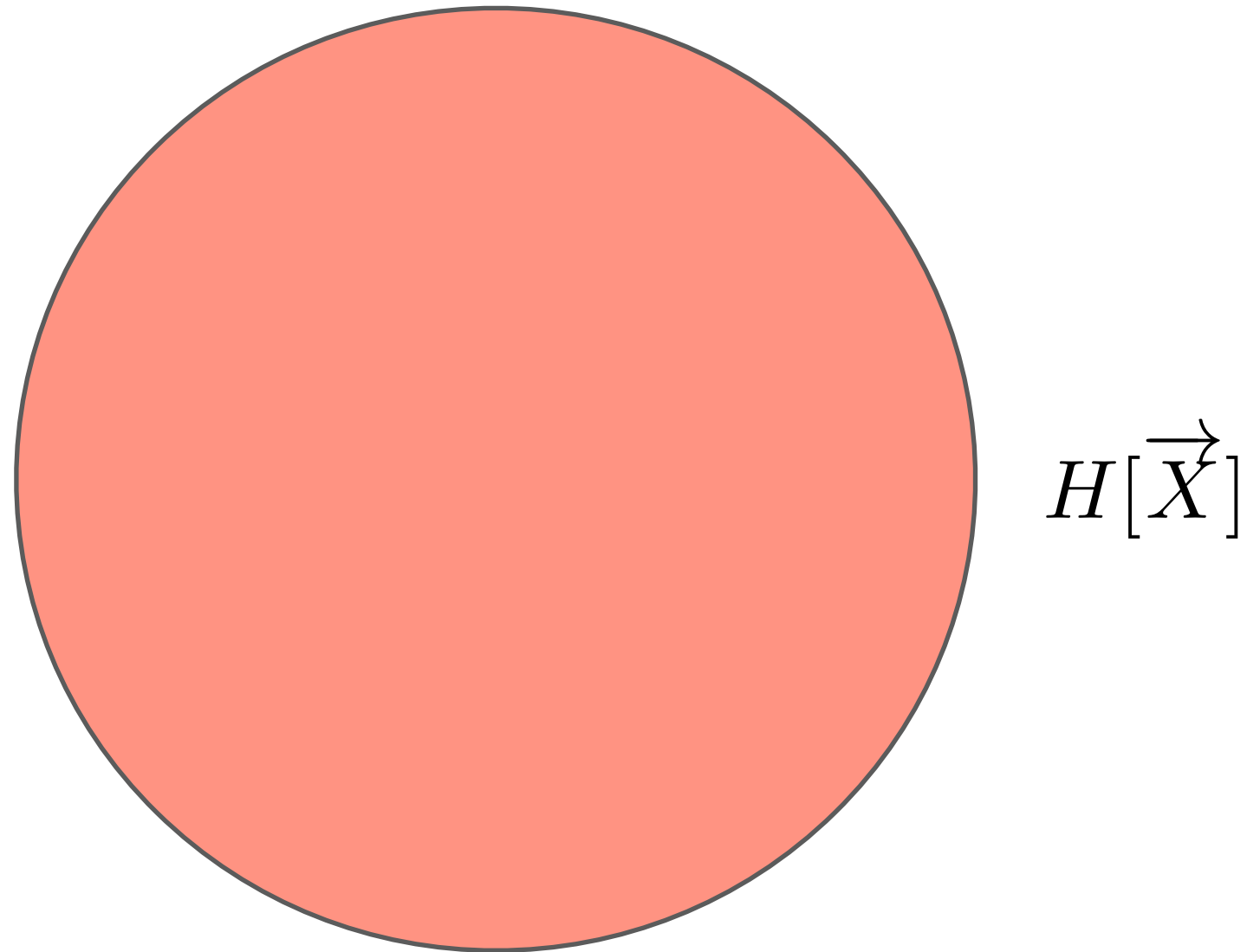
Information Diagrams for Processes

Information Diagrams for Processes

Process I-diagram:

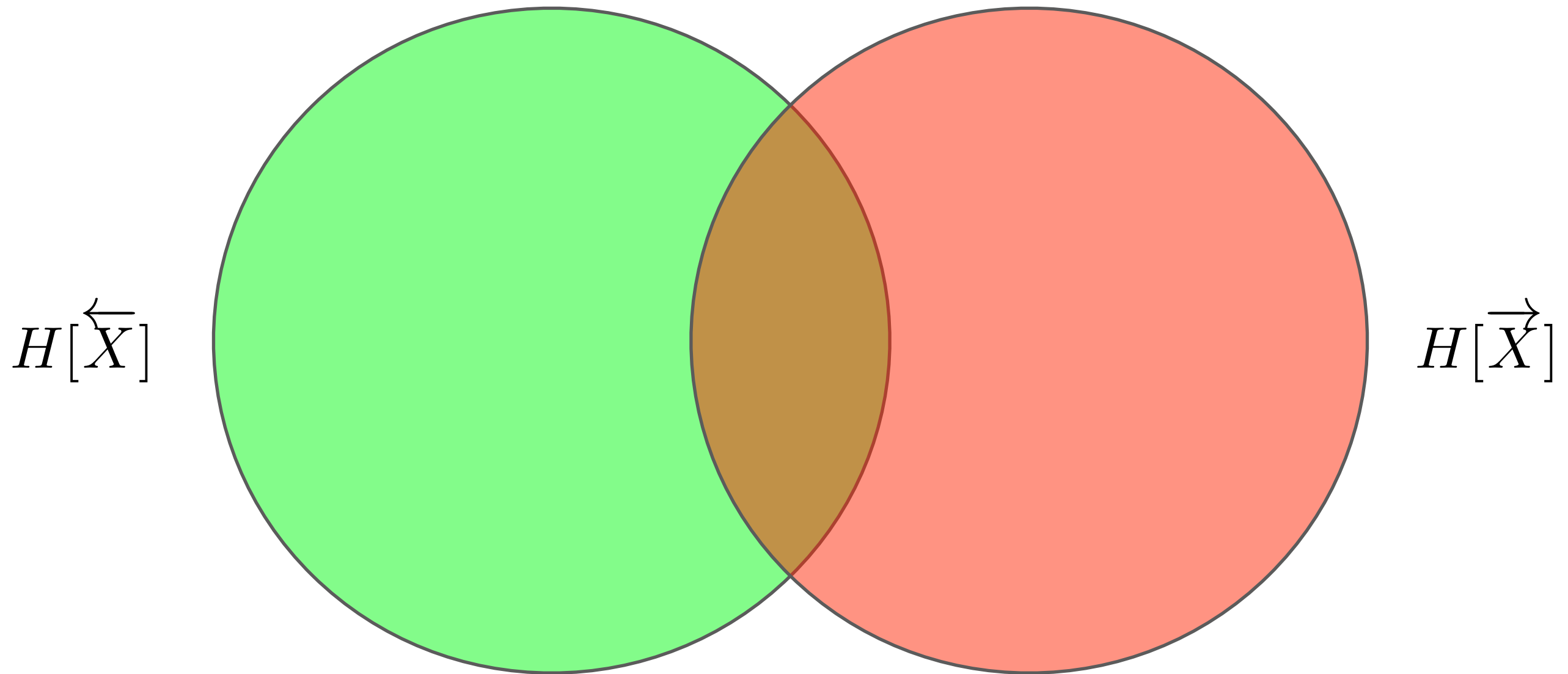
Information Diagrams for Processes

Process I-diagram:



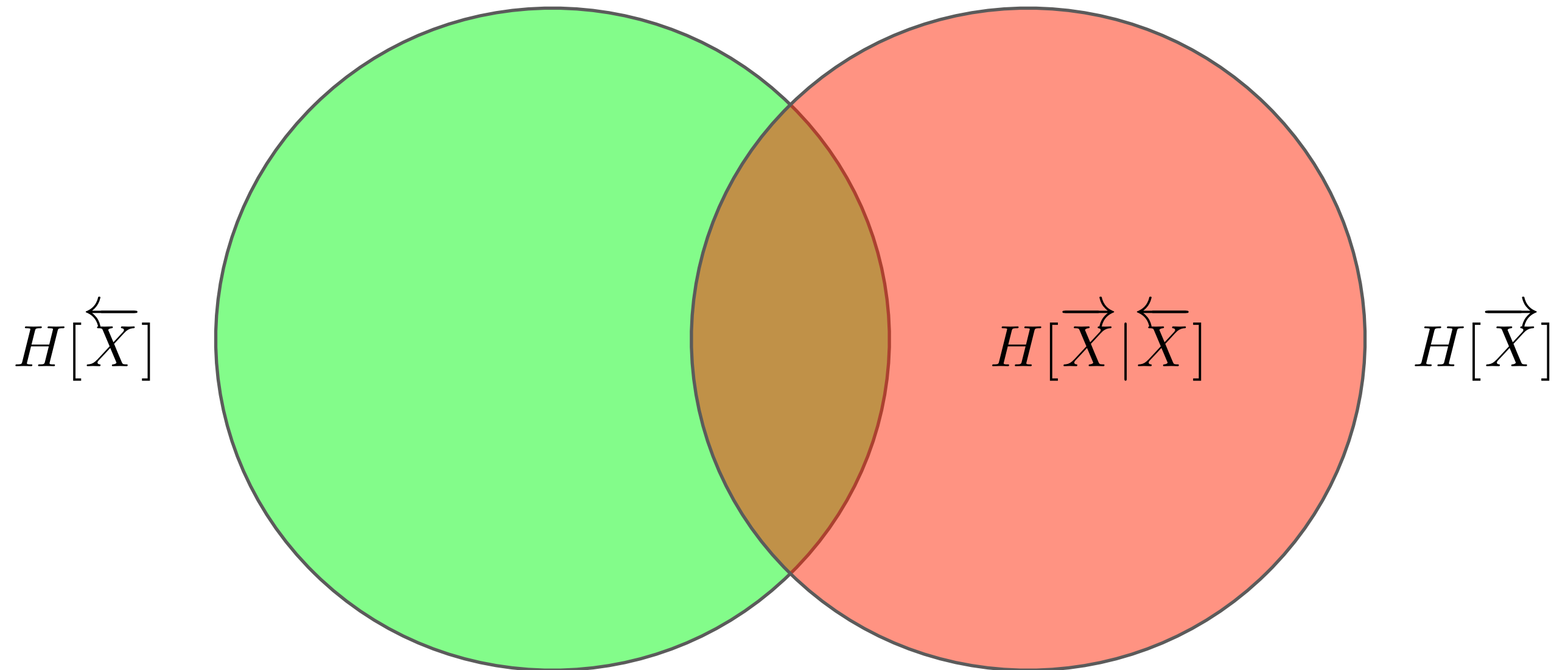
Information Diagrams for Processes

Process I-diagram:



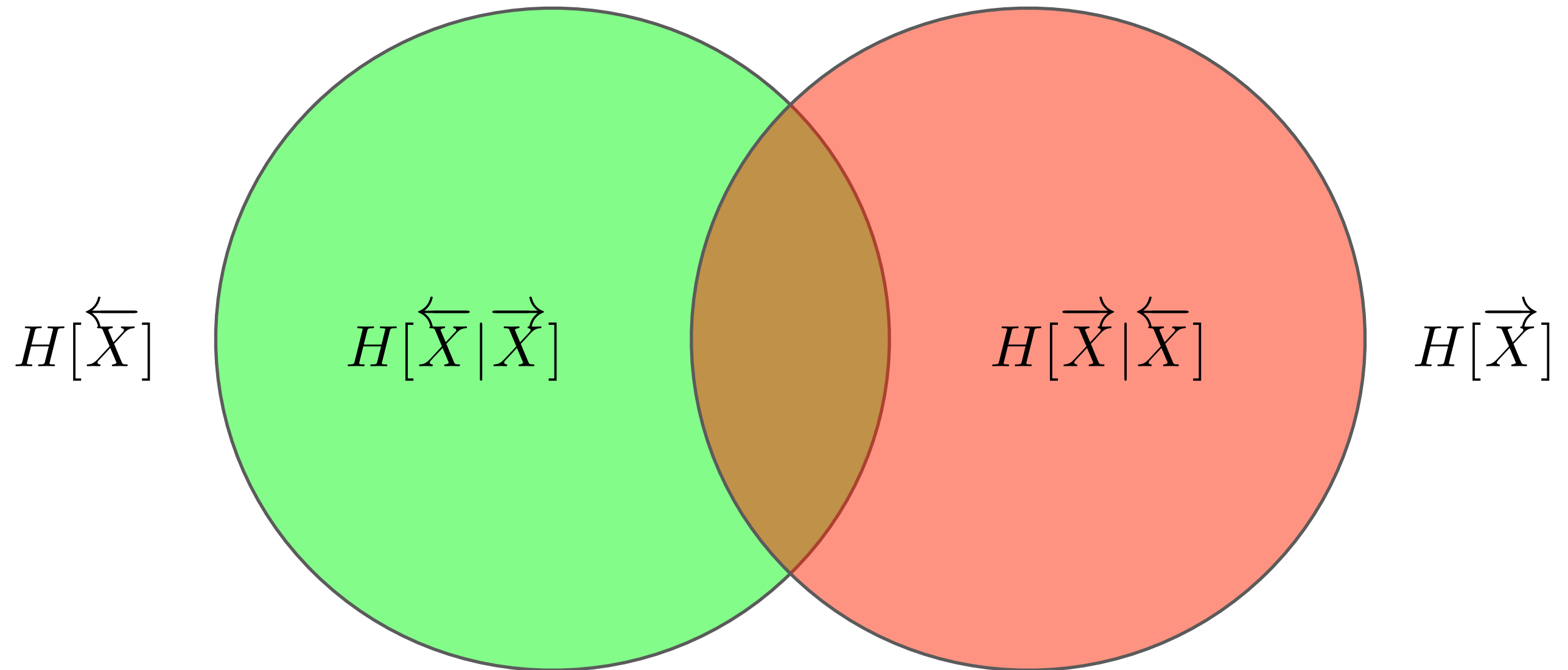
Information Diagrams for Processes

Process I-diagram:



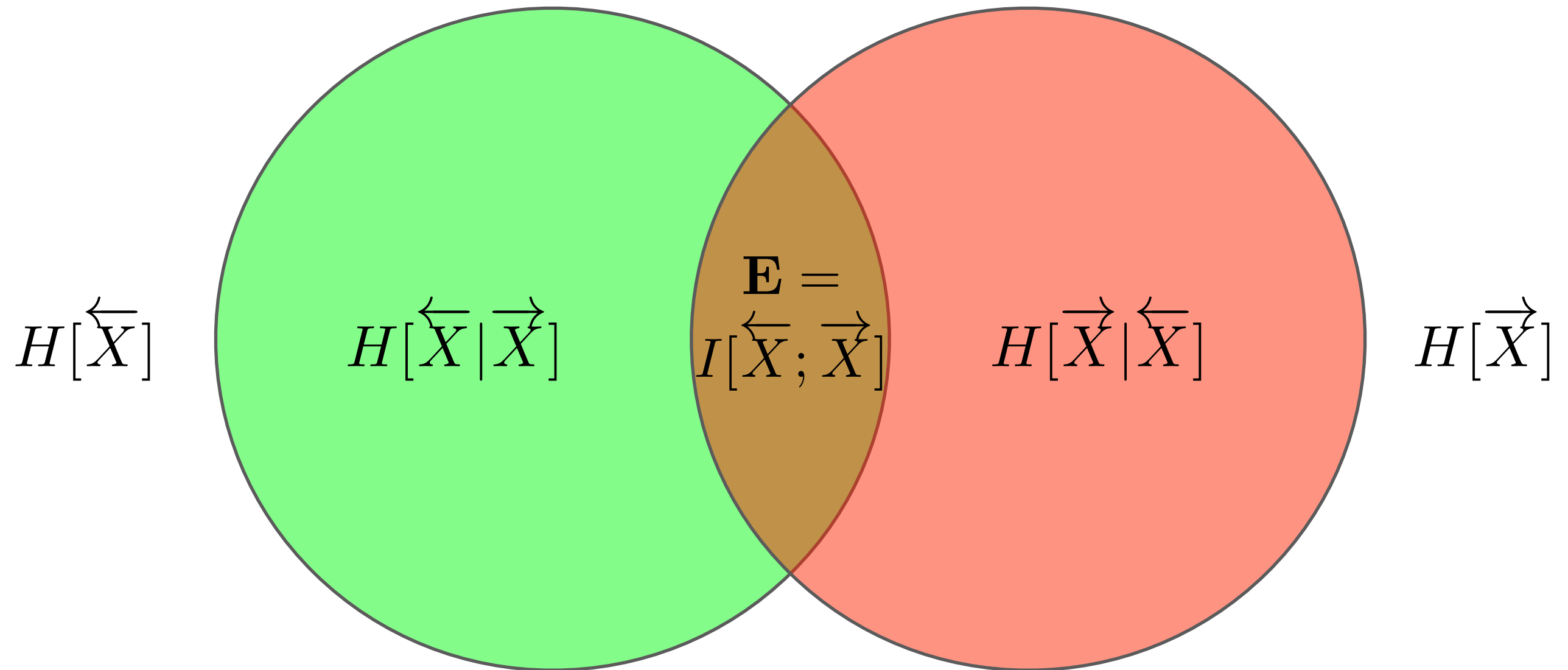
Information Diagrams for Processes

Process I-diagram:



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: $\mathcal{S} \in \mathcal{S}$

Information measures:

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

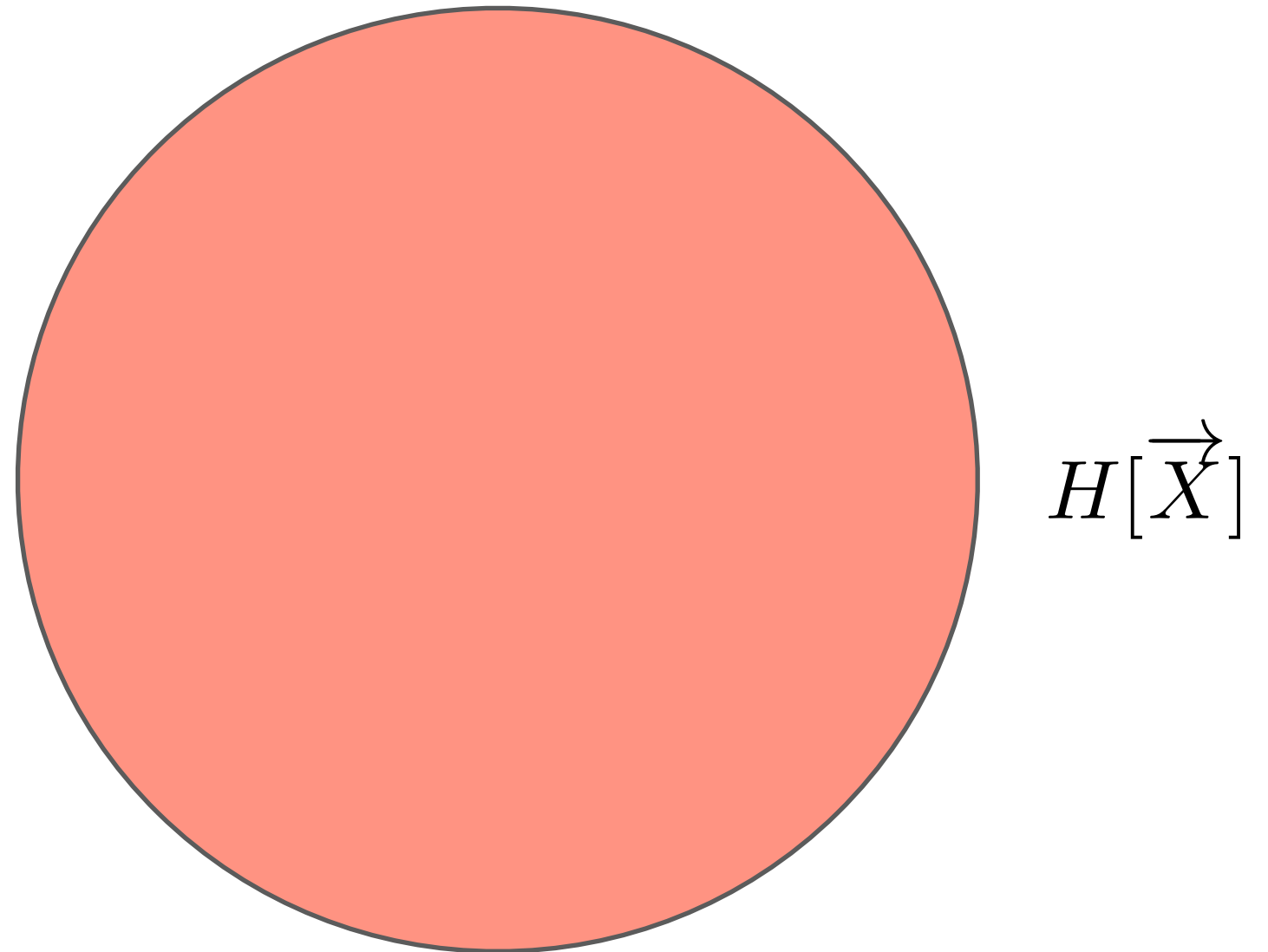
There are $8 = 2^3$ atomic information measures.

Information Diagrams for Processes

Process I-diagram using ε -machine ...

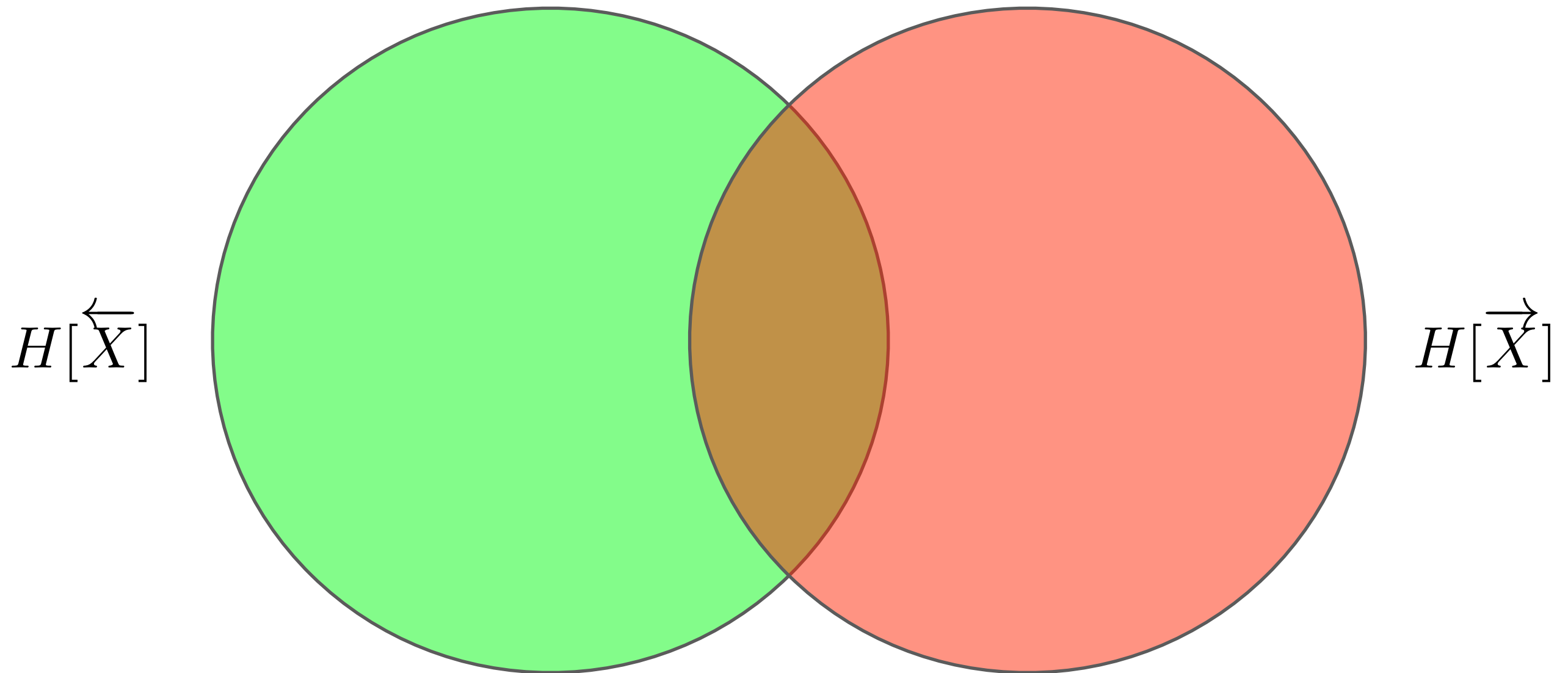
Information Diagrams for Processes

Process I-diagram using ε -machine ...



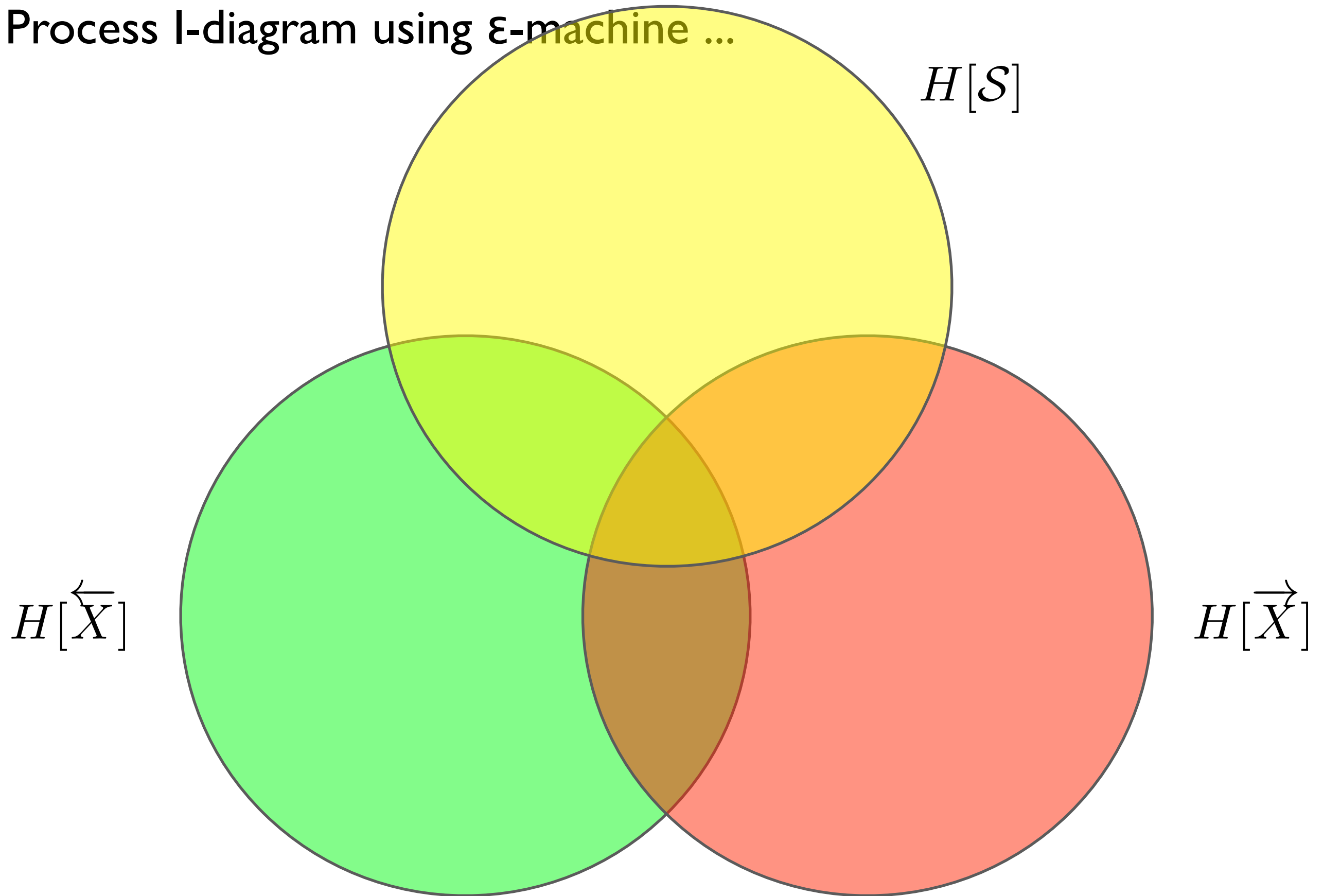
Information Diagrams for Processes

Process I-diagram using ε -machine ...



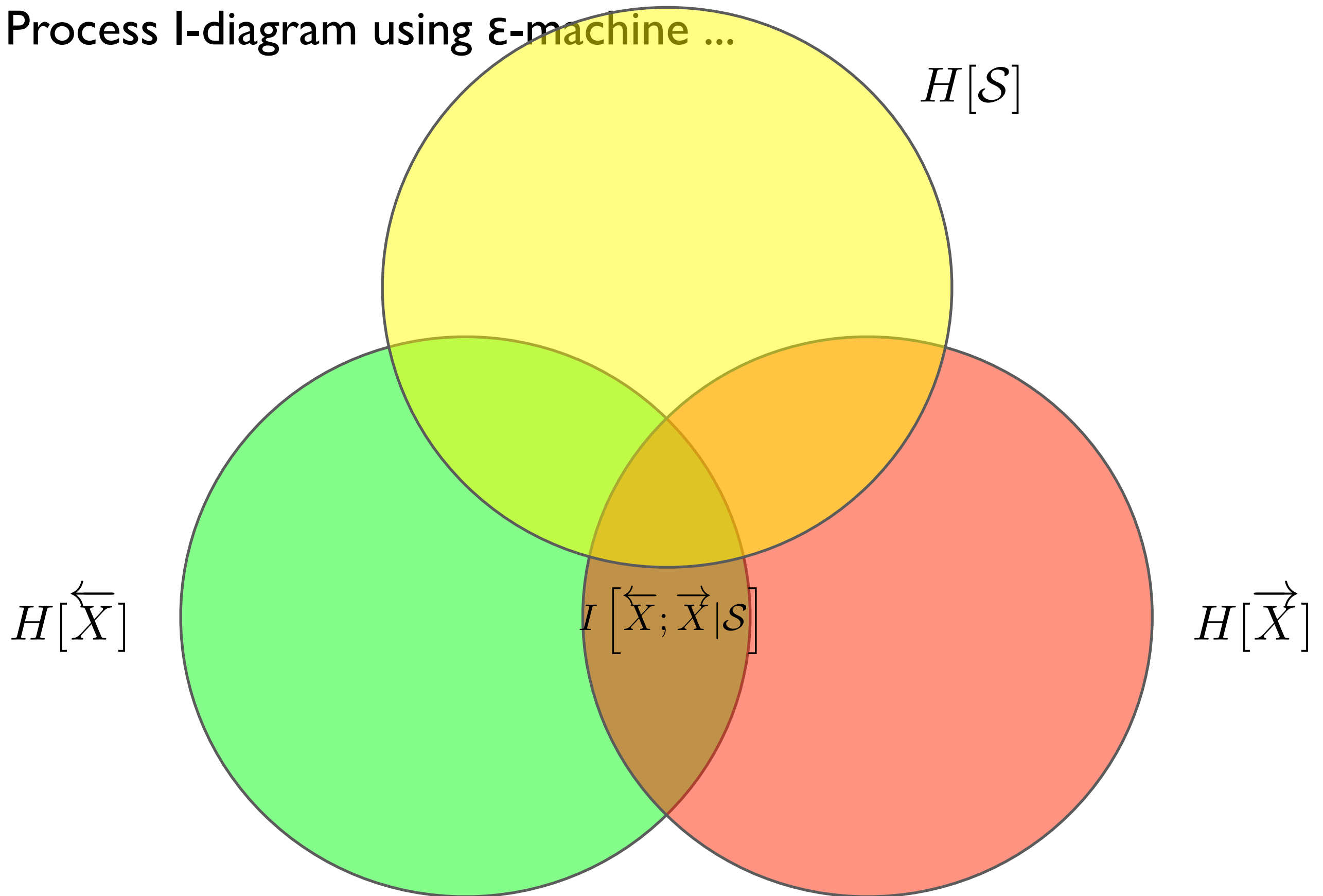
Information Diagrams for Processes

Process I-diagram using ε -machine ...



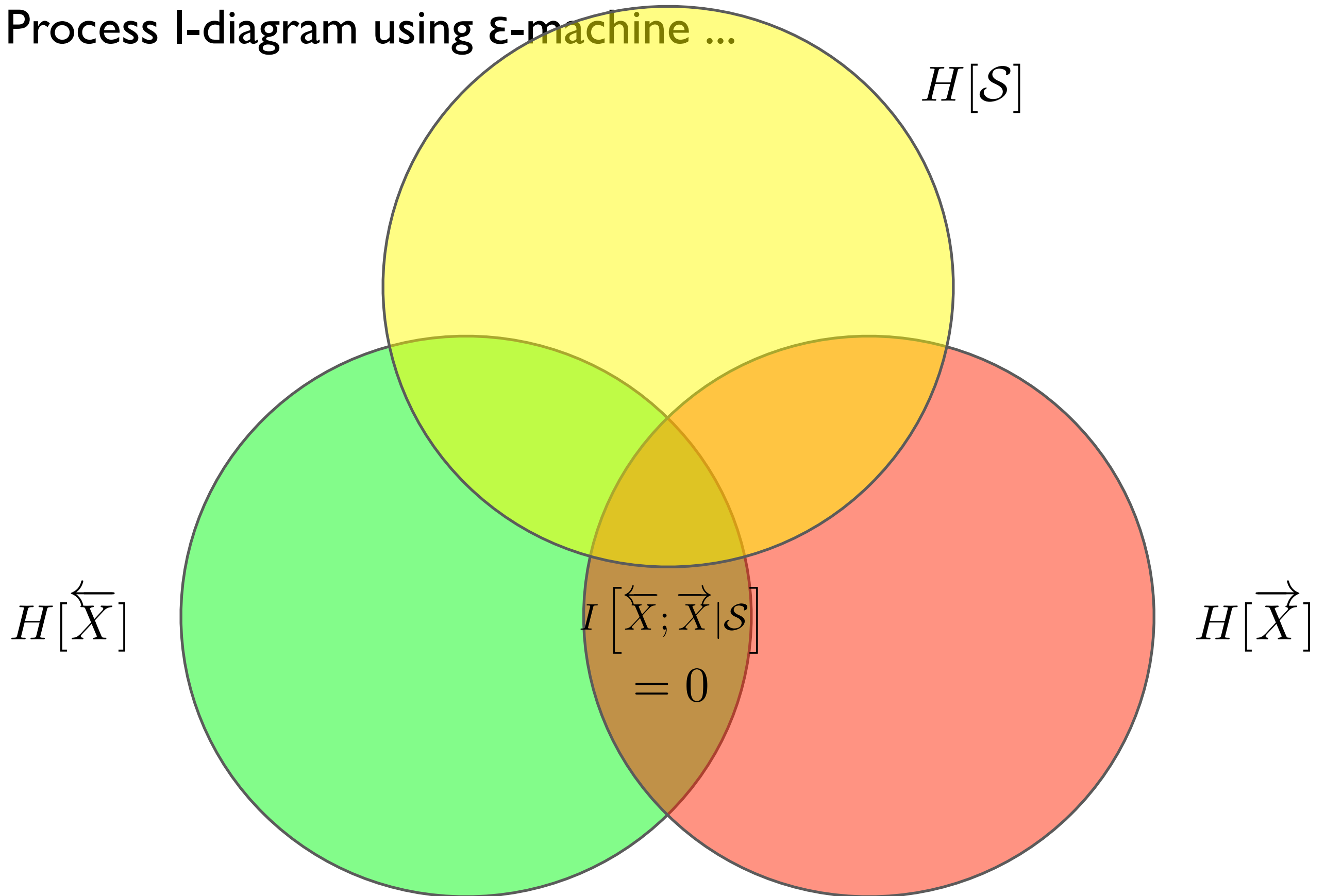
Information Diagrams for Processes

Process I-diagram using ε -machine ...



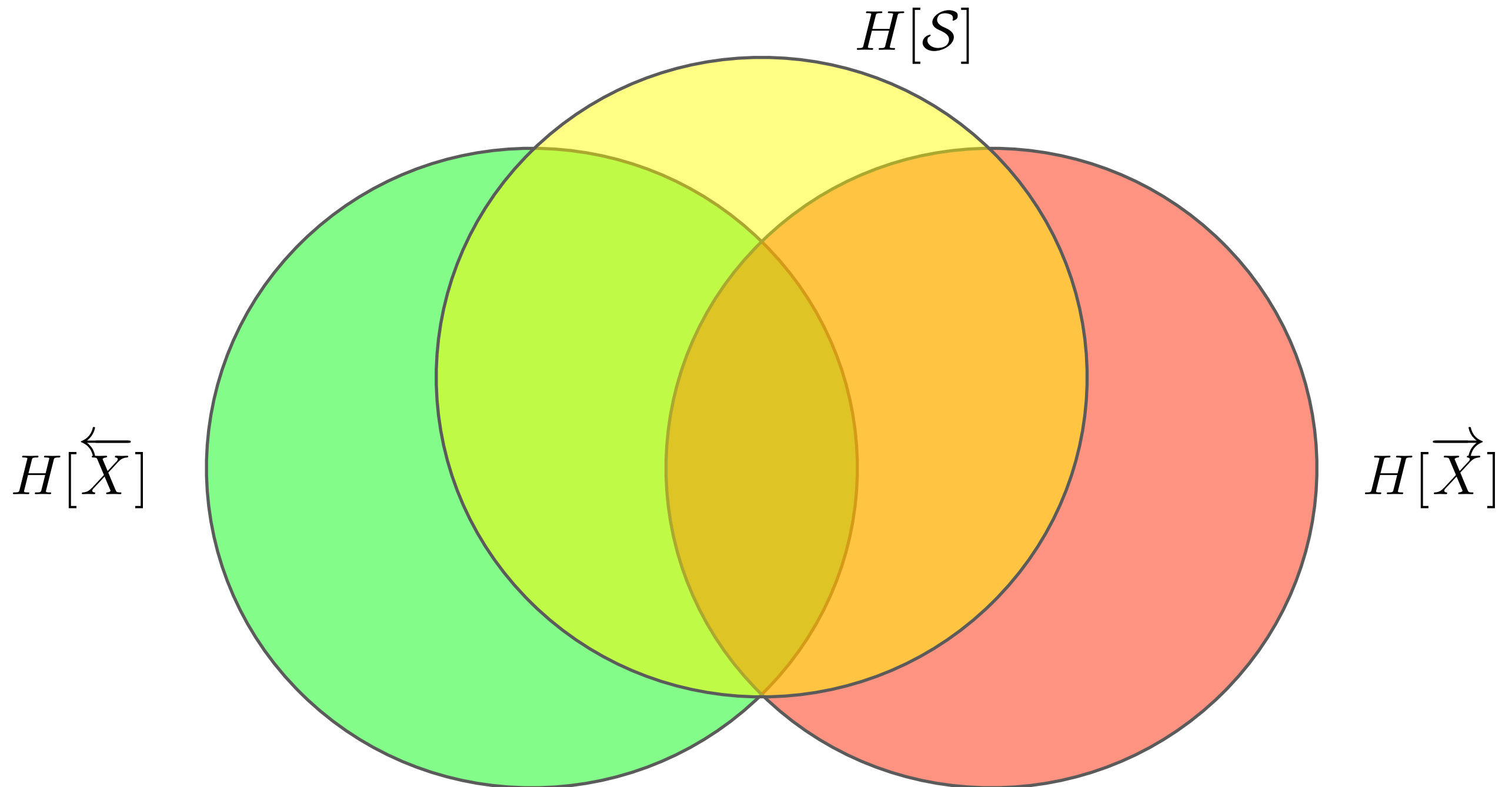
Information Diagrams for Processes

Process I-diagram using ε -machine ...



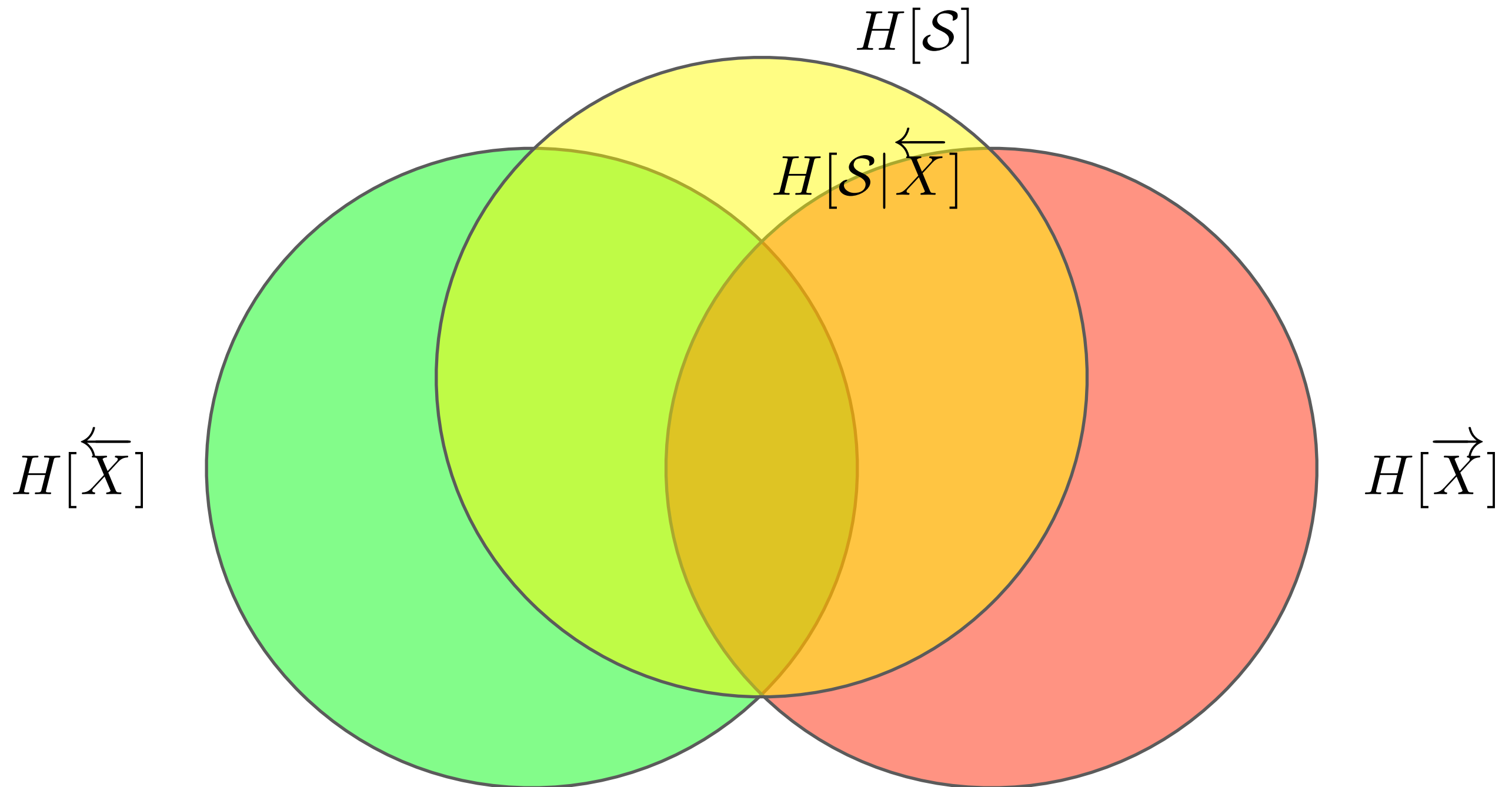
Information Diagrams for Processes

Process I-diagram using ε -machine ...



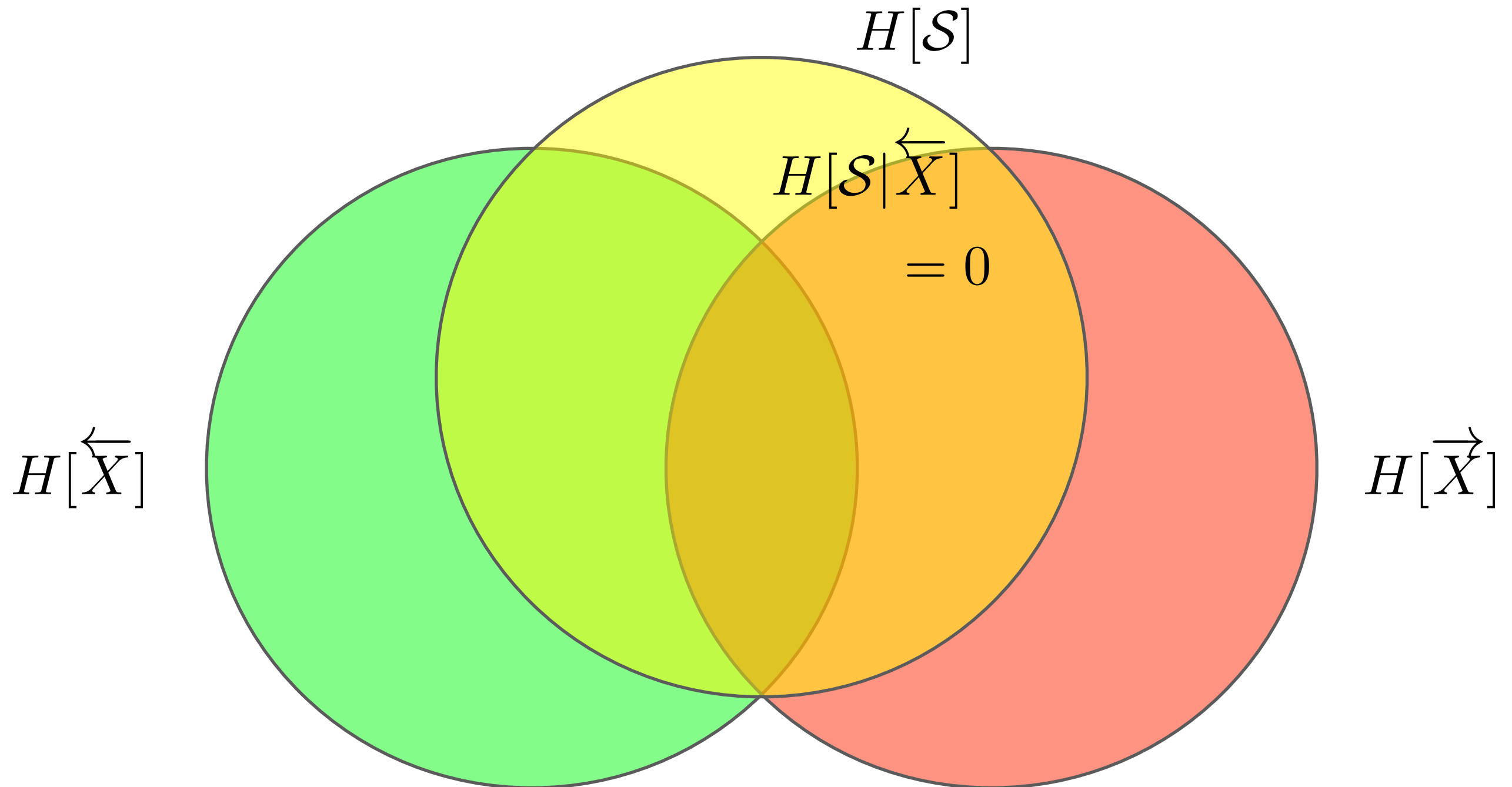
Information Diagrams for Processes

Process I-diagram using ε -machine ...



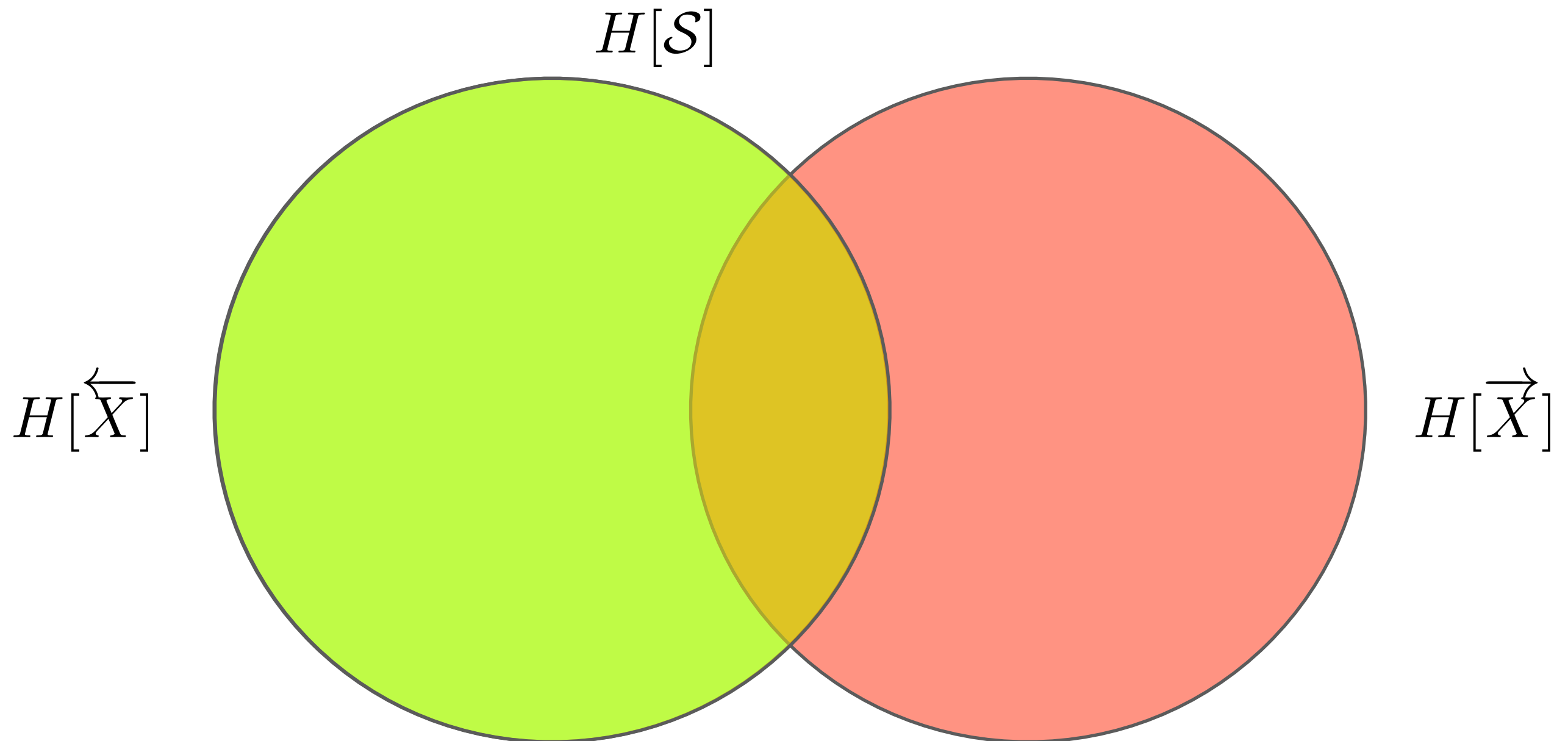
Information Diagrams for Processes

Process I-diagram using ε -machine ...



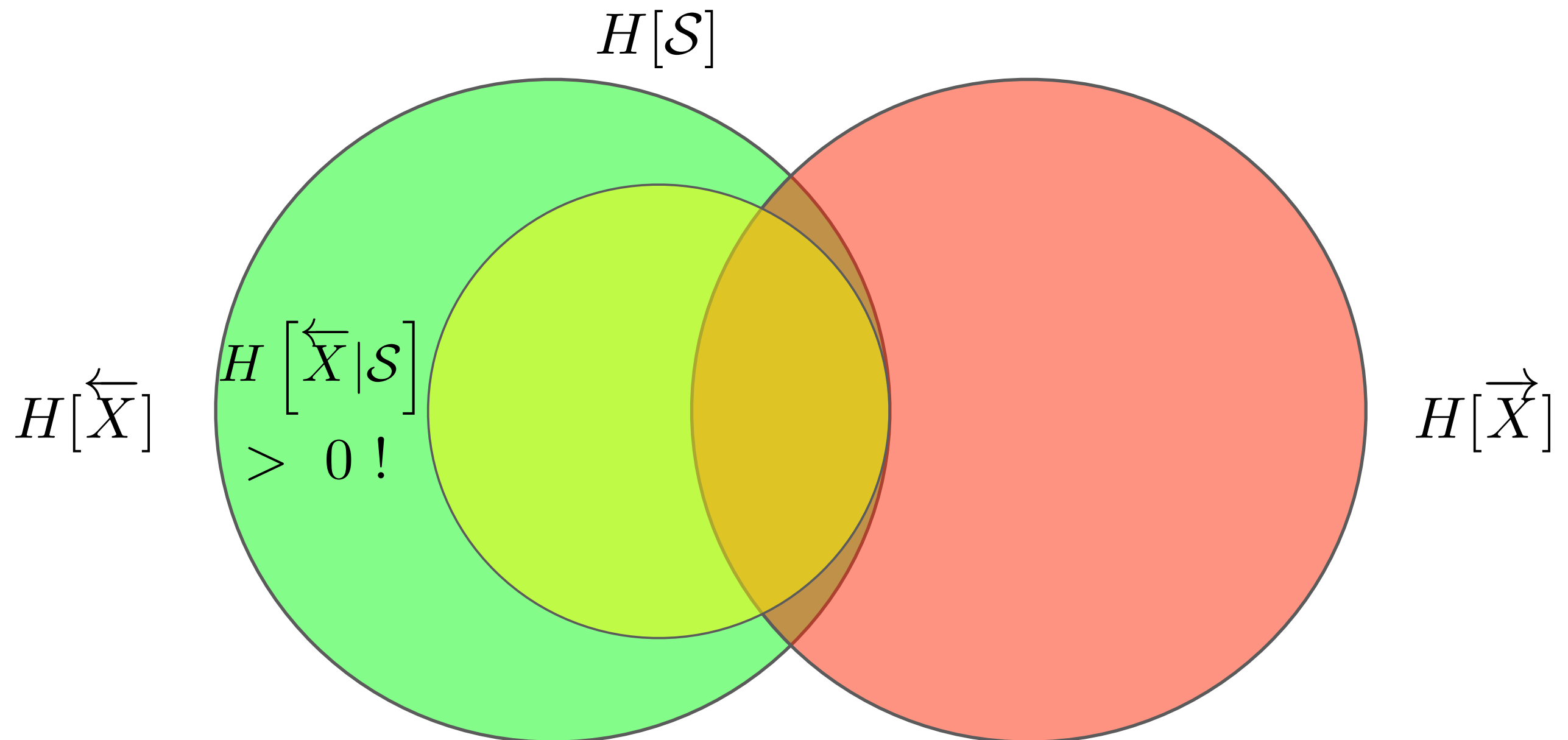
Information Diagrams for Processes

Process I-diagram using ε -machine ...



Information Diagrams for Processes

Process I-diagram using ε -machine ...

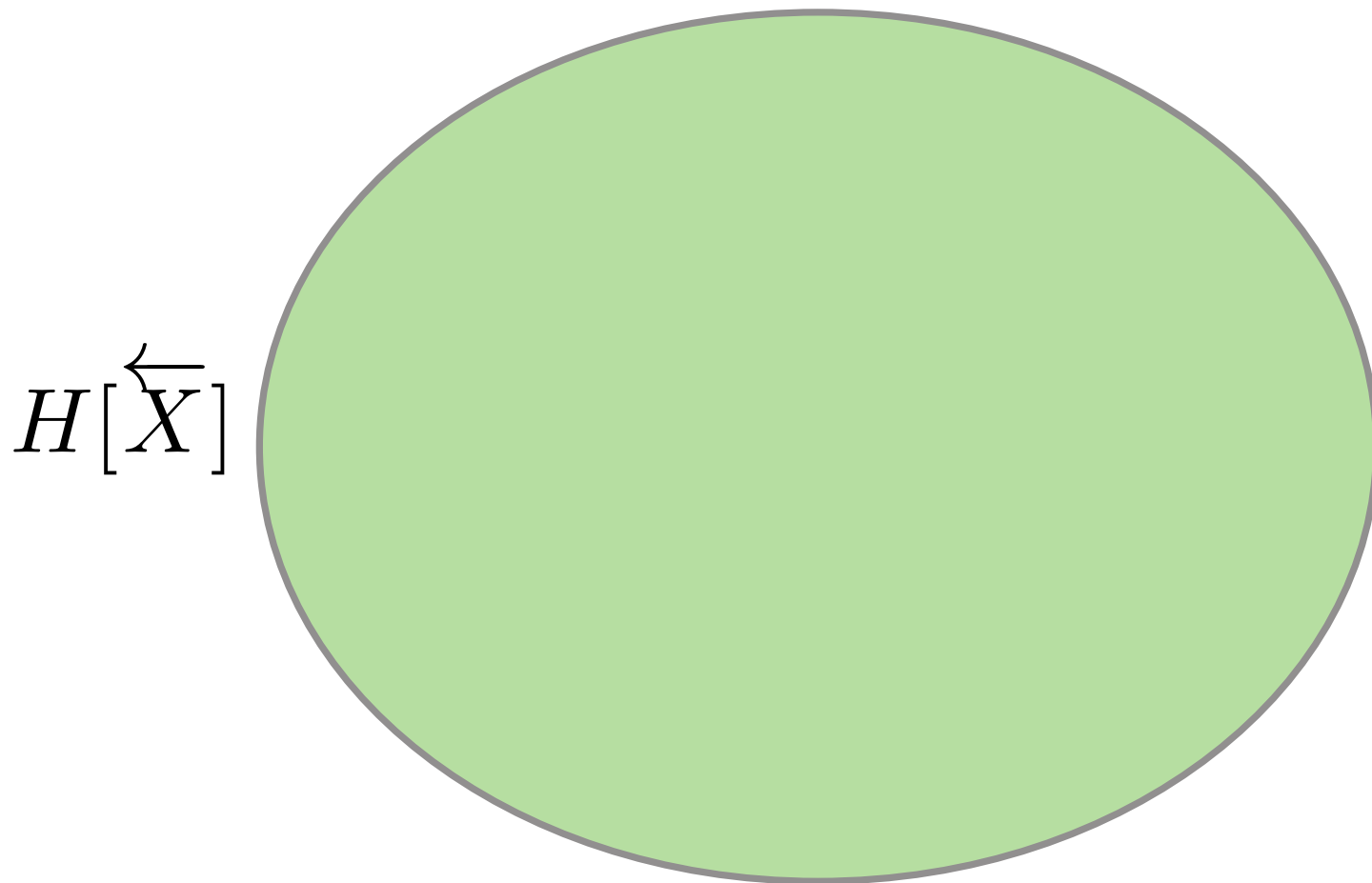


Information Diagrams for Processes

ε -Machine I-diagram:

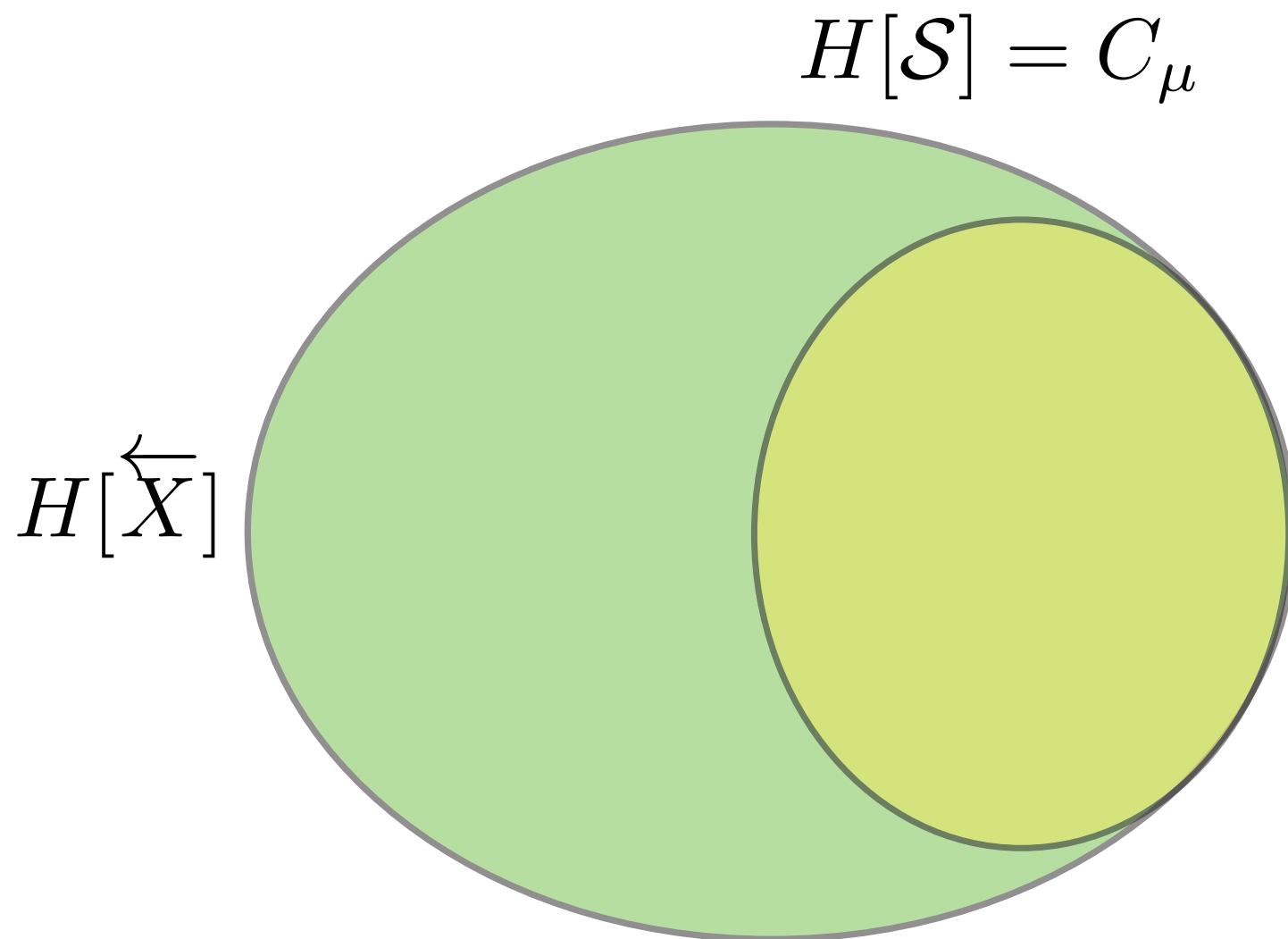
Information Diagrams for Processes

ε -Machine I-diagram:



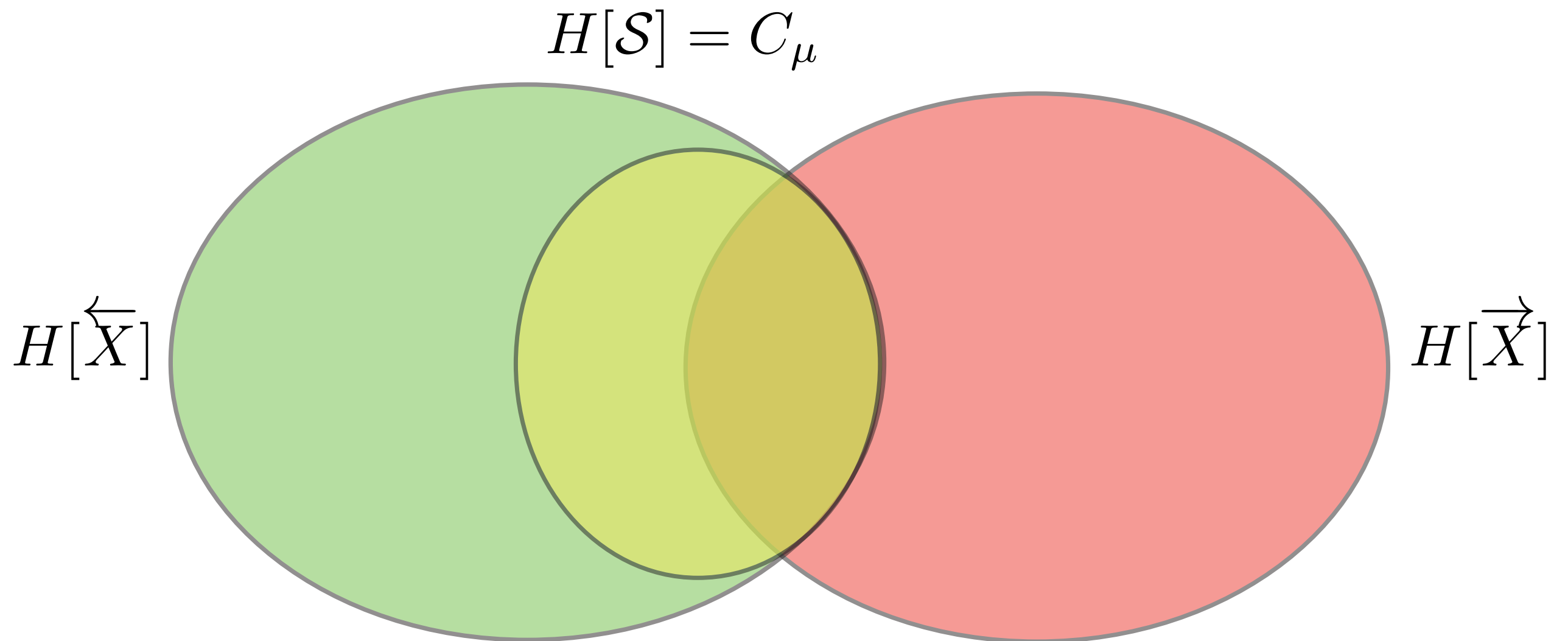
Information Diagrams for Processes

ε -Machine I-diagram:



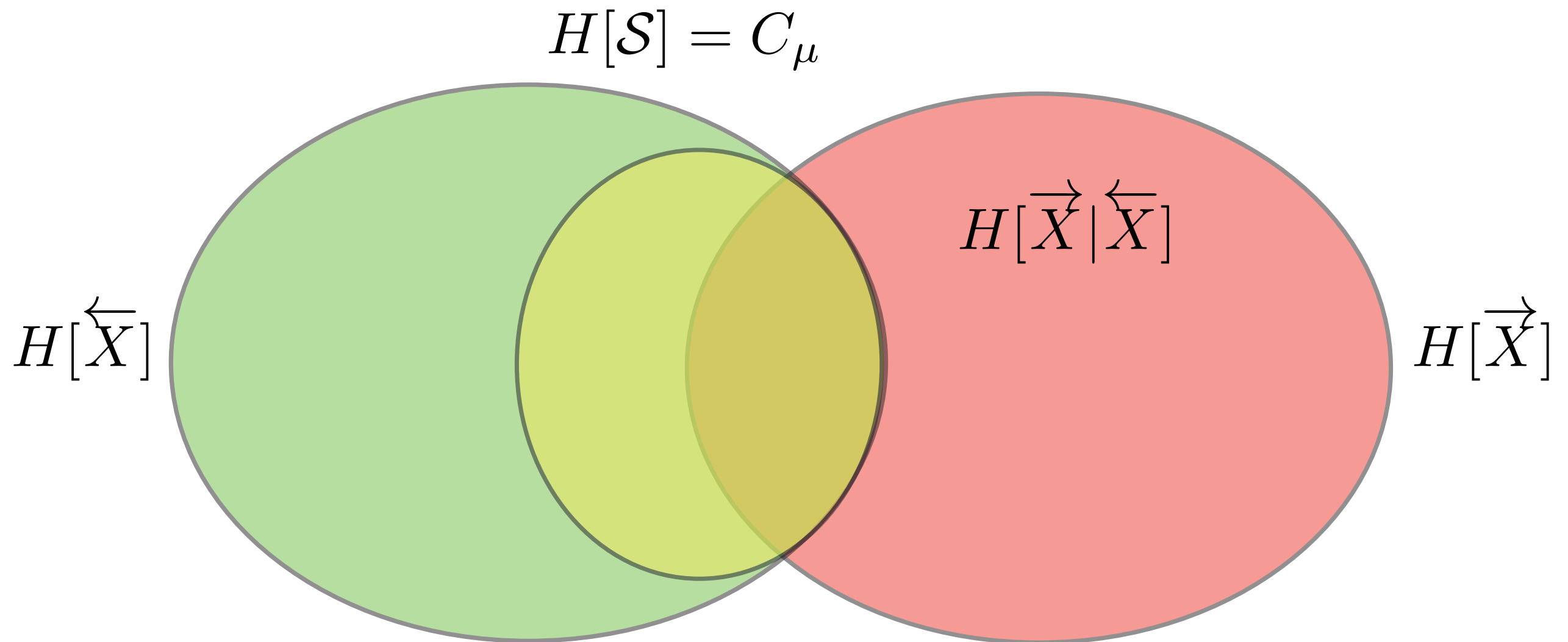
Information Diagrams for Processes

ε -Machine I-diagram:



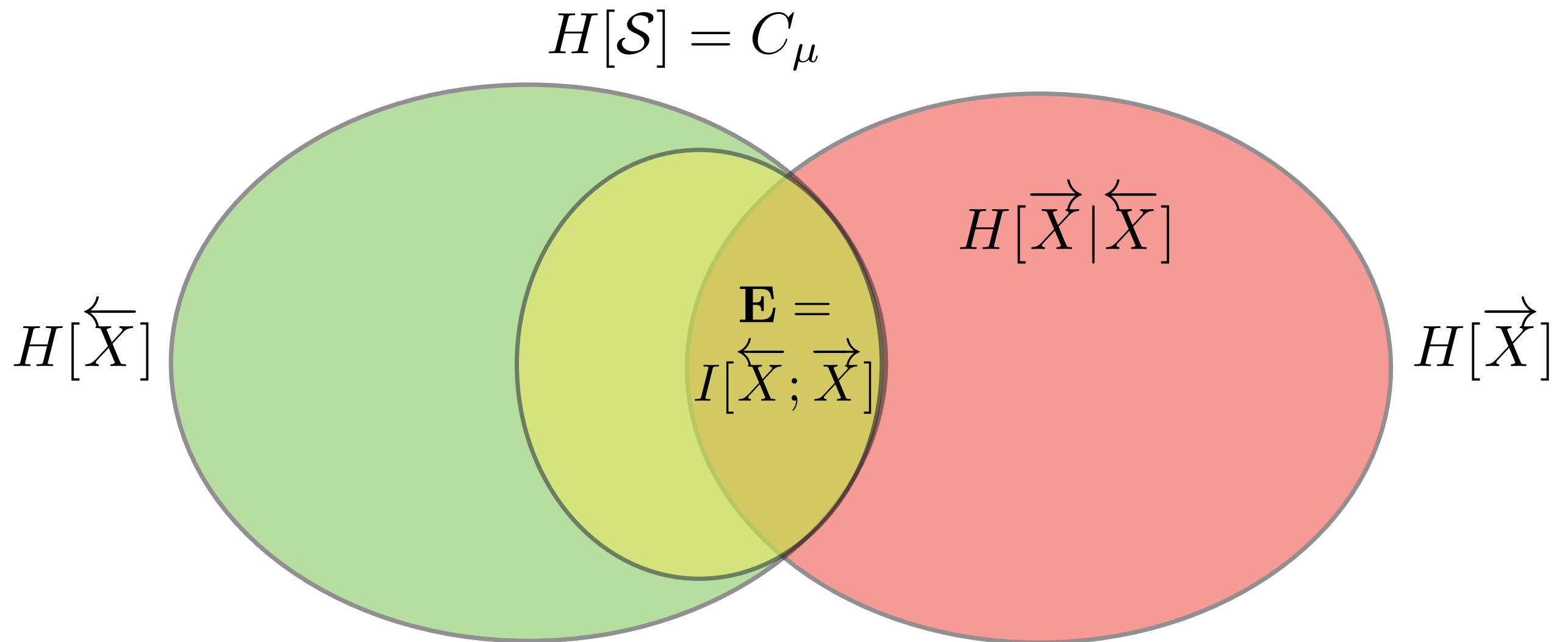
Information Diagrams for Processes

ε -Machine I-diagram:



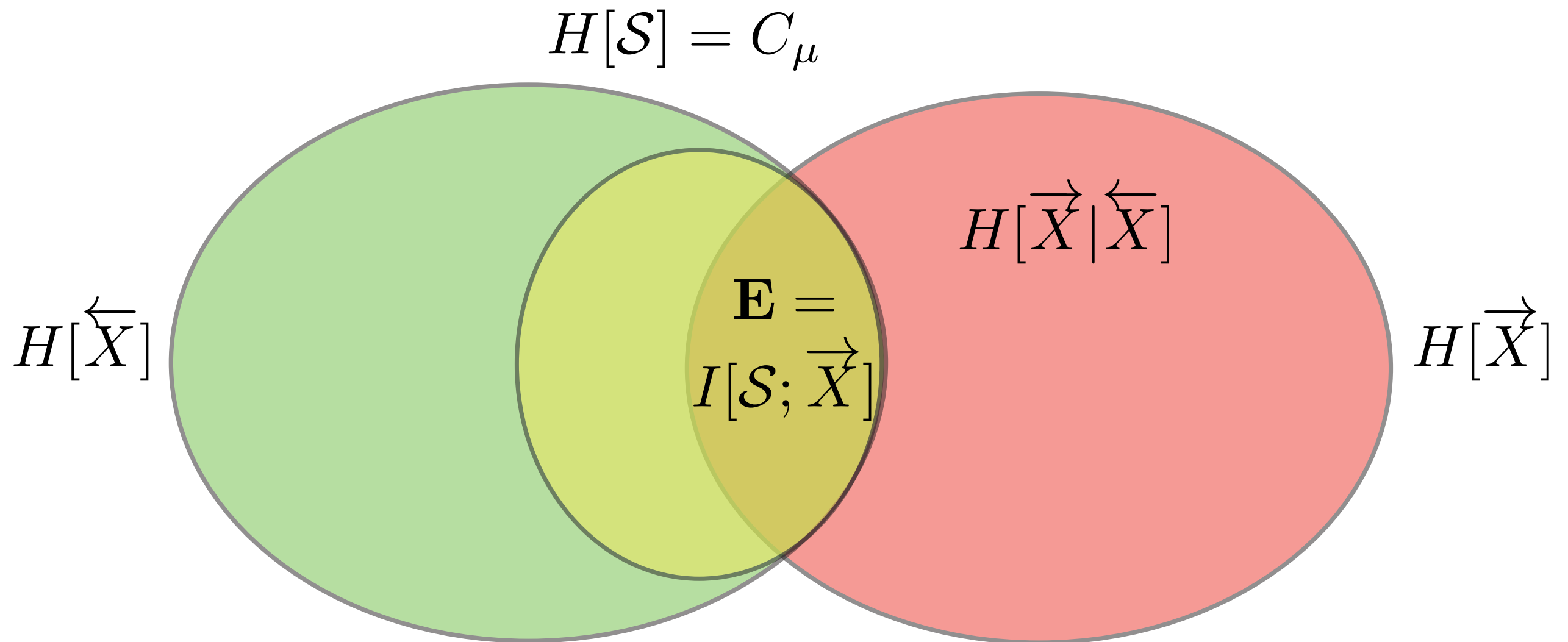
Information Diagrams for Processes

ε -Machine I-diagram:



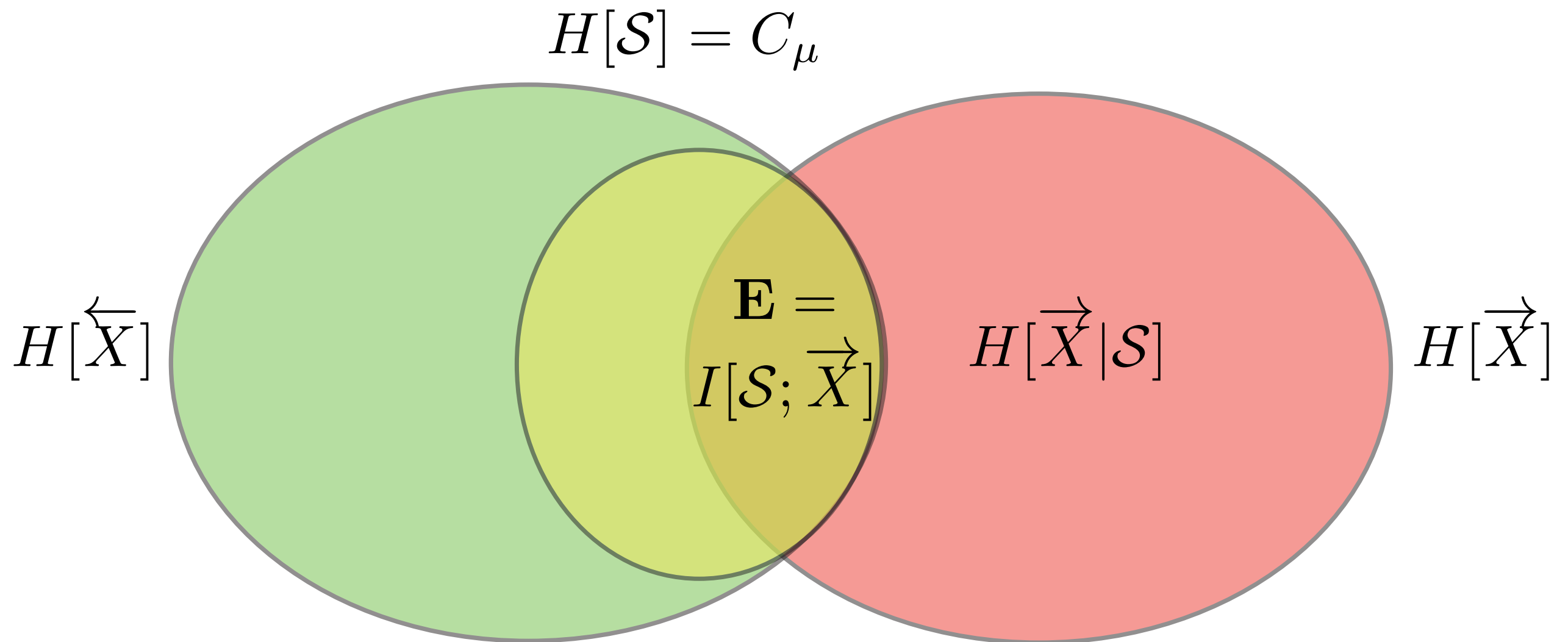
Information Diagrams for Processes

ε -Machine I-diagram:



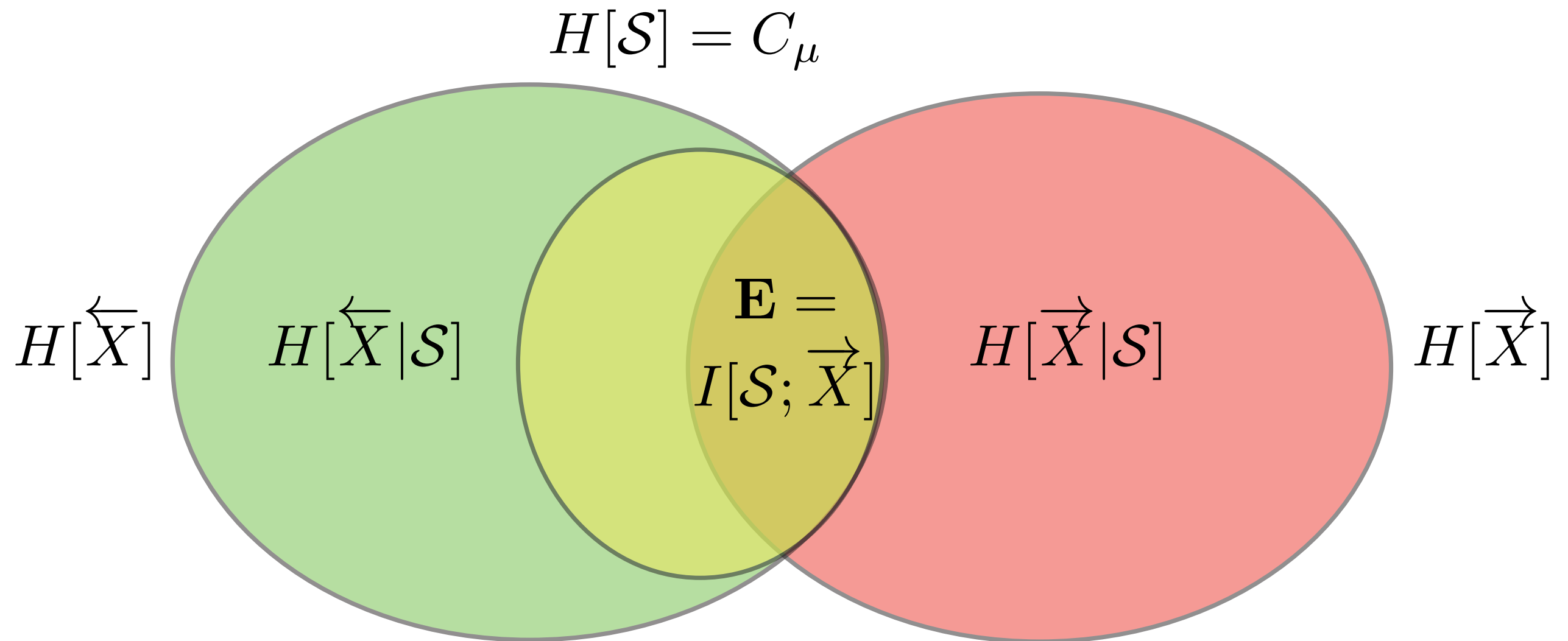
Information Diagrams for Processes

ε -Machine I-diagram:



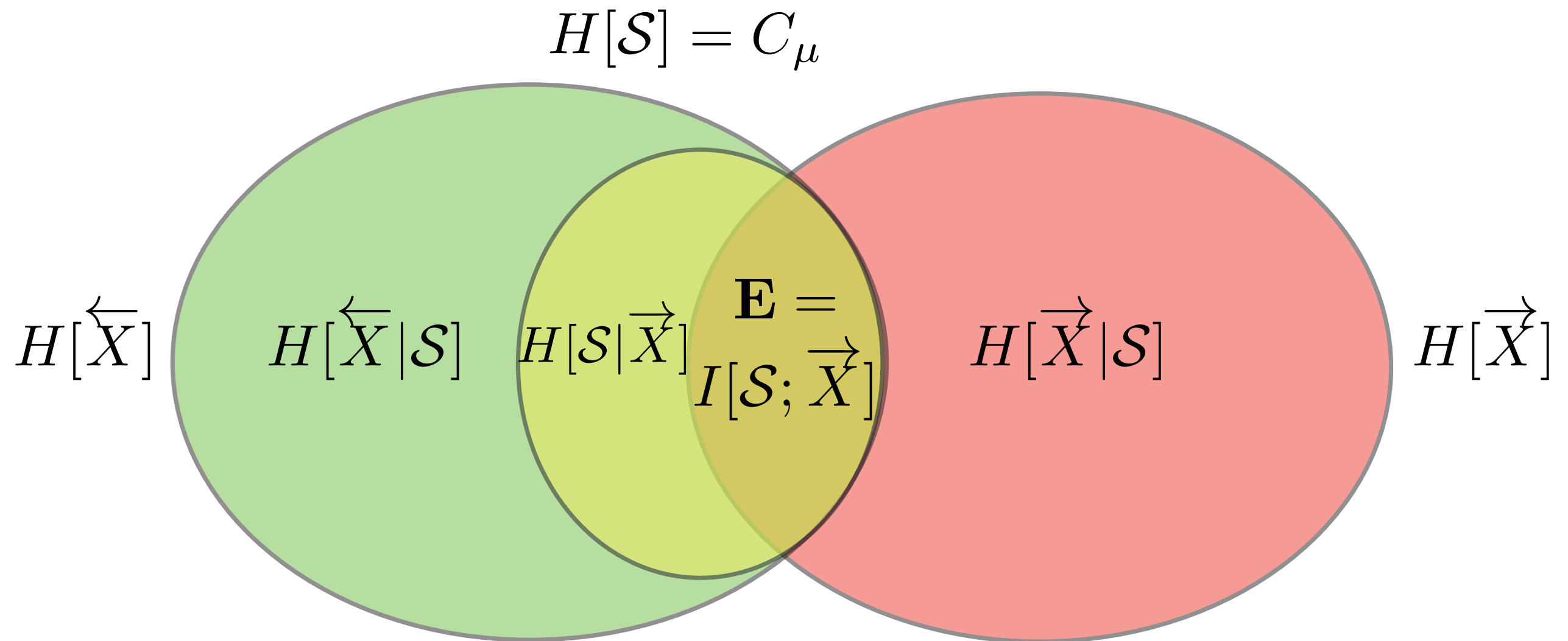
Information Diagrams for Processes

ε -Machine I-diagram:



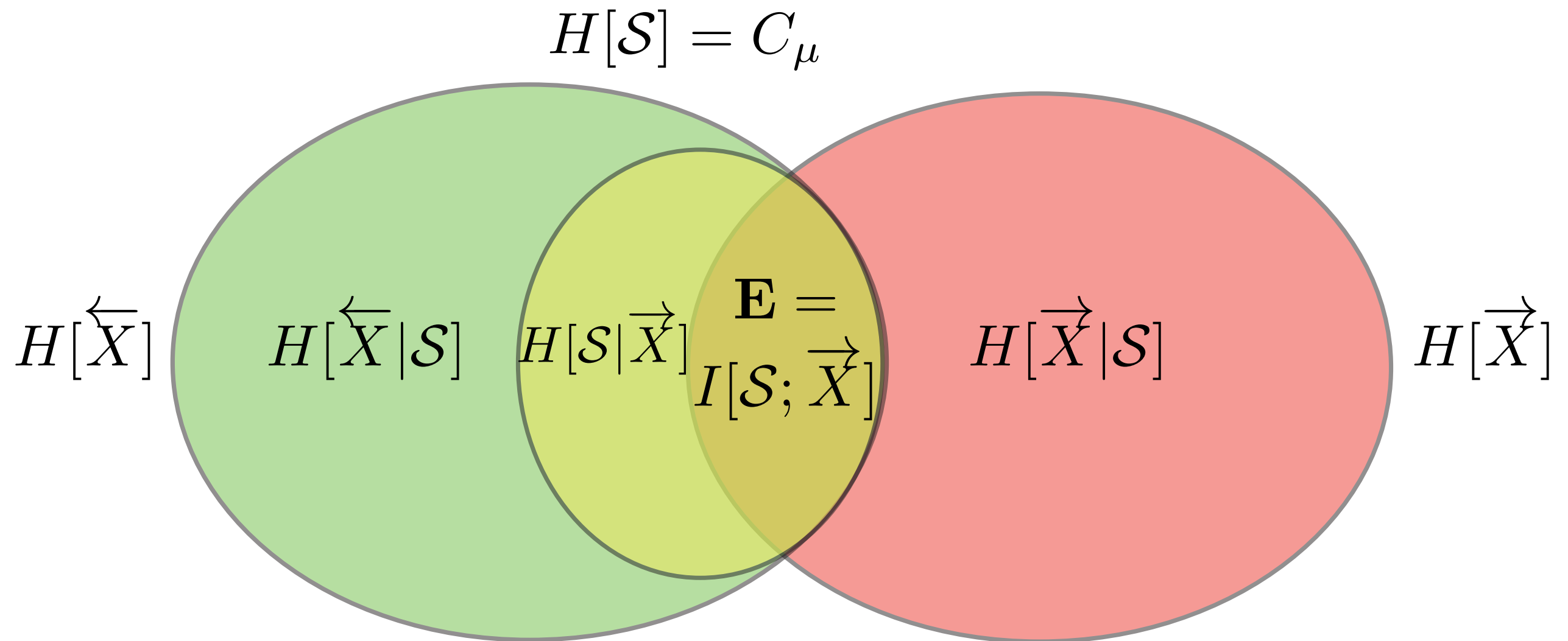
Information Diagrams for Processes

ε -Machine I-diagram:



Information Diagrams for Processes

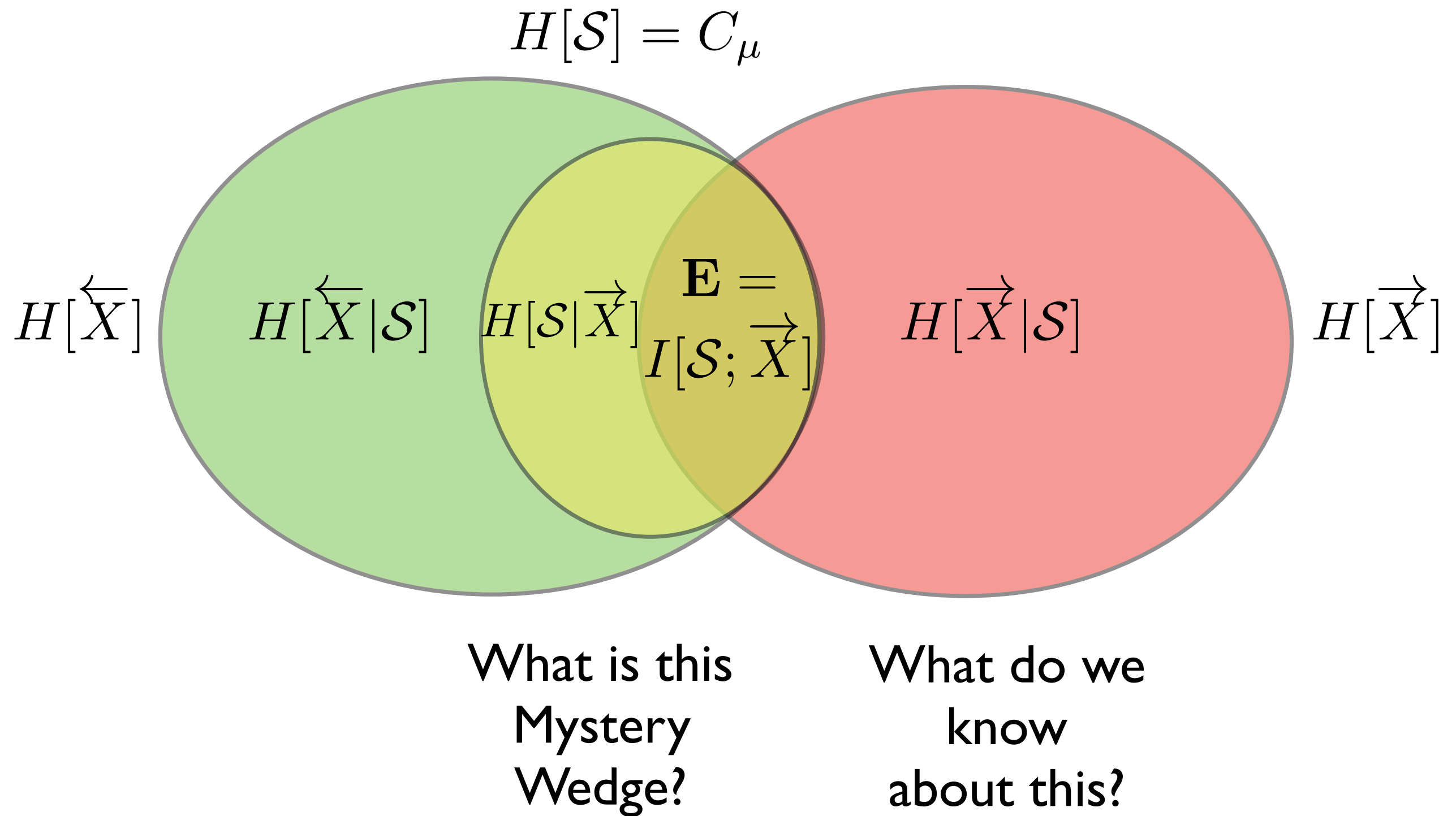
ε -Machine I-diagram:



What do we
know
about this?

Information Diagrams for Processes

ε -Machine I-diagram:



Information Diagrams for Processes

What is $H[\vec{X}|\mathcal{S}]$?

Unpredictability: $H[\vec{X}^L|\mathcal{S}] = Lh_\mu$

Proof Sketch:

$$\begin{aligned} H[\vec{X}^L|\mathcal{S}] &= H[\vec{X}^L|\overleftarrow{X}] \\ &= H[X_0X_1 \dots X_{L-1}|\overleftarrow{X}] \\ &= H[X_1 \dots X_{L-1}|\overleftarrow{X}X_0] + H[X_0|\overleftarrow{X}] \\ &= H[X_1 \dots X_{L-1}|\overleftarrow{X}] + H[X_0|\overleftarrow{X}] \\ &\vdots \\ &= H[X_{L-1}|\overleftarrow{X}] + \dots + H[X_1|\overleftarrow{X}] + H[X_0|\overleftarrow{X}] \\ &= LH[X_0|\overleftarrow{X}] \\ &= Lh_\mu \end{aligned}$$

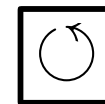
Information Diagrams for Processes

What is Mystery Wedge? $H[\mathcal{S}|\vec{X}]$

Uncertainty of causal state given future. Implications?

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

$$\begin{aligned}\text{Proof sketch: } \mathbf{E} &= I[\overleftarrow{X}; \vec{X}] \\ &= H[\vec{X}] - H[\vec{X}|\overleftarrow{X}] \\ &= H[\vec{X}] - H[\vec{X}|\mathcal{S}] \\ &= I[\vec{X}; \mathcal{S}] \\ &= H[\mathcal{S}] - H[\mathcal{S}|\vec{X}] \\ &\leq H[\mathcal{S}] \\ &= C_\mu\end{aligned}$$



Information Diagrams for Processes

What is Mystery Wedge? $H[S|\vec{X}]$

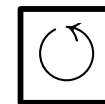
Uncertainty of causal state given future. Implications?

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

Proof sketch: $\mathbf{E} = I[\overleftarrow{X}; \vec{X}]$

$$\begin{aligned} &= H[\vec{X}] - H[\vec{X}|\overleftarrow{X}] \\ &= H[\vec{X}] - H[\vec{X}|S] \\ &= I[\vec{X}; S] \\ &= H[S] - H[S|\vec{X}] \\ &\leq H[S] \\ &= C_\mu \end{aligned}$$

I am the
Mystery Wedge!



Information Diagrams for Processes

What is Mystery Wedge? $H[\mathcal{S}|\vec{X}]$

Wedge is the inaccessibility of hidden state information!

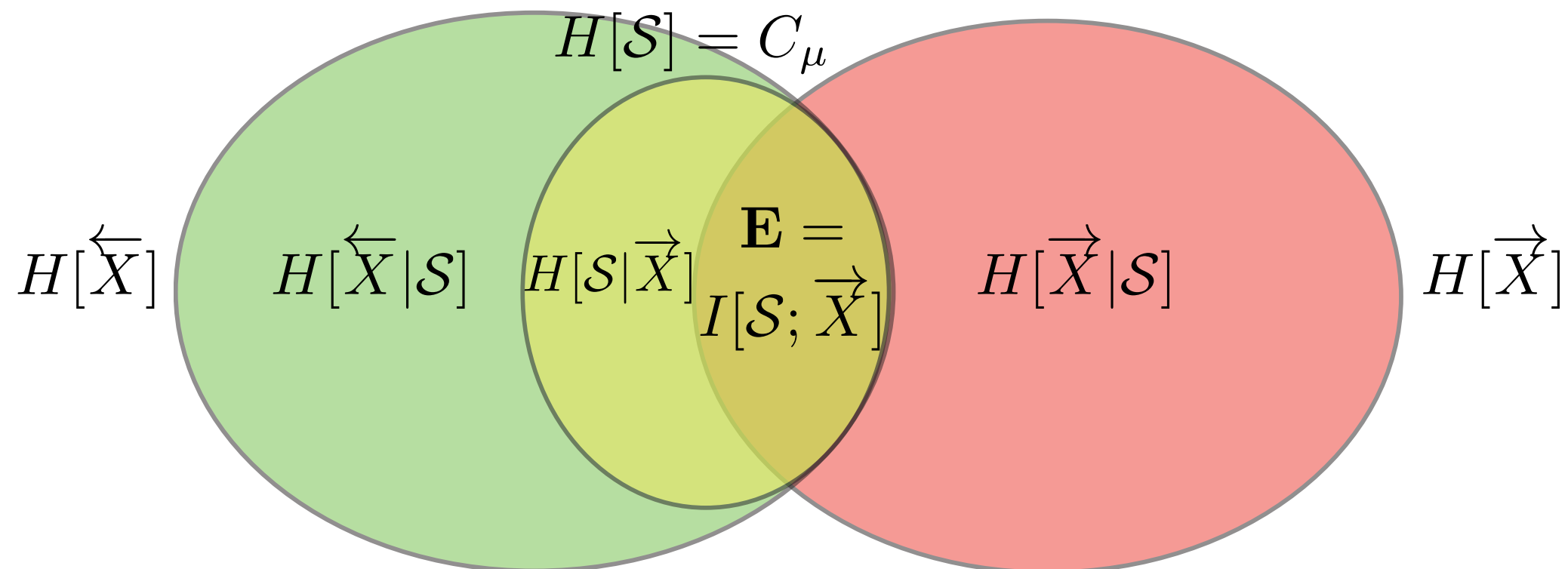
$$H[\mathcal{S}|\vec{X}] = C_\mu - \mathbf{E}$$

Wedge controls Internal - Observed

The **Process Crypticity**:

$$\chi = C_\mu - \mathbf{E}$$

Controls how much internal state information is observable.



How to get $\mathbf{E}$ from ϵM ?

DIRECTIONAL COMPUTATIONAL MECHANICS

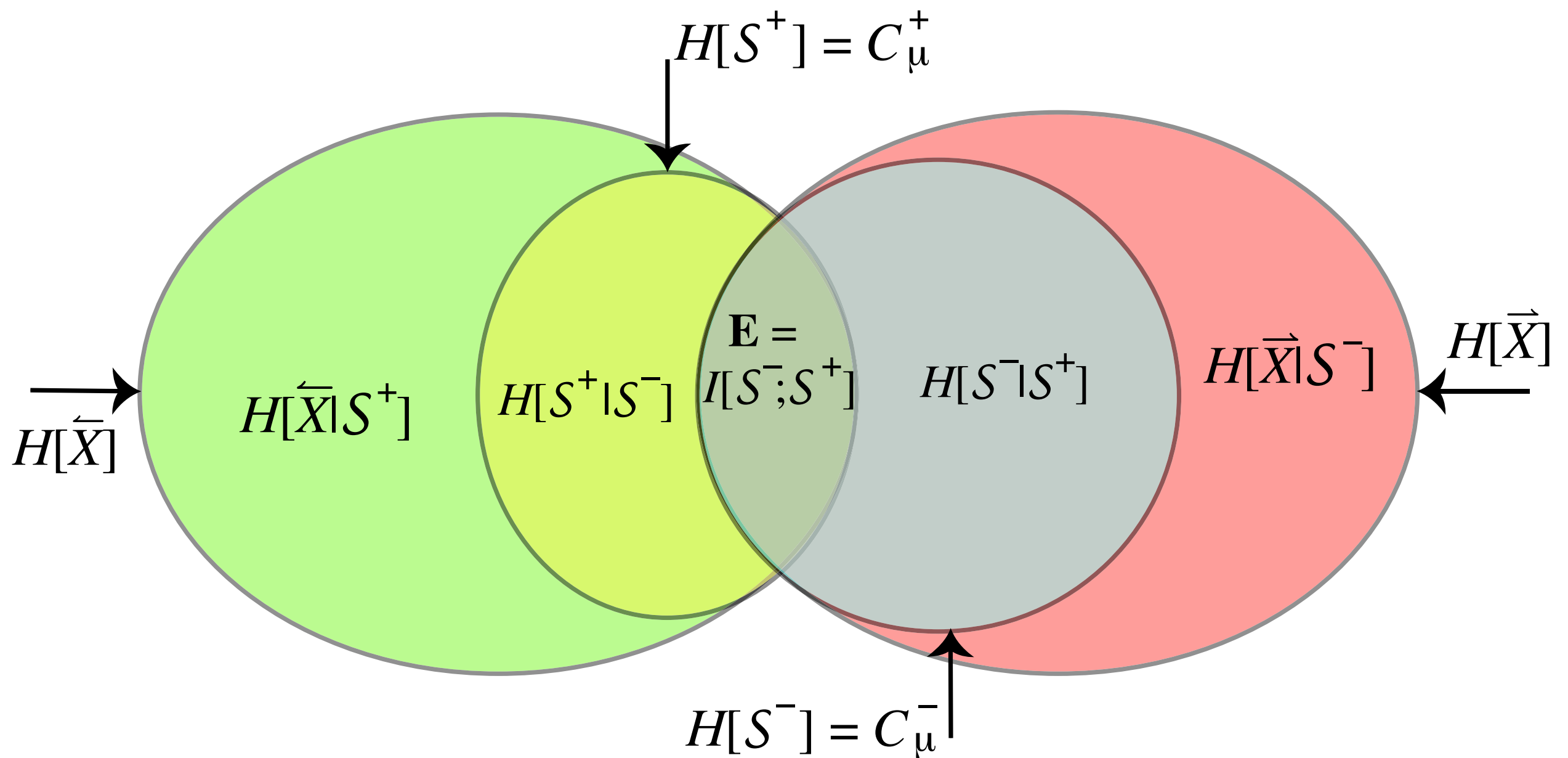
- Theorem:

$$\mathbf{E} = I[\mathcal{S}^+; \mathcal{S}^-]$$

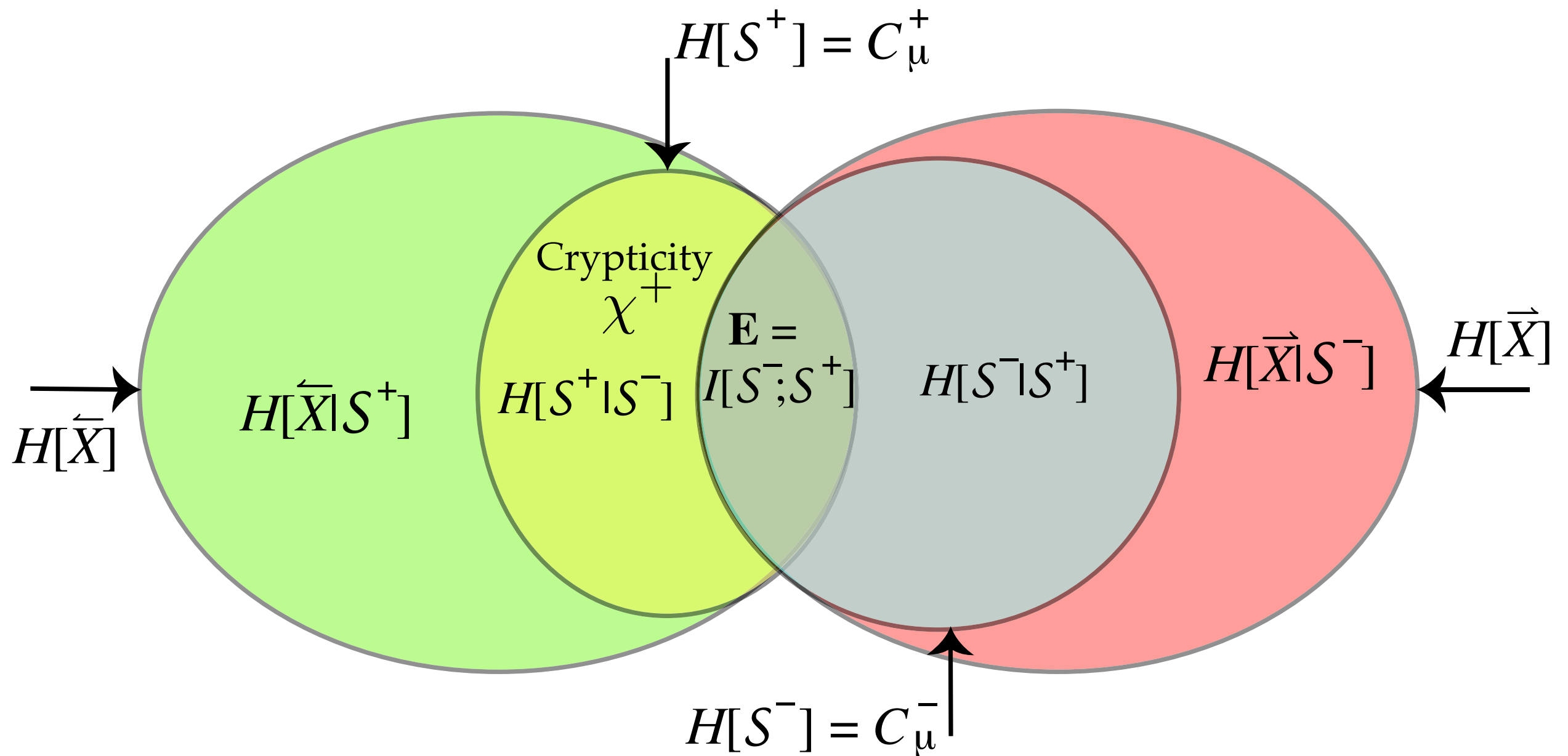
- Effective transmission capacity of channel between forward and reverse processes.
- Time agnostic representation: The **BiMachine**.

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.

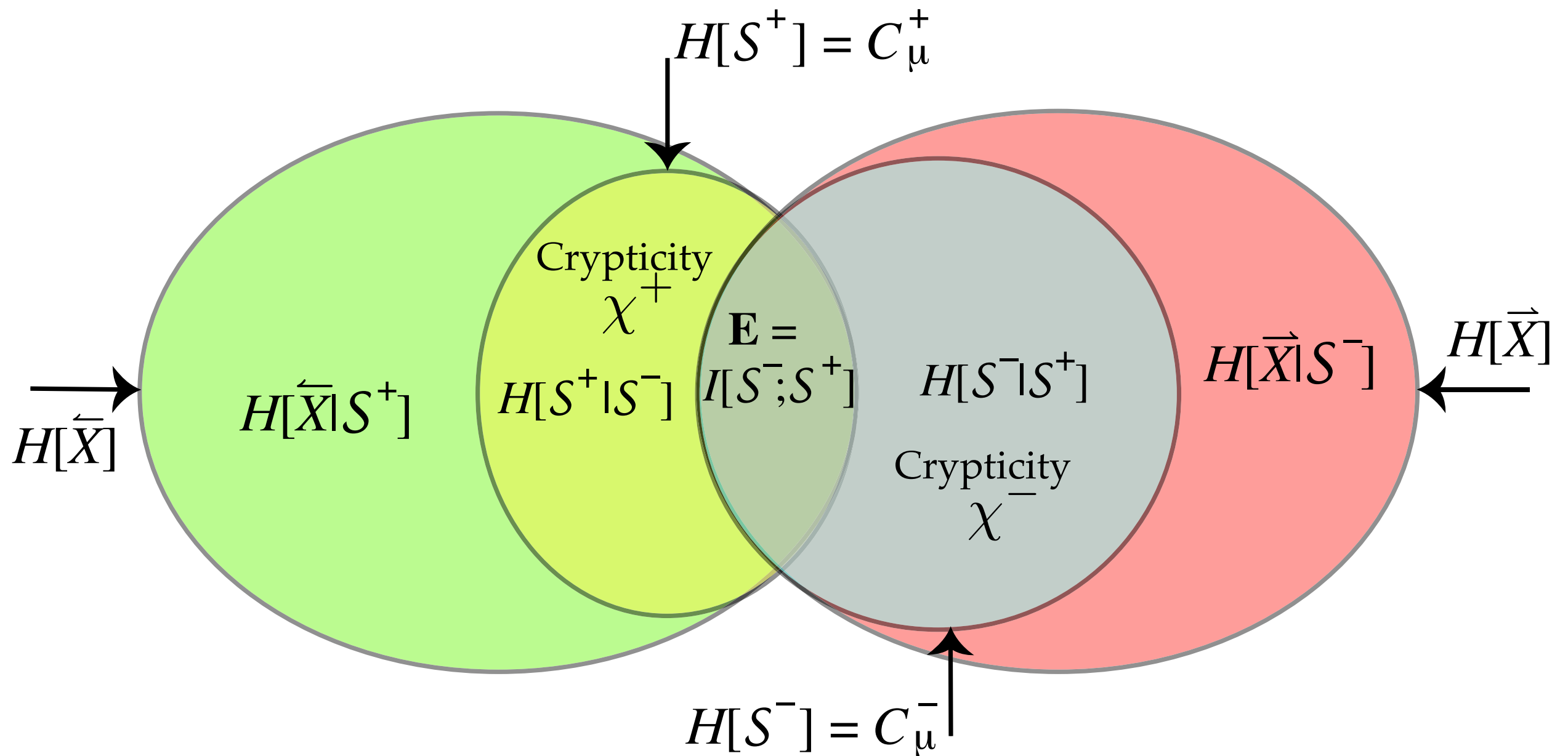
Forward-Reverse ε -Machine Information Diagram



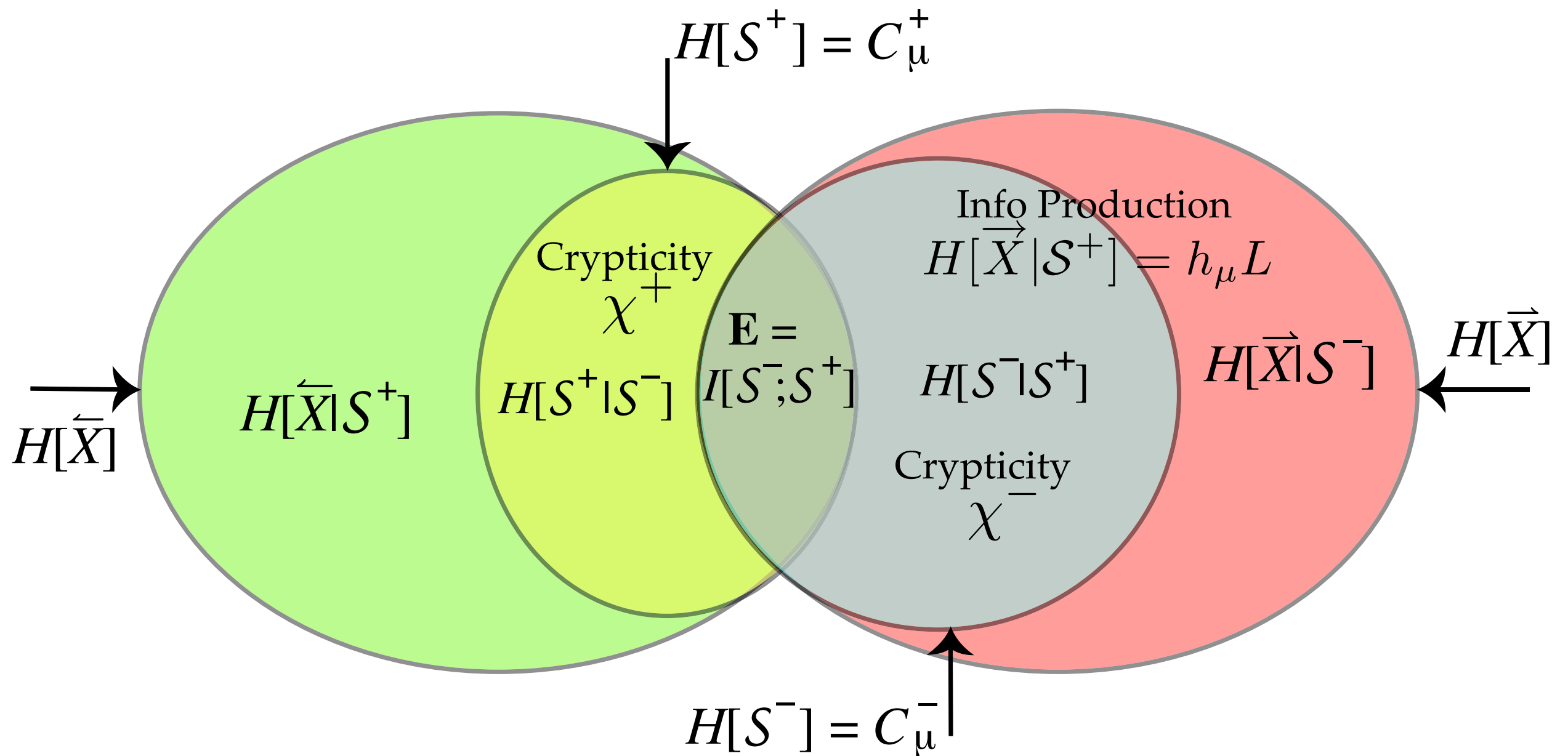
Forward-Reverse ε -Machine Information Diagram



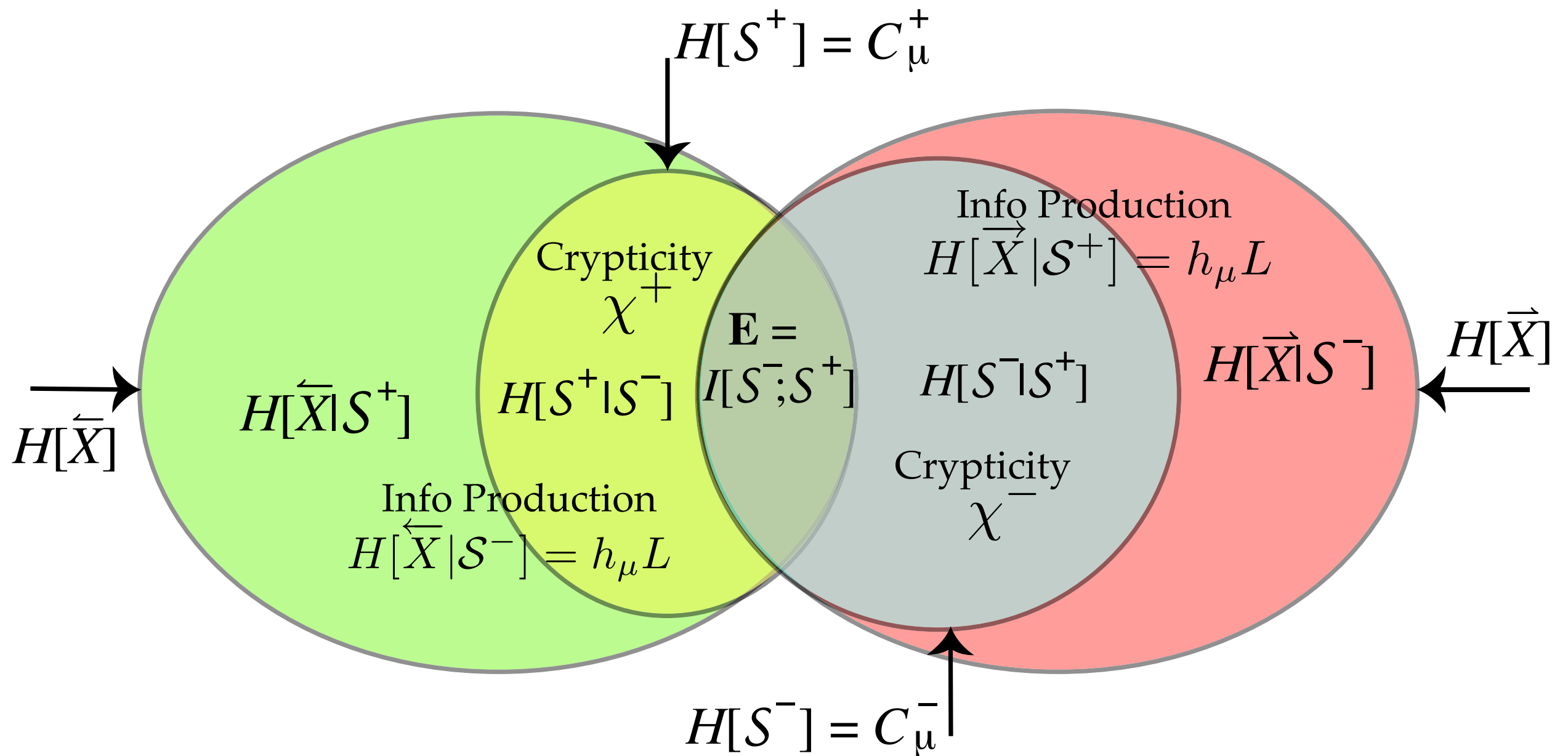
Forward-Reverse ε -Machine Information Diagram



Forward-Reverse ε -Machine Information Diagram

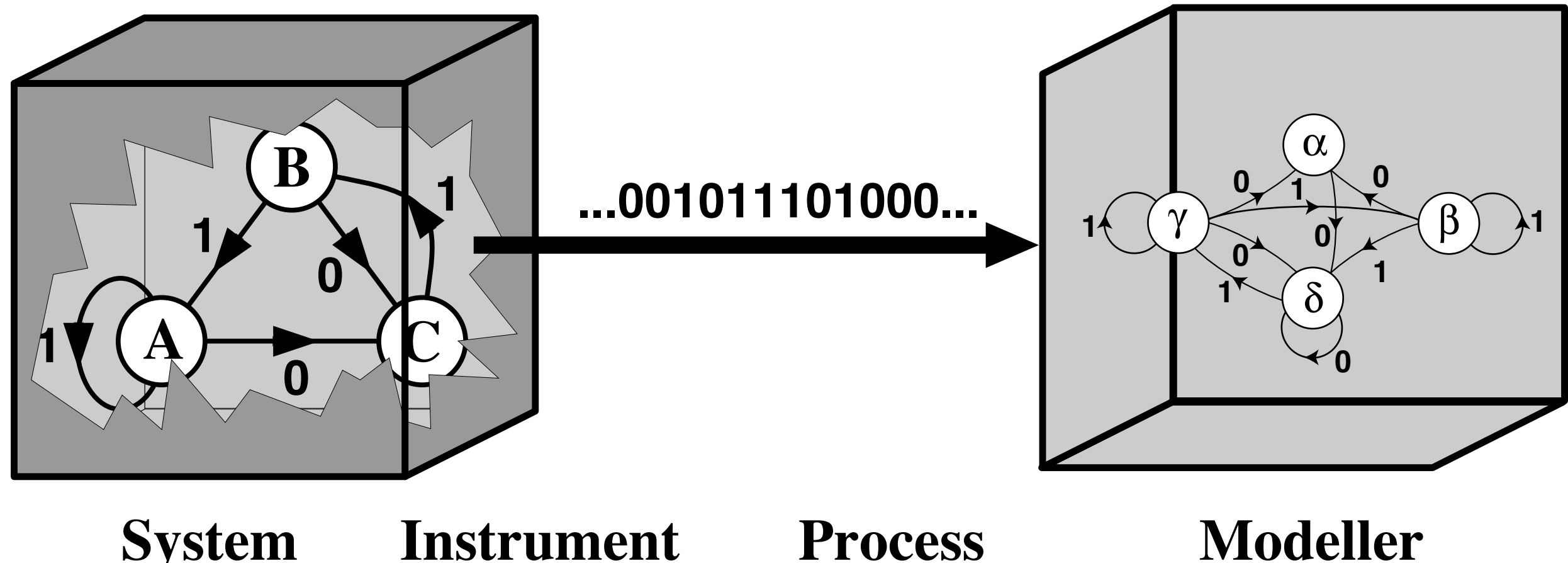


Forward-Reverse ε -Machine Information Diagram



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:
 $\max_{\{P\}} h_{\mu}$
 $\min_{\{P\}} C_{\mu}$

How random?
 $\lambda, H(L), h_{\mu}$
How structured?
 $C_{\mu}, \mathbf{E}, \mathbf{T}, \mathbf{G}, \mathcal{R}$

Universal model:
 ϵ – 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 $H(\ell)$:

K quantifies the amount of information observed as ℓ gets large, whereas C_μ quantifies how much information it takes to generate 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

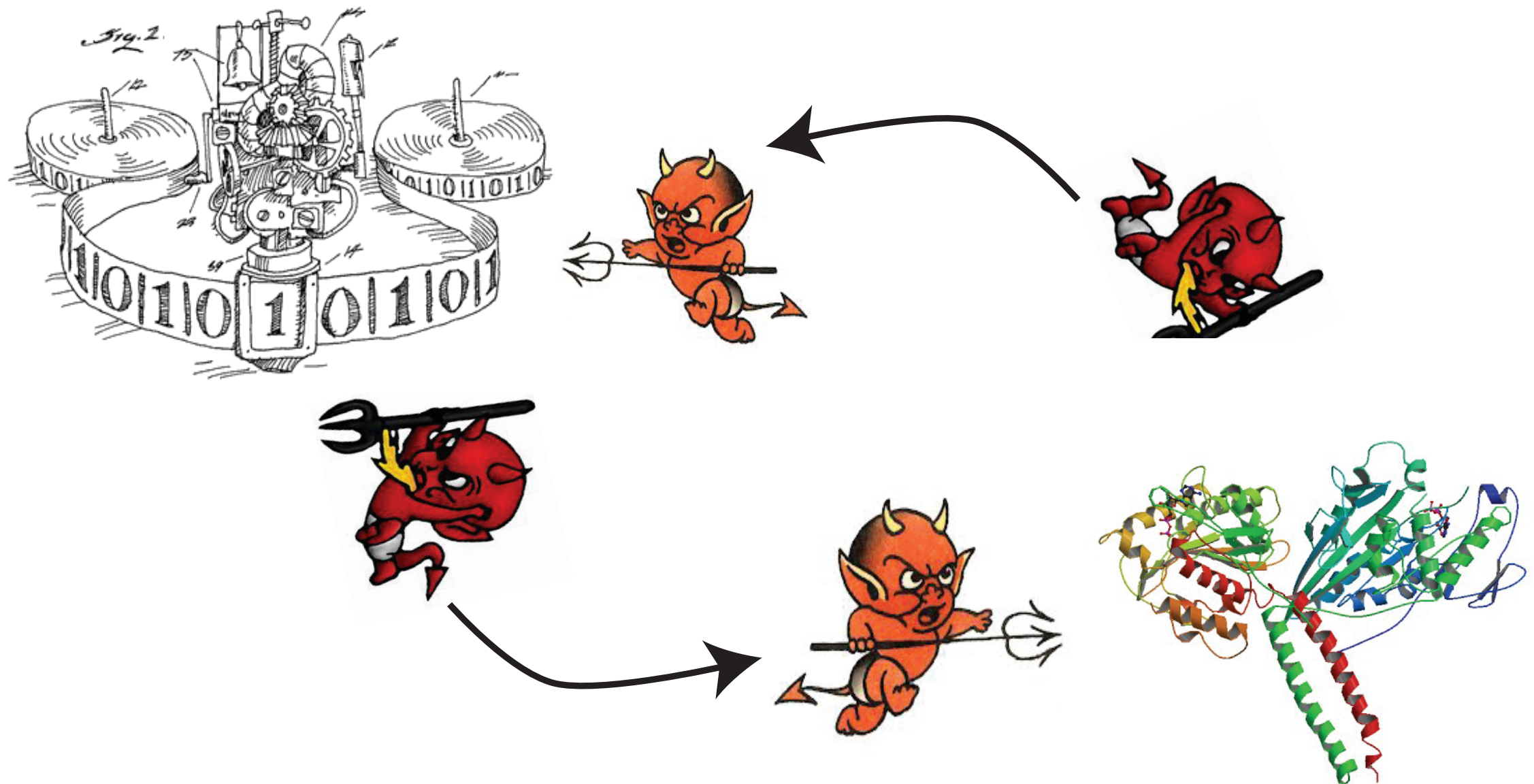
Computational Mechanics was introduced to be a calculable, quantitative version of KC Complexity Theory.

Constructive! For finite eMs, all complexity/information measures

- can be calculated in closed form.
- $O(1)$ computational complexity.

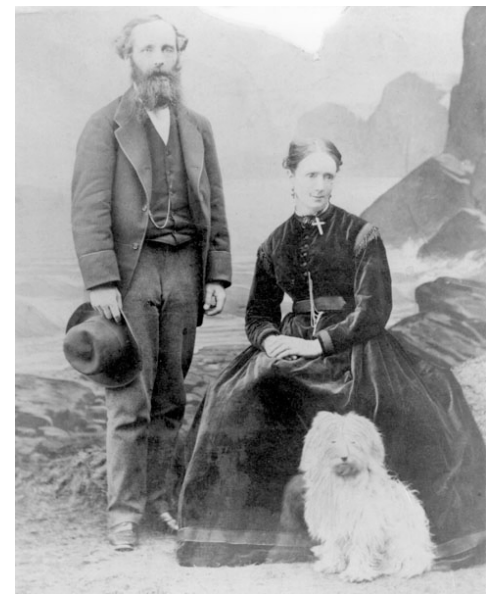
So, much computational complexity in KC Theory and in Information Theory obviated.

Thermodynamics of Adaptive Complex Systems: An Application

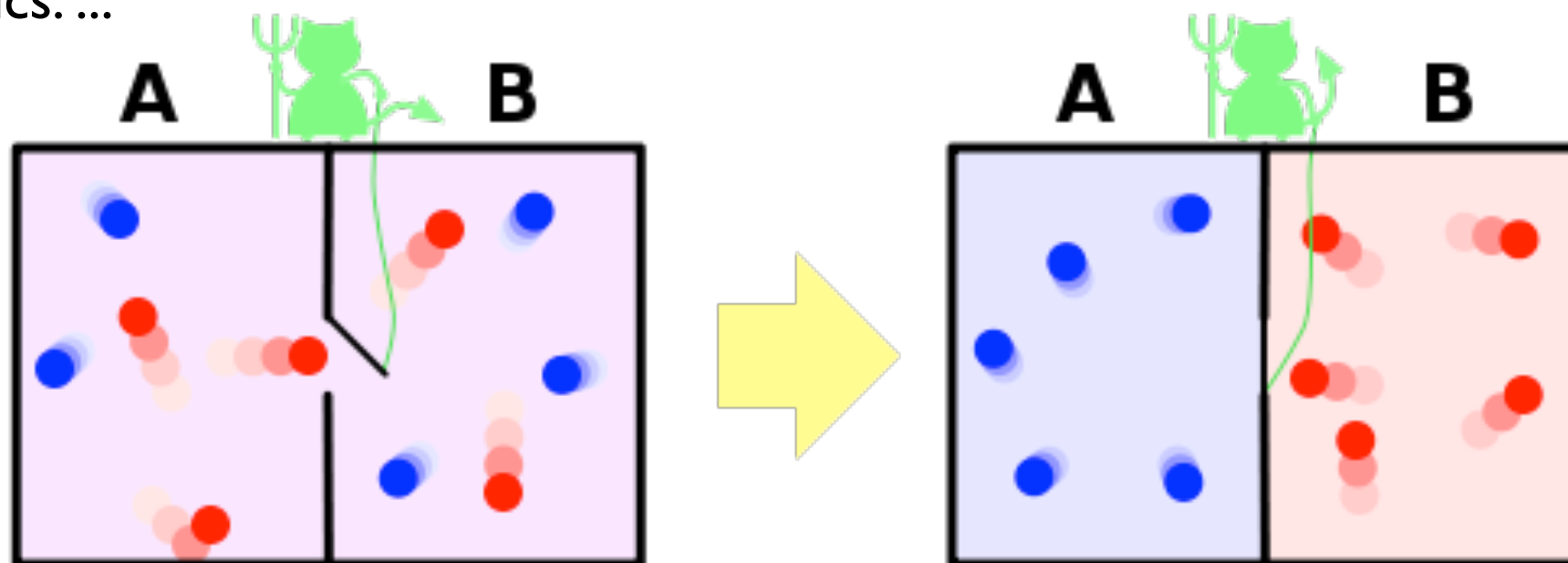


Maxwell's Demon

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.



MAXWELL'S DEMON

- Demon creates order out of chaos.



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.

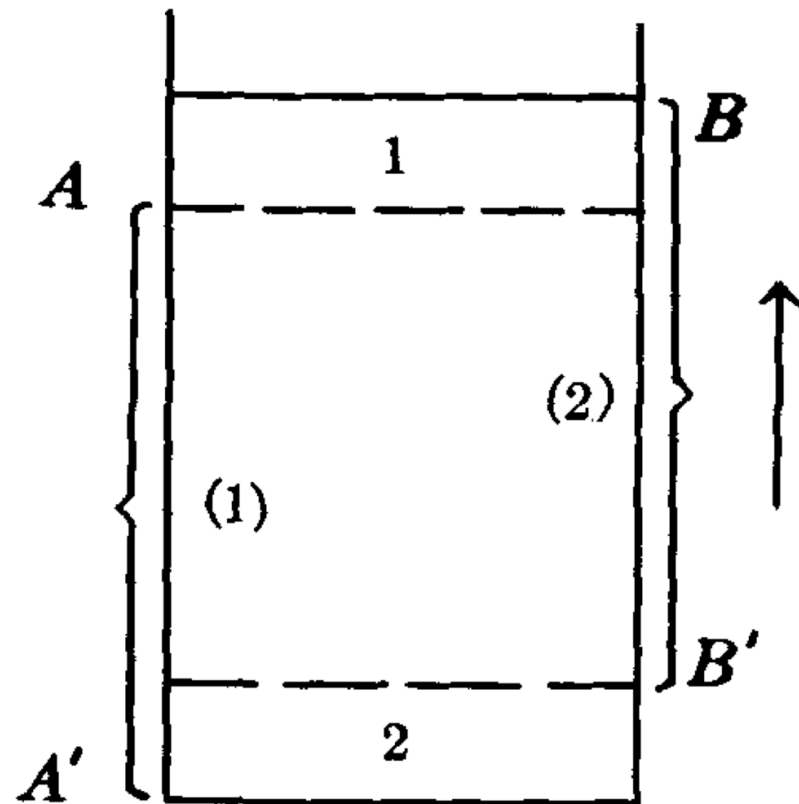


FIG. 1

... 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.

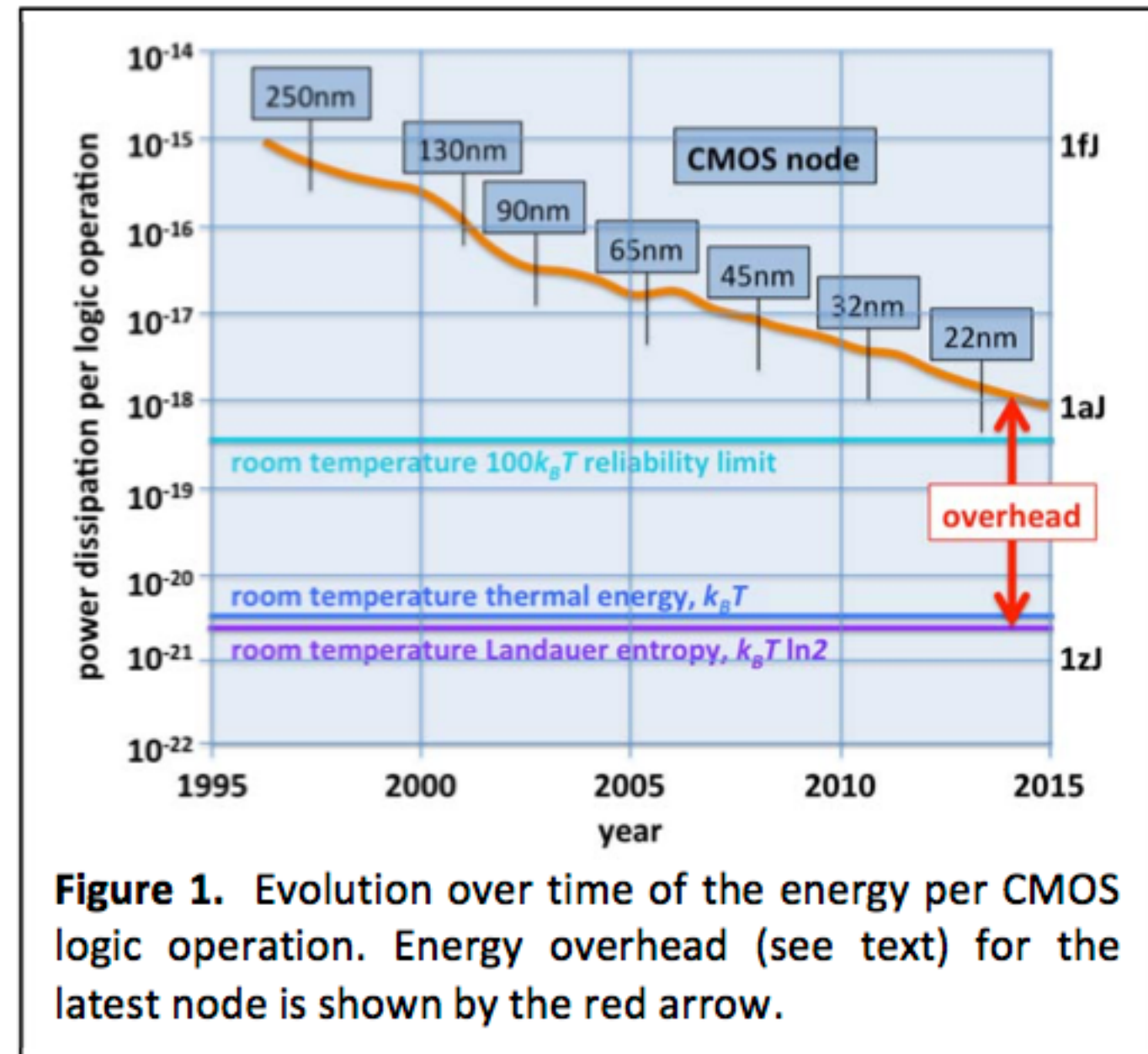
... a simple inanimate device can achieve the same essential result as would be achieved by the intervention of intelligent beings. We have examined the “biological phenomena” of a nonliving device and have seen that it generates exactly that quantity of entropy which is required by thermodynamics.

Information Engines

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

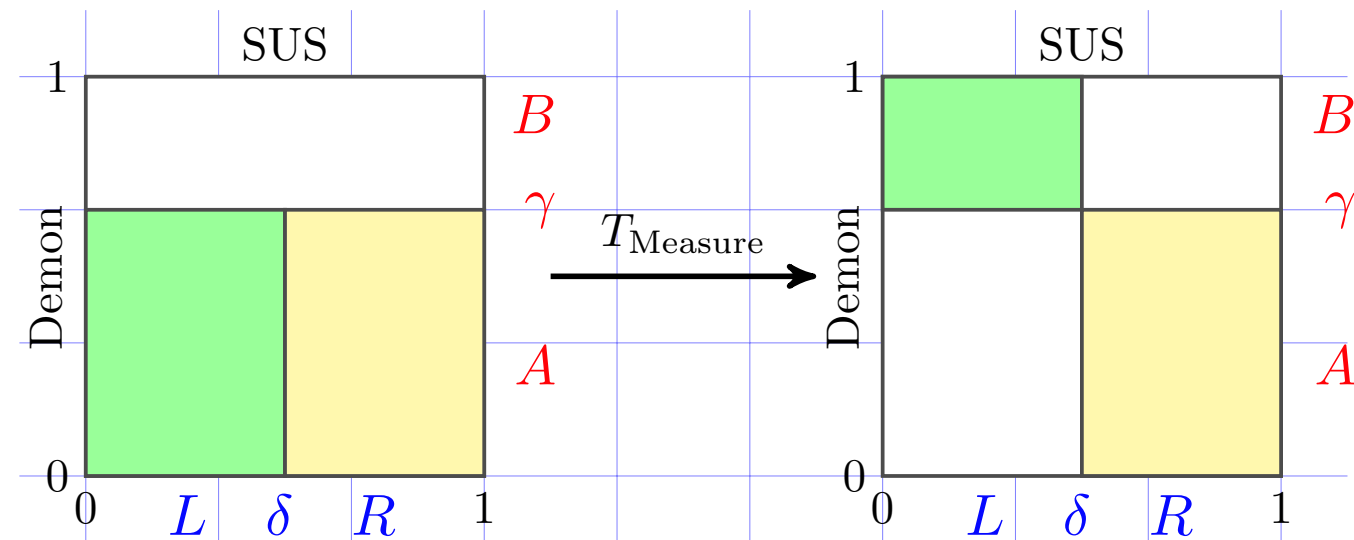
Landauer Principle: $Q_{\text{erase}} \geq k_{\text{Boltzmann}} T \ln 2$

Erasing a bit dissipates energy to the heat bath.

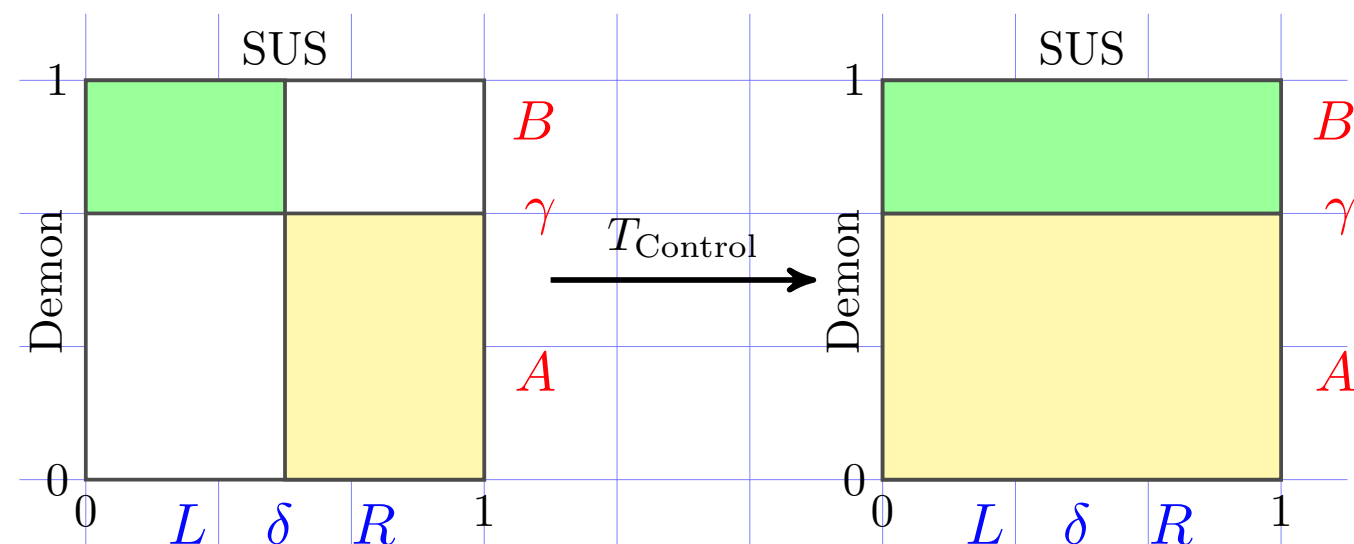


Szilard's Engine is a Chaotic Dynamical System

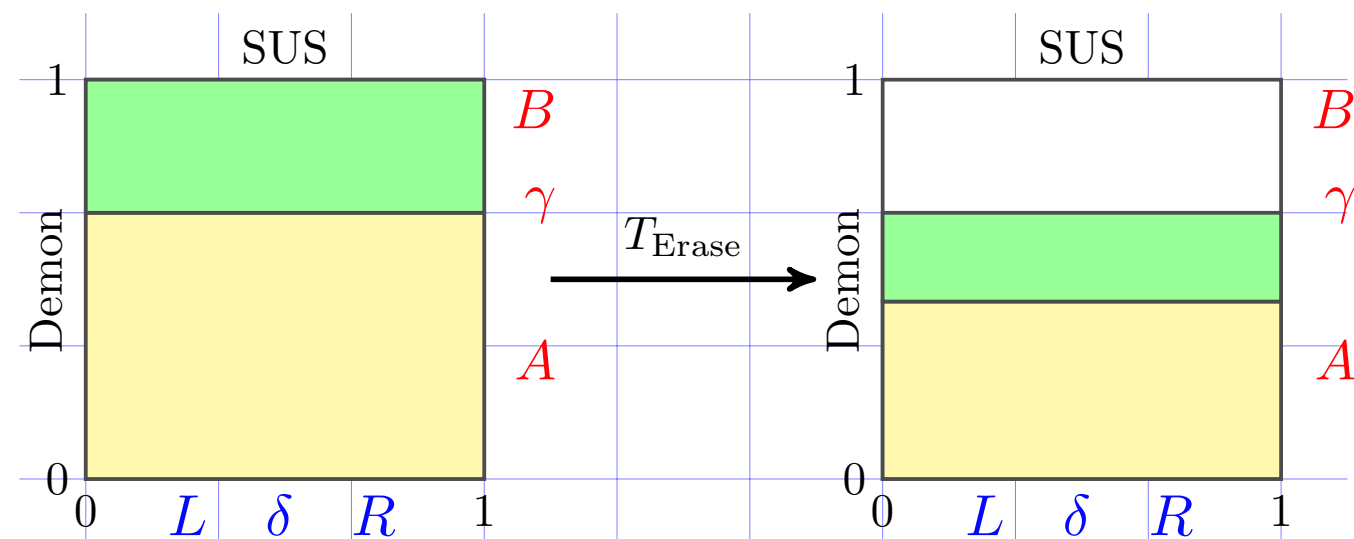
Measure



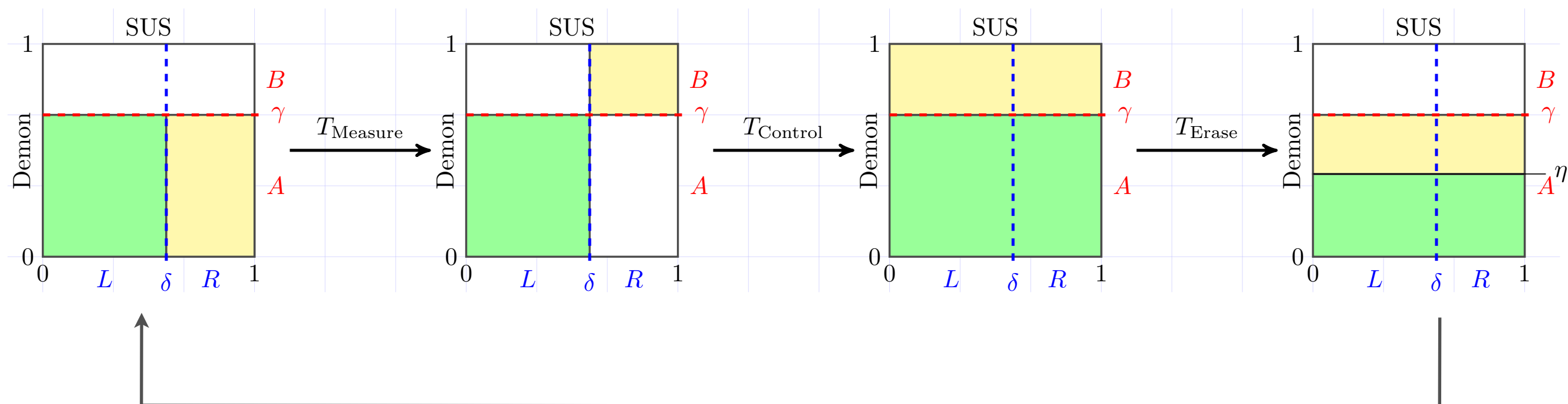
Extract Energy
from Heat Bath



Erase Demon
Memory

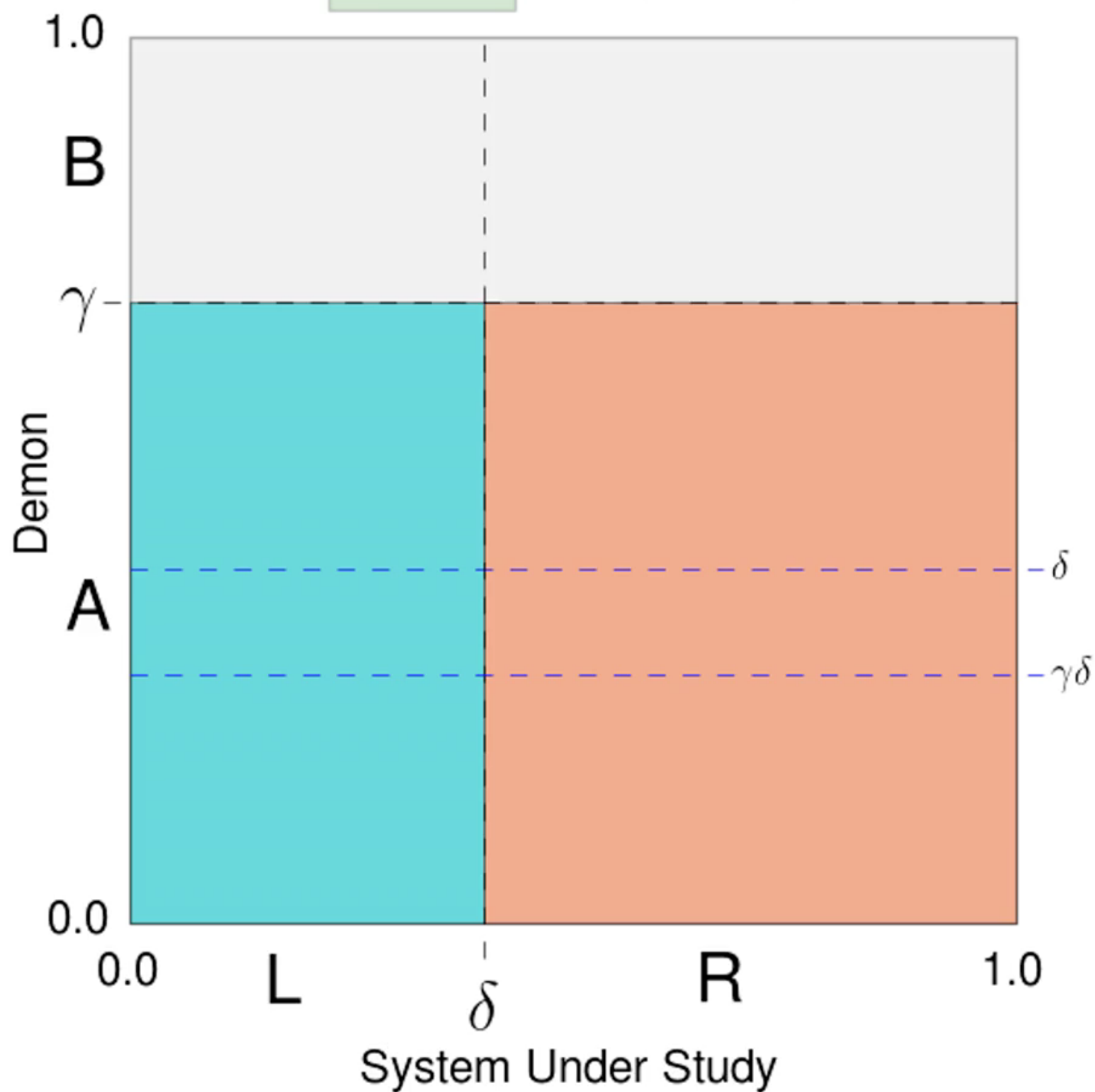


Szilard's Engine is a Chaotic Dynamical System



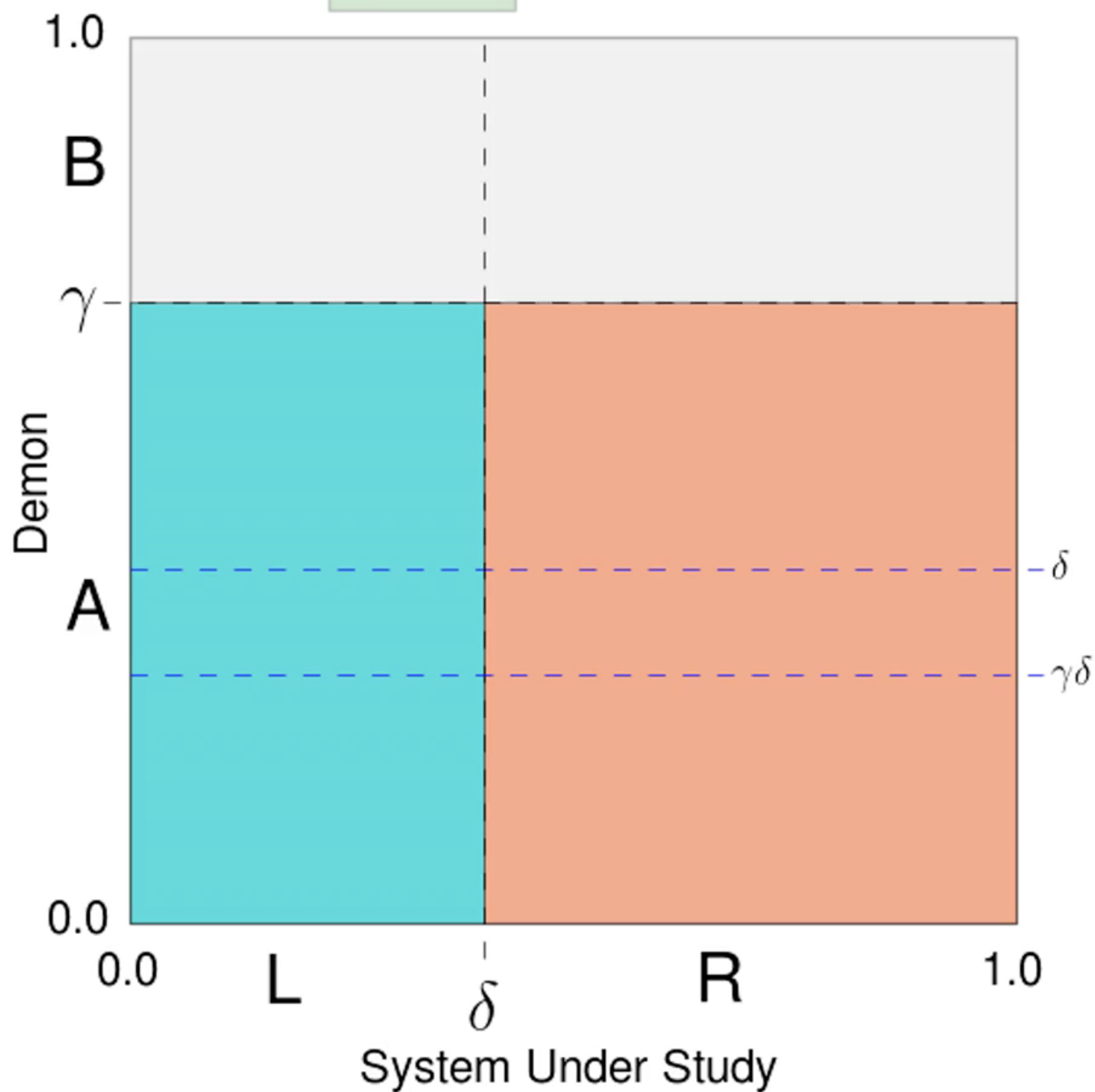
Szilard Engine is a Chaotic Map

Measure Control Erase



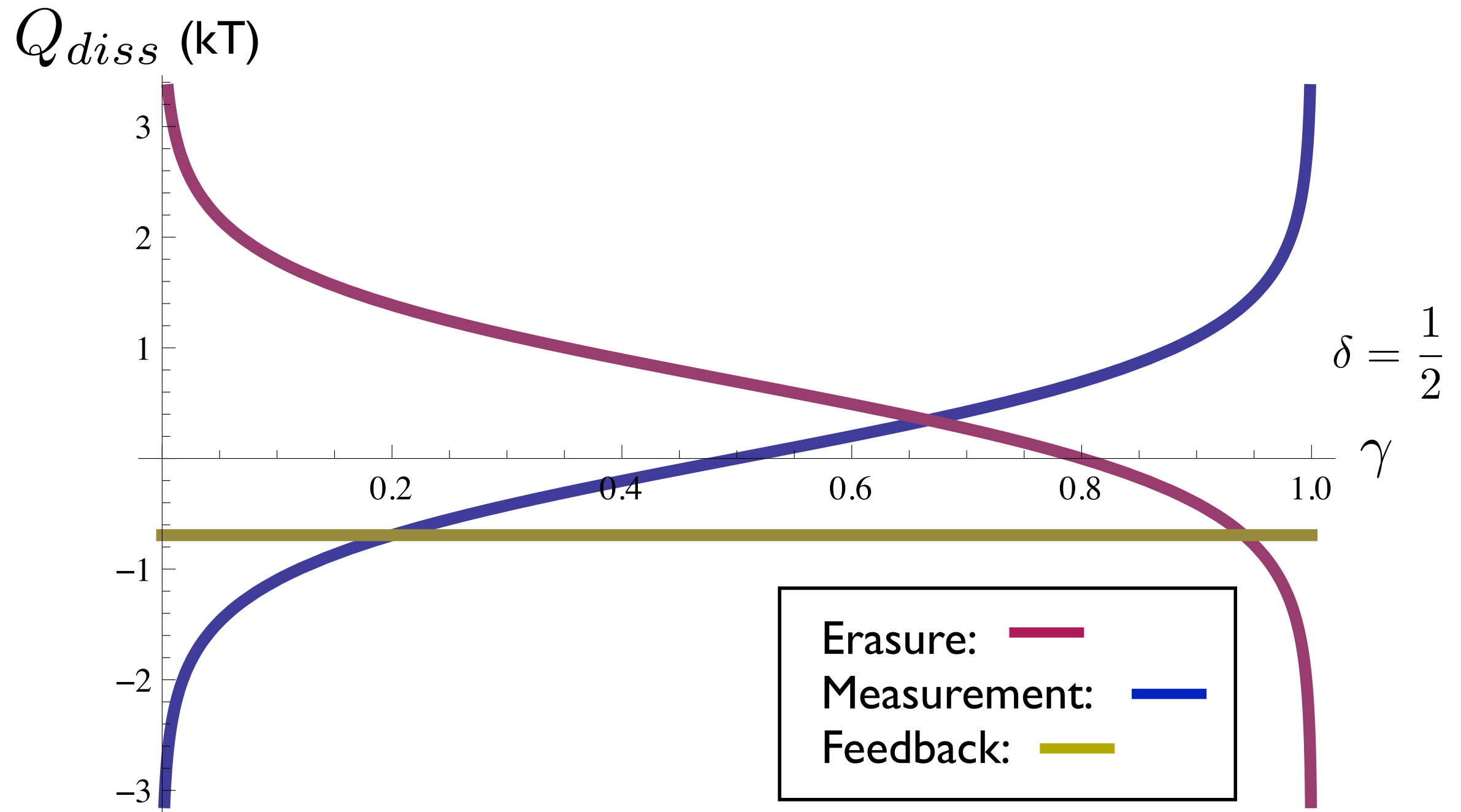
Szilard Engine is a Chaotic Map

Measure Control Erase



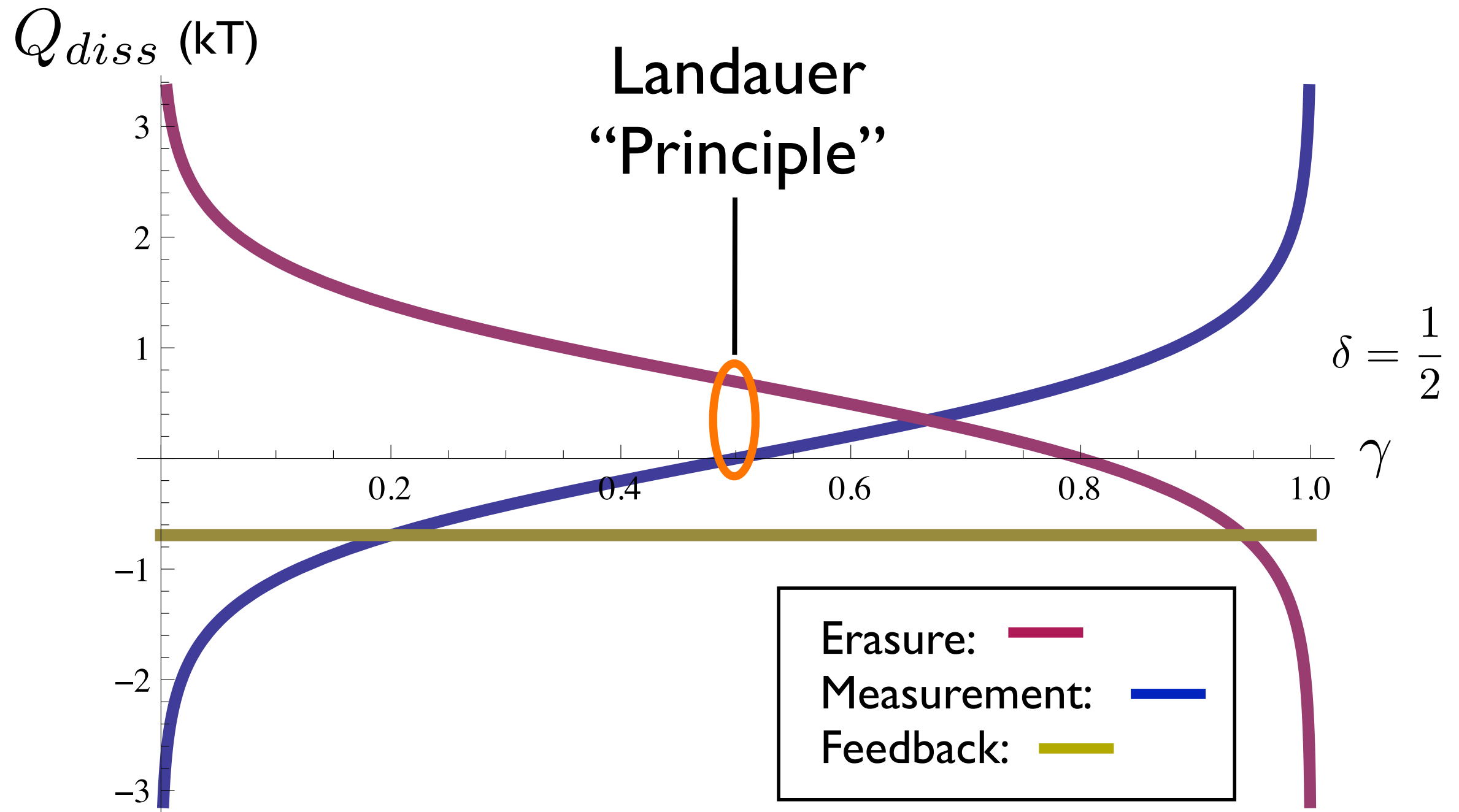
Beyond Landauer:

Energy Dissipation during Erasure and Measurement!



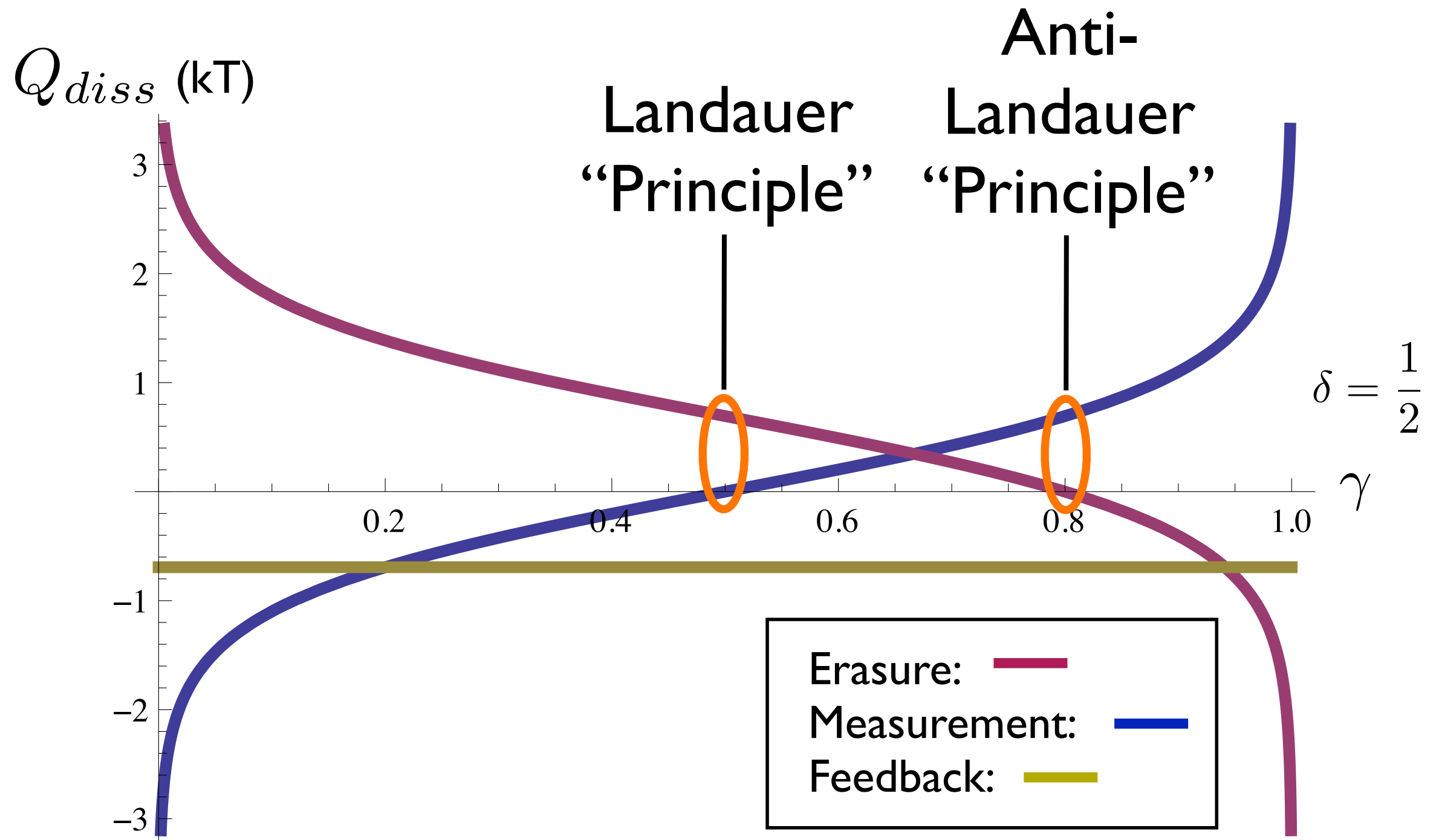
Beyond Landauer:

Energy Dissipation during Erasure and Measurement!



Beyond Landauer:

Energy Dissipation during Erasure and Measurement!



Complexity!

Information Theory for Complex Systems

Yesterday:

Complex Processes

Information in Processes

Today:

Memory in Processes

Intrinsic Computation

Measuring Structure

Intrinsic Computation

Optimal Models

Physics of Information

See online course:

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