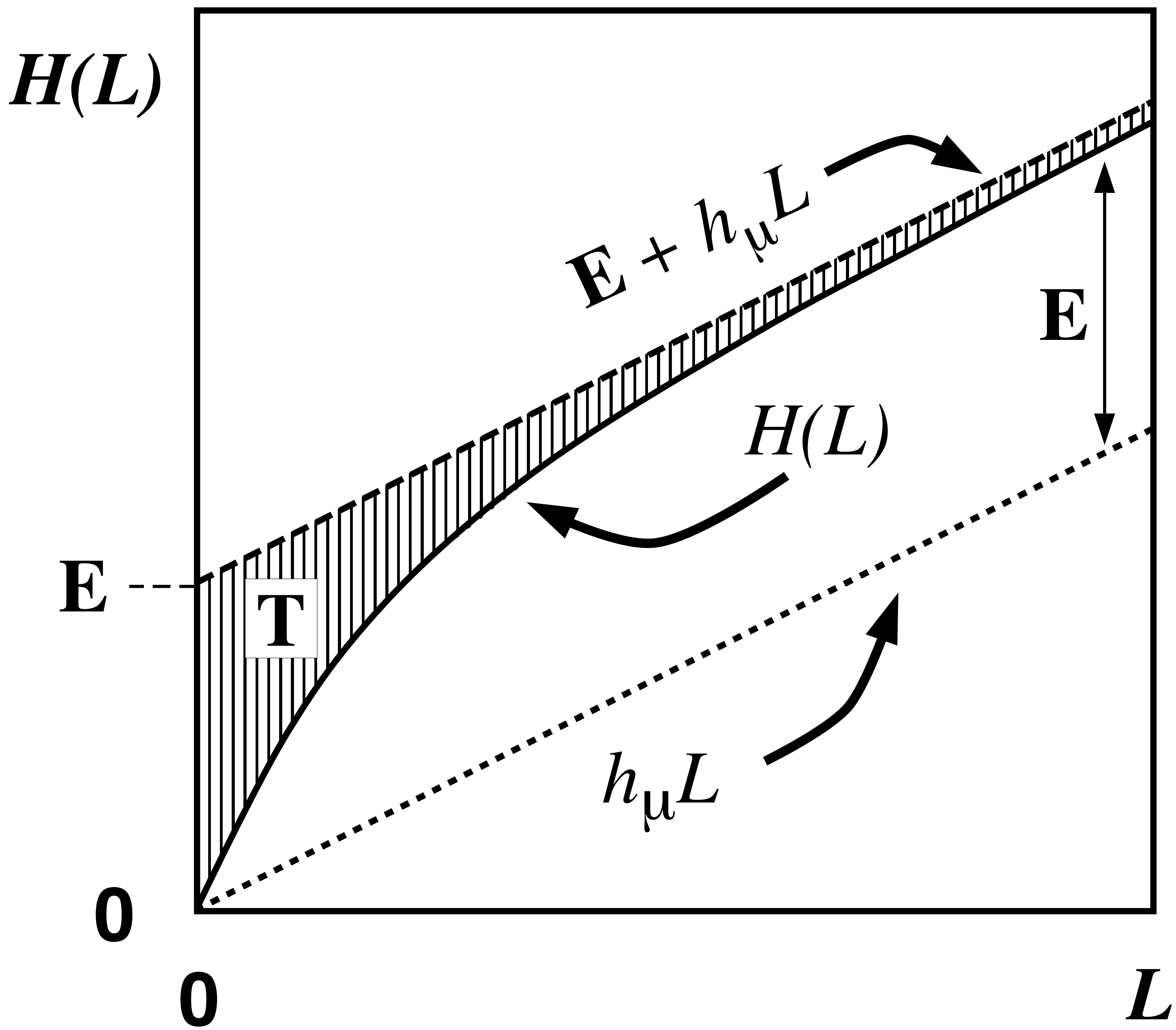
Intrinsic Computation

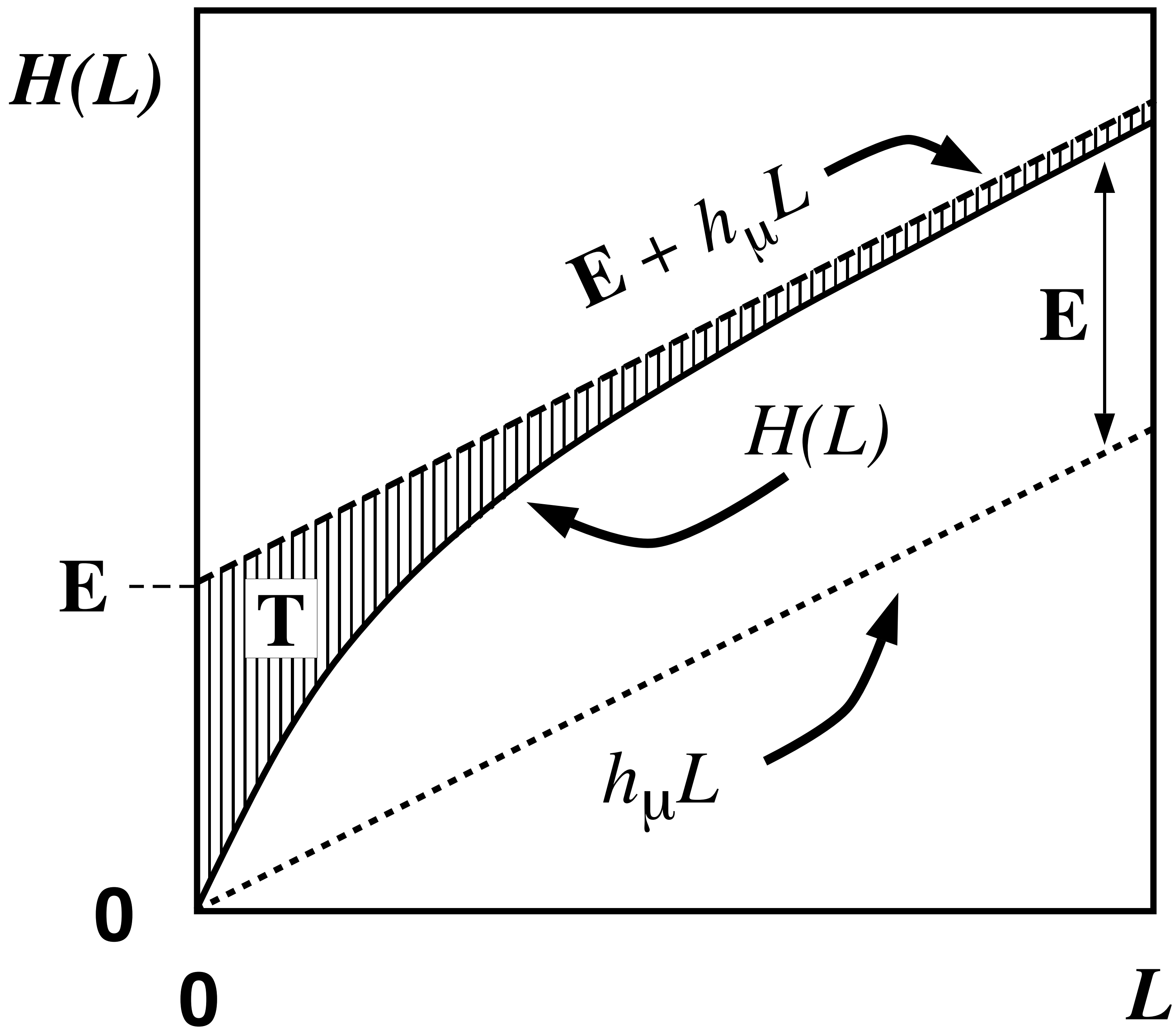
Jim Crutchfield & Nix Barnett
Complexity Sciences Center
Physics Department
University of California at Davis

Complex Systems Summer School
Santa Fe Institute
Santa Fe, NM
14 June 2013

Information Roadmap for a Complex Process

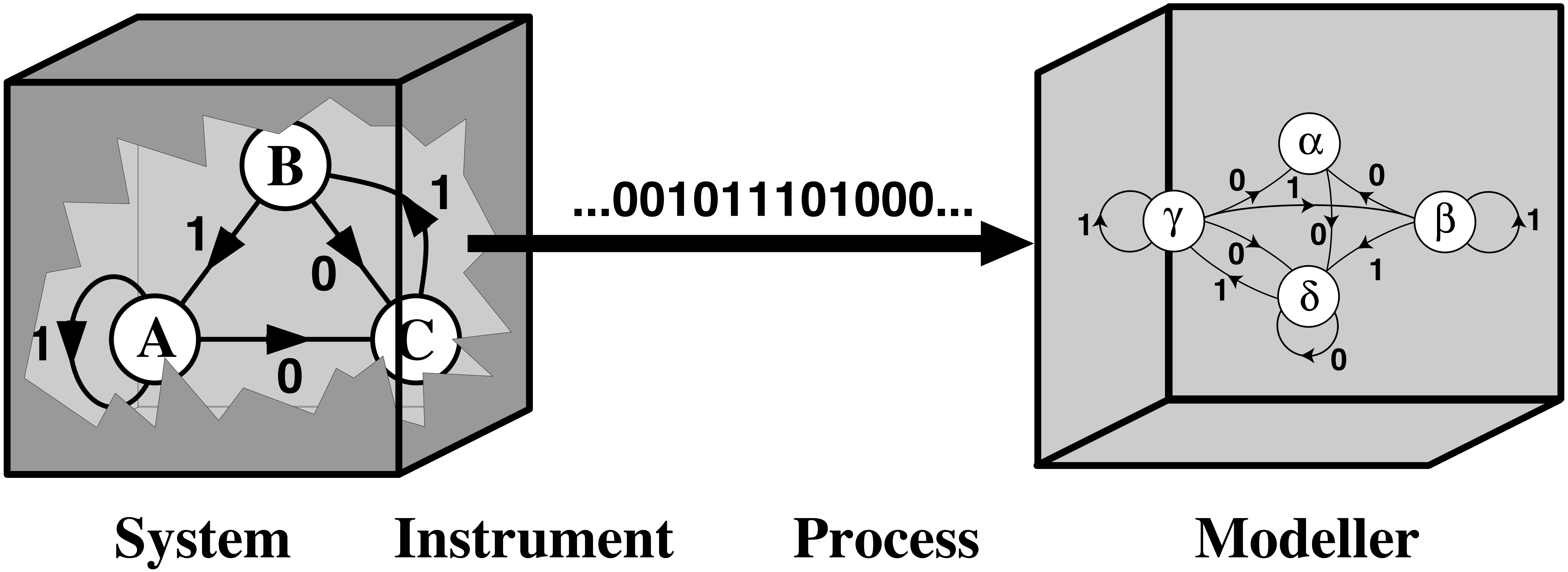


Information Roadmap for a Complex Process



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: $\ldots 111111111111$

The Learning Channel ...

The Prediction Game ...

Process I:

Past: $\ldots 111111111111$

Your prediction is?

The Learning Channel ...

The Prediction Game ...

Process I:

Past: $\ldots 111111111111$

Your prediction is?

Future: $111111111111\ldots$

The Prediction Game ...

Process I:

Past: $\ldots 111111111111$

Your prediction is?

Future: $111111111111 \ldots$

Your model (states & dynamic) is?

The Learning Channel ...

The Prediction Game ...

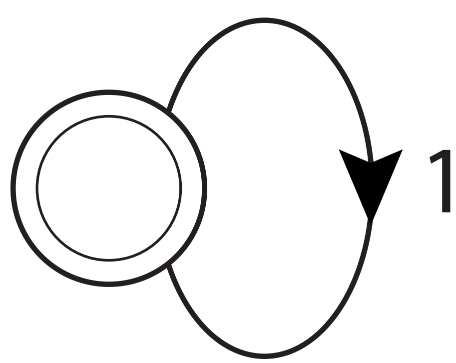
Process I:

Past: $\ldots 111111111111$

Your prediction is?

Future: $111111111111 \ldots$

Your model (states & dynamic) is?



The Learning Channel ...

The Prediction Game ...

Process II:

Complexity Lecture 2: Intrinsic Computation (CSSS 2013) Jim Crutchfield

The Learning Channel ...

The Prediction Game ...

Process II:

Past: $\ldots 10110010001101110$

The Learning Channel ...

The Prediction Game ...

Process II:

Past: $\ldots 10110010001101110$

Your prediction is?

The Learning Channel ...

The Prediction Game ...

Process II:

Past: $\ldots 10111001000 1101110$

Your prediction is?

Analysis: All words of length L occur & equally often

The Learning Channel ...

The Prediction Game ...

Process II:

Past: $\ldots 10111001000110110$

Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

The Prediction Game ...

Process II:

Past: $\ldots 1011001000110111 0$

Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

$$0101011101000110 1 \ldots$$

The Learning Channel ...

The Prediction Game ...

Process II:

Past: $\ldots 10110010001101110$

Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

$$0101011101000110 1 \ldots$$

Your model is?

The Learning Channel ...

The Prediction Game ...

Process II:

Past: $\ldots 1011001000110111 0$

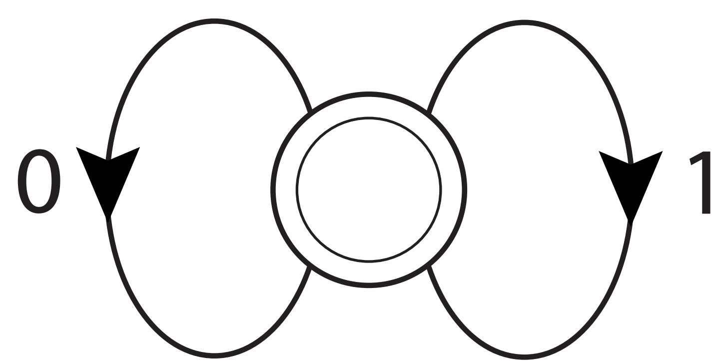
Your prediction is?

Analysis: All words of length L occur & equally often

Future: Well, anything can happen, how about?

$0101011101000110 1 \ldots$

Your model is?



The Learning Channel ...

The Prediction Game ...

Process III:

The Learning Channel ...

The Prediction Game ...

Process III:

Past: $\ldots 1010101010101010$

The Prediction Game ...

Process III:

Past: $\ldots 1010101010101010$

Your prediction is?

The Learning Channel ...

The Prediction Game ...

Process III:

Past: $\ldots 1010101010101010$

Your prediction is?

Future: $1010101010101010\ldots$

The Learning Channel ...

The Prediction Game ...

Process III:

Past: $\ldots 1010101010101010$

Your prediction is?

Future: $101010101010101\ldots$

Your model is?

The Prediction Game ...

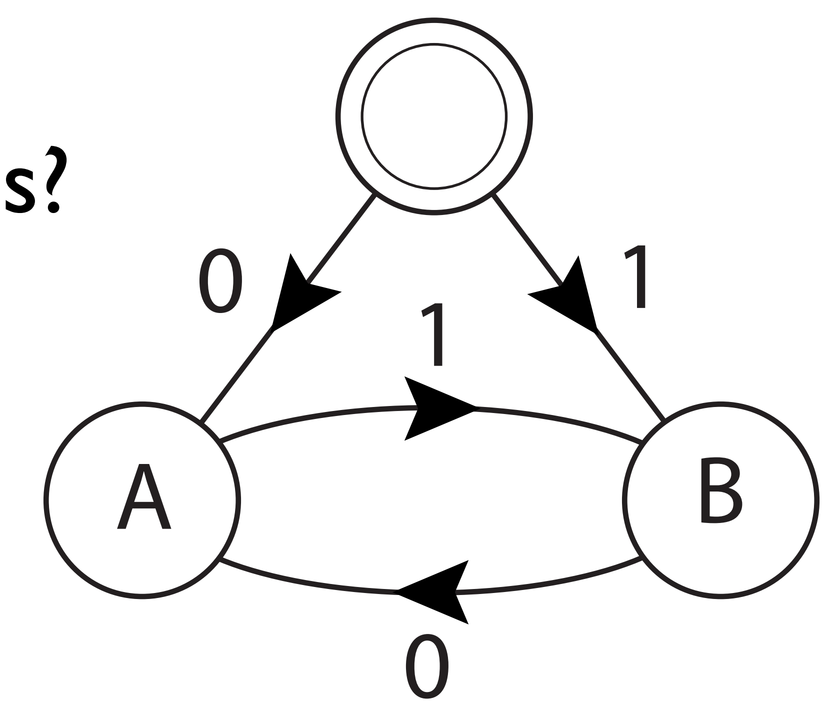
Process III:

Past: $\ldots 1010101010101010$

Your prediction is?

Future: $1010101010101010\ldots$

Your model is?



The Learning Channel ...

Theory? Algorithms?

Computational Mechanics

The Learning Channel ...

Goal:

Predict the future $\overrightarrow{S}$

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

But what "information" to use?

We want to find the effective "states"

and the dynamic (state-to-state mapping)

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

All we have are sequences of observations

Over some measurement alphabet $\mathcal{A}$

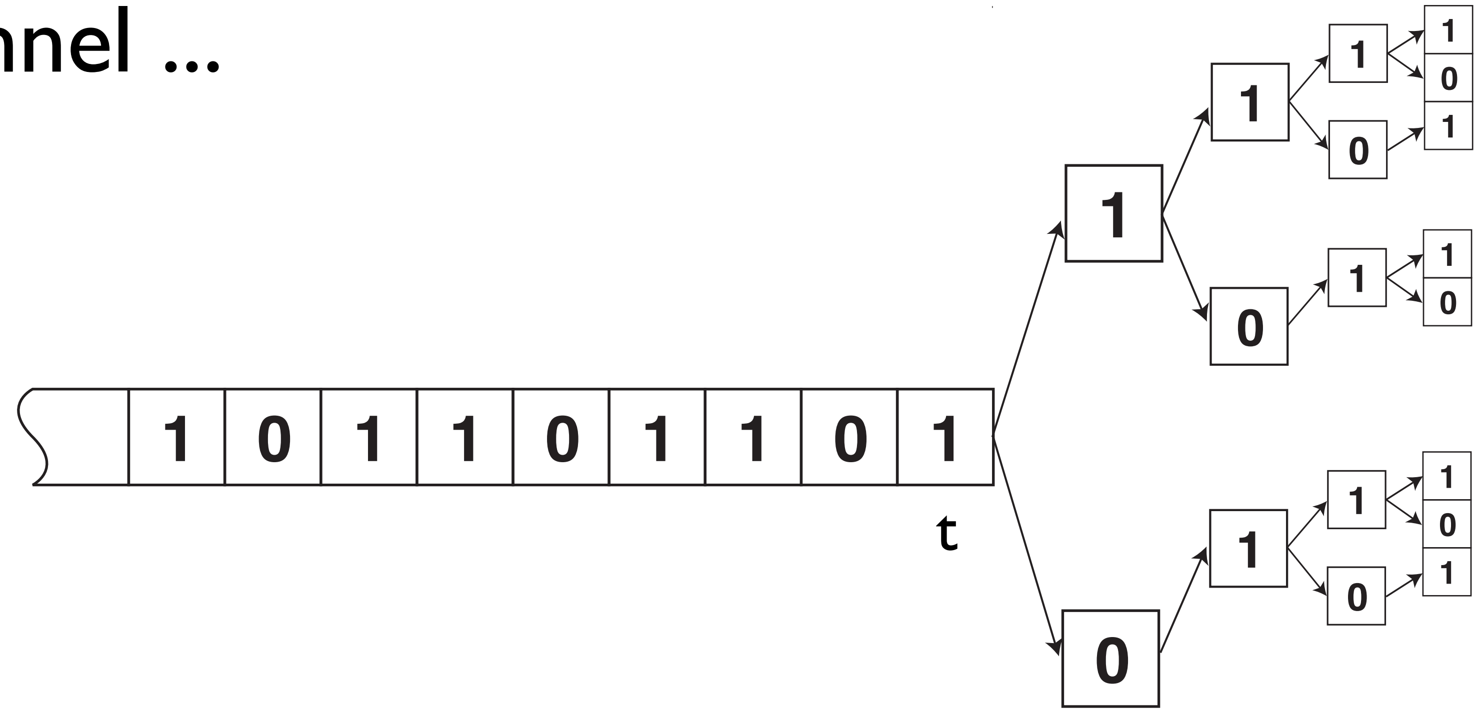
These symbols only indirectly reflect the hidden states

The Learning Channel ...

Effective States:

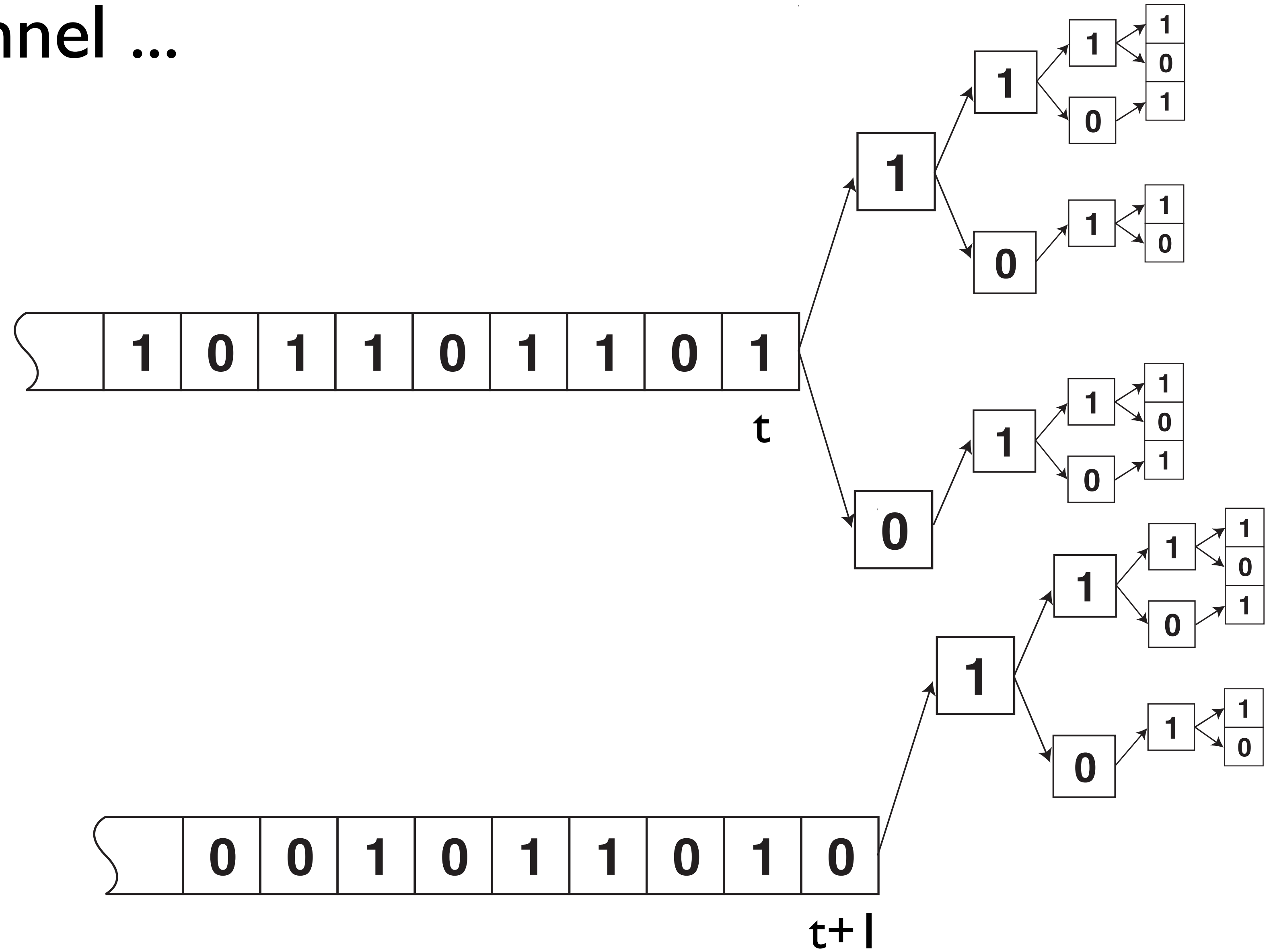
The Learning Channel ...

Effective States:



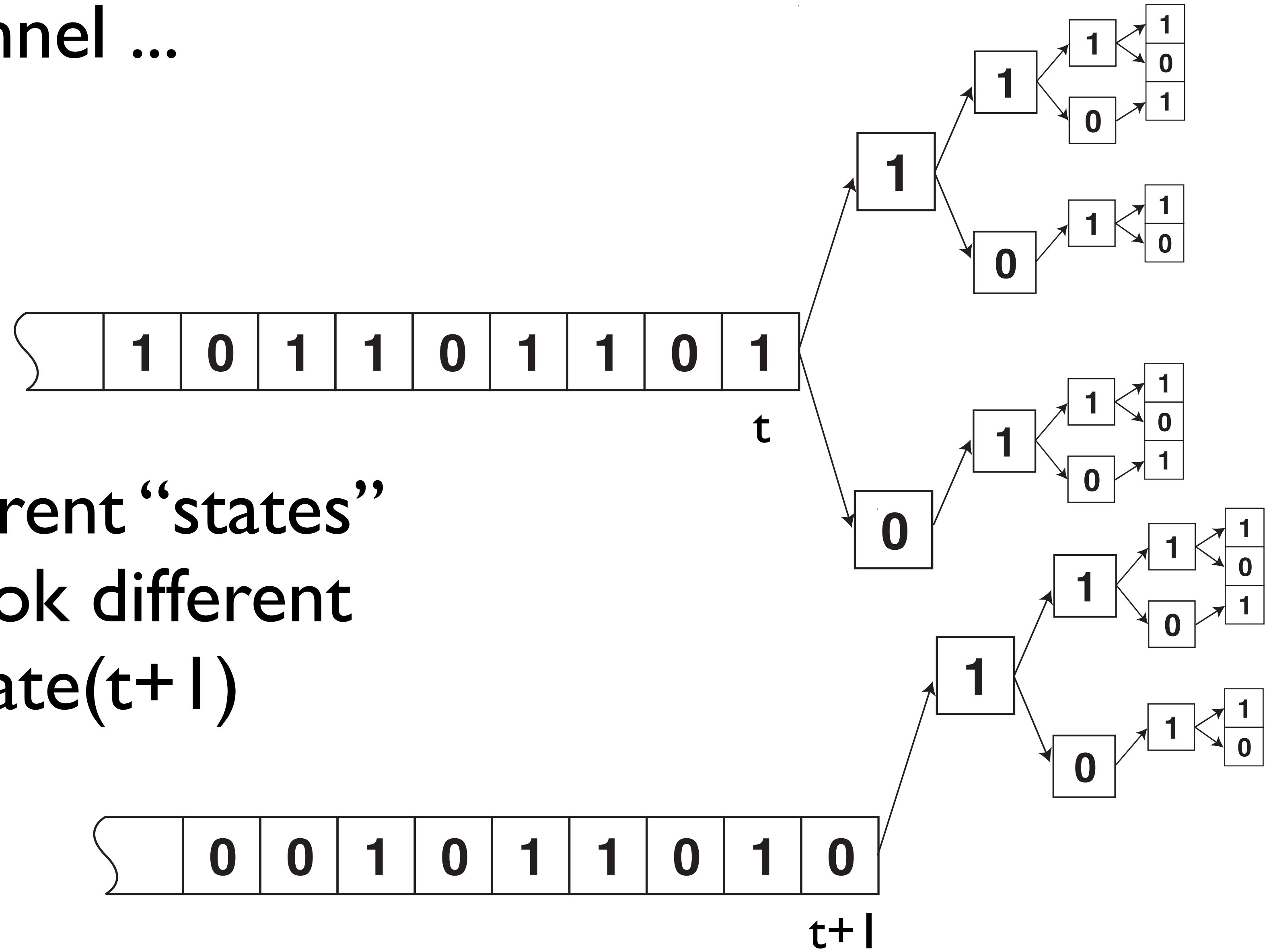
The Learning Channel ...

Effective States:



The Learning Channel ...

Effective States:

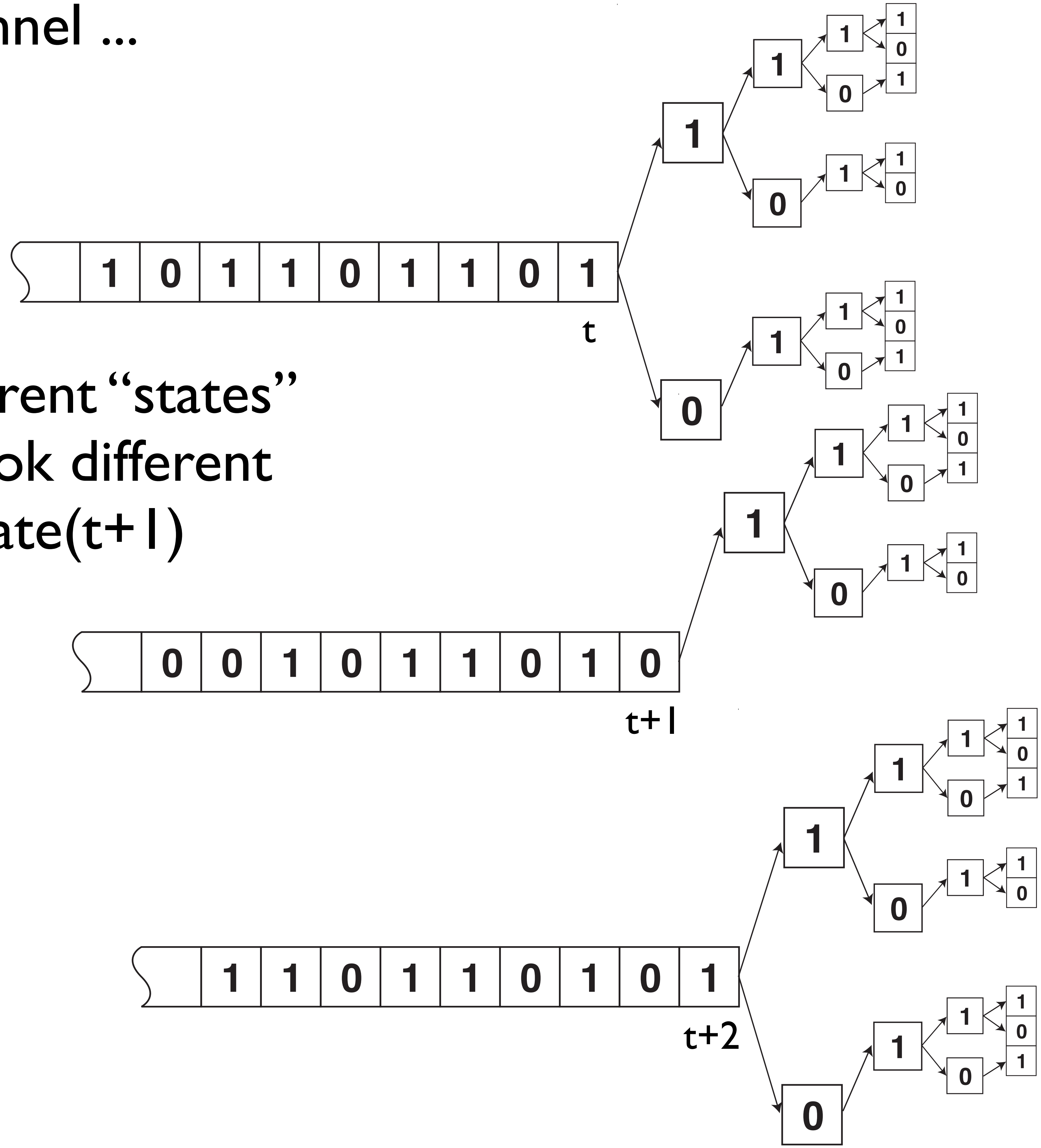


Process is in different "states"
when futures look different
State(t) $\nsim$ State(t+1)

The Learning Channel ...

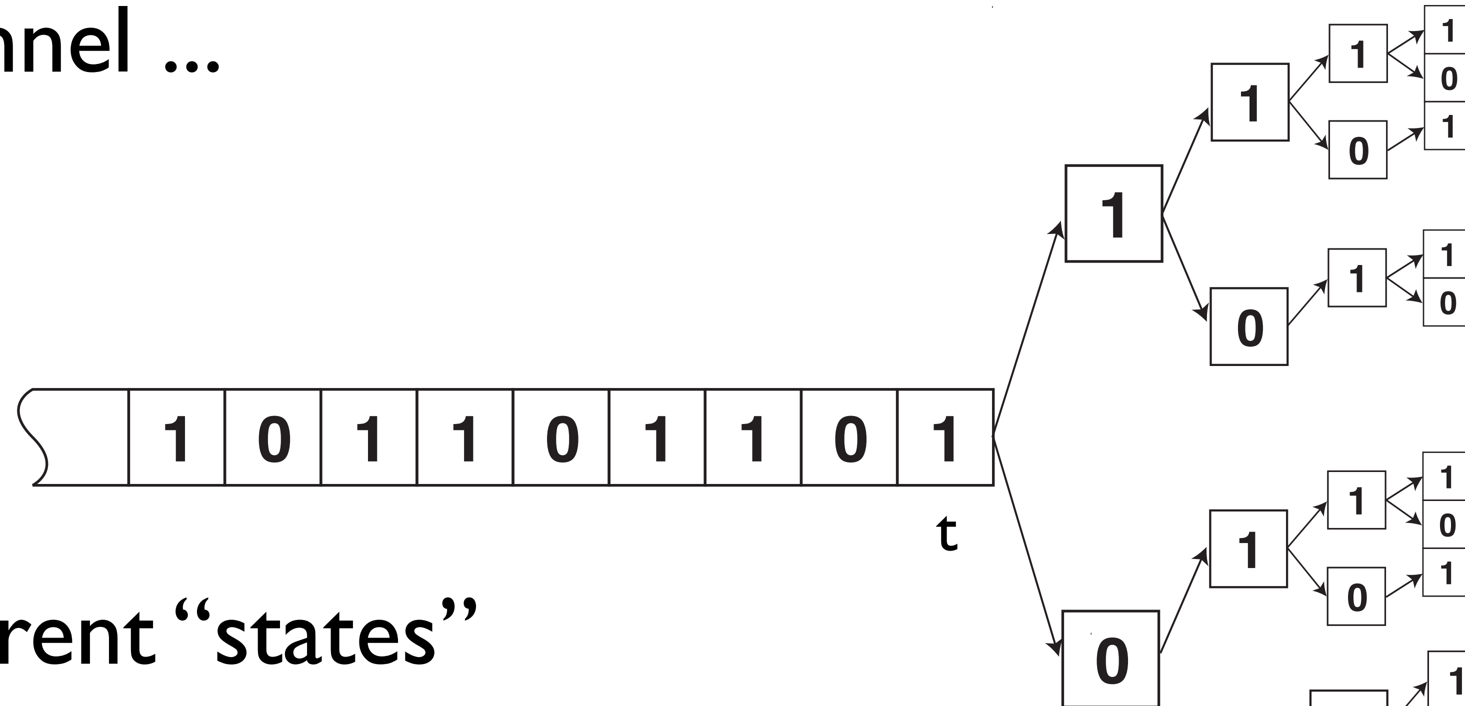
Effective States:

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

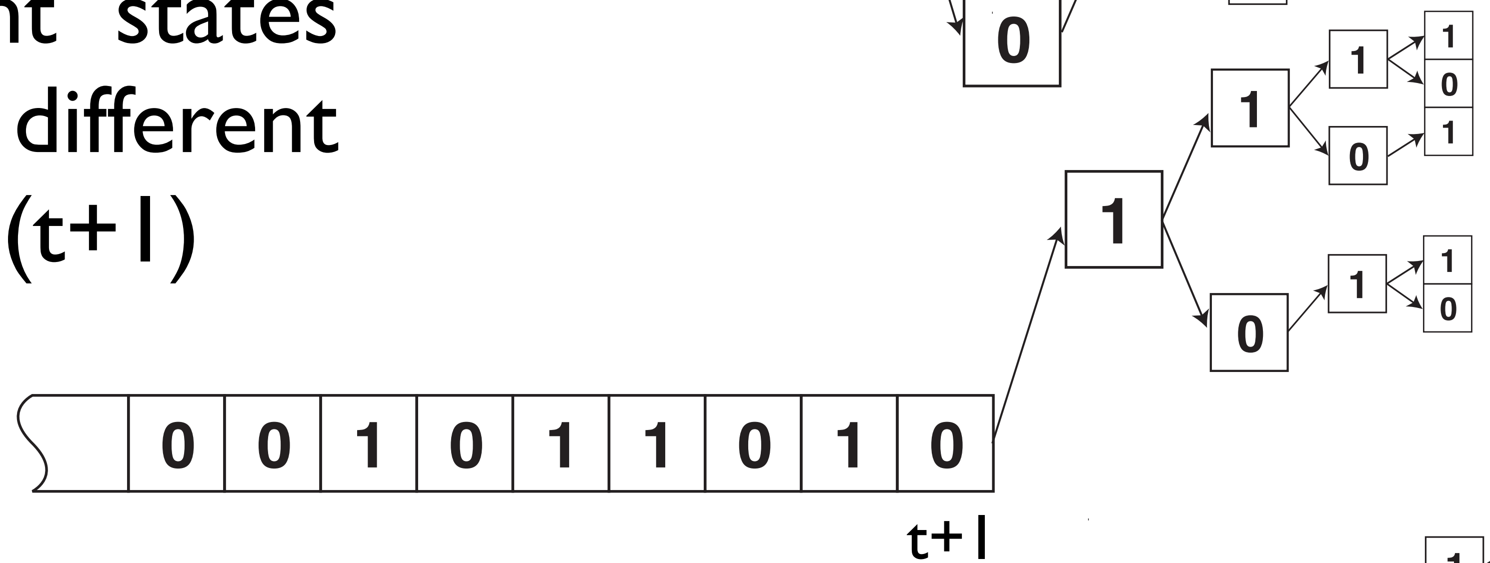


The Learning Channel ...

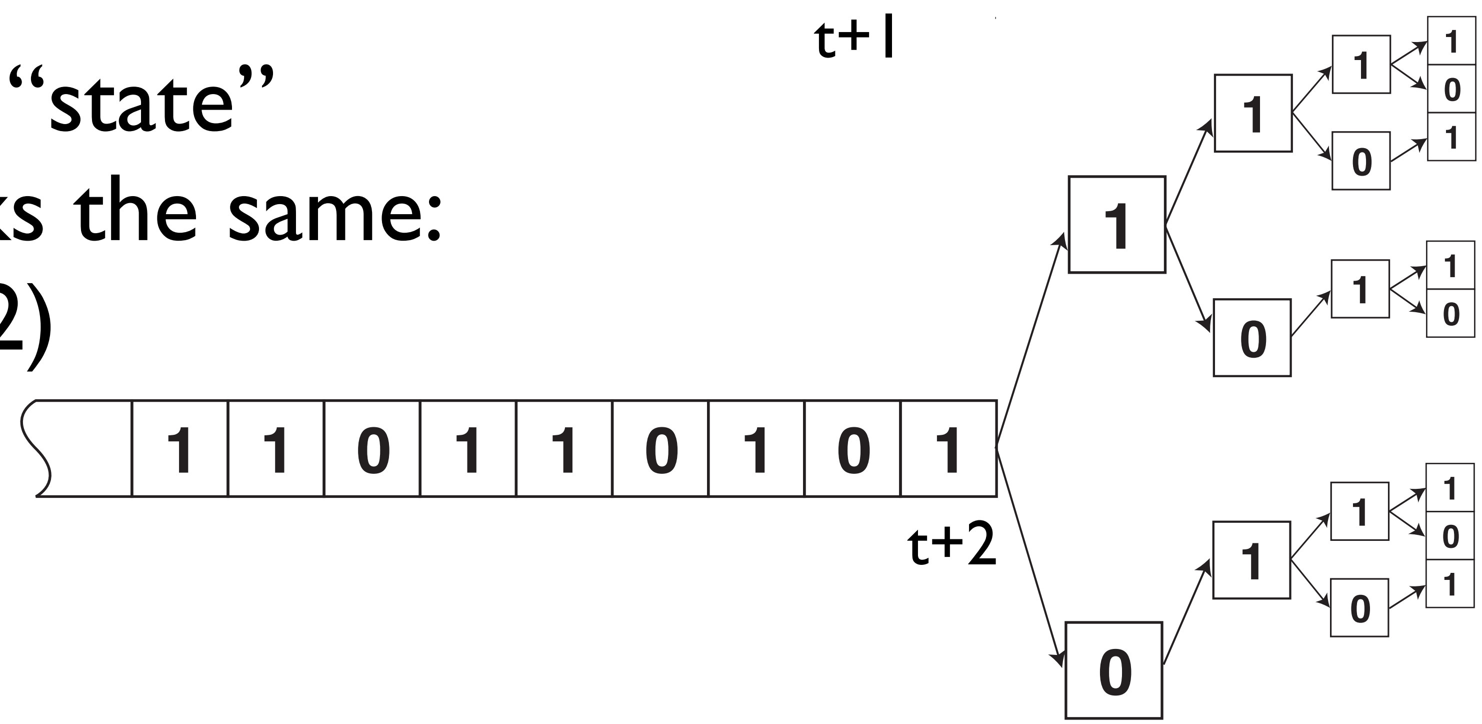
Effective States:



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



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



Complexity Lecture 2: Intrinsic Computation (CSSS 2013) Jim Crutchfield

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: $\Pr(\overrightarrow{S}^L \mid \overleftarrow{s})$ (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:

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

Uncertainty about future given effective states

Effective Prediction Error Rate:

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

Entropy rate given effective states

How Effective are the Effective States?

Statistical Complexity of the Effective States:

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

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?

$$H[\overrightarrow{S}^L | R] = H[\overrightarrow{S}^L \mid \overleftarrow{S}]$$

or

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

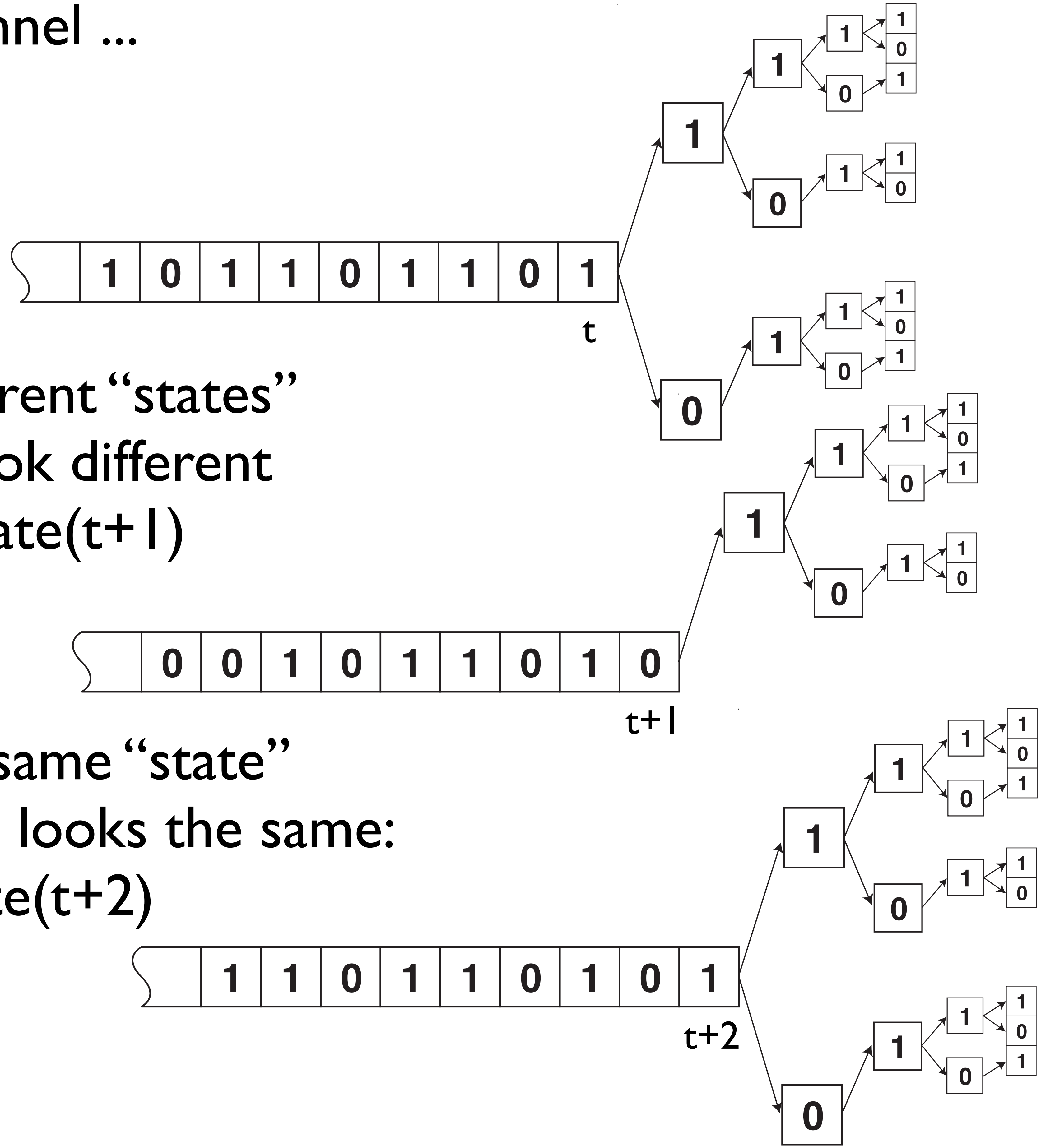
Question 2:

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
State(t) $\nsim$ State(t+1)

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

Causal States:

Causal State:

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

Set of histories that lead to same predictions.

Predictive equivalence relation:

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

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

The Learning Channel ...

Causal State Components

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

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

Set of causal states:

$$\boldsymbol{\mathcal{S}} = \overleftarrow{\mathbf{S}} / \sim \quad = \{\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots\}$$

Causal state map:

$$\epsilon : \overleftarrow{\mathbf{S}} \to \boldsymbol{\mathcal{S}}$$

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

Causal States ...

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

We have the effective states!

Now,

What is the dynamic?

Causal State Dynamic ...

Causal-state Filtering:

$$\overleftrightarrow{s} = \ldots\ s_{-3}\quad s_{-2}\quad s_{-1}\quad s_0\quad s_1\quad s_2\quad s_3\quad \ldots$$

$$\epsilon(\overleftrightarrow{s}) = \ldots \epsilon(\overleftarrow{s}_{-3})\epsilon(\overleftarrow{s}_{-2})\epsilon(\overleftarrow{s}_{-1})\epsilon(\overleftarrow{s}_{0})\epsilon(\overleftarrow{s}_{1})\epsilon(\overleftarrow{s}_{2})\epsilon(\overleftarrow{s}_{3}) \ldots$$

$$\overleftrightarrow{\mathcal{S}} = \ldots\quad \mathcal{S}_{t=-3}\ \mathcal{S}_{t=-2}\ \mathcal{S}_{t=-1}\ \mathcal{S}_{t=0}\ \mathcal{S}_{t=1}\ \mathcal{S}_{t=2}\ \mathcal{S}_{t=3}\quad \ldots$$

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:

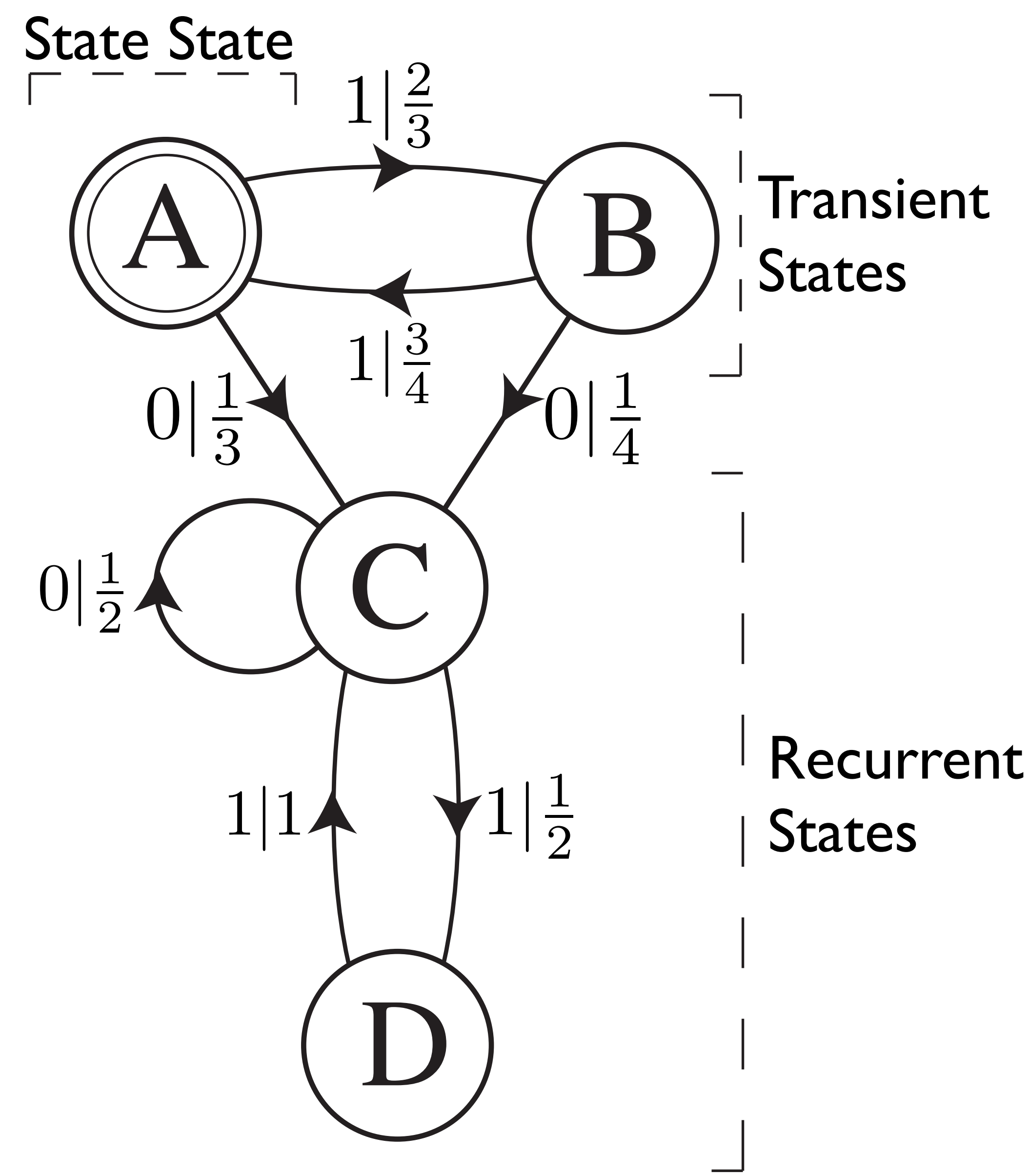
$$\{T_{ij}^{(s)} : s \in \mathcal{A},\ i, j = 0, 1, \ldots, |\boldsymbol{\mathcal{S}}|\}$$

The ϵ-Machine ...

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

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

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



Unique Start State:

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

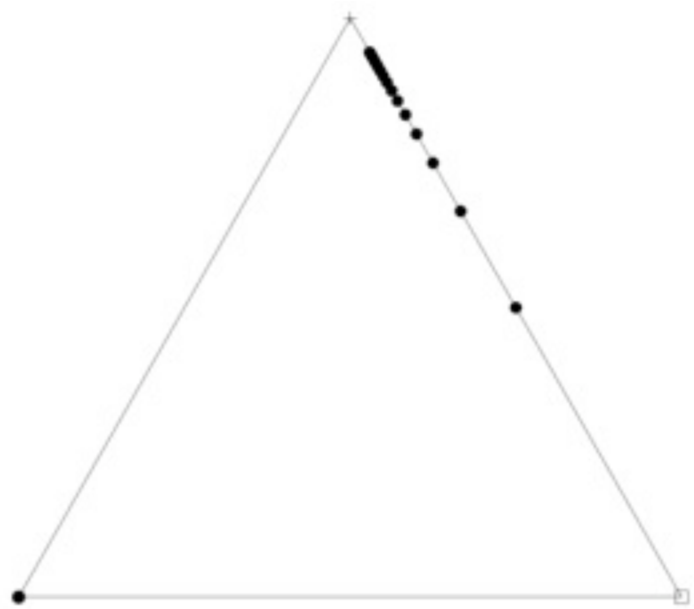
$$\Pr(\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots) = (1, 0, 0, \ldots)$$

Transient States

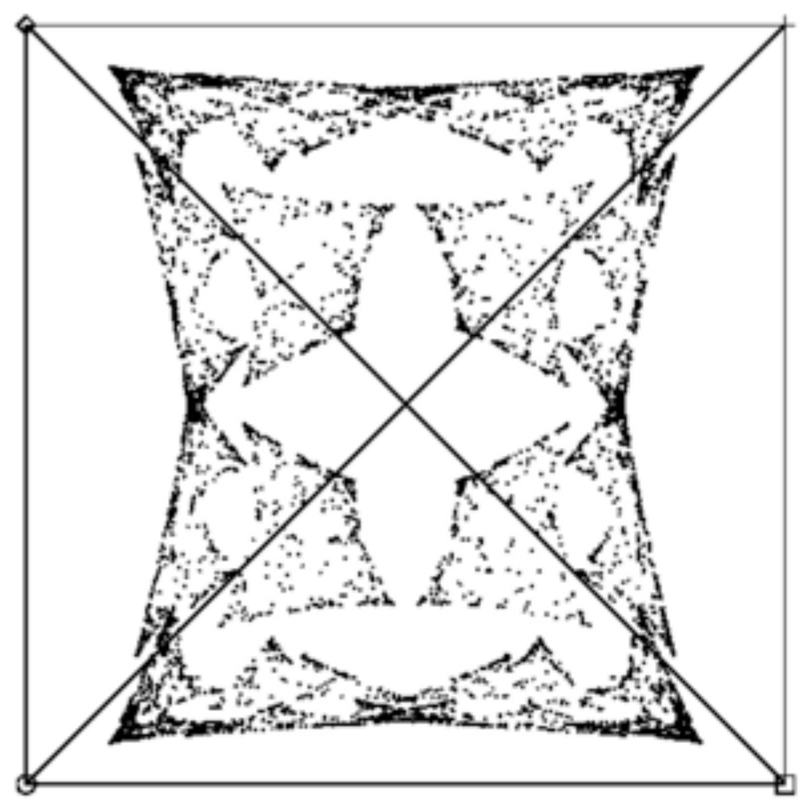
Recurrent States

The Learning Channel ...

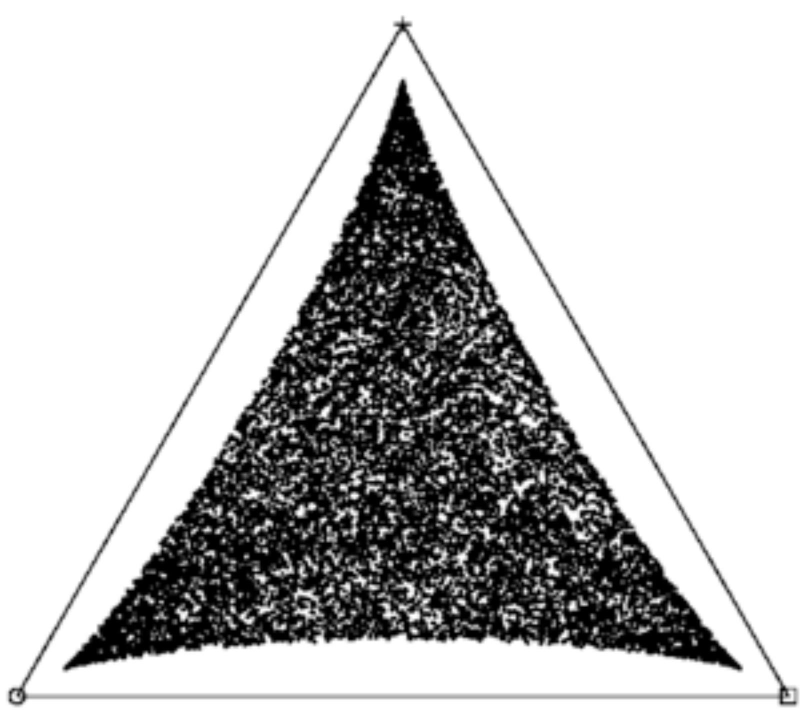
The ϵ-Machine of a Process ...



Denumerable Causal States



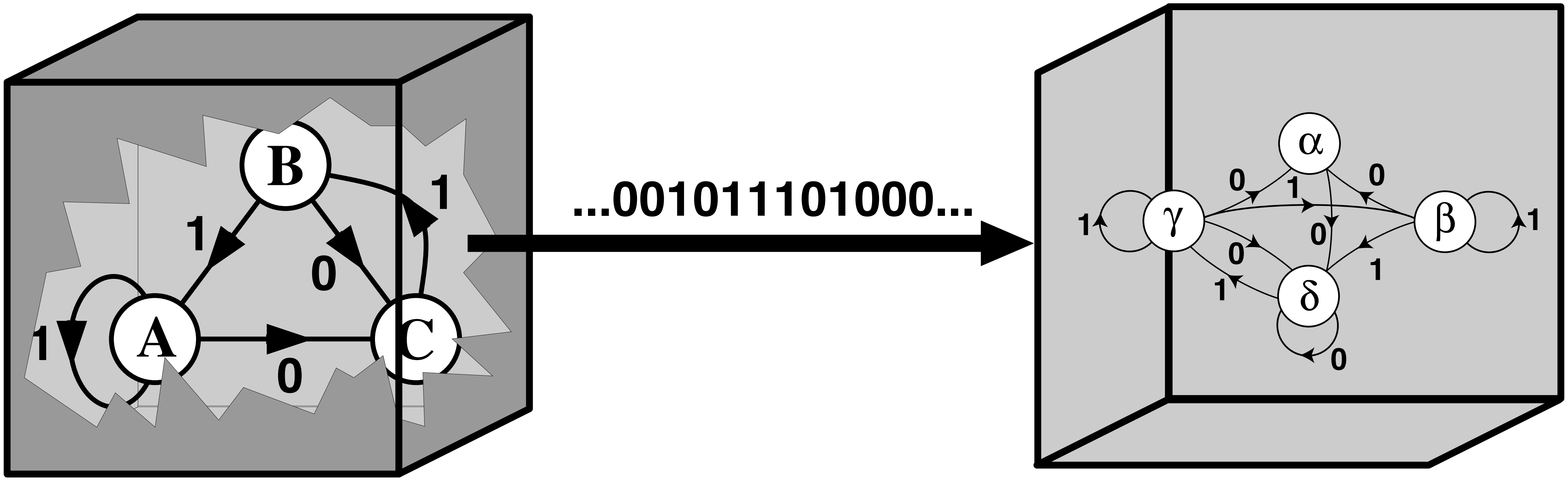
Fractal



Continuous

Complexity Lecture 2: Intrinsic Computation (CSSS 2013) Jim Crutchfield

The Learning Channel:



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

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

ϵ-Machine reproduces the process's word distribution:

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

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

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

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

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

The ϵ-Machine ...

Causal shielding:
Past and future are independent given causal state

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

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

Causal states shield past & future from each other.

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

The ϵ-Machine ...

$\epsilon\mathrm{Ms}$ are Unifilar: $(\mathcal{S}_t, s) \rightarrow$ 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 $\epsilon\mathrm{M}$ to calculate entropy rate h_μ.

(Any unifilar presentation will do.)

The ϵ-Machine ...

ϵMs are Optimal Predictors:

Compared to any rival effective states R:

$$H\left[\overrightarrow{S}^{L}|R\right] \geq H\left[\overrightarrow{S}^{L}|\mathcal{S}\right]$$

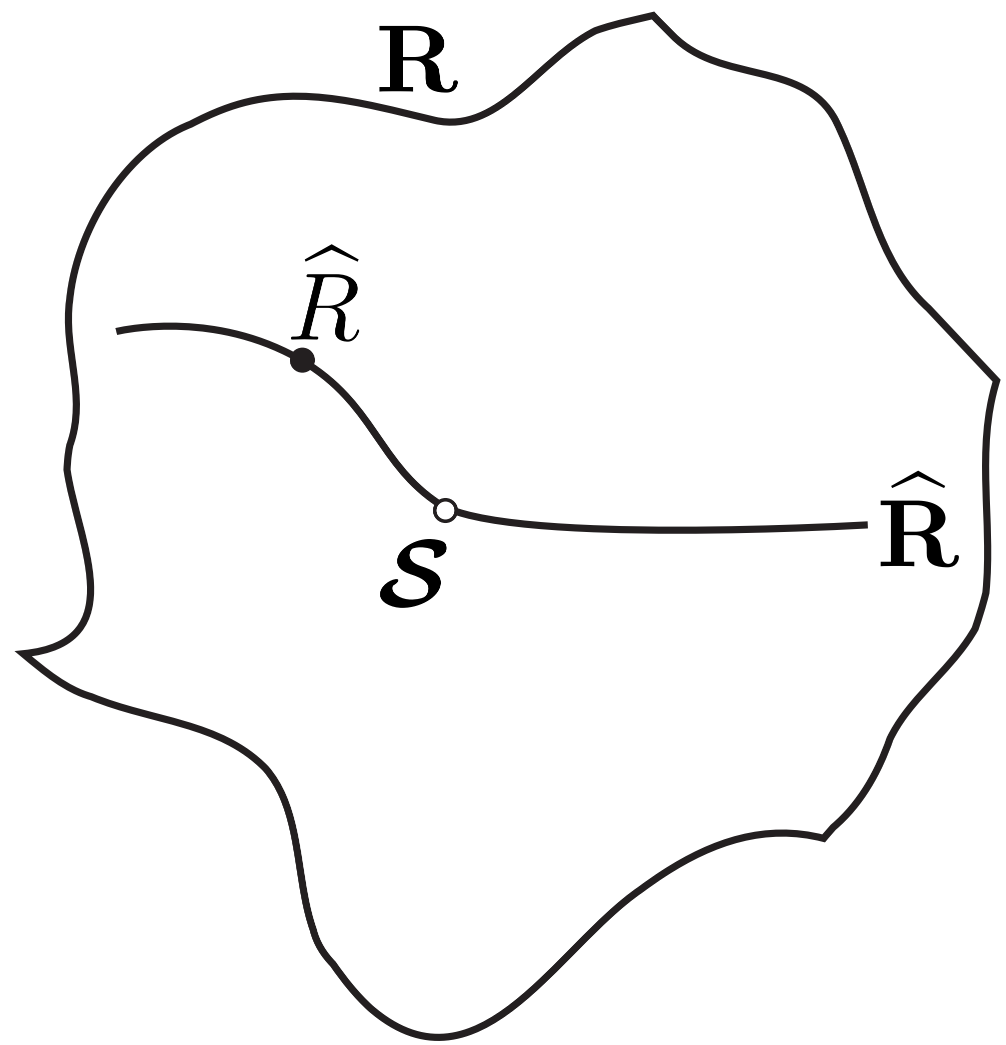
The ϵ-Machine ...

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

Alternative models that are optimal predictors

$$\widehat{R} \in \widehat{\mathbf{R}}$$

$$H[\overrightarrow{S}^L | \widehat{R}] = H[\overrightarrow{S}^L | \mathcal{S}]$$



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

The ϵ-Machine ...

Minimal Statistical Complexity:

For all prescient rivals, $\epsilon\mathrm{M}$ is the smallest:

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

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:

$\epsilon\mathrm{M}$:

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

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

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

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

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

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?

A complex process's intrinsic computation:

(1) How much of past does process store?

$$C_\mu$$

(2) In what architecture is that information stored?

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

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

A complex process's intrinsic computation:

(1) How much of past does process store?

$$C_\mu$$

(2) In what architecture is that information stored?

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

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

$$h_\mu$$

Measures of Intrinsic Computation ...

Measures of Structural Complexity:

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

How related to statistical complexity C_μ?

How to get from $\epsilon\mathrm{M}$?

Measures of Intrinsic Computation ...

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

Entropy Rate of a Process:

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

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

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

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

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

1-1 mapping between measurement sequences & internal paths.

Measures of Intrinsic Computation ...

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

Statistical Complexity of a Process:

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

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

Meaning:

Shannon information in the causal states.

The amount of historical information a process stores.

The amount of structure in a process.

Measures of Intrinsic Computation ...

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

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

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

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

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

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

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

General case: Need a new framework.

Measures of Intrinsic Computation ...

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

Bound on Excess Entropy:

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

Proof sketch:

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

(2) Causal States: $H[\overrightarrow{S} \mid \overleftarrow{S}] = H[\overrightarrow{S} \mid \mathcal{S}]$

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

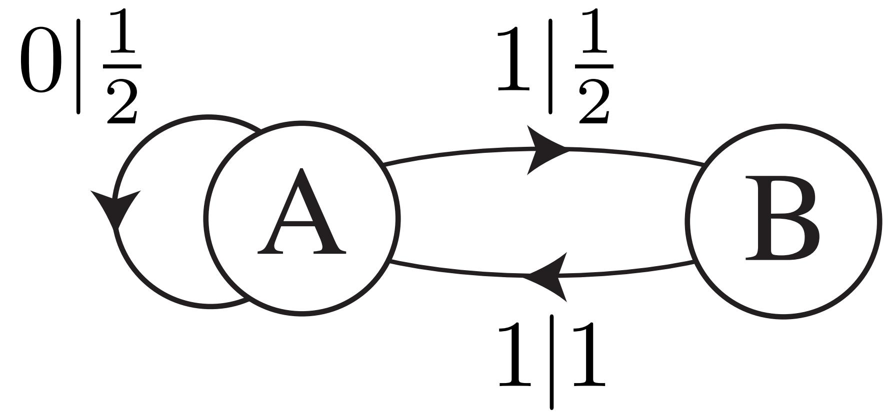
$$= I[\overrightarrow{S}; \mathcal{S}]$$

$$= H[\mathcal{S}] - H[\mathcal{S} \mid \overrightarrow{S}]$$

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

Measures of Intrinsic Computation ...

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



Bound on Excess Entropy ...

But, the bound is saturated!

Even Process:

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

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

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

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

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

When does this occur?

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

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

Bound on Excess Entropy ...

Consequence: The Cryptographic Limit

Can have $\mathbf{E} \to 0$ 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*

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

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

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

Consequence:

The inequality is Why We Must Model.

Cannot simply use sequences as states.

There is internal structure not expressed by this.

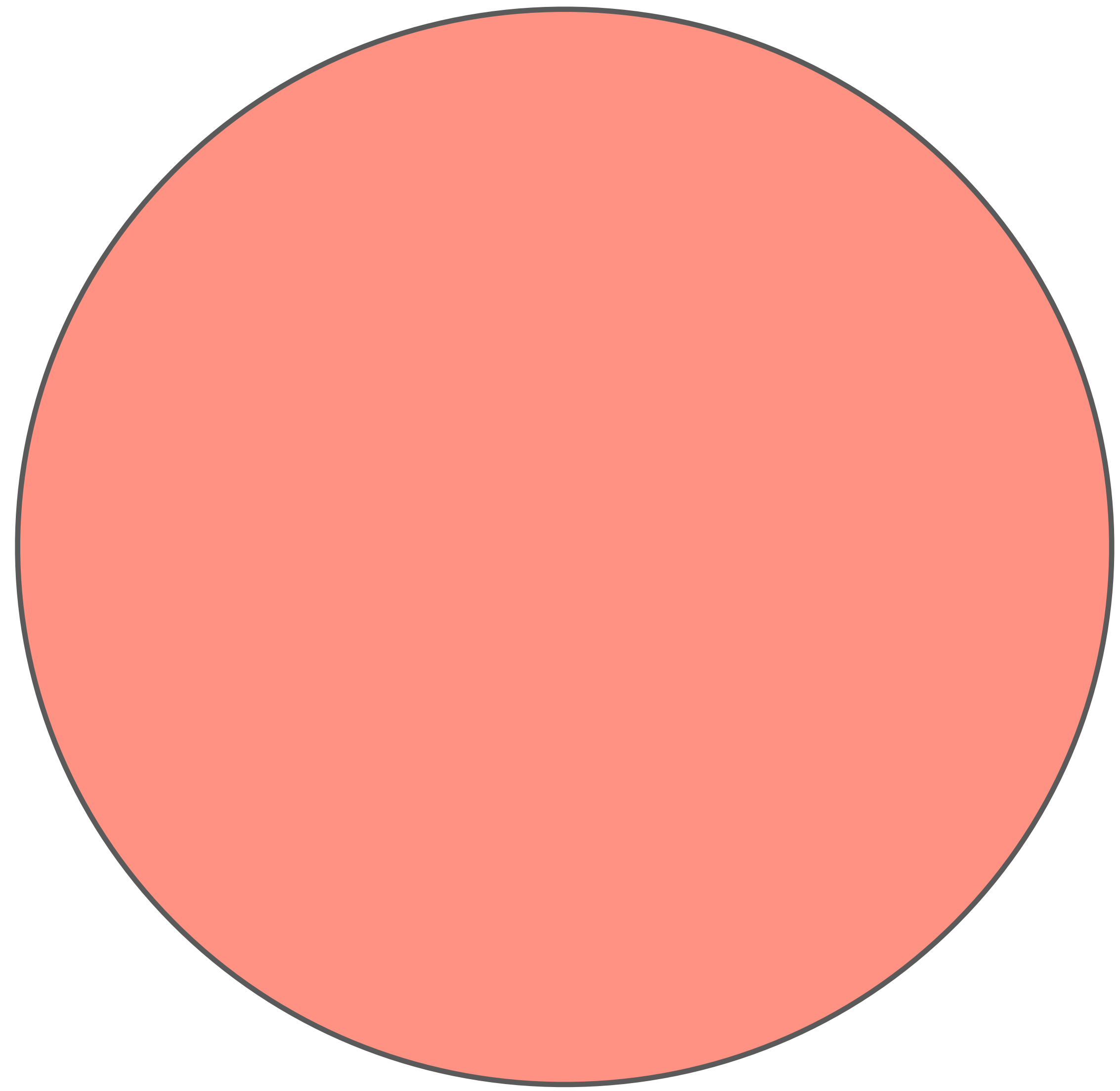
Information Diagrams for Processes

Information Diagrams for Processes

Process I-diagram:

Information Diagrams for Processes

Process I-diagram:



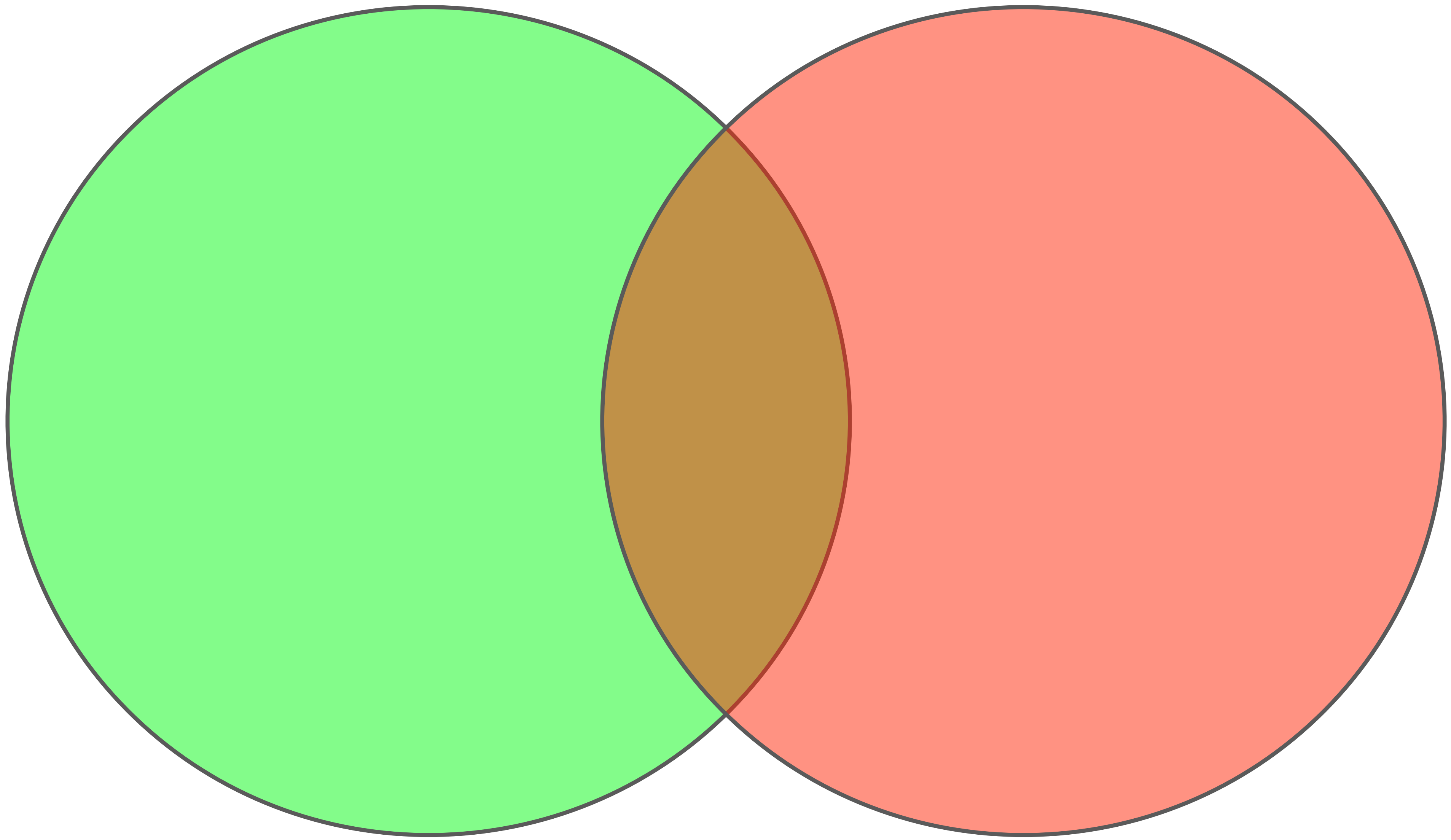
$H[\overrightarrow{X}]$

Complexity Lecture 2: Intrinsic Computation (CSSS 2013) Jim Crutchfield

Information Diagrams for Processes

Process I-diagram:

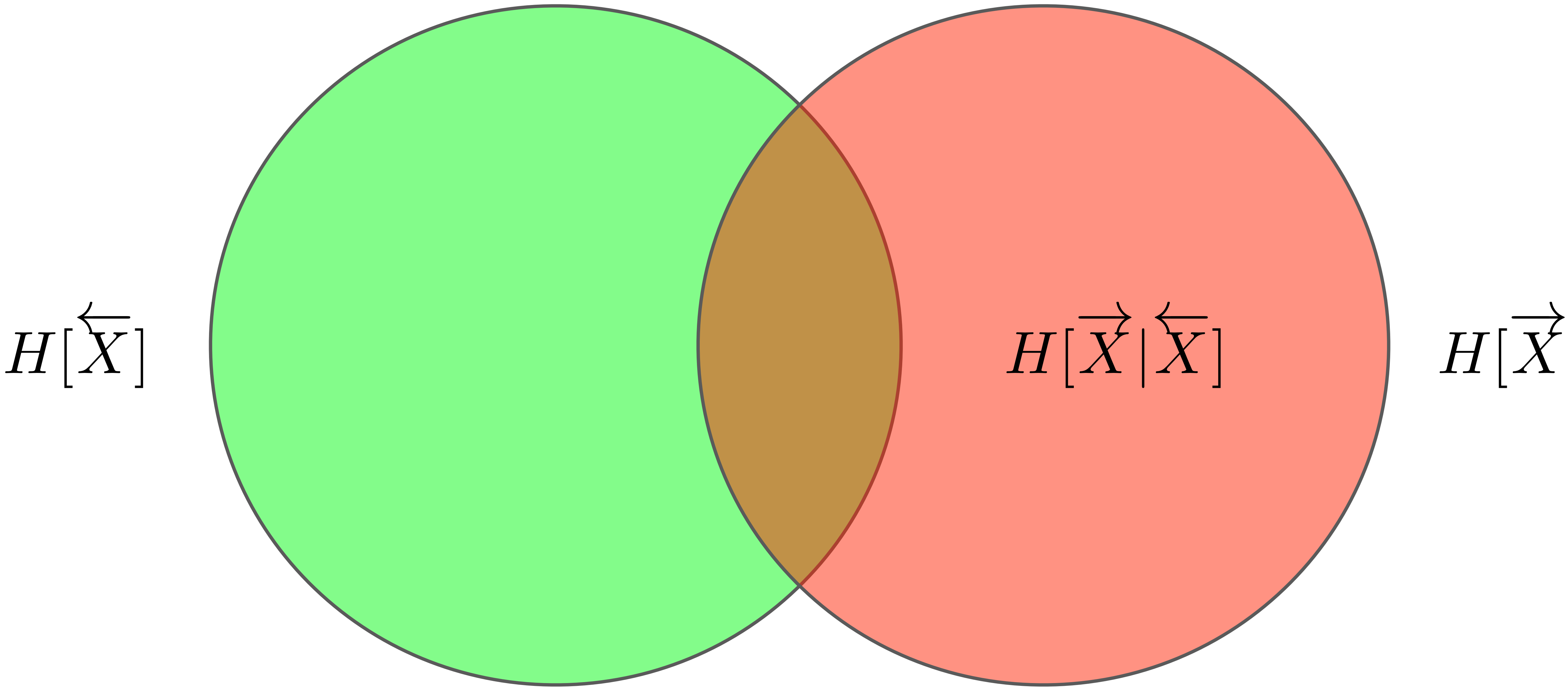
$H[\overleftarrow{X}]$



$H[\overrightarrow{X}]$

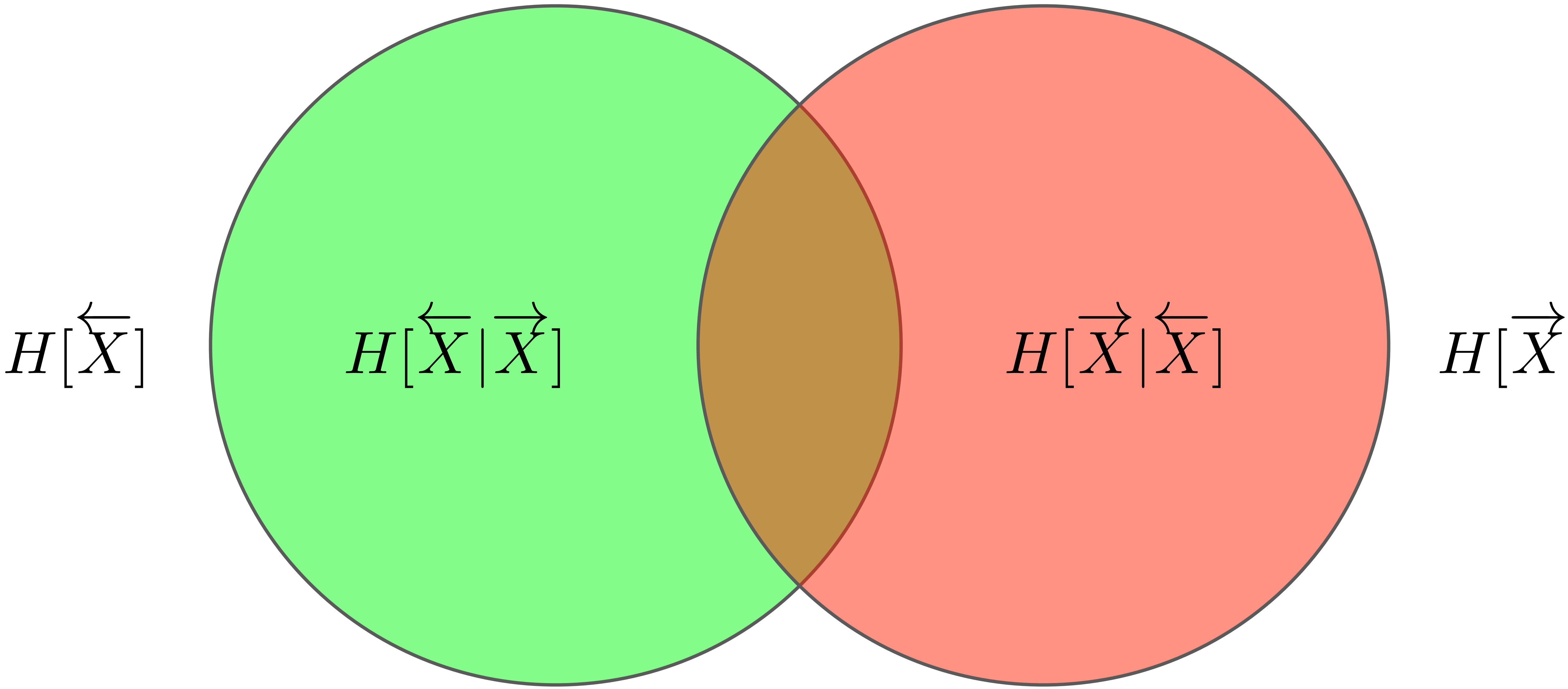
Information Diagrams for Processes

Process I-diagram:



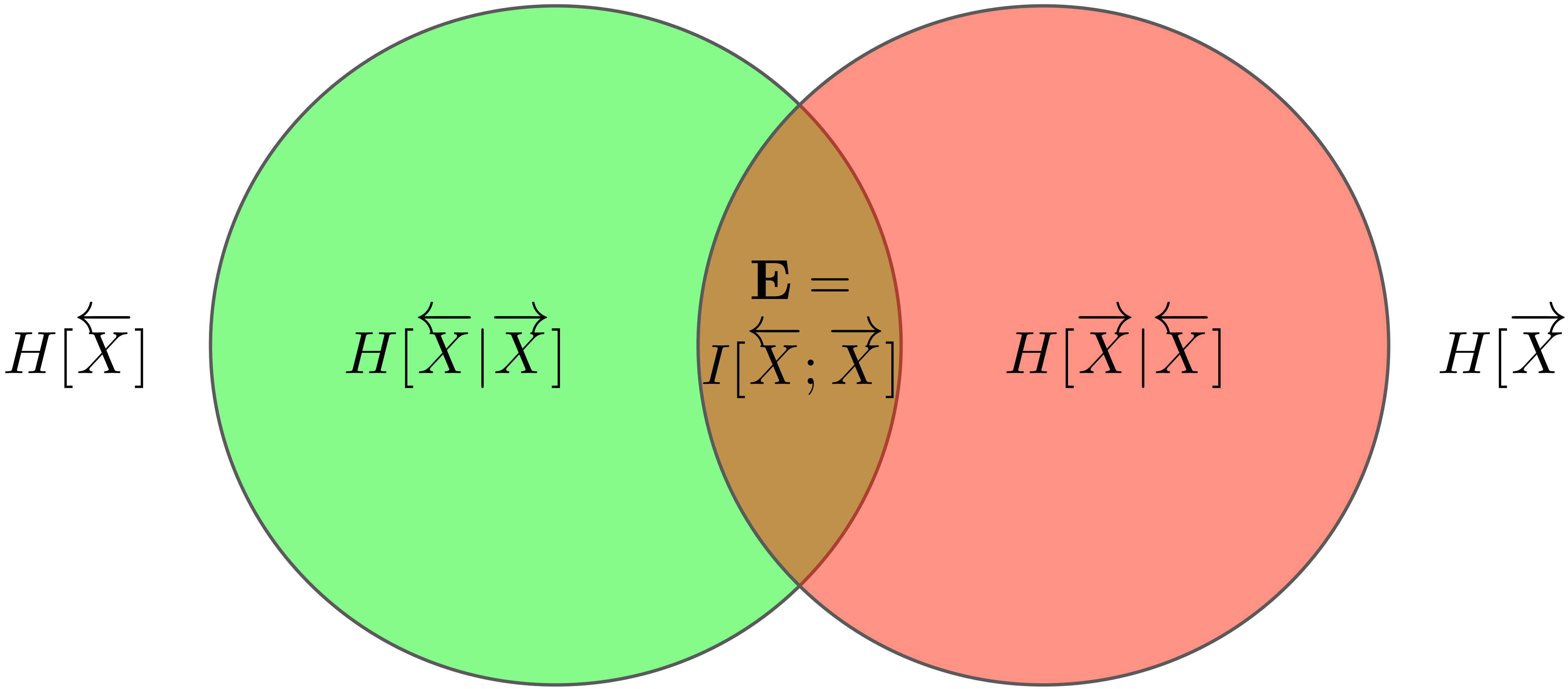
Information Diagrams for Processes

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

Information measures:

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

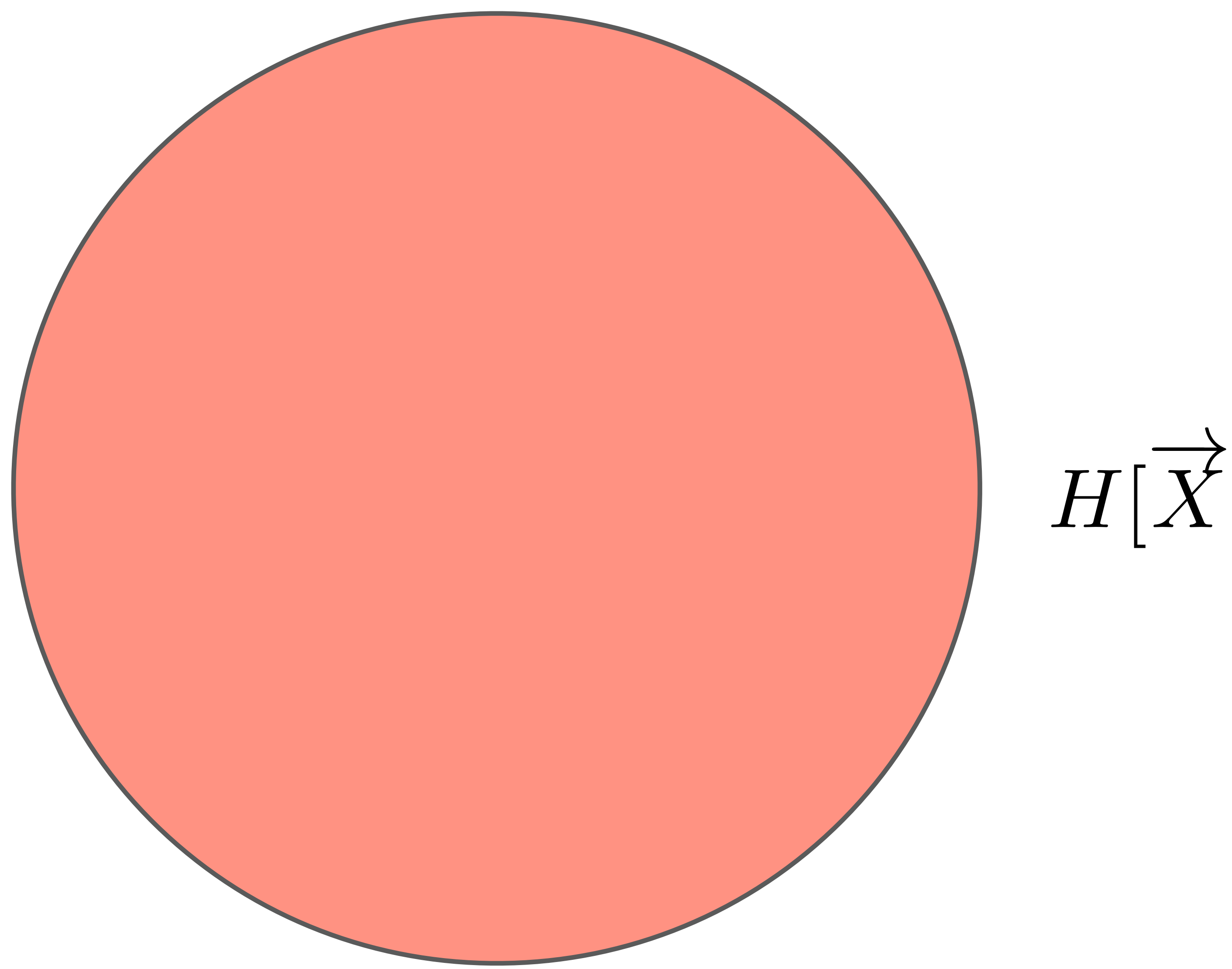
There are $8 = 2^3$ atomic information measures.

Information Diagrams for Processes

Process I-diagram using ε-machine ...

Information Diagrams for Processes

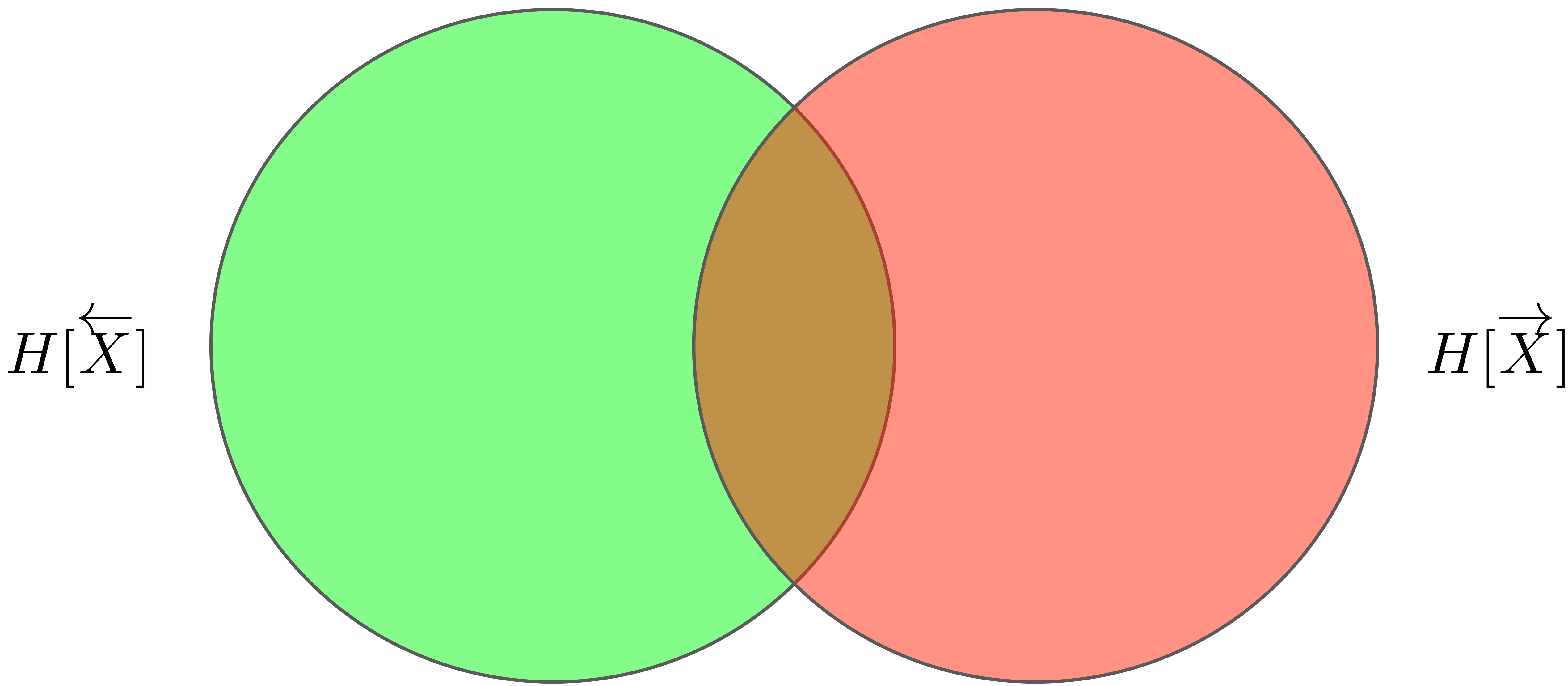
Process I-diagram using ε-machine ...



Complexity Lecture 2: Intrinsic Computation (CSSS 2013) Jim Crutchfield

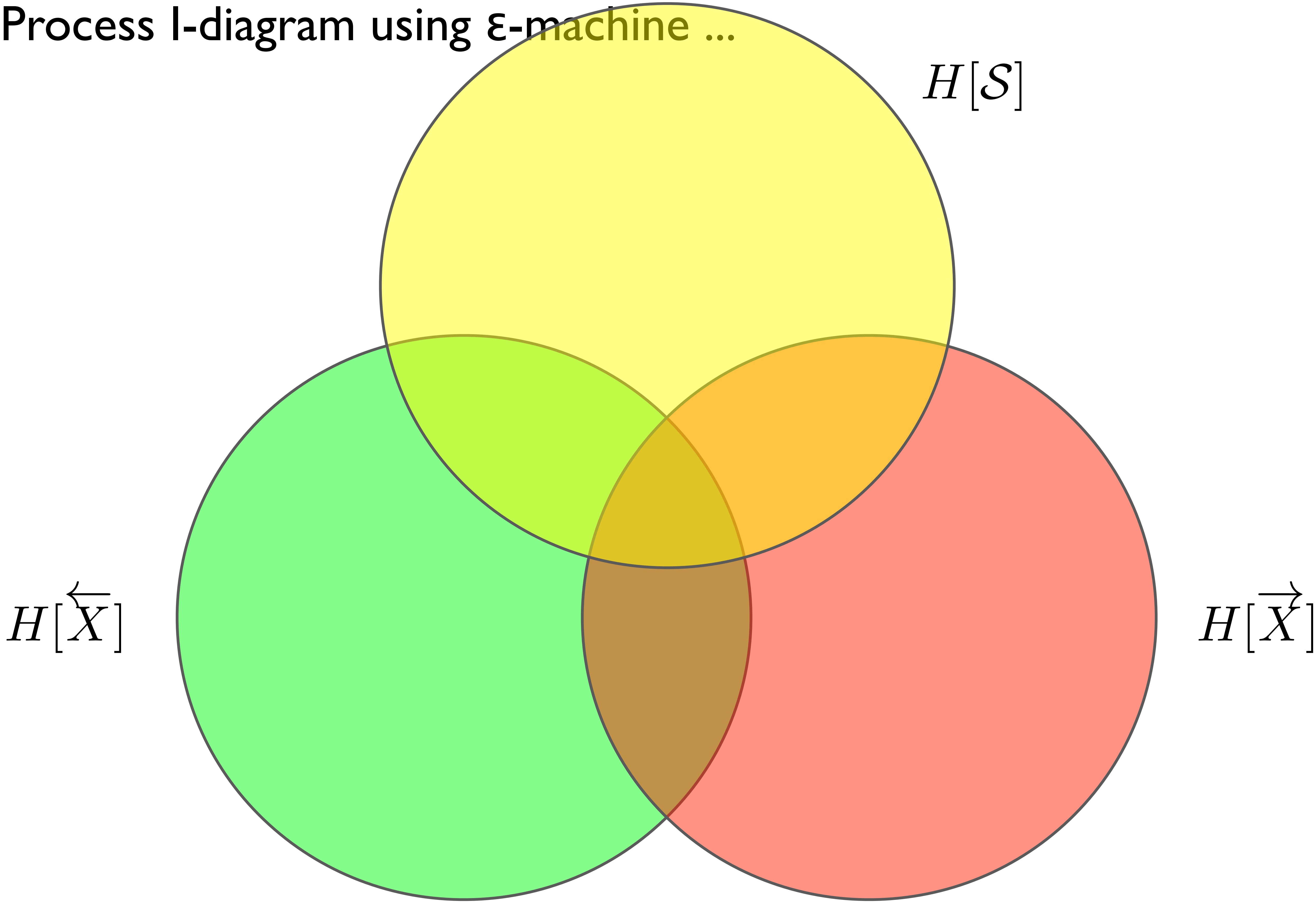
Information Diagrams for Processes

Process I-diagram using ε-machine ...



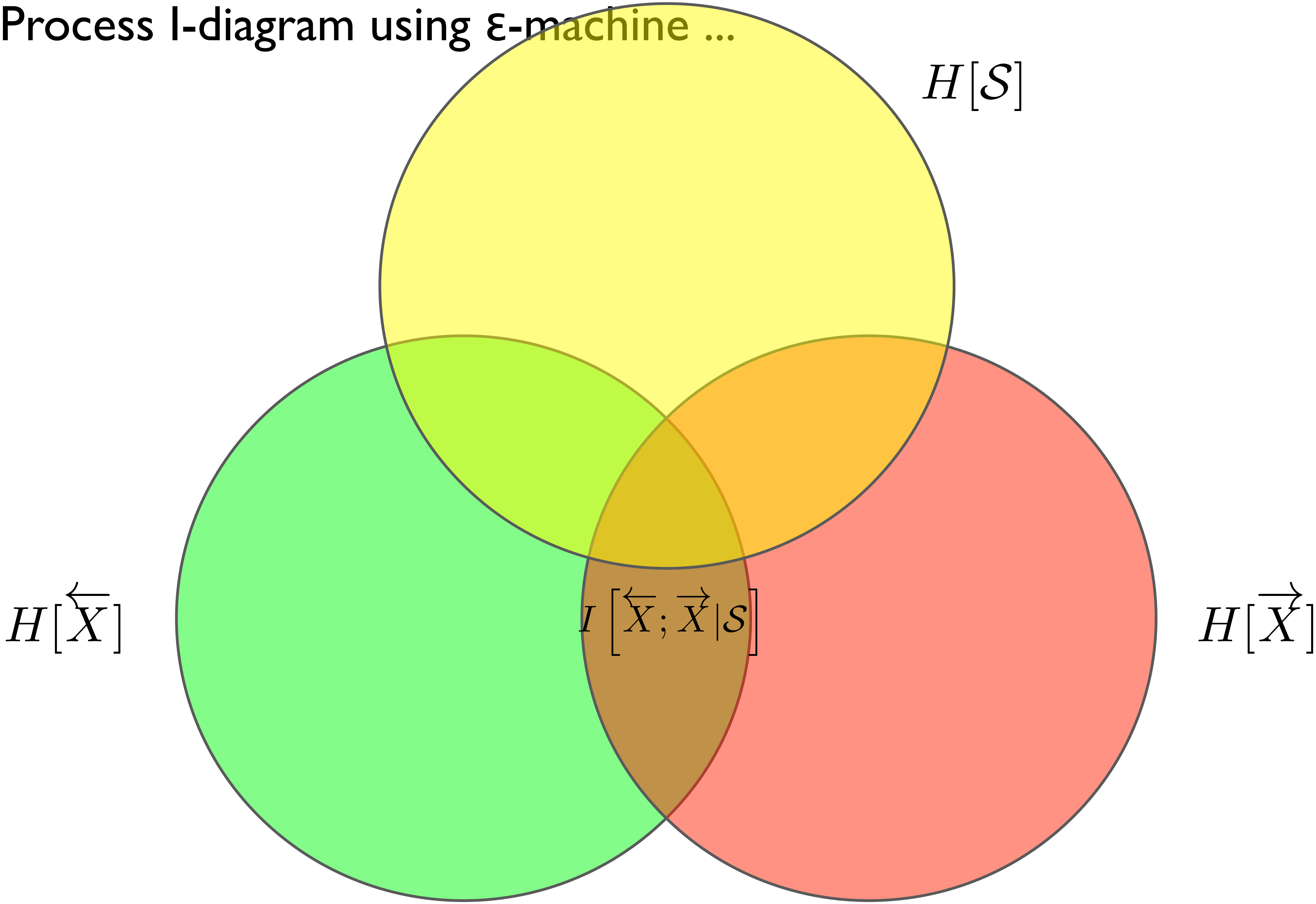
Information Diagrams for Processes

Process I-diagram using ε-machine ...



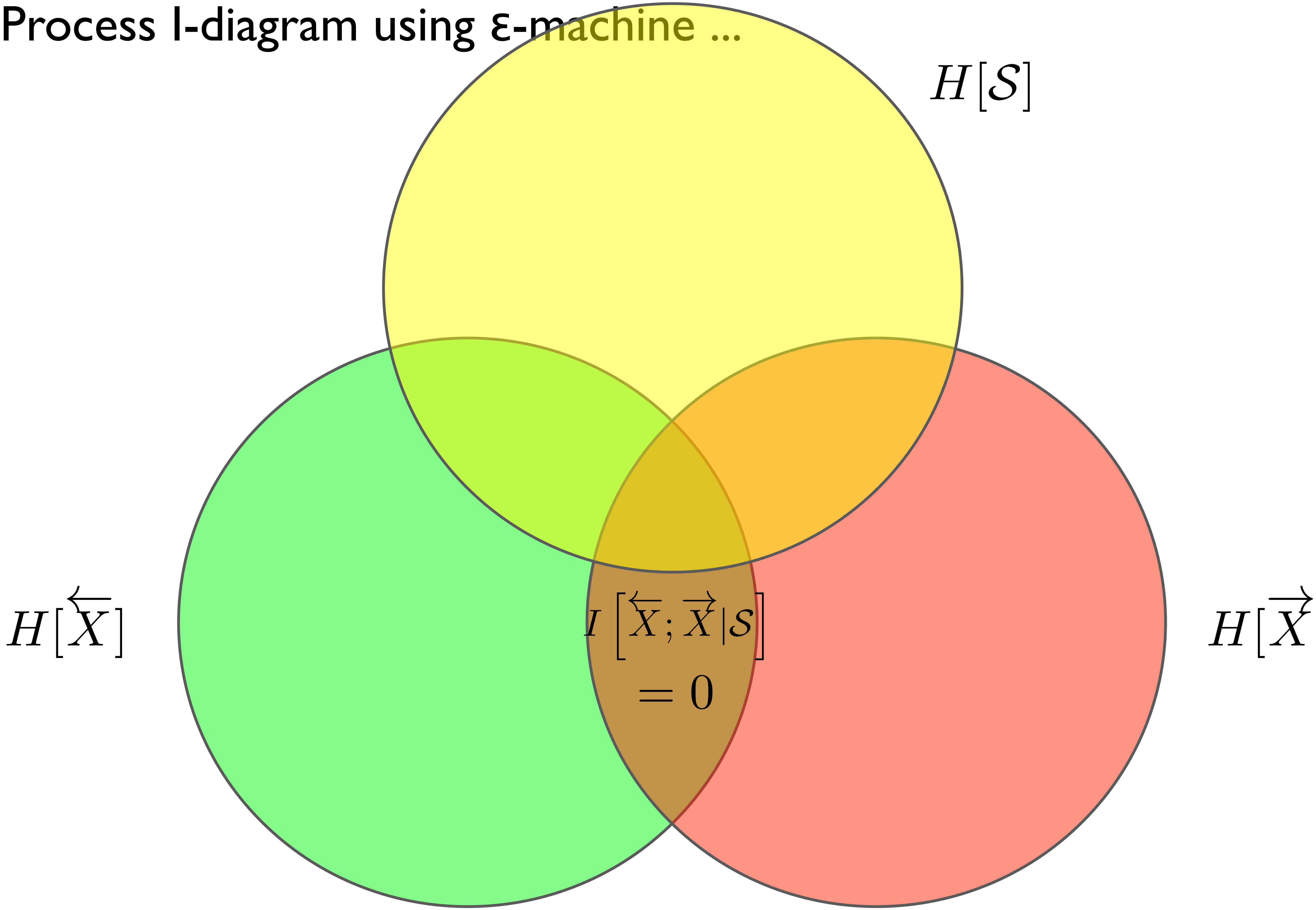
Information Diagrams for Processes

Process I-diagram using ε-machine ...



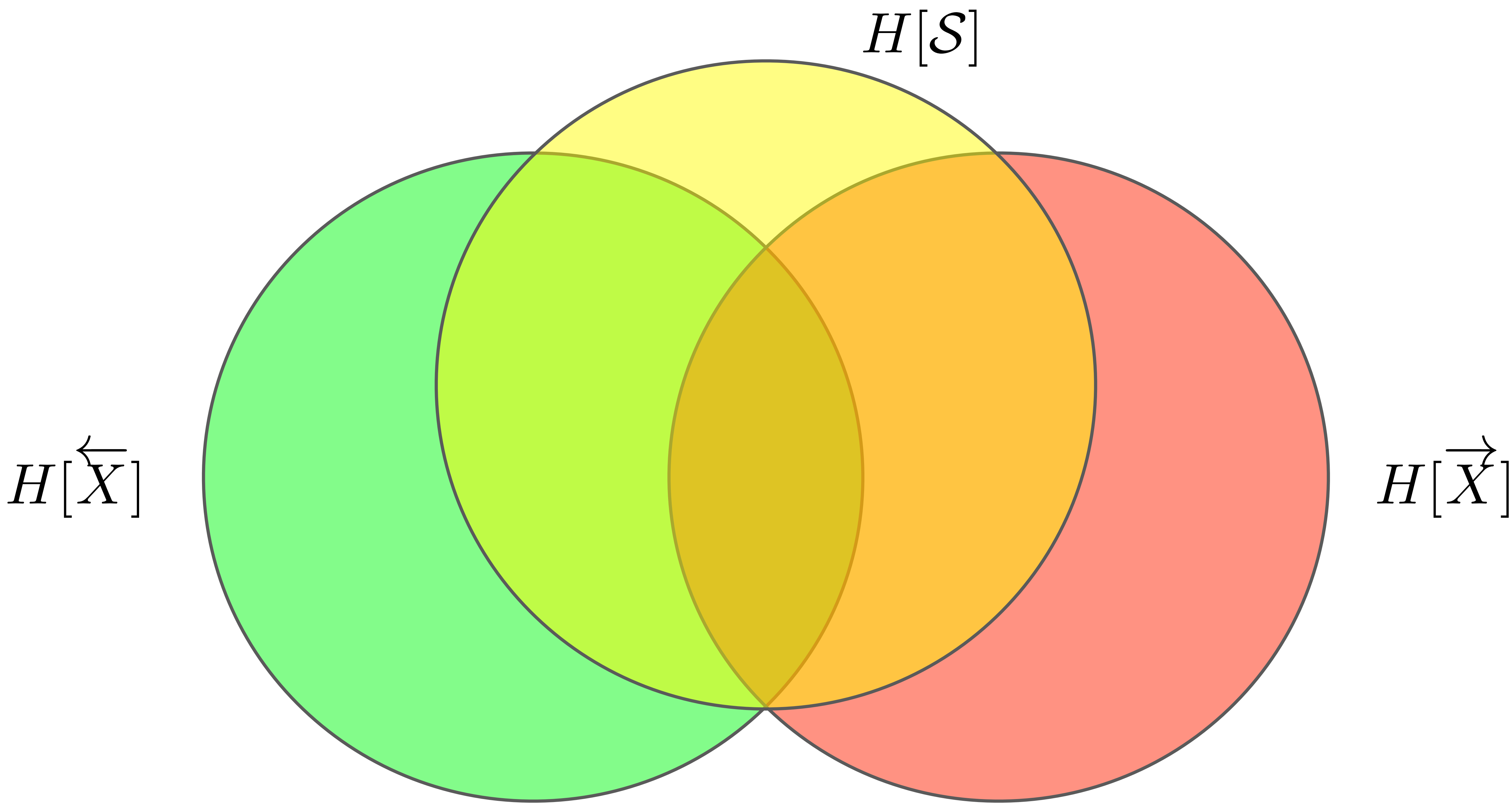
Information Diagrams for Processes

Process I-diagram using ε-machine ...



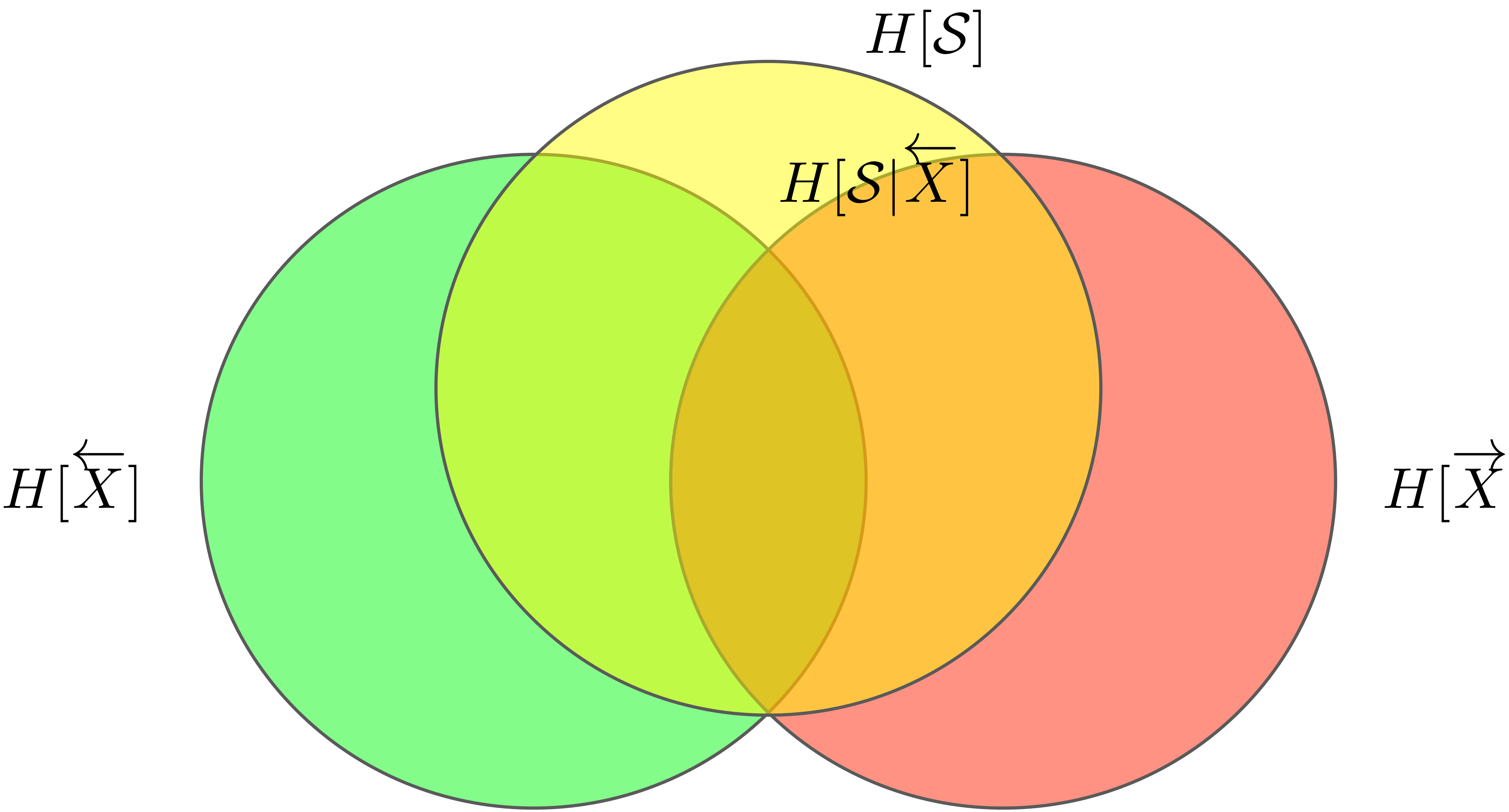
Information Diagrams for Processes

Process I-diagram using ε-machine ...



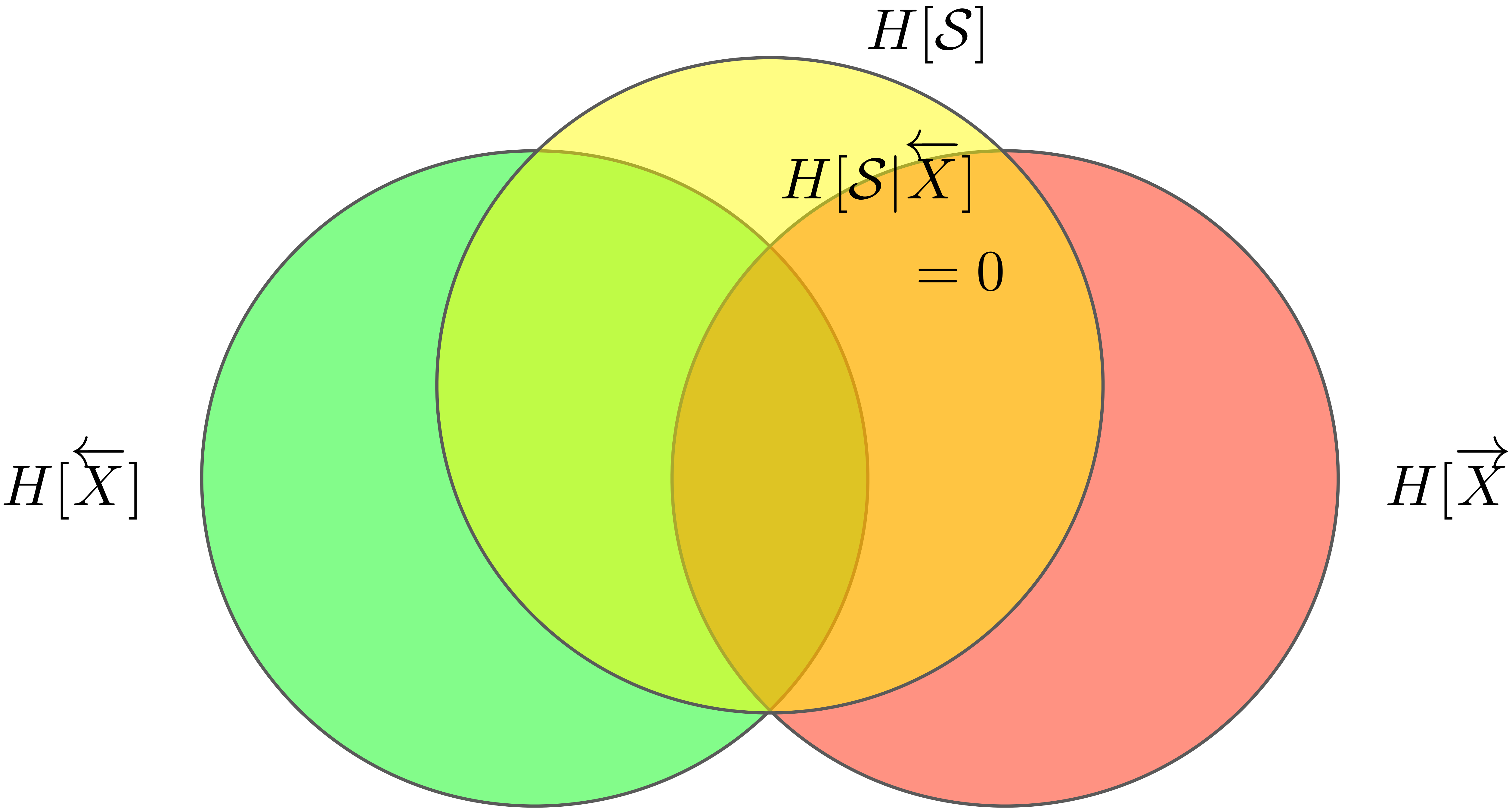
Information Diagrams for Processes

Process I-diagram using ε-machine ...



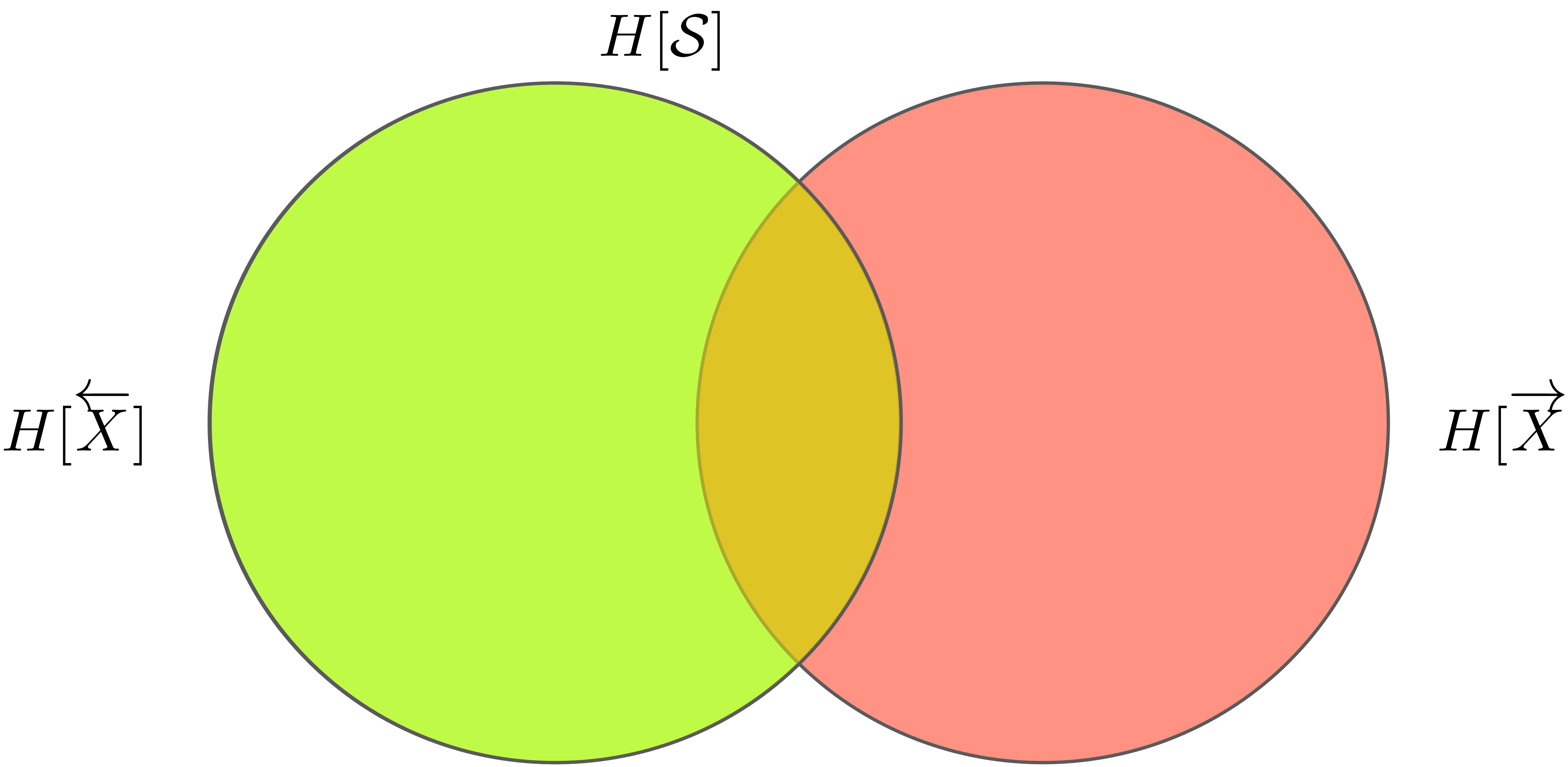
Information Diagrams for Processes

Process I-diagram using ε-machine ...



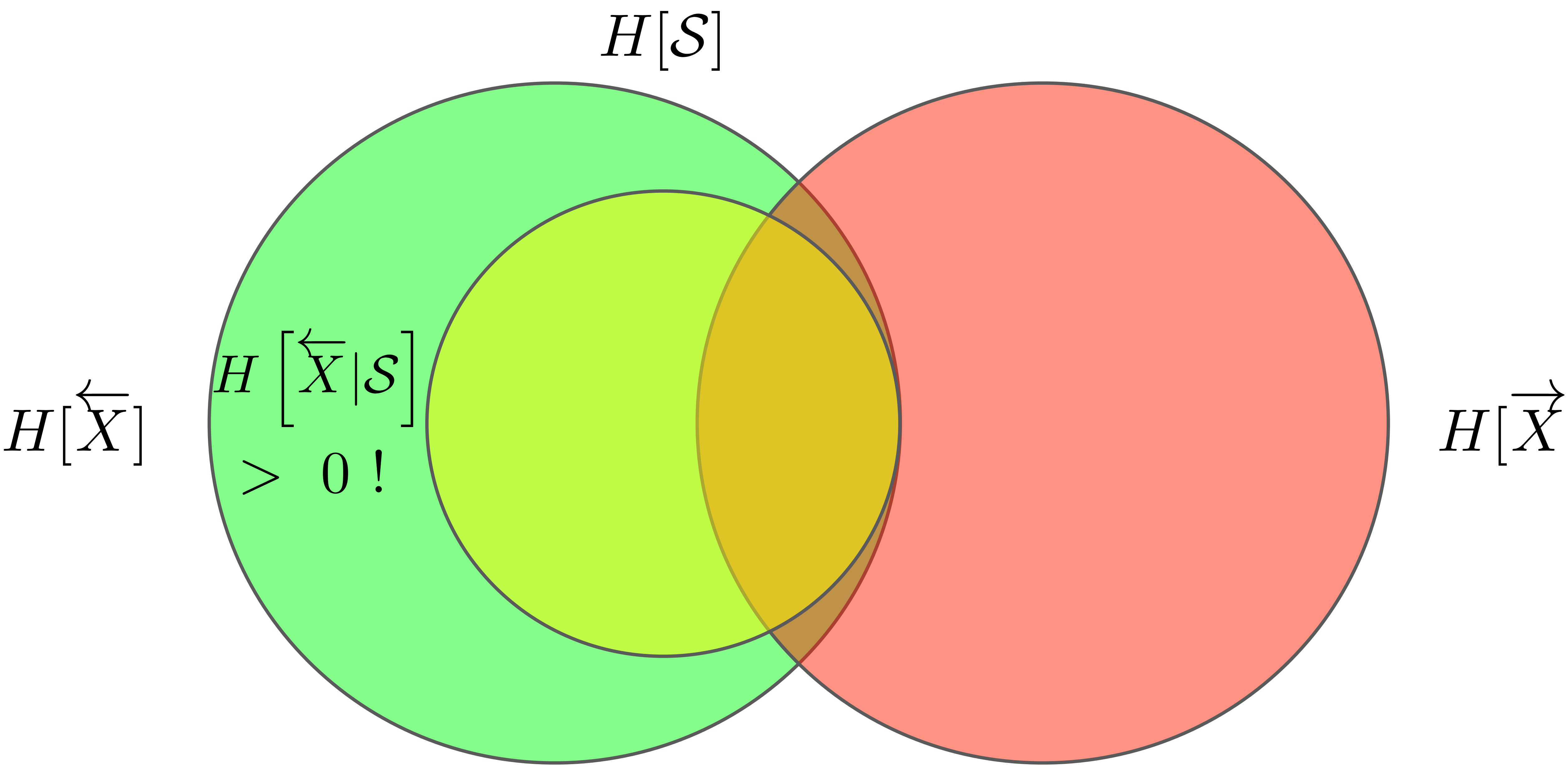
Information Diagrams for Processes

Process I-diagram using ε-machine ...



Information Diagrams for Processes

Process I-diagram using ε-machine ...



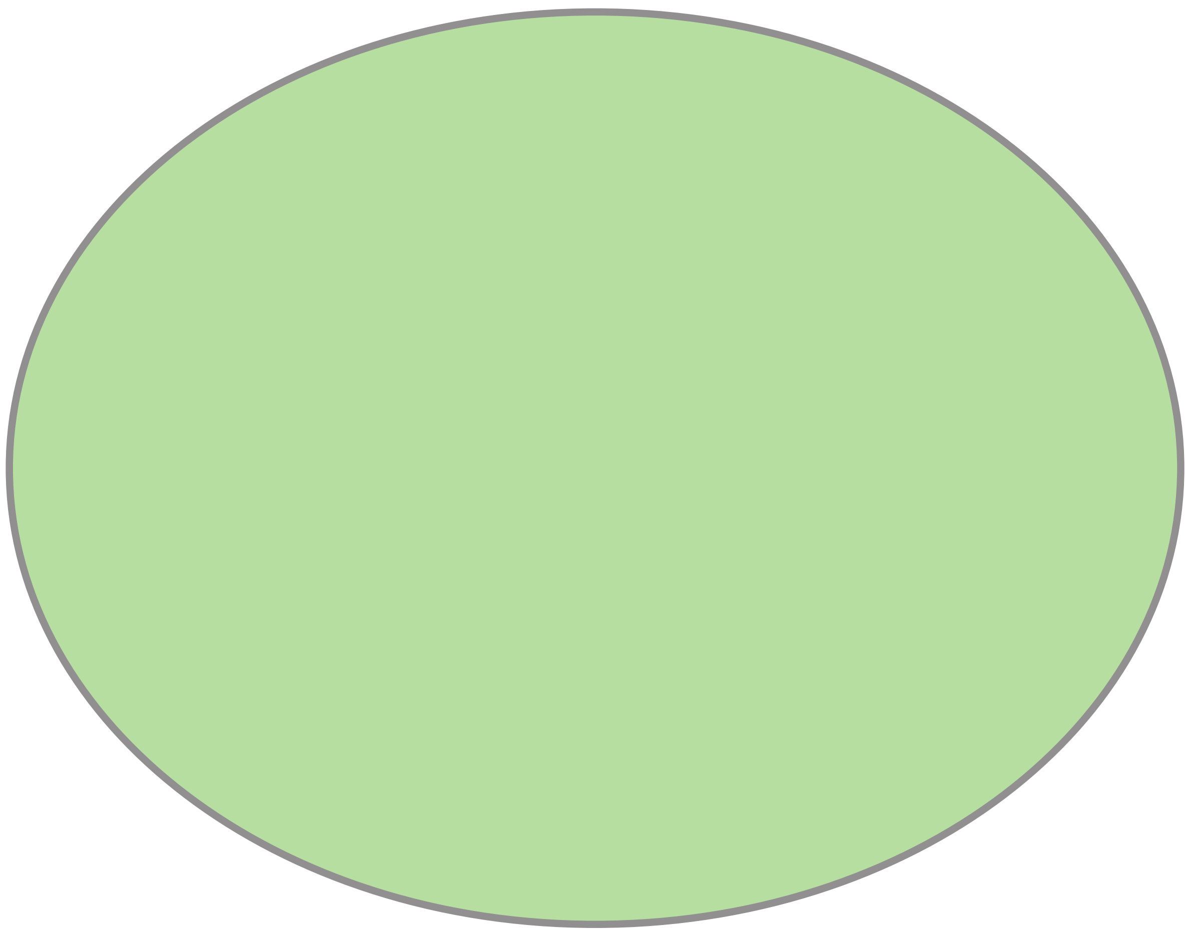
Information Diagrams for Processes

ε-Machine I-diagram:

Information Diagrams for Processes

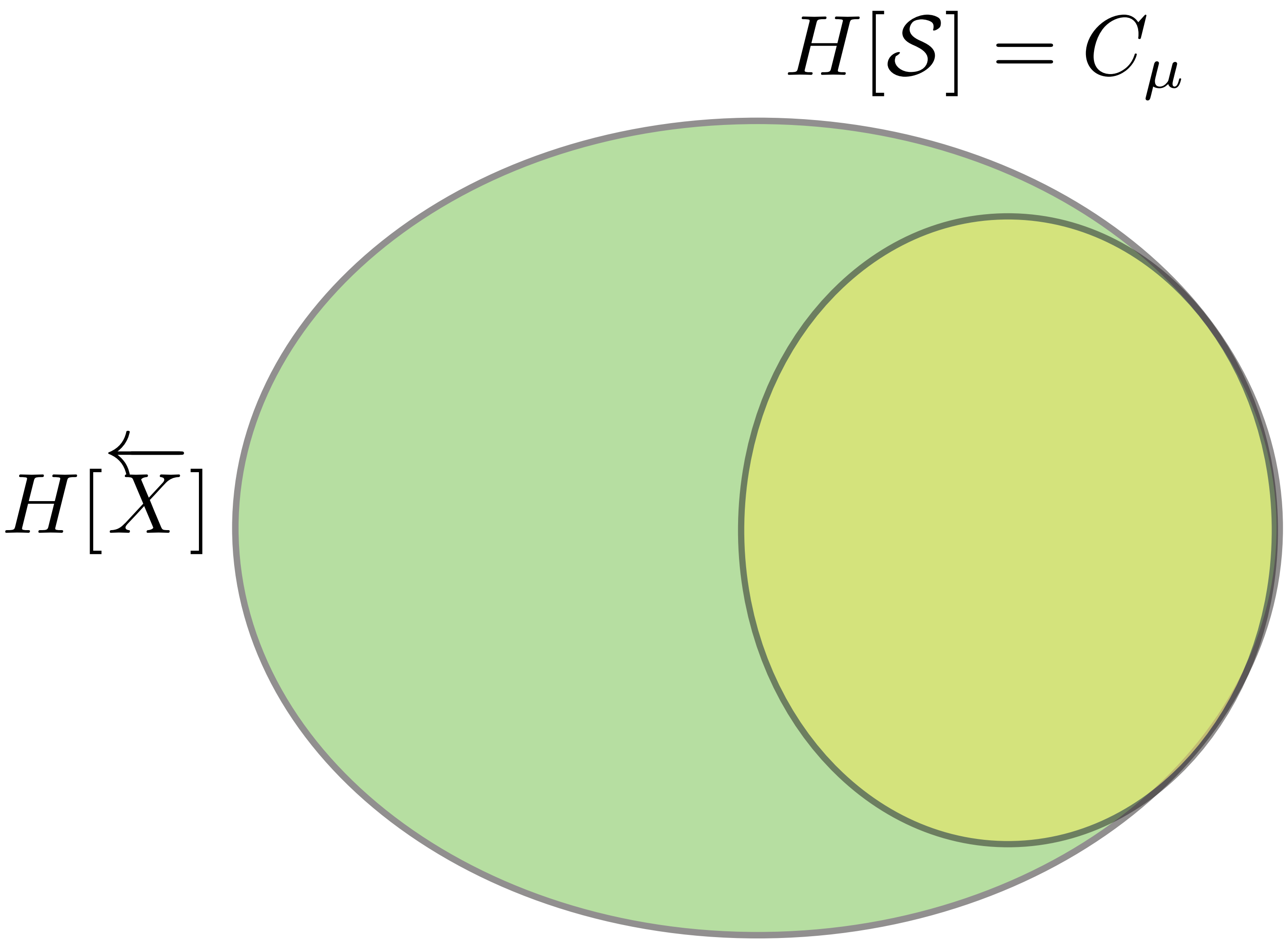
ε-Machine I-diagram:

$H[\overleftarrow{X}]$



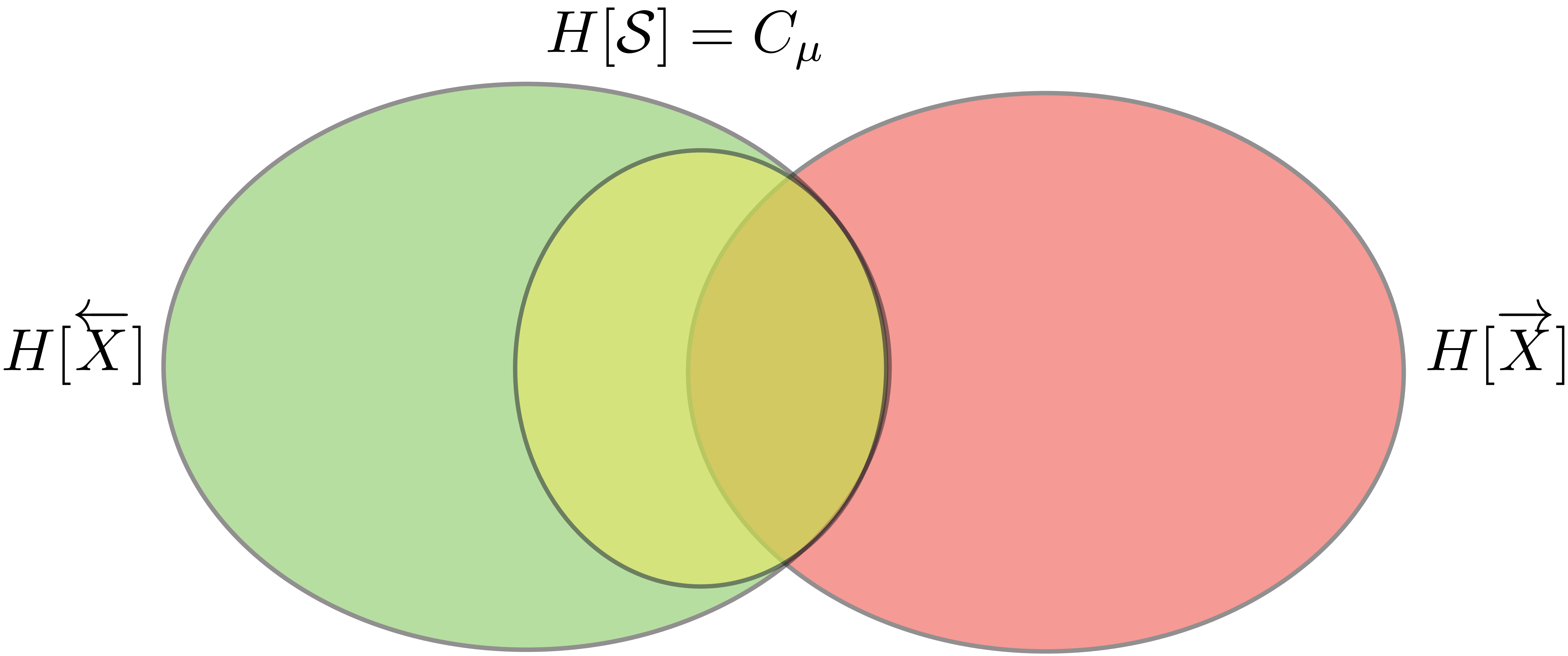
Information Diagrams for Processes

ε-Machine I-diagram:



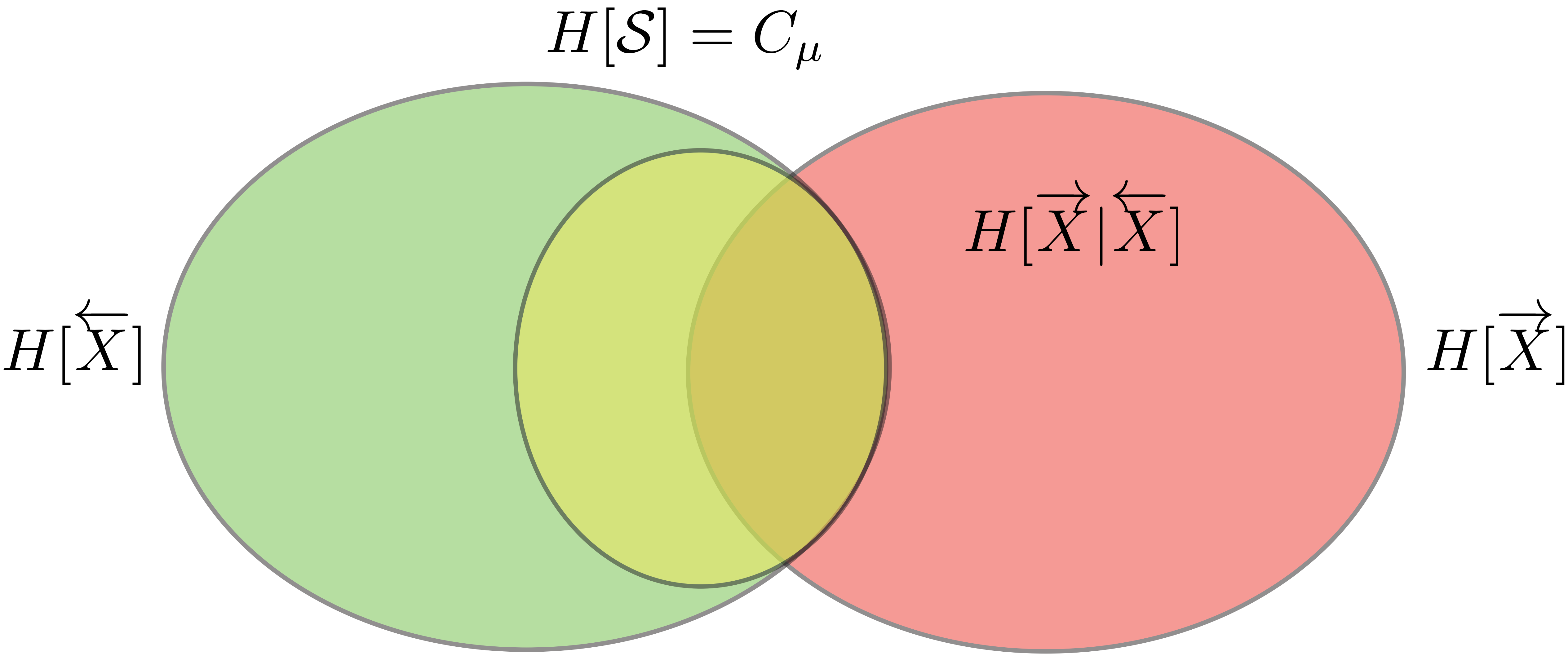
Information Diagrams for Processes

ε-Machine I-diagram:



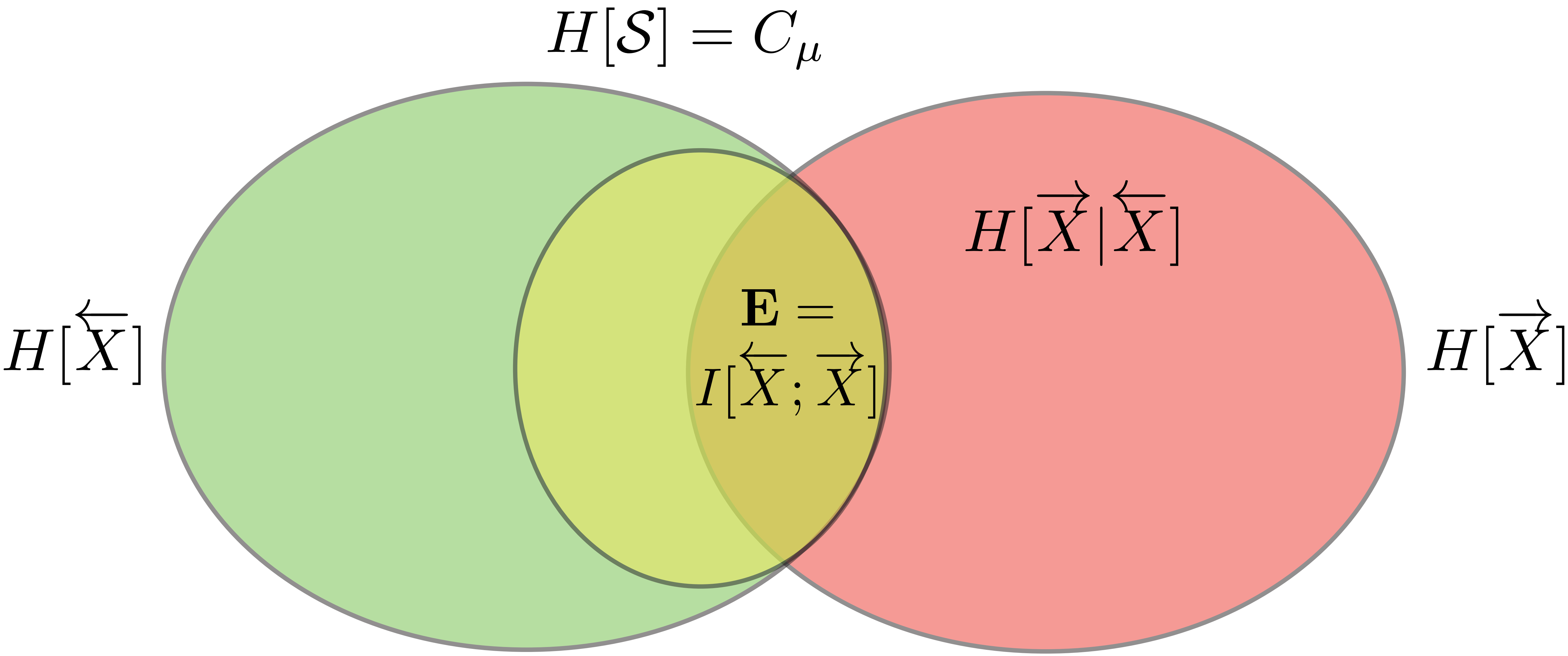
Information Diagrams for Processes

ε-Machine I-diagram:



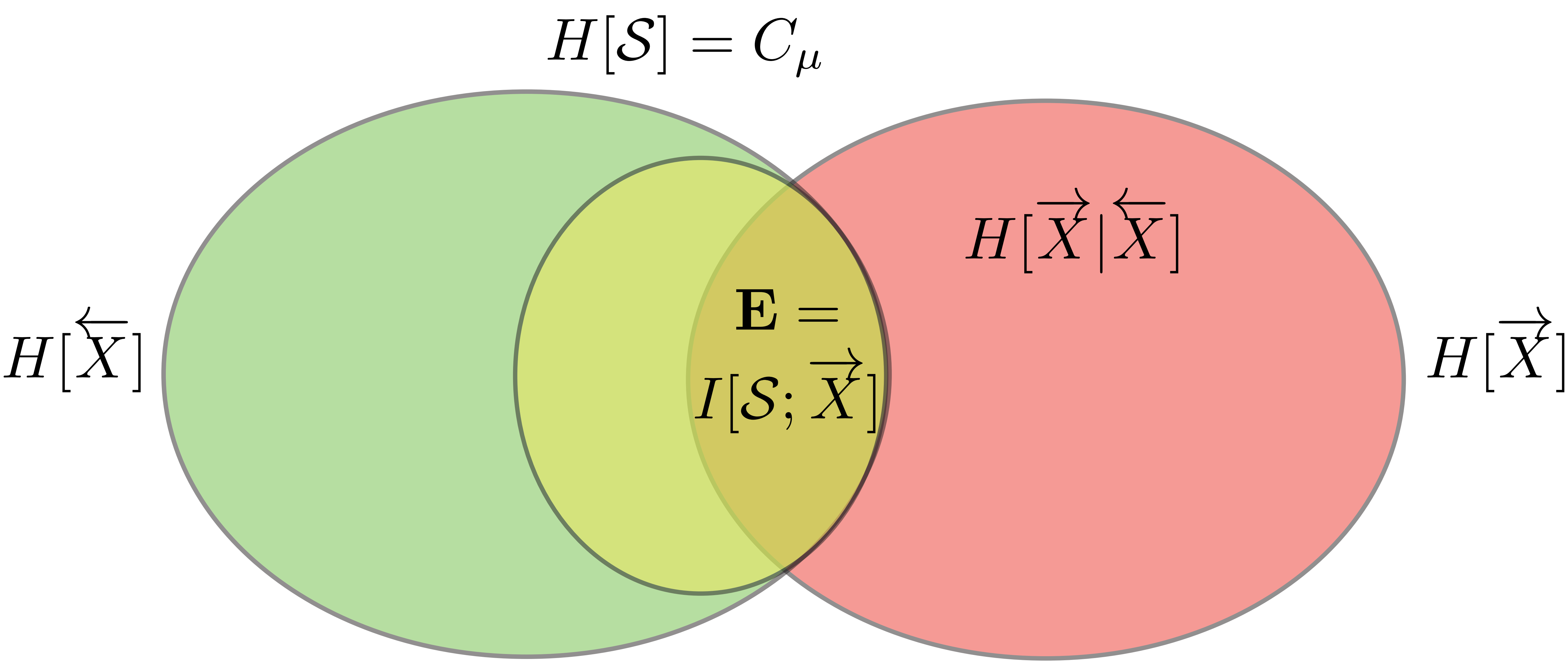
Information Diagrams for Processes

ε-Machine I-diagram:



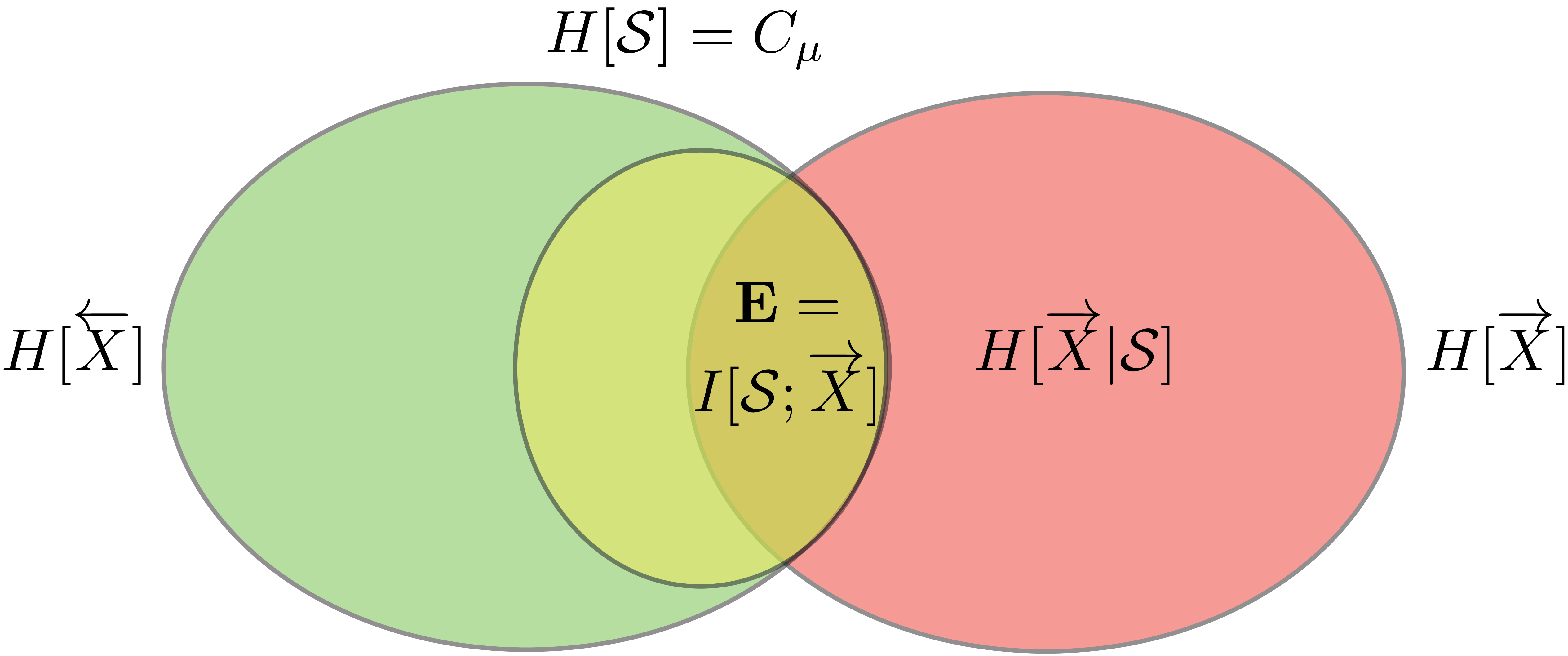
Information Diagrams for Processes

ε-Machine I-diagram:



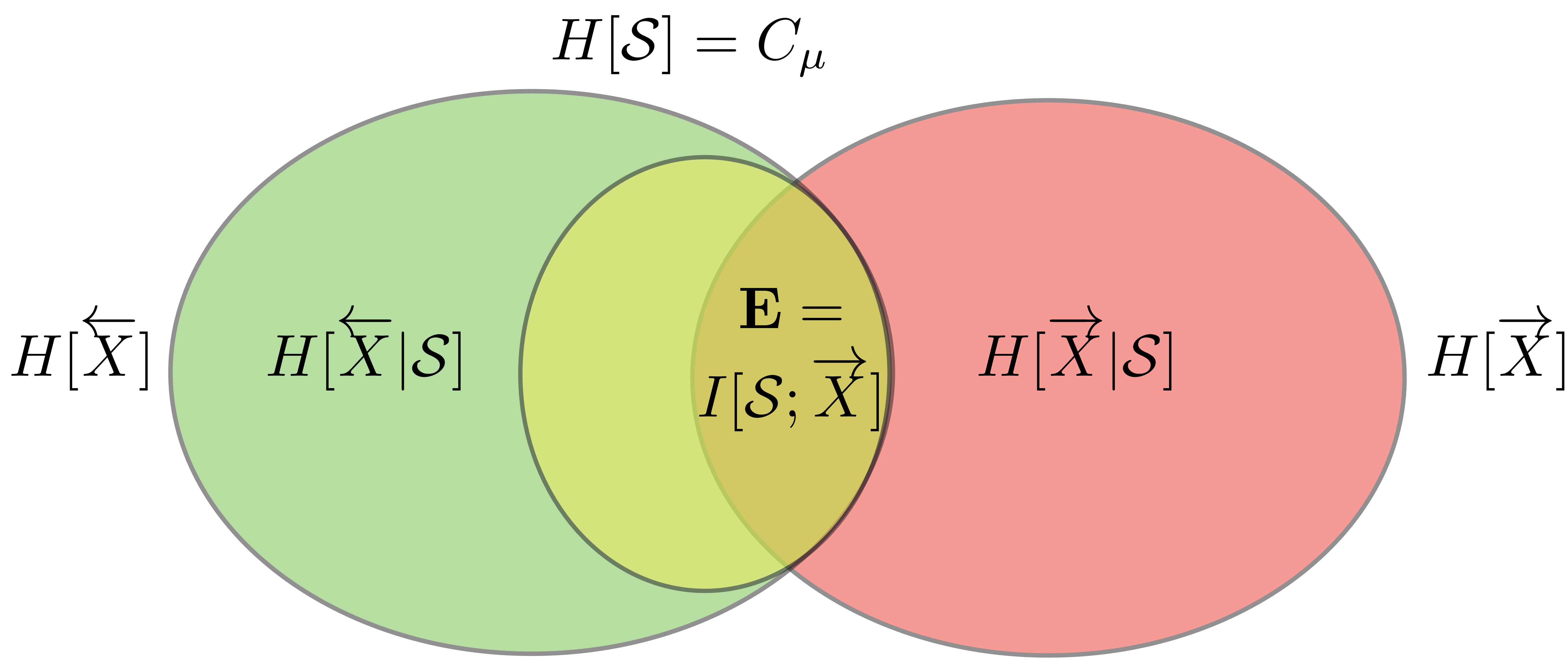
Information Diagrams for Processes

ε-Machine I-diagram:



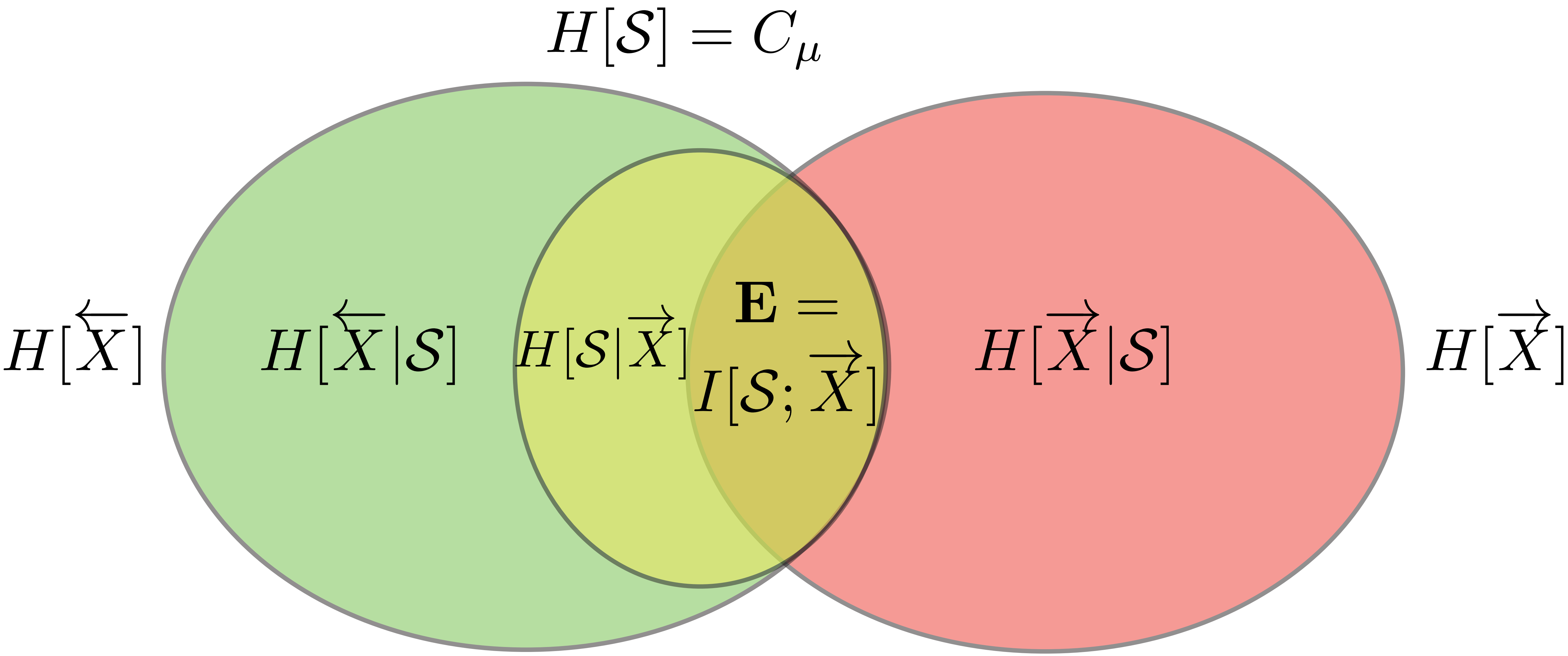
Information Diagrams for Processes

ε-Machine I-diagram:



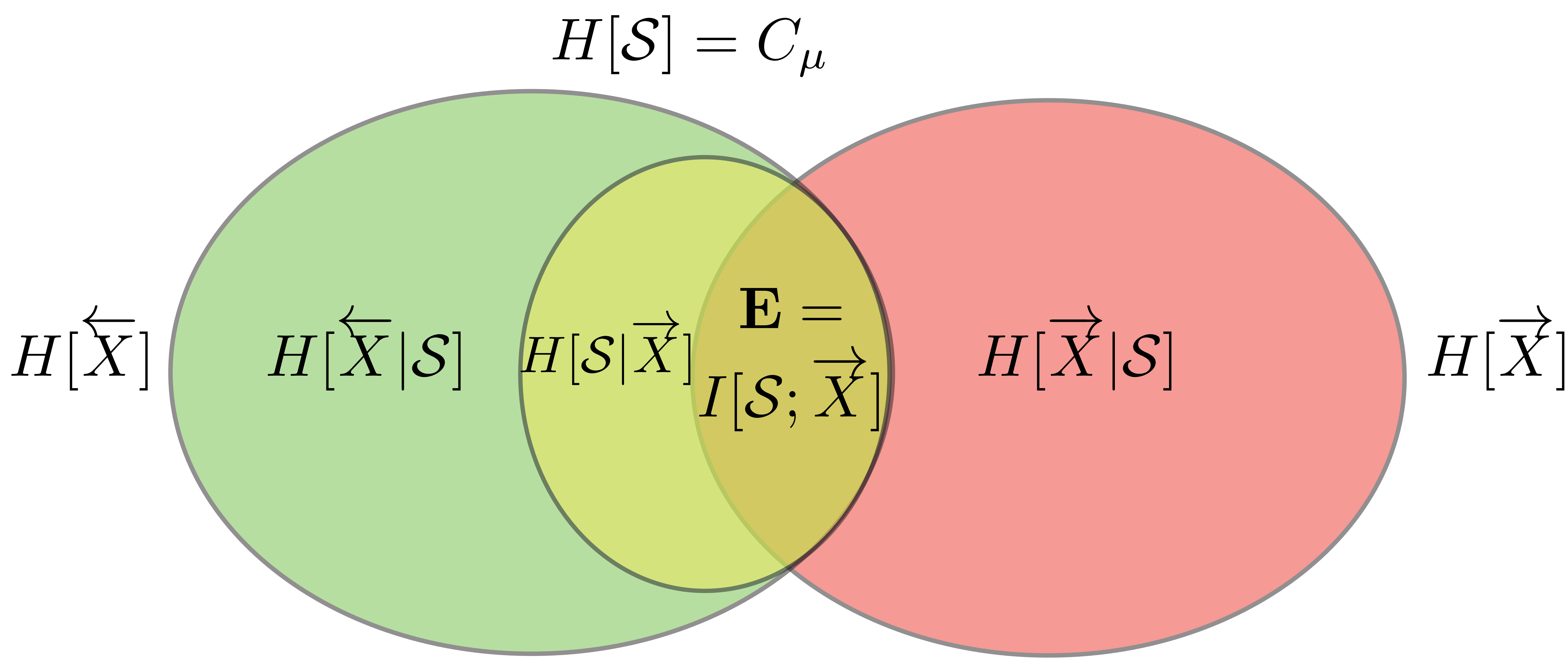
Information Diagrams for Processes

ε-Machine I-diagram:



Information Diagrams for Processes

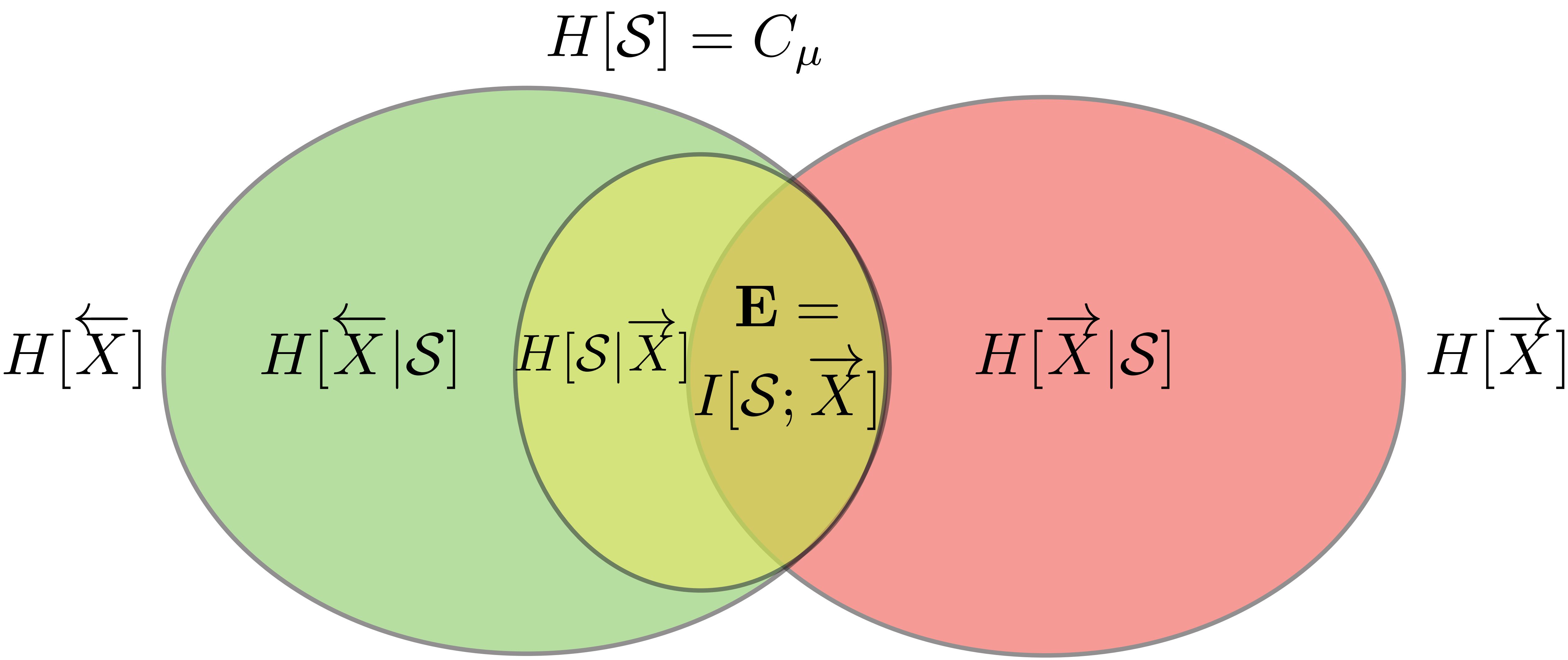
ε-Machine I-diagram:



What do we know about this?

Information Diagrams for Processes

ε-Machine I-diagram:



What is this Mystery Wedge?

What do we know about this?

Information Diagrams for Processes

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

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

Proof Sketch:

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

Information Diagrams for Processes

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

Uncertainty of causal state given future. Implications?

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

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

Information Diagrams for Processes

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

Uncertainty of causal state given future. Implications?

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

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

I am the Mystery Wedge!

Information Diagrams for Processes

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

Wedge is the inaccessibility of hidden state information!

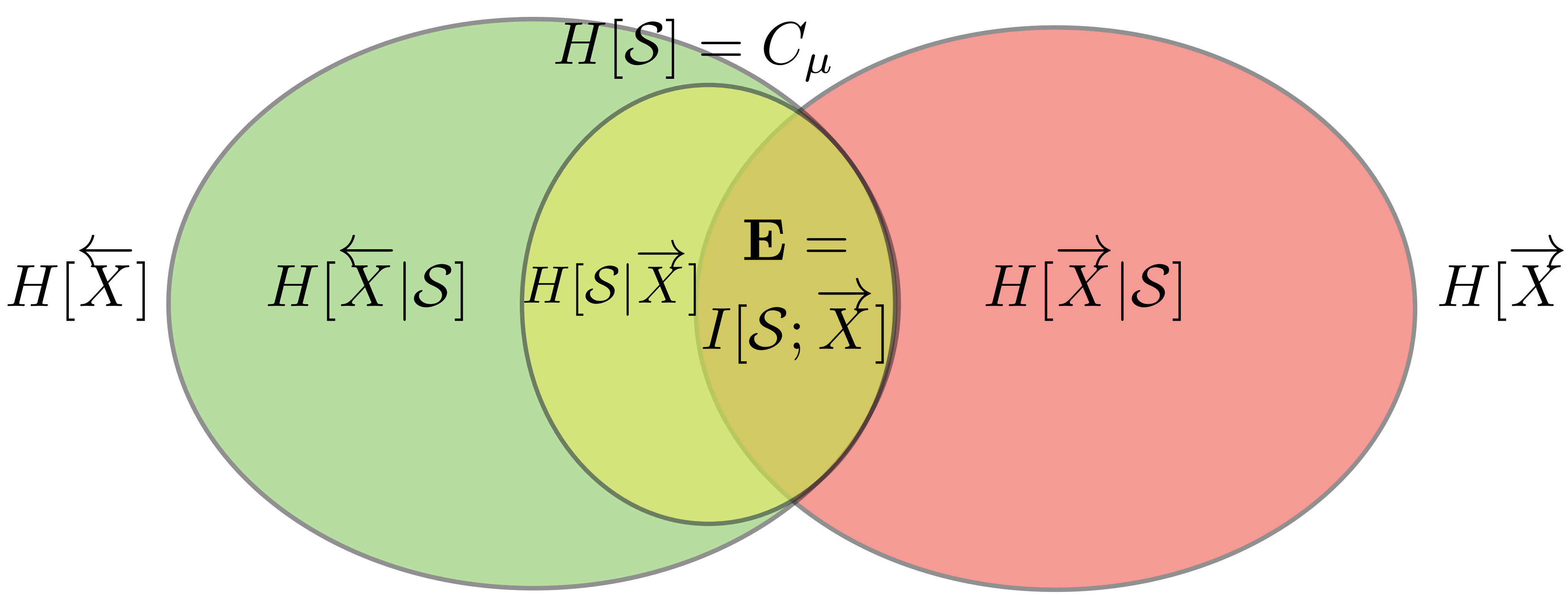
$$H[\mathcal{S}|\overrightarrow{X}] = C_\mu - \mathbf{E}$$

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

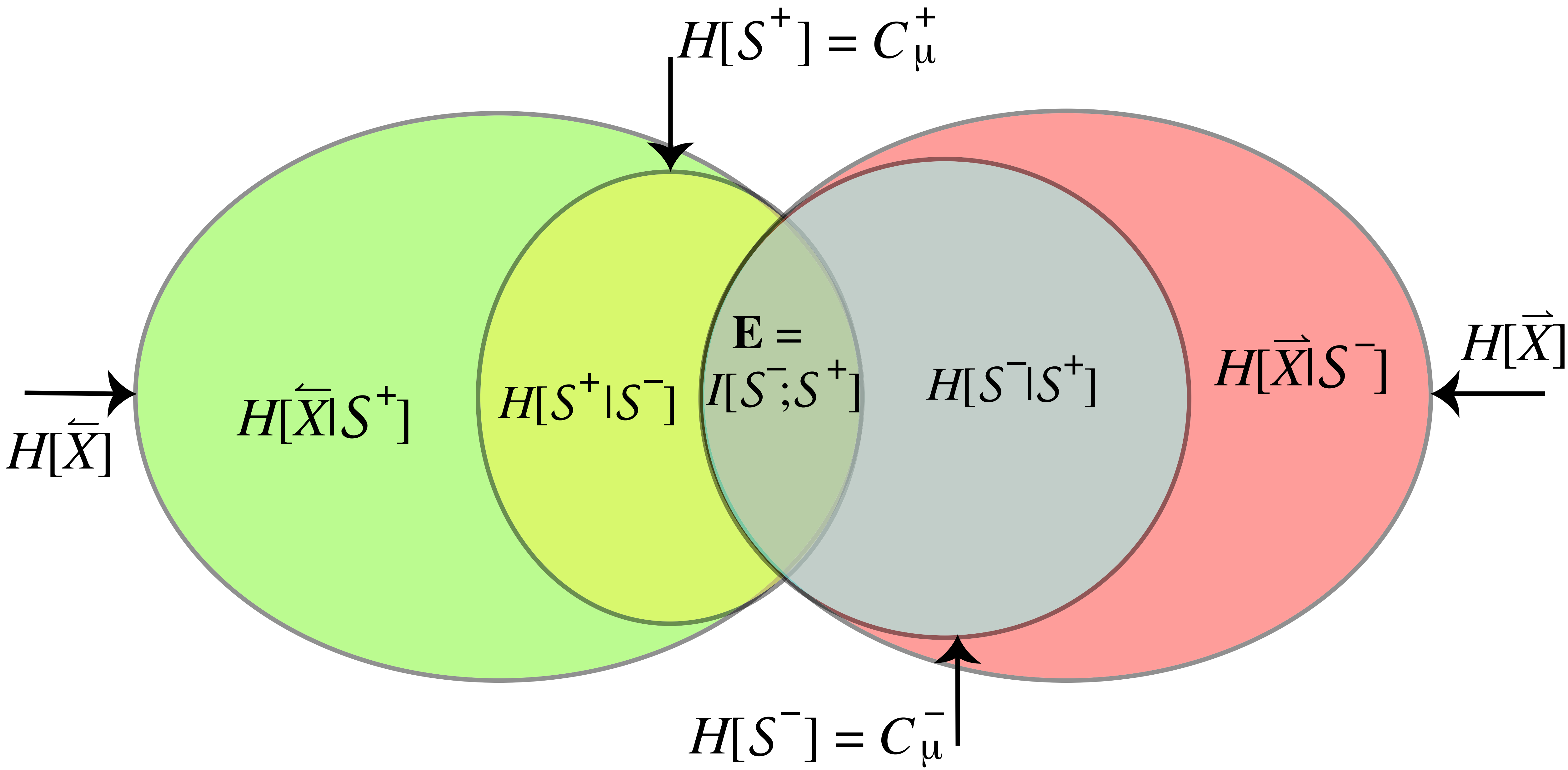
- Theorem:

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

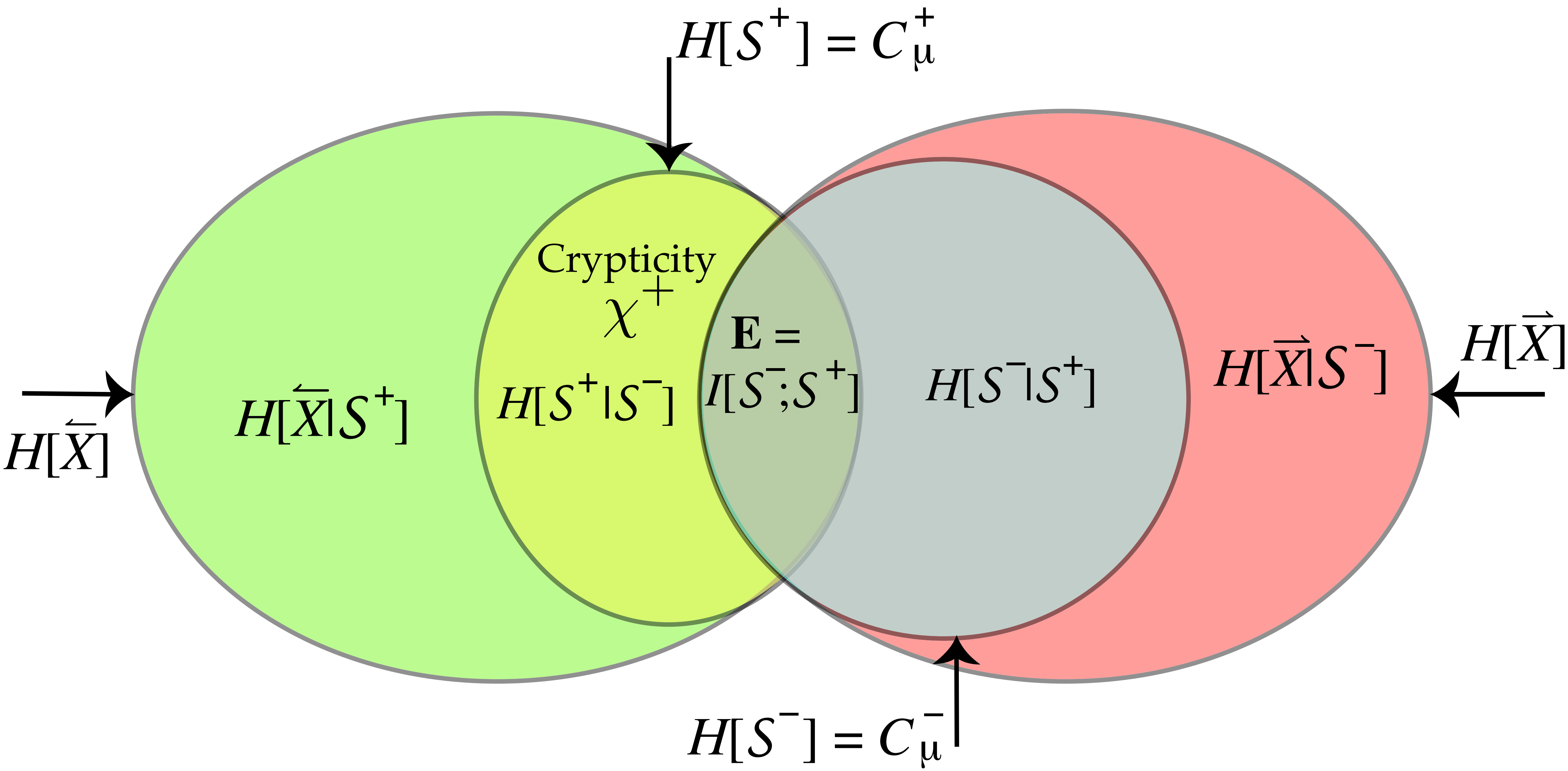
- Effective transmission capacity of channel between forward and reverse processes.
- Time agnostic representation: The **BiMachine**.

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.

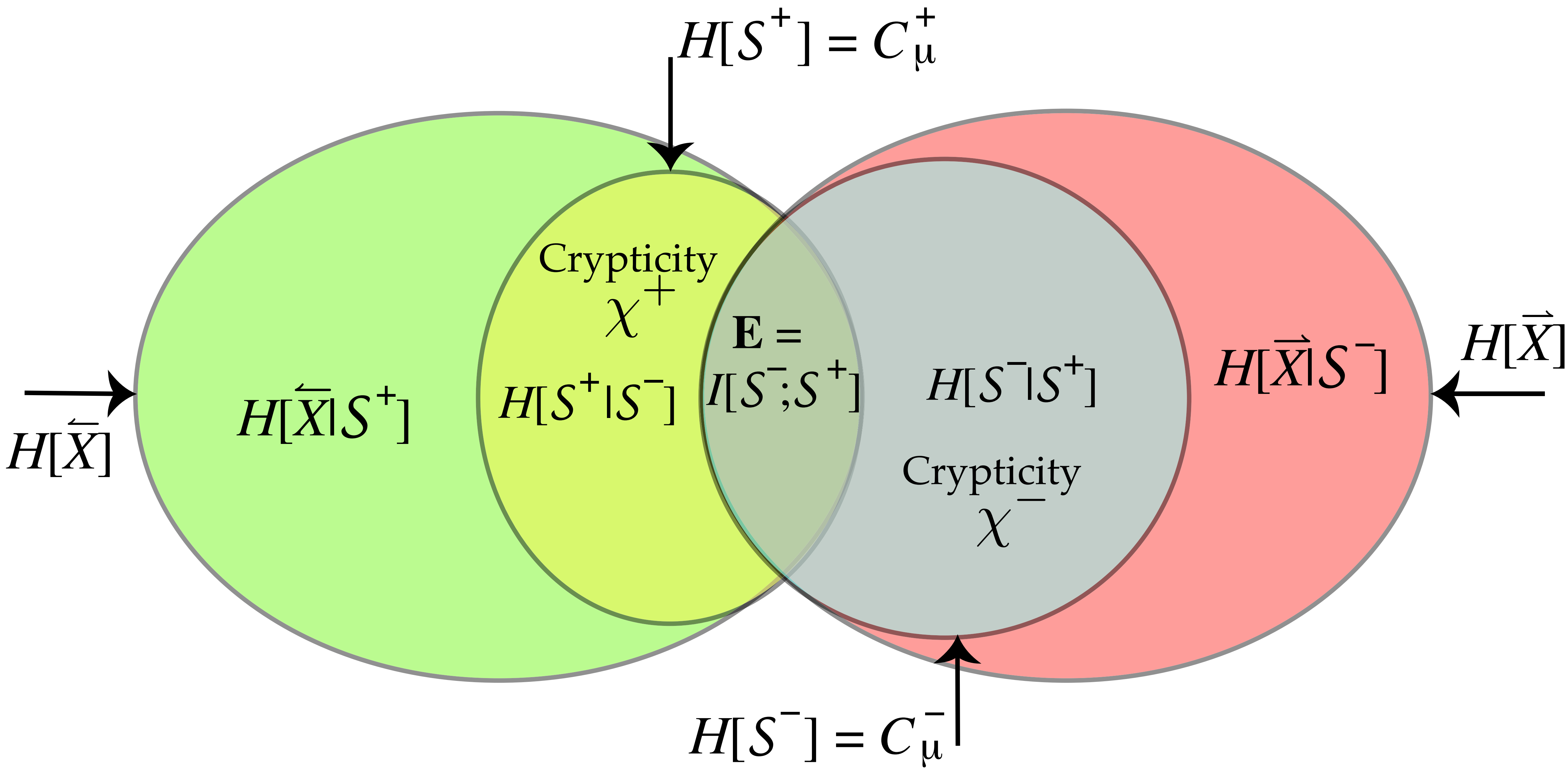
Forward-Reverse ε-Machine Information Diagram



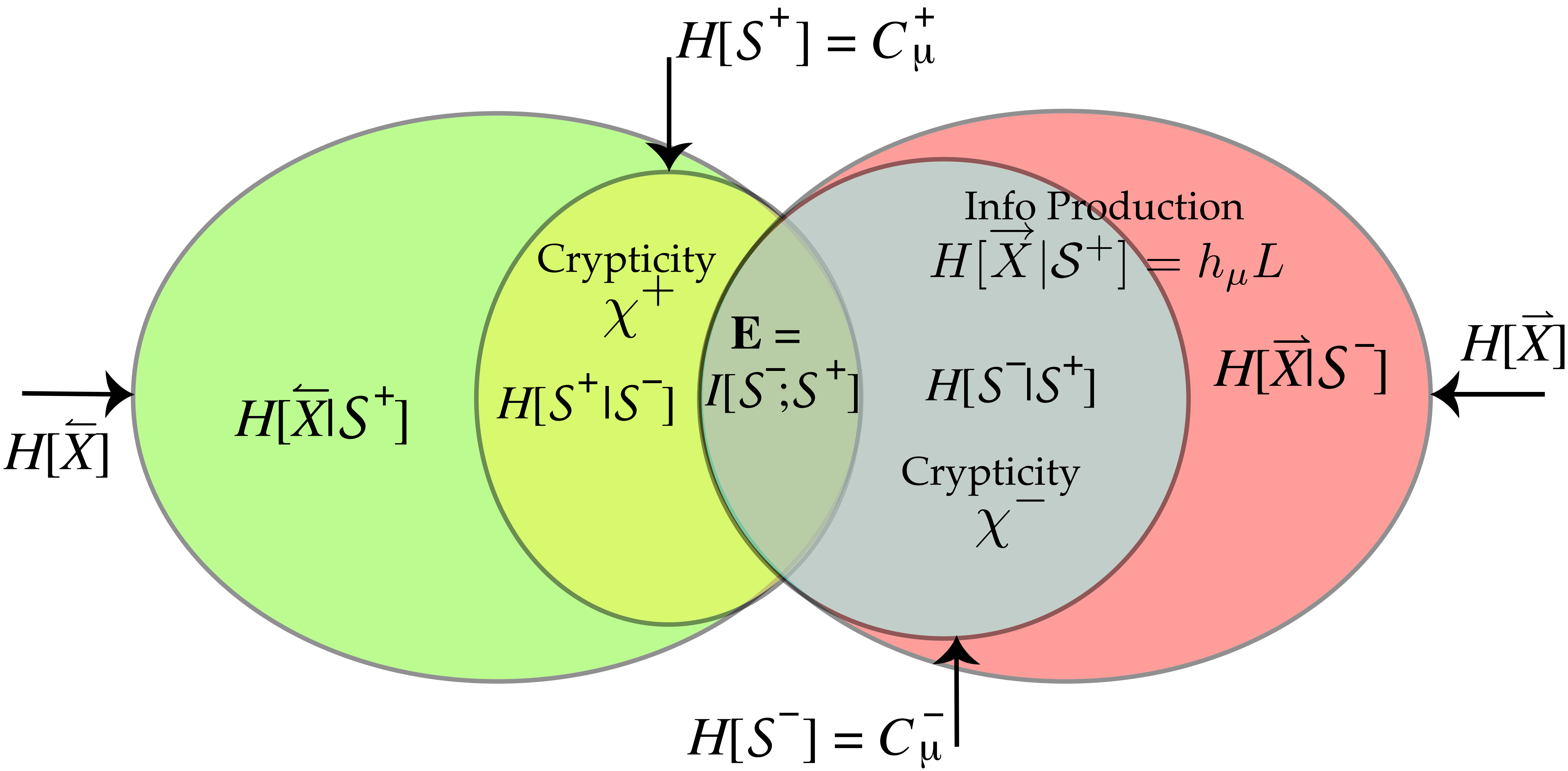
Forward-Reverse ε-Machine Information Diagram



Forward-Reverse ε-Machine Information Diagram

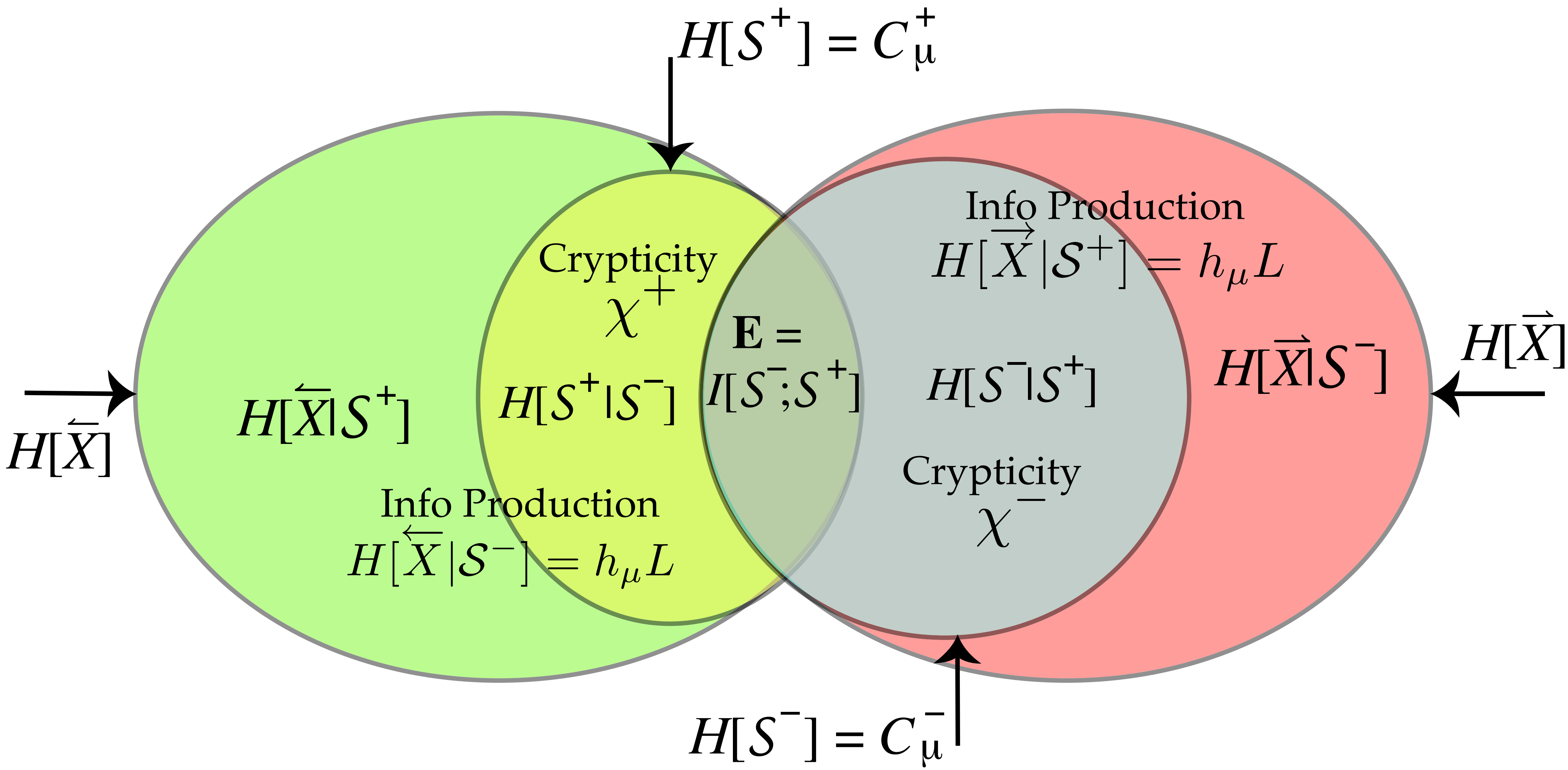


Forward-Reverse ε-Machine Information Diagram



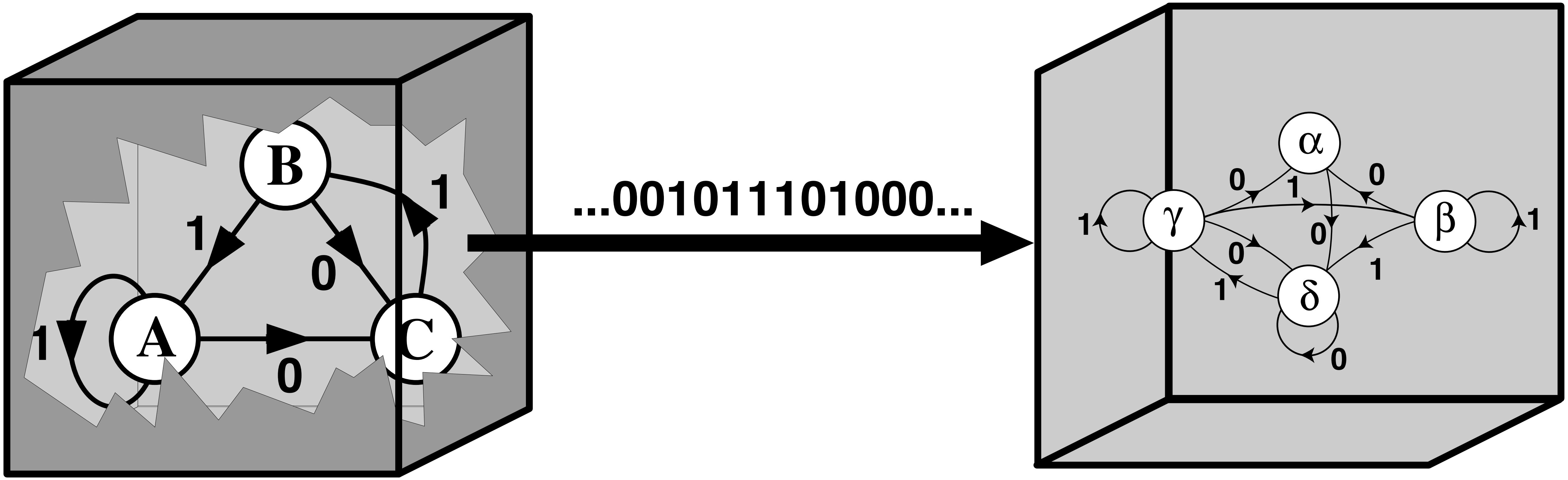
Complexity Lecture 2: Intrinsic Computation (CSSS 2013) Jim Crutchfield

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:

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

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

How random?

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

How structured?

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

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

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: $\chi > 0$

Complexity!

Thursday: Information Theory for Complex Systems

- Complex Processes
- Information & Memory in Processes
- Interactive Labs: Nix

Friday: Intrinsic Computation

- Measuring Structure
- Intrinsic Computation
- Optimal Models
- Interactive Labs: Nix

See online course:

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