Computational Mechanics: A Primer

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

Tutorial Presented at
Unconventional Computation 2009
8th International Conference
Ponta Delgada (Azores), Portugal

7, 8, & 10 September 2009

Agenda
In Three Lectures

Goal: Intrinsic Computation

- Lecture 1: Motivations, Examples, & Setting
- Lecture 2: Information in Processes
- Lecture 3: Computational Mechanics

Lecture 1: Motivations, Examples, & Setting

- Randomness
- Pattern
- Learning
- Emergence
- Examples
- Mathematical Nesting

Lecture 2: Information In Processes

- Shannon Information
- Processes
- Degrees of Randomness
- Measures of Organization
- Information Hierarchy
- Limitations of Information Theory

Lecture 3: Computational Mechanics

- Pattern
- Causal Architecture
- Intrinsic Computation
- Modeling as Decryption

Lessons

- Can define & measure degree of organization
- Emergence = Increase in stored information
- How natures computes is how nature is structured

References

- Information Theory:

  - C. E. Shannon, “A Mathematical Theory of Communication”, Bell Systems Technical Journal **27** (1948) 379-423; 623-656.
  - C. E. Shannon, “Communication Theory of Secrecy Systems”, Bell Sys. Tech. J. **28** (1949) 656-715.
  - T. Cover and J. Thomas, **Elements of Information Theory**, Second Edition, Wiley-Interscience (2006).
  - R. W. Yeung, **Information Theory and Network Coding**, Springer (New York, 2008).
  - R. W. Yeung, **A First Course in Information Theory**, Kluwer Academic Press (New York, 2002).
  - J. P. Crutchfield and D. P. Feldman, “Regularities Unseen, Randomness Observed: Levels of Entropy Convergence”, CHAOS **13**:1 (2003) 25-54.

References

- Computation Theory:
  - J. E. Hopcroft and J. D. Ullman, **Introduction to Automata Theory, Languages, and Computation**, Addison-Wesley, Reading, Masschusetts (1979).
  - J. G. Brookshear, **Theory of computation: Formal languages, automata, and complexity**, Benjamin/Cummings, Redwood City, California(1989).
  - C. H. Papadimitriou, **Computational Complexity**, Addison-Wesley, Reading, Massachusetts (1994).
  - S. Mertens and C. Moore, **The Nature of Computation**, Oxford University Press, Oxford, UK (2009).

References

- Dynamical Systems Theory:
  - S. H. Strogatz, **Nonlinear Dynamics and Chaos: With applications to physics, biology, chemistry, and engineering**, Addison-Wesley, Reading, Massachusetts (1994).
  - R. Gilmore and Marc Lefranc, **The Topology of Chaos**, Wiley-VCH, New York (2002).
  - D. Lind and B. Marcus, **An Introduction to Symbolic Dynamics and Coding,** Cambridge University Press, New York (1995).
  - ... many more

References

- Computational Mechanics:
  - J. P. Crutchfield and K. Young, “Inferring Statistical Complexity”, Phys. Rev. Let. **63** (1989) 105-108.
  - J. P. Crutchfield, “The Calculi of Emergence: Computation, Dynamics, and Induction”, Physica D 75 (1994) 11-54.
  - J. P. Crutchfield and C. R. Shalizi, “Thermodynamic Depth of Causal States: Objective Complexity via Minimal Representations,” Physical Review E **59** (1999) 275-283.
  - C. R. Shalizi and J. P. Crutchfield, “Computational Mechanics: Pattern and Prediction, Structure and Simplicity”, J. Stat. Phys. **104** (2001) 817-879.
  - Archive: http://cse.ucdavis.edu/~cmg/
  - Course website: http://cse.ucdavis.edu/~chaos/courses/ncaso/

Recent Citations

- J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney, “Time's Barbed Arrow: Irreversibility, Crypticity, and Stored Information”, Physical Review Letters **103**:10 (2009) 10xxxx.
- C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield, “Prediction, Retrodiction, and the Amount of Information Stored in the Present”, Journal of Statistical Physics (2009) in press.
- J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield, “Information Accessibility and Cryptic Processes”, Journal of Physics A: Math. Theo. **42** (2009) 362002.
- J. Mahoney, C. J. Ellison, and J. P. Crutchfield, “Information Accessibility and Cryptic Processes: Linear Combinations of Causal States”, arxiv.org:0906.5099 [cond-mat].

Lecture 1: Motivations, Examples, & Setting

- Randomness
- Pattern
- Learning
- Emergence
- Examples
- Mathematical Nesting

Lecture 1...

Order versus Disorder Pattern in the Middle Ground

Deterministic Chaos & Structural Complexity

Innovation in Art and Society

Determinism:

The present state of the system of nature is evidently a consequence of what it was in the preceding moment, and if we conceive of an intelligence which at a given instant comprehends all the relations of the entities of this universe, it could state the respective positions, motions, and general affects of all these entities at any time in the past or future.

PIERRE SIMON DE LAPLACE (1776)

Origin of randomness?

But ignorance of the different causes involved in the production of events, as well as their complexity, taken together with the imperfection of analysis, prevents our reaching the same certainty about the vast majority of phenomena. Thus there are things that are uncertain for us, things more or less probable, and we seek to compensate for the impossibility of knowing them by determining their different degrees of likelihood. So it is that we owe to the weakness of the human mind one of the most delicate and ingenious of mathematical theories, the science of chance or probability.

Pierre Simon de Laplace
Calculus of Probabilities (1776)

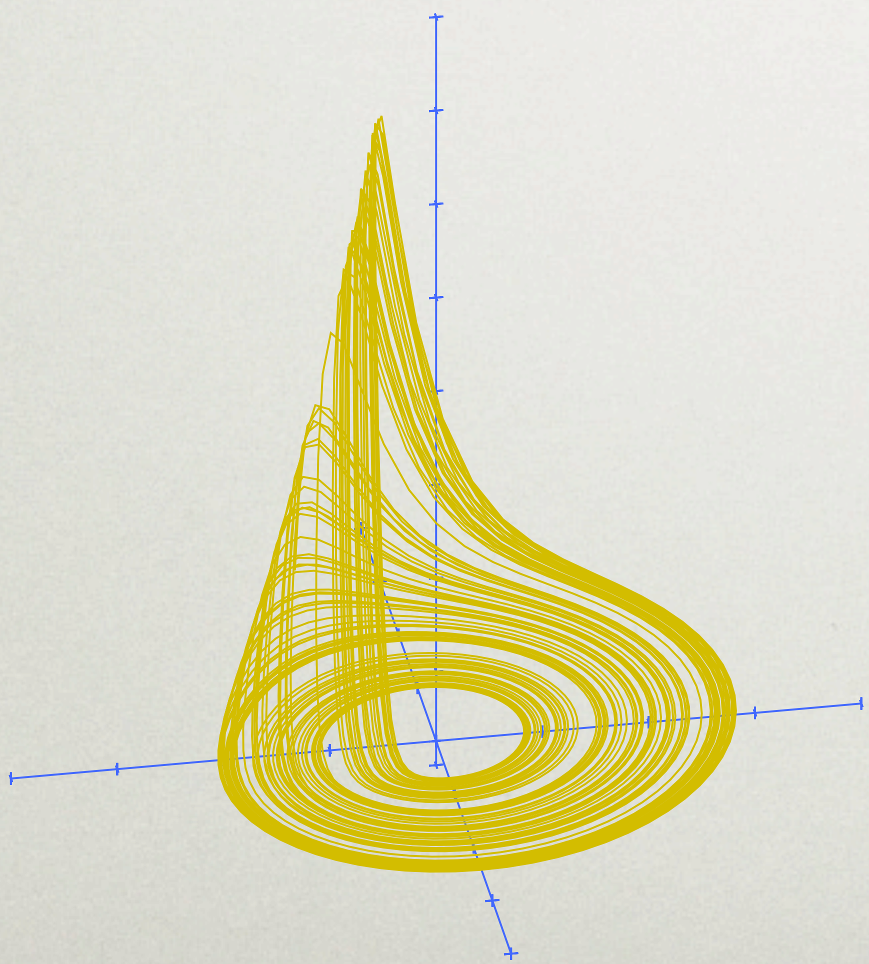
The Dynamical Extremes

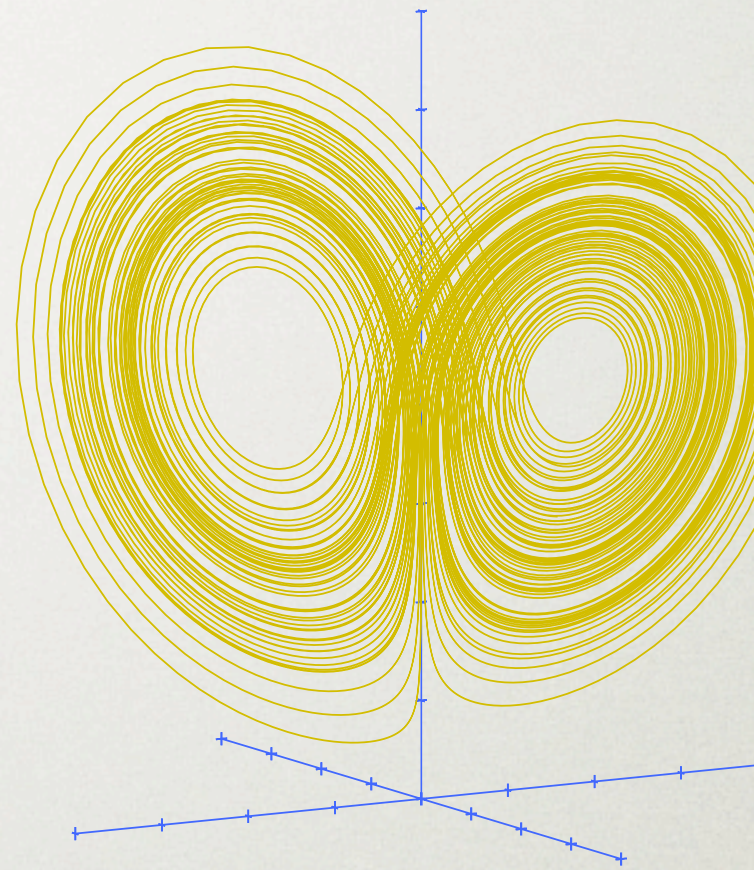
- Clock: Ideal predictability

- Fair Coin: Ideal unpredictability

Sounds of Chaos

Sounds of Chaos





Intrinsic Randomess:

And often, such small things can cause very important changes. I used to say a fly can change the whole state, in case it should buzz around a great king's head while he is weighing important counsels of state (...). And even this effect of small things causes those who do not consider things correctly to imagine some things happen accidentally and are not determined by destiny, for this distinction arises not in the facts but in our understanding.

Gottfried Leibniz (1646-1716)
Selections, Scribner's Sons, New York (1951) p. 571

Intrinsic Randomness (modern):

But even if it were the case that natural laws had no longer any secret for us, we could still only know the initial situation approximately. If that enabled us to predict the succeeding situation with the same approximation, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws. But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon.

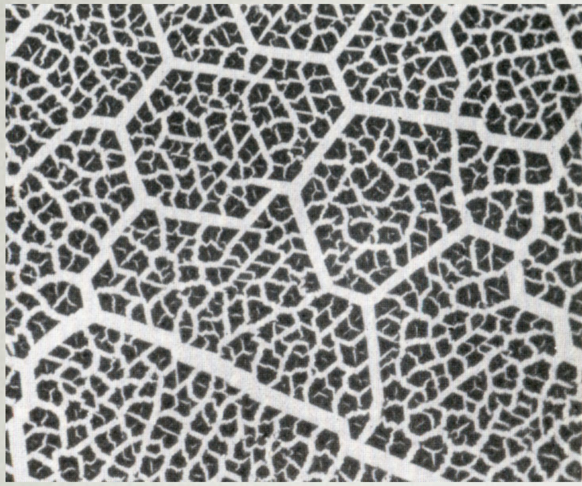
Henri Poincaré
Les Methodes Nouvelles de la Mecanique Celeste (1892)

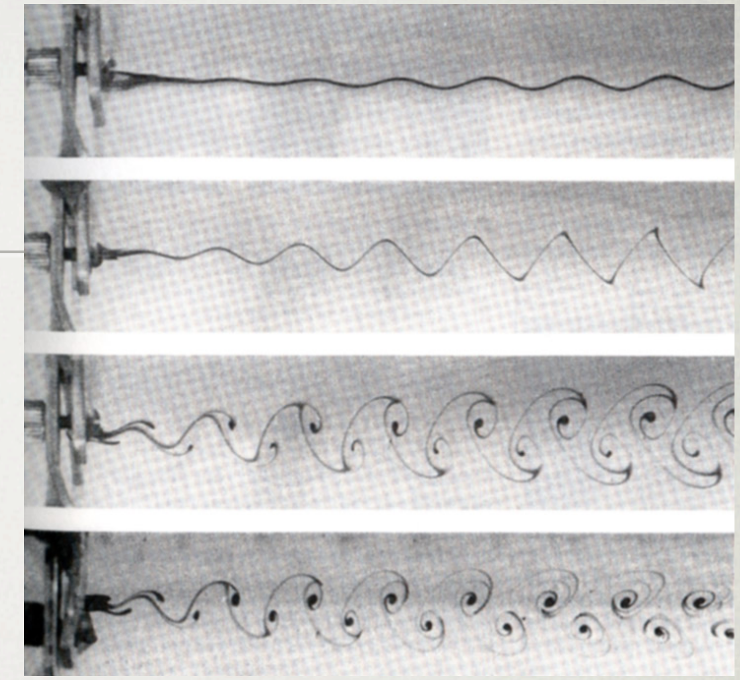
The other side of the coin ...

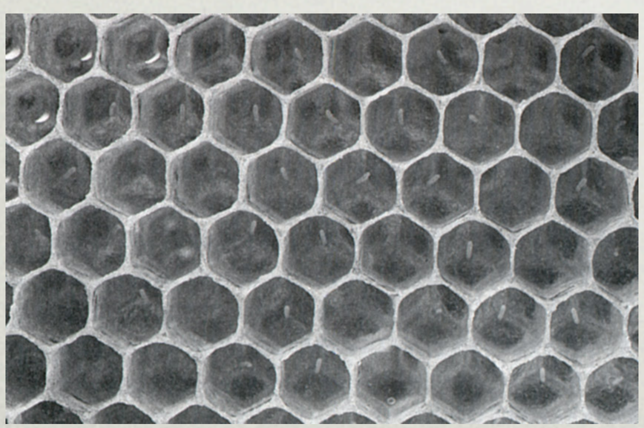
- First side:
  - Disorder from order
  - Unpredictability from simple systems
- Now:
  - Order from complication
  - Organization in large-scale systems

What is a Pattern?

Patterns







A Pattern?

These ambiguities, redundances, and deficiencies recall those attributed by Dr. Franz Kuhn to a certain Chinese encyclopedia entitled Celestial Emporium of Benevolent Knowledge. On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's brush hair, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance.

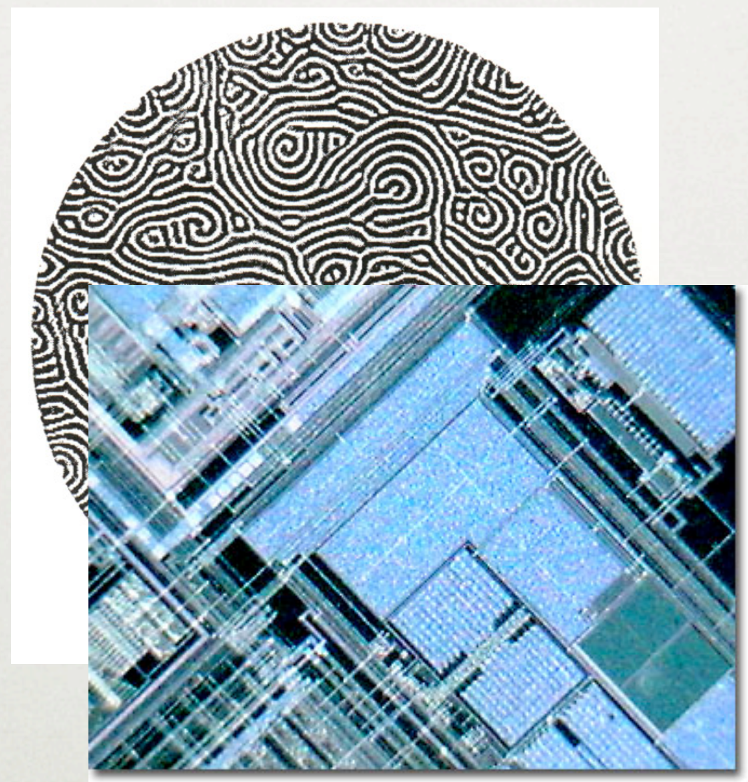
JORGE LUIS BORGES,
"THE ANALYTICAL LANGUAGE OF JOHN WILKINS"
IN *Other Inquisitions 1937-1952* (1964) 103.

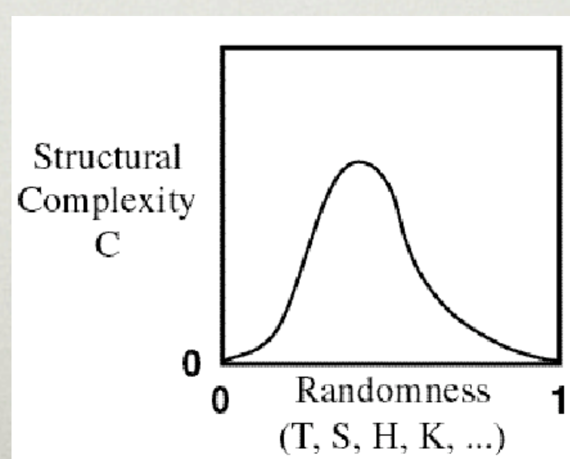
Complication versus Structure

Boredom

Variatio Delectat

Confusion





Complexity as the Middle Ground

The social history of mankind exhibits great organizations in their alternating functions of conditions for progress, and of contrivances for stunting humanity. The history of the Mediterranean lands, and of western Europe, is the history of the blessing and the curse of political organizations, of religious organizations, of schemes of thought, of social agencies for large purposes. The moment of dominance, prayed for, worked for, sacrificed for, by generations of the noblest spirits, marks the turning point where the blessing passes into the curse. Some new principle of refreshment is required. The art of progress is to preserve order amid change, and to preserve change amid order. Life refuses to be embalmed alive. The more prolonged the halt in some unrelieved system of order, the greater the crash of the dead society.

Alfred North Whitehead
"Ideal Opposites"
in *Process and Reality* (1920)

Complexity as the Middle Ground …

The same principle is exhibited by the tedium arising from the unrelieved dominance of fashion in art. Europe, having covered itself with treasures of Gothic architecture, entered upon generations of satiation. These jaded epochs seem to have lost all sense of that particular form of loveliness. It seems as though the last delicacies of feeling require some element of novelty to relieve their massive inheritance from bygone system. Order is not sufficient. What is required, is something much more complex. It is order entering upon novelty; so that the massiveness of order does not degenerate into mere repetition; and so that the novelty is always reflected upon a background of system.

Alfred North Whitehead
"Ideal Opposites"
in *Process and Reality* (1920)

Natural Lessons

- Simple systems can be complicated
- Complicated systems can be structured
- Organization arises from the interplay of order and randomness

Pattern Discovery

- All we can express is given in the language of our current understanding.
- How do we extend our vocabularies?
- Empirical: We do do this!
- We are Pattern Discovery Engines

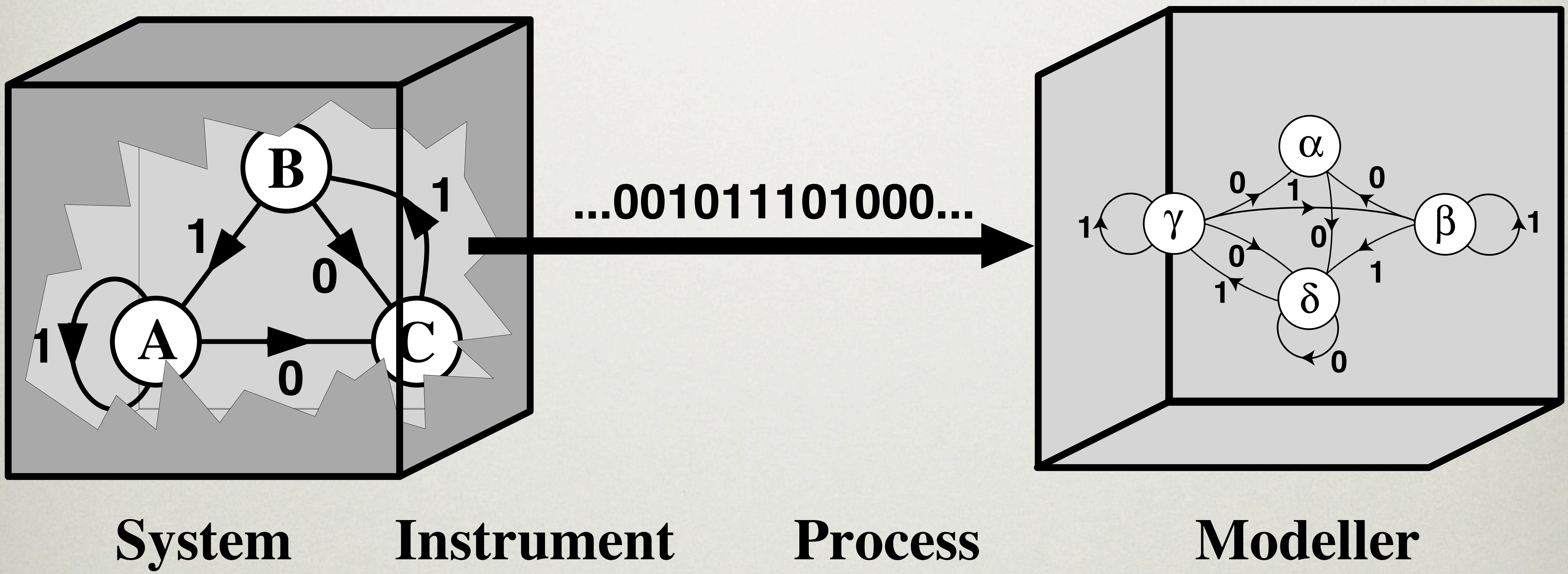
Lecture 1...

- The fundamental problem:

  Emergence

- How to avoid the inherent observer-subjectivity?

The Learning Channel



Next?

- Lecture 2: Information in Processes

  What is information?

- Lecture 3: Computational Mechanics

  What is computation?

Lecture 2: Information In Processes

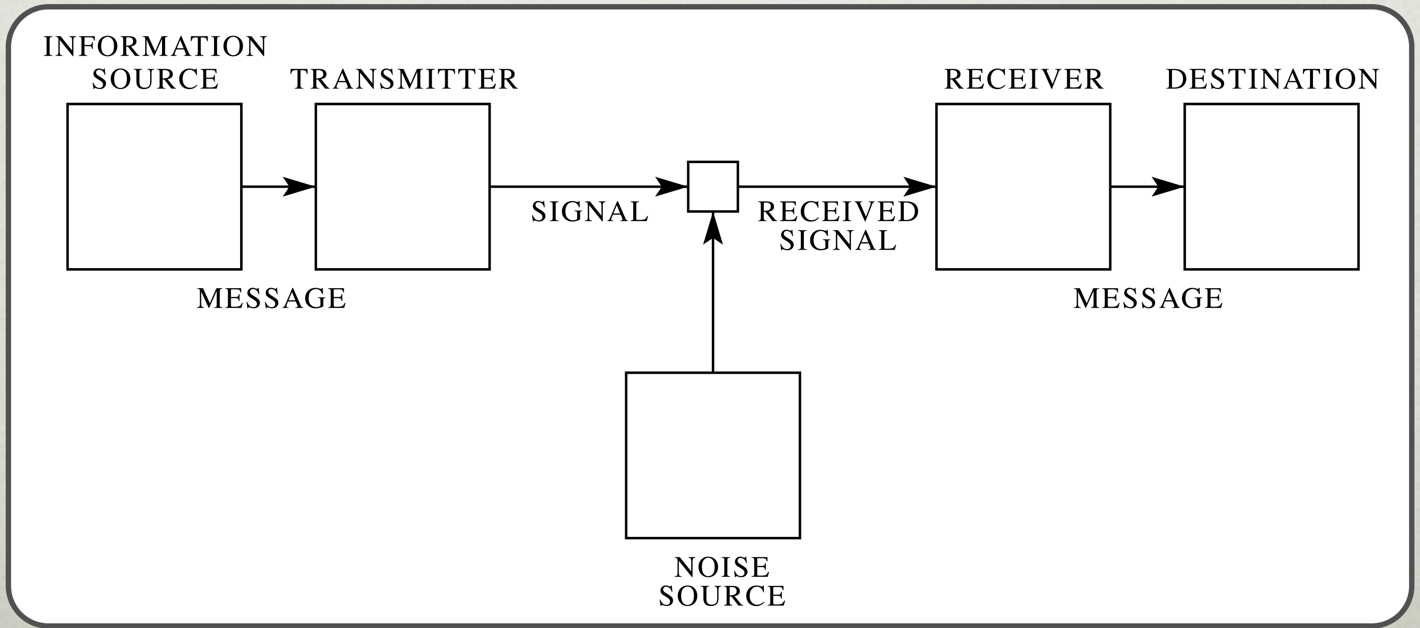
- Shannon Information
- Processes
- Degrees of Randomness
- Measures of Organization
- Information Hierarchy
- Limitations of Information Theory

Setting

- Shannon's original:

  A Mathematical Theory of Communication,
  Bell Systems Technical Journal **27** (1948) 379-423; 623-656.

- Two component theories:
  - Information Theory: What is information?
  - Communication Theory: How is it transmitted?



Setting ...

- First connects to dynamical systems theory & statistical mechanics: Information & its production
- Source entropy rate as key dynamical property:
  - A. N. Kolmogorov, "Entropy per Unit Time as a Metric Invariant of Automorphisms", Dokl. Akad. Nauk. SSSR **124** (1959) 754.
  - Ja. G. Sinai, "On the Notion of Entropy of a Dynamical System", Dokl. Akad. Nauk. SSSR **124** (1959) 768.
- Prediction $\Leftrightarrow$ Entropy rate:
  - ... degree of surprise
  - ... rate of information production
  - ... irreducible prediction error
  - ... mixing of state space regions

Information-Theoretic Analysis of Complex Systems

- Chain: $\overleftrightarrow{X} = \ldots X_{-2} X_{-1} X_0 X_1 X_2 \ldots$
- Random variable: X_t
- Realization: $\ldots x_{-2} x_{-1} x_0 x_1 x_2 \ldots$
- Event: $x_t \in \mathcal{A}$
- Measurement alphabet: $\mathcal{A}$

Information-Theoretic Analysis of Complex Systems

- Past: $\overleftarrow{X}_t = \ldots X_{t-3} X_{t-2} X_{t-1}$

- Future: $\overrightarrow{X}_t = X_t X_{t+1} X_{t+2} \ldots$

- L-Block: $X_t^L = X_t X_{t+1} \ldots X_{t+L-1}$

Information-Theoretic Analysis of Complex Systems ...

- Process:

$$\Pr(\overleftrightarrow{X}) = \Pr(\ldots X_{-2} X_{-1} X_0 X_1 X_2 \ldots)$$

- Block distributions:

$$\{\Pr(X_t^L) = \Pr(X_t X_{t+1} \ldots X_{t+L-1})\}$$

- Process specification:

$$\{\Pr(X_t^L) : \forall t, L\}$$

Information-Theoretic Analysis of Complex Systems ...

- Consistent specification:

$$\Pr(X_t^{L-1}) = \sum_{\{X_{t+L-1}\}} \Pr(X_t^L) \text{ and } \Pr(X_t^{L-1}) = \sum_{\{X_t\}} \Pr(X_t^L)$$

- Stationary process:

$$\Pr(X_t X_{t+1} \ldots X_{t+L-1}) = \Pr(X_0 X_1 \ldots X_{L-1})$$

Information-Theoretic Analysis of Complex Systems ...

- Block Entropy: $$H(L) \equiv H[\Pr(X^L)] = -\sum_{\{x^L \in \mathcal{A}^L\}} \Pr(x^L) \log_2 \Pr(x^L)$$

No measurements, no information:

$$H(0) = 0$$

Monotonic increasing:

$$H(L) \geq H(L-1)$$

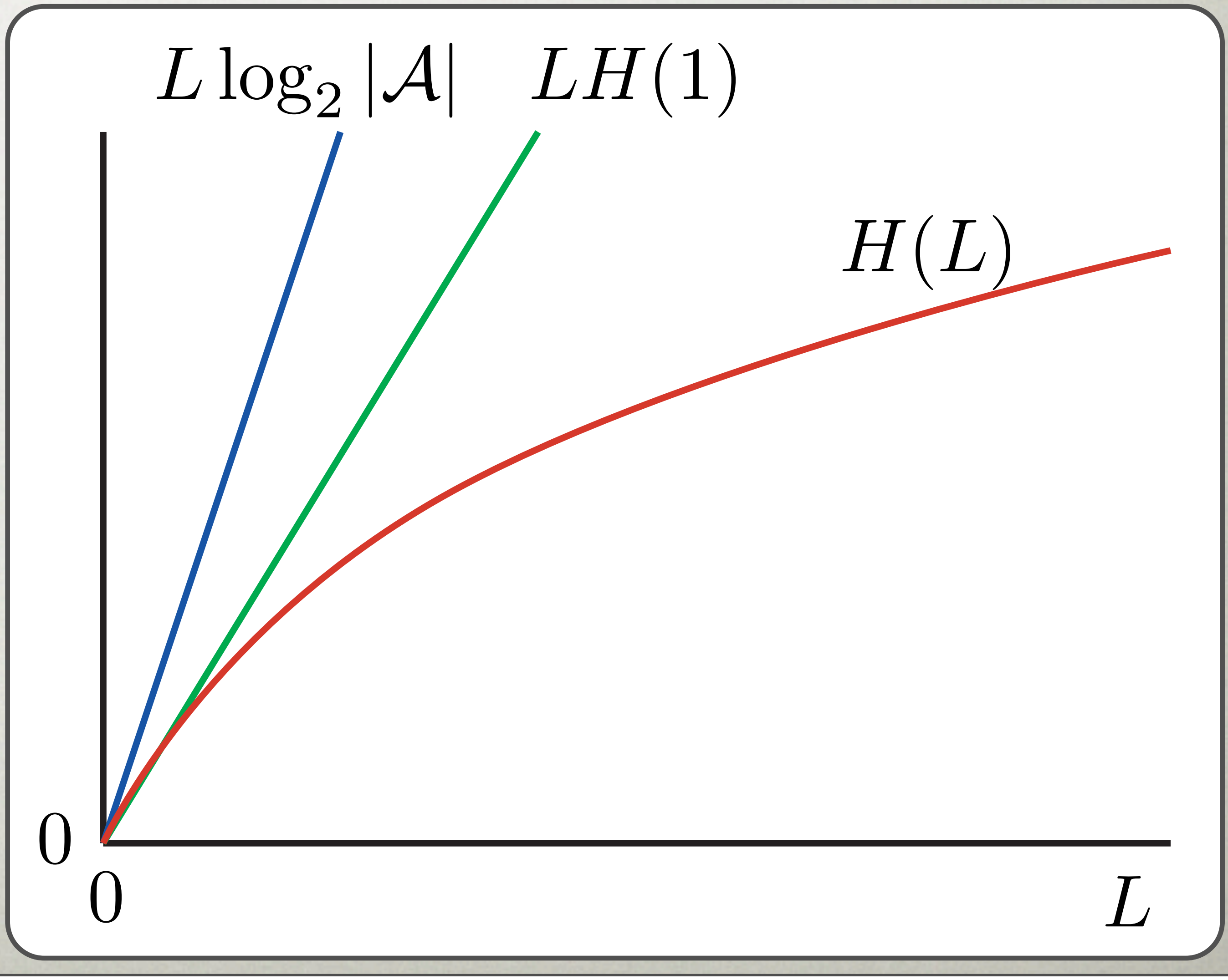
Bounds:

(1) Crude:

$$H(L) \leq L \log_2 |\mathcal{A}|$$

(2) 1-block Markov:

$$H(L) \leq L H(1)$$



Information-Theoretic Analysis of Complex Systems ...

- Entropy Rate:

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

Interpretations:

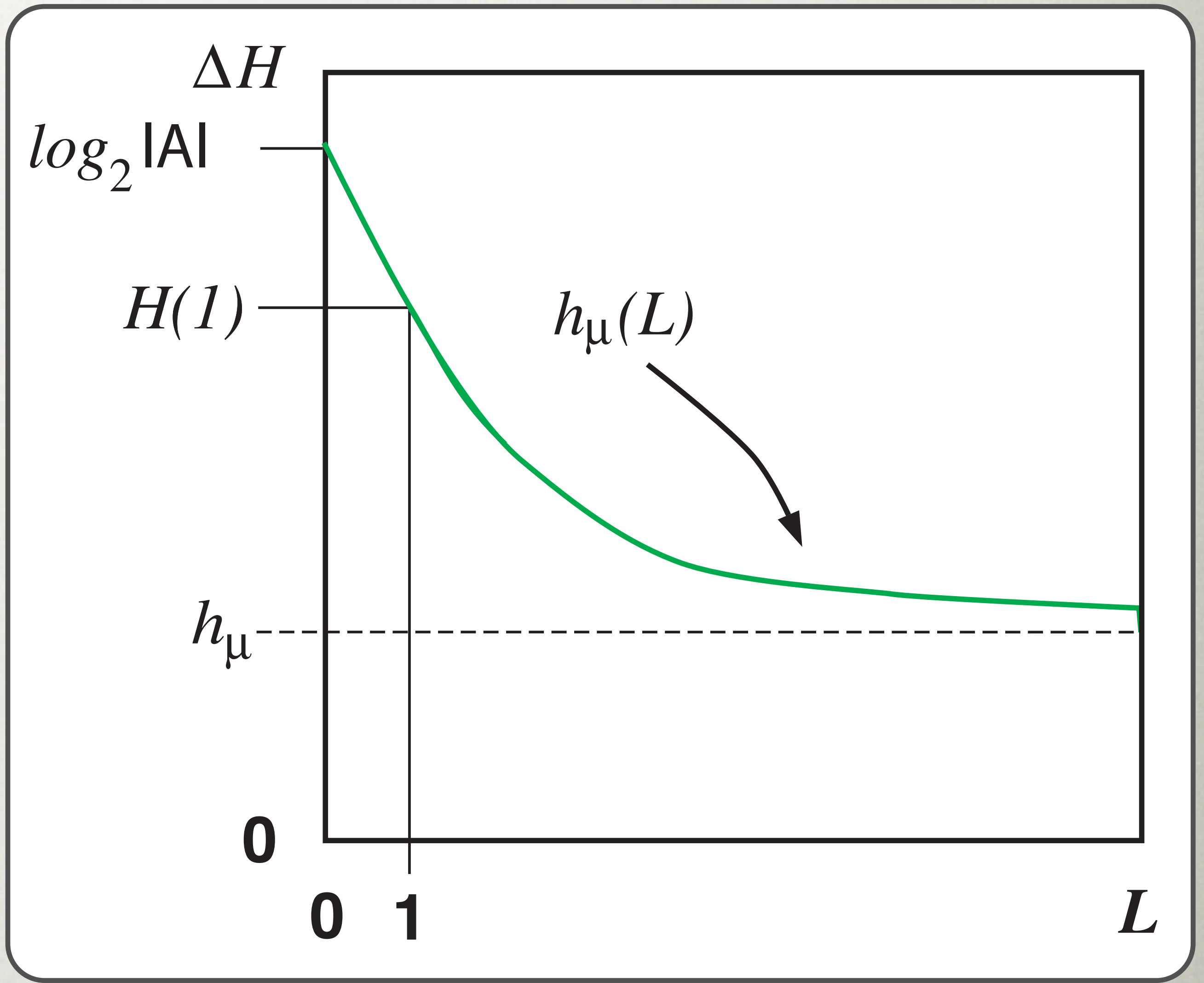
Asymptotic growth rate of entropy

Irreducible randomness of process

Information-Theoretic Analysis of Complex Systems ...

- Length-L Estimate of Entropy Rate:

$$\widehat{h}_\mu(L) = H(s_L | s_1 \cdots s_{L-1})$$



Interpretations:

Uncertainty in next measurement, given past

A measure of unpredictability

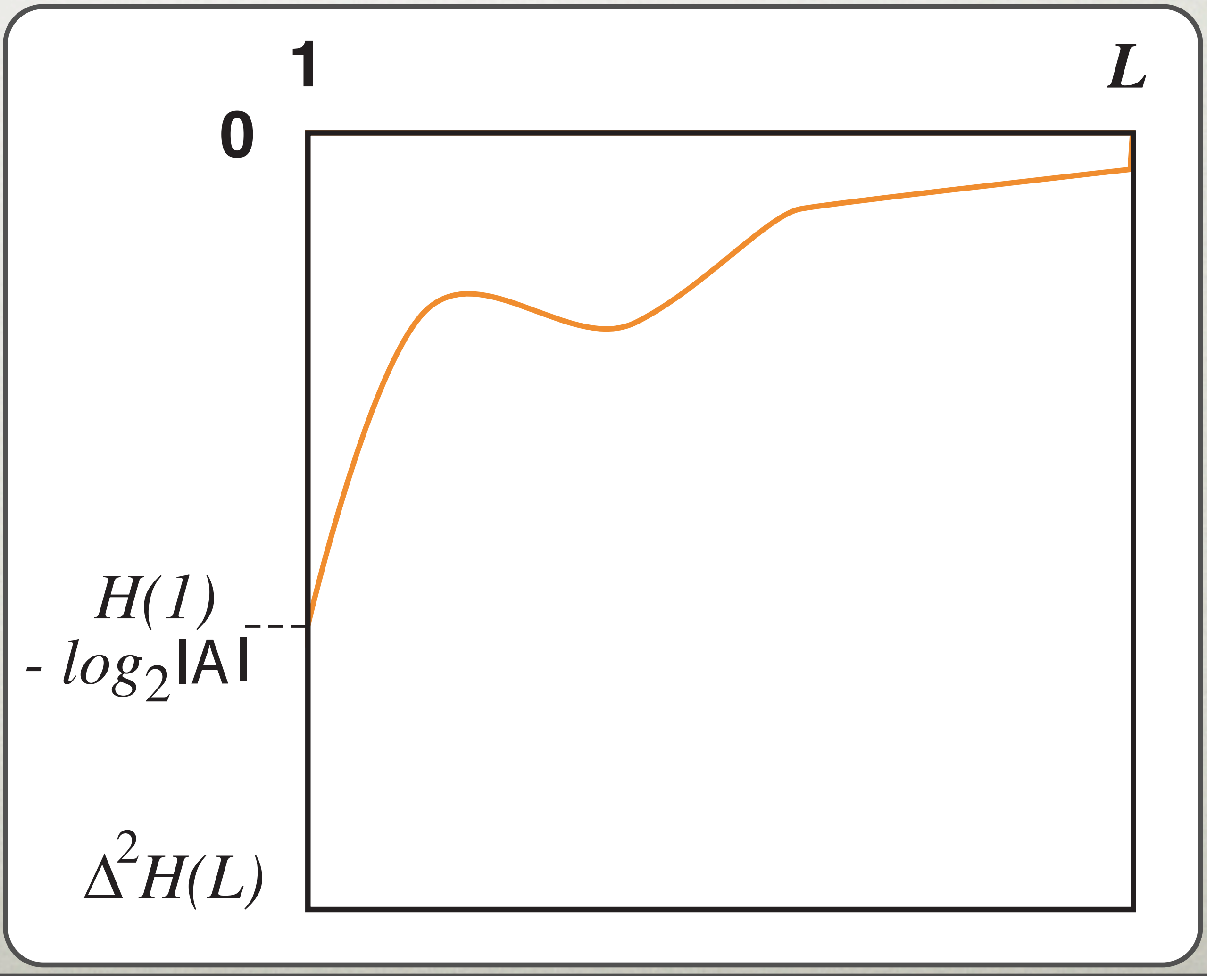
Asymptotic slope of block entropy

Information-Theoretic Analysis of Complex Systems ...

- Predictability Gain:

$$\Delta^2 H(L) = h_\mu(L) - h_\mu(L-1)$$

Rate at which unpredictability is lost



Information-Theoretic Analysis of Complex Systems ...

- Entropy Hierarchy:

Took derivatives:

(1) Block entropy: $H(L)$

(2) Entropy rate: $h_\mu(L) = \Delta H(L)$

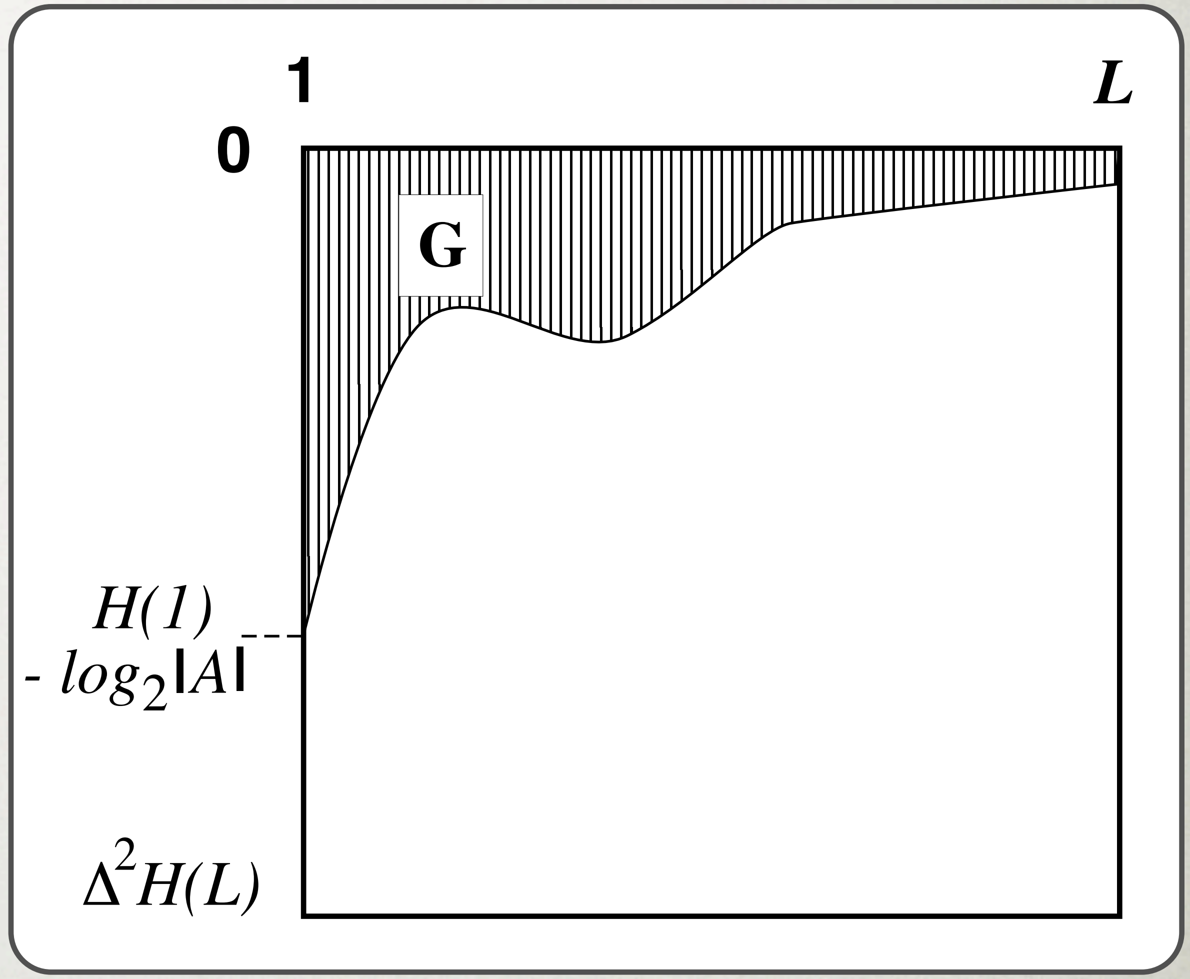
(3) Predictability gain: $\Delta h_\mu(L) = \Delta^2 H(L)$

Now take integrals!

Information-Theoretic Analysis of Complex Systems ...

- Total Predictability:

$$\mathbf{G} = \sum_{L=1}^{\infty} \Delta^2 H(L)$$



- Measurement Redundancy:

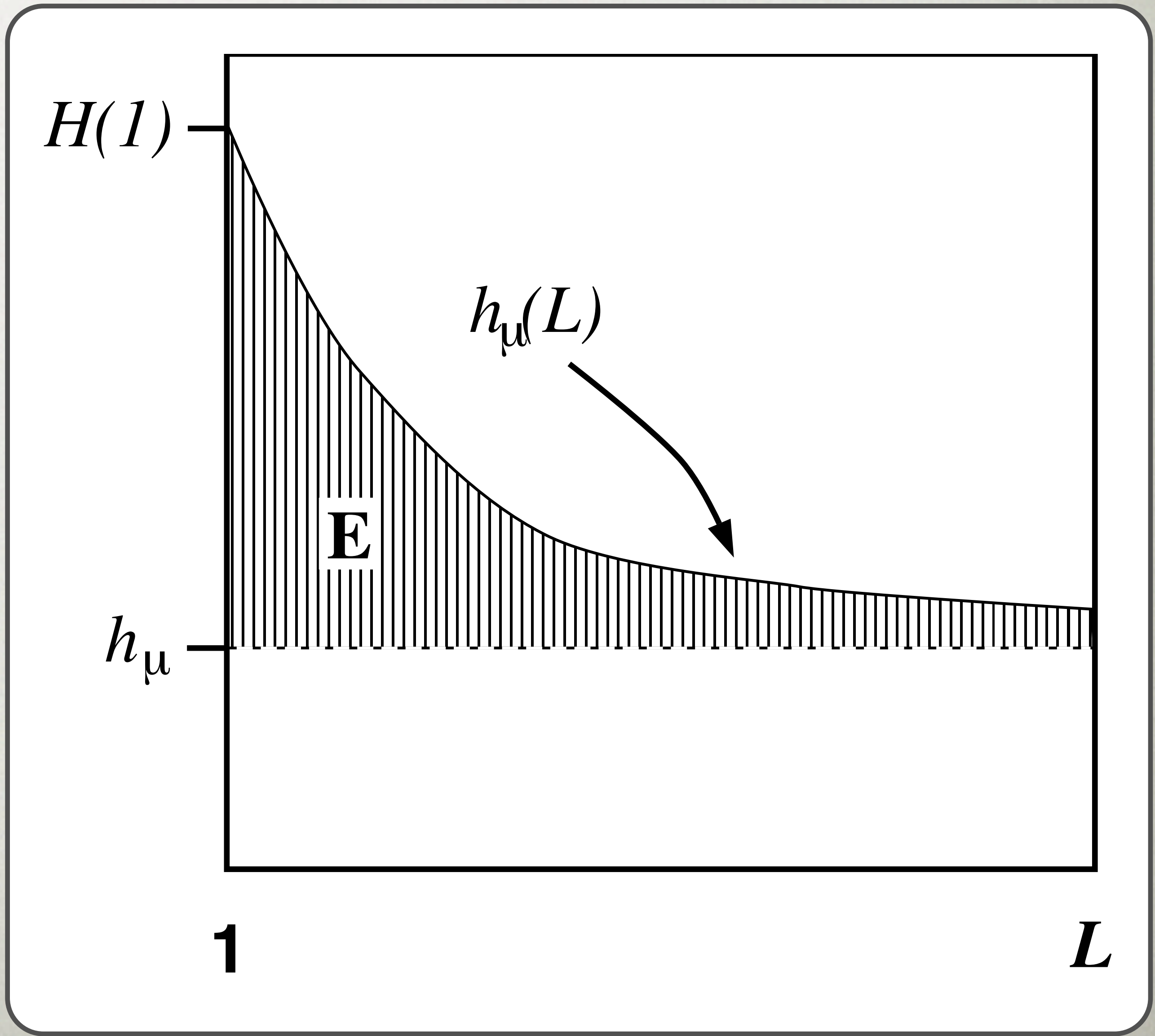
$$-\mathbf{G} = \mathcal{R} = \log_2 |\mathcal{A}| - h_\mu$$

Information-Theoretic Analysis of Complex Systems ...

- Excess Entropy:

  As entropy convergence:

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

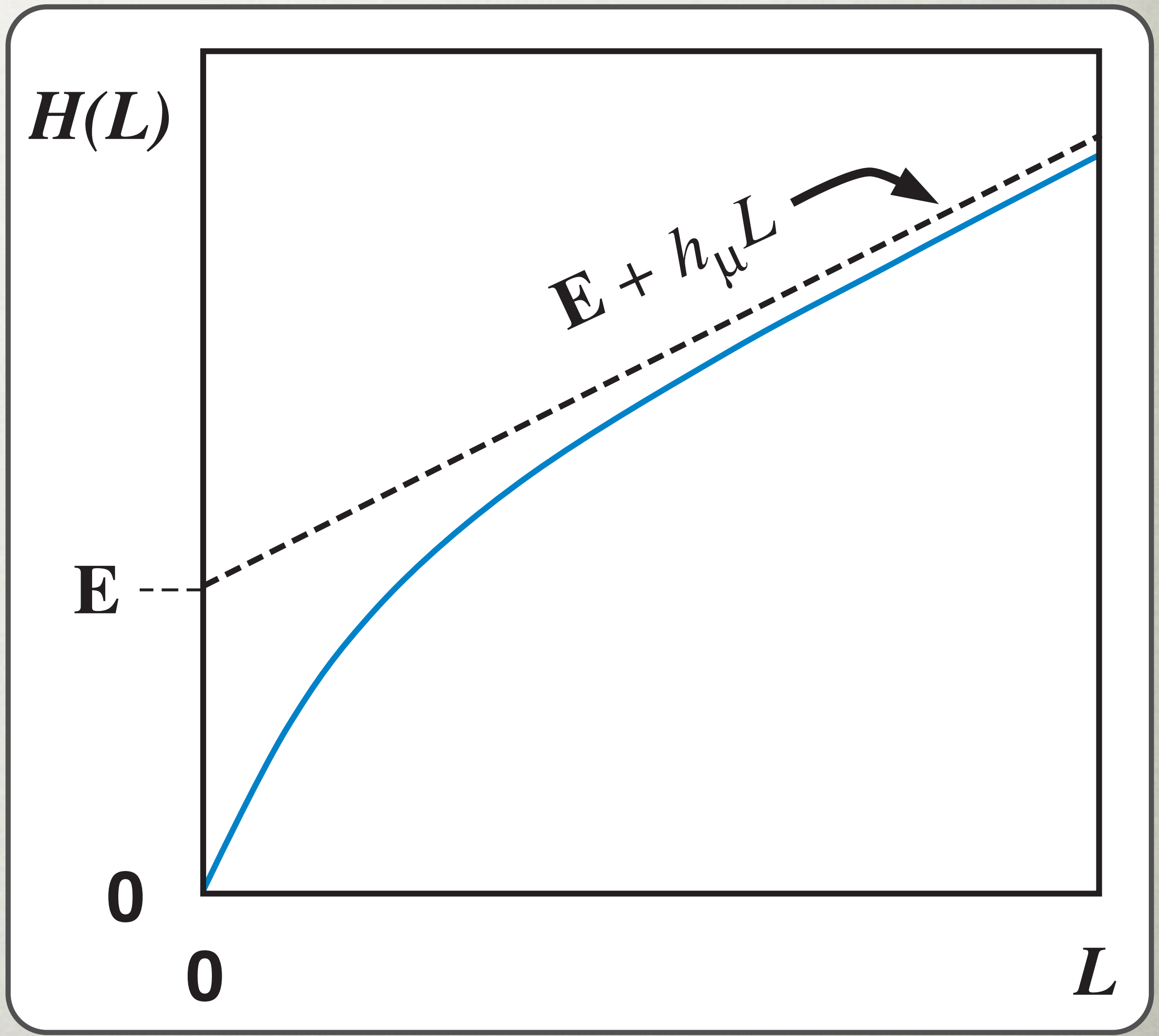


Information-Theoretic Analysis of Complex Systems ...

- Excess Entropy

  As asymptote of entropy growth:

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



That is,

$$H(L) \propto \mathbf{E} + h_\mu L$$

Y-Intercept of entropy growth

Information-Theoretic Analysis of Complex Systems ...

- Excess Entropy:

  As mutual information between past and future:

  View process as a communication channel

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

$$\text{Past} \rightarrow \underbrace{\boxed{\text{Present}}}_{\text{Channel}} \rightarrow \text{Future}$$

Property:

Symmetric in time

Interpretation:

Effective channel utilization

Information that process communicates from past to future.

Reduction in uncertainty about the future, given the past.

Reduction in uncertainty about the past, given the future.

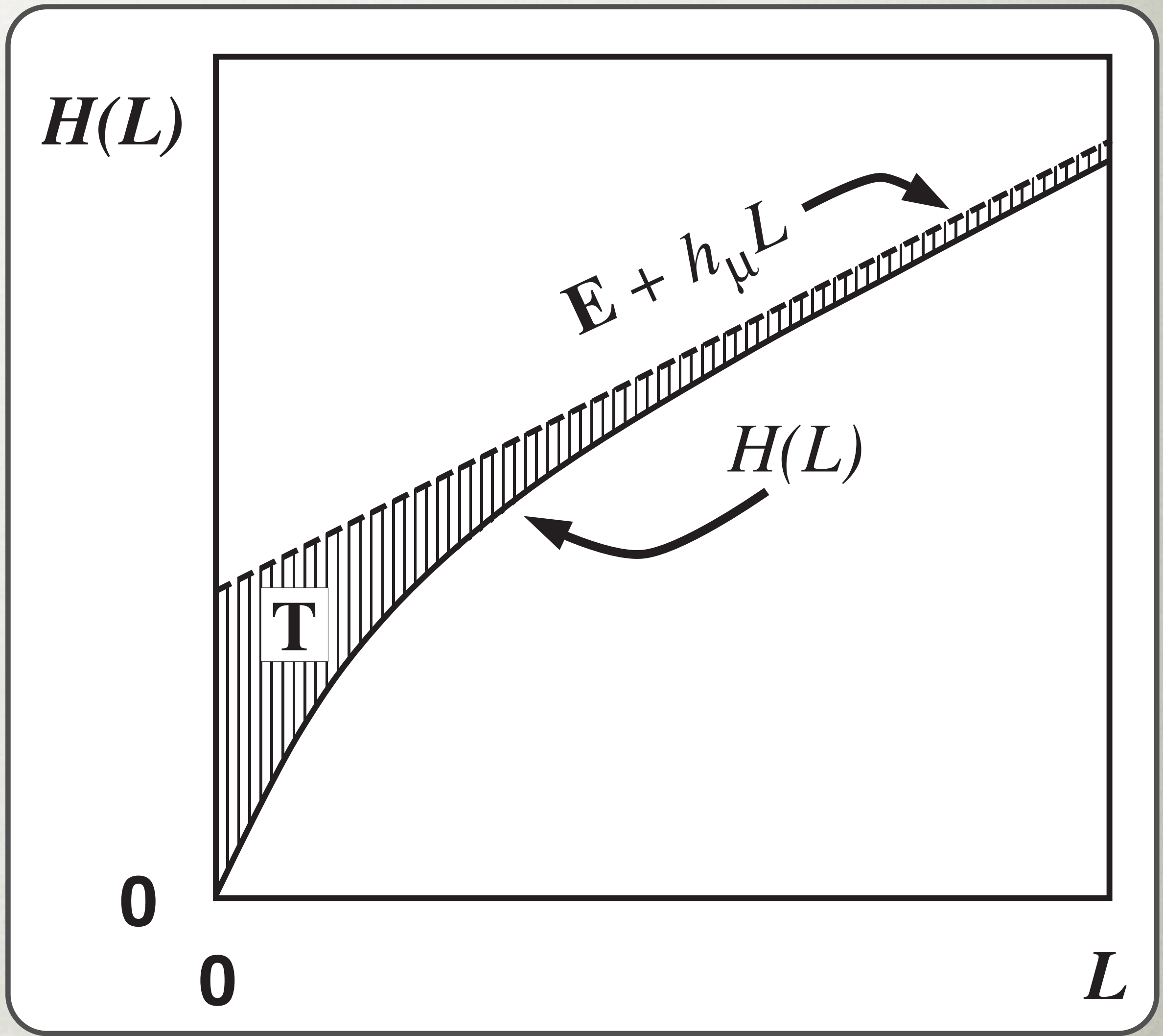
Information-Theoretic Analysis of Complex Systems ...

- Synchronized: $L \geq L'$

$$H(L) \approx \mathbf{E} + h_\mu L$$

- Transient Information

$$\mathbf{T} = \sum_{L=0}^{\infty} [\mathbf{E} + h_\mu L - H(L)]$$



Interpretations:

Total uncertainty observed while synchronizing.

Information to extract to be synchronized.

Kinds of Information

- Entropy Rate: h_μ
- Excess Entropy: $\mathbf{E}$
- Predictability Gain: $\mathbf{G}$
- Transient Information: $\mathbf{T}$

Kinds of Information

- Entropy Rate: h_μ info production; prediction error
- Excess Entropy: $\mathbf{E}$
- Predictability Gain: $\mathbf{G}$
- Transient Information: $\mathbf{T}$

Kinds of Information

- Entropy Rate: h_μ — info production; prediction error
- Excess Entropy: $\mathbf{E}$ — correlation of past and future
- Predictability Gain: $\mathbf{G}$
- Transient Information: $\mathbf{T}$

Kinds of Information

- Entropy Rate: h_μ — info production; prediction error
- Excess Entropy: $\mathbf{E}$ — correlation of past and future
- Predictability Gain: $\mathbf{G}$ — info to extract to predict optimally
- Transient Information: $\mathbf{T}$

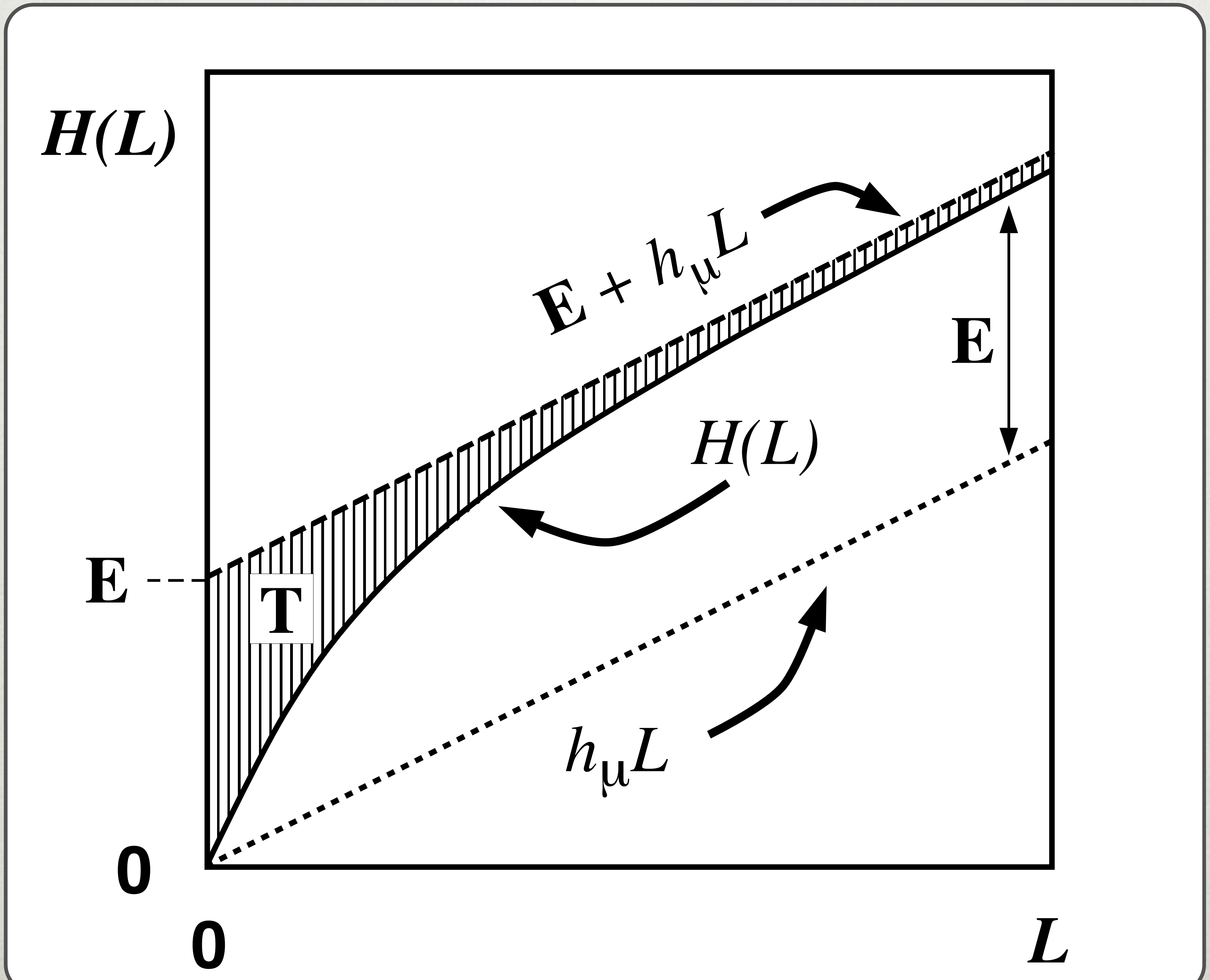
Kinds of Information

- Entropy Rate: h_μ — info production; prediction error
- Excess Entropy: $\mathbf{E}$ — correlation of past and future
- Predictability Gain: $\mathbf{G}$ — info to extract to predict optimally
- Transient Information: $\mathbf{T}$ — info to extract to synchronize

Calculus of the Entropy Hierarchy:

Level	Gain (Derivative)	Information (Integral)
0	Block Entropy $H(L)$	Transient Information $\mathbf{T} = \sum_{L=1}^{\infty} [\mathbf{E} + h_\mu L - H(L)]$
1	Entropy Rate Loss $h_\mu(L) = \Delta H(L)$	Excess Entropy $\mathbf{E} = \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu]$
2	Predictability Gain $\Delta^2 H(L)$	Total Predictability $\mathbf{G} = -\mathcal{R}$
...	...	...

Roadmap to Information(s)



J. P. Crutchfield and D. P. Feldman, "Regularities Unseen, Randomness Observed: Levels of Entropy Convergence", CHAOS **13**:1 (2003) 25-54.

Is Information Theory Sufficient?

- No!
- Measurements are process's states
- No direct measure of structure

Lecture 3: Computational Mechanics

- So far answered: What is information?
- Now: What is pattern?

Lecture 3: Computational Mechanics

- Pattern
- Causal Architecture
- Intrinsic Computation
- Modeling as Decryption

Is Information Theory Sufficient?

- No!
- Measurements are process's states (Wrong!)
- No direct measure of structure

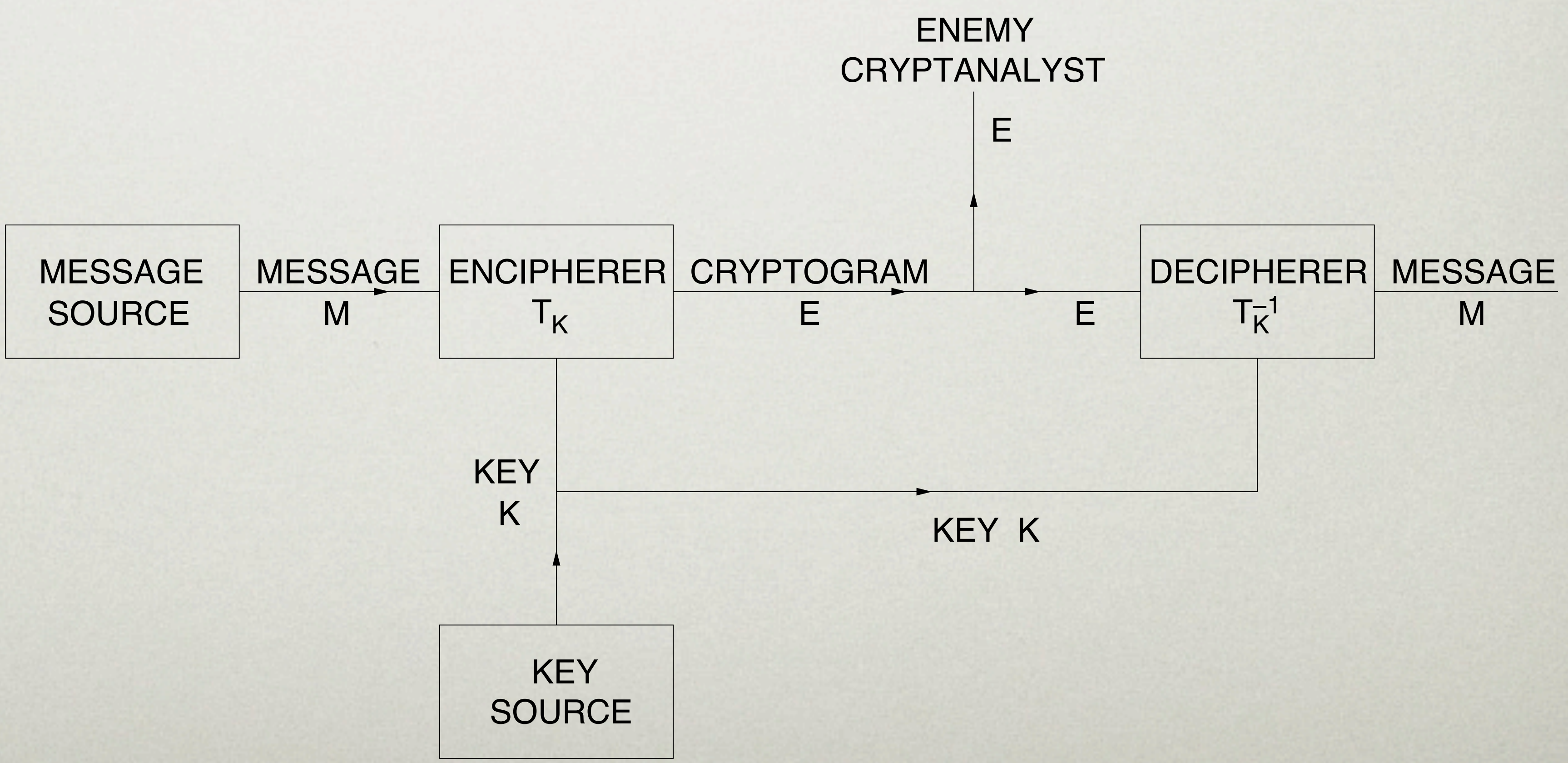
- Need to use communication theory!
- ... and ...

Setting ...

- Shannon's second paper: Key for modeling!

  Communication Theory of Secrecy Systems, Bell Sys. Tech. J. **28** (1949) 656-715.

  Original effort on "information".



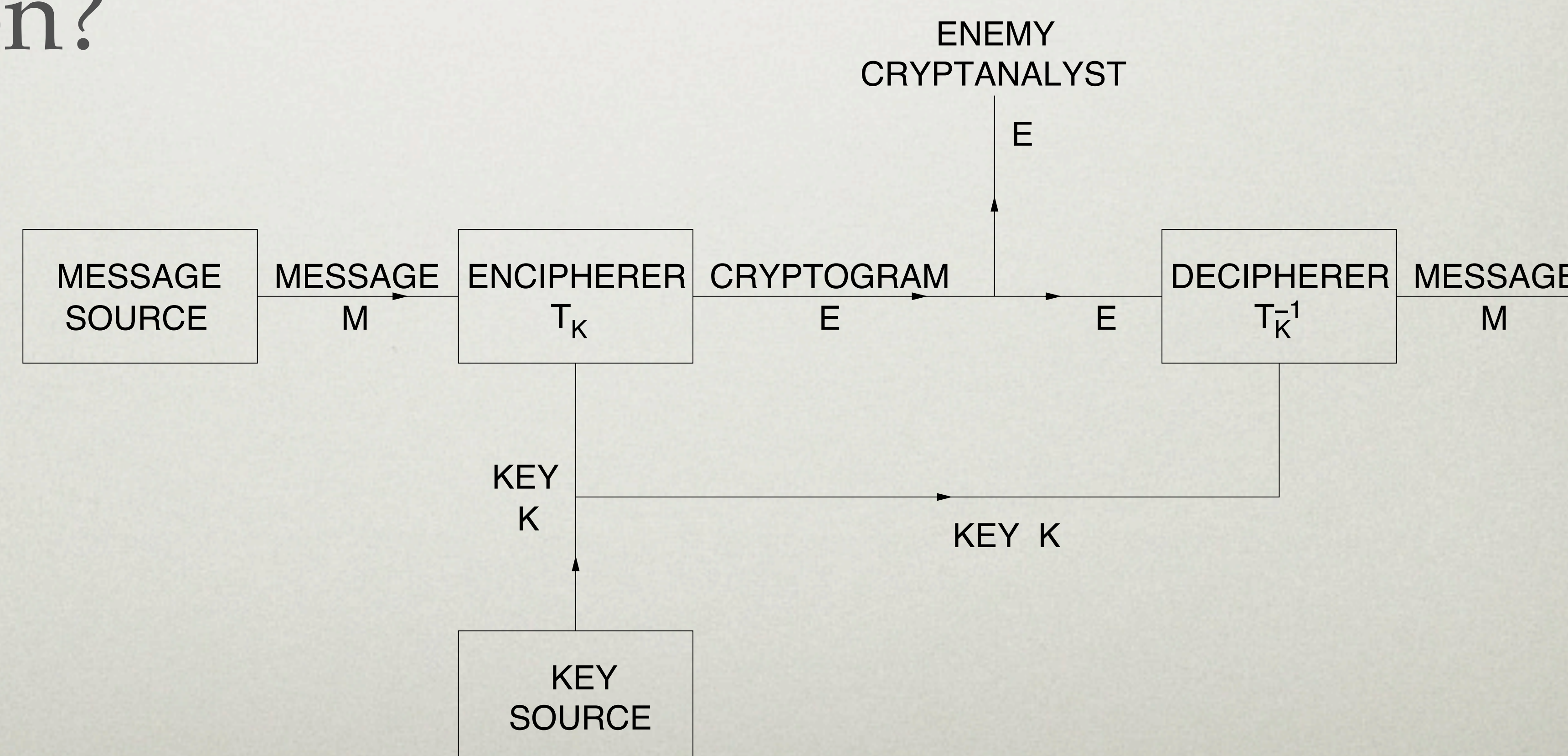
Setting ...

- Shannon's second paper: Key for modeling!

  Communication Theory of Secrecy Systems, Bell Sys. Tech. J. **28** (1949) 656-715.

  Original effort on "information".

- Modeling: How to decrypt nature's hidden information?



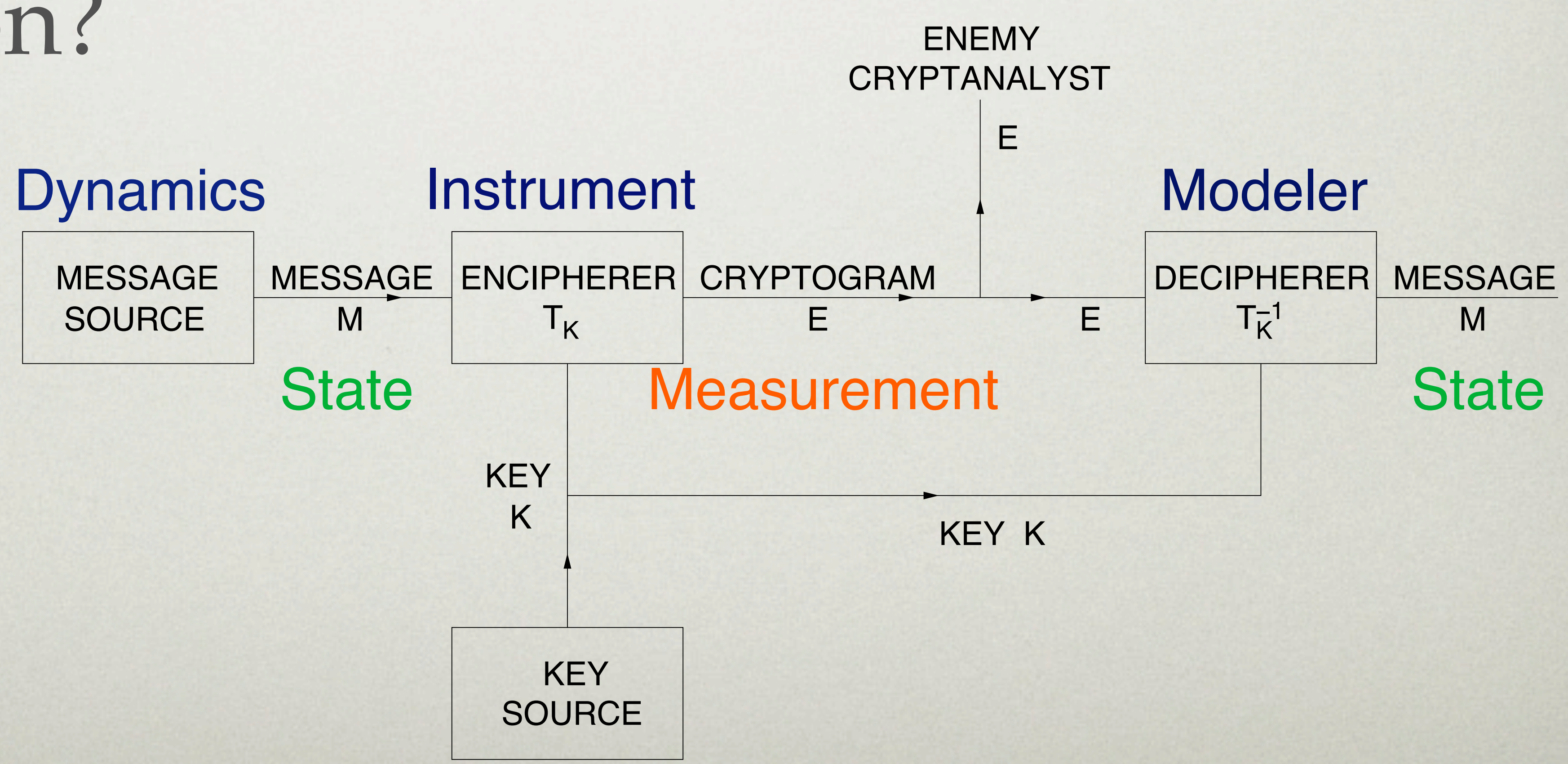
Setting ...

- Shannon's second paper: Key for modeling!

Communication Theory of Secrecy Systems, Bell Sys. Tech. J. **28** (1949) 656-715.

Original effort on "information".

- Modeling: How to decrypt nature's hidden information?



Computational Mechanics

- Intrinsic computation:

1. How much historical information is stored in the present?
2. In what architecture is that information stored?
3. How does stored information produce future behavior?

J. P. Crutchfield and K. Young, “Inferring Statistical Complexity”,
Phys. Rev. Let. **63** (1989) 105-108.

Computational Mechanics

- Process $\Pr(\overleftarrow{X}, \overrightarrow{X})$ is a communication channel from the past $\overleftarrow{X}$ to the future $\overrightarrow{X}$:

Computational Mechanics

- Prediction: Map from a past to possible futures

$$\Pr(\overrightarrow{X} | \overleftarrow{x})$$

- A good predictor captures all of the information between past and future:

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

Process's effective channel utilization.

- Modeling:
  - Make good predictions, but also
  - Represent underlying mechanisms

Computational Mechanics

- Group all histories that give same prediction:

$$\epsilon(\overleftarrow{x}) = \{\overleftarrow{x}' : \Pr(\overrightarrow{X}|\overleftarrow{x}) = \Pr(\overrightarrow{X}|\overleftarrow{x}')\}$$

- Equivalence relation: $\overleftarrow{x} \sim \overleftarrow{x}'$

- Equivalence classes are causal states:

$$\boldsymbol{\mathcal{S}} = \Pr(\overleftarrow{X}, \overrightarrow{X})/ \sim$$

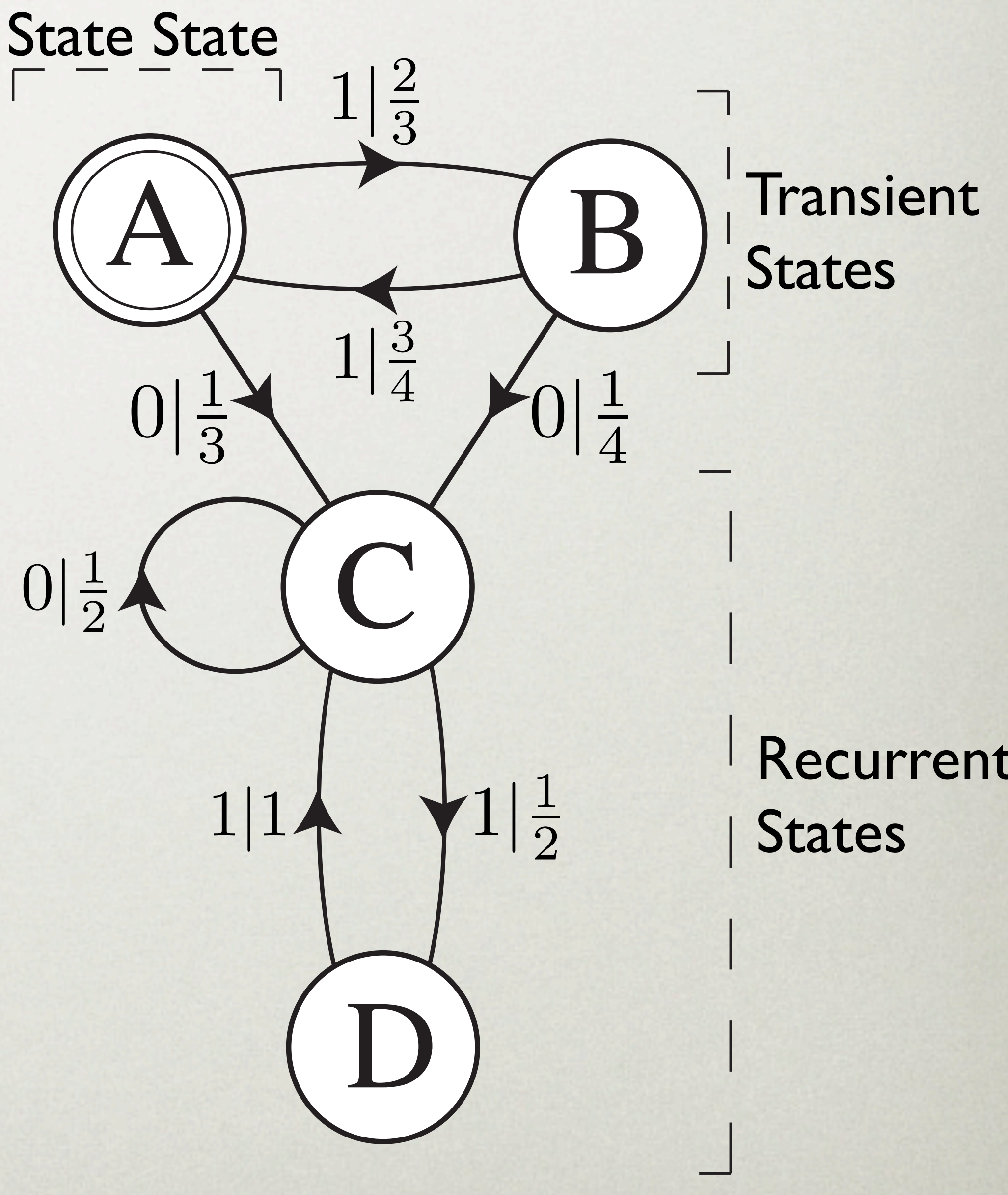
Computational Mechanics

- ε-Machine:

$$M = \{\boldsymbol{\mathcal{S}}, \{T^{(x)}, x \in X\}\}$$

- Dynamic:

$$T^{(x)}_{\sigma,\sigma'} = \Pr(\sigma'|\sigma, x)$$



Computational Mechanics

- Causal Shielding:

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

- Optimal Prediction:

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

- Capture all of the shared information:

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

Computational Mechanics

- Rival models: $\mathcal{R}$

- Prescient rivals: $I[\widehat{\mathcal{R}}; \overrightarrow{X}] = \mathbf{E}$

- ε-Machine smallest prescient model:

$$C_\mu \equiv H[\mathcal{S}] \leq H[\widehat{\mathcal{R}}]$$

Kinds of Intrinsic Computing

- Directly from ε-Machine:
  - Stored information (Statistical complexity):

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

  - Information production (Entropy rate):

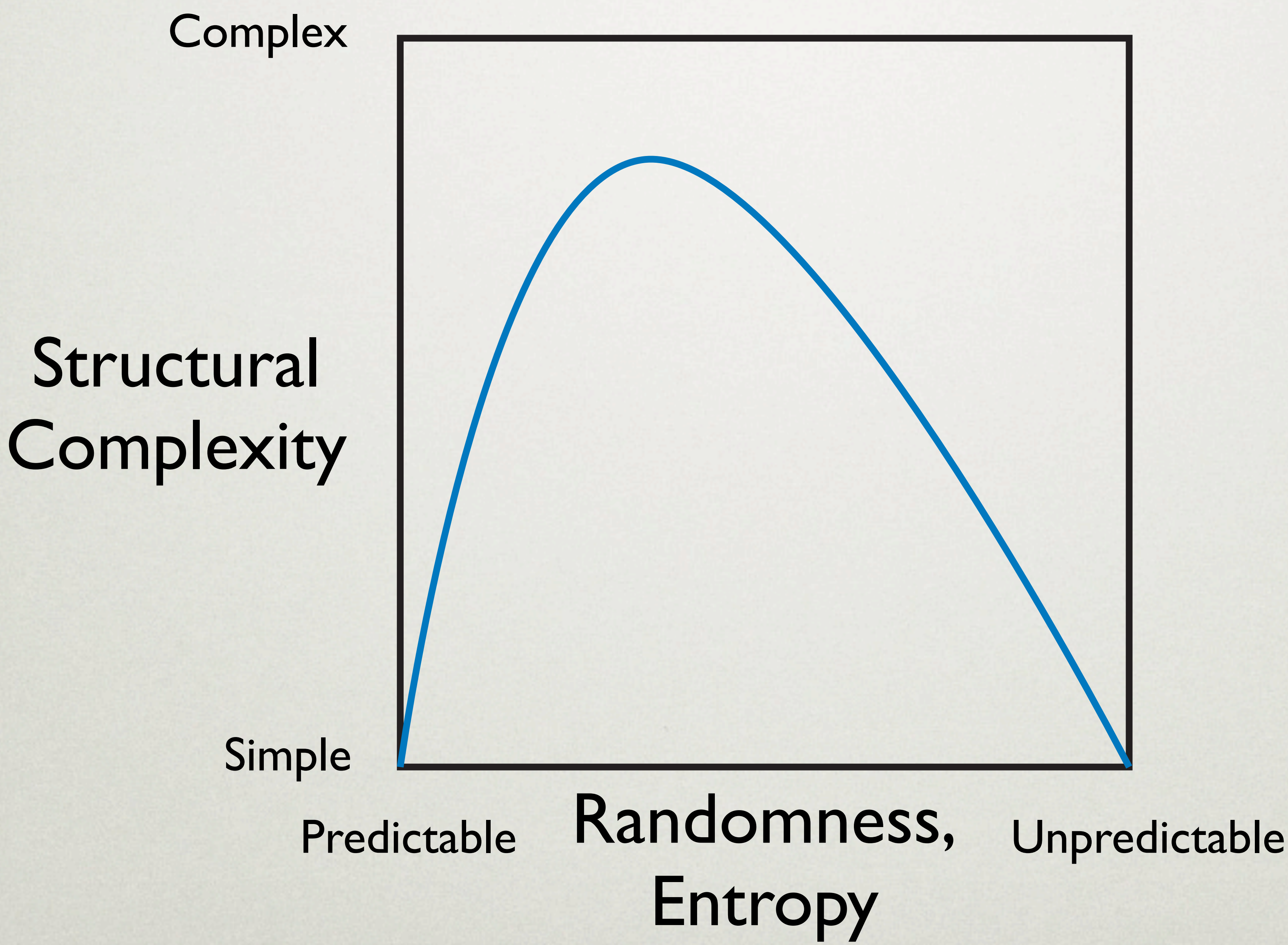
$$h_\mu = - \sum_{\sigma \in \boldsymbol{\mathcal{S}}} \Pr(\sigma) \sum_{\sigma' \in \boldsymbol{\mathcal{S}}, s \in \mathcal{A}} \Pr(\sigma \rightarrow_s \sigma') \log_2 \Pr(\sigma \rightarrow_s \sigma')$$

Stored Information

- New "invariant"—Statistical complexity: C_μ
  - J. P. Crutchfield and K. Young, "Inferring Statistical Complexity", Physical Review Letters **63** (1989) 105-108.

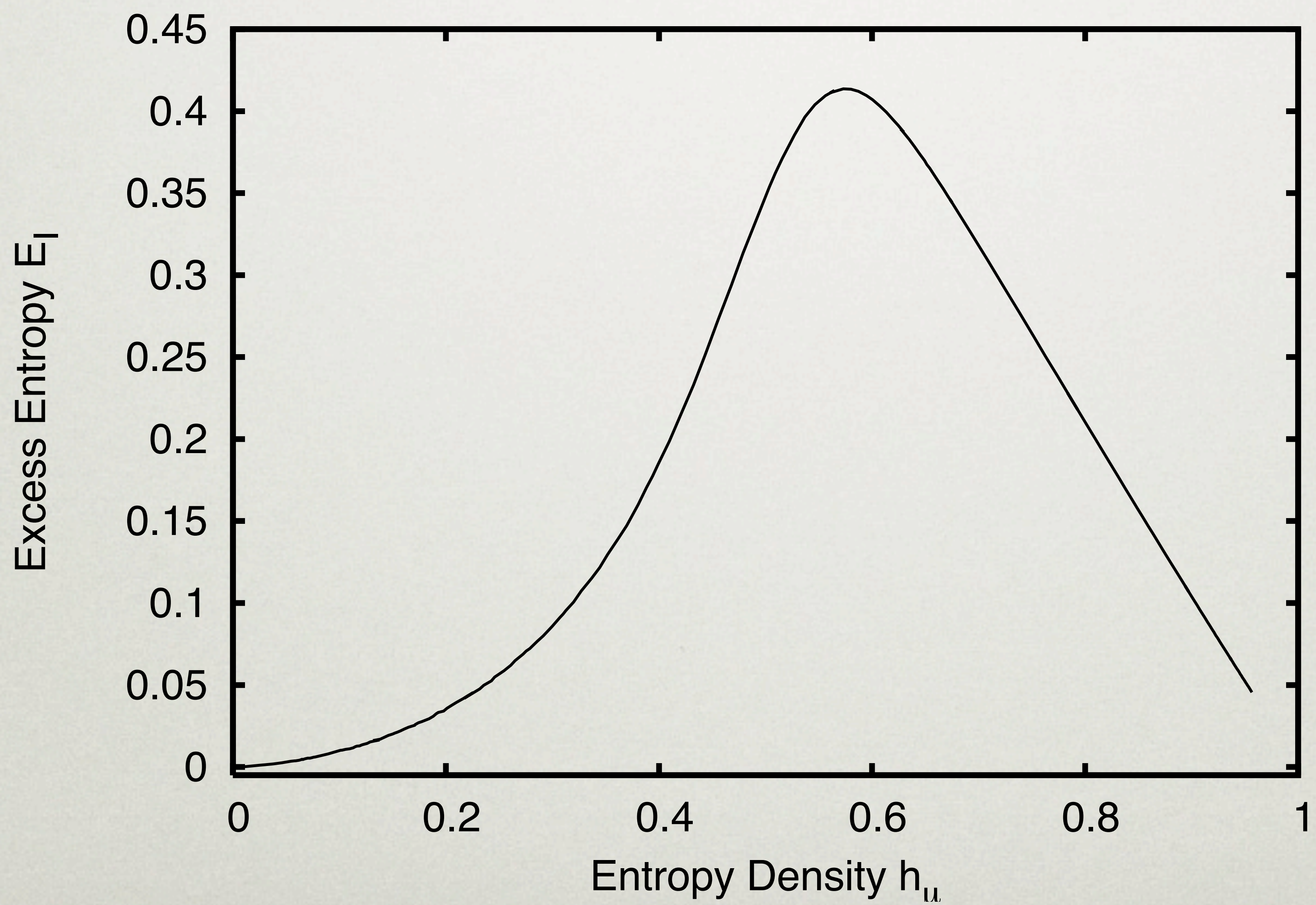
- The amount of information a process stores
- ... producing its unpredictability (entropy rate)

Complexity-Entropy Diagrams



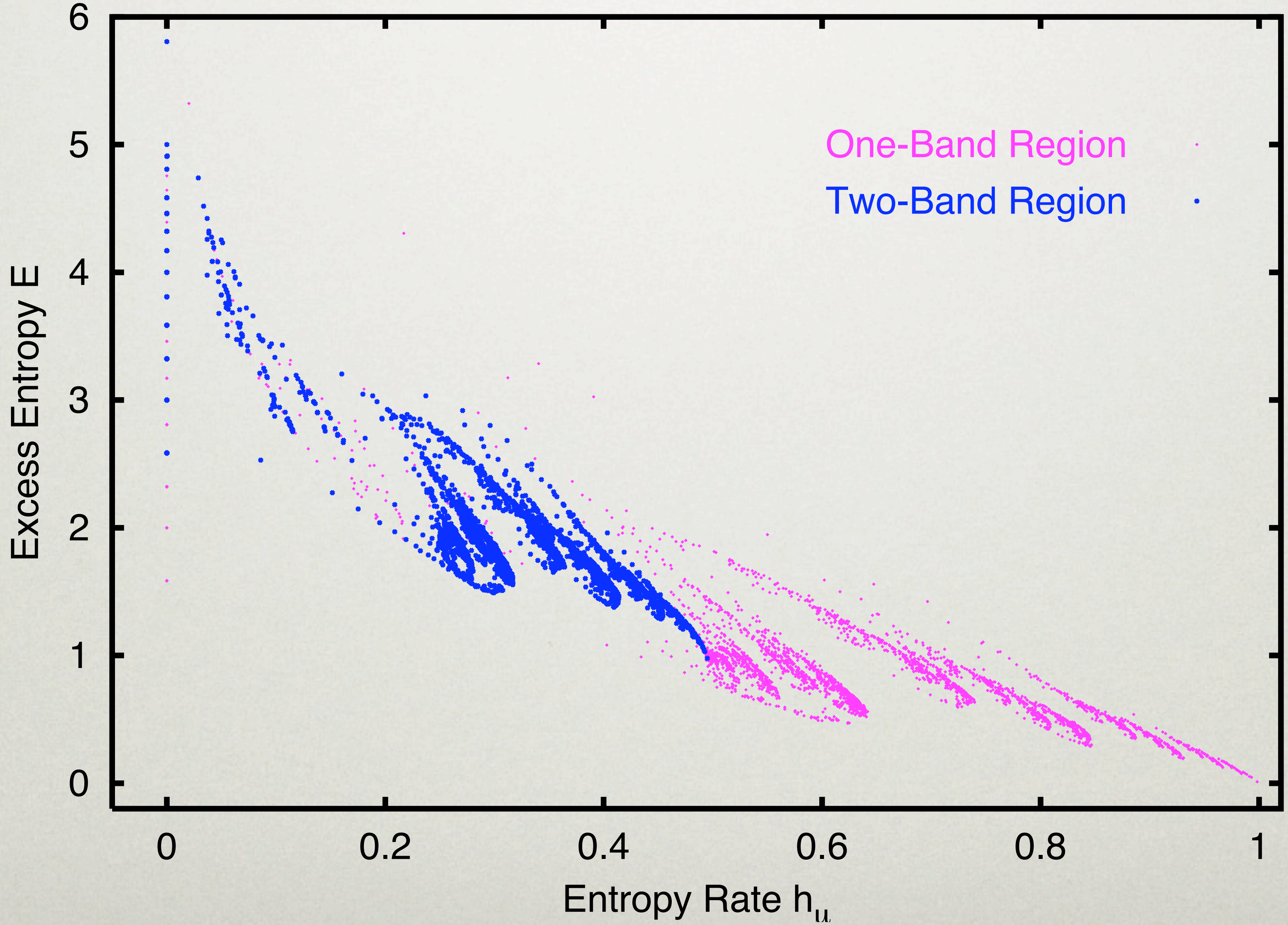
Complexity-Entropy Diagrams ... a Variety

2D Ising Spin System

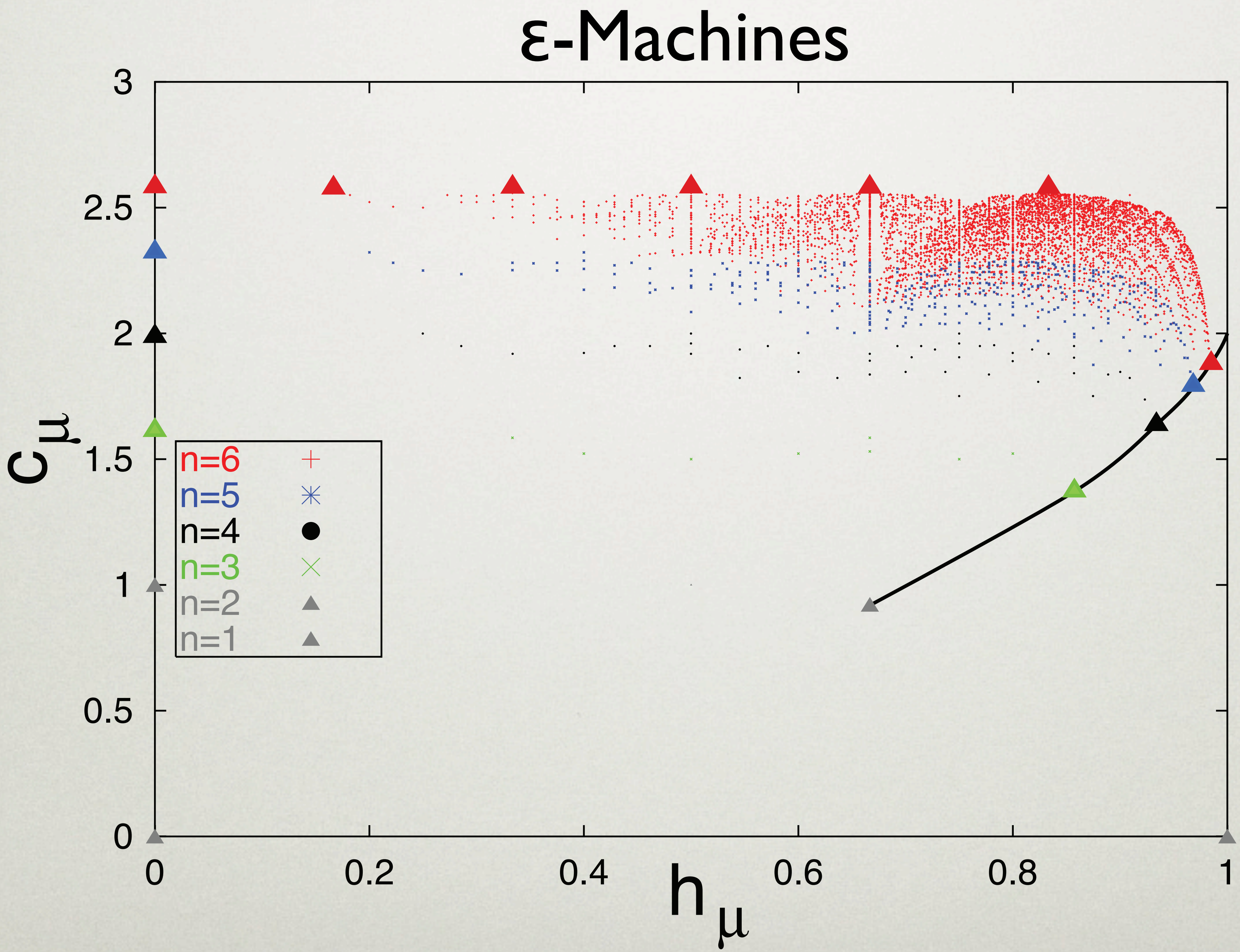


Complexity-Entropy Diagrams ... a Variety

Logistic Map



Complexity-Entropy Diagrams ... a Variety



D. P. Feldman, C. S. McTague, and J. P. Crutchfield, "The Organization of Intrinsic Computation: Complexity-Entropy Diagrams and the Diversity of Natural Information Processing", CHAOS **18**:4 (2008) 59-73.

What's New?

- Hidden processes: State information only available via measurement.
- So, how accessible is information?
- How are measurements connected to internal states?
- How is prediction connected to modeling?

Focus Problem

- Quantitative version:
  - Prediction ~ $\mathbf{E}$
  - Modeling ~ C_μ
- So, how are these related?

Focus Problem

- Known:
  - Spin systems:

$$C_\mu = \mathbf{E} + R h_\mu$$

J. P. Crutchfield and D. P. Feldman, "Statistical Complexity of Simple One-Dimensional Spin Systems", Physical Review E **55**:2 (1997) R1239-R1243.

  - Generally,

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

C. R. Shalizi and J. P. Crutchfield, "Computational Mechanics: Pattern and Prediction, Structure and Simplicity", J. Stat. Phys. **104** (2001) 817-879.

Focus Problem

- Can get h_μ and C_μ from ε-Machine.
- How to calculate $\mathbf{E}$ from ε-Machine?
- Back to the big issues at the beginning (relating modeling and prediction), but with a new "invariant": information accessibility.

Directional Computational Mechanics

- Previously, "forward" ε-Machines:
  - Equivalence relation: $\epsilon^+ \qquad \overrightarrow{x} \sim^+ \overrightarrow{x}'$
  - States: $\boldsymbol{\mathcal{S}}^+$
  - Machine: M^+
  - Entropy Rate: h_μ^+
  - Statistical Complexity: C_μ^+

$$\overleftrightarrow{X} = \ldots X_{-2} X_{-1} X_0 X_1 X_2 \ldots$$

$\longrightarrow$

Scan direction

Directional Computational Mechanics

- Now, "reverse" ε-Machines:

$$\overleftrightarrow{X} = \ldots X_{-2} X_{-1} X_0 X_1 X_2 \ldots$$

$\xleftarrow{\hspace{6em}}$

Scan direction

- Retrodictive equivalence relation: $\overrightarrow{x} \sim^{-} \overrightarrow{x}'$

$$\epsilon^{-}(\overrightarrow{x}) = \{\overrightarrow{x}' : \Pr(\overleftarrow{X}|\overrightarrow{x}) = \Pr(\overleftarrow{X}|\overrightarrow{x}')\}$$

Directional Computational Mechanics

- Retrodictive causal states: $\boldsymbol{\mathcal{S}}^- = \Pr(\overleftarrow{X}, \overrightarrow{X}) / \sim^-$

- Reverse ε-Machine: M^-

- Retrodictive entropy rate: h_μ^-

- Reverse statistical complexity: $C_\mu^- \equiv H[\mathcal{S}^-]$

Directional Computational Mechanics

- Which time direction is most predictable?

$$h_\mu^- = h_\mu^+ \qquad \text{Neither!}$$

- How much information must be stored?

$$C_\mu^- \neq C_\mu^+$$

- Temporal asymmetry in structure!

Directional Computational Mechanics

			Past	Present	Future				
			$\overleftarrow{X}$		$\overrightarrow{X}$				
$\ldots$	X_{-3}	X_{-2}	X_{-1}		X_0	X_1	X_2	$\ldots$	
$\ldots \mathcal{S}^+_{-3}$	$\mathcal{S}^+_{-2}$	$\mathcal{S}^+_{-1}$		$\mathcal{S}^+_0$		$\mathcal{S}^+_1$	$\mathcal{S}^+_2$	$\mathcal{S}^+_3 \ldots$	
$\ldots \mathcal{S}^-_{-3}$	$\mathcal{S}^-_{-2}$	$\mathcal{S}^-_{-1}$		$\mathcal{S}^-_0$		$\mathcal{S}^-_1$	$\mathcal{S}^-_2$	$\mathcal{S}^-_3 \ldots$	

Hidden Process Lattice

Directional Computational Mechanics

- Theorem:

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

- Effective transmission capacity of communication channel between forward and reverse processes.

Directional Computational Mechanics

- Corollaries:

$$\begin{aligned}\mathbf{E} &= C_\mu^+ - H[\mathcal{S}^+|\overrightarrow{X}] \\ &= C_\mu^+ - H[\mathcal{S}^+|\mathcal{S}^-]\end{aligned}$$

$$\begin{aligned}\mathbf{E} &= C_\mu^- - H[\mathcal{S}^-|\overleftarrow{X}] \\ &= C_\mu^- - H[\mathcal{S}^-|\mathcal{S}^+]\end{aligned}$$

Directional Computational Mechanics

- Corollaries:

$$\begin{aligned}\mathbf{E} &= C_\mu^+ - H[\mathcal{S}^+|\overrightarrow{X}] \\ &= C_\mu^+ - H[\mathcal{S}^+|\mathcal{S}^-]\end{aligned}$$

What are these terms?

$$\begin{aligned}\mathbf{E} &= C_\mu^- - H[\mathcal{S}^-|\overleftarrow{X}] \\ &= C_\mu^- - H[\mathcal{S}^-|\mathcal{S}^+]\end{aligned}$$

Directional Computational Mechanics

- Bidirectional Machine: $M^{\pm}$
- Equivalence relation:

$$\epsilon^{\pm}(\overleftrightarrow{x}) = \{(\overleftarrow{x}', \overrightarrow{x}') : \overleftarrow{x}' \in \epsilon^{+}(\overleftarrow{x}) \text{ and } \overrightarrow{x}' \in \epsilon^{-}(\overrightarrow{x})\}$$

- Bidirectional States:

$$\begin{aligned} \boldsymbol{\mathcal{S}}^{\pm} &= \Pr(\overleftarrow{X}, \overrightarrow{X})/\!\sim^{\pm} \\ &\subset \boldsymbol{\mathcal{S}}^{+} \times \boldsymbol{\mathcal{S}}^{-} \end{aligned}$$

Directional Computational Mechanics

- Bidirectional Machine:

$$M^{\pm} = \{\boldsymbol{\mathcal{S}}^{\pm}; \mathcal{T}^{(x)}, x \in \mathcal{A}\}$$

- Statistical Complexity:

$$C_\mu^{\pm} \equiv H[\mathcal{S}^{\pm}] = H[\mathcal{S}^{+}, \mathcal{S}^{-}]$$

Directional Computational Mechanics

- Excess entropy: $\mathbf{E} = C_\mu^+ + C_\mu^- - C_\mu^\pm$
- Only when $\mathbf{E} = 0$: $C_\mu^\pm = C_\mu^+ + C_\mu^-$
- $M^\pm$ is efficient: $C_\mu^\pm \leq C_\mu^+ + C_\mu^-$
- Bounds: $C_\mu^+ \leq C_\mu^\pm \qquad C_\mu^- \leq C_\mu^\pm$

Directional Computational Mechanics

- Temporal asymmetry:

$$C_\mu^- \neq C_\mu^+$$

- Causal Irreversibility:

$$\begin{aligned} \Xi &\equiv C_\mu^+ - C_\mu^- \\ &= H[\mathcal{S}^+|\mathcal{S}^-] - H[\mathcal{S}^-|\mathcal{S}^+] \end{aligned}$$

- Time-symmetric component ($\mathbf{E}$) cancels!

J. P. Crutchfield, "Semantics and Thermodynamics", in **Nonlinear Modeling and Forecasting**, M. Casdagli and S. Eubank, editors, Addison-Wesley, Reading, Massachusetts (1992) 317-359.

Directional Computational Mechanics

$$C_\mu^\pm = \mathbf{E} + H[\mathcal{S}^+|\mathcal{S}^-] + H[\mathcal{S}^-|\mathcal{S}^+]$$

- Crypticity:

$$\chi \equiv H[\mathcal{S}^+|\mathcal{S}^-] + H[\mathcal{S}^-|\mathcal{S}^+]$$

Distance between measurements & model:

$$d(X,Y) = H[X|Y] + H[Y|X]$$

Information inaccessibility!

How much internal information is hidden.

Directional Computational Mechanics

- Directional crypticities:

$$\begin{aligned}\chi^+ &= H[\mathcal{S}^+|\overrightarrow{X}] \\ &= H[\mathcal{S}^+|\mathcal{S}^-]\end{aligned}$$

$$\begin{aligned}\chi^- &= H[\mathcal{S}^-|\overleftarrow{X}] \\ &= H[\mathcal{S}^-|\mathcal{S}^+]\end{aligned}$$

k-Cryptic Processes

- k-Cryptic:

$$H[\mathcal{S}_k|\overrightarrow{X}_0] = 0$$

- No general way to know the order k.
- k-Cryptic approximation:

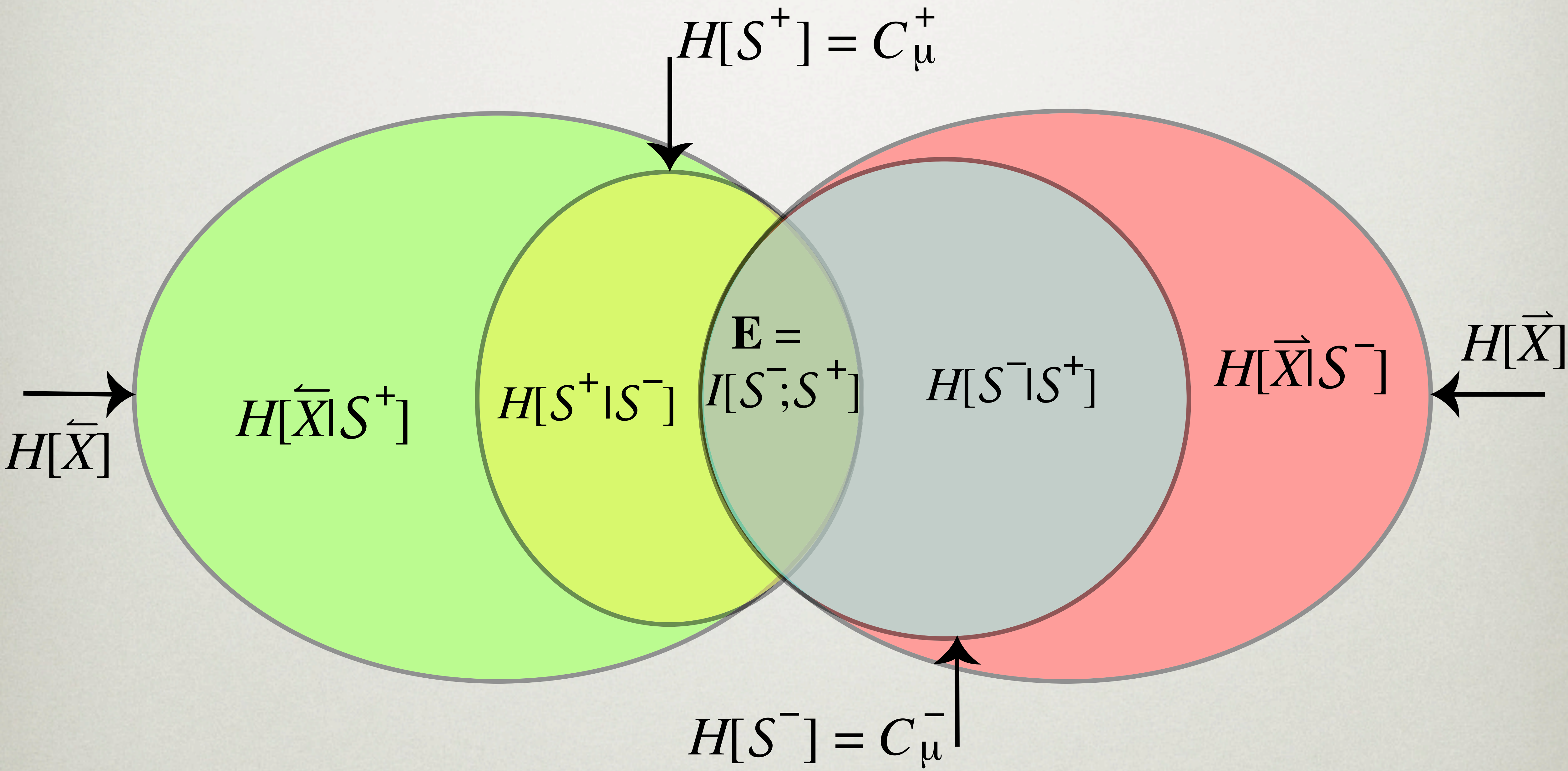
$$\chi(k) = H[\mathcal{S}_0|X_0^k, \mathcal{S}_k]$$

- That is,

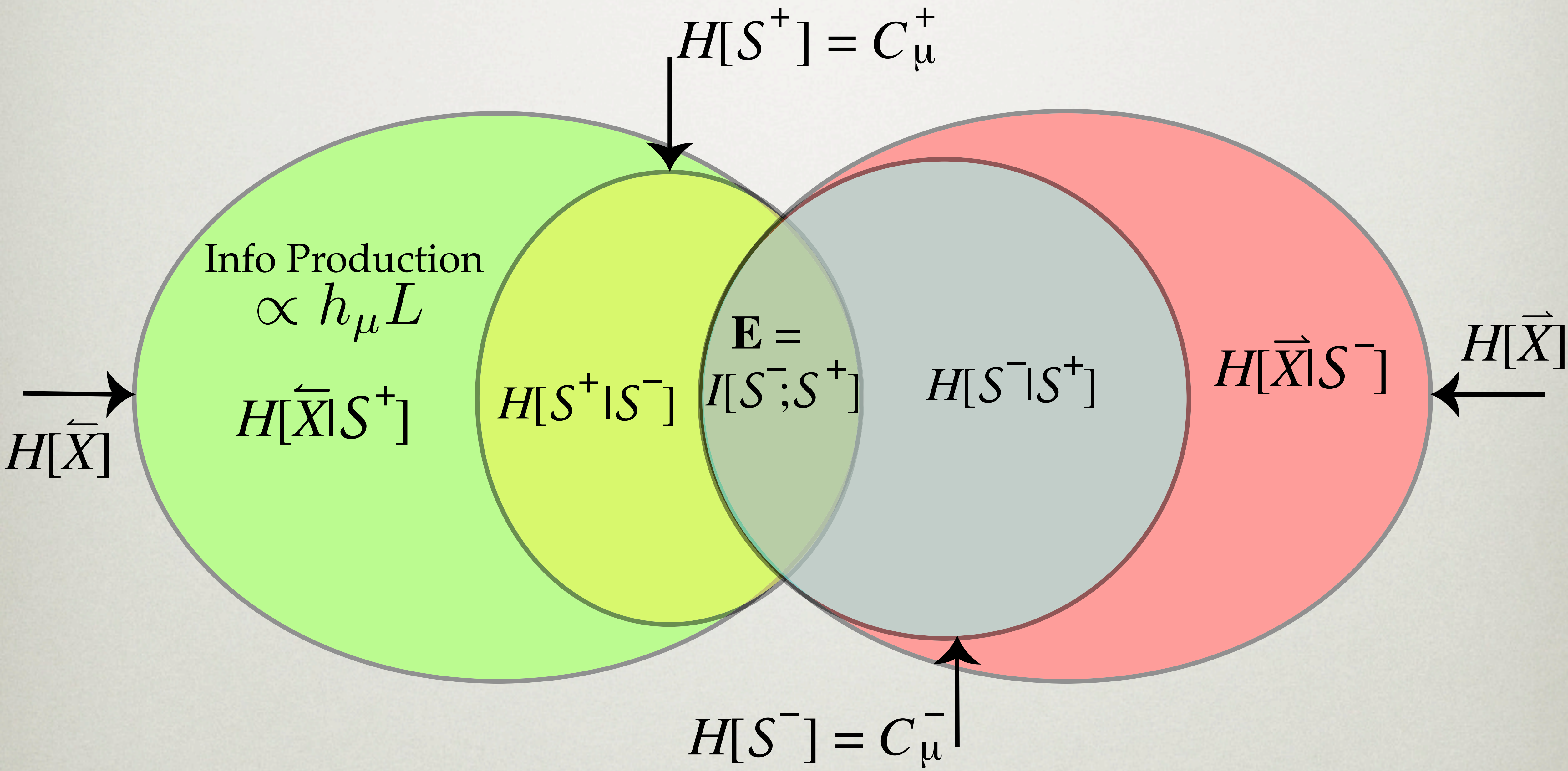
$$\chi = \chi(k) + H[\mathcal{S}_k|\overrightarrow{X}_0]$$

Error term

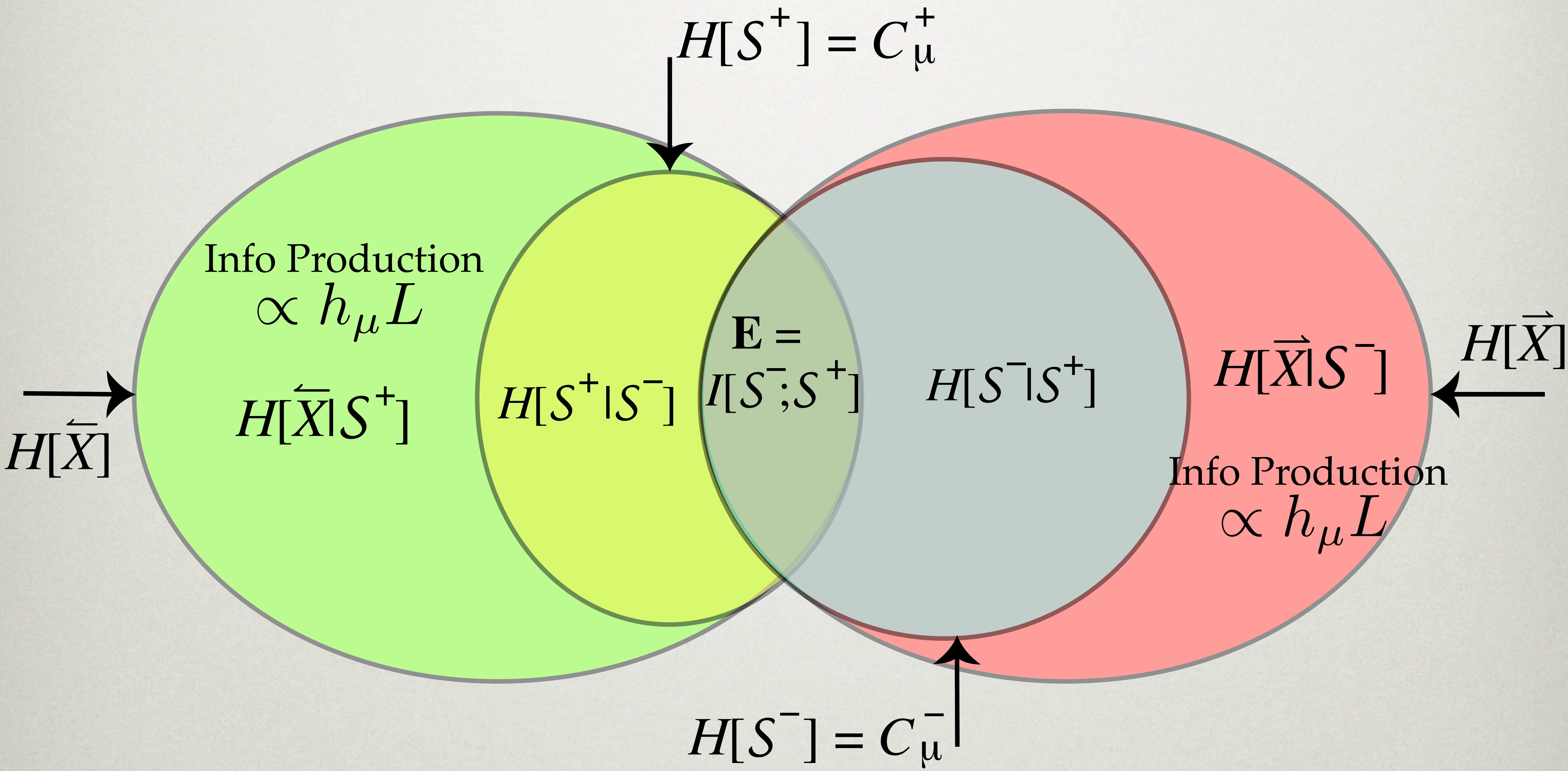
ϵ-machine information Diagram



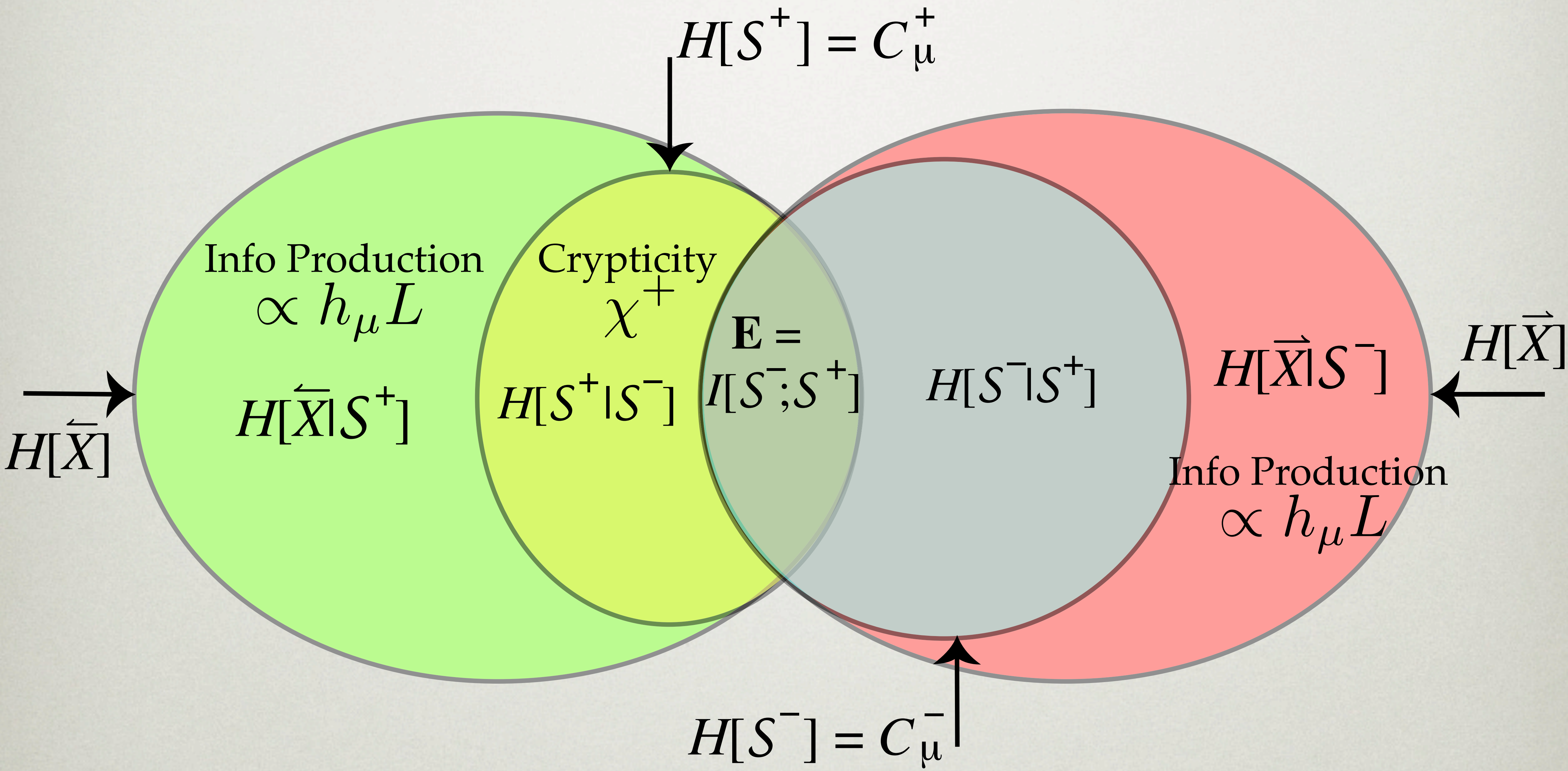
ϵ-machine Information Diagram



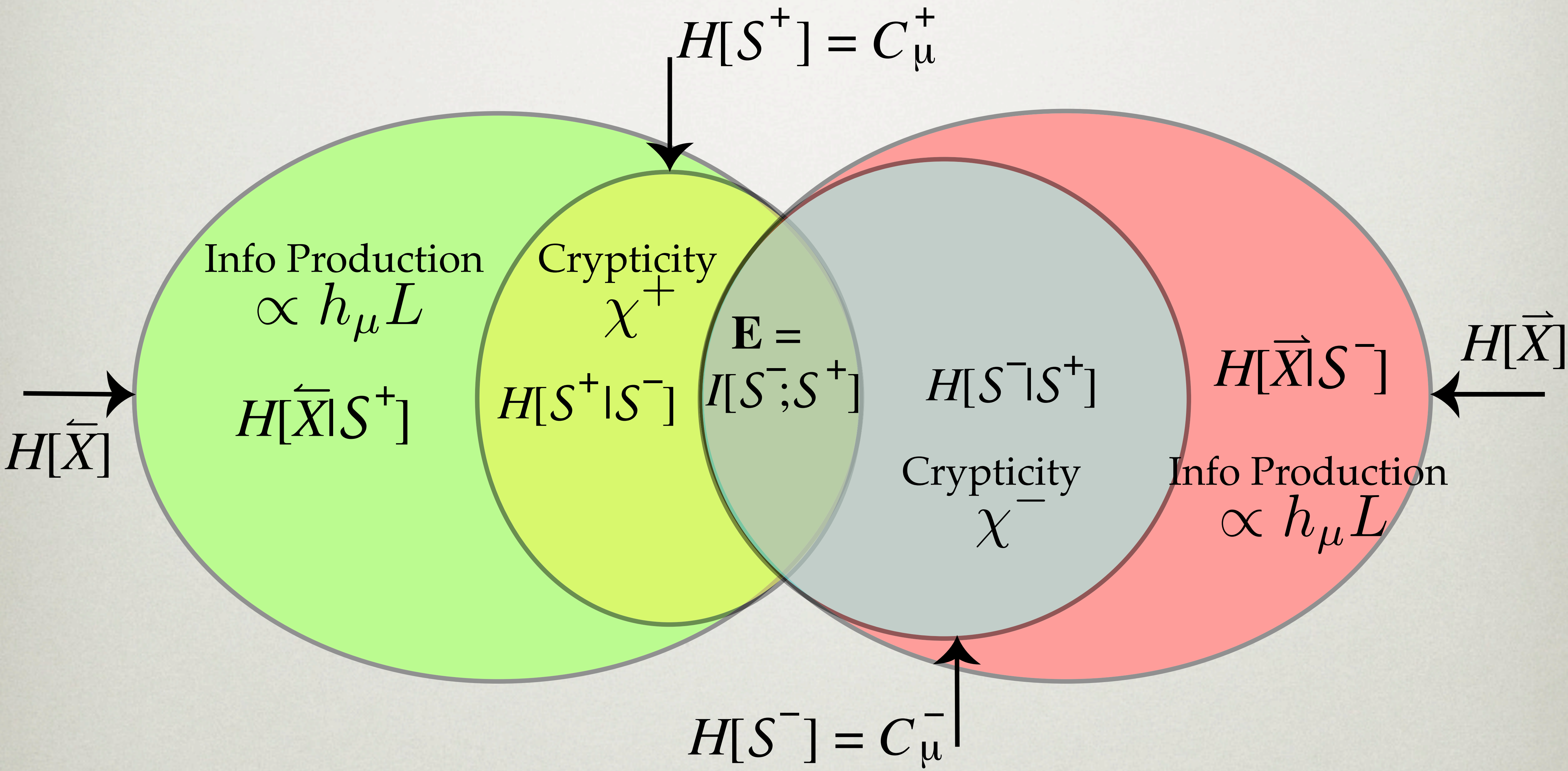
ϵ-machine information Diagram



ϵ-MACHINE INFORMATION DIAGRAM



ϵ-MACHINE INFORMATION DIAGRAM



The Past and the Future in the Present

- What right to use the "future"? It hasn't happened yet!

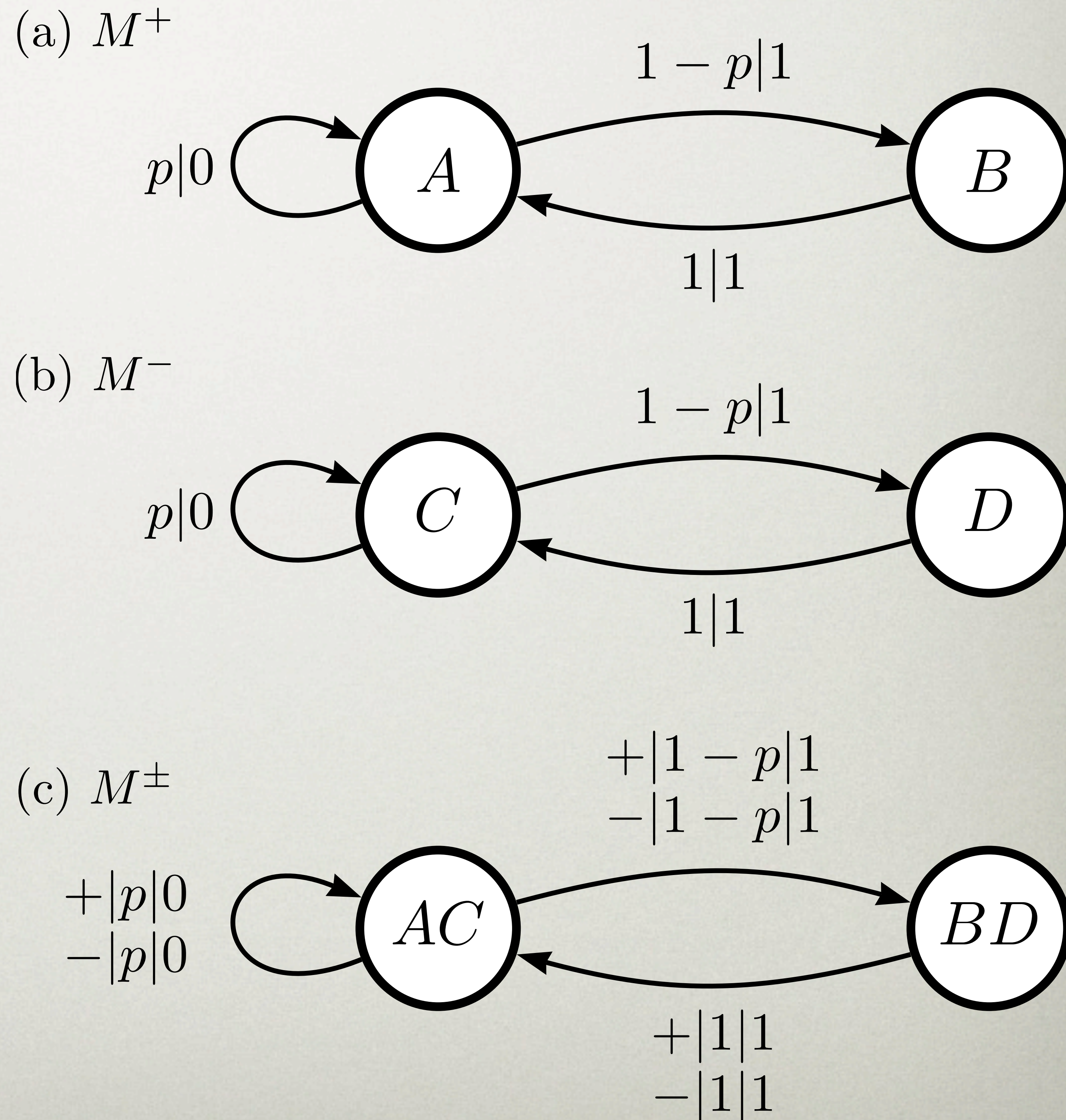
- Any time you write down a model, you make a commitment not only to what histories can occur, but also what futures can be generated.
- A model (equations of motion) has both past and future built in.
- Ditto empirics: Measure time series; once measured, then "futures" are available.

Examples

- Even Process

$$T^{(0)} = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} p & 0 \\ 0 & 0 \end{pmatrix} \end{array}$$

$$T^{(1)} = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 0 & 1-p \\ 1 & 0 \end{pmatrix} \end{array}$$



Examples

- Even Process

$$\Pr(\mathcal{S}^+) = \begin{array}{cc} A & B \\ \left(\frac{1}{2-p} \right. & \left. \frac{1-p}{2-p} \right) \end{array} \qquad \Pr(\mathcal{S}^-) = \begin{array}{cc} C & D \\ \left(\frac{1}{2-p} \right. & \left. \frac{1-p}{2-p} \right) \end{array}$$

Causally reversible: $\Xi = 0$

$$C_\mu^+ = H\left(1/(2-p)\right)$$

$$h_\mu = H(p)/(2-p)$$

$\chi^\pm = 0 \Rightarrow \mathbf{E} = C_\mu^\pm$ A 0-cryptic process

Examples

- Even Process bidirectional machine:

$$\Pr(\mathcal{S}^-|\mathcal{S}^+) \quad = \begin{array}{c c} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \qquad \Pr(\mathcal{S}^+|\mathcal{S}^-) \quad = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}$$

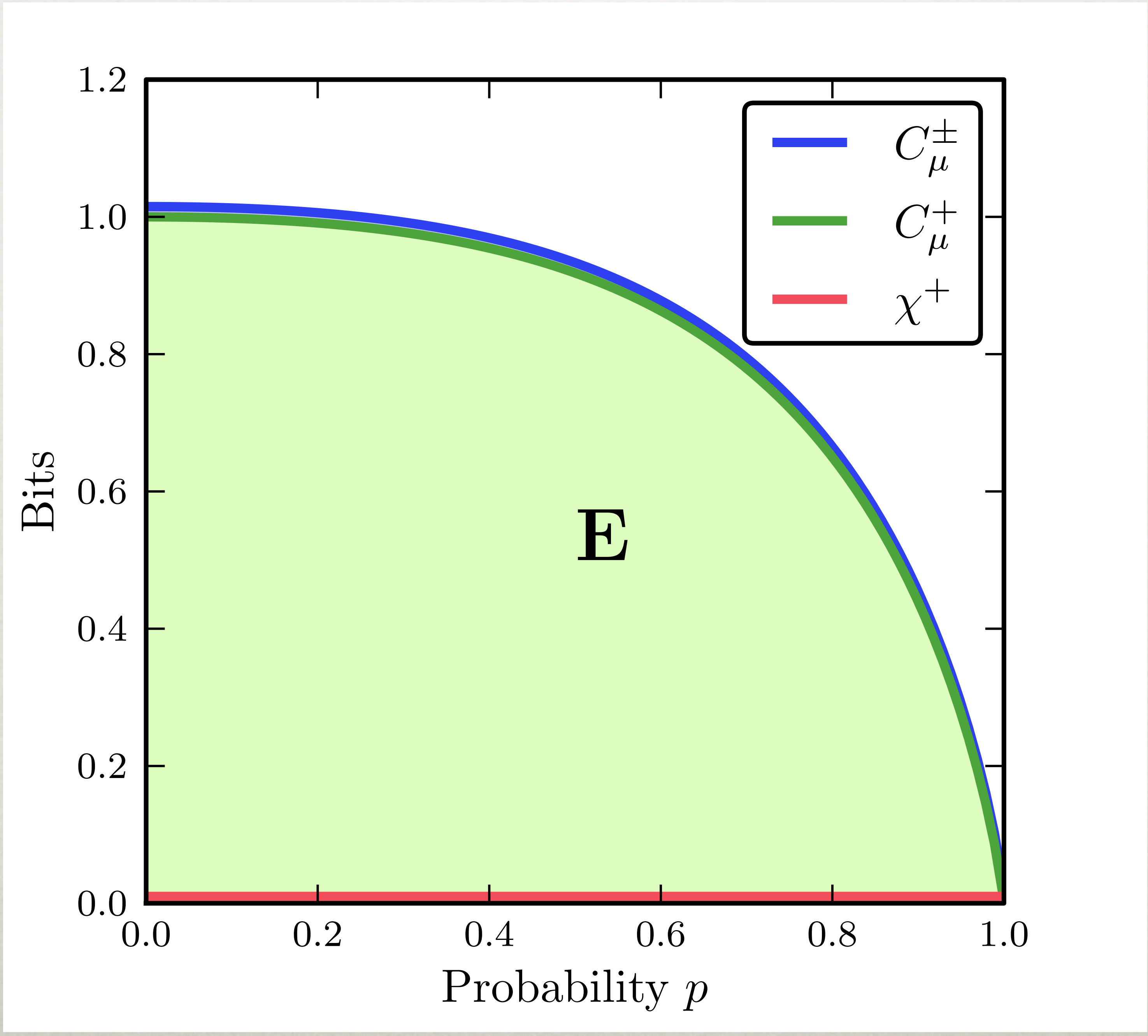
$$p = 1/2 \qquad \Pr(\mathcal{S}^\pm) = \Pr(AC, BD) = (2/3, 1/3)$$

$$C_\mu^\pm = H[\mathcal{S}^\pm] = H(2/3) \approx 0.9183 \text{ bits}$$

$$\mathbf{E} = I[\mathcal{S}^+; \mathcal{S}^-] \approx 0.9183 \text{ bits}$$

Examples

- Even Process

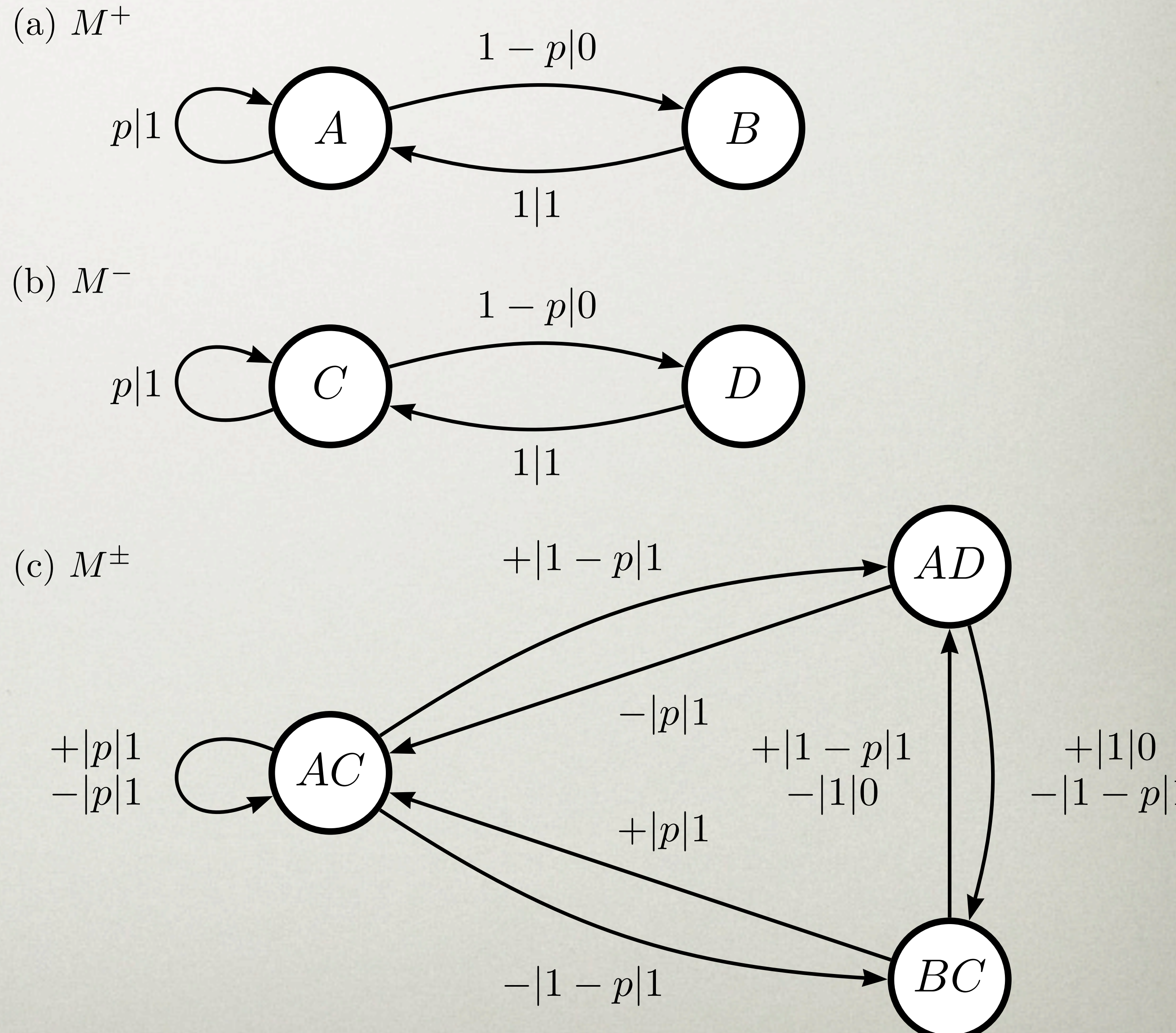


Examples

- Golden Mean Process

$$T^{(0)} = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 0 & 1-p \\ 0 & 0 \end{pmatrix} \end{array}$$

$$T^{(1)} = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} p & 0 \\ 1 & 0 \end{pmatrix} \end{array}$$



Examples

- Golden Mean Process

$$\Pr(\mathcal{S}^+) = \begin{array}{cc} A & B \\ \left(\frac{1}{2-p}\right. & \left.\frac{1-p}{2-p}\right) \end{array} \quad \Pr(\mathcal{S}^-) = \begin{array}{cc} C & D \\ \left(\frac{1}{2-p}\right. & \left.\frac{1-p}{2-p}\right) \end{array}$$

Causally reversible: $\Xi = 0$

$$C_\mu^+ = H\left(1/(2-p)\right)$$

$$h_\mu = H(p)/(2-p)$$

Just as with Even Process! But ...

$$\chi^\pm \neq 0 \qquad \mathbf{E} = H\left(\frac{1}{2-p}\right) - \frac{H(p)}{2-p}$$

$$\chi^\pm = H(p)/(2-p)$$

Examples

- Golden Mean Process bidirectional machine

$$\Pr(\mathcal{S}^+|\mathcal{S}^-) = \begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} p & 1-p \\ 1 & 0 \end{pmatrix} \end{array}$$

$$\Pr(\mathcal{S}^-|\mathcal{S}^+) = \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} p & 1-p \\ 1 & 0 \end{pmatrix} \end{array}$$

$$\Pr(\mathcal{S}^\pm) = \Pr(AC, AD, BC) = (p, 1-p, 1-p)\,/(2-p)$$

Examples

- Golden Mean Process

$$p = 1/2$$

$$C_\mu^\pm = H[\mathcal{S}^\pm] = \log_2 3 \approx 1.5850 \text{ bits}$$

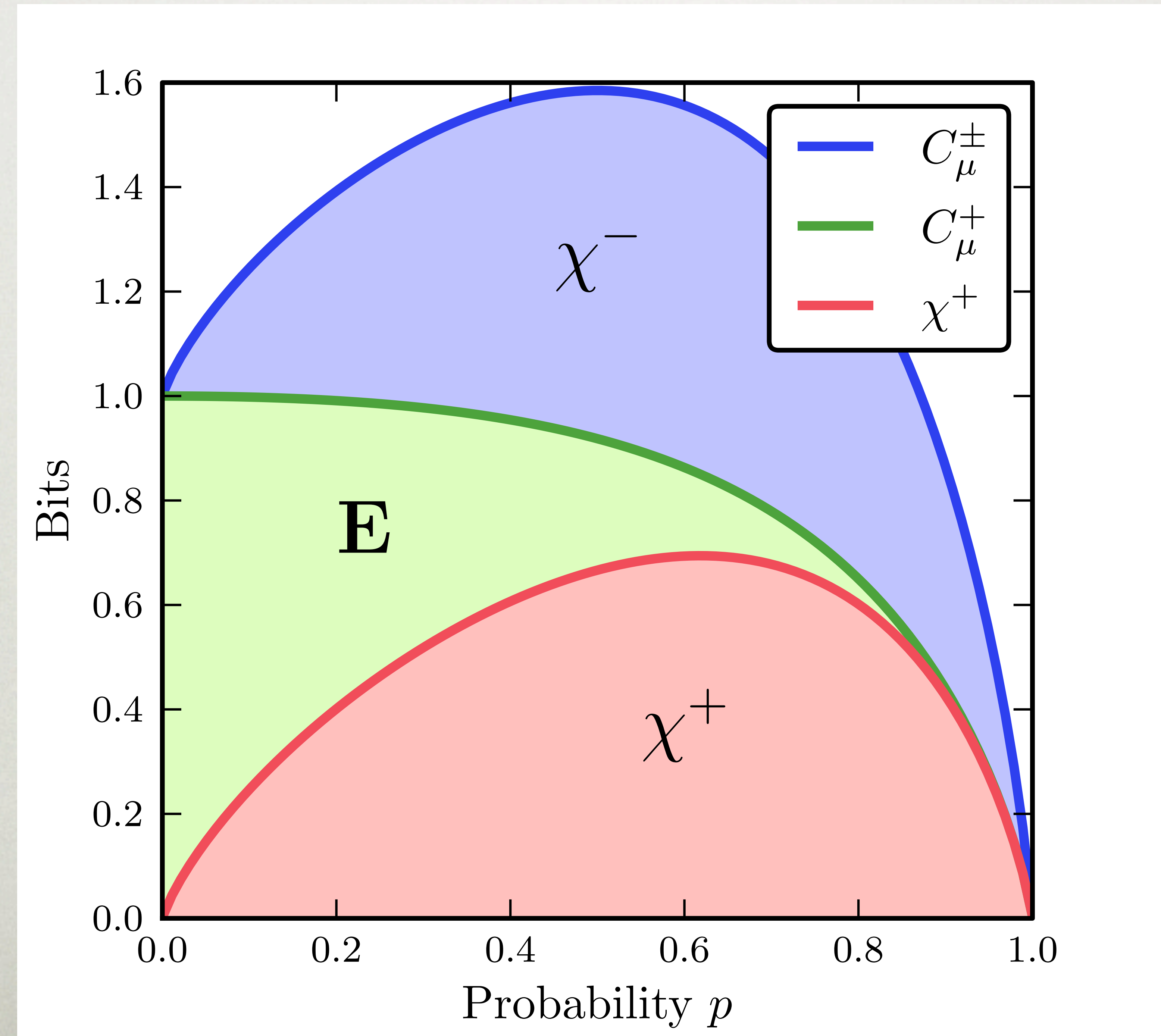
$$\mathbf{E} = I[\mathcal{S}^+;\mathcal{S}^-] \approx 0.2516 \text{ bits}$$

$$\chi^\pm \approx 1.3334 \text{ bits}$$

A cryptic process, with order k = 1.

Examples

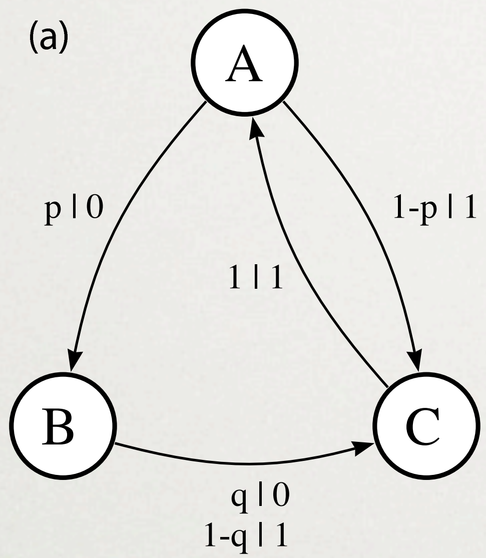
- Golden Mean Process



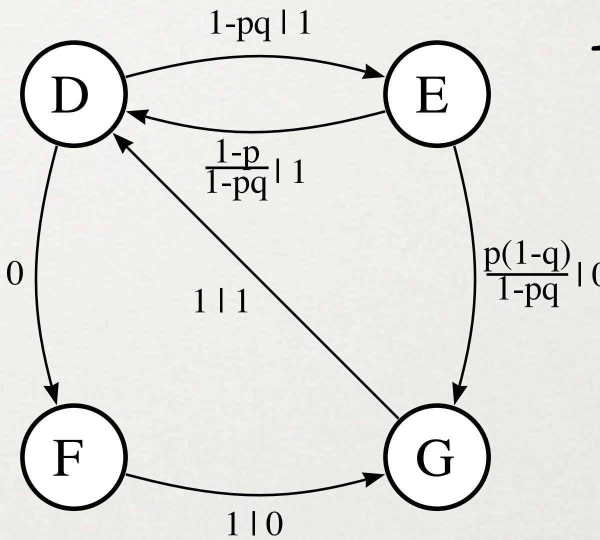
Examples ... Random Insertion Process

M^{+}

(a)



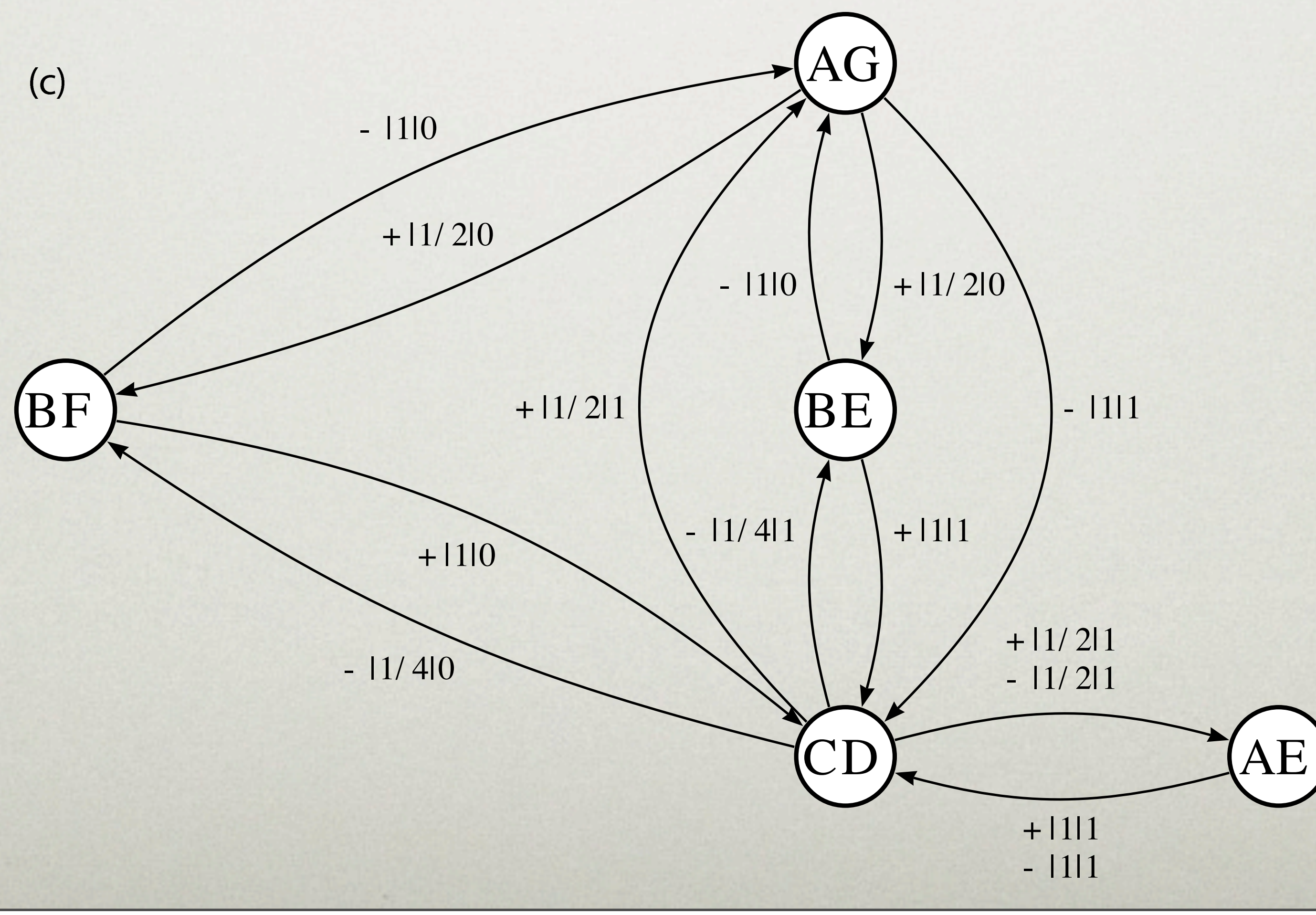
(b)



M^{-}

$M^{\pm}$

(c)



Examples ... Random Insertion Process

M^+ $\quad \Pr(A, B, C) = (1, p, 1)/(p+2)$

M^- $\quad \Pr(D, E, F, G) = (1, 1 - pq, pq, p)/(p+2)$

$p = q = 1/2 \quad C_\mu^+ \approx 1.5219 \text{ bits}$

$C_\mu^- \approx 1.8464 \text{ bits}$

$h_\mu = 3/5 \text{ bits/measurement}$

Causally irreversible:

$$\Xi \approx 0.3245 \text{ bits}$$

Examples ... Random Insertion Process

$M^{\pm}$

$$\Pr(\mathcal{S}^+, \mathcal{S}^-) = \tfrac{1}{(p+2)} \times \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{cccc} D & E & F & G \end{array} \\ \begin{pmatrix} 0 & 1-p & 0 & p \\ 0 & p(1-q) & pq & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\mathbf{E} = \underbrace{\log_2(p+2) - \frac{p \log_2 p}{p+2}}_{C_\mu} - \underbrace{\frac{1-pq}{p+2} H\left(\frac{1-p}{1-pq}\right)}_{\chi^+}$$

Examples ...
Random Insertion Process

$$M^{\pm} \qquad \begin{aligned}\Pr(\mathcal{S}^{\pm}) &= \Pr(AE, AG, BE, BF, CD) \\ &= (1/5, 1/5, 1/10, 1/10, 2/5)\end{aligned}$$

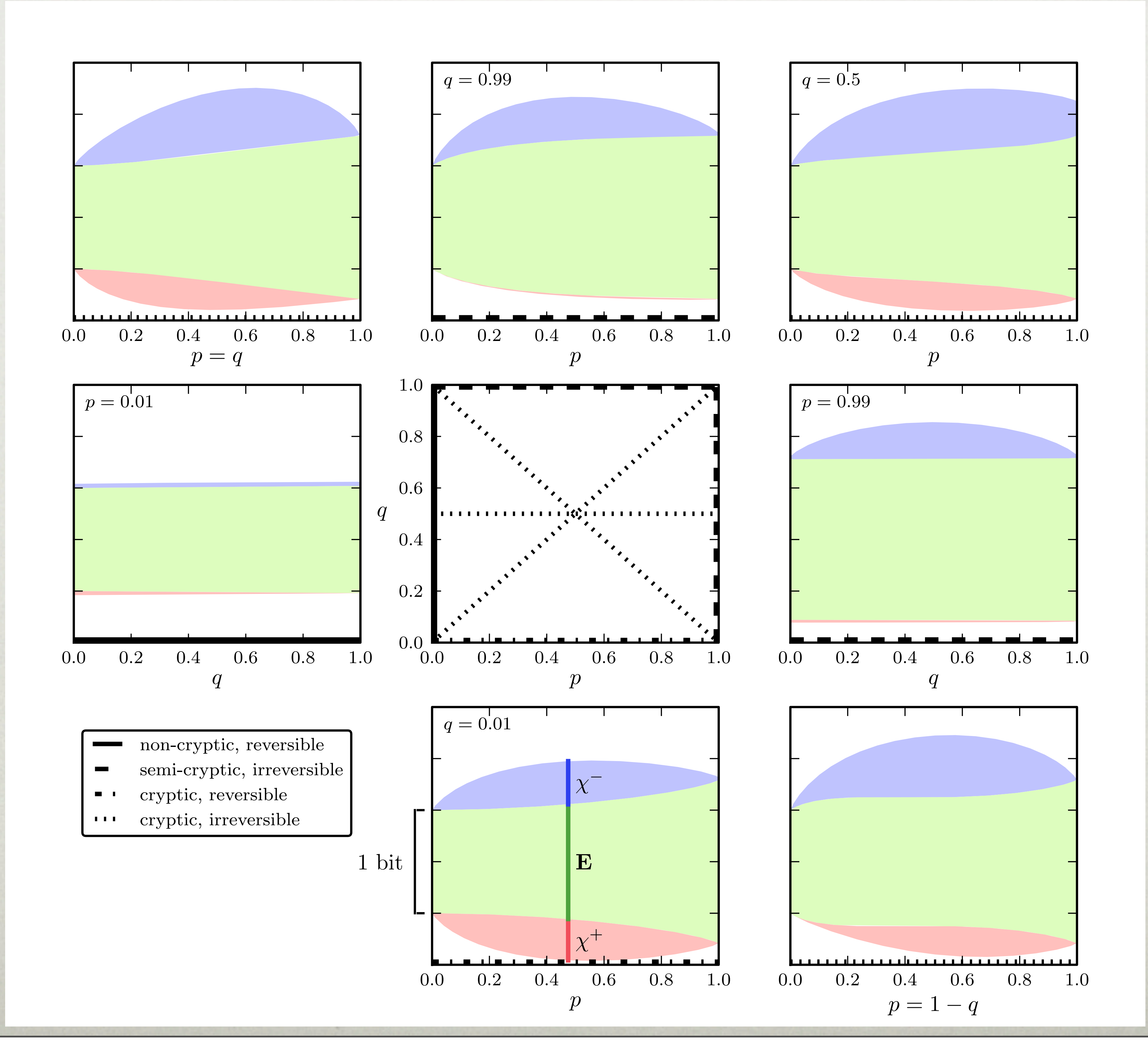
$$p = q = 1/2$$

$$C_{\mu}^{\pm} = H[\mathcal{S}^{\pm}] \approx 2.1219 \text{ bits}$$

$$\mathbf{E} = I[\mathcal{S}^{+}; \mathcal{S}^{-}] \approx 1.2464 \text{ bits}$$

$\chi \approx 0.8755$ bits   A rather cryptic process!

Examples ... Random Insertion Process



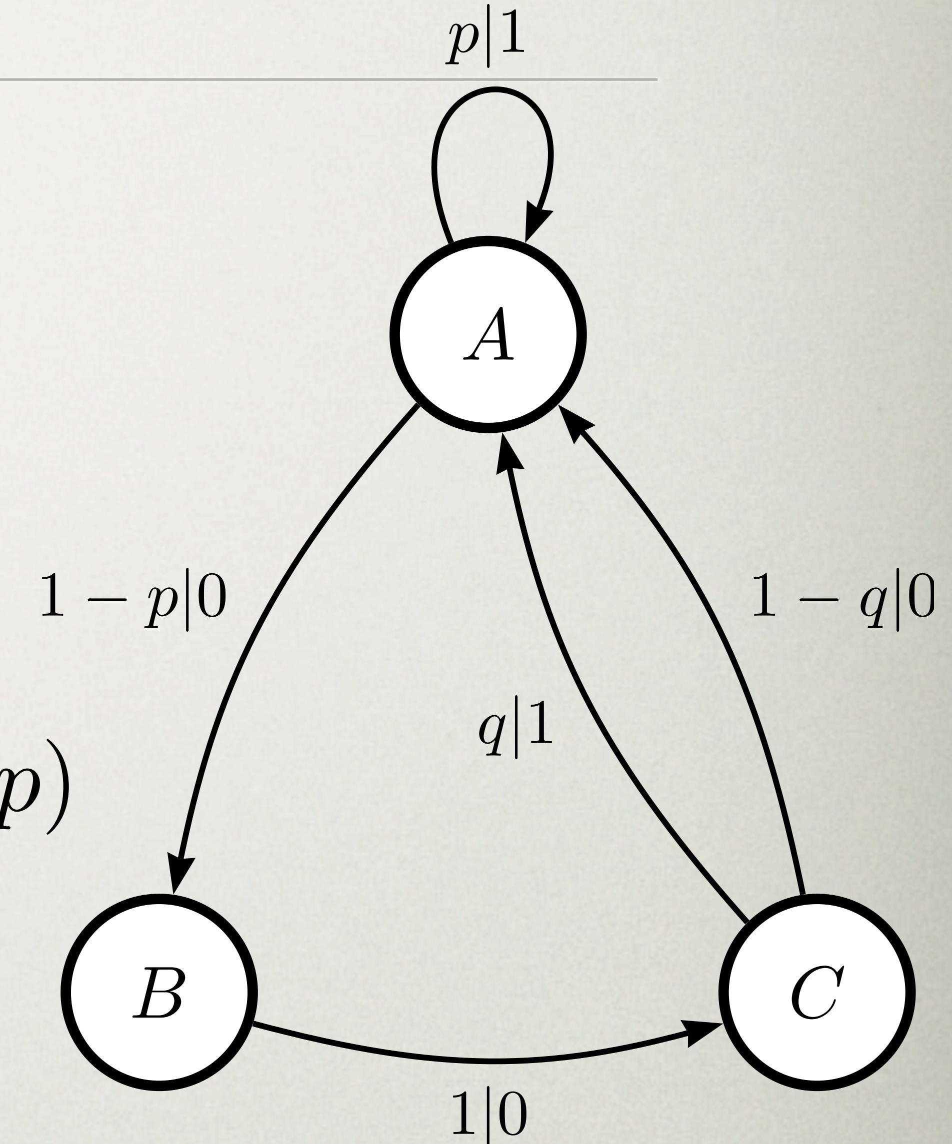
Examples

- Nemo Process: ∞-cryptic

$$\Pr(\mathcal{S}) = \frac{1}{3-2p} \begin{array}{ccc} A & B & C \\ \big(1 & 1-p & 1-p \big) \end{array}$$

$$C_\mu = \log_2(3-2p) - \tfrac{2(1-p)}{3-2p} \log_2(1-p)$$

Causally reversible: $C_\mu^+ = C_\mu^-$



∞-Cryptic: $H[\mathcal{S}_k|\overrightarrow{X}_0] > 0, k = 0, 1, 2, \ldots$

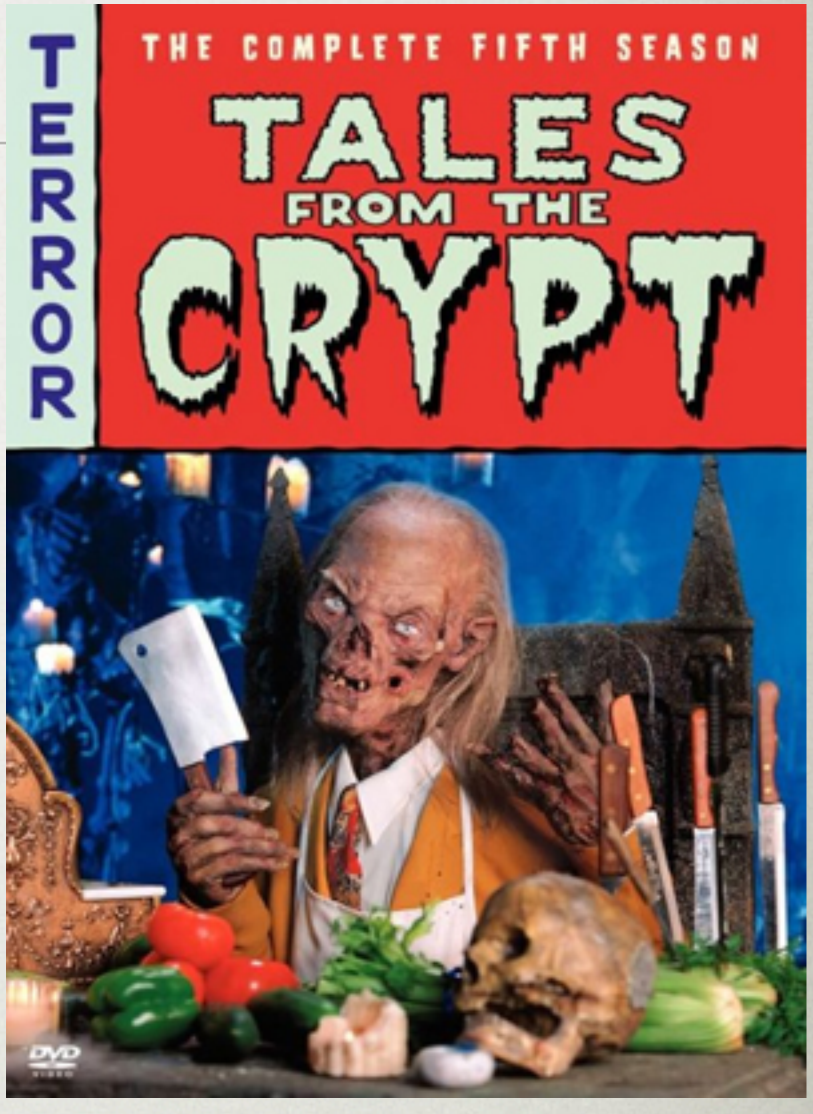
Examples

- Nemo Process: ∞-cryptic

$$
\Pr(\mathcal{S}^+|\mathcal{S}^-) = \frac{1}{p+q-pq}
\begin{array}{c c}
 & \begin{array}{ccc} A & B & C \end{array} \\
\begin{array}{c} D \\ E \\ F \end{array} &
\begin{pmatrix}
p & 0 & q(1-p) \\
0 & q & p(1-q) \\
q & p(1-q) & 0
\end{pmatrix}
\end{array}
$$

$$
\begin{aligned}
\mathbf{E} &= C_\mu - H[\mathcal{S}^+|\mathcal{S}^-] \\
&= \log_2(3-2p) - \frac{2(1-p)}{3-2p}\log_2(1-p) \\
&\quad - \frac{1}{3-2p}\left[\frac{p}{p+q-pq}\log_2\left(\frac{p+q-pq}{p}\right) + \frac{q(1-p)}{p+q-pq}\log_2\left(\frac{p+q-pq}{q(1-p)}\right)\right] \\
&\quad + \frac{2(1-p)}{3-2p}\left[\frac{q}{p+q-pq}\log_2\left(\frac{p+q-pq}{q}\right) + \frac{p(1-q)}{p+q-pq}\log_2\left(\frac{p+q-pq}{p(1-q)}\right)\right]
\end{aligned}
$$

Cautionary Tales From the Cryptic



- Cryptic Processes: Excess entropy can be arbitrarily small ($\mathbf{E} \approx 0$).
- Even for very structured ($C_\mu \gg 1$) processes.
- **Care** when applying informational analyses to complex systems; esp. mutual information.
- Best to focus on causal architecture, then calculate what you need.

Immediate Conclusions

- New meaning to excess entropy:

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

- New level of analytical calculation possible.
- New algorithms for complexity measures.

Broad Conclusions

- Prediction connected to modeling.
- Lesson: Must be very careful in empirically estimating complex systems organization: Very slow convergence!
- Temporal Computational Mechanics complete!

Lessons

Lessons

- Can define & measure degree of pattern

Lessons

- Can define & measure degree of pattern
- Emergence = Increase in stored information

Lessons

- Can define & measure degree of pattern

- Emergence = Increase in stored information

- How natures computes is

  how nature is structured

Computational Mechanics

Intrinsic computation:

1. How much historical information is stored in the present?
2. In what architecture is that information stored?
3. How is the stored information used to produce future behavior?

Computational Mechanics

Intrinsic computation:

1. How much historical information is stored in the present? C_μ
2. In what architecture is that information stored?
3. How is the stored information used to produce future behavior?

Computational Mechanics

Intrinsic computation:

1. How much historical information is stored in the present? C_μ
2. In what architecture is that information stored?

$$\epsilon - \mathrm{Machine}$$

3. How is the stored information used to produce future behavior?

Computational Mechanics

Intrinsic computation:

1. How much historical information is stored in the present? C_μ
2. In what architecture is that information stored?

$$\epsilon - \text{Machine}$$

3. How is the stored information used to produce future behavior? h_μ

Thanks!

http://cse.ucdavis.edu/~chaos/