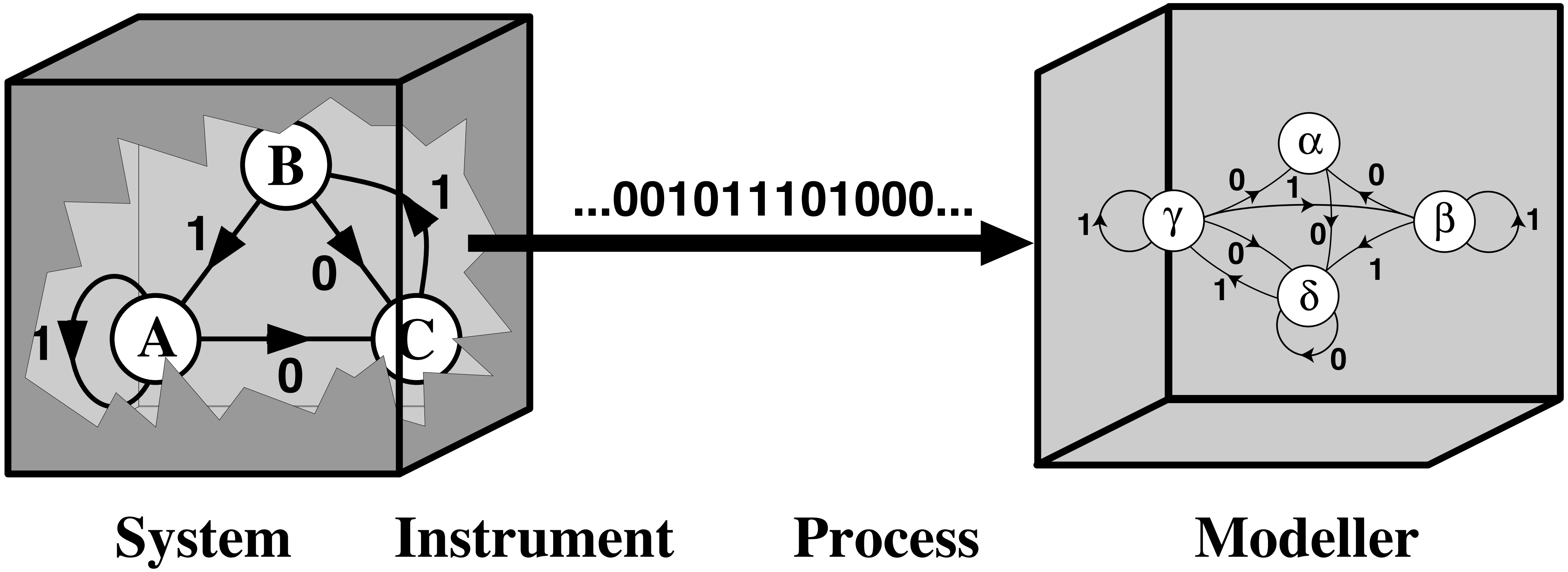
Complexity

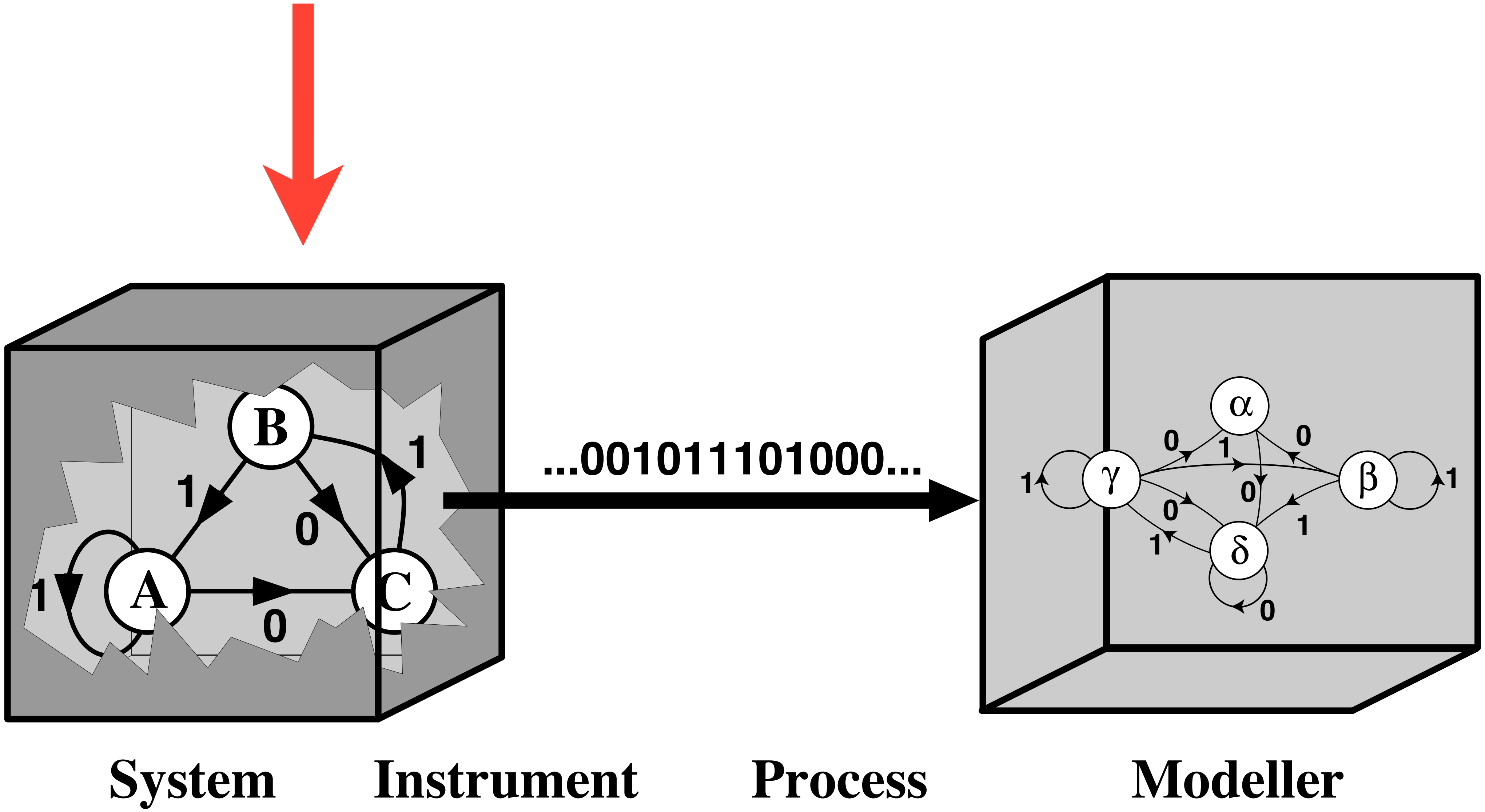
Jim Crutchfield & Nix Barnett
Complexity Sciences Center
Physics Department
University of California at Davis

Complex Systems Summer School
Santa Fe Institute
St. John's College, Santa Fe, NM
13 June 2013



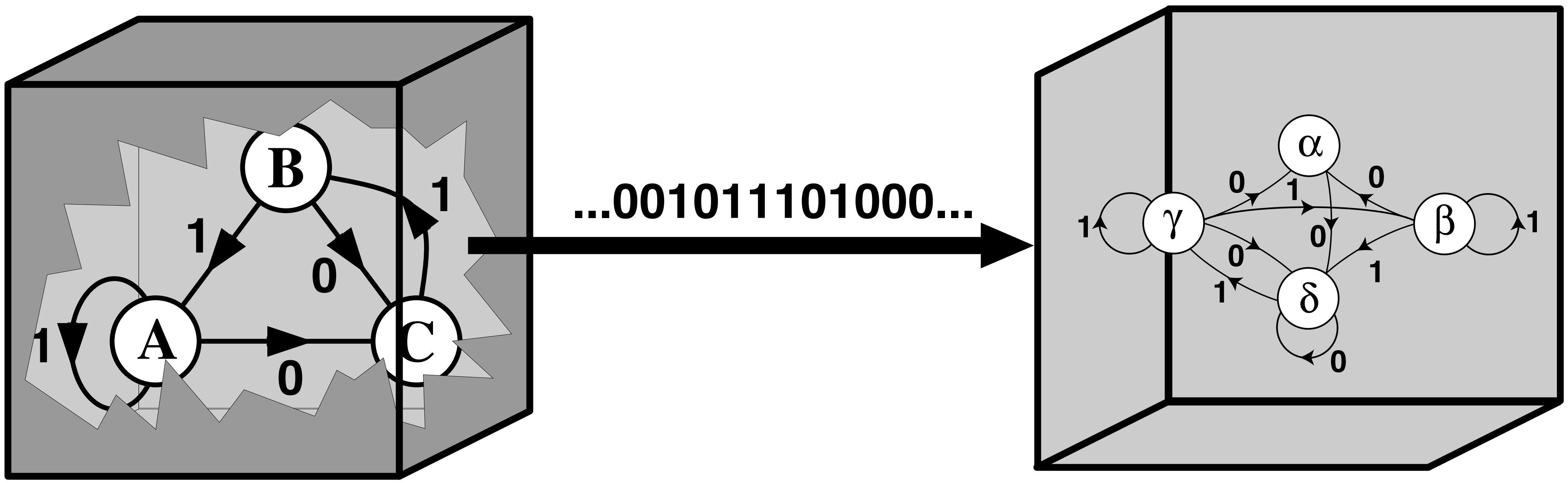
The Learning Channel

Last Week



The Learning Channel

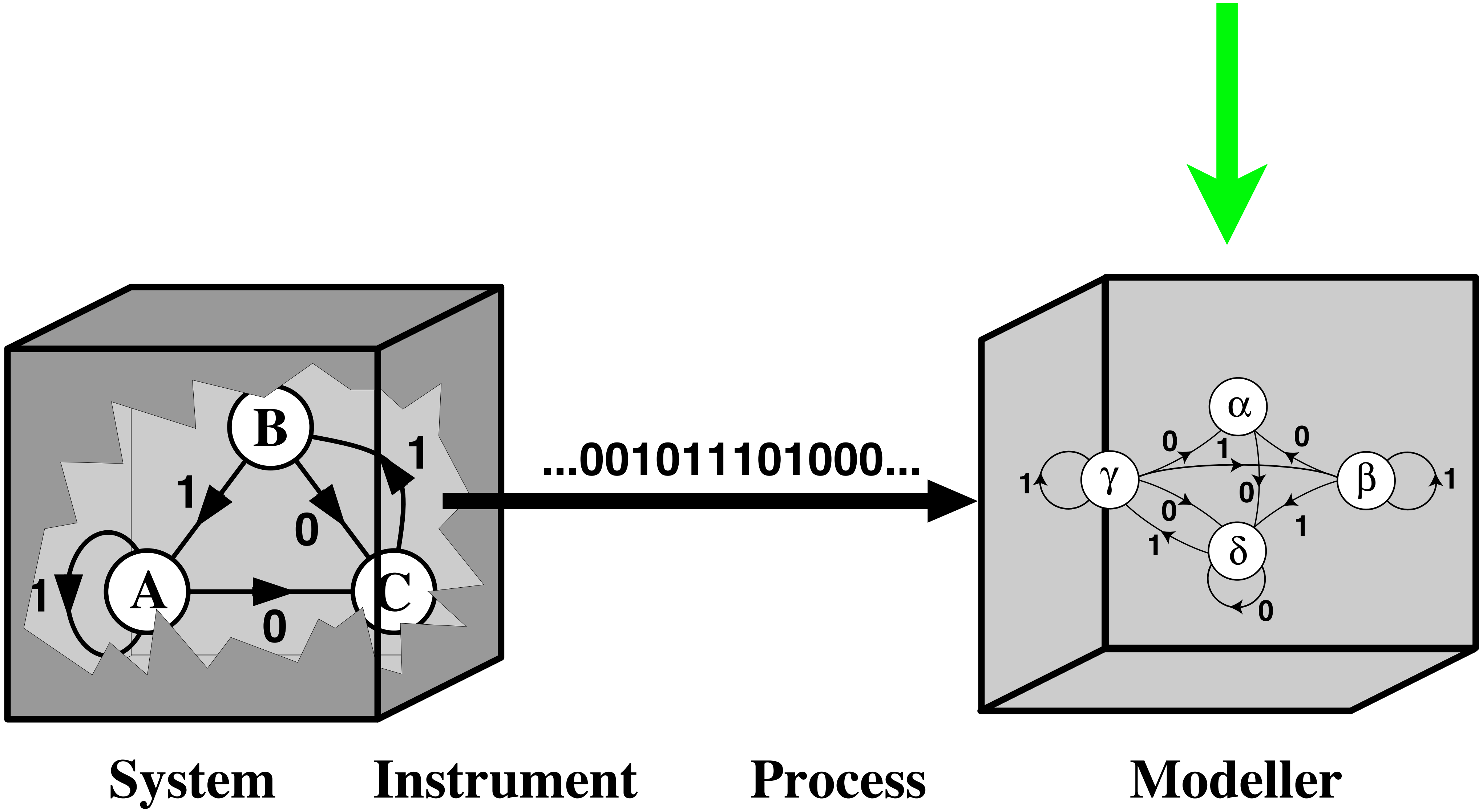
Today



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

The Learning Channel

Tomorrow



The Learning Channel

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/

Complexity

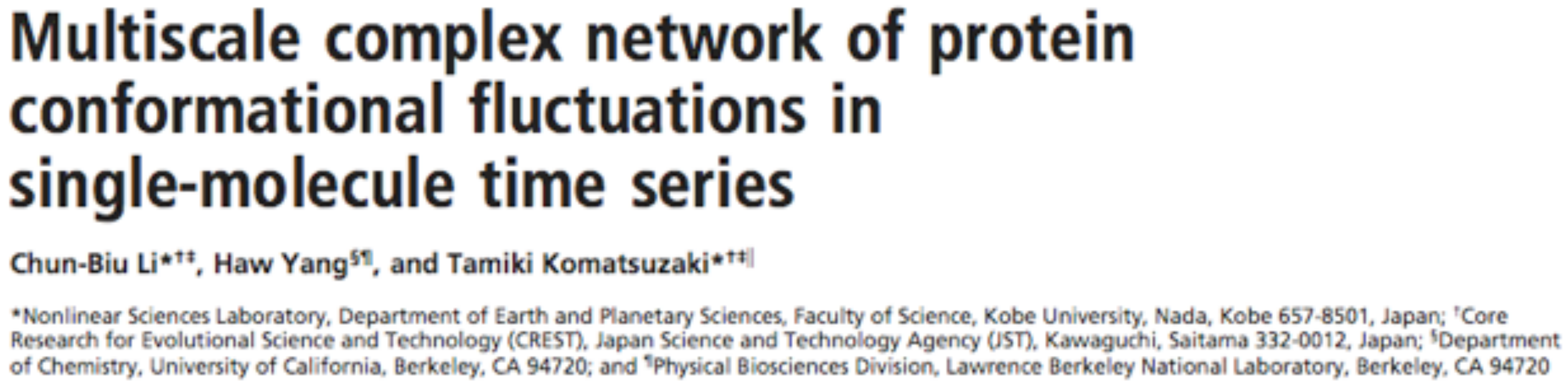
References? For example:

Stanislaw Lem, *Chance and Order*, New Yorker **59** (1984) 88-98.
T. Cover and J. Thomas, *Elements of Information Theory*,
    Wiley, Second Edition (2006) Chapters 1 - 7.
M. Li and P.M.B. Vitanyi, *An Introduction to Kolmogorov Complexity and its Applications*,
    Springer, New York (1993).
J. P. Crutchfield and D. P. Feldman,
    "Regularities Unseen, Randomness Observed: Levels of Entropy Convergence",CHAOS **13**:1 (2003) 25-54.
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.
R. G. James, C. J. Ellison, and J. P. Crutchfield,
    "Anatomy of a Bit: Information in a Time Series Observation",CHAOS **21**:1 (2011) 037109.
J. P. Crutchfield,
    "Between Order and Chaos", Nature Physics **8** (January 2012) 17-24.

See http://csc.ucdavis.edu/~cmg/

Applications?

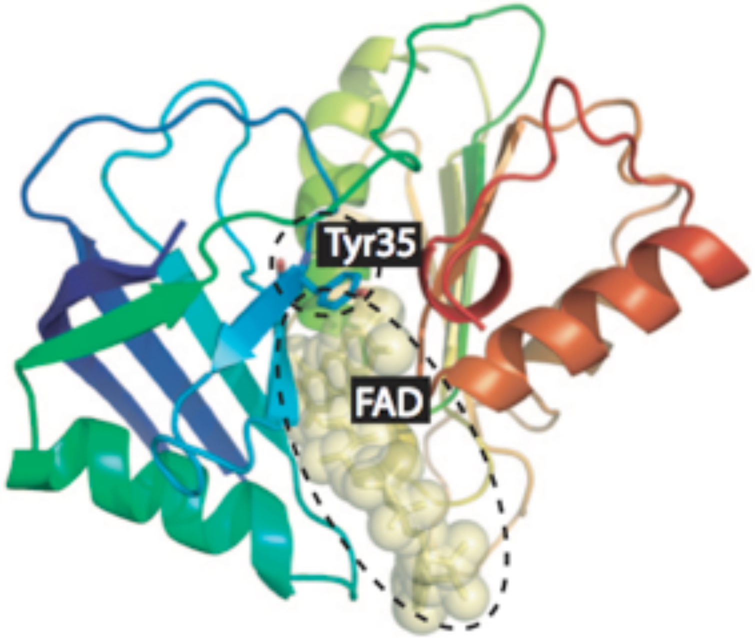
Computational Mechanics: Application to Experimental Molecular Dynamics Spectroscopy

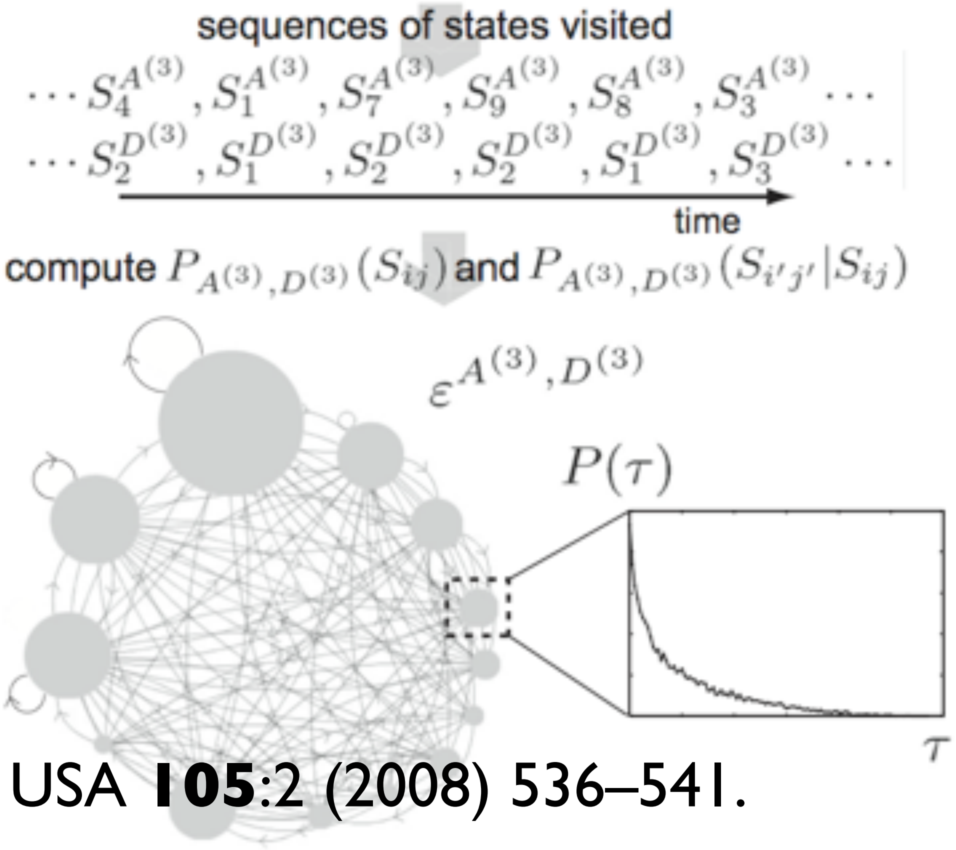


Multiscale complex network of protein conformational fluctuations in single-molecule time series

Chun-Biu Li*[†][‡], Haw Yang[§][¶], and Tamiki Komatsuzaki*[†][‡][‖]

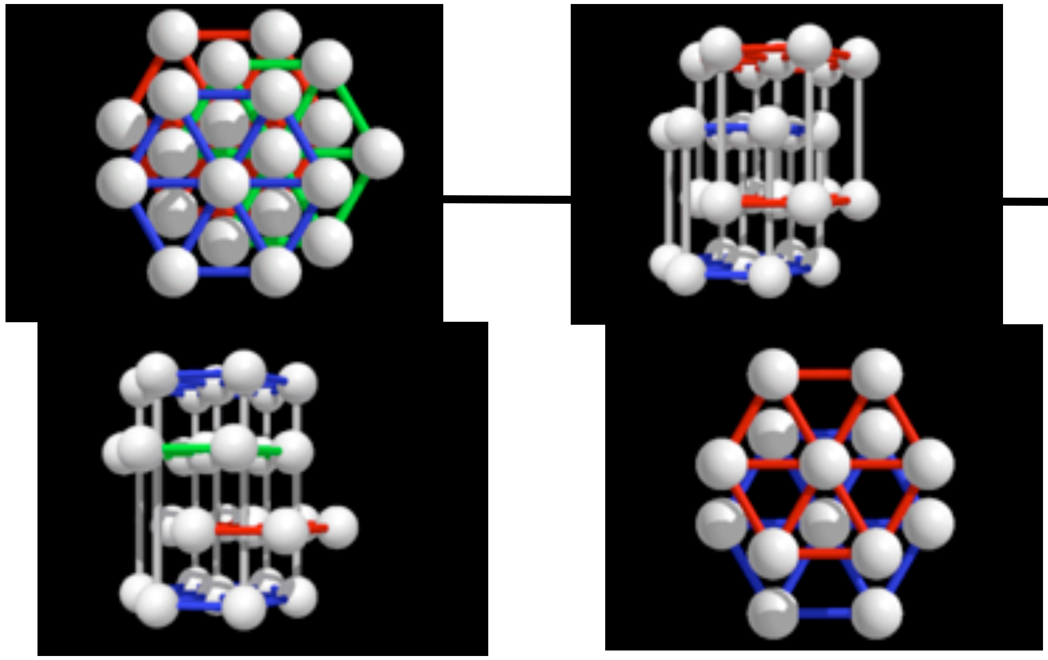
*Nonlinear Sciences Laboratory, Department of Earth and Planetary Sciences, Faculty of Science, Kobe University, Nada, Kobe 657-8501, Japan; [†]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Kawaguchi, Saitama 332-0012, Japan; [§]Department of Chemistry, University of California, Berkeley, CA 94720; and [¶]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

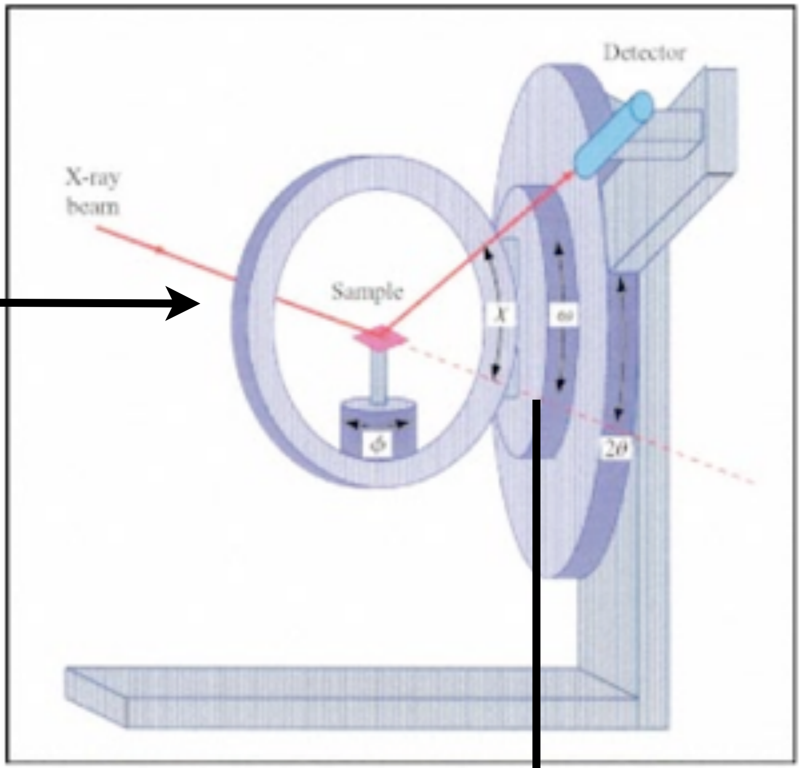


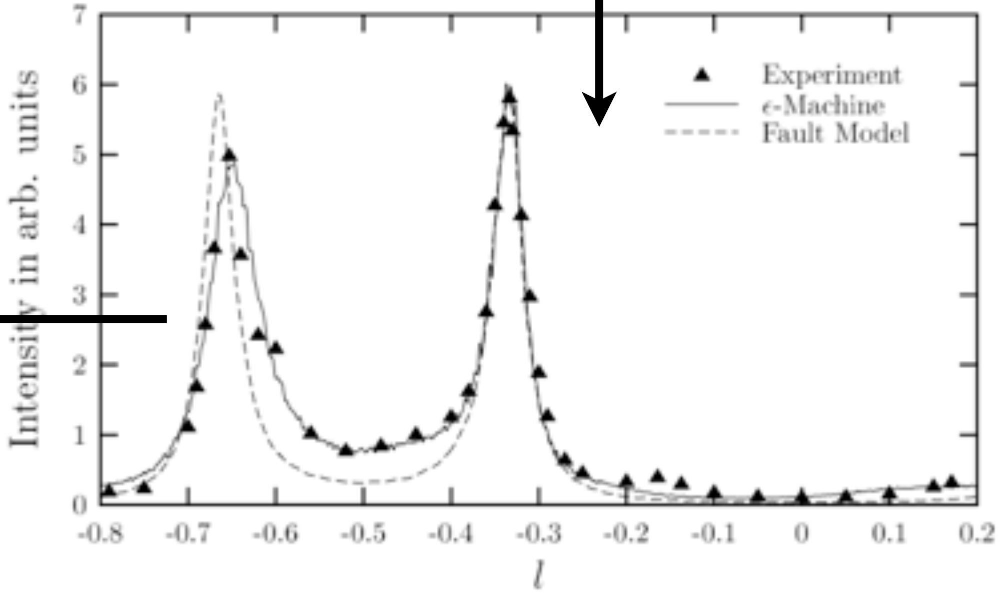


C.-B. Li, H. Yang, & T. Komatsuzaki, Proc. Natl. Acad. Sci USA **105**:2 (2008) 536–541.

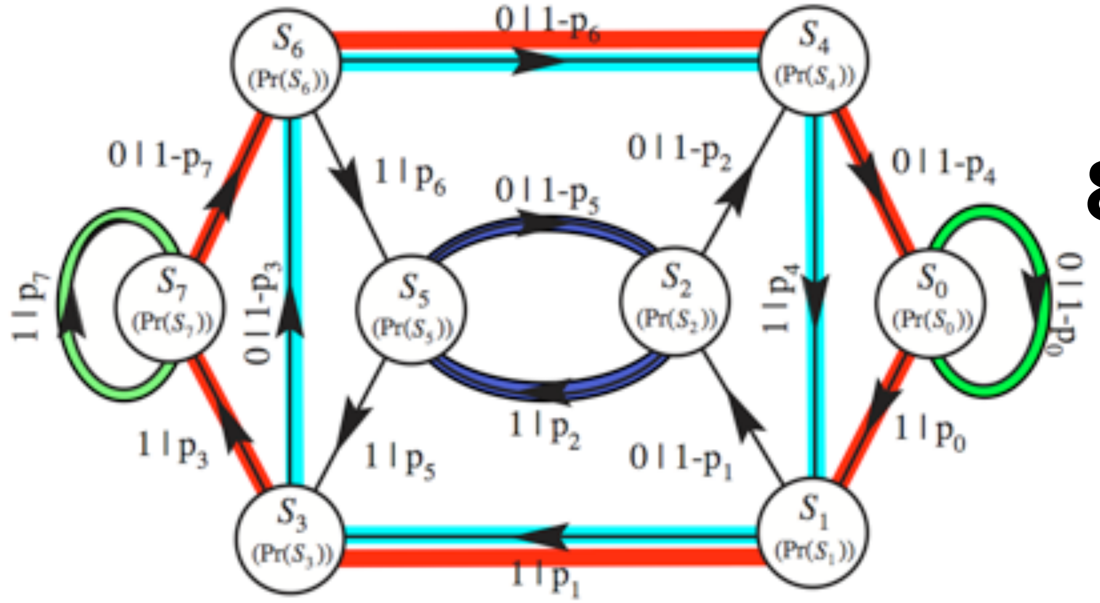
Computational Mechanics: Application to Experimental X-Ray Diffraction





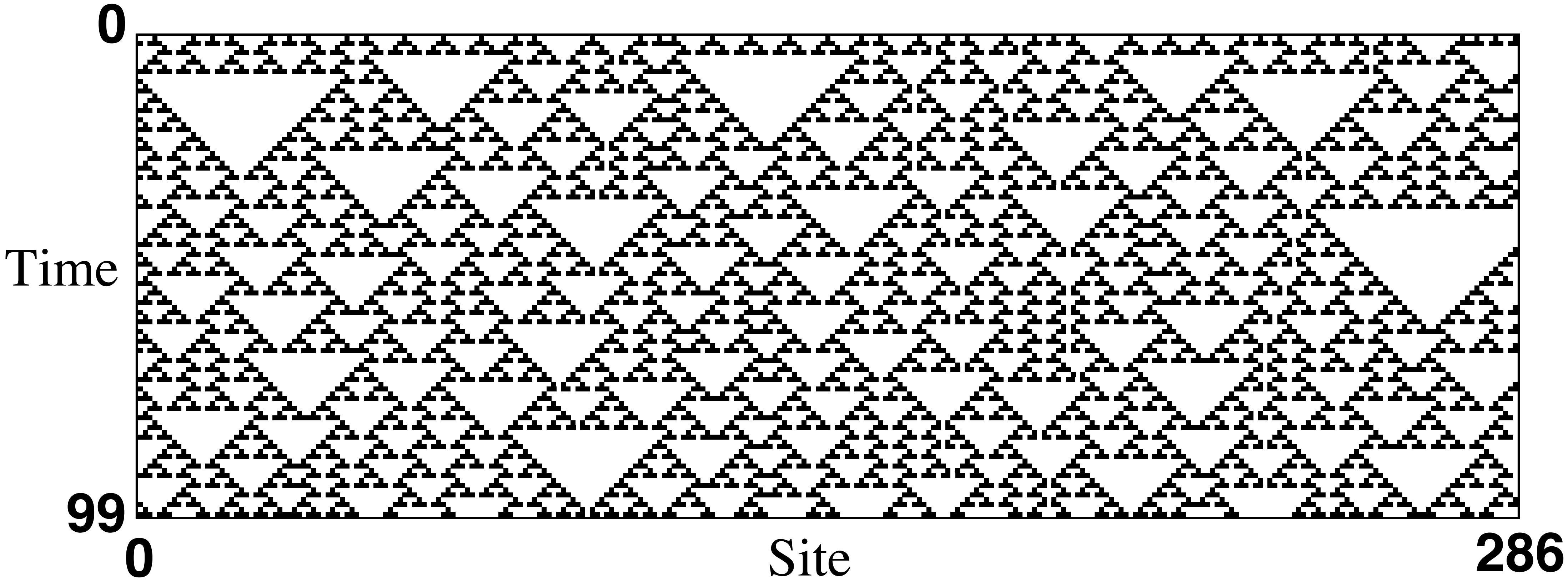


ε-MSR

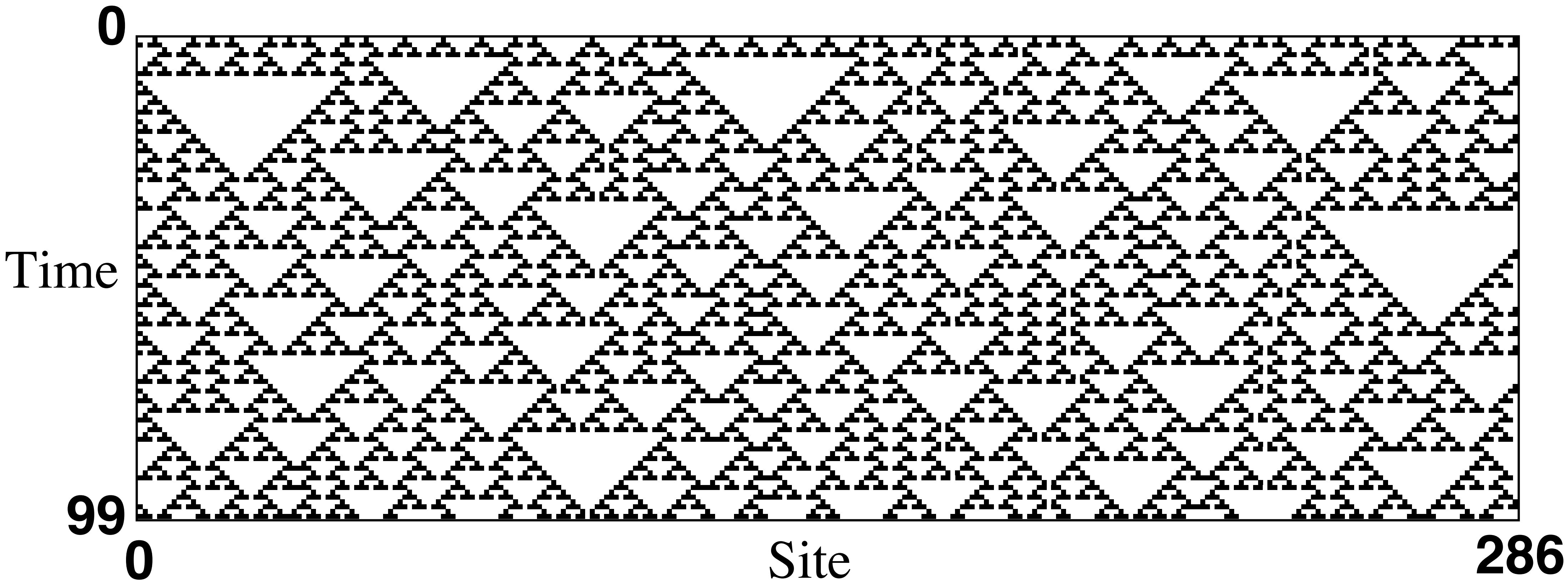


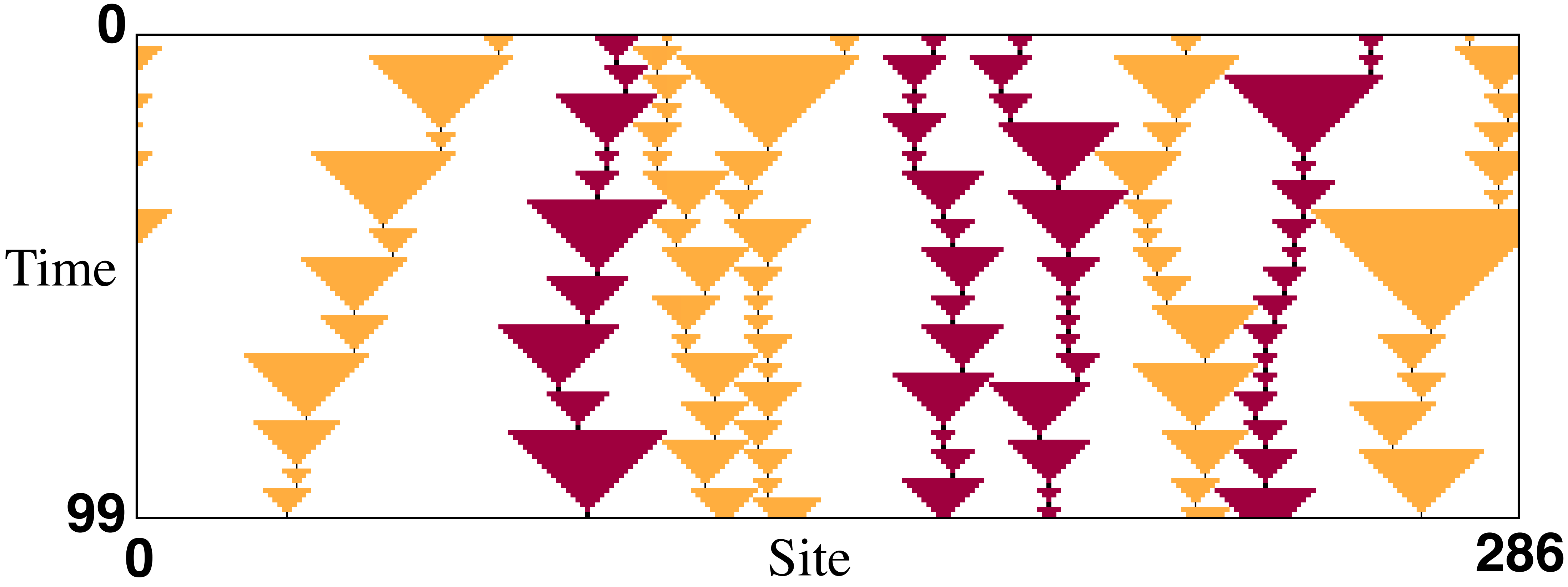
D. P. Varn, G. S. Canright, J. P. Crutchfield, "Discovering Planar Disorder in Close-Packed Structures from X-Ray Diffraction: Beyond the Fault Model", Phys. Rev. B **66**: 17 (2002) 174110-2.

Cellular Automata Computational Mechanics

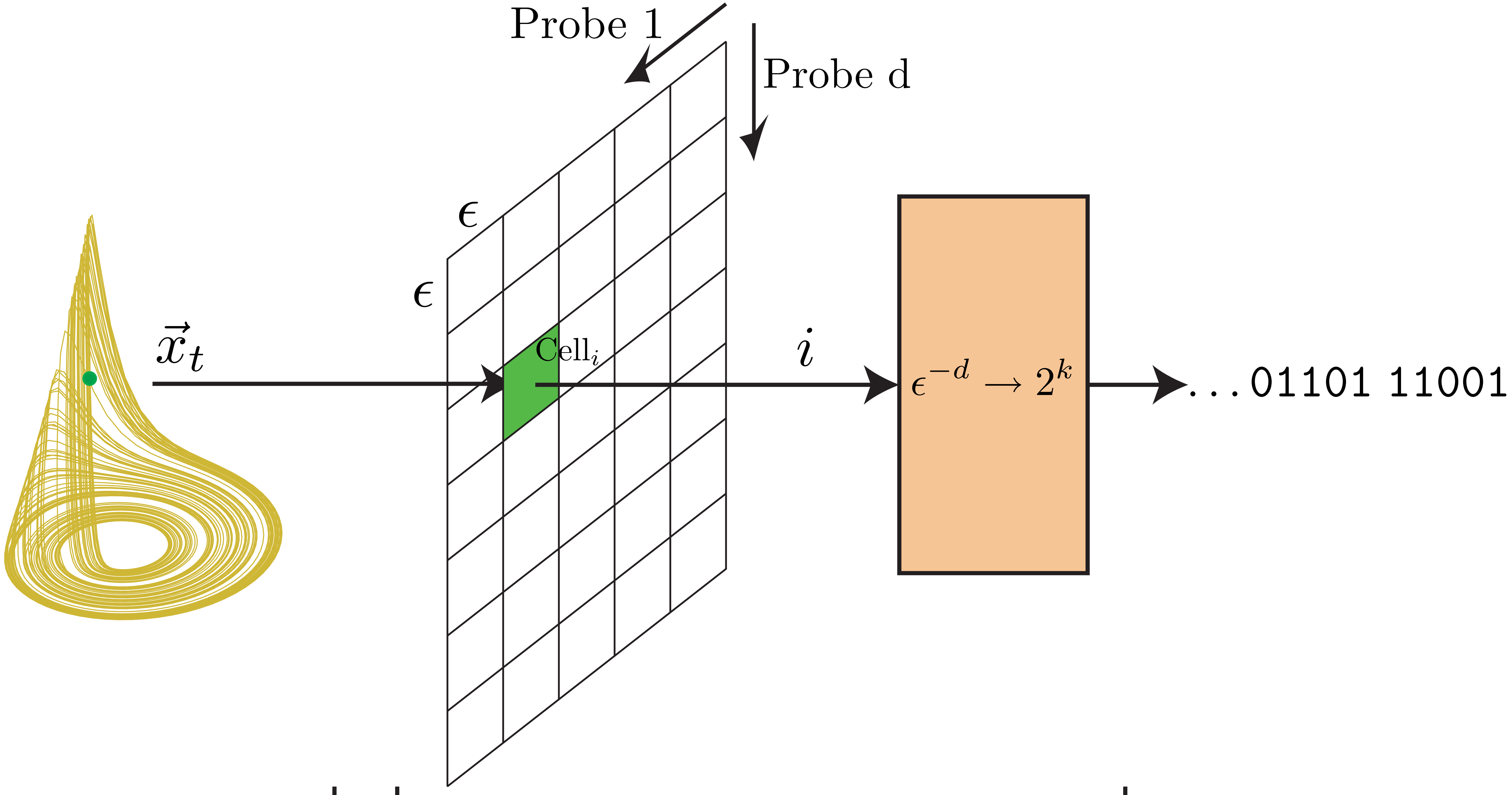


Cellular Automata Computational Mechanics





Processes and Their Models



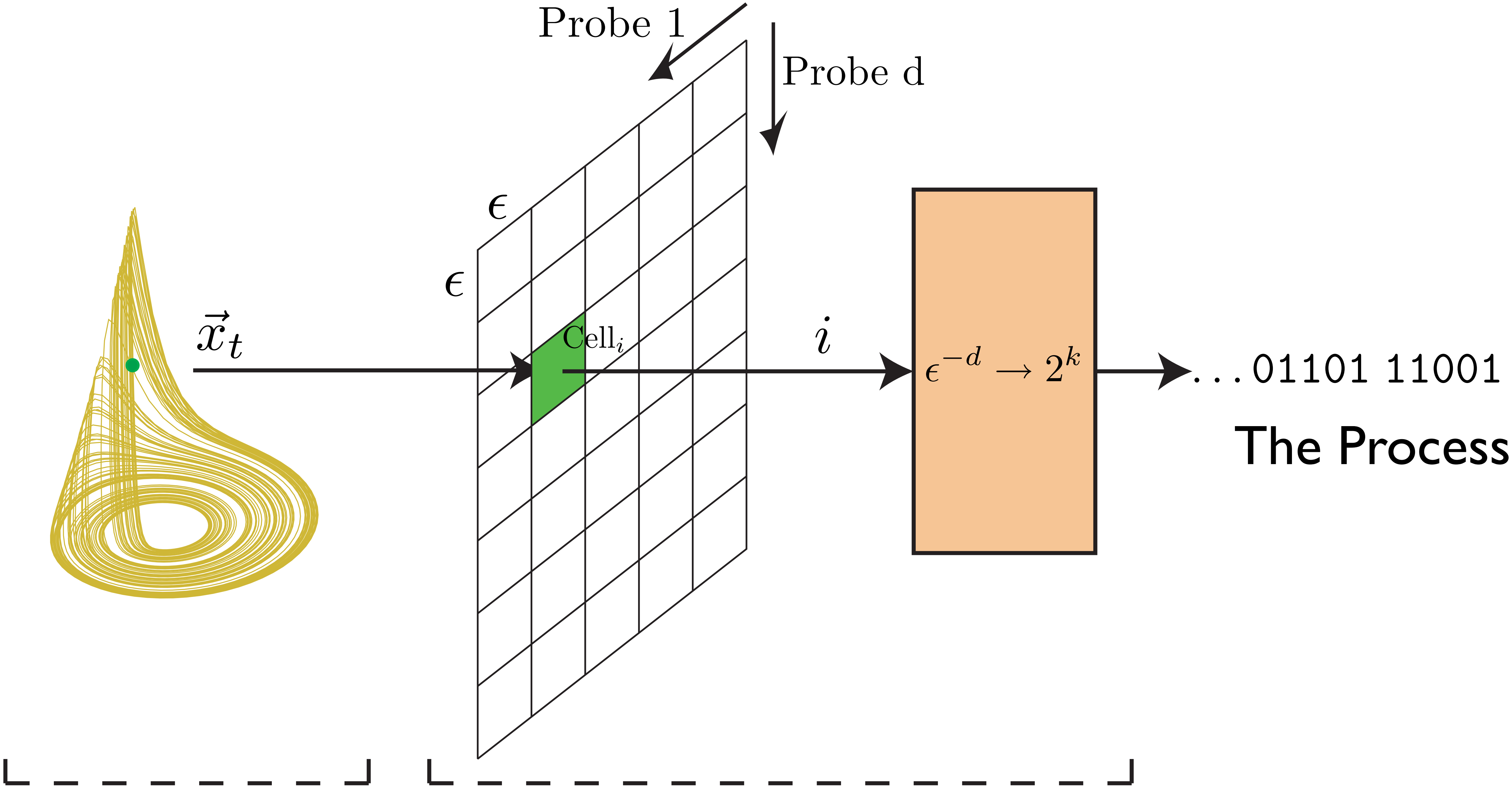
System of Interest

Instrument

Measurement Channel

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Processes and Their Models



System of Interest

Instrument

Measurement Channel

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Processes and Their Models ...

Main questions now:

- How do we characterize the resulting process?
  - Measure degrees of unpredictability & randomness.
  - What correlational structure is there?
- How do we build a model from the process itself?
- How much can we reconstruct about the hidden internal dynamics?

Processes and Their Models ...

Stochastic Processes:

Chain of random variables: $\overleftrightarrow{S} \equiv \ldots S_{-2} S_{-1} S_0 S_1 S_2 \ldots$

Random variable: S_t Alphabet: $\mathcal{A}$

Past: $\overleftarrow{S}_t = \ldots S_{t-3} S_{t-2} S_{t-1}$

Future: $\overrightarrow{S}_t = S_t S_{t+1} S_{t+2} \ldots$

L-Block: $S_t^L \equiv S_t S_{t+1} \ldots S_{t+L-1}$

Word: $s_t^L \equiv s_t s_{t+1} \ldots s_{t+L-1} \in \mathcal{A}^L$

Processes and Their Models ...

Stochastic Processes ...

Process:

$$\Pr(\overleftrightarrow{S}) = \Pr(\ldots S_{-2} S_{-1} S_0 S_1 S_2 \ldots)$$

Sequence (or word) distributions:

$$\{\Pr(S_t^L) = \Pr(S_t S_{t+1} \ldots S_{t+L-1}) : S_t \in \mathcal{A}\}$$

Process:

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

Consistency conditions:

$$\Pr(S_t^{L-1}) = \sum_{S_{t+L-1}} \Pr(S_t^L) \qquad \Pr(S_{t+1}^{L-1}) = \sum_{S_t} \Pr(S_t^L)$$

Processes and Their Models ...

Types of Stochastic Process:

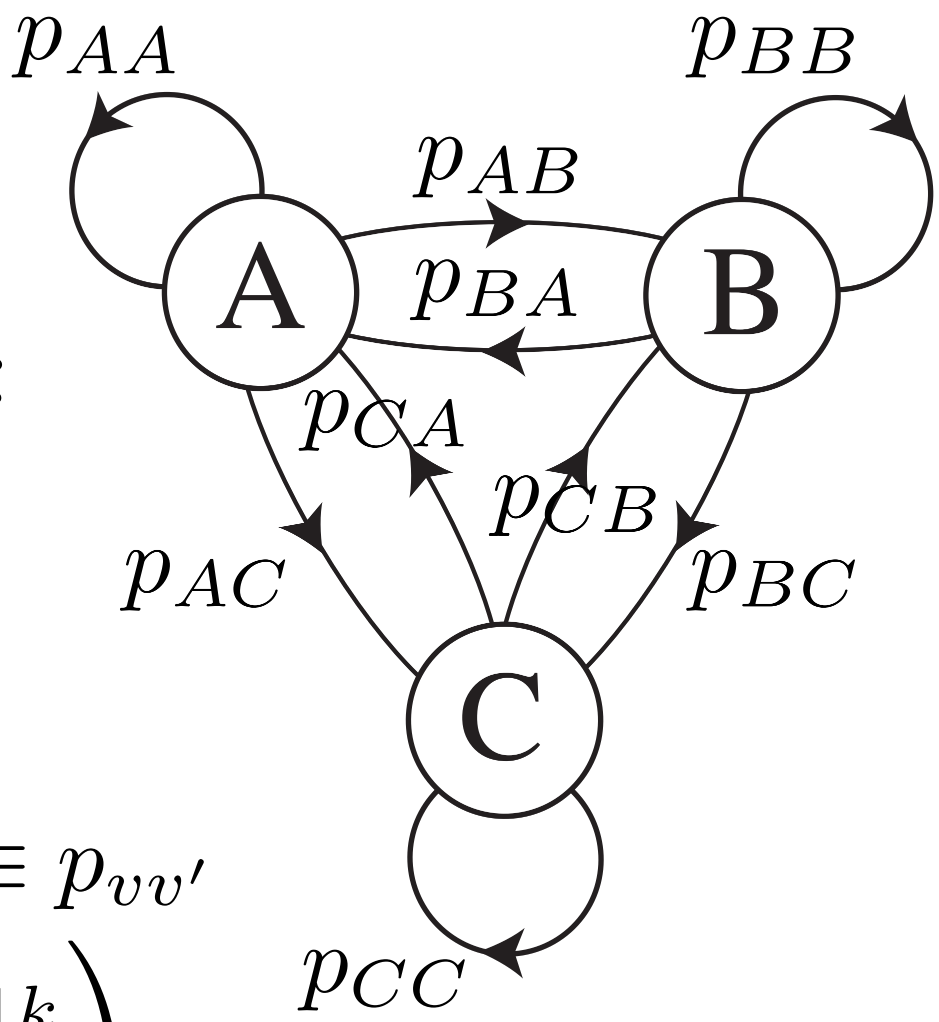
Stationary process:

$$\Pr(S_t S_{t+1} \ldots S_{t+L-1}) = \Pr(S_0 S_1 \ldots S_{L-1})$$

Assume stationarity, unless otherwise noted.

Notation: Drop time indices.

Processes and Their Models ...



Models of Stochastic Processes:

Markov chain model of a Markov process:

States: $v \in \mathcal{A} = \{1, \ldots, k\}$

$$\overleftrightarrow{V} = \ldots V_{-2} V_{-1} V_0 V_1 \ldots$$

Transition matrix: $T_{ij} = \Pr(v_{t+1} | v_t) \equiv p_{vv'}$

$$T = \begin{pmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{pmatrix}$$

Stochastic matrix: $\sum_{j=1}^{k} T_{ij} = 1$

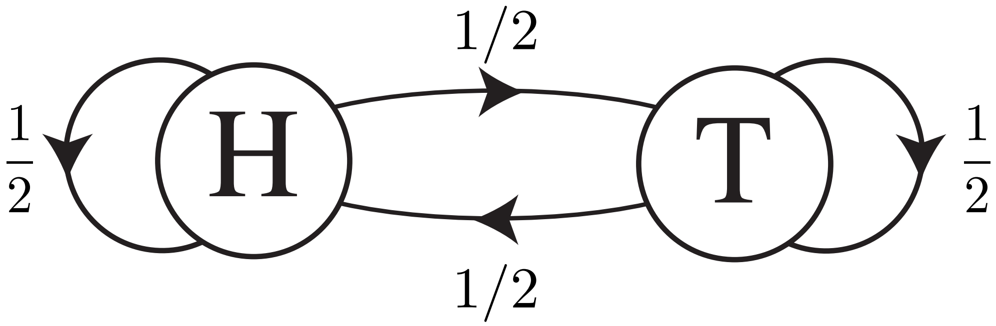
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Fair Coin: $\mathcal{A} = \{H, T\}$

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$



$$\Pr(H) = \Pr(T) = 1/2$$

Asymptotic invariant distribution: $\pi \equiv \Pr(H, T)$

$$\pi = \pi T$$

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Processes and Their Models ...

Models of Stochastic Processes ...

Example: Sequence Distribution: $\Pr(v^L) = 2^{-L}$

Fair Coin ...

Word as binary fraction:

$$s^L = s_1 s_2 \ldots s_L$$

$$\text{``}s^L\text{''} = \sum_{i=1}^{L} \frac{s_i}{2^i}$$

$$s^L \in [0,1]$$

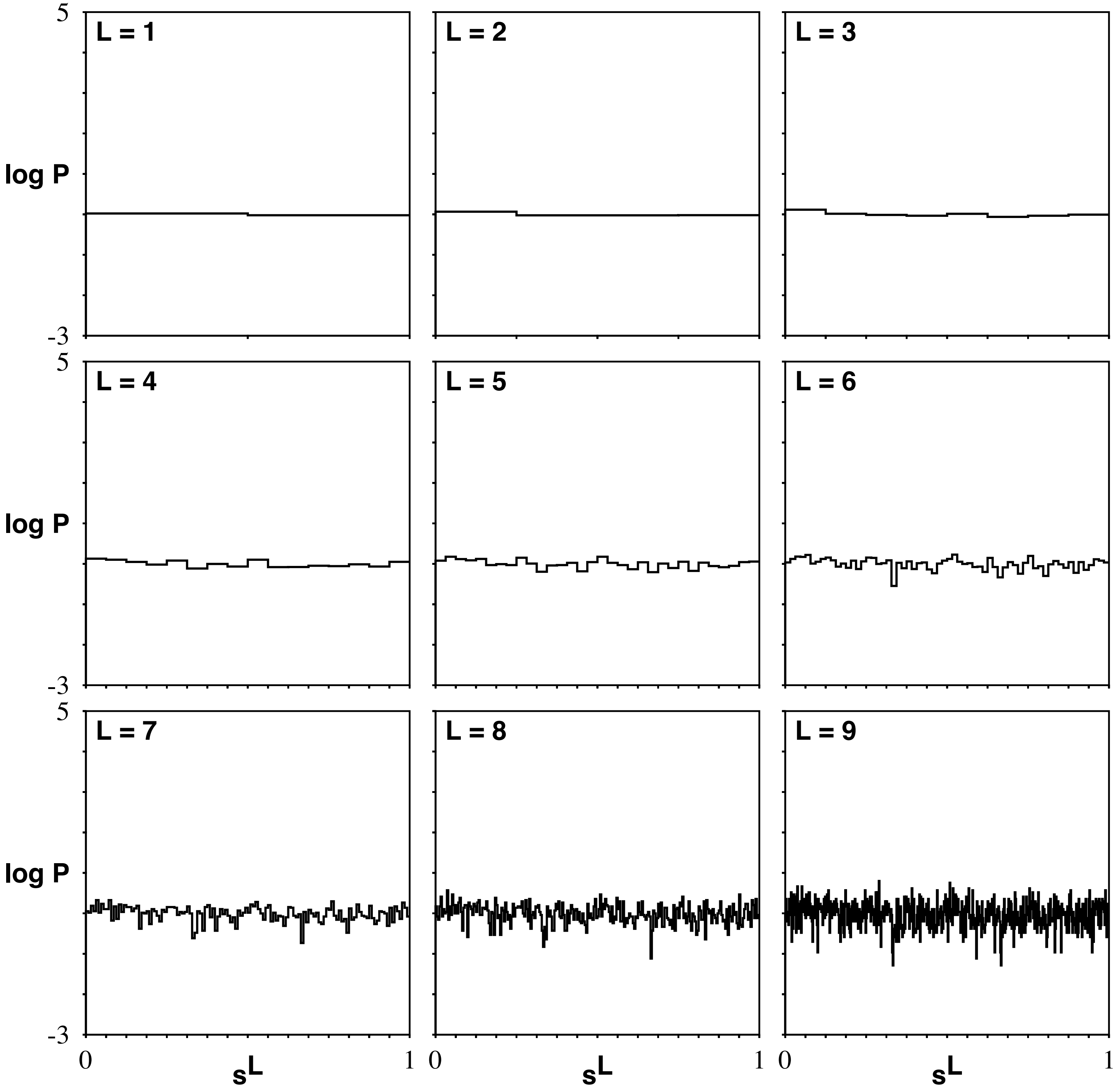
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Fair Coin ...

Sequence Distribution: $\Pr(v^L) = 2^{-L}$



Word as binary fraction:

$$s^L = s_1 s_2 \ldots s_L$$

$$\text{“}s^L\text{”} = \sum_{i=1}^{L} \frac{s_i}{2^i}$$

$$s^L \in [0, 1]$$

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

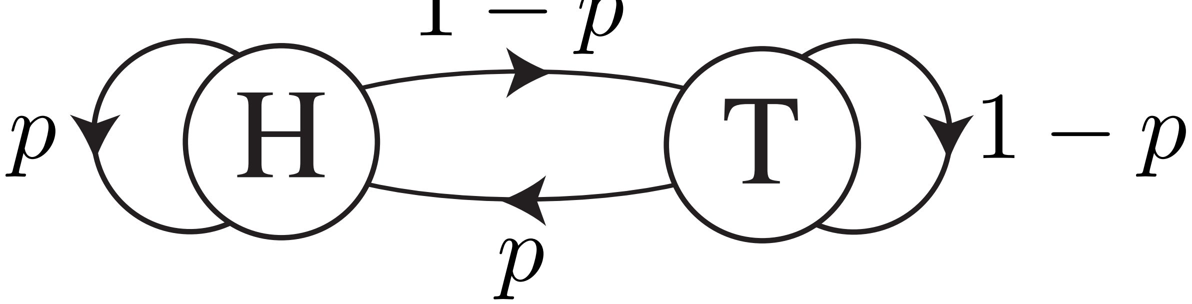
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Biased Coin: $\mathcal{A} = \{H, T\}$

$$T = \begin{pmatrix} p & 1-p \\ p & 1-p \end{pmatrix}$$



$$\Pr(H) = p$$

$$\Pr(T) = 1 - p$$

$$\pi = \Pr(p, 1-p)$$

Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Biased Coin ...

Sequence Distribution:

$$\Pr(s^L) = p^n (1-p)^{L-n},$$

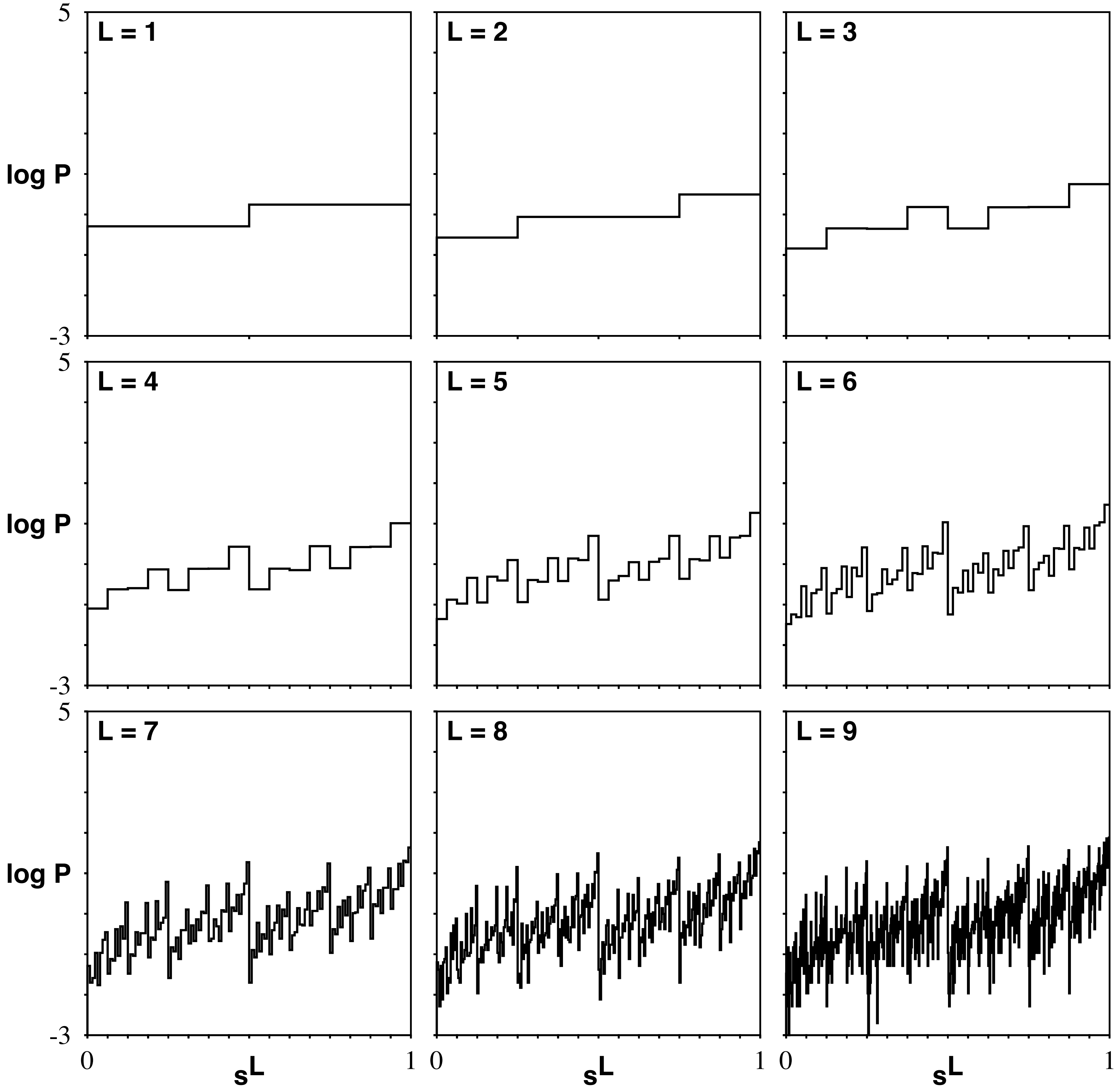
$$n = \text{Number } H\text{s in } s^L$$

Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Biased Coin ...



Sequence Distribution:

$$\Pr(s^L) = p^n (1-p)^{L-n},$$

$$n = \text{Number } H\text{s in } s^L$$

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

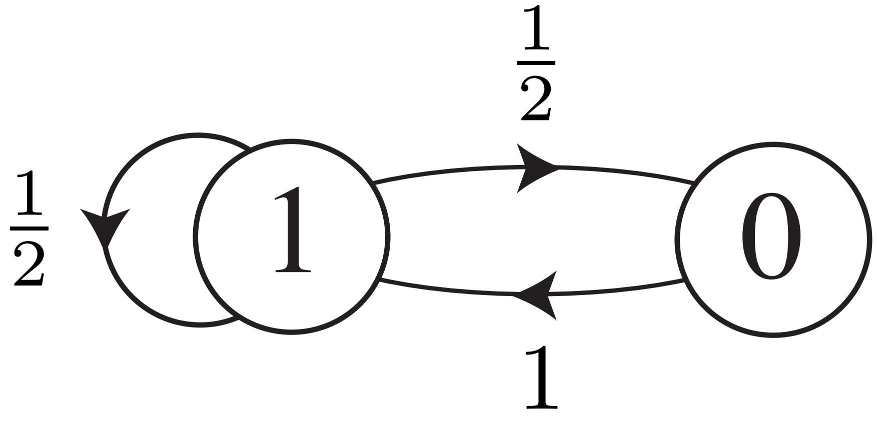
Processes and Their Models ...

Models of Stochastic Processes ...

Example: Golden Mean Process = “No consecutive 0s”

Markov chain over 1-Blocks: $\mathcal{A} = \{0, 1\}$

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



$$\pi = \Pr(V = 1, V = 0)$$
$$= \left(\tfrac{2}{3}, \tfrac{1}{3}\right)$$

As an order-1 Markov chain.

A minimal-order model of the GM Process.

Processes and Their Models ...

Models of Stochastic Processes ...

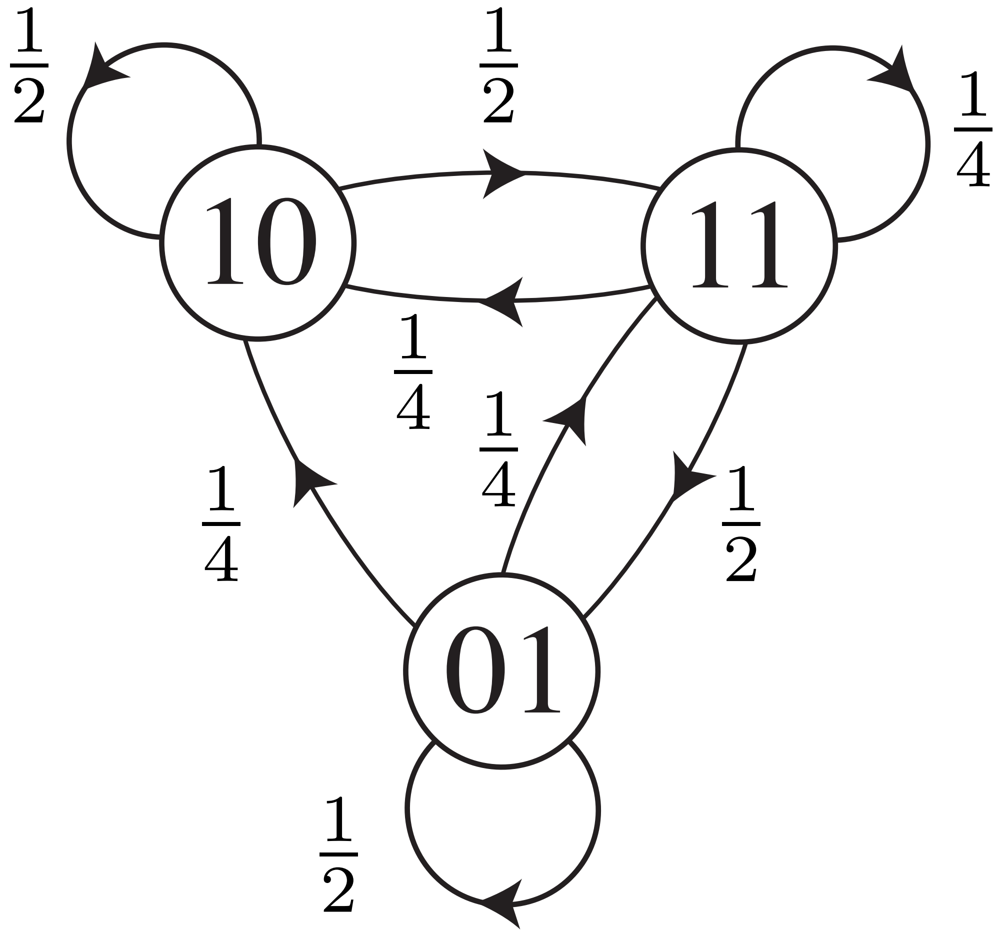
Example: Golden Mean Process ...

as a Markov chain over 2-Blocks: $\mathcal{A} = \{10, 01, 11\}$

$$T = \begin{array}{c} \\ \\ \end{array}\begin{array}{ccc} 10 & 01 & 11 \\ \end{array}$$

$$T = \left(\begin{array}{ccc} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}\right)$$

$$\pi = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right)$$



Previous model and this:

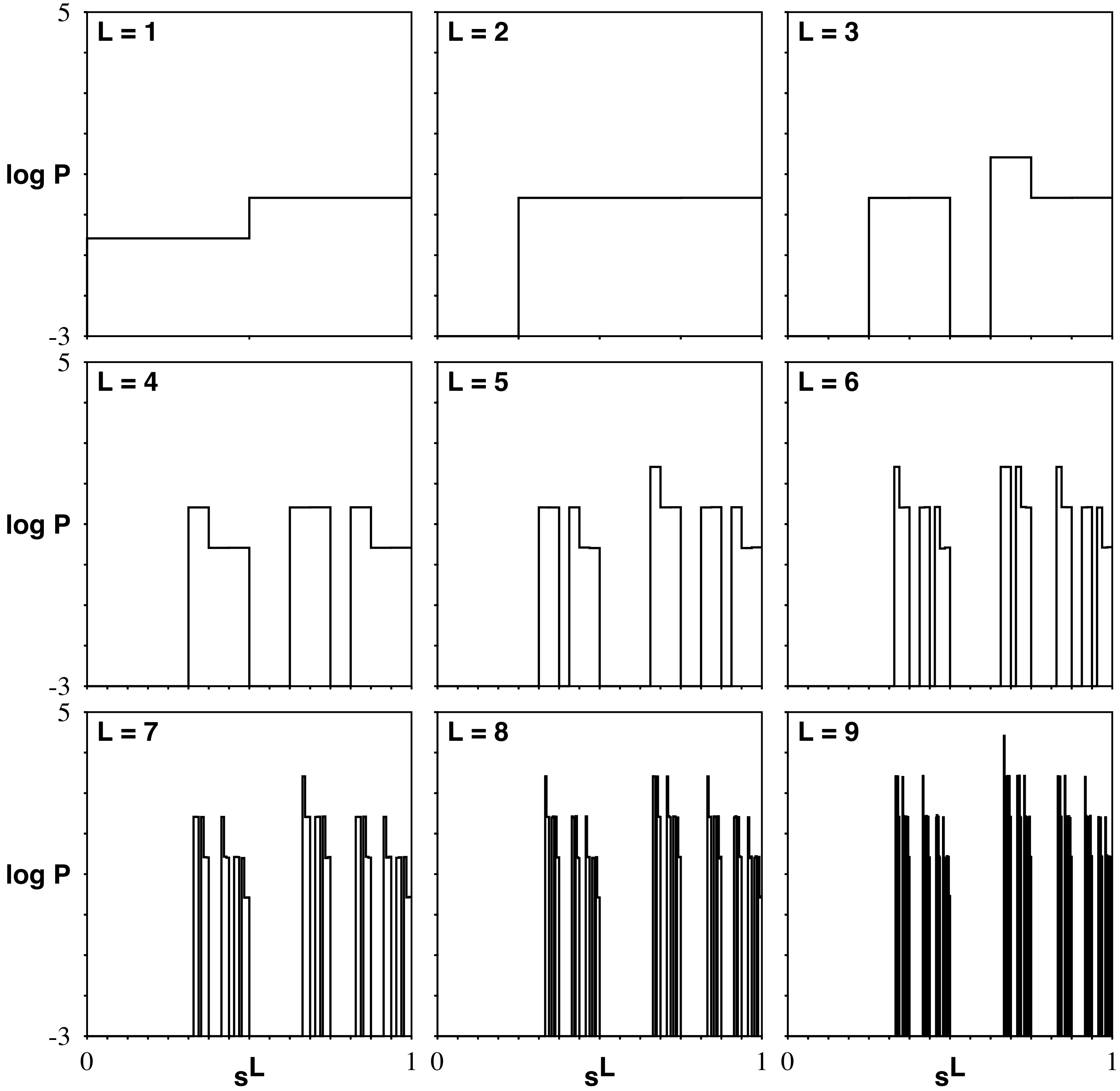
Different presentations of the same Golden Mean Process

Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Golden Mean:



Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Processes and Their Models ...

Models of Stochastic Processes ...

Two Lessons:

Structure in the behavior: $\text{supp } \Pr(s^L)$

Structure in the distribution of behaviors: $\Pr(s^L)$

Processes and Their Models ...

Models of Stochastic Processes ...

Hidden Markov Models of Processes:

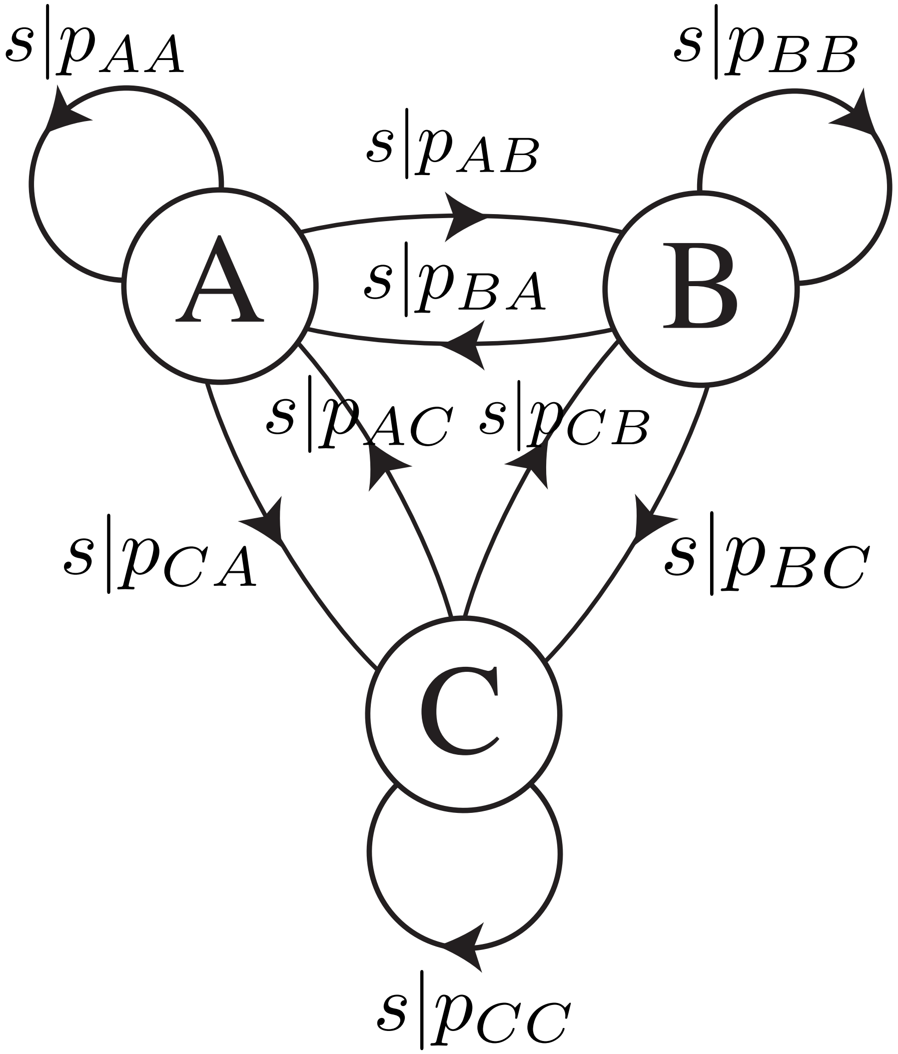
Internal: $\mathcal{A} = \{A, B, C\}$

$$T = \begin{pmatrix} p_{AA} & p_{AB} & p_{AC} \\ p_{BA} & p_{BB} & p_{BC} \\ p_{CA} & p_{CB} & p_{CC} \end{pmatrix}$$

Observed: $\mathcal{B} = \{0, 1\}$

$$T^{(s)} = \begin{pmatrix} p_{AA;s} & p_{AB;s} & p_{AC;s} \\ p_{BA;s} & p_{BB;s} & p_{BC;s} \\ p_{CA;s} & p_{CB;s} & p_{CC;s} \end{pmatrix}$$

$$p_{AA} = \sum_{s \in \mathcal{B}} p_{AA;s}$$



symbol | transition probability

Processes and Their Models ...

Models of Stochastic Processes ...

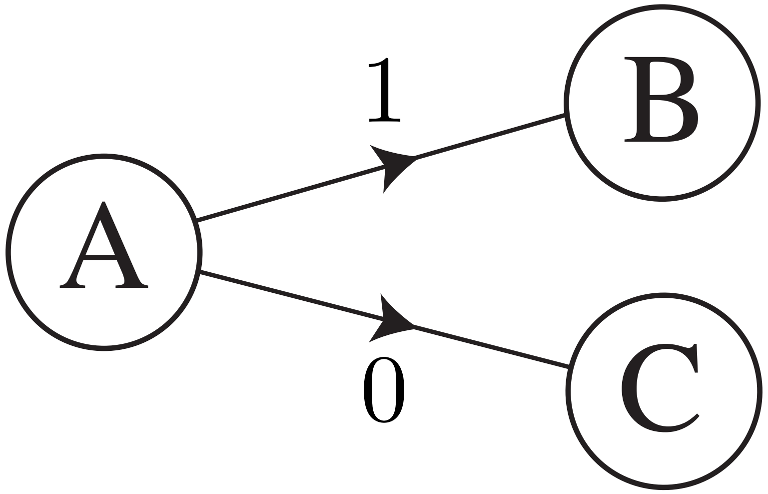
Types of Hidden Markov Model:

"Unifilar": current state + symbol "determine" next state

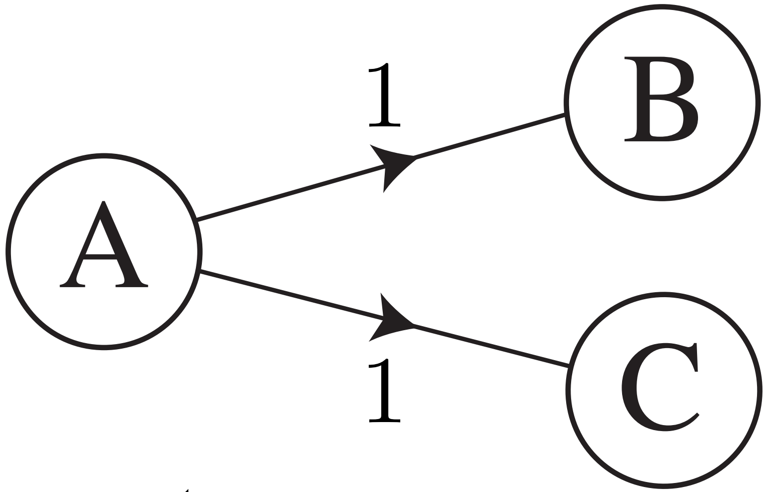
$$\Pr(v'|v,s) = \begin{cases} 1 \\ 0 \end{cases}$$

$$\Pr(v',s|v) = p(s|v)$$

$$\Pr(v'|v) = \sum_{s \in \mathcal{A}} p(s|v)$$



"Nonunifilar": no restriction



Multiple internal edge paths can generate same observed sequence.

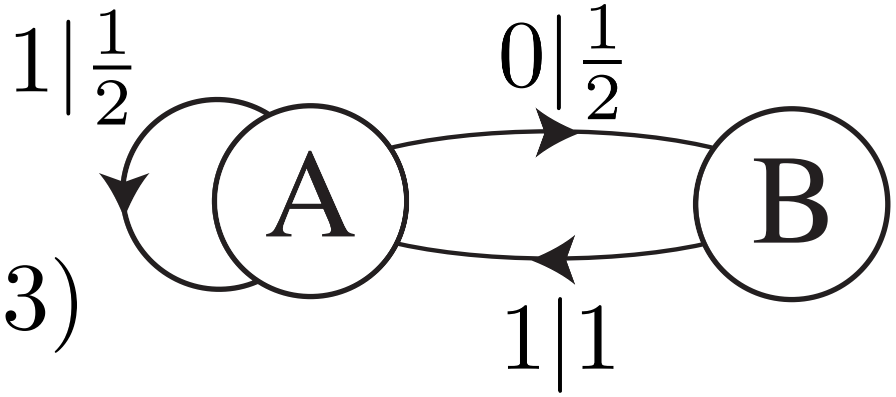
Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Golden Mean Process as a unifilar HMM:

Internal: $\mathcal{A} = \{A, B\}$



$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \quad \pi_V = (2/3, 1/3)$$

Observed: $\mathcal{B} = \{0, 1\}$

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

Initial ambiguity only: At most 2-to-1 mapping

$$BA^n = 1^n$$

$$AA^n = 1^n$$

Sync'd: $s = 0 \Rightarrow v = B$
$s = 1 \Rightarrow v = A$

Irreducible forbidden words: $\mathcal{F} = \{00\}$

Processes and Their Models ...

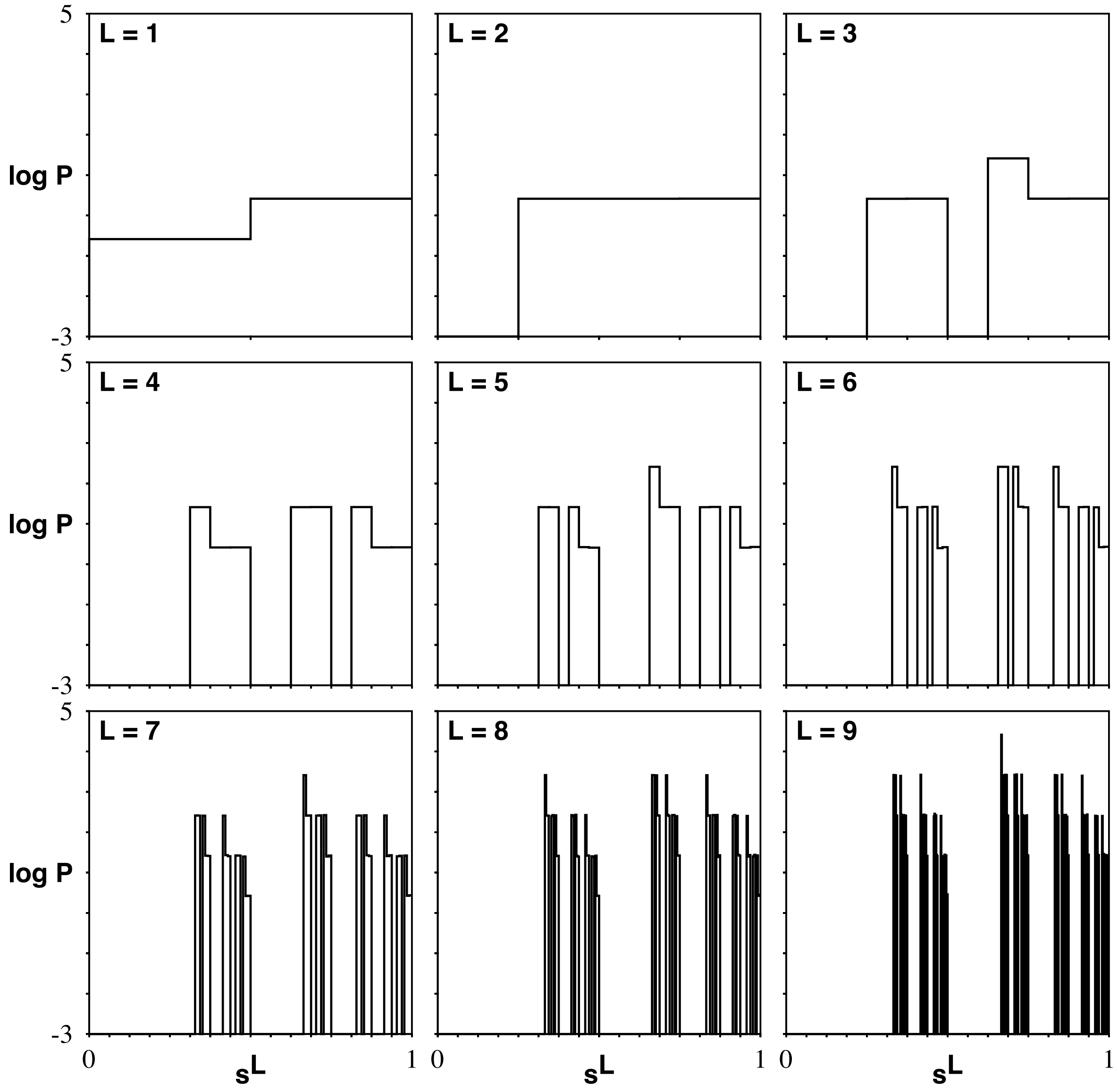
Models of Stochastic Processes ...

Example:

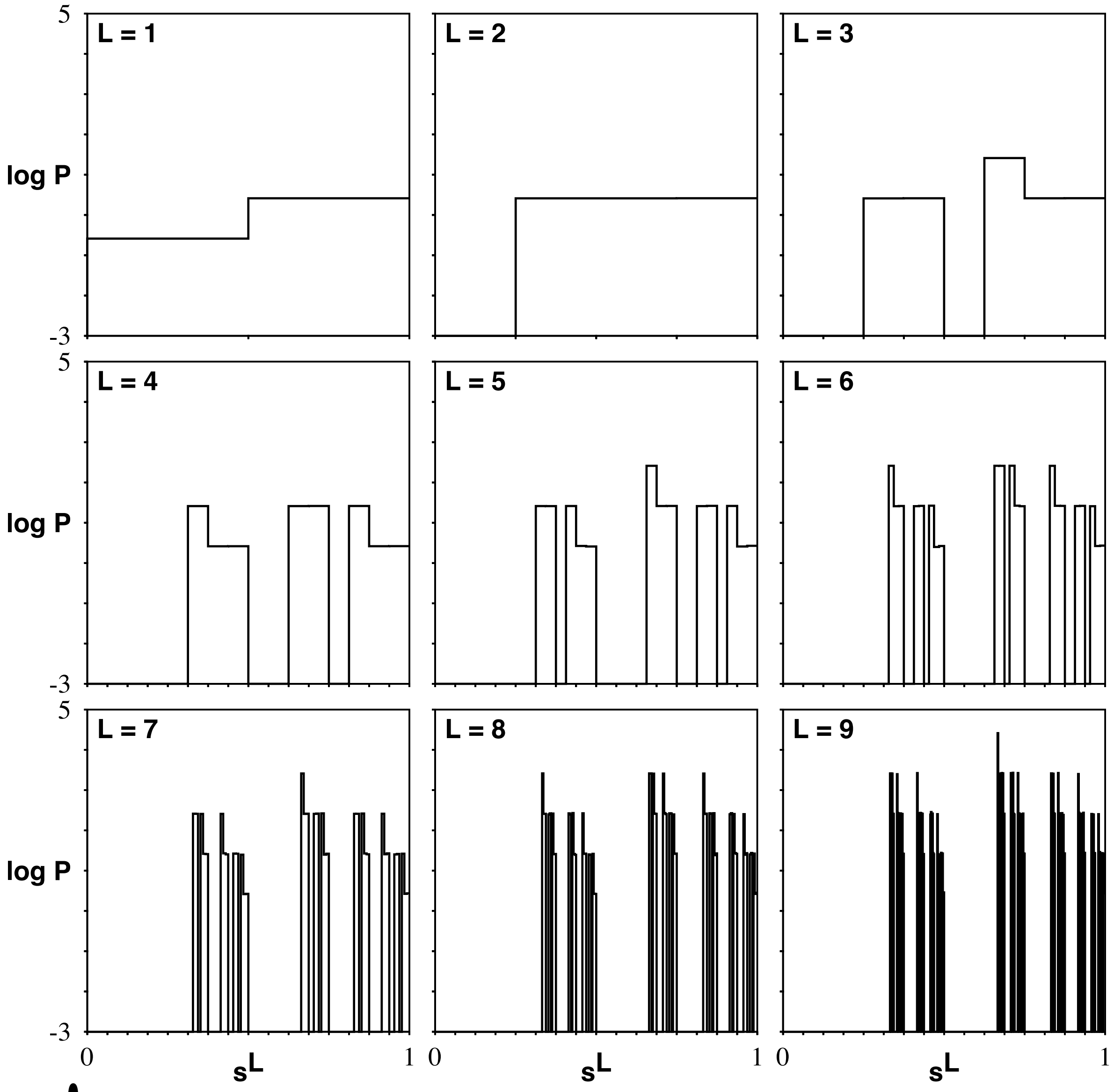
Golden Mean Process ... Sequence distributions:

Internal state sequences

(A = 1; B = 0)



Observed sequences

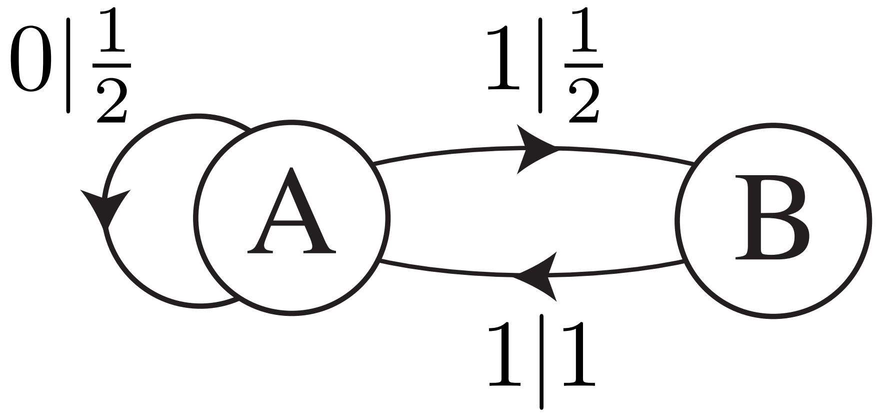


Same!

Processes and Their Models ...

Models of Stochastic Processes ...

Example: Even Process = Even #1s



As a unifilar HMM:

Internal (= GMP): $\mathcal{A} = \{A, B\}$

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \quad \pi_V = (2/3, 1/3)$$

Observed: $\mathcal{B} = \{0, 1\}$

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

$$v^L = \ldots AABAABABAA \ldots$$

$$s^L = \ldots.\ 0\ 1\ 1\ 0\ 1\ 1\ 1\ 1\ 0 \ldots \qquad s^L = \{\ldots 01^{2n}0 \ldots\}$$

Irreducible forbidden words: $\mathcal{F} = \{010, 01110, 0111110, \ldots\}$

No finite-order Markov process can model the Even process!

Lesson: Finite Markov Chains are a subset of HMMs.

Processes and Their Models ...

Models of Stochastic Processes ...

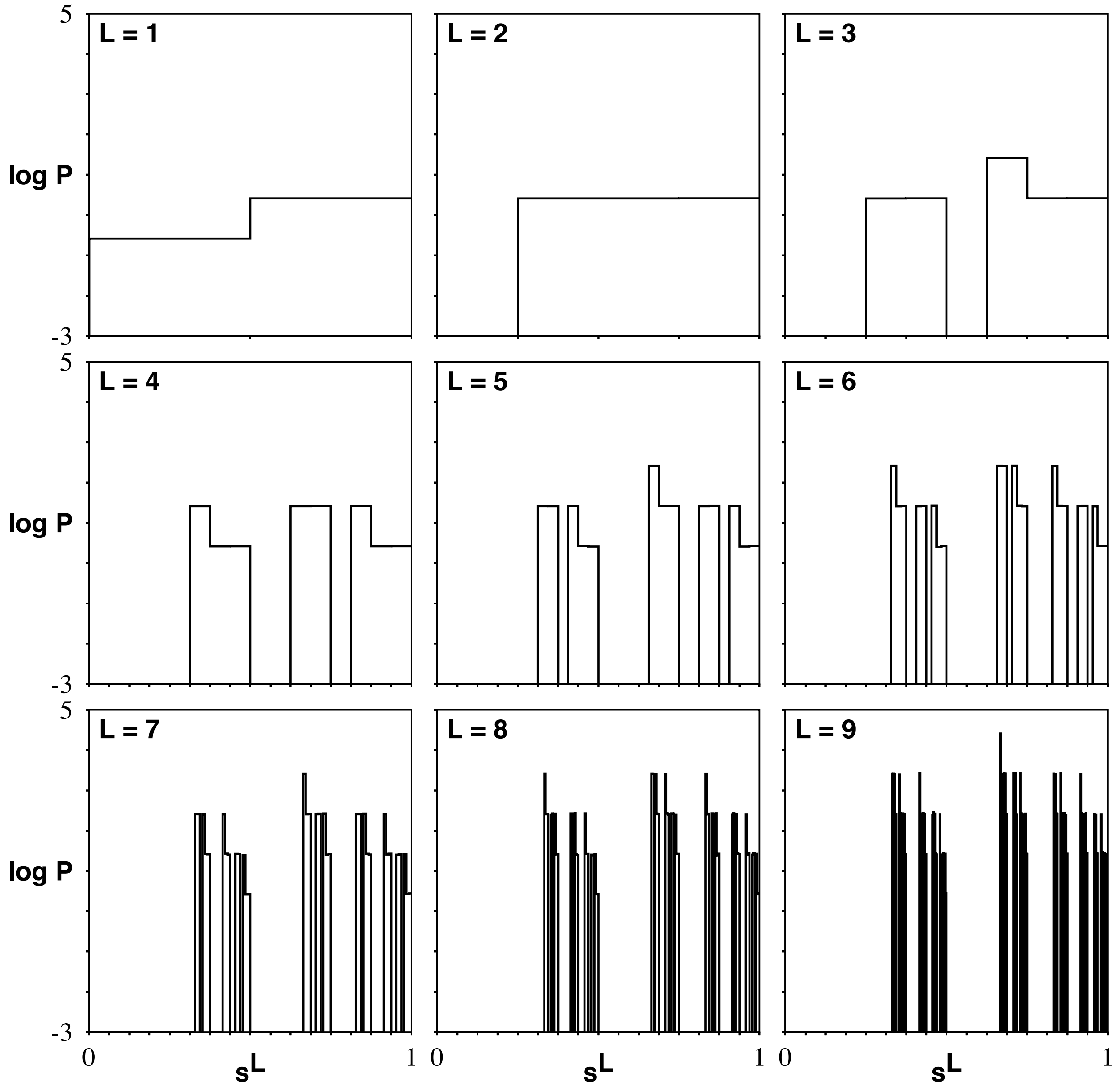
Example:

Even Process ... Sequence distributions:

Internal states (= GMP)

(A = 1; B = 0)

Observed sequences



Rather different!

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Processes and Their Models ...

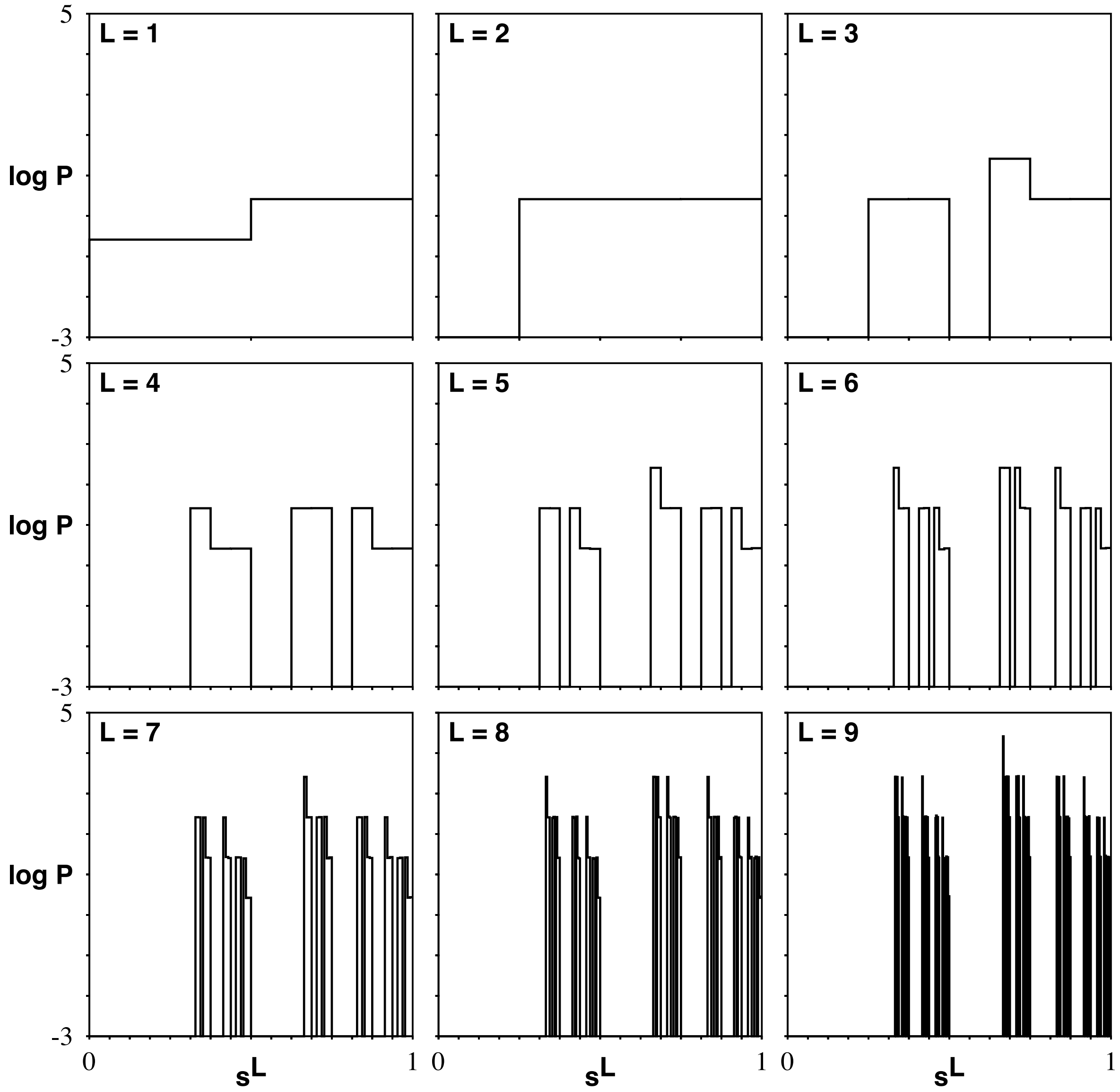
Models of Stochastic Processes ...

Example:

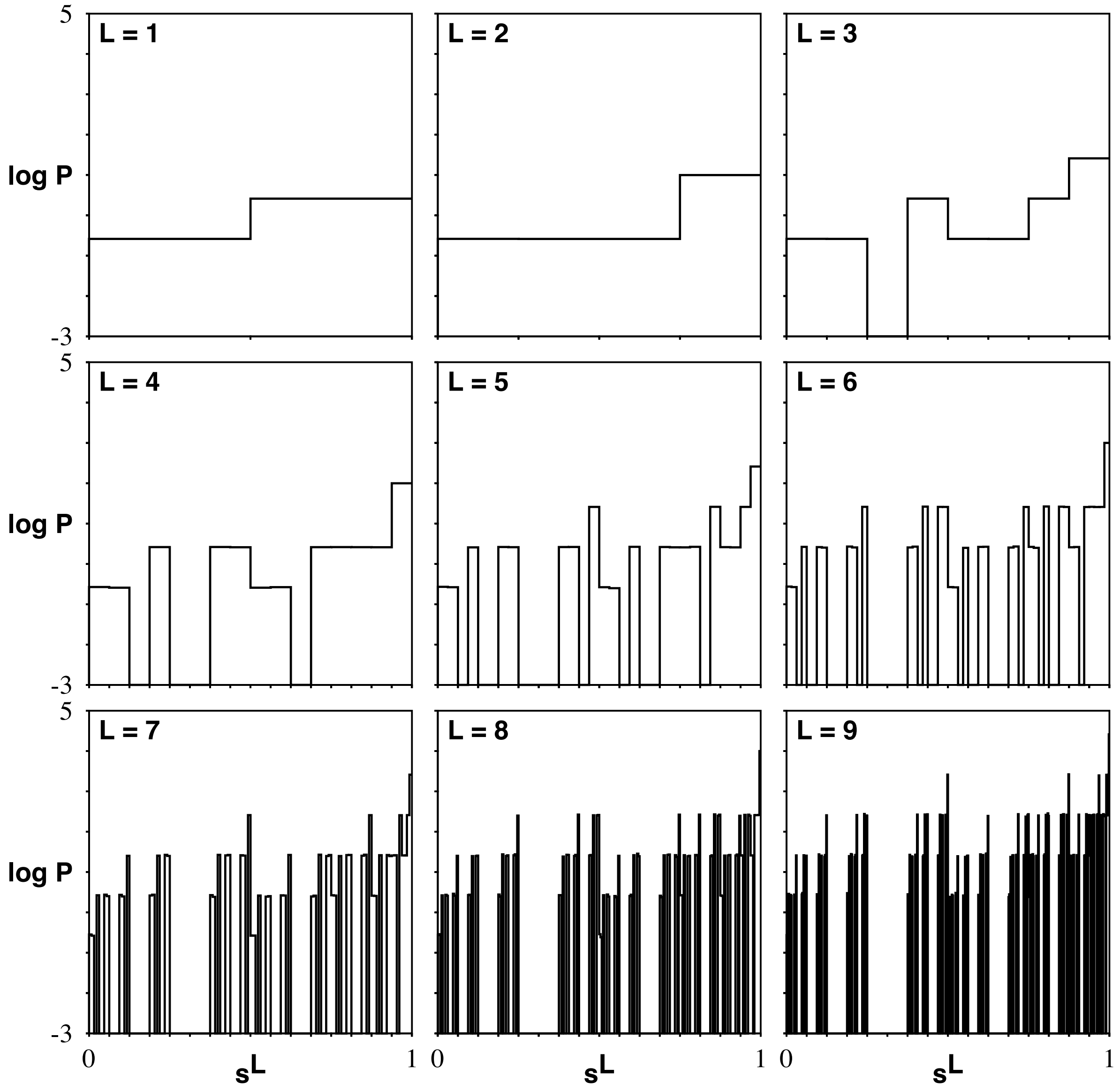
Even Process ... Sequence distributions:

Internal states (= GMP)

(A = 1; B = 0)



Observed sequences



Rather different!

Processes and Their Models ...

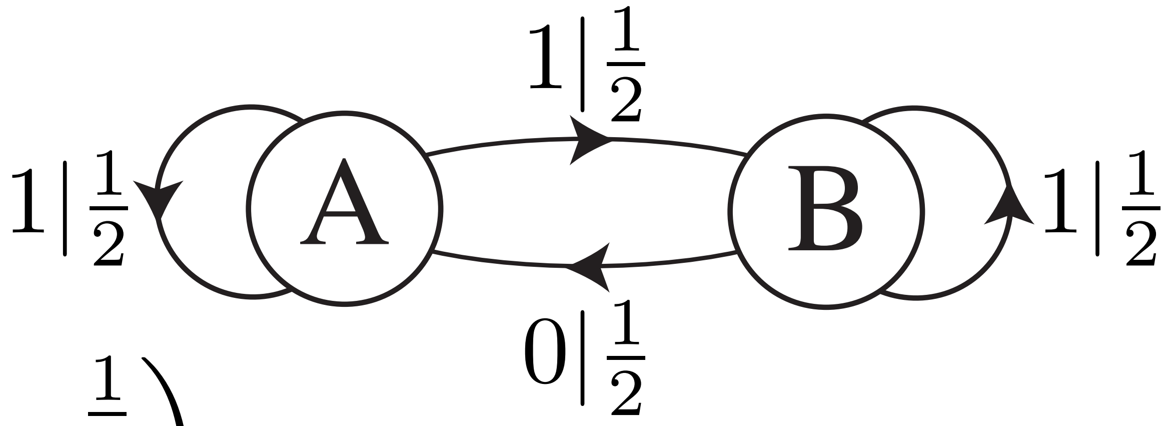
Models of Stochastic Processes ...

Example:

Simple Nonunifilar Source:

Internal (= Fair Coin): $\mathcal{A} = \{A, B\}$

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \pi_V = \left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$



Observed: $\mathcal{B} = \{0, 1\}$

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

Many to one: $1111111 \Leftarrow \begin{cases} AAAAAAAA\ldots \\ ABBBBBBB\ldots \\ AABBBBBB\ldots \\ AAABBBBB\ldots \\ \ldots \\ BBBBBBBB\ldots \end{cases}$

Is there a unifilar HMM presentation of the observed process?

Processes and Their Models ...

Models of Stochastic Processes ...

Example:

Simple Nonunifilar Process ...

Internal states (= Fair coin)

(A = 1; B = 0)

Observed sequences

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Processes and Their Models ...

Models of Stochastic Processes ...

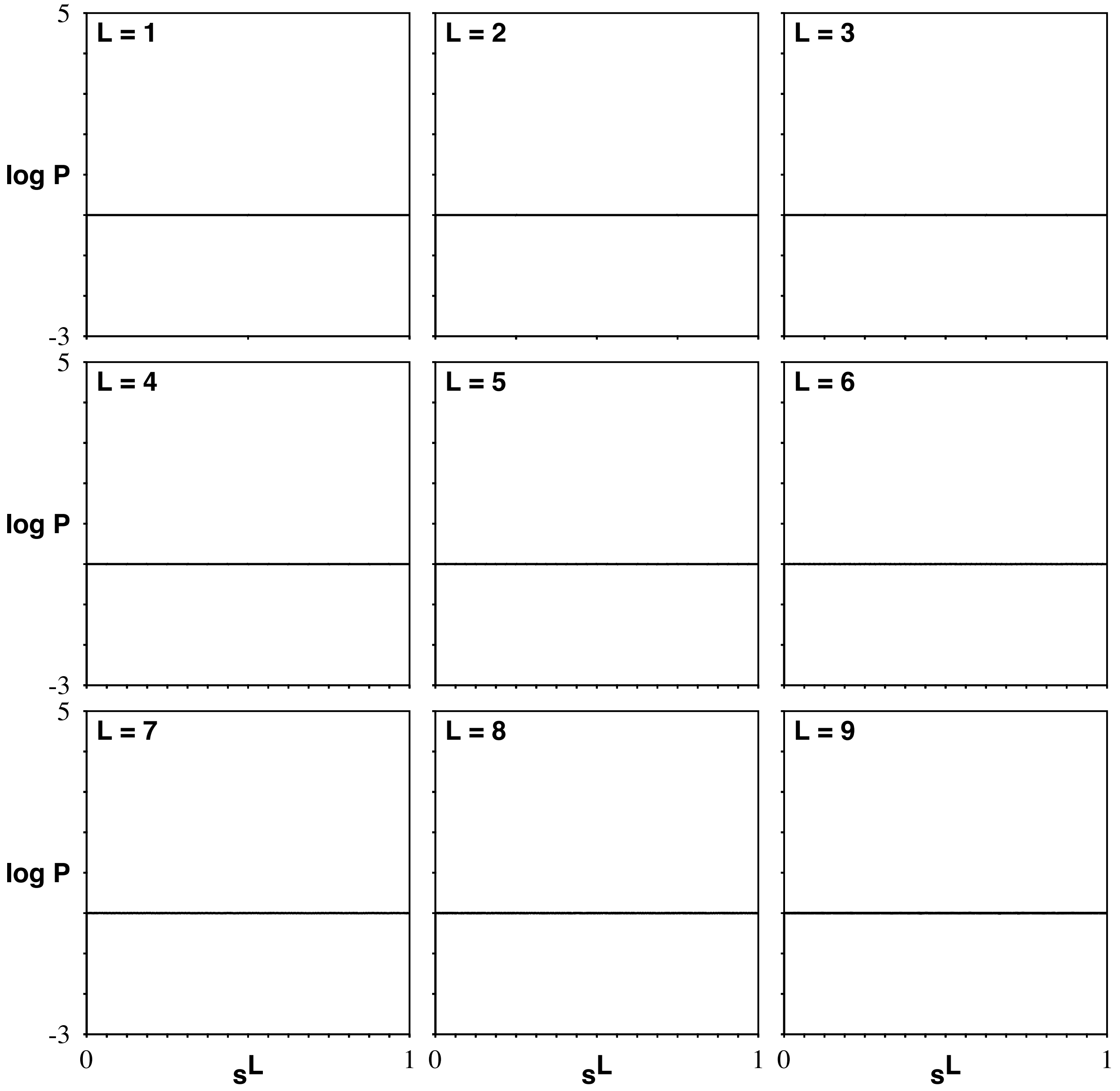
Example:

Simple Nonunifilar Process ...

Internal states (= Fair coin)

(A = 1; B = 0)

Observed sequences



Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Processes and Their Models ...

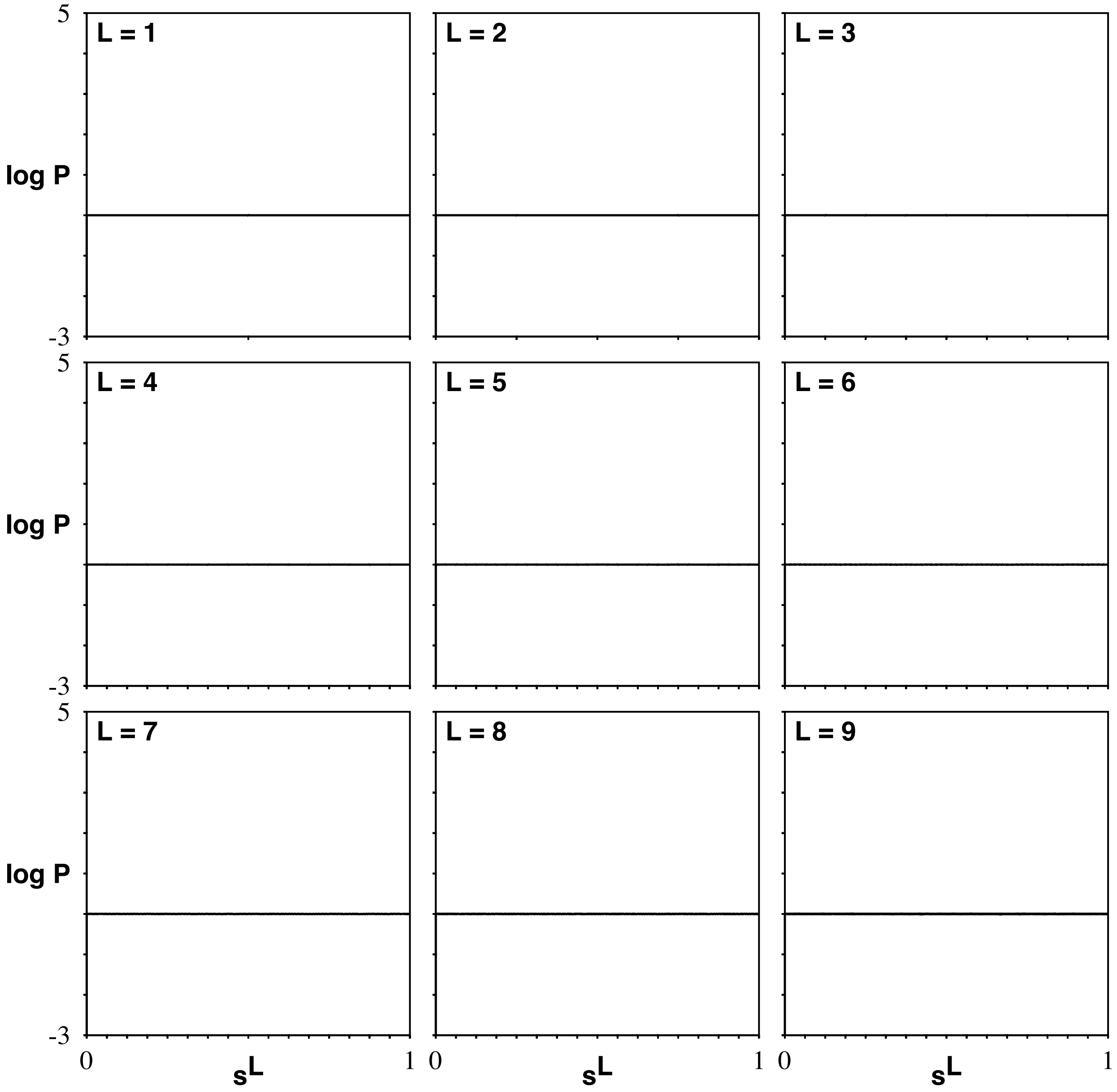
Models of Stochastic Processes ...

Example:

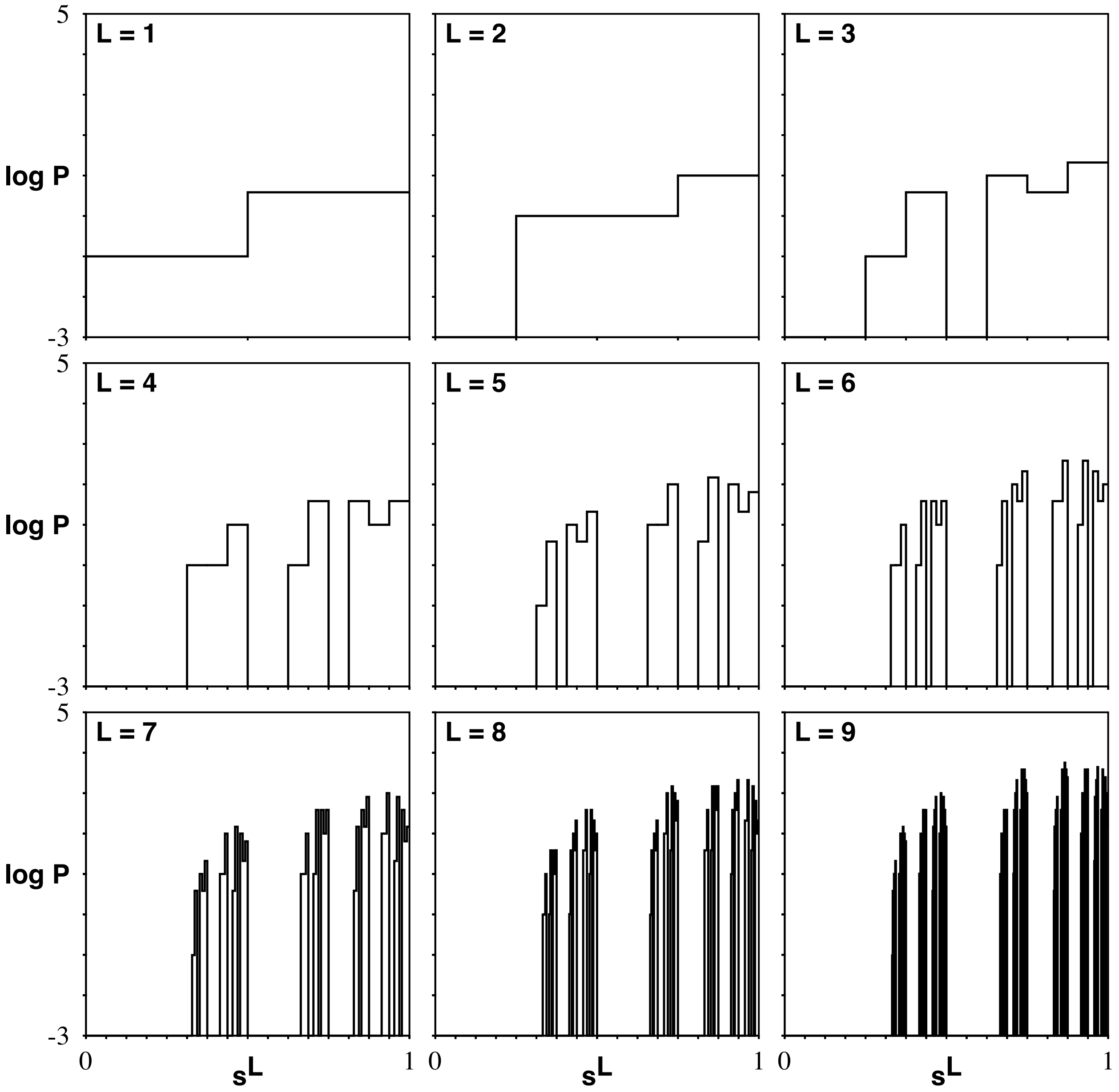
Simple Nonunifilar Process ...

Internal states (= Fair coin)

(A = 1; B = 0)



Observed sequences



Processes and Their Models ...

What to do with all of this complicatedness?

1. Information theory for complex processes
2. Measures of complexity
3. Optimal models and how to build them

Labs:

Track these topics.

Nix will give a tour in evening session.

Work through labs on your own.

Information!

Sources of Information:

Apparent randomness:
- Uncontrolled initial conditions
- Actively generated: Deterministic chaos

Hidden regularity:
- Ignorance of forces
- Limited capacity to model structure

Information ...

Information as uncertainty and surprise:

Observe something unexpected:
Gain information

Bateson: "A difference that makes a difference"

Information ...

Information as uncertainty and surprise ...

How to formalize?

- Shannon's approach:
  - A measure of surprise.
  - Connection with Boltzmann's thermodynamic entropy

Self-information of an event $\propto -\log \Pr(\text{event})$.

Predictable: No surprise $-\log 1 = 0$

Completely unpredictable: Maximally surprised

$$-\log \frac{1}{\text{Number of Events}} = \log(\text{Number of Events})$$

Information ...

Shannon Entropy: $X \sim P$

$$x \in \mathcal{X} = \{1, 2, \ldots, k\}$$

$$P = \{\Pr(x = 1), \Pr(x = 2), \ldots\}$$

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$$

Note: $0 \log 0 = 0$

Units:

Log base 2: $H(X) = [\mathrm{bits}]$

Natural log: $H(X) = [\mathrm{nats}]$

Properties:

1. Positivity: $H(X) \geq 0$
2. Predictive: $H(X) = 0 \;\Leftrightarrow\; p(x) = 1$ for one and only one x
3. Random: $H(X) = \log_2 k \;\Leftrightarrow\; p(x) = U(x) = 1/k$

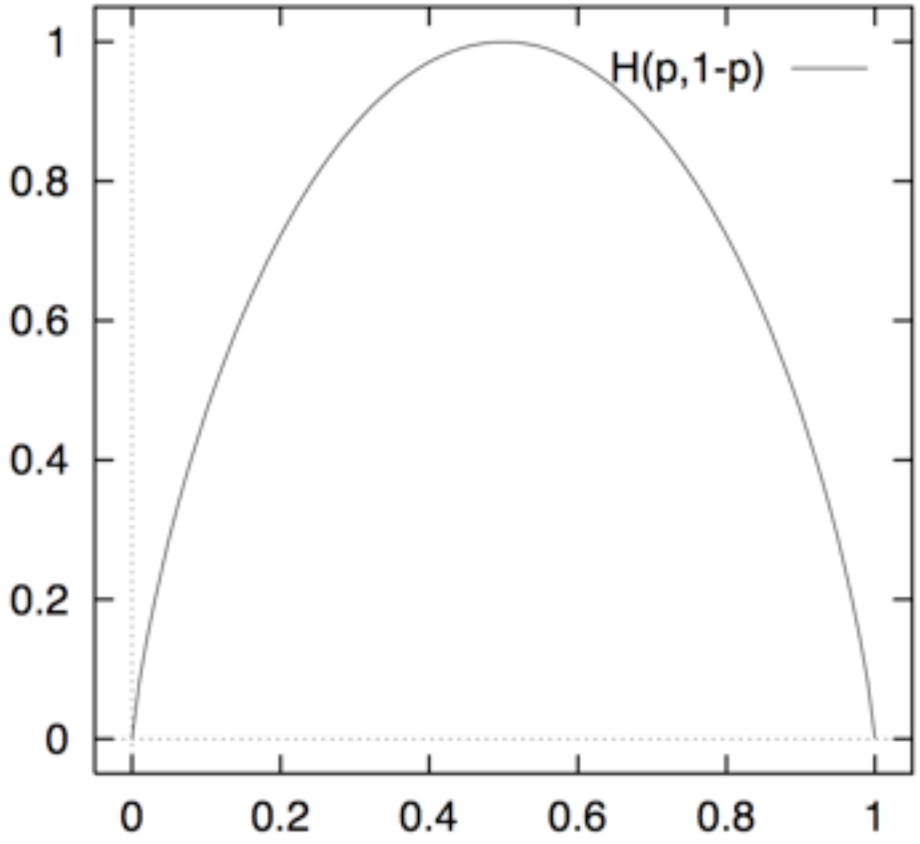
Information ...

Example: Binary random variable X (Biased Coin)

$$\mathcal{X} = \{0, 1\} \qquad \Pr(1) = p \;\&\; \Pr(0) = 1 - p$$

$H(X)$?

Binary entropy function:



$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

Fair coin: $p = \frac{1}{2}$

$$H(p) = 1 \text{ bit}$$

Completely biased coin: $p = 0 \text{ (or } 1)$

$$H(p) = 0 \text{ bits}$$

Recall: $0 \cdot \log 0 = 0$

Information ...

Example: Independent, Identically Distributed (IID) Process over four events

$$\mathcal{X} = \{a, b, c, d\} \qquad \Pr(X) = (\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8})$$

Entropy: $H(X) = \frac{7}{4}$ bits

Number of questions to identify the event?

x = a? (must always ask at least one question)
x = b? (this is necessary only half the time)
x = c? (only get this far a quarter of the time)

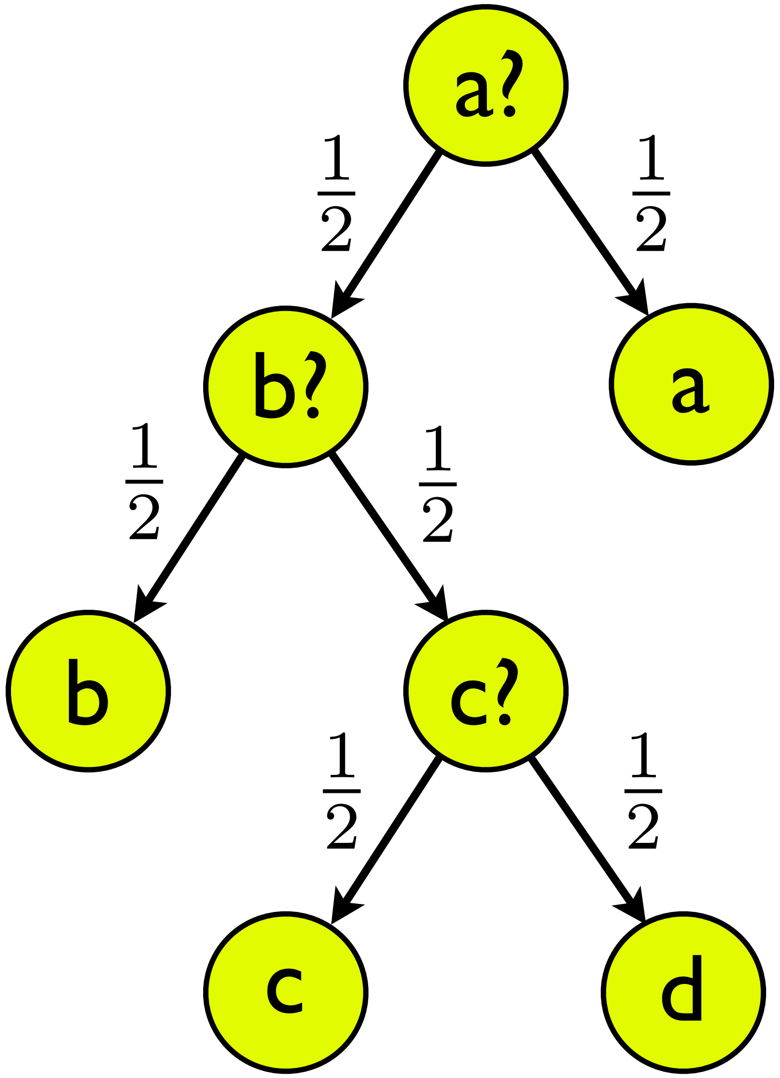
Average number: $1 \cdot 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 1.75$ questions

Interpretation? Optimal way to ask questions.

Information ...

Example: IID Process over four events ...

Average number: $1 \cdot 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} = 1.75$ questions



$$\Pr(X) = (\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8})$$

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Information ...

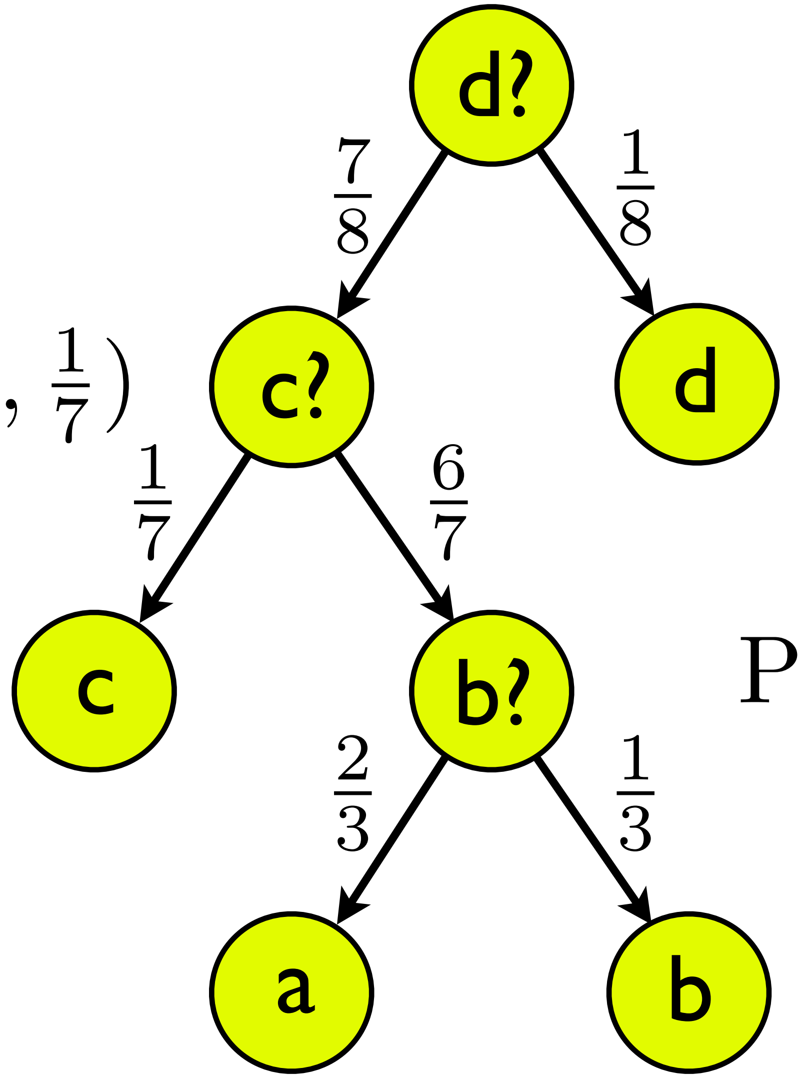
Example: IID Process over four events ...

Query in a different order:

Average number: $1 \cdot 1 + 1 \cdot \frac{7}{8} + 1 \cdot \frac{6}{7} \approx 2.7$ questions

$$\Pr(X) = (\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8})$$

$$\Pr(a, b, c) = (\tfrac{4}{7}, \tfrac{2}{7}, \tfrac{1}{7})$$



$$\Pr(a, b) = (\tfrac{2}{3}, \tfrac{1}{3})$$

Information ...

Example: IID Process over four events

Entropy: $H(X) = \frac{7}{4}$ bits

At each stage, ask questions that are most informative.

Choose partitions of event space that give "most random" measurements.

Theorem:
Entropy gives the smallest number of questions
to identify an event, on average.

Information ...

Interpretations of Shannon Entropy:

Observer's *degree of surprise* in outcome of a random variable

Uncertainty *in* random variable

Information required to *describe* random variable

A measure of *flatness* of a distribution

Information ...

Two random variables: $(X, Y) \sim p(x, y)$

Joint Entropy: Average uncertainty in X and Y occurring

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 p(x, y)$$

Independent:

$$X \perp Y \Rightarrow H(X, Y) = H(X) + H(Y)$$

Information ...

Conditional Entropy: Average uncertainty in X, knowing Y

$$H(X|Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log_2 p(x|y)$$

$$H(X|Y) = H(X,Y) - H(Y)$$

Not symmetric: $H(X|Y) \neq H(Y|X)$

Information ...

Common Information Between Two Random Variables:

$$X \sim p(x) \ \& \ Y \sim p(y)$$

$$(X, Y) \sim p(x, y)$$

Mutual Information:

$$I(X;Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$

Information ...

Mutual Information ...

Properties:

(1) $I(X;Y) \geq 0$

(2) $I(X;Y) = I(Y;X)$

(3) $I(X;Y) = H(X) - H(X|Y)$

(4) $I(X;Y) = H(X) + H(Y) - H(X,Y)$

(5) $I(X;X) = H(X)$

(6) $X \perp Y \Rightarrow I(X;Y) = 0$

Interpretations:

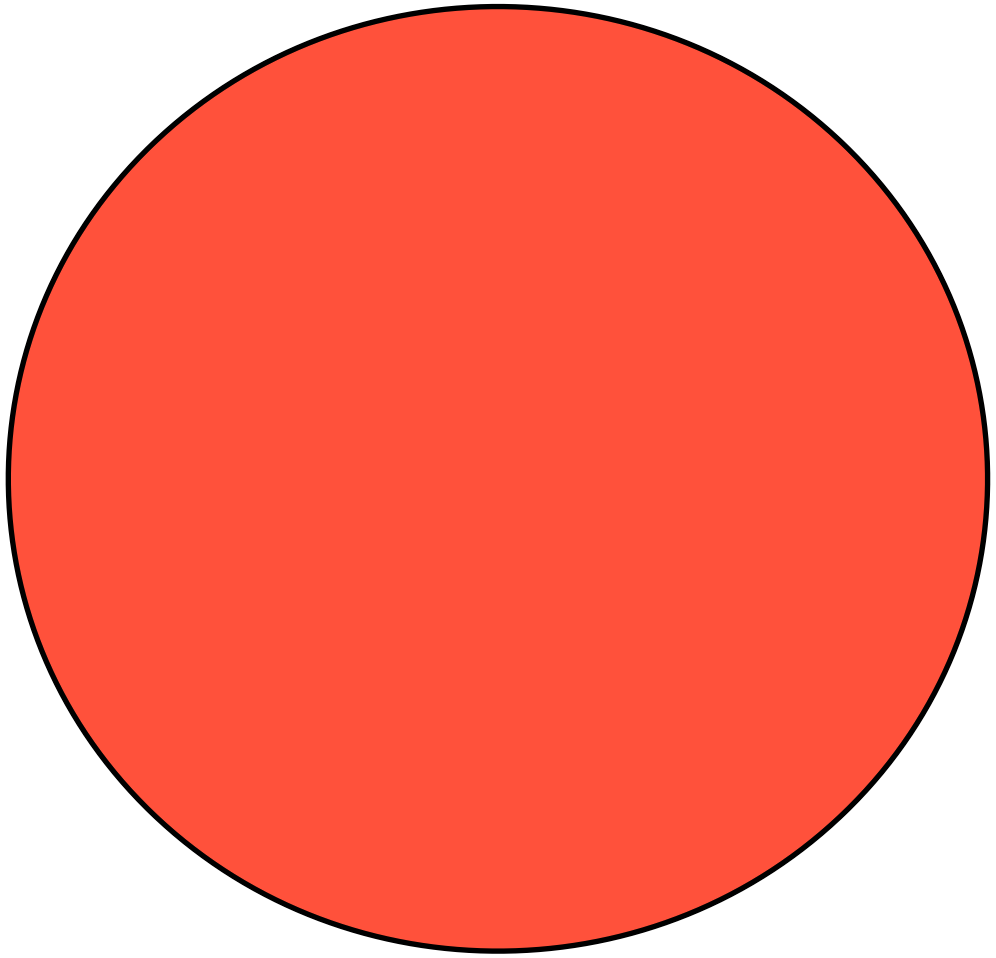
Information one variable has about another
Information shared between two variables
Measure of dependence between two variables

Information ...

Event Space Relationships of Information Quantifiers:

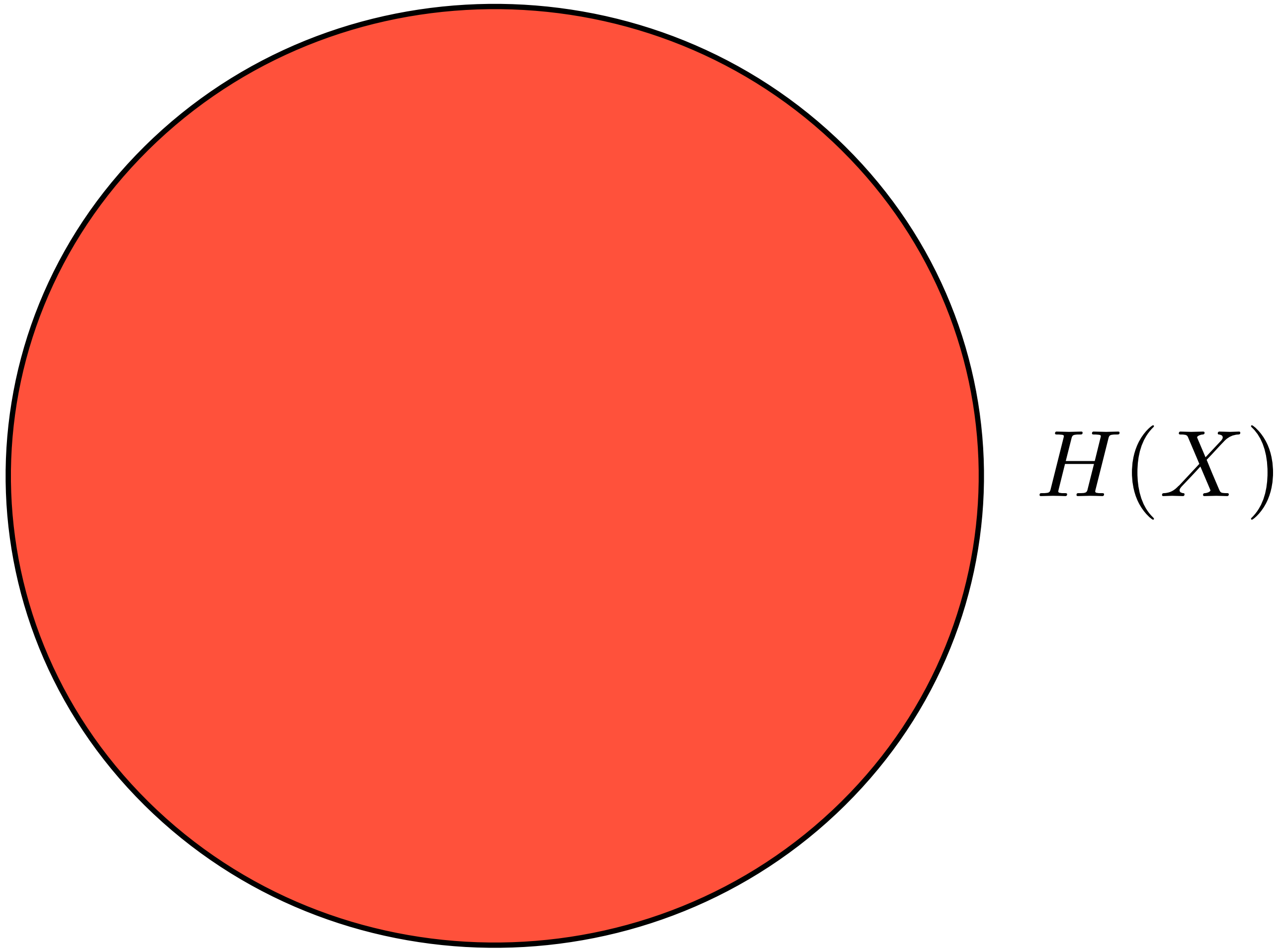
Information ...

Event Space Relationships of Information Quantifiers:



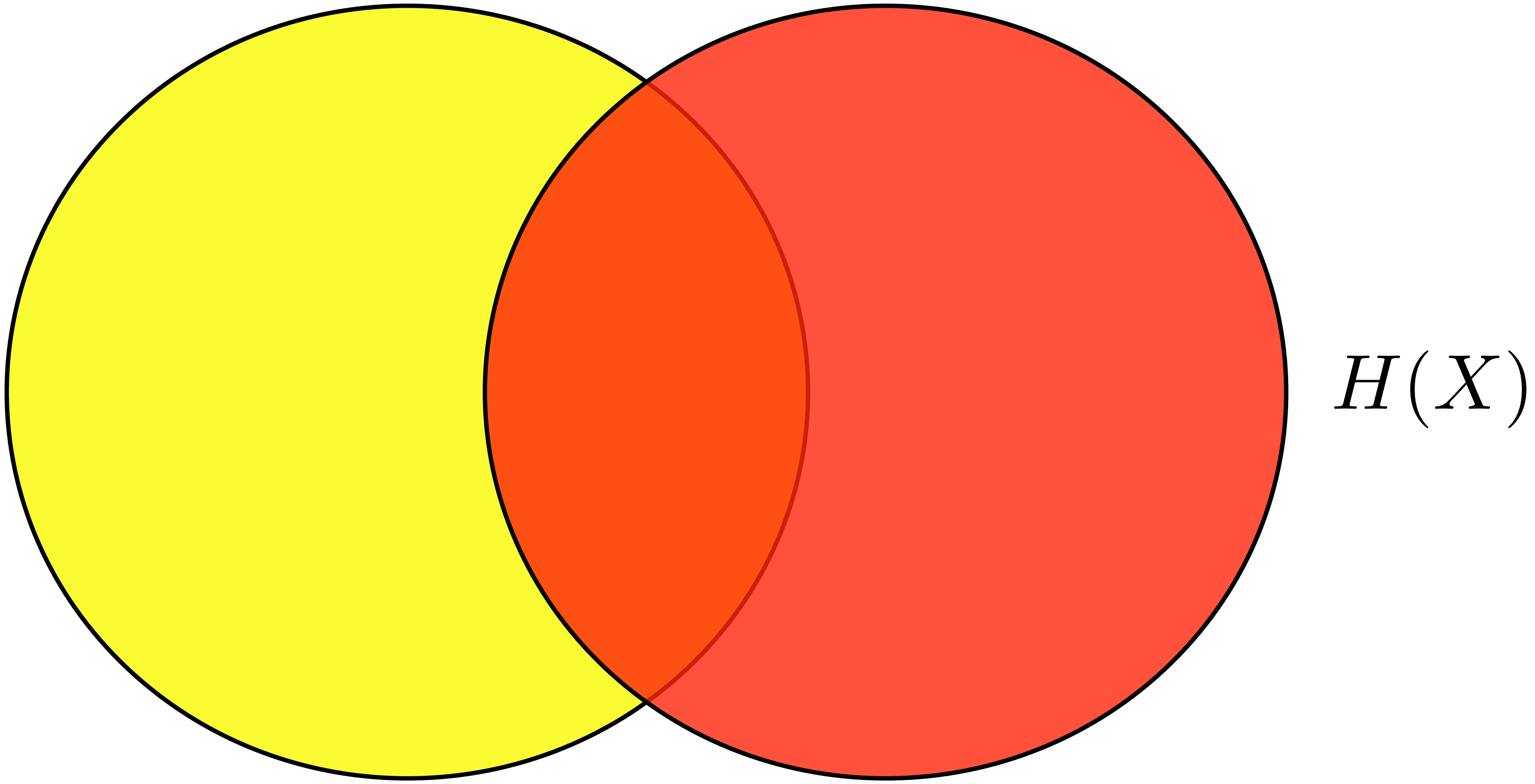
Information ...

Event Space Relationships of Information Quantifiers:



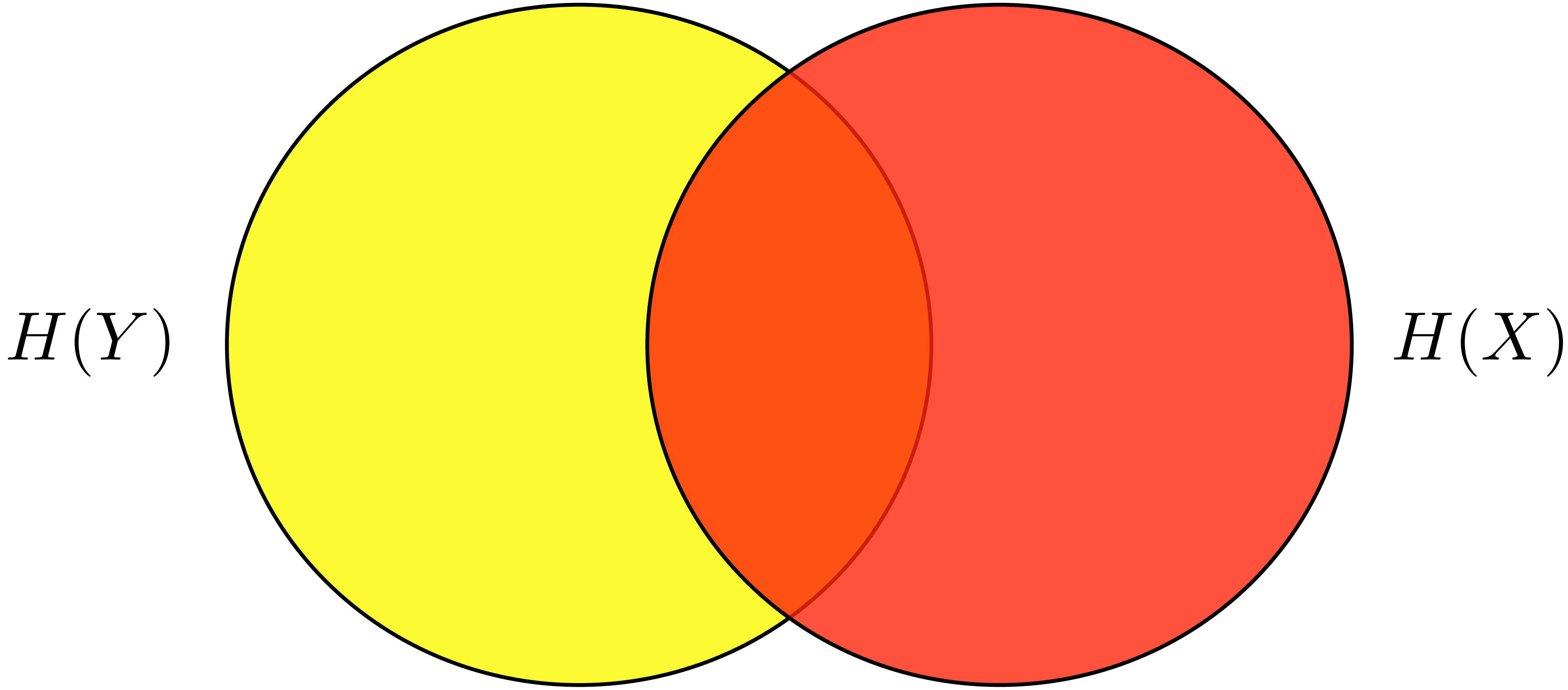
Information ...

Event Space Relationships of Information Quantifiers:



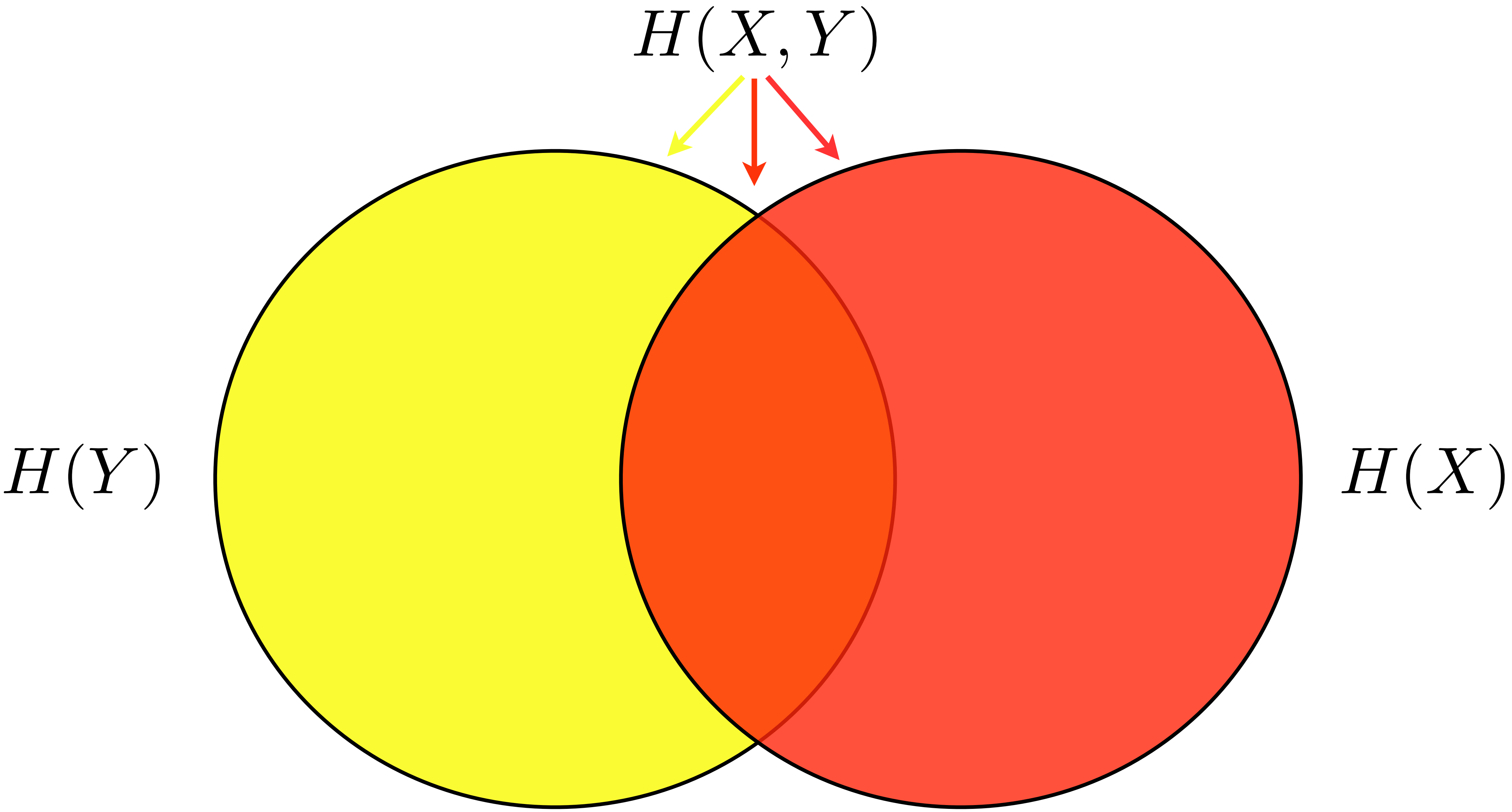
Information ...

Event Space Relationships of Information Quantifiers:



Information ...

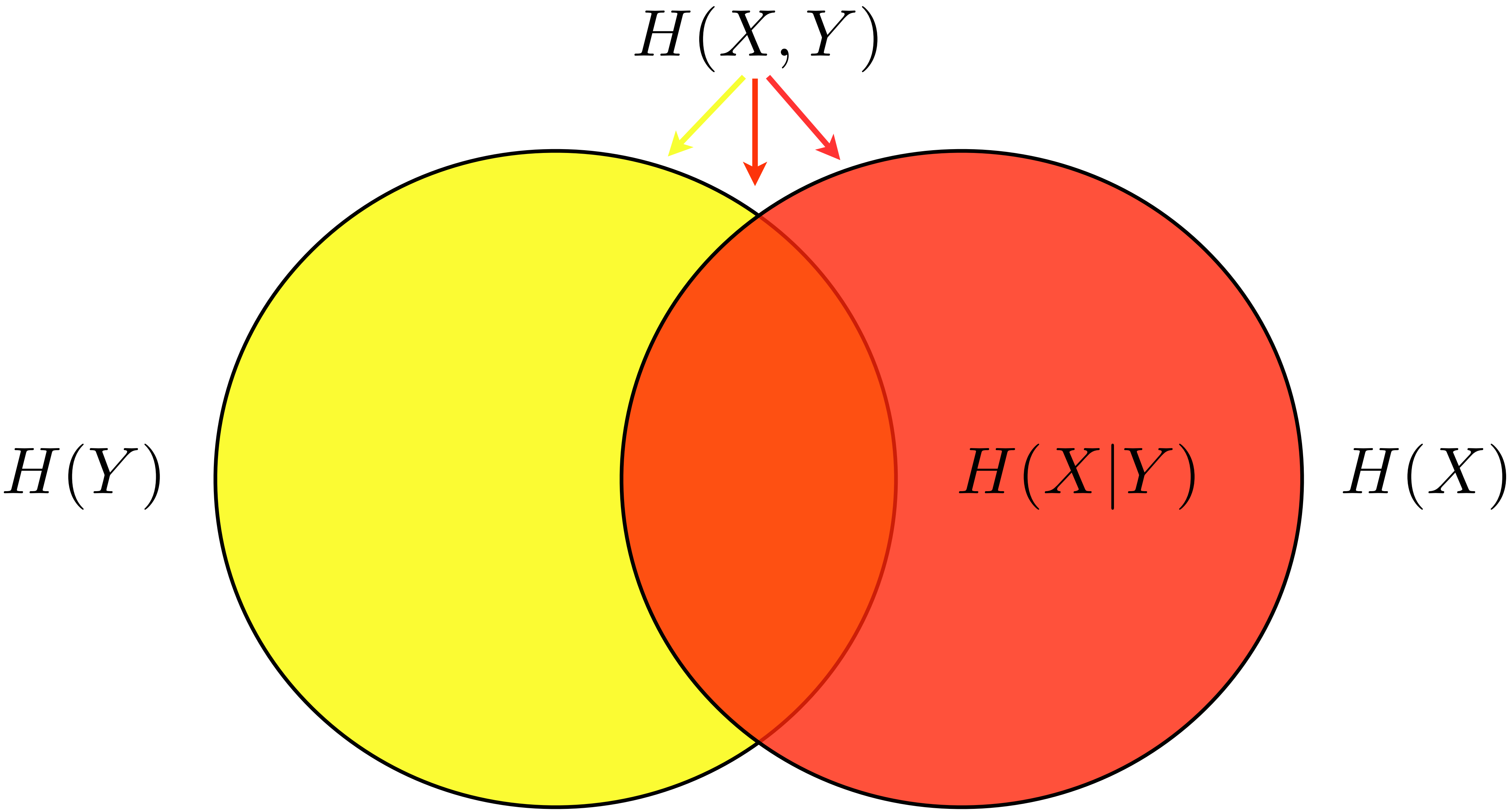
Event Space Relationships of Information Quantifiers:



Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

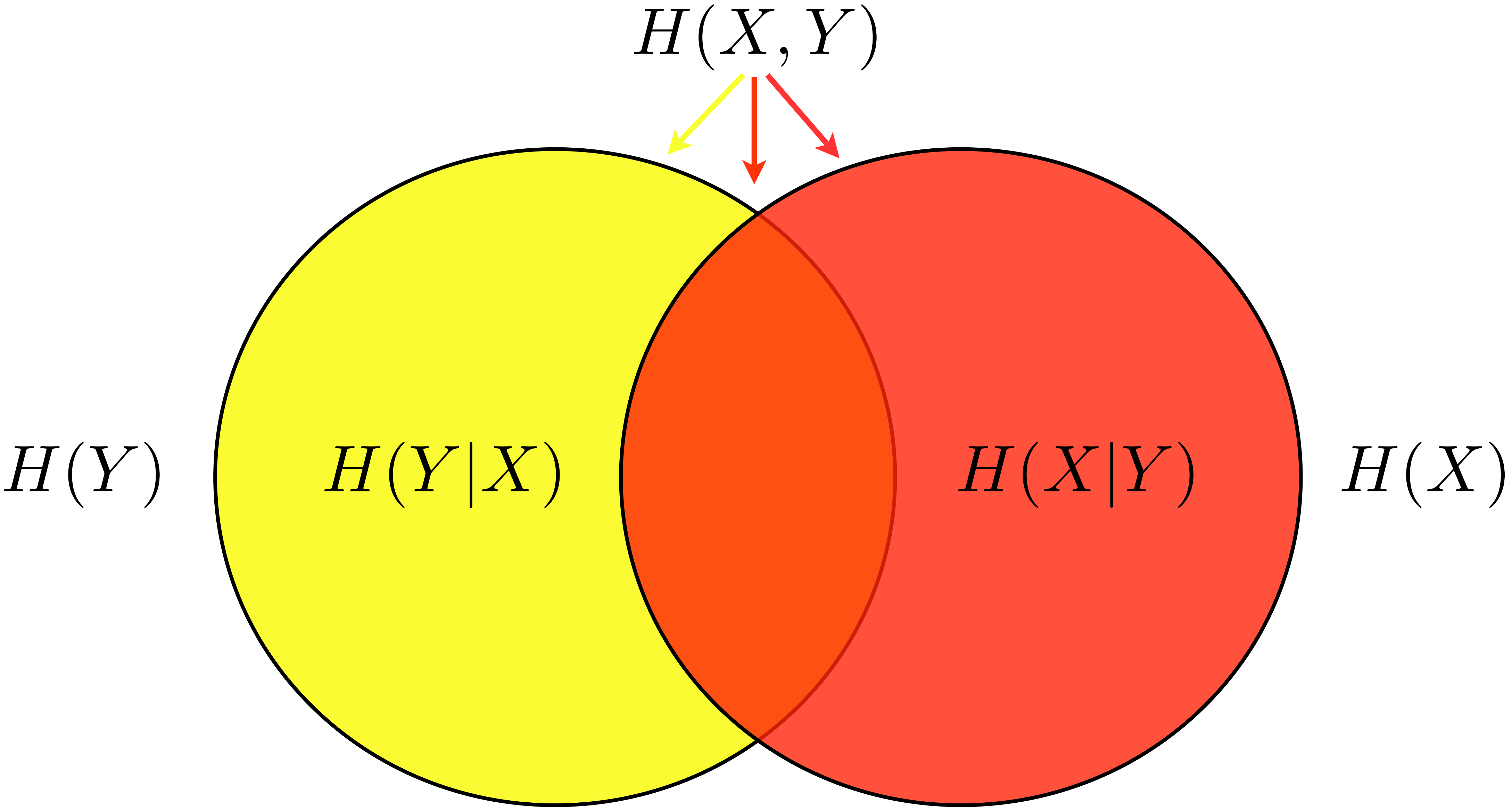
Information ...

Event Space Relationships of Information Quantifiers:



Information ...

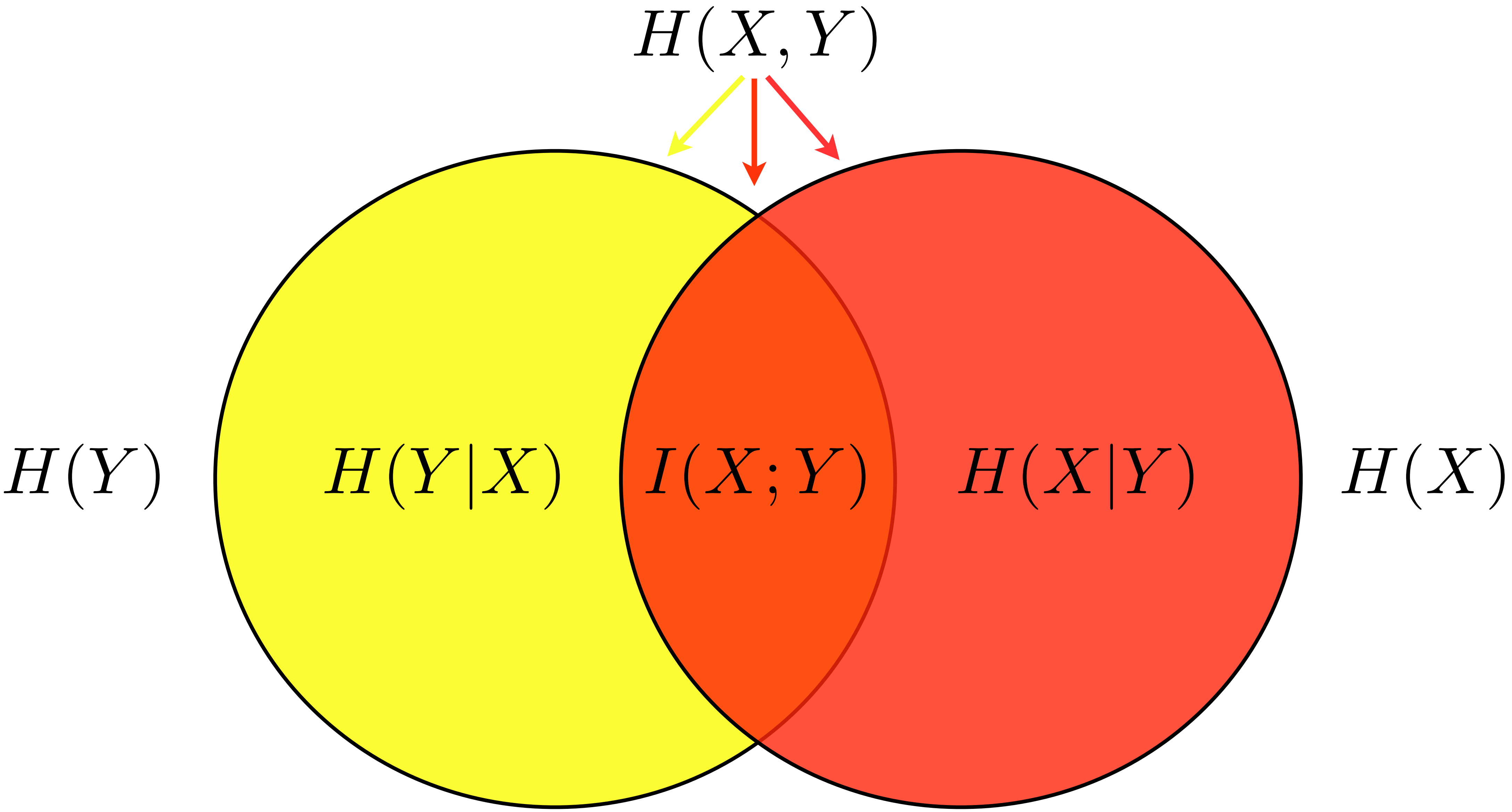
Event Space Relationships of Information Quantifiers:



Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Information ...

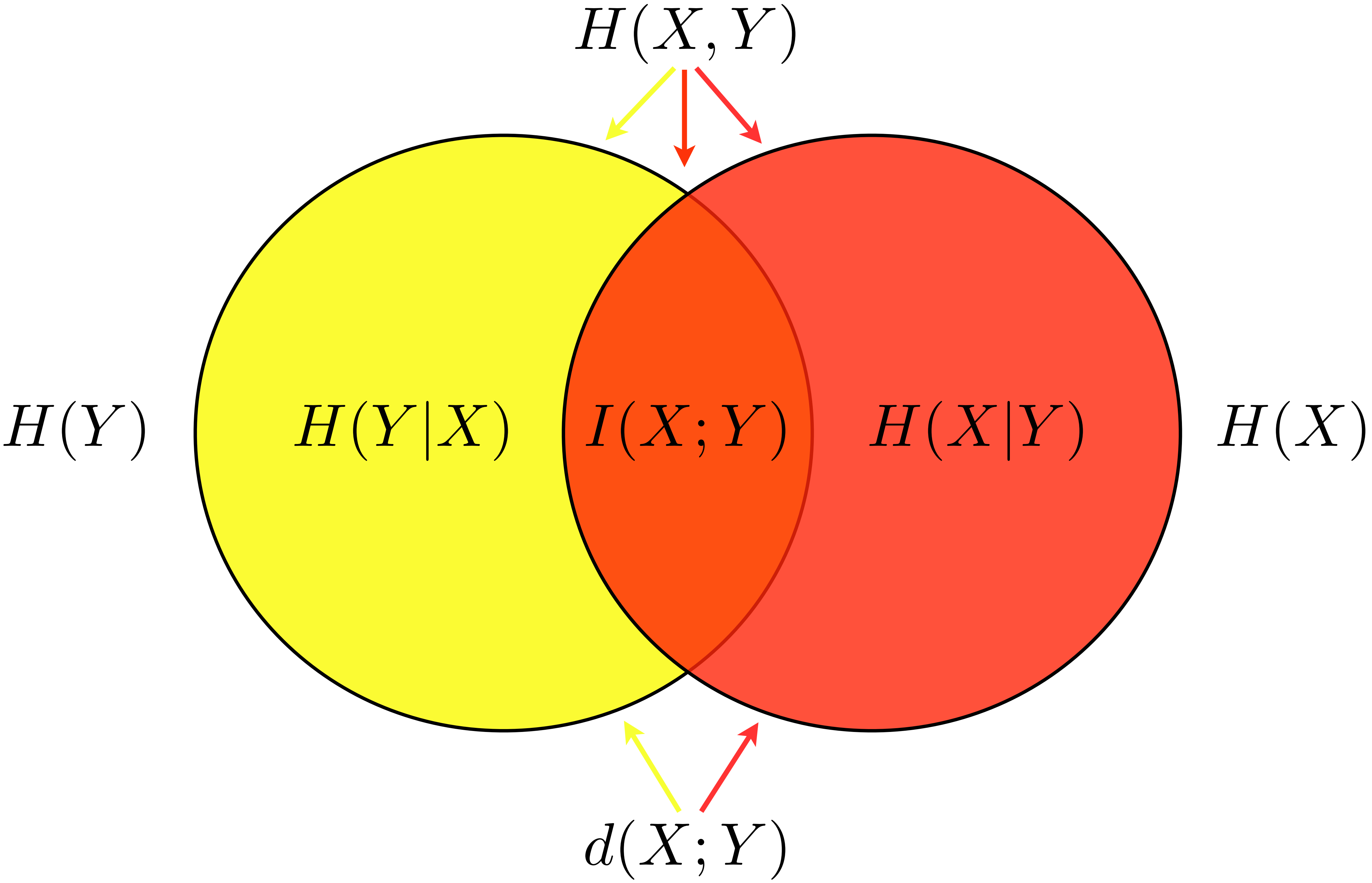
Event Space Relationships of Information Quantifiers:



Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Information ...

Event Space Relationships of Information Quantifiers:

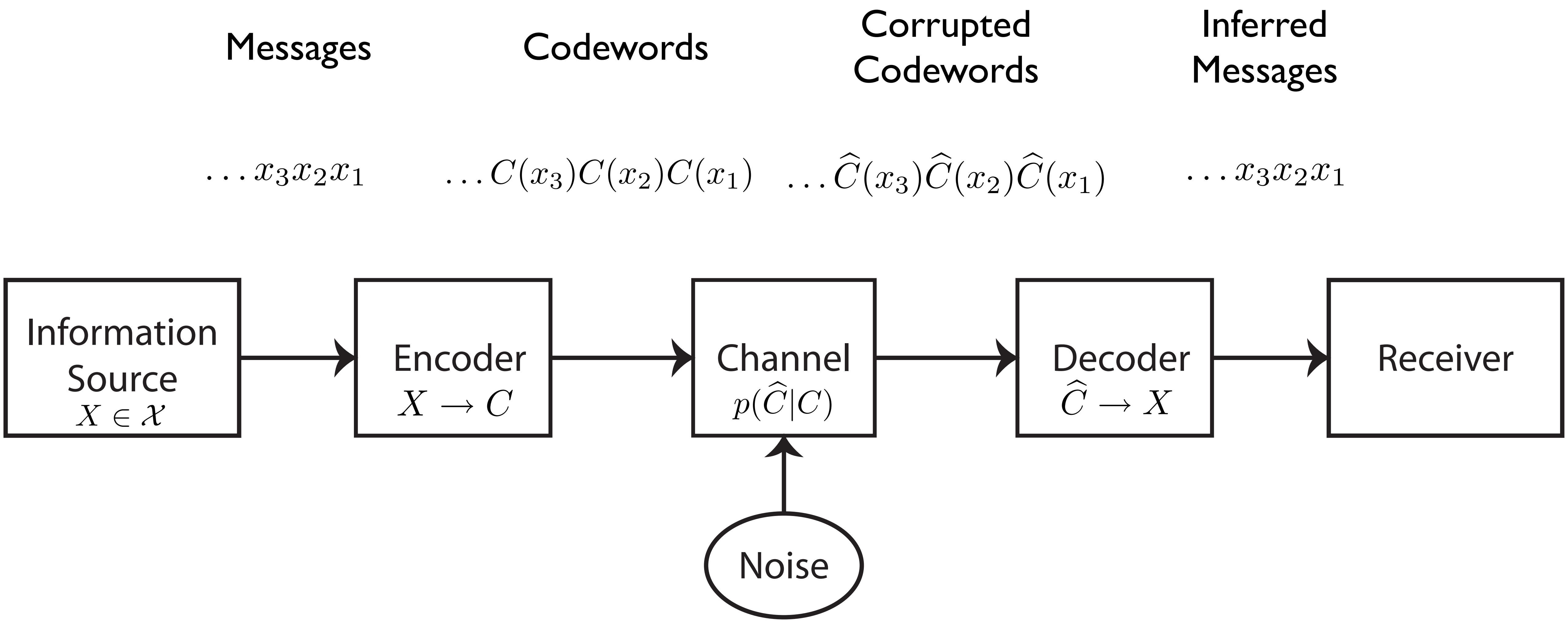


Why information?

1. Accounts for any type of co-relation
   - Statistical correlation ~ linear only
   - Information measures nonlinear correlation
2. Broadly applicable:
   - Many systems don't have "energy", physical modeling precluded
   - Information defined: social, biological, engineering, ... systems
3. Comparable units across different systems:
   - Correlation: Meters v. volts v. dollars v. ergs v. ...
   - Information: bits.
4. Probability theory ~ Statistics ~ Information
5. Complex systems:
   - Emergent patterns!
   - We don't know these ahead of time

Information in Processes ...

Communication channel:



Information in Processes ...

Real Information Theory:

How to compress a process:

Can't do better than $H(X)$

(Shannon's First Theorem)

How to communicate a process's data: $H(X) \leq \mathcal{C}$

Can transmit error-free at rates up to channel capacity

(Shannon's Second Theorem)

Both results give operational meaning to entropy.

Previously, entropy motivated as a measure of surprise.

Information in Processes

Information in Processes ...

Entropy Growth for Stationary Stochastic Processes: $\Pr(\overleftrightarrow{S})$

Block Entropy:

$$H(L) = H(\Pr(s^L)) = -\sum_{s^L \in \mathcal{A}} \Pr(s^L) \log_2 \Pr(s^L)$$

Monotonic increasing: $H(L) \geq H(L-1)$

Adding a random variable cannot decrease entropy:

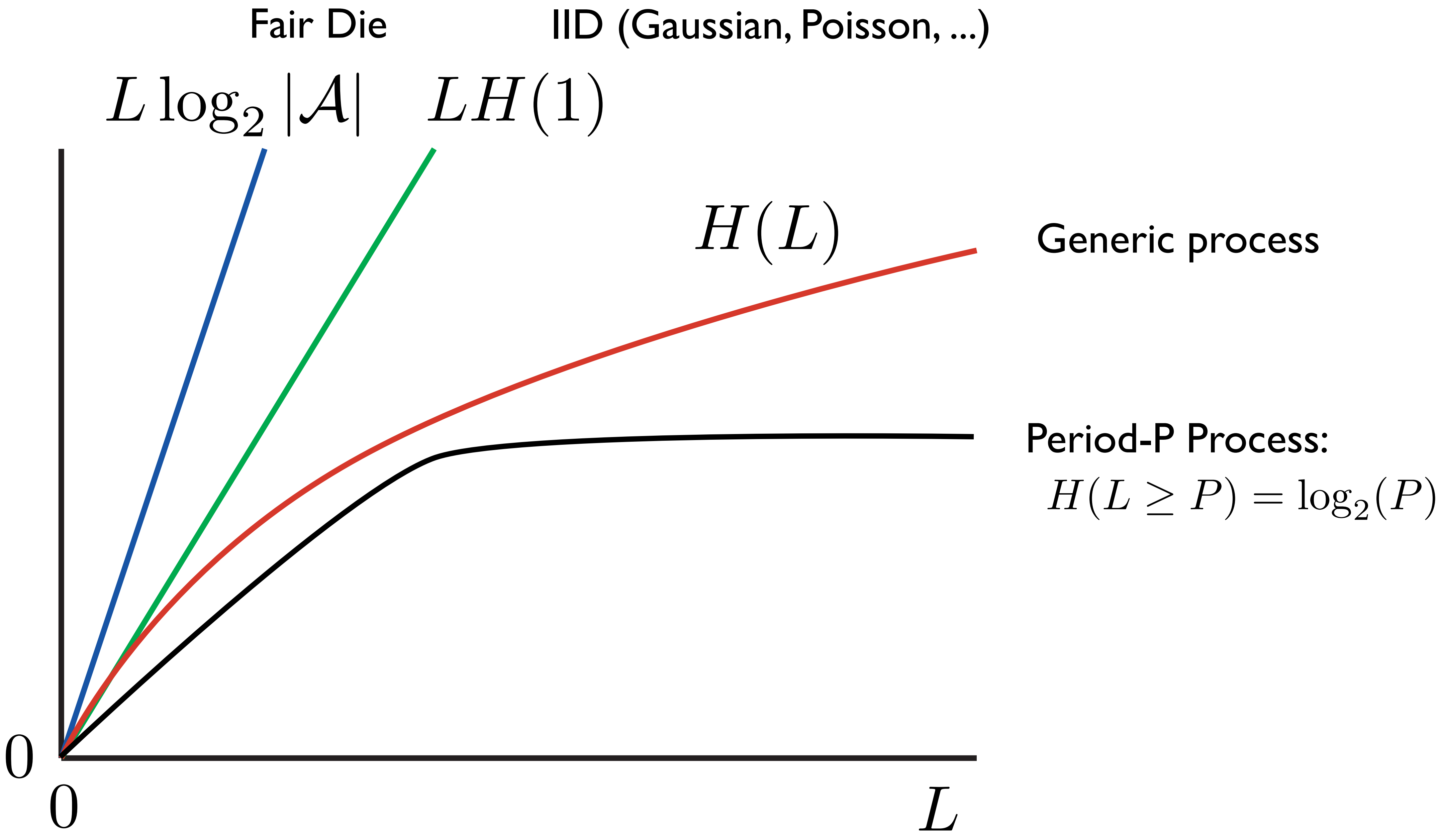
$$H(S_1, S_2, \ldots, S_L) \leq H(S_1, S_2, \ldots, S_L, S_{L+1})$$

No measurements, no information: $H(0) = 0$

Information in Processes ...

Entropy Growth for Stationary Stochastic Processes ...

Block Entropy ...



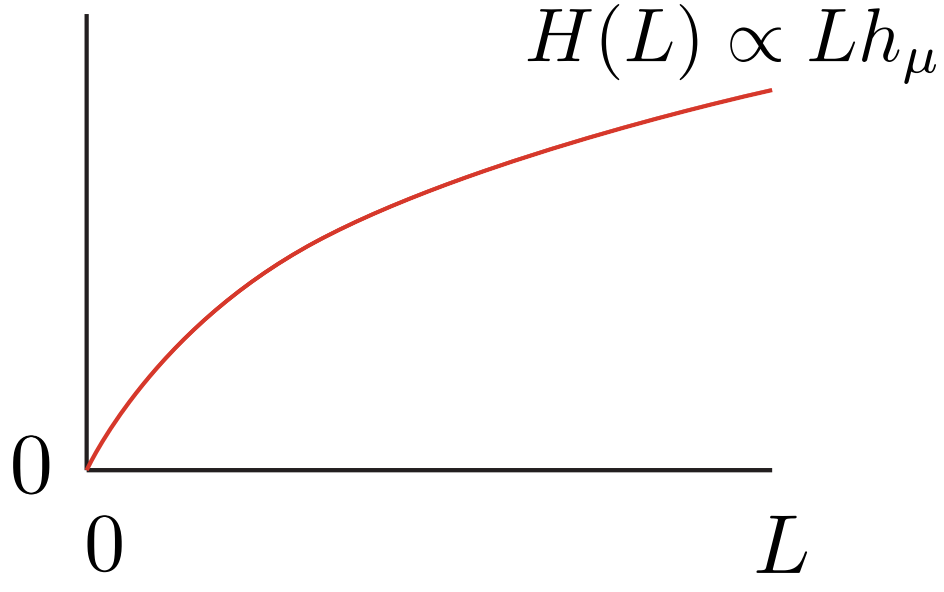
Information in Processes ...

Entropy Rates for Stationary Stochastic Processes:

Entropy per symbol is given by the Source Entropy Rate:

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

(When limits exists.)



Interpretations:

- Asymptotic growth rate of entropy
- Irreducible randomness of process
- Average description length (per symbol) of process

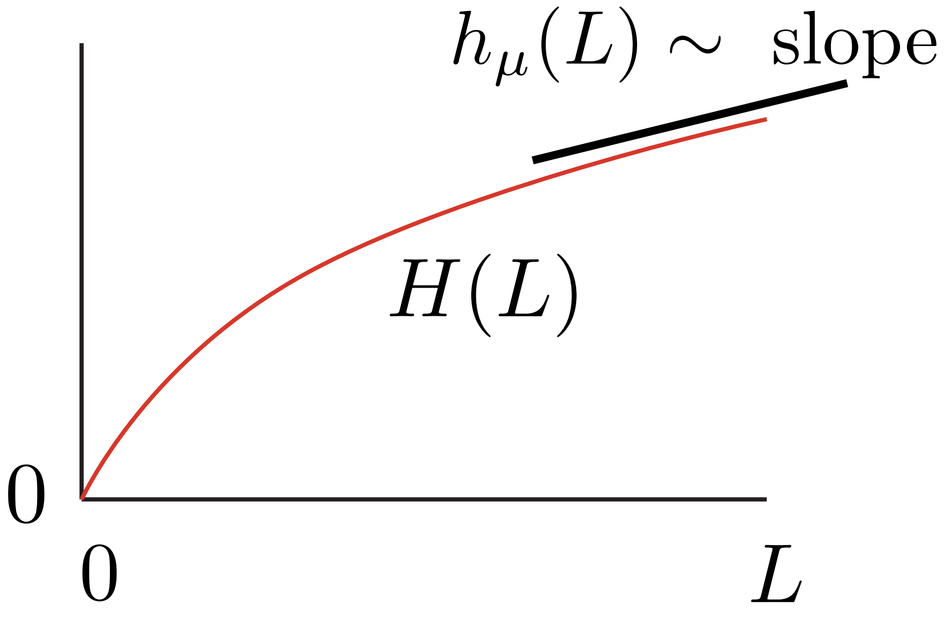
Information in Processes ...

Entropy Rates for Stationary Stochastic Processes ...

Length-L Estimate of Entropy Rate:

$$\widehat{h}_\mu(L) = H(L) - H(L-1)$$

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



Monotonic decreasing: $\widehat{h}_\mu(L) \leq \widehat{h}_\mu(L-1)$

Conditioning cannot increase entropy:

$$H(s_L | s_1 \cdots s_{L-1}) \leq H(s_L | s_2 \cdots s_{L-1}) = H(s_{L-1} | s_1 \cdots s_{L-2})$$

Information in Processes ...

Entropy Rates for Stationary Stochastic Processes:

Entropy rate ...

$$\widehat{h}_\mu = \lim_{L \to \infty} \widehat{h}_\mu(L) = \lim_{L \to \infty} H(s_0 | \overleftarrow{s}^L) = H(s_0 | \overleftarrow{s})$$

Interpretations:

- Uncertainty in next measurement, given past
- A measure of unpredictability
- Asymptotic slope of block entropy

Alternate entropy rate definitions agree:

$$\widehat{h}_\mu = h_\mu$$

Information in Processes ...

Entropy Rate for a Markov chain: $\{V, T\}$

$$
\begin{aligned}
h_\mu &= \lim_{L\to\infty} h_\mu(L) \\
&= \lim_{L\to\infty} H(v_L | v_1 \cdots v_{L-1}) \\
&= \lim_{L\to\infty} H(v_L | v_{L-1})
\end{aligned}
$$

Assuming asymptotic state distribution:

- Process in statistical equilibrium
- Process running for a long time
- Forgotten it's initial distribution

Closed-form:

$$h_\mu = - \sum_{v \in V} p_v(\infty) \sum_{v' \in V} T_{vv'} \log_2 T_{vv'}$$

$$\vec{p}(n) = \vec{p}(0) T^n$$

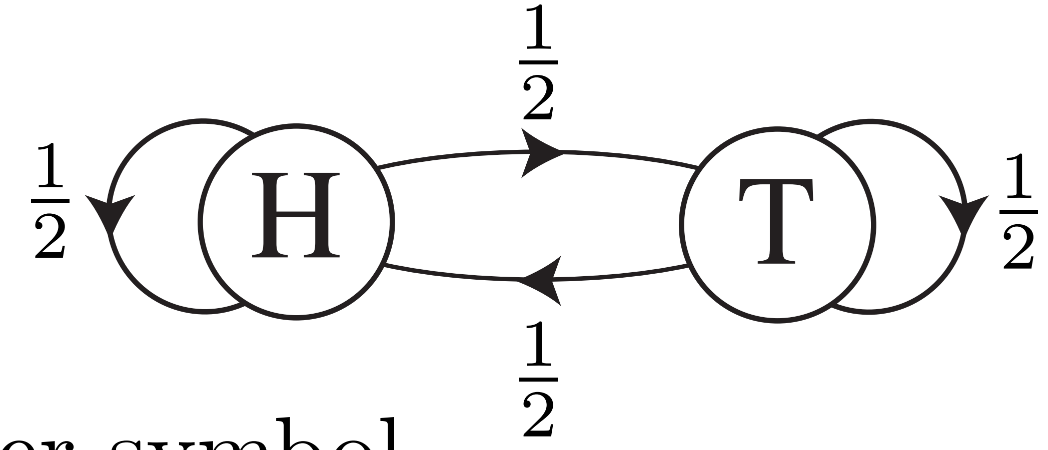
$$\vec{p}(\infty) = \vec{p}(\infty) T^n$$

Information in Processes ...

Entropy Rate for Markov chains ...

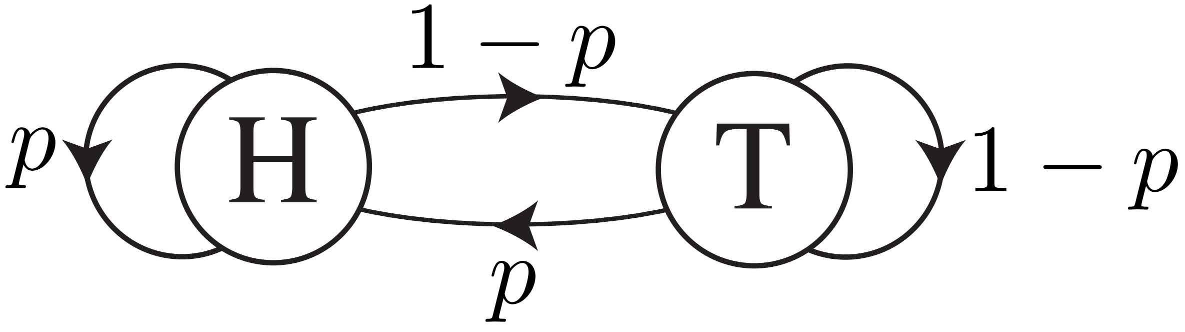
Examples:

(1) Fair Coin:



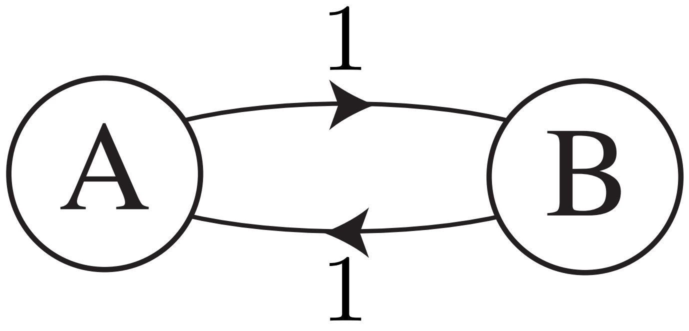
$$h_\mu = 1 \text{ bit per symbol}$$

(2) Biased Coin:



$$h_\mu = H(p) \text{ bits per symbol}$$

(3) Period-2 Process:



$$h_\mu = 0 \text{ bits per symbol}$$

Information in Processes ...

Entropy Rate for Unifilar Hidden Markov Chain:

Internal: $\{V, T\}$
Observed: $\{T^{(s)} : s \in \mathcal{A}\}$

Closed-form for entropy rate:

$$h_\mu = - \sum_{v \in V} p_v(\infty) \sum_{s \in \mathcal{A}} \sum_{v' \in V} T^{(s)}_{vv'} \log_2 T^{(s)}_{vv'}$$

Due to unifilarity:

Observed sequences are (effectively) unique state paths

Information in Processes ...

Entropy Rate for Nonunifilar Hidden Markov Chain:

Internal: $\{V, T\}$
Observed: $\{T^{(s)} : s \in \mathcal{A}\}$

Entropy rate: No closed-form!

$$h_\mu \neq - \sum_{v \in V} \pi_v \sum_{s \in \mathcal{A}} \sum_{v' \in V} T^{(s)}_{vv'} \log_2 T^{(s)}_{vv'} \qquad \pi_v = p_v(\infty)$$

Upper and Lower Bounds:

$$H(S_L | V_1 S_1 \cdots S_{L-1}) \leq h_\mu(L) \leq H(S_L | S_1 \cdots S_{L-1})$$

Unrealistic for inference: Must know about internal states.
Unrealistic for analysis: Simulate chain, do empirical estimate.

Entropy rate? But there exists a way ... stay tuned!

Information in Processes ...

Entropy Convergence:

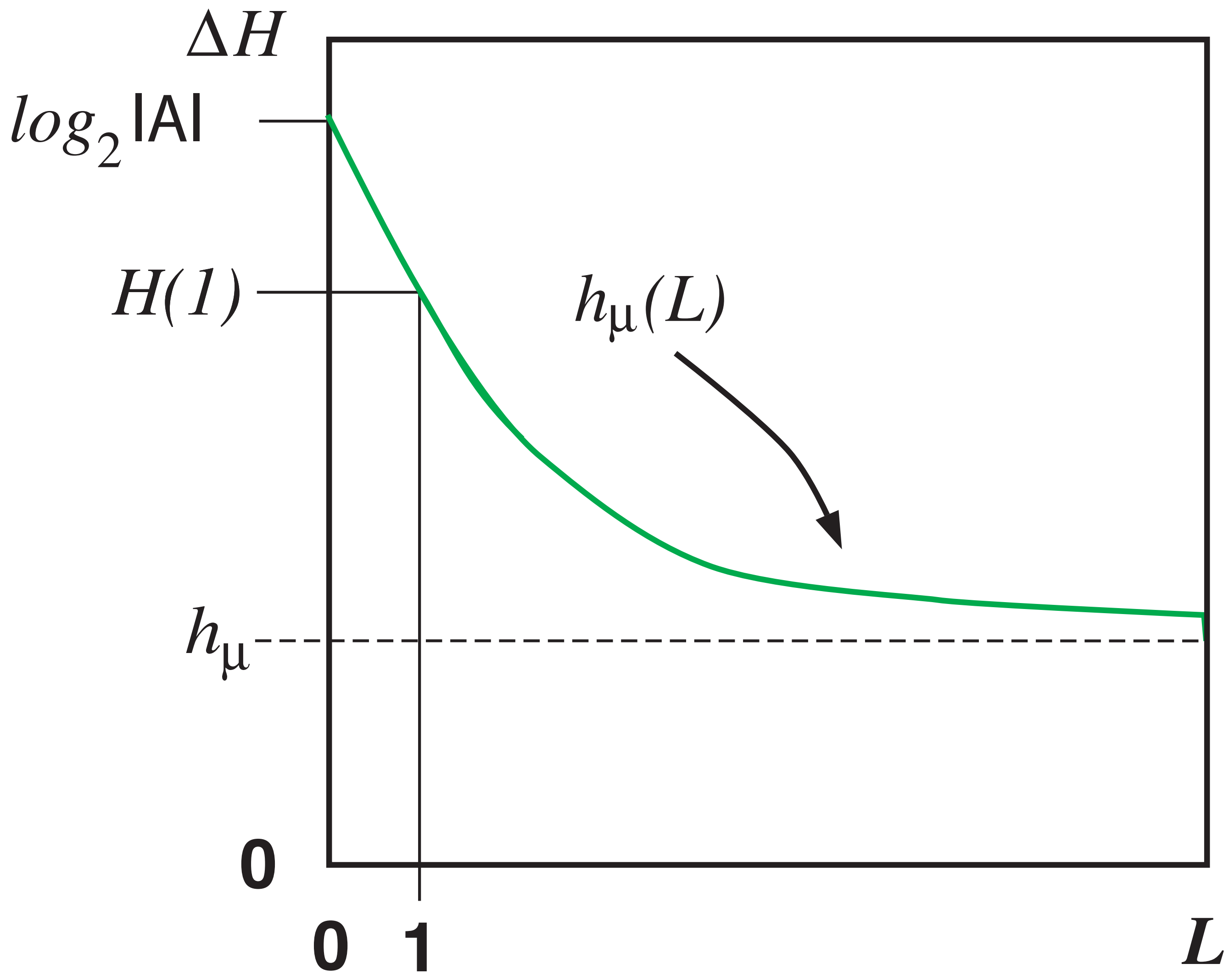
Length-L entropy rate estimate:

$$h_\mu(L) = H(L) - H(L-1)$$

$$h_\mu(L) = \Delta H(L)$$

Monotonic decreasing:

$$h_\mu(L) \leq h_\mu(L-1)$$



Process appears less random
as account for longer correlations

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Memory in Processes

Information in Processes ...

Motivation:

Previous: Measures of randomness of information source
  Block entropy $H(L)$
  Entropy rate h_μ

Current target point:
  Measures of memory & information storage

Big Picture:
  Complementary.

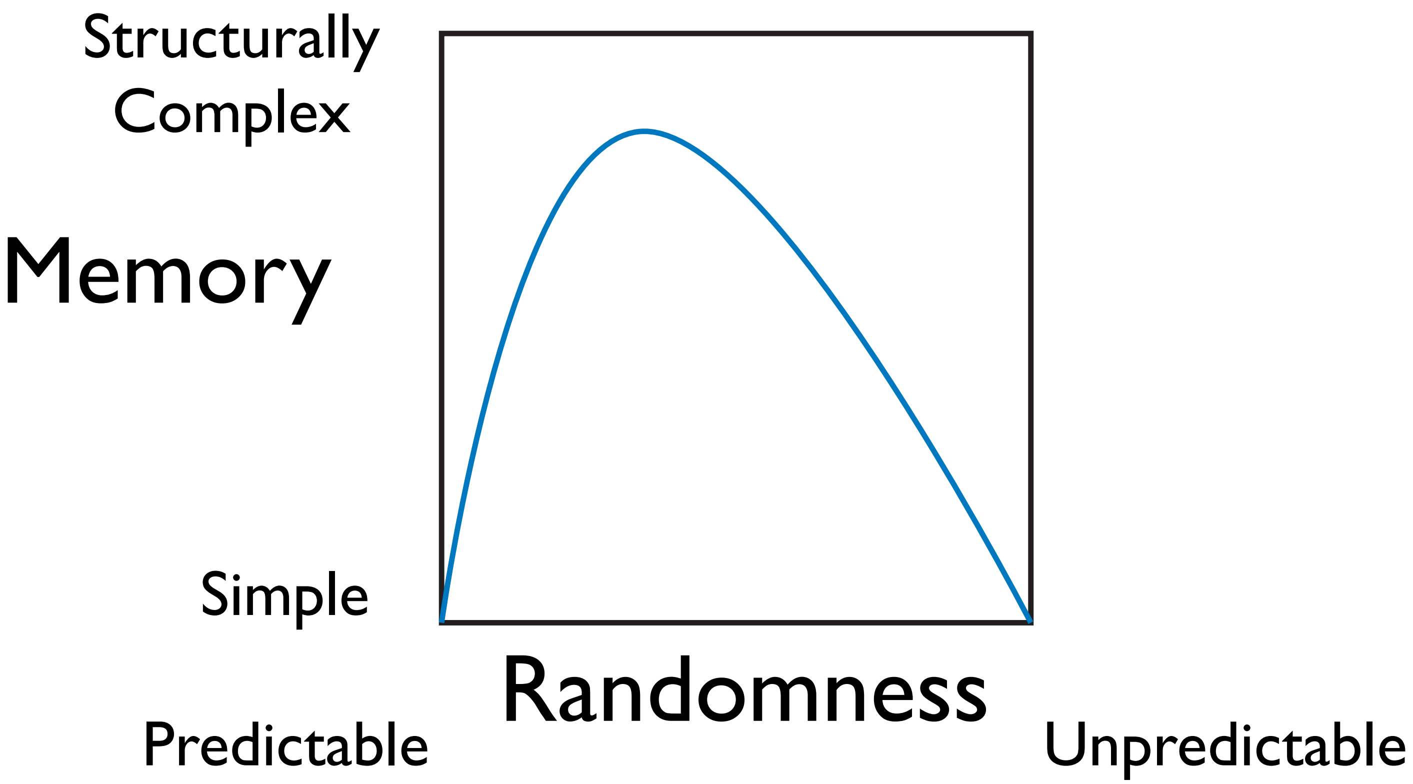
Information in Processes ...

Motivation:

Previous: Measures of randomness of information source
  Block entropy $H(L)$
  Entropy rate h_μ

Current target point:
  Measures of memory & information storage

Big Picture:
  Complementary.



Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

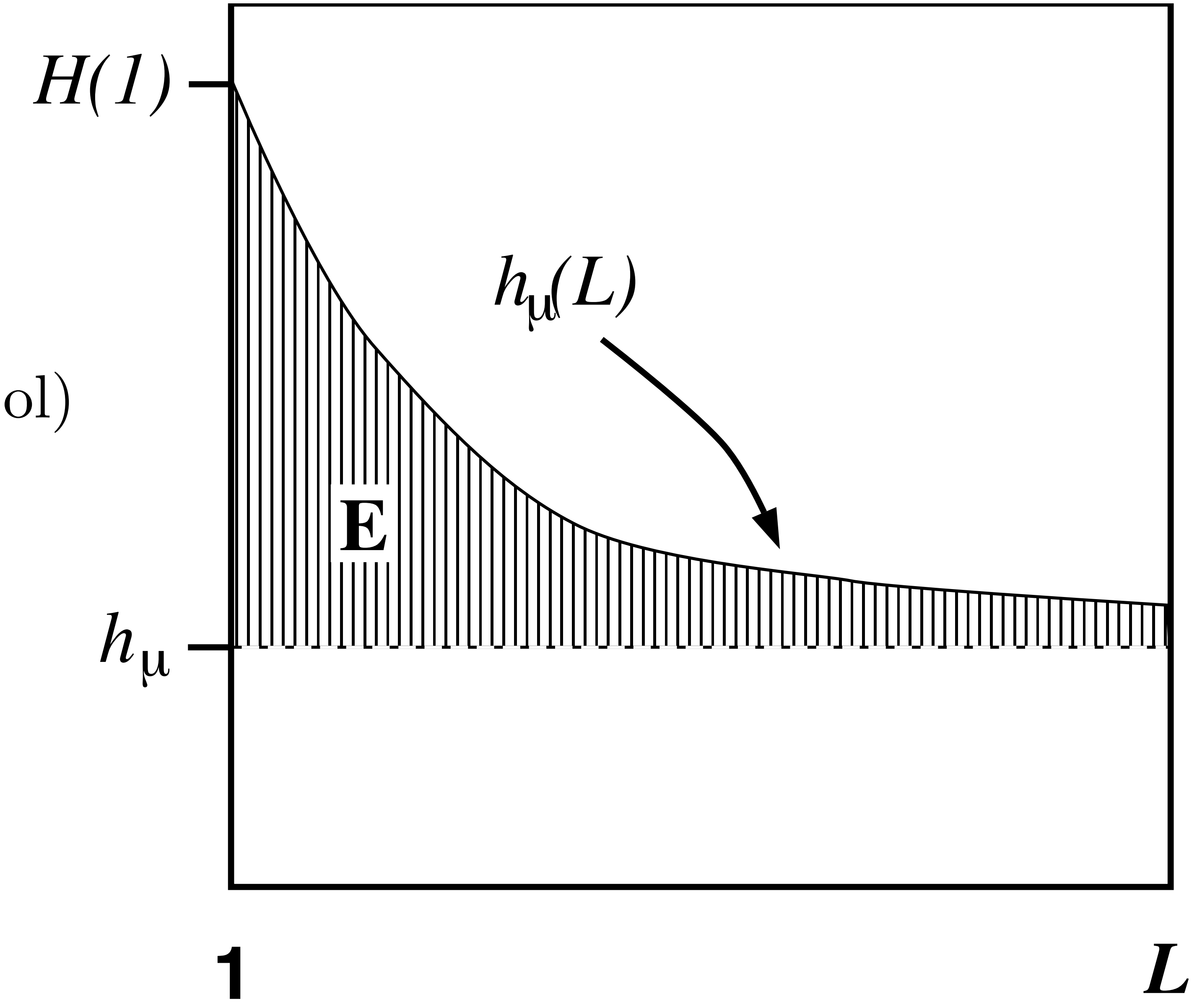
Memory in Processes ...

Excess Entropy:

As entropy convergence:

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

$(\Delta L = 1 \text{ symbol})$



As intrinsic redundancy:

$$\mathbf{E} = \sum_{L=1}^{\infty} r(L)$$

Properties:

(1) Units: $\mathbf{E} = [\text{bits}]$

(2) Positive: $\mathbf{E} \geq 0$

(3) Controls convergence to actual randomness.

(4) Slow convergence ⇔ Correlations at longer words.

(5) Complementary to entropy rate.

Memory in Processes ...

Excess Entropy ...

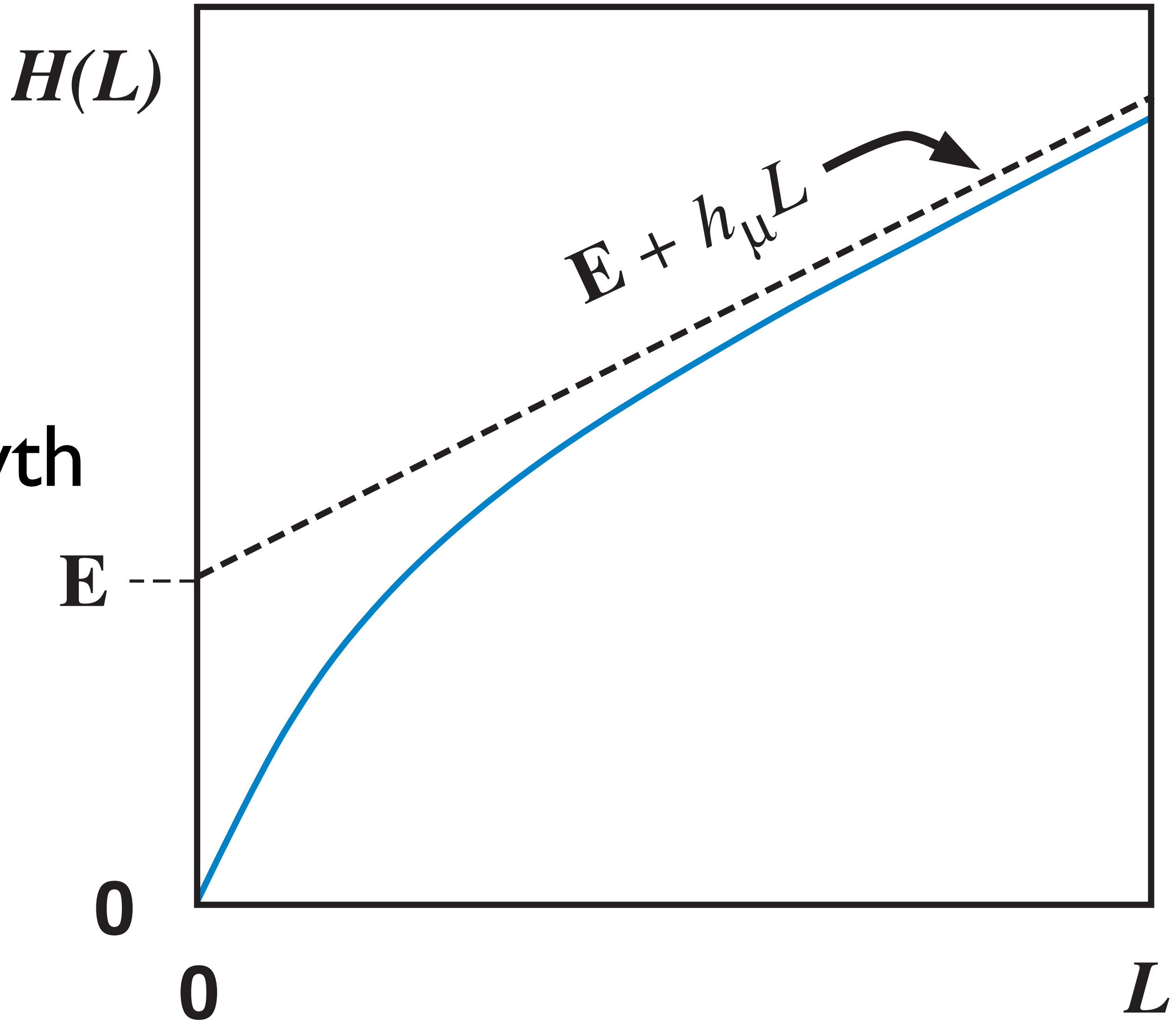
Asymptote of entropy growth:

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

That is,

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

Y-Intercept of entropy growth



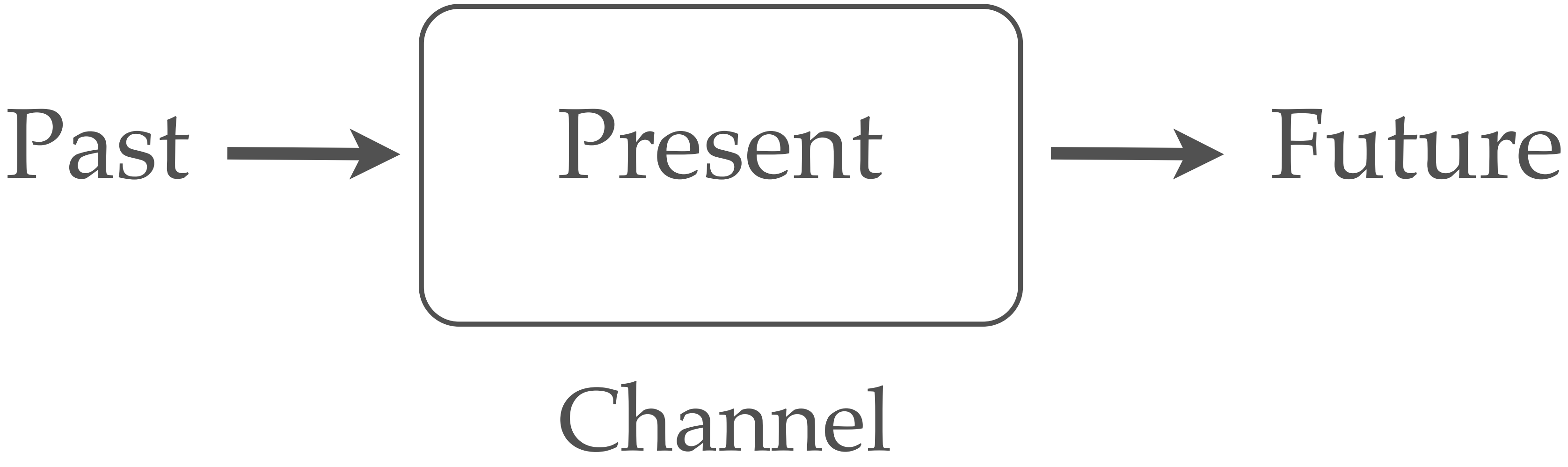
Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Memory in Processes ...

Excess Entropy ...

Mutual information between past and future: Process as channel

Process $\Pr(\overleftarrow{X}, \overrightarrow{X})$ communicates past $\overleftarrow{X}$ to future $\overrightarrow{X}$:



Excess Entropy as Channel Utilization:

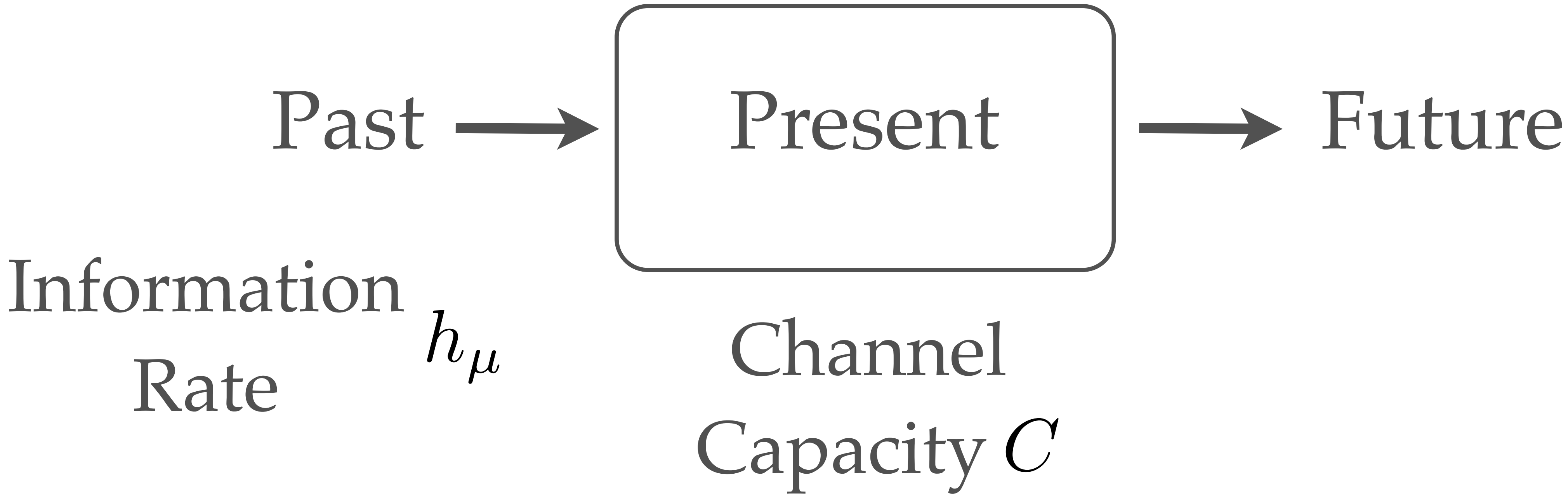
$$\mathbf{E} = I[\overleftarrow{X}; \overrightarrow{X}]$$

Memory in Processes ...

Excess Entropy ...

Mutual information between past and future: Process as channel

Process $\Pr(\overleftarrow{X}, \overrightarrow{X})$ communicates past $\overleftarrow{X}$ to future $\overrightarrow{X}$:



Past → Present → Future

Information Rate h_μ

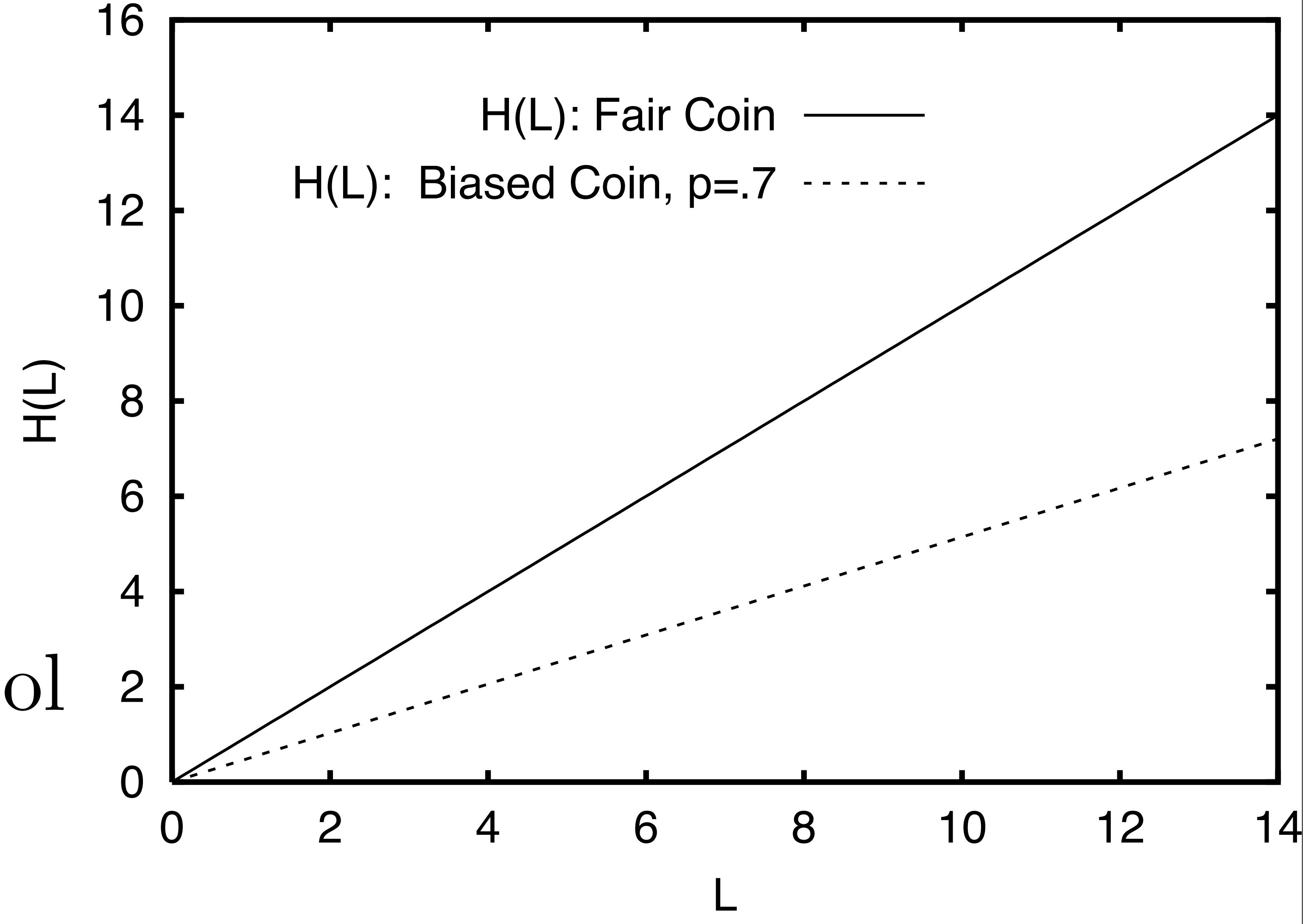
Channel Capacity C

Excess Entropy as Channel Utilization:

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

Memory in Processes ...

Examples of Excess Entropy:



Fair Coin:

$$h_\mu = 1 \text{ bit per symbol}$$

$$\mathbf{E} = 0 \text{ bits}$$

Biased Coin:

$$h_\mu = H(p) \text{ bits per symbol}$$

$$\mathbf{E} = 0 \text{ bits}$$

Any IID Process:

$$h_\mu = H(X) \text{ bits per symbol}$$

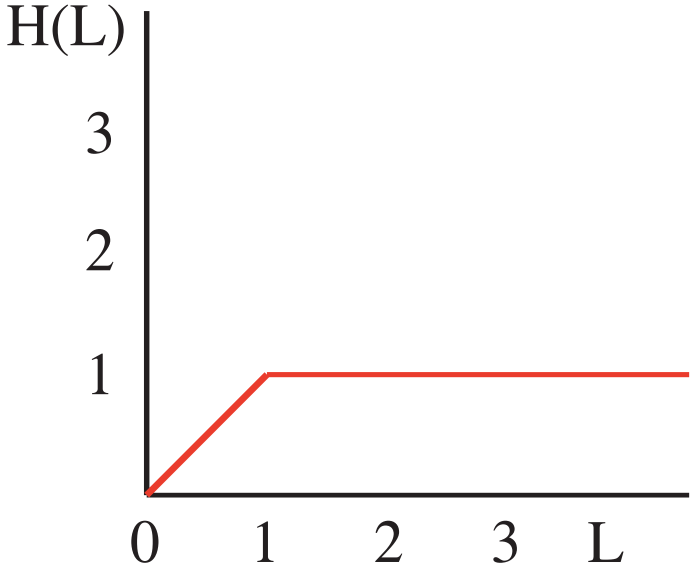
$$\mathbf{E} = 0 \text{ bits}$$

Memory in Processes ...

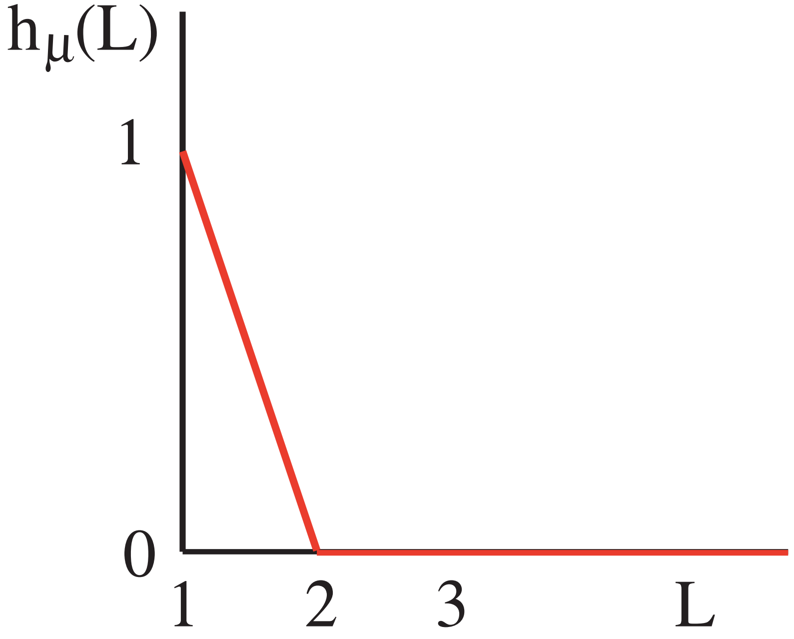
Examples of Excess Entropy ...

Period-2 Process: 010101010101

$$H(1) = 1$$
$$H(2) = 1$$
$$H(3) = 1$$



$$h_\mu(1) = 1$$
$$h_\mu(2) = 0$$
$$h_\mu(3) = 0$$



$$h_\mu = 0 \text{ bits per symbol}$$

$$\mathbf{E} = 1 \text{ bit}$$

Meaning:

1 bit of phase information

0-phase or 1-phase?

Memory in Processes ...

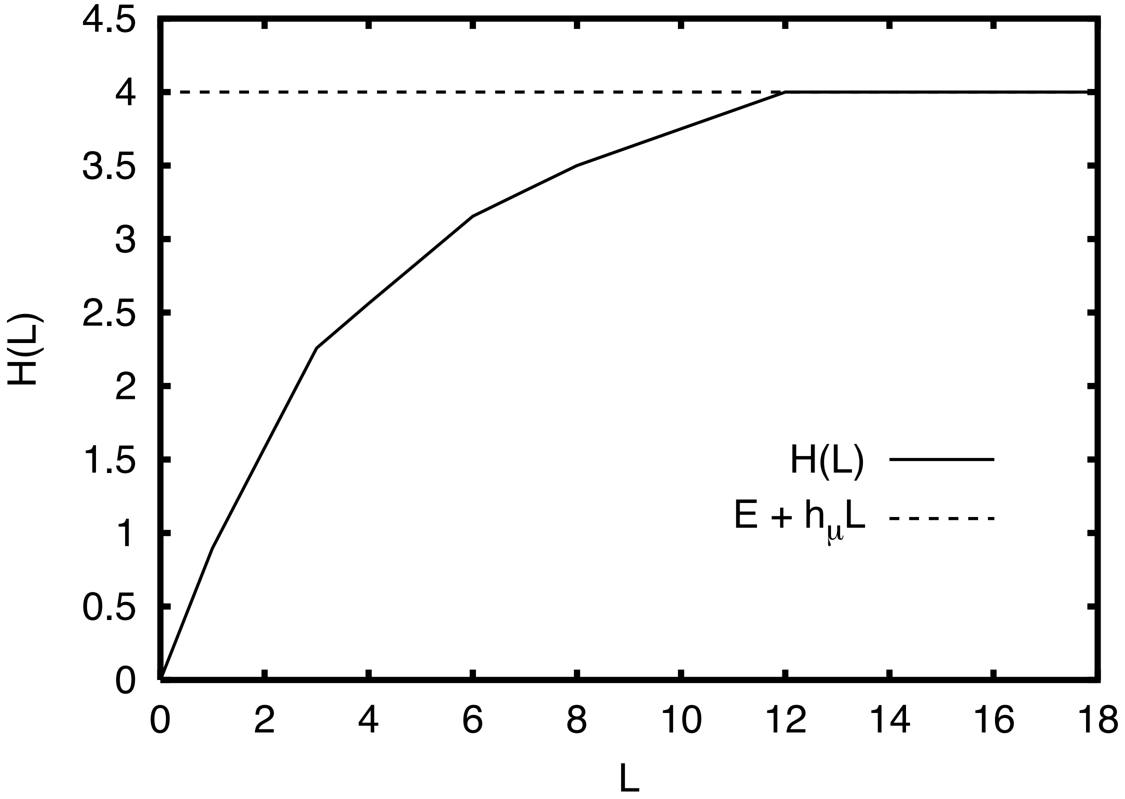
Examples of Excess Entropy ...

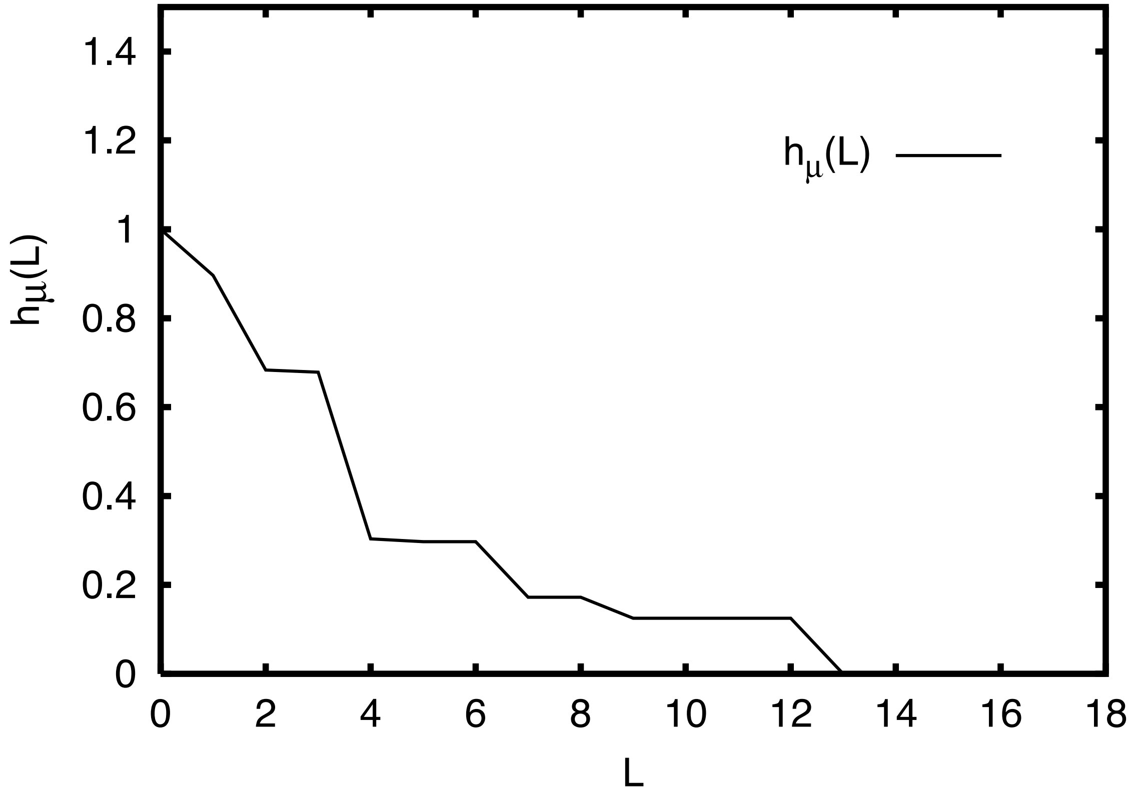
Period-16 Process:

$$(1010111011101110)^\infty$$

$$h_\mu = 0 \text{ bits per symbol}$$

$$\mathbf{E} = 4 \text{ bits}$$





Period-P Processes:

$$h_\mu = 0 \text{ bits per symbol}$$

$$\mathbf{E} = \log_2 P \text{ bits}$$

Cf., entropy rate does not distinguish periodic processes!

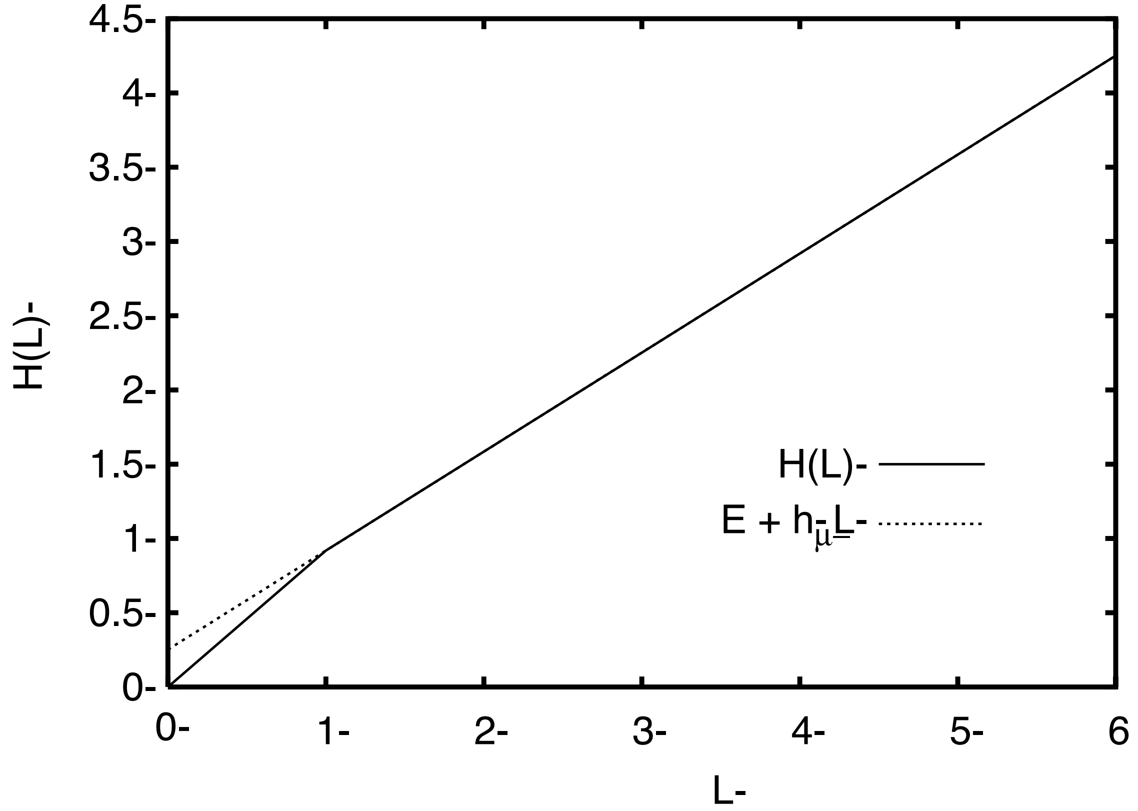
Memory in Processes ...

Examples of Excess Entropy ...

Golden Mean Process:

$$h_\mu = \frac{2}{3} \text{ bits per symbol}$$

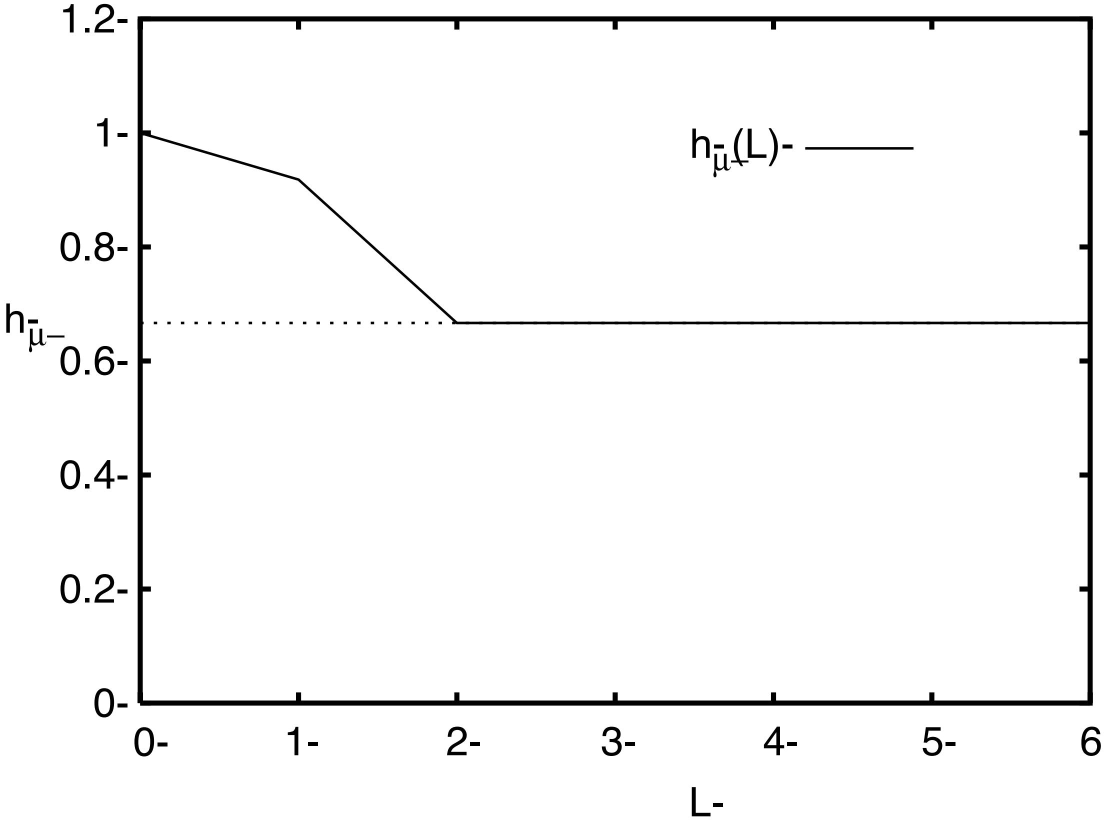
$$\mathbf{E} \approx 0.2516 \text{ bits}$$



R-Block Markov Chain:

$$\mathbf{E} = H(R) - R \cdot h_\mu$$

(E.g., 1D Ising Spin System)



Memory in Processes ...

Examples of Excess Entropy:

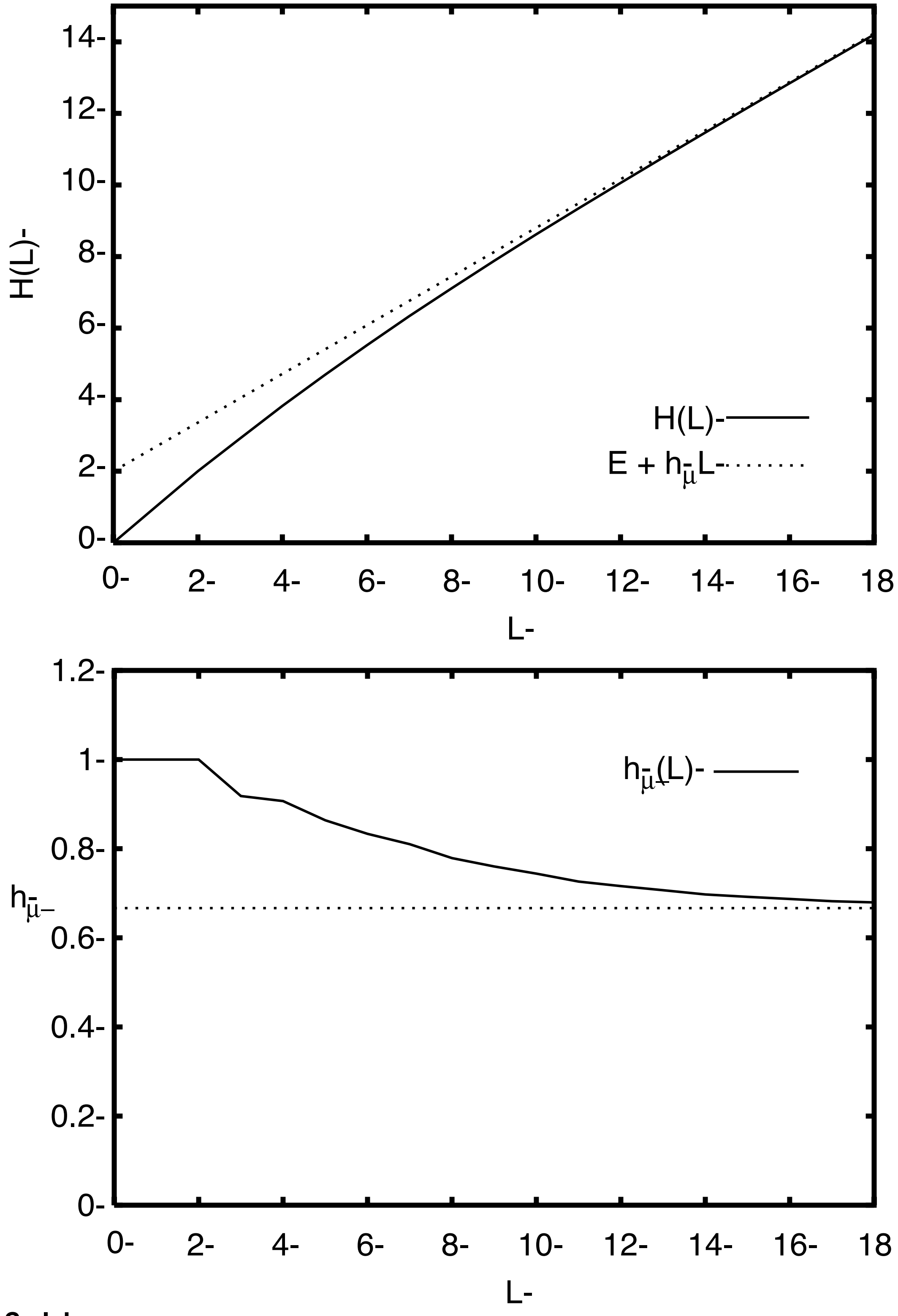
Finitary Processes: Exponential entropy convergence

Random-Random XOR (RRXOR) Process:

$$S_t = S_{t-1} \text{ XOR } S_{t-2}$$

$$h_\mu = \frac{2}{3} \text{ bits per symbol}$$

$$\mathbf{E} \approx 2.252 \text{ bits}$$



Finitary processes: Exponential convergence:

$$h_\mu(L) - h_\mu \approx 2^{-\gamma L}$$

$$\mathbf{E} = \frac{H(1) - h_\mu}{1 - 2^{-\gamma}}$$

$$\gamma \approx 0.30$$

Memory in Processes ...

Examples of Excess Entropy:

Infinitary Processes:

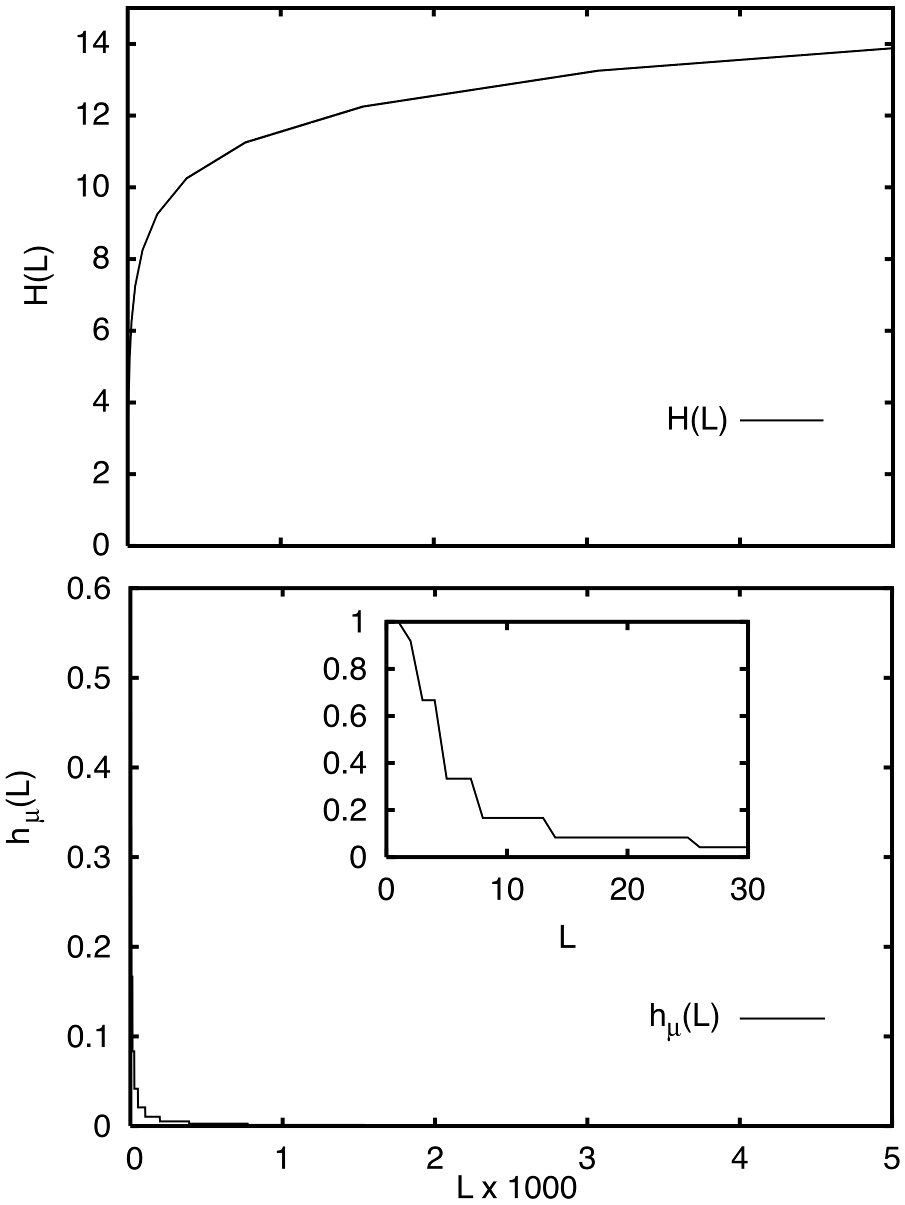
$$\mathbf{E} \rightarrow \infty$$

Excess entropy can diverge:

- Slow entropy convergence
- Long-range correlations
  - (e.g., at phase transitions)

Morse-Thue Process:

- A context-free language
- From Logistic map at onset of chaos

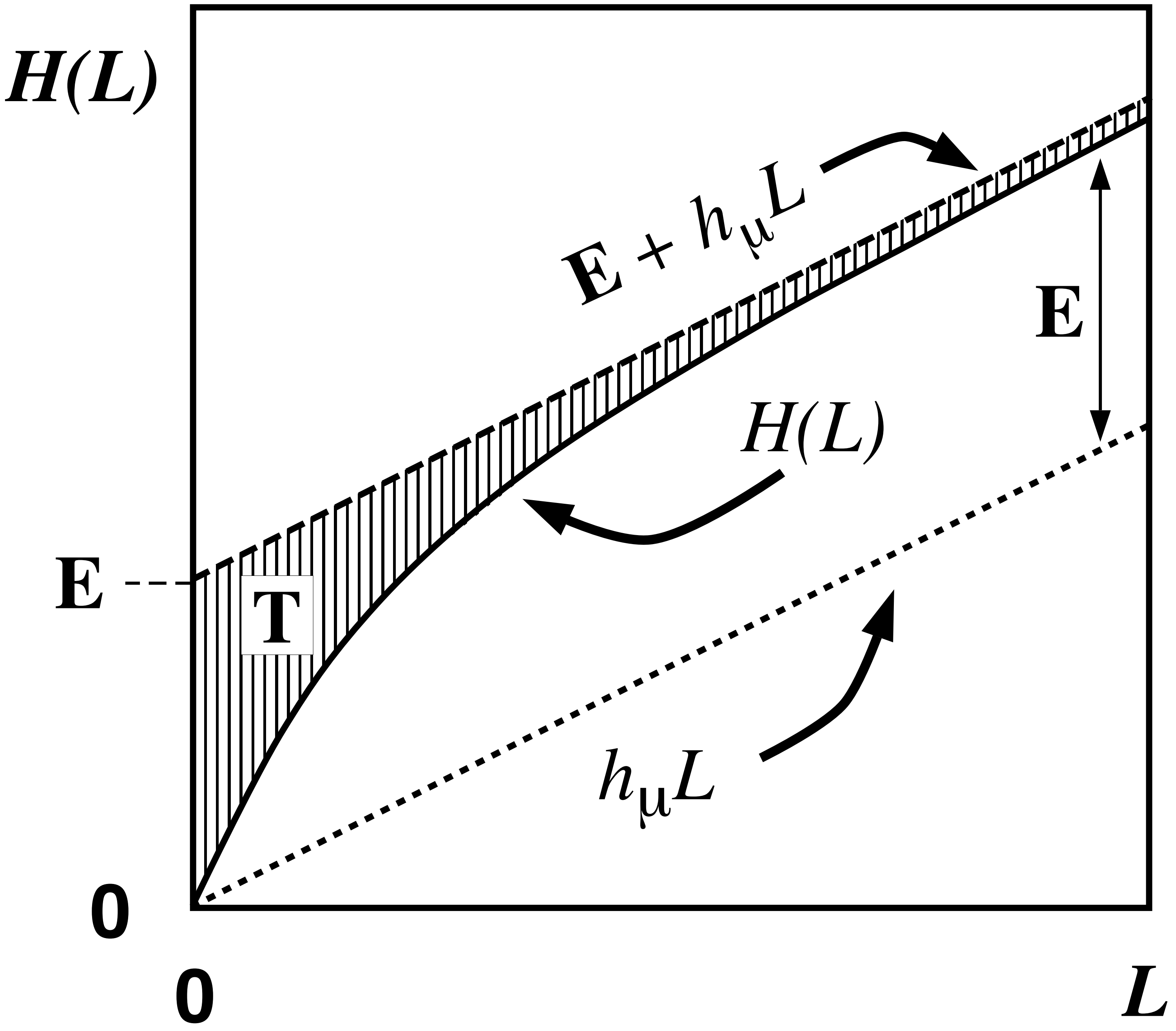


$$h_\mu = 0 \text{ bits per symbol}$$

Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Memory in Processes ...

Information-Entropy Roadmap for a Stochastic Process:



Complexity Lecture 1: Processes and information (CSSS 2013) Jim Crutchfield

Memory in Processes ...

What is information?

Depends on the question!

Uncertainty, surprise, randomness, ...
Compressibility.
Transmission rate.
Memory, apparent stored information, ...
Synchronization.
...

Complexity

Thursday: Information Theory for Complex Systems

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

Friday: Intrinsic Computation

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

See online course:

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