The Complexity of Simplicity

James P Crutchfield
Complexity Sciences Center
Physics Department
University of California, Davis

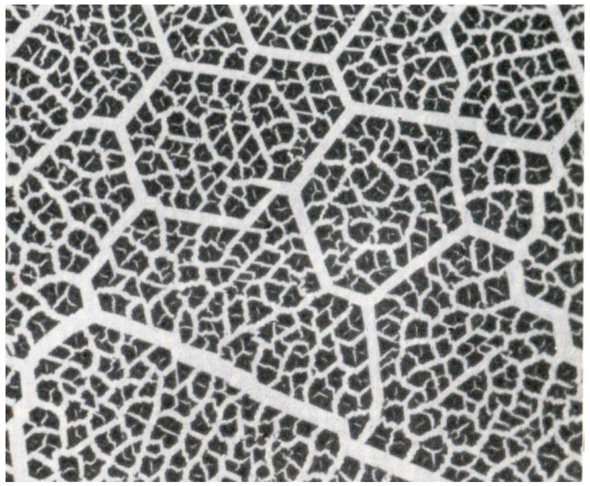
Workshop on the
Interdisciplinary Frontiers of
Quantum and Complexity Science
8–12 January 2017
Singapore

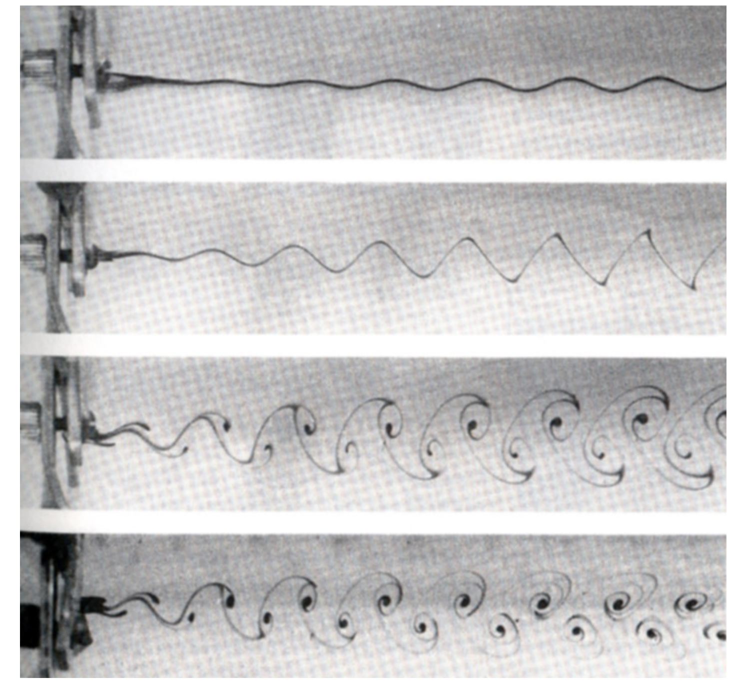
http://csc.ucdavis.edu/~chaos/

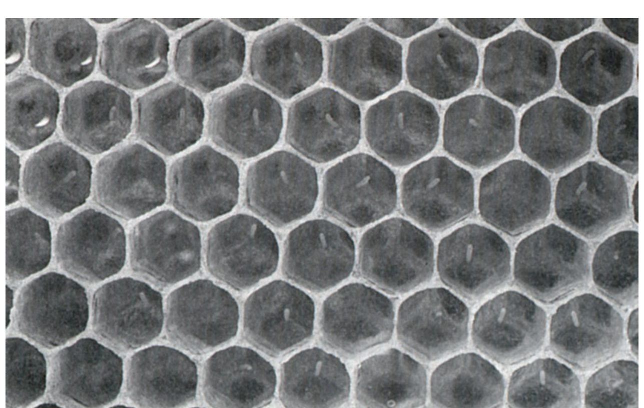
Why Information?

Joint work by 100s

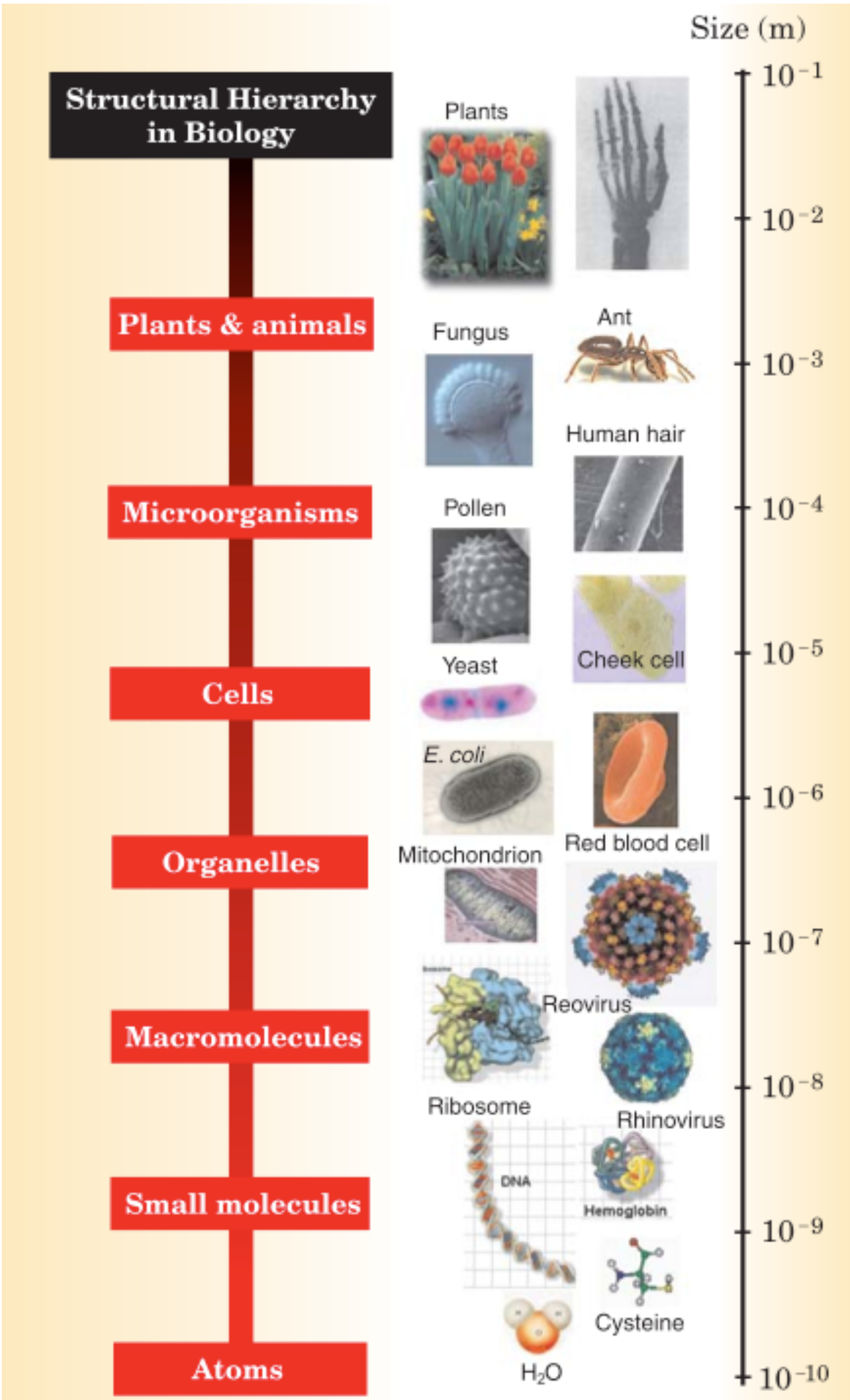
Why Information?







Why Information?



In what ways does nature organize?
(Phenomenology)

How does it organize?
(Mechanism)

Are these levels real or merely convenient?
(Objectivity)

Why does nature organize?
(Optimization versus chance versus ....)

Physics Today, March 2006

Why Information?

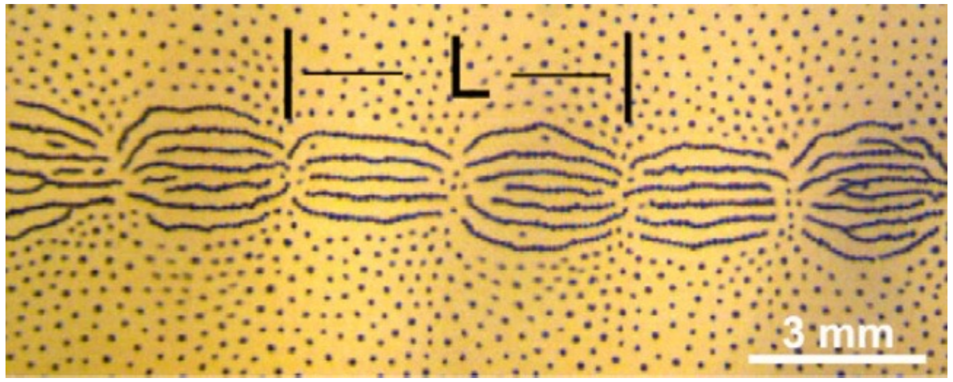
M. M. Hanczyc et al, J. Am. Chem. Soc. **129** (2007) 9386-9391. (Packard)

Why Information?

M. M. Hanczyc et al, J. Am. Chem. Soc. **129** (2007) 9386-9391. (Packard)

Why Information?

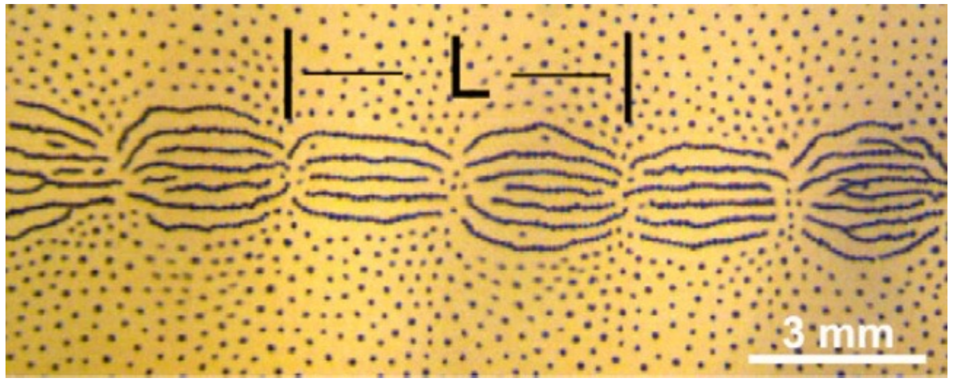
Magnetic 90 μm nickel spheres are suspended over the surface of water being supported by surface tension



A. Snezhko, M. Belkin, I.S. Aranson, and W.-K. Kwok, Phys. Rev. Lett., **102**, 118103 (2009).

Why Information?

Magnetic 90 μm nickel spheres are suspended over the surface of water being supported by surface tension



A. Snezhko, M. Belkin, I.S. Aranson, and W.-K. Kwok, Phys. Rev. Lett., **102**, 118103 (2009).

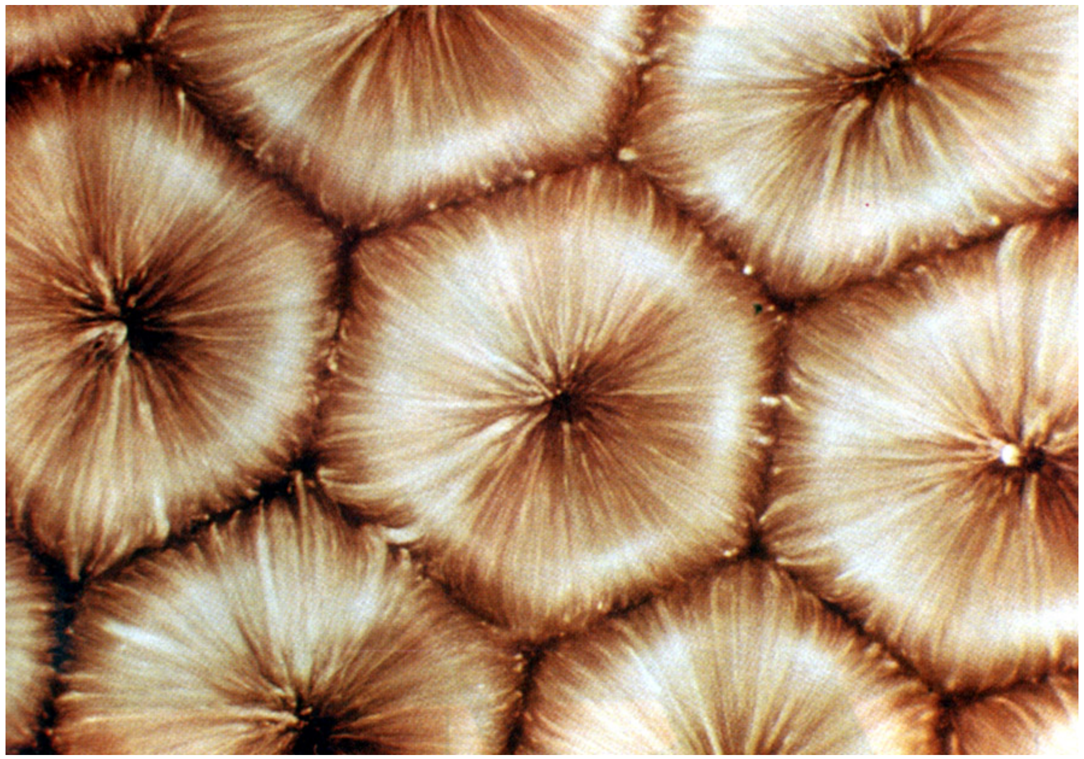
Pattern Discovery!

Seen from the outside, the Amazonian forest seems like a mass of congealed bubbles, a vertical accumulation of green swellings; it is as if some pathological disorder had attacked the riverscape over its whole extent. But once you break through the surface-skin and go inside, everything changes: seen from within, the chaotic mass becomes a monumental universe. The forest ceases to be a terrestrial distemper; it could be taken for a new planetary world, as rich as our world, and replacing it.

As soon as the eye becomes accustomed to recognizing the forest's various closely adjacent planes, and the mind has overcome its first impression of being overwhelmed, a complex system can be perceived.

Claude Levi-Strauss, *Triste Tropiques* (1955).

Pattern Discovery!



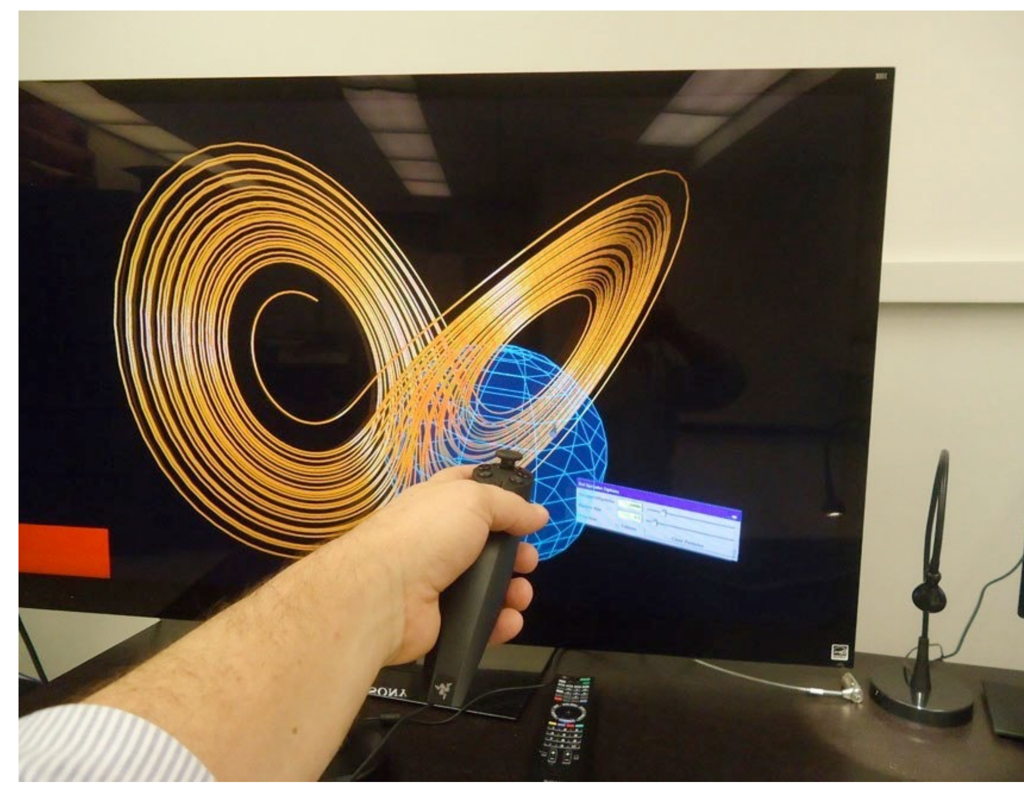
Bénard Convection

VOLUME 45, NUMBER 9 PHYSICAL REVIEW LETTERS 1 SEPTEMBER 1980

Geometry from a Time Series

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw
Dynamical Systems Collective, Physics Department, University of California, Santa Cruz, California 95064
(Received 13 November 1979)

It is shown how the existence of low-dimensional chaotic dynamical systems describing turbulent fluid flow might be determined experimentally. Techniques are outlined for reconstructing phase-space pictures from the observation of a single coordinate of any dissipative dynamical system, and for determining the dimensionality of the system's attractor. These techniques are applied to a well-known simple three-dimensional chaotic dynamical system.



VOLUME 51, NUMBER 16 PHYSICAL REVIEW LETTERS 17 OCTOBER 1983

Low-Dimensional Chaos in a Hydrodynamic System

A. Brandstäter, J. Swift, Harry L. Swinney, and A. Wolf
Department of Physics, University of Texas, Austin, Texas 78712

and

J. Doyne Farmer and Erica Jen
Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

and

P. J. Crutchfield
Physics Department, University of California, Berkeley, Berkeley, California 94720
(Received 21 July 1983)

Evidence is presented for low-dimensional strange attractors in Couette-Taylor flow data. Computations of the largest Lyapunov exponent and metric entropy show that the system displays sensitive dependence on initial conditions. Although the phase space is very high dimensional, analysis of experimental data shows that motion is restricted to an attractor of dimension 5 for Reynolds numbers up to 30% above the onset of chaos. The Lyapunov exponent, entropy, and dimension all generally increase with Reynolds number.

Why Information?

$$X \sim \Pr(X),\ Y \sim \Pr(Y):$$

$$\mathrm{H}[X] = -\sum_{x \in \mathcal{X}} \Pr(x) \log_2 \Pr(x)$$

$$\mathrm{H}[X,Y] = -\sum_{(x,y) \in (\mathcal{X},\mathcal{Y})} \Pr(x,y) \log_2 \Pr(x,y)$$

$$\mathrm{I}[X:Y] = \mathrm{H}[X] + \mathrm{H}[Y] - \mathrm{H}[X,Y]$$

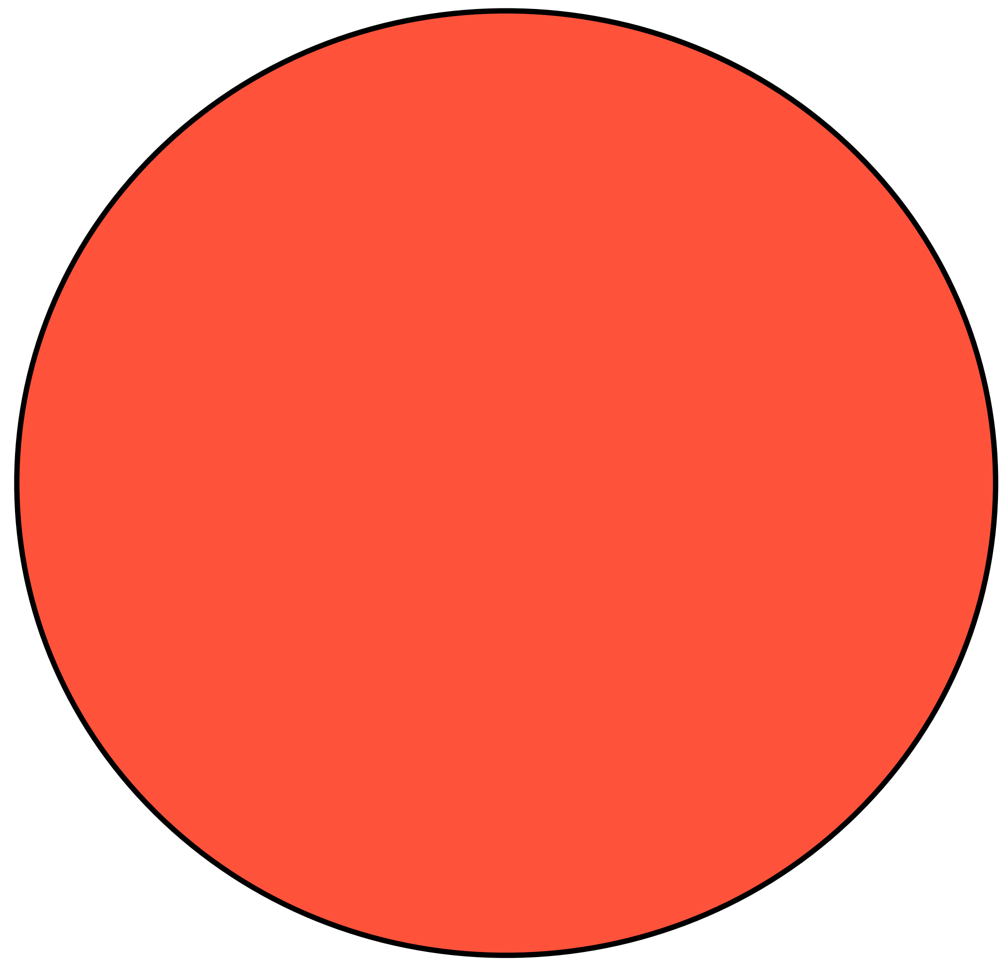
1. Accounts for any type of co-relation
   - Statistical correlation ~ linear only
   - Information measures nonlinear correlation
2. Broadly applicable:
   - Many systems don’t have “energy”, physical modeling precluded
   - Information defined: social, biological, engineering, ... systems
3. Comparable 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
   - Architecture of information storage and flows

Anatomy of a Bit

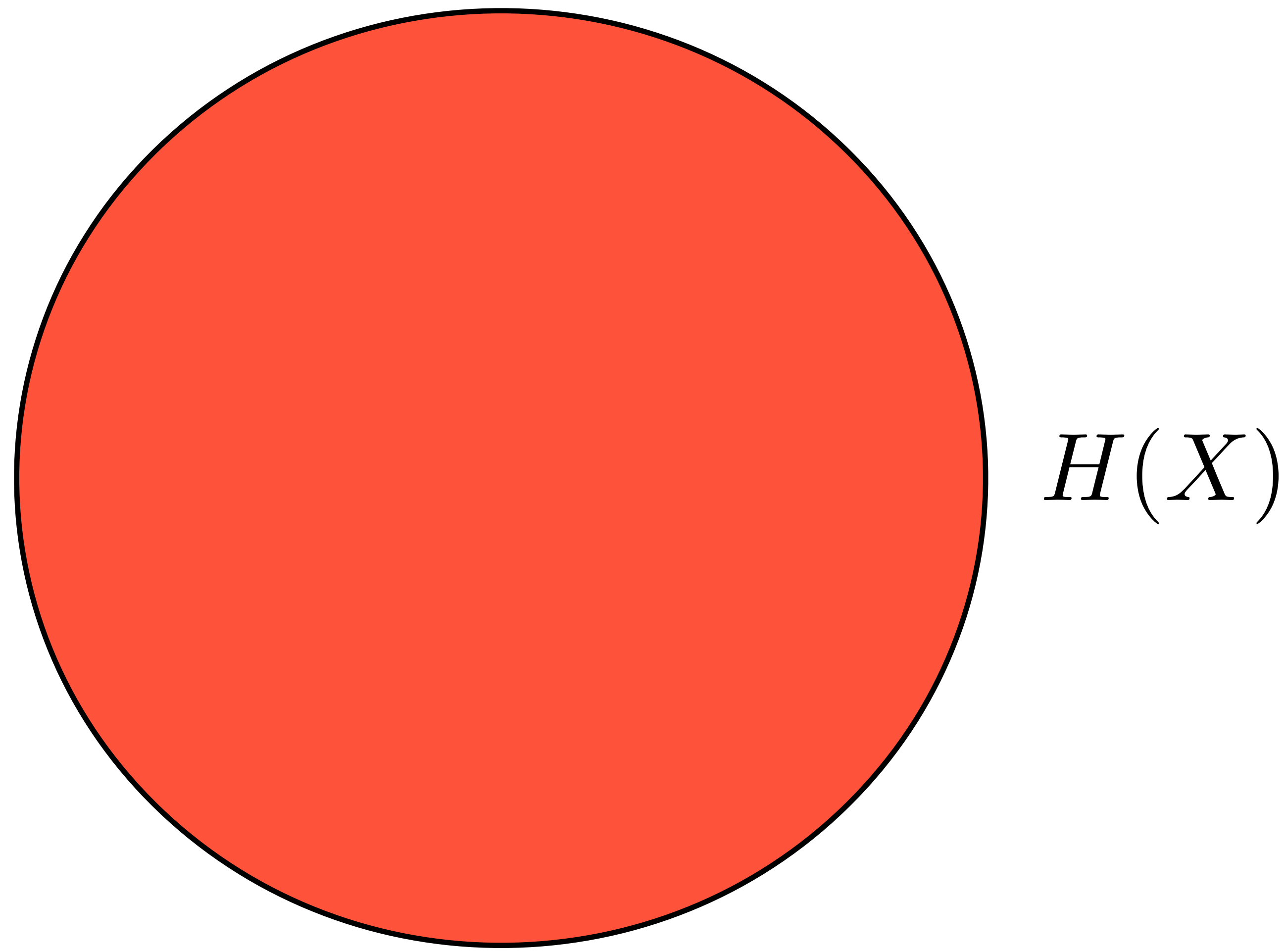
Joint work with Ryan James

Anatomy of a Bit

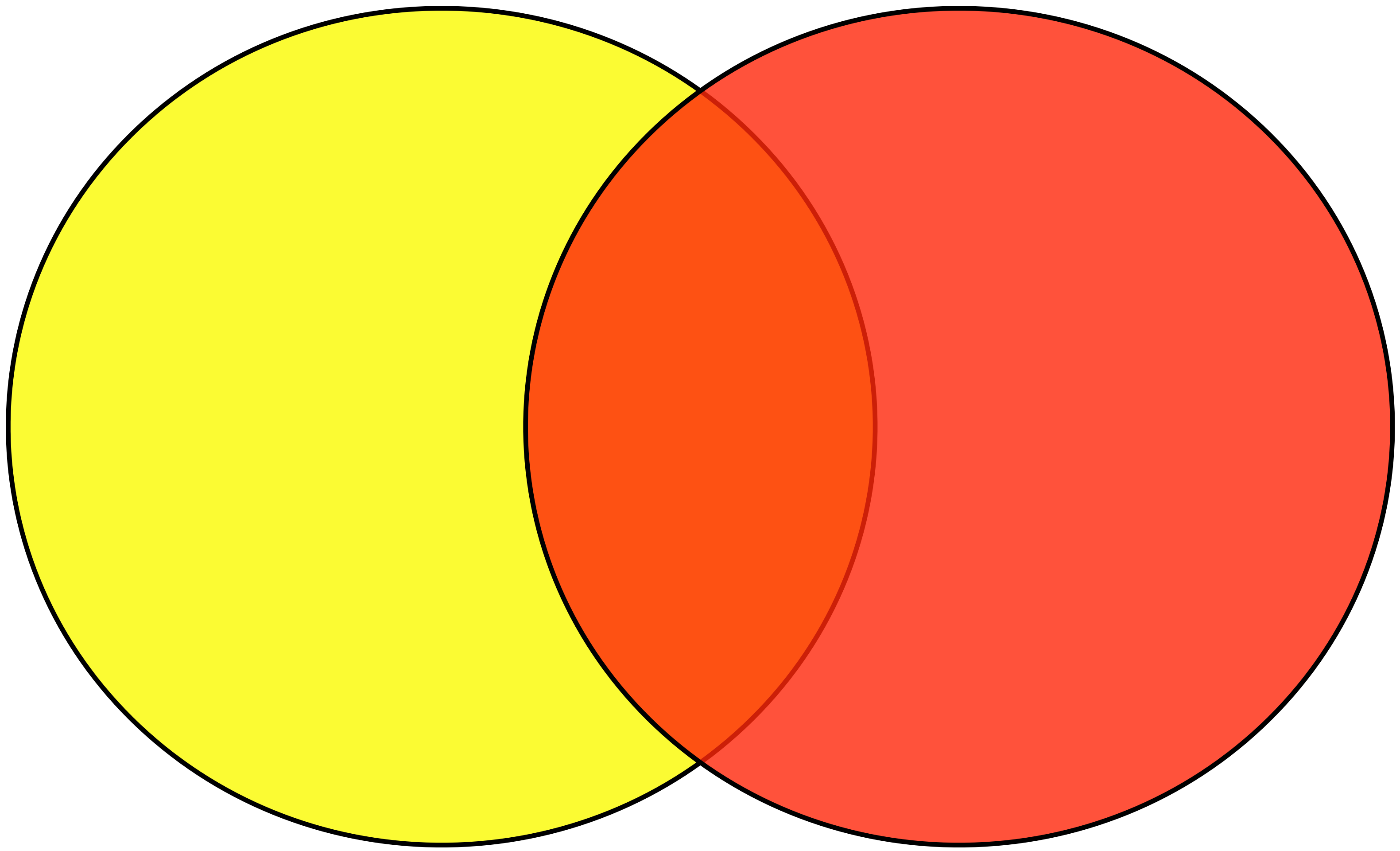
Anatomy of a Bit



Anatomy of a Bit

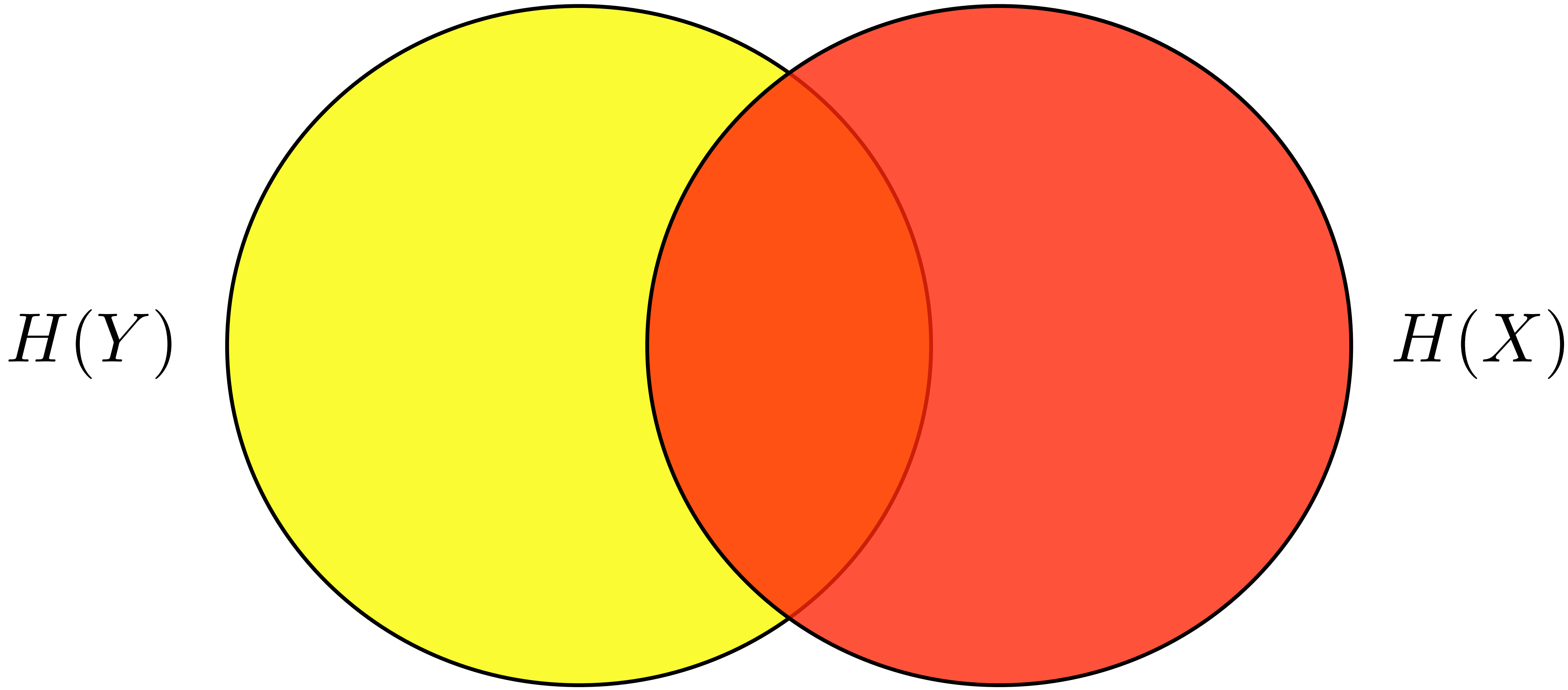


Anatomy of a Bit

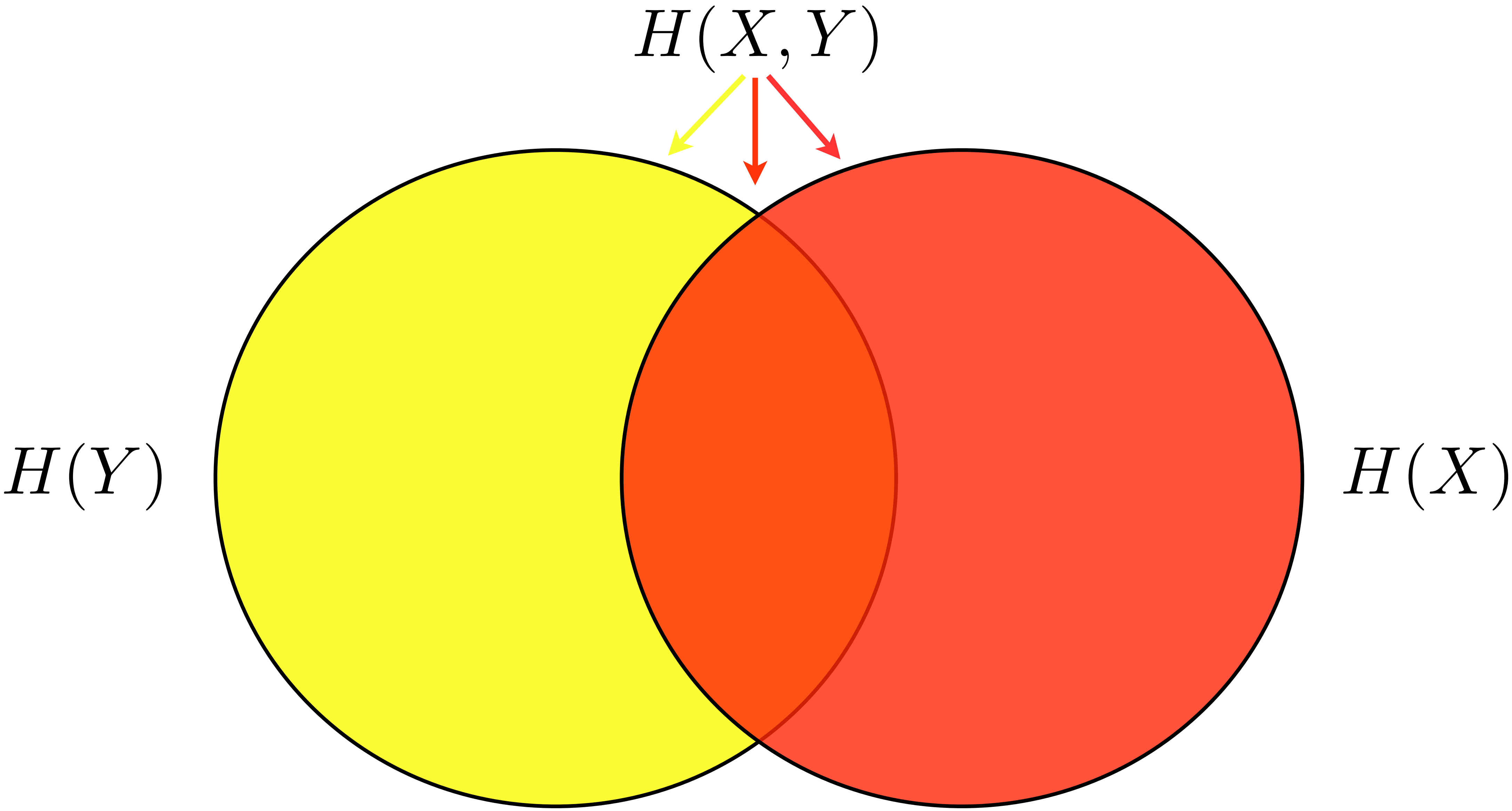


$H(X)$

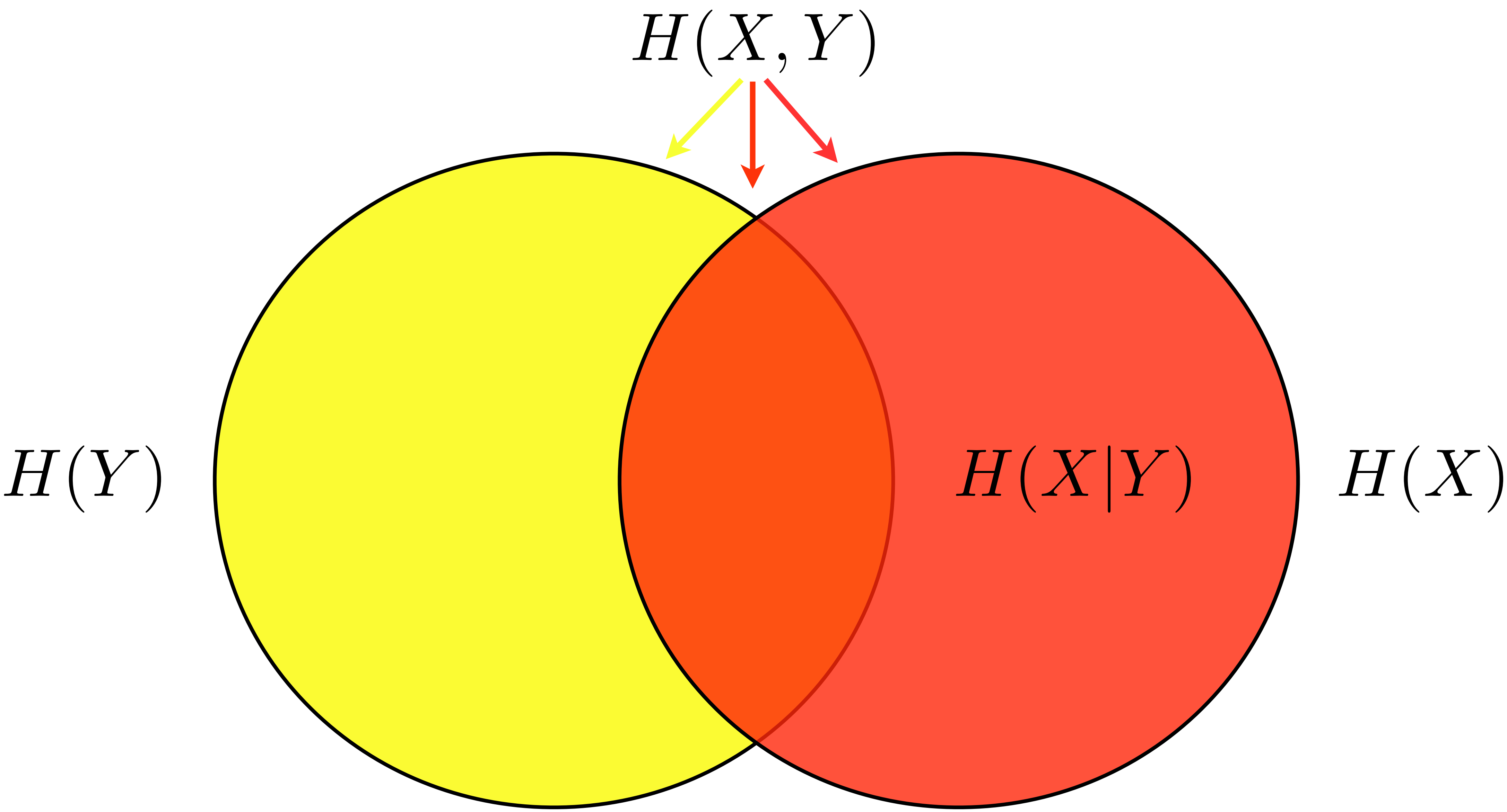
Anatomy of a Bit



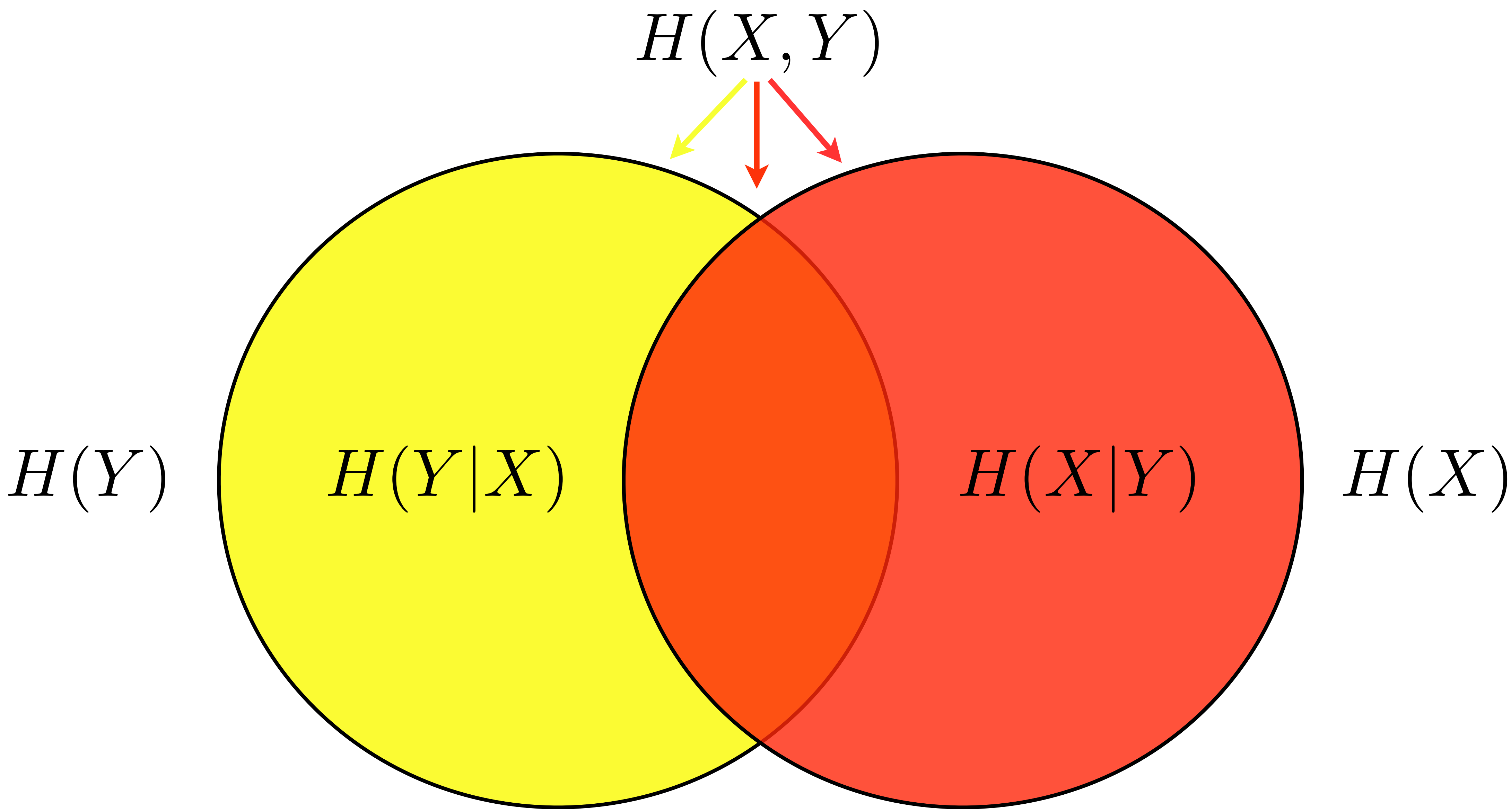
Anatomy of a Bit



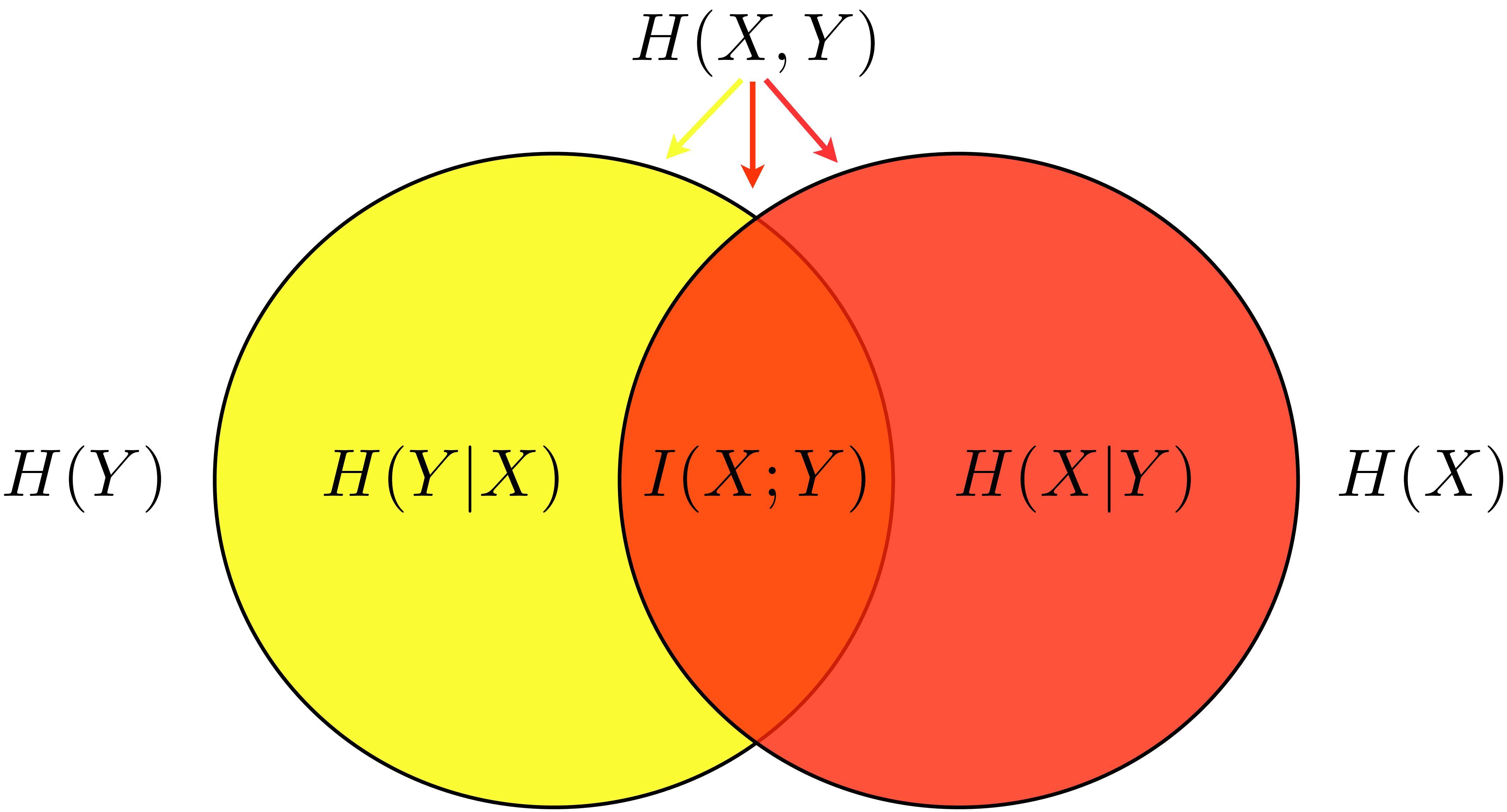
Anatomy of a Bit



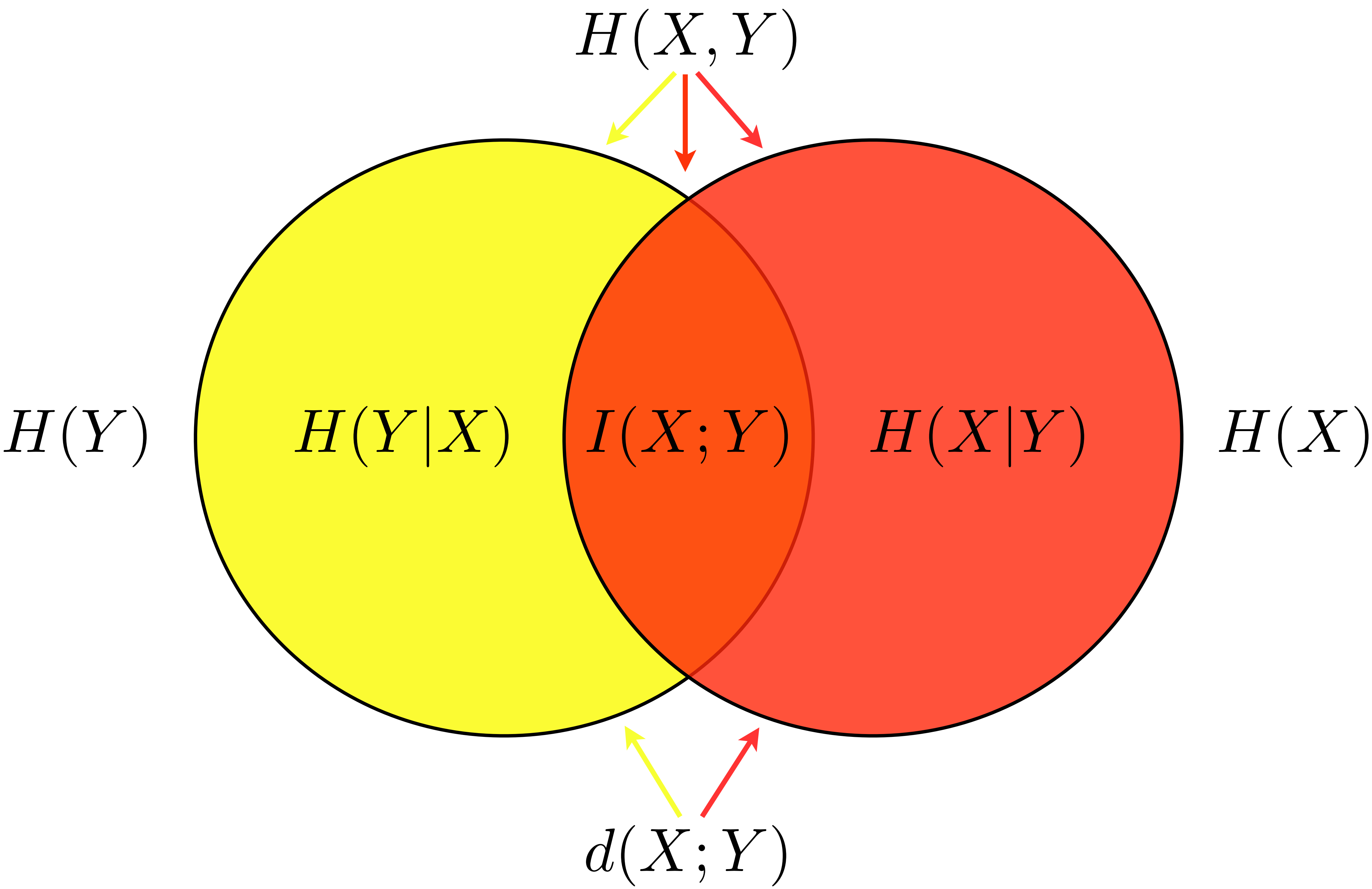
Anatomy of a Bit



Anatomy of a Bit

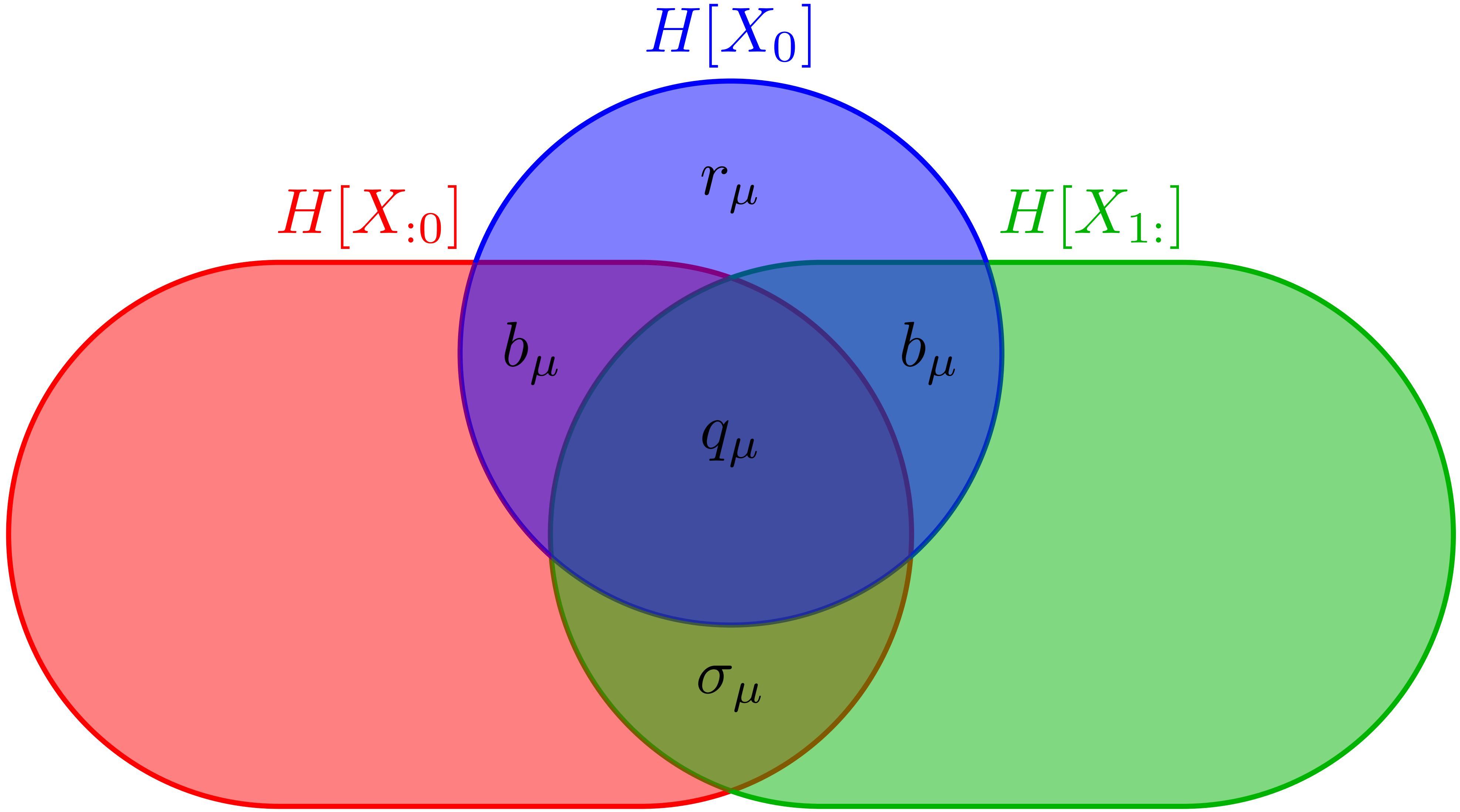


Anatomy of a Bit



Anatomy of a Bit

Process: Past Present Future



Anatomy of a Bit

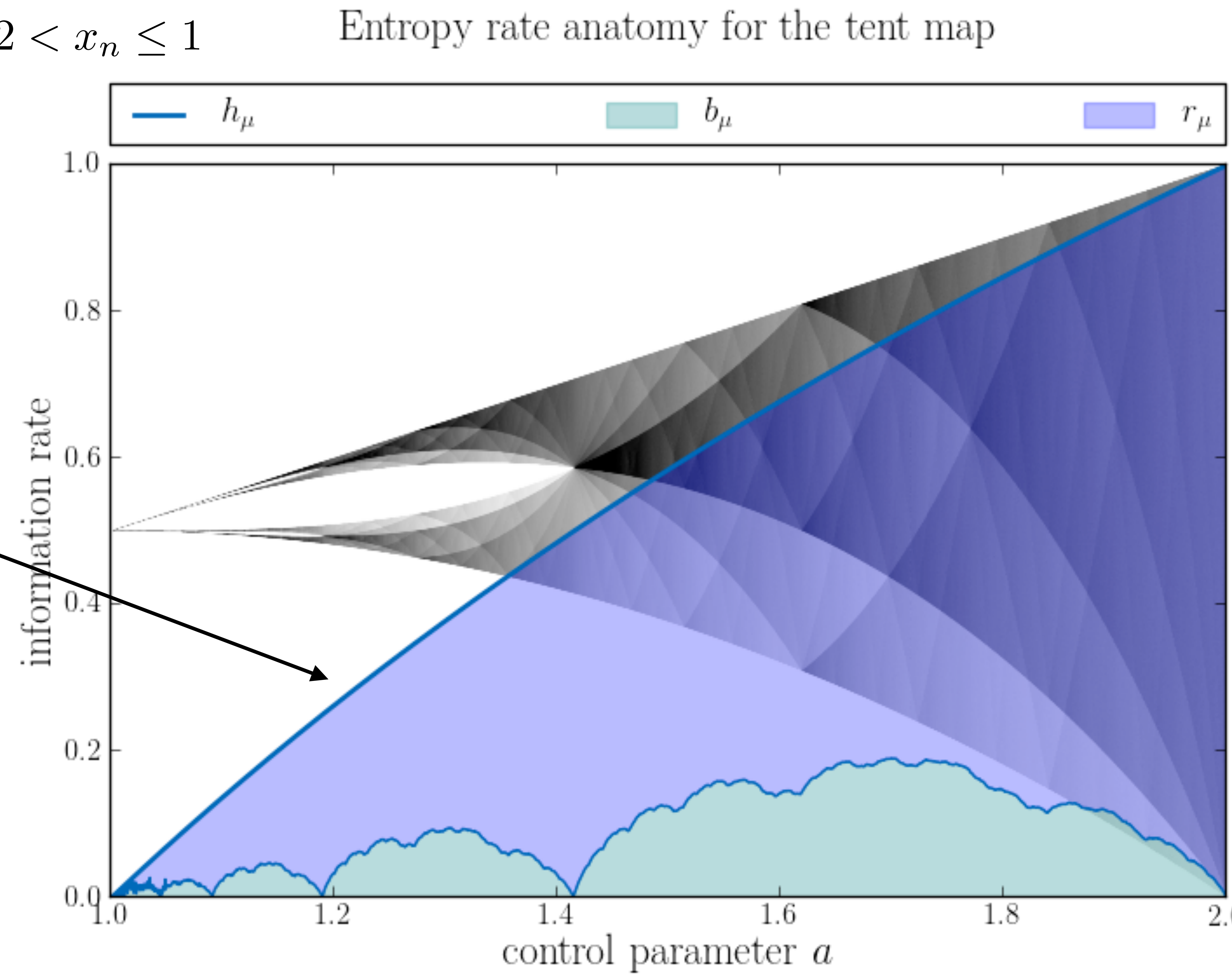
$$x_{n+1} = \begin{cases} a x_n, & 0 \le x_n \le 1/2 \\ a(1 - x_n), & 1/2 < x_n \le 1 \end{cases}$$

$$x_0 \in [0, 1]$$

$$a \in [0, 2]$$

Entropy-rate

$$h_\mu = \log_2 a$$



Shannon Information Measures

Information measures to set-theoretic operations

$$\begin{aligned}
\mu^*(\widetilde{X} \cup \widetilde{Y}) &= H[X, Y] \\
\mu^*(\widetilde{X}) &= H[X] \\
\mu^*(\widetilde{Y}) &= H[Y] \\
\mu^*(\widetilde{X} \cap \widetilde{Y}) &= I[X; Y] \\
\mu^*(\widetilde{X} - \widetilde{Y}) &= H[X|Y] \\
\mu^*(\widetilde{Y} - \widetilde{X}) &= H[Y|X] \\
\mu^*((\widetilde{X} \cap \widetilde{Y})^c) &= H[X|Y] + H[Y|X] \\
\mu^*(\emptyset) &= 0
\end{aligned}$$

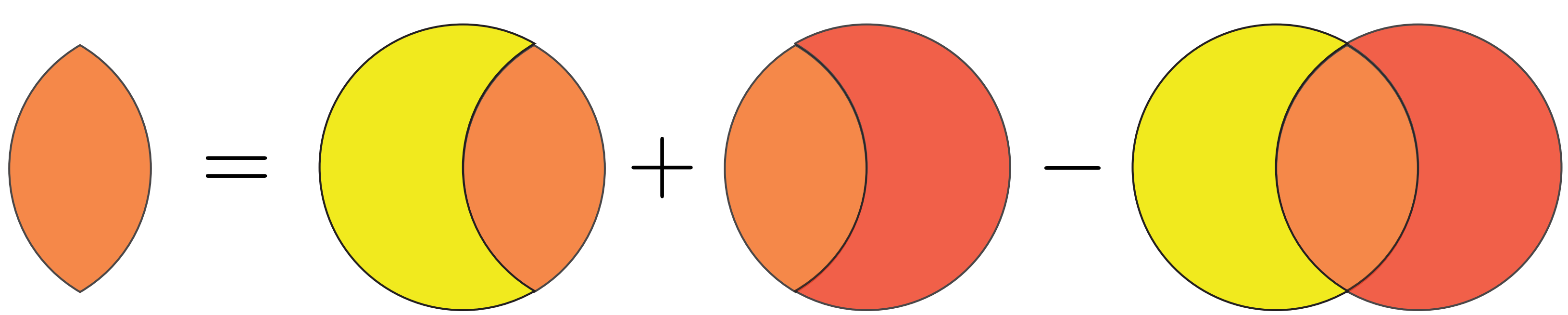
Roadmap:

$$\begin{array}{rcl}
H \text{ and } I & \rightarrow & \mu^* \\
, & \rightarrow & \cup \\
; & \rightarrow & \cap \\
| & \rightarrow & -
\end{array}$$

For example

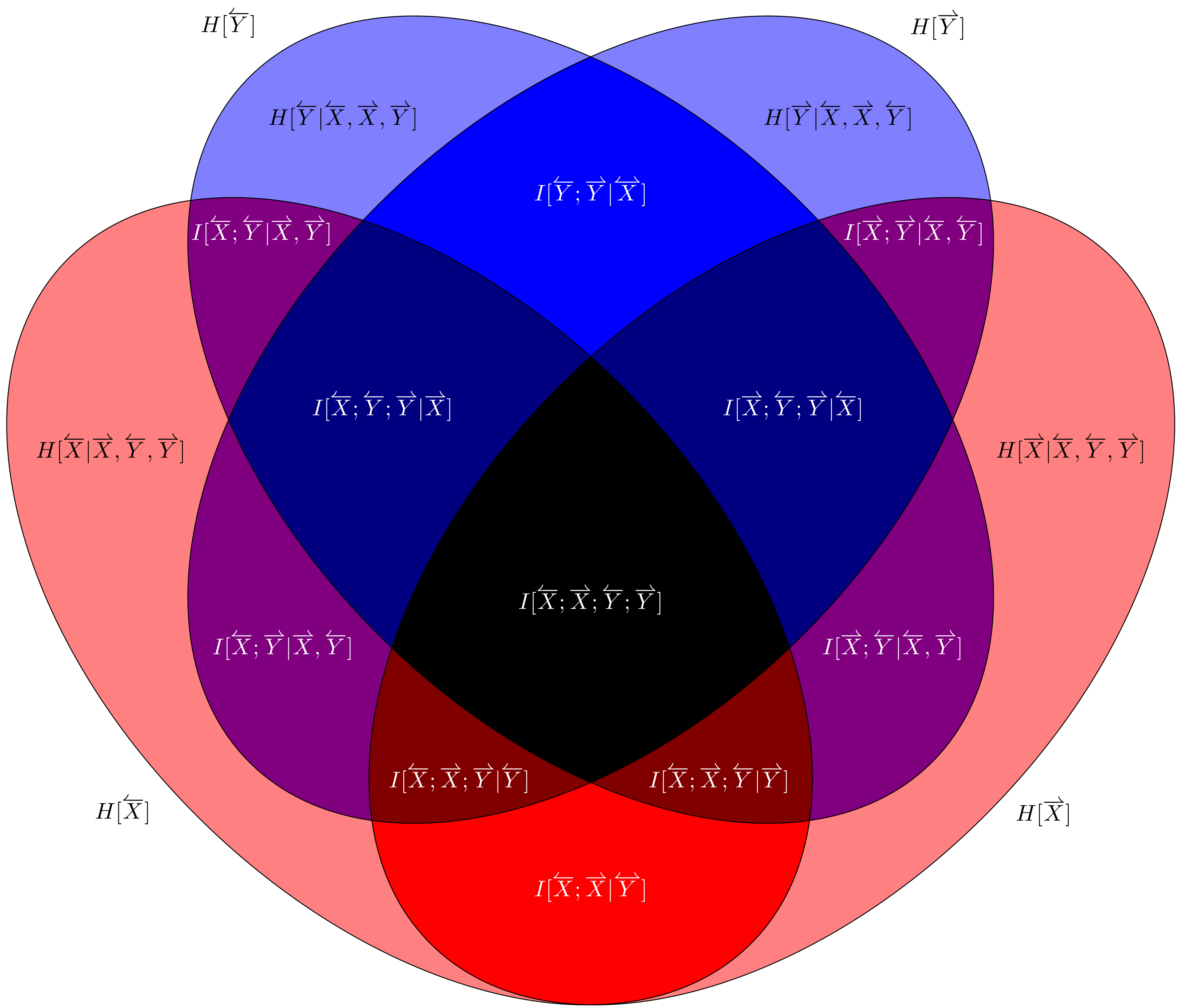
$$I[X; Y] = H[X] + H[Y] - H[X, Y]$$

$$\mu^*(\widetilde{X} \cap \widetilde{Y}) = \mu^*(\widetilde{X}) + \mu^*(\widetilde{Y}) - \mu^*(\widetilde{X} \cup \widetilde{Y})$$



Shannon Information Measures

$$(\overleftarrow{X}, \overrightarrow{X}) \Rightarrow \ \text{Channel} \ \Rightarrow (\overleftarrow{Y}, \overrightarrow{Y})$$



N. Barnett and J. P. Crutchfield, "Computational Mechanics of Input-Output Processes: Structured transformations and the ε-transducer", Journal of Statistical Physics **161**:2 (2015) 404-451.

Information Flows?

Joint work with Ryan James and Nix Barnett

Information Flows?

- Goal: Pattern discovery
  Detect structure & organization in complex systems
  via the internal & external flows of information.

- *Information flow* from process X to process Y:
  Existence of information currently in Y,
  "Cause" of which solely attributed to X's past.

- If information can be solely attributed:
  it is *localized*.

Information Flows?

- *Transfer entropy* (Schreiber, PRL 2000):

$$\mathrm{T}_{X \to Y} = \mathrm{I}[Y_t : X_{0:t} | Y_{0:t}]$$

- How much better one predicts Y_t

  using both $X_{0:t}$ and $Y_{0:t}$ over using $Y_{0:t}$ alone.

Information Flows?

Dyadic Distribution

- Two times series

$$X_t \sim \begin{cases} 0 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases} ,$$

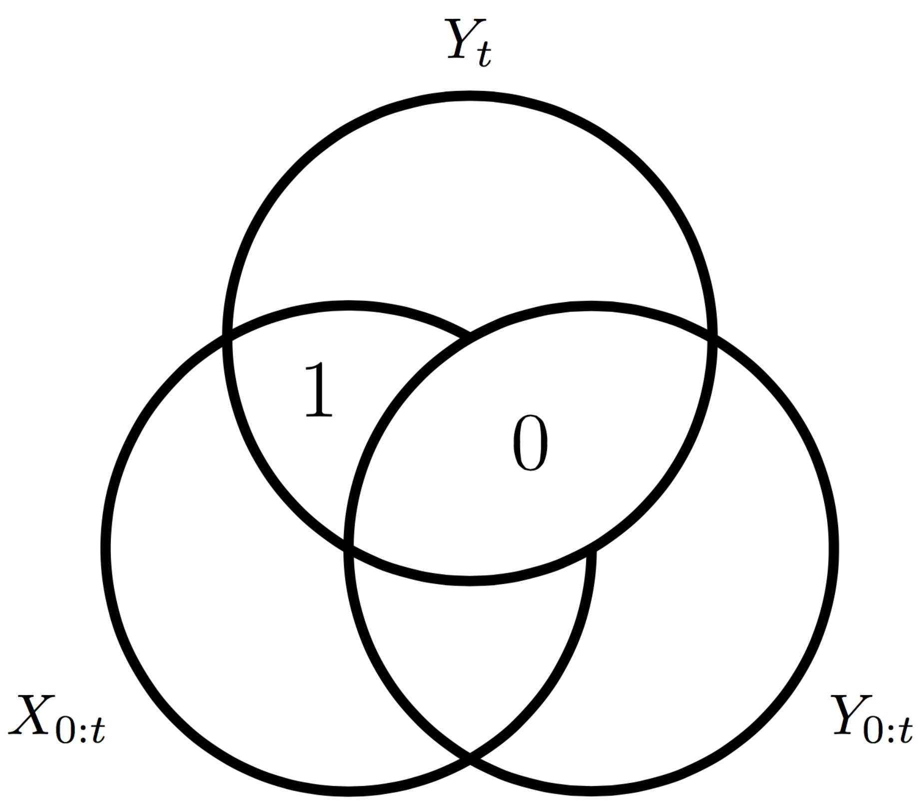
$$Y_0 \sim \begin{cases} 0 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases} , \text{ and}$$

$$Y_t = X_{t-1} \oplus Y_{t-1} \; ;$$

$$\mathrm{T}_{X \to Y} = 1 \text{ bit}$$

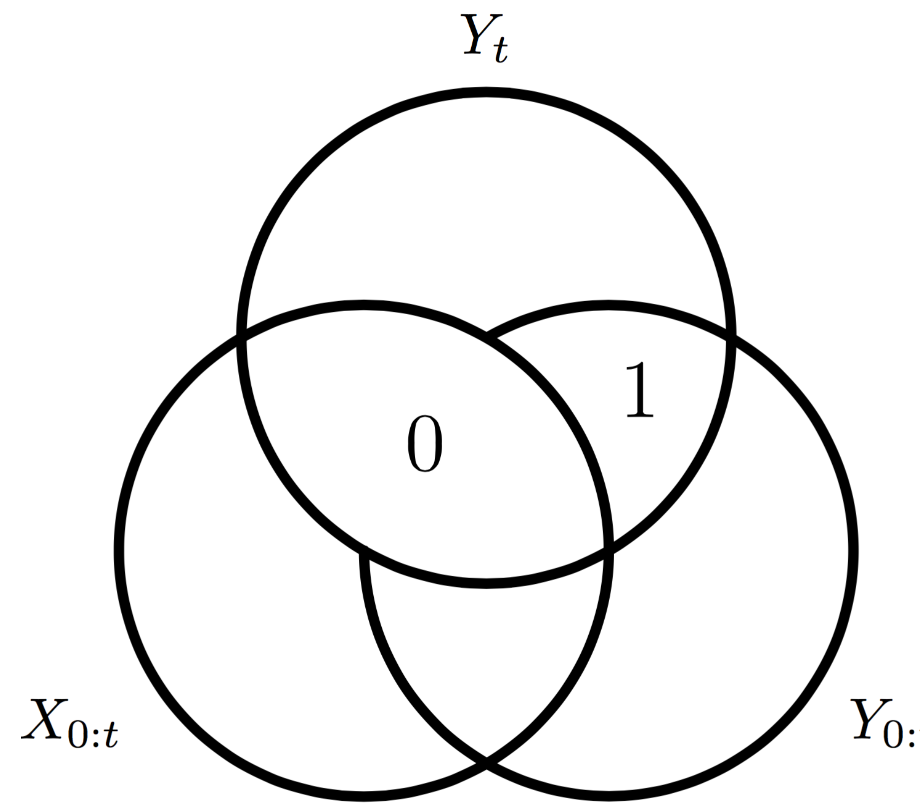
- One bit of information transferred from X to Y at each time?

Information Flows?



$Y_{0:t}$ alone does not predict Y_t. However, $Y_{0:t}$ and $X_{0:t}$ completely predict its value.

$\neq$
?



X's past $X_{0:t}$ alone does not predict Y_t. Given knowledge of $X_{0:t}$, then $Y_{0:t}$ predicts Y_t.

The bit of information about Y_t does not come from *either* time series individually, but rather from *both* of them simultaneously.

The bit of reduction in uncertainty $H[Y_t]$ should not be *localized* to either time series.
Transfer entropy erroneously localizes this information to $X_{0:t}$.
The transfer entropy *overestimates* information flow.
Transfer entropy can be positive due not to information flow,
but rather to nonlocalizable influence—a *conditional dependence* between variables.

Information Flows?

- Three times series

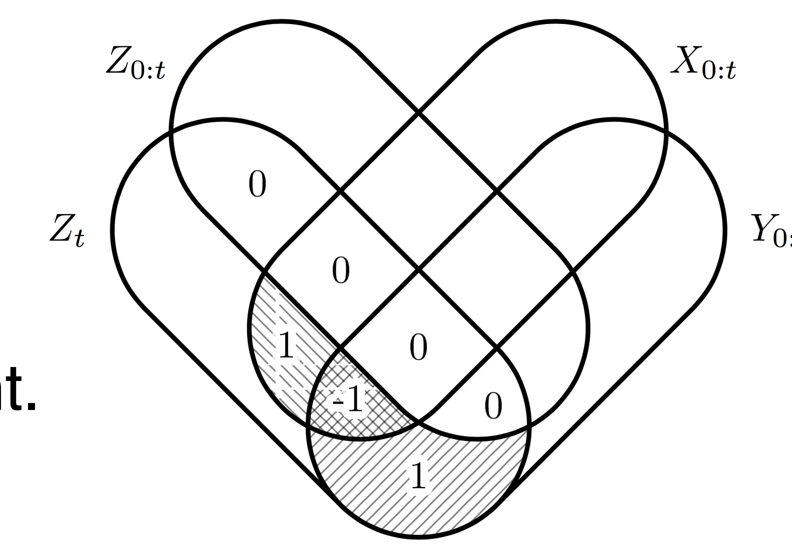
Triadic Distribution

$$X_t \sim \begin{cases} 0 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases} ,$$

$$Y_t \sim \begin{cases} 0 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases} , \text{ and}$$

$$Z_t = X_{t-1} \oplus Y_{t-1} ,$$

$$\mathrm{T}_{X \to Z} = 0 \text{ bit}$$
$$\mathrm{T}_{Y \to Z} = 0 \text{ bit}$$



The transfer entropy *underestimates* influence in the present.

The time series are pairwise independent: $\mathrm{I}[Z_t : X_{0:t}] = 0$

$$\mathrm{I}[Z_t : Y_{0:t}] = 0$$
$$\mathrm{I}[Z_t : Z_{0:t}] = 0$$

Information Flows?

- Misunderstanding of conditional mutual information.
- Conditioning is *not* subtractive.
- $\mathrm{I}[X : Y|Z]$ is information shared by X and Y *taking into account Z*.
- Conditioning can increase information shared between two processes:

$$\mathrm{I}[X : Y] < \mathrm{I}[X : Y|Z]$$

Information Flows?

- Synergy, Uniqueness, and Redundancy:

"Inputs"

$$\mathrm{I}[(X_{0:t}, Y_{0:t}) : Y_t] = R + U_1 + U_2 + S$$

"Output"

$$\mathrm{T}_{X \to Y}$$

Partial Info Decomposition (Williams, Beer 2011)

- Transfer entropy conflates unique & synergistic informations.

Information Flows?

- Do not conflate conditional independence & conditional dependence.
- Do not conflate unique information & synergistic informations.
- Still need a measure of information flow!

Beyond Shannon

Joint work with Ryan James

Beyond Shannon

X	Y	Z	Pr
0	0	0	$1/8$
0	2	1	$1/8$
1	0	2	$1/8$
1	2	3	$1/8$
2	1	0	$1/8$
2	3	1	$1/8$
3	1	2	$1/8$
3	3	3	$1/8$

(a) Dyadic

X	Y	Z	Pr
0	0	0	$1/8$
1	1	1	$1/8$
0	2	2	$1/8$
1	3	3	$1/8$
2	0	2	$1/8$
3	1	3	$1/8$
2	2	0	$1/8$
3	3	1	$1/8$

(b) Triadic

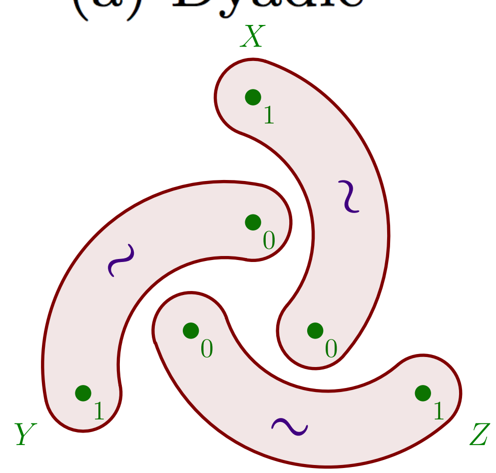
Beyond Shannon

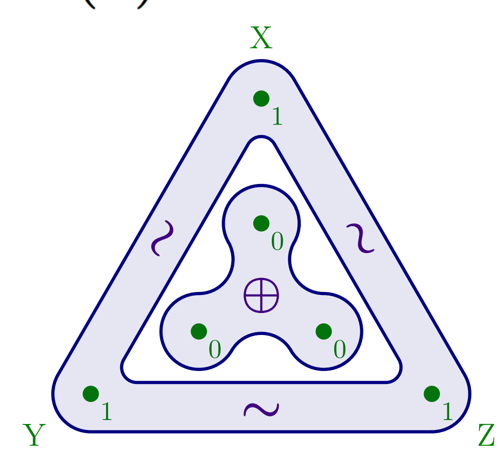
X		Y		Z		
X_0	X_1	Y_0	Y_1	Z_0	Z_1	Pr
0	0	0	0	0	0	$^1/_8$
0	0	1	0	0	1	$^1/_8$
0	1	0	0	1	0	$^1/_8$
0	1	1	0	1	1	$^1/_8$
1	0	0	1	0	0	$^1/_8$
1	0	1	1	0	1	$^1/_8$
1	1	0	1	1	0	$^1/_8$
1	1	1	1	1	1	$^1/_8$

(a) Dyadic

X		Y		Z		
X_0	X_1	Y_0	Y_1	Z_0	Z_1	Pr
0	0	0	0	0	0	$^1/_8$
0	1	0	1	0	1	$^1/_8$
0	0	1	0	1	0	$^1/_8$
0	1	1	1	1	1	$^1/_8$
1	0	0	0	1	0	$^1/_8$
1	1	0	1	1	1	$^1/_8$
1	0	1	0	0	0	$^1/_8$
1	1	1	1	0	1	$^1/_8$

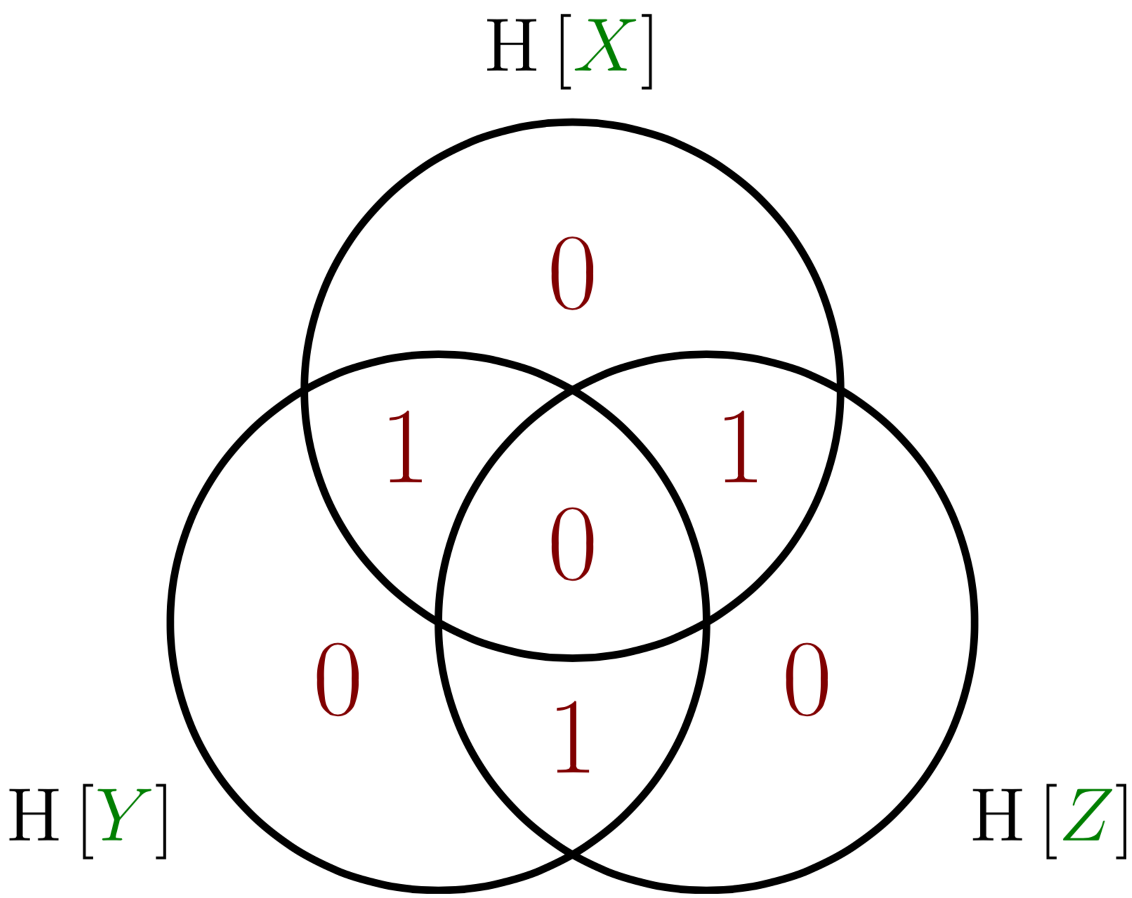
(b) Triadic



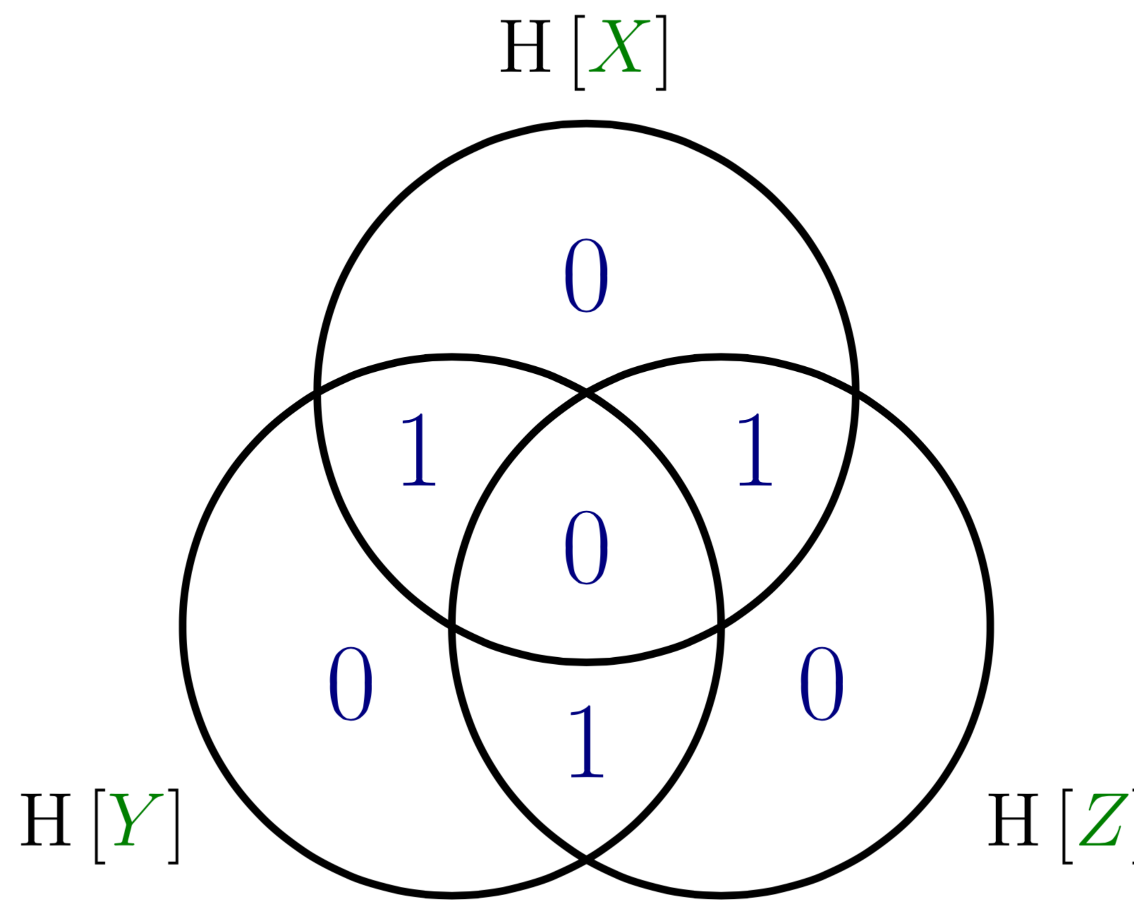


Beyond Shannon

- Shannon Information Diagrams



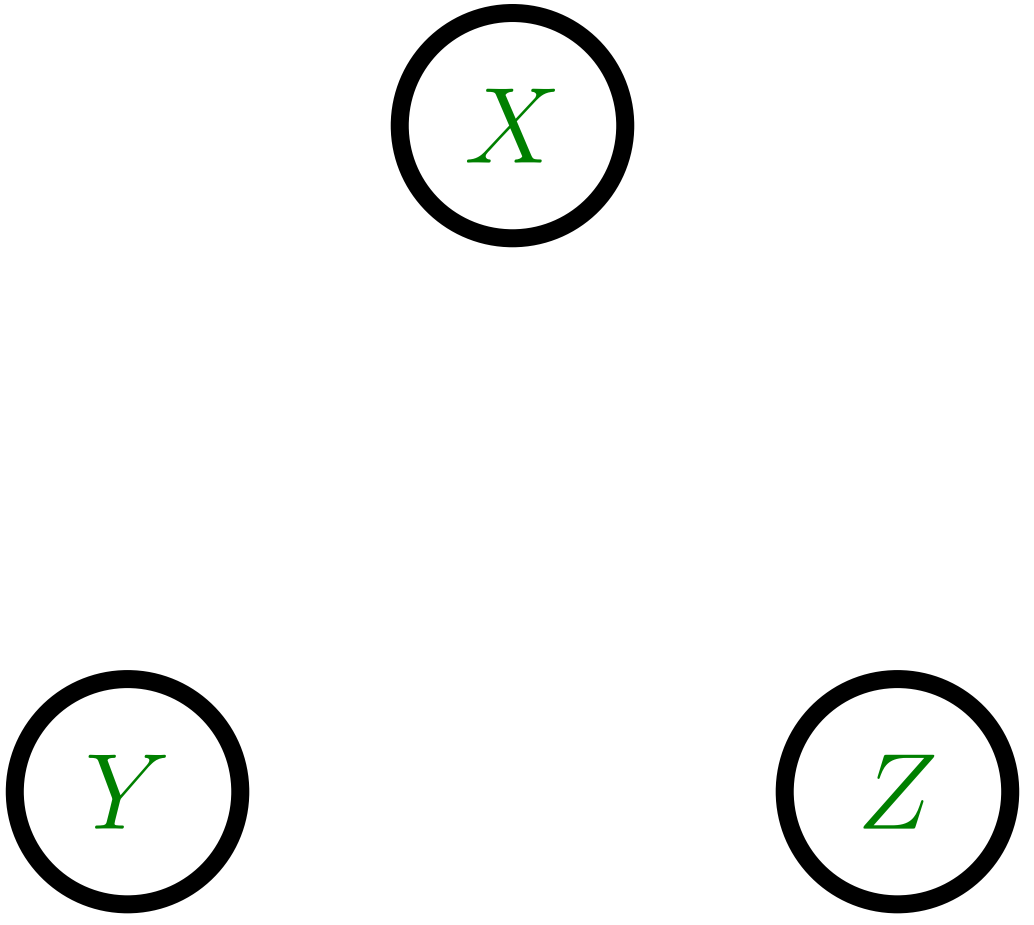
(a) Dyadic



(b) Triadic

Beyond Shannon

- Bayes net inference



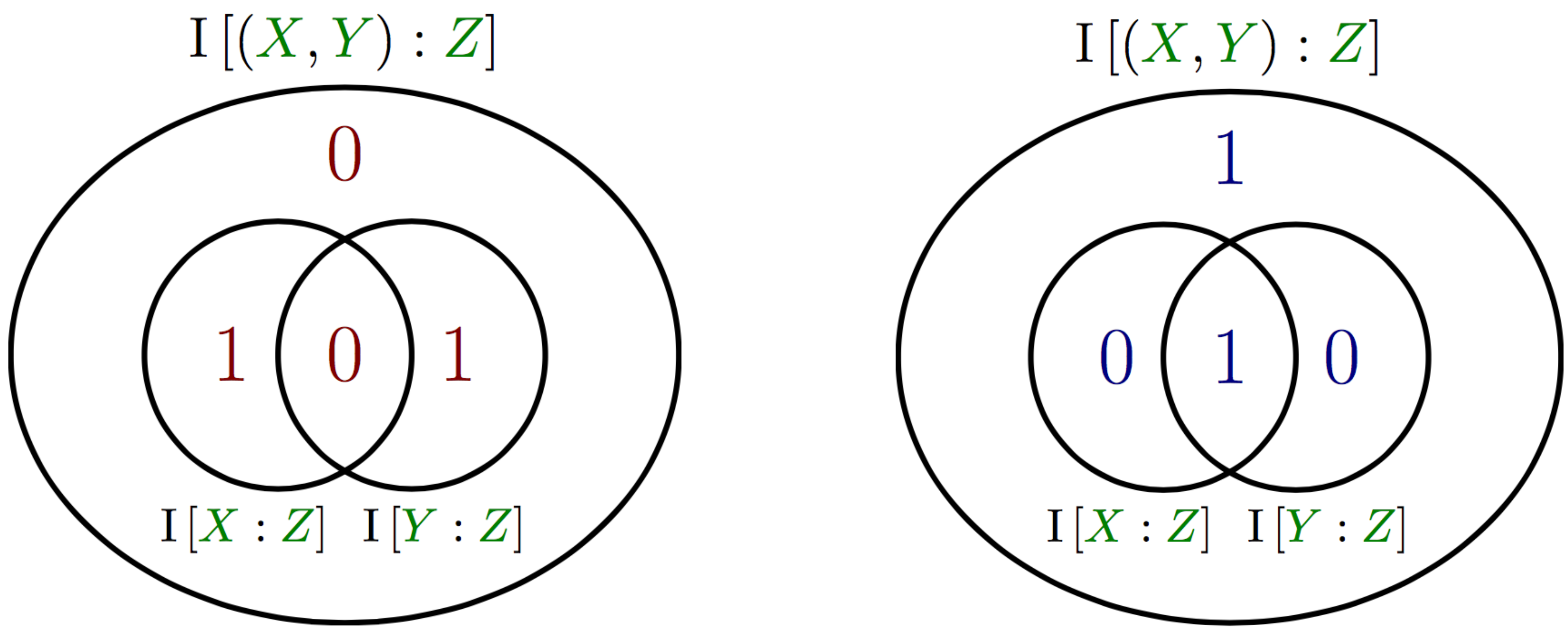
Beyond Shannon

Measures	Dyadic	Triadic
$\mathrm{H}[X, Y, Z]$	3 bit	3 bit
$\mathrm{H}_2[X, Y, Z]$	3 bit	3 bit
$\mathrm{S}_2[X, Y, Z]$	0.875 bit	0.875 bit
$\mathrm{I}[X : Y : Z]$	0 bit	0 bit
$\mathrm{T}[X : Y : Z]$	3 bit	3 bit
$\mathrm{B}[X : Y : Z]$	3 bit	3 bit
$\mathrm{J}[X : Y : Z]$	1.5 bit	1.5 bit
$\mathrm{II}[X : Y : Z]$	0 bit	0 bit
$\mathrm{K}[X : Y : Z]$	0 bit	1 bit
$\mathrm{C}[X : Y : Z]$	3 bit	3 bit
$\mathrm{G}[X : Y : Z]$	3 bit	3 bit
$\mathrm{F}[X : Y : Z]$[a]	3 bit	3 bit
$\mathrm{M}[X : Y : Z]$[b]	3 bit	3 bit
$\mathrm{I}[X : Y \downarrow Z]$[c]	1 bit	0 bit
$\mathrm{I}[X : Y \Downarrow Z]$[cd]	1 bit	0 bit
$\mathrm{X}[X, Y, Z]$	1.349 bit	1.349 bit
$\mathrm{R}[X : Y : Z]$	0 bit	0 bit
$\mathrm{P}[X, Y, Z]$	8	8
$\mathrm{D}[X, Y, Z]$	0.761 bit	0.761 bit
$\mathrm{C}_{\mathrm{LMRP}}[X, Y, Z]$	0.381 bit	0.381 bit
$\mathrm{TSE}[X : Y : Z]$	2 bit	2 bit

TABLE III. Suite of information measures applied to the dyadic and triadic distributions, where: $\mathrm{H}[\bullet]$ is the Shannon entropy [32], $\mathrm{H}_\alpha[\bullet]$ is the Rényi entropy [61], $\mathrm{S}_q[\bullet]$ is the Tsallis entropy [62], $\mathrm{I}[\bullet]$ is the co-information [44], $\mathrm{T}[\bullet]$ is the total correlation [47], $\mathrm{B}[\bullet]$ is the dual total correlation [48, 63], $\mathrm{J}[\bullet]$ is the CAEKL mutual information [49], $\mathrm{II}[\bullet]$ is the interaction information [64], $\mathrm{K}[\bullet]$ is the Gács-Körner common information [57], $\mathrm{C}[\bullet]$ is the Wyner common information [65, 66], $\mathrm{G}[\bullet]$ is the exact common information [67], $\mathrm{F}[\bullet]$ is the functional common information[a], $\mathrm{M}[\bullet]$ is the MSS common information[b], $\mathrm{I}[\bullet \downarrow \bullet]$ is the intrinsic mutual information [26][c], $\mathrm{I}[\bullet \Downarrow \bullet]$ is the reduced intrinsic mutual information [27][cd], $\mathrm{X}[\bullet]$ is the extropy [68], $\mathrm{R}[\bullet]$ is the residual entropy or erasure entropy [60, 63], $\mathrm{P}[\bullet]$ is the perplexity [69], $\mathrm{D}[\bullet]$ is the disequilibrium [51], $\mathrm{C}_{\mathrm{LMRP}}[\bullet]$ is the LMRP complexity [51], and $\mathrm{TSE}[\bullet]$ is the TSE complexity [59]. Only the Gács-Körner common information and the intrinsic mutual informations, highlighted, are able to distinguish the two distributions; the Gács-Körner common information via the construction of a subvariable ($X_1 = Y_1 = Z_1$) common to X, Y, and Z, and the intrinsic mutual informations via the relationship $X_0 = Y_1$ being independent of Z.

Beyond Shannon

- Partial Information Decomposition



(a) Dyadic (b) Triadic

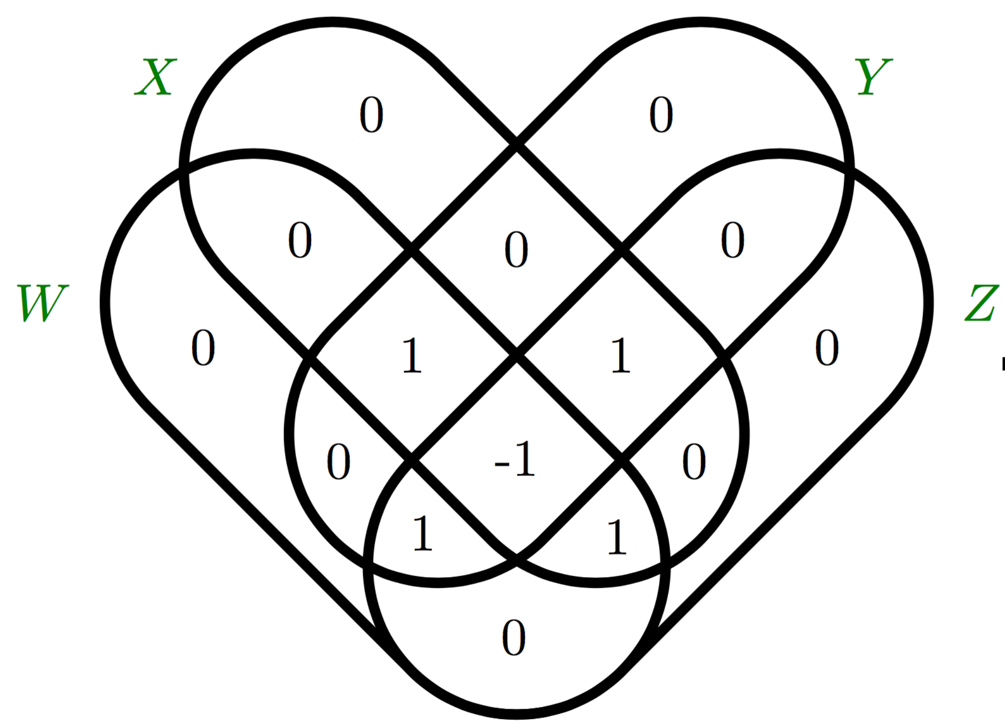
Beyond Shannon

- Arbitrary, higher dimensions, too!

- Dyadic Camouflage Distribution

W	X	Y	Z	Pr
0	0	0	0	$1/8$
0	1	3	1	$1/8$
1	0	2	2	$1/8$
1	1	1	3	$1/8$
2	2	3	3	$1/8$
2	3	0	2	$1/8$
3	2	1	1	$1/8$
3	3	2	0	$1/8$

(a) Distribution



(b) I-diagram

Only dyadic dependencies.

NO!

- Dependency diffusion:
  Cryptographic embedding NP-Hard to discover.

Beyond Shannon

- Shannon information measures cannot capture even simple dependency structures.
- Shannon extensions cannot either.
- Except for several.
  (There are dependency structures for which they fail!)
- Whither information?

Ambiguity of Simplicity

Joint work with Cina Aghamohammadi & John Mahoney

Ambiguity of Simplicity

- *Process*: $X_{-\infty:\infty} = \ldots X_{-2}X_{-1}X_0X_1X_2\ldots$
- *ε-Machine* $\left\{\boldsymbol{\mathcal{S}}, \{T^{(x)} : x \in \mathcal{A}\}, |\eta_0\rangle\right\}$: Minimal, optimal predictor.
- *Causal states*: $\sigma \in \boldsymbol{\mathcal{S}}$

$$\overleftarrow{x} \sim \overleftarrow{x}' \iff \Pr(\overrightarrow{X}|\overleftarrow{x}) = \Pr(\overrightarrow{X}|\overleftarrow{x}')$$

- Process memory: *Statistical complexity*

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

- Process A is ***simpler*** than B: $C_\mu(A) < C_\mu(B)$
- Process is ***emergent***: $C_\mu(X_t) > C_\mu(X_0)$

Crutchfield & Young *Inferring Statistical Complexity*, Physical Review Letters (1989)

Ambiguity of Simplicity

- Alice wants to send Bob information to predict

$$X_{-\infty:\infty} = \dots X_{-2} X_{-1} X_0 X_1 X_2 \dots$$

- How much information must be sent?

- Classical channel?

  Minimal amount is the statistical complexity C_μ.

- Quantum channel (transmits qubits)?

Ambiguity of Simplicity

Gu et al, Nat Physics (2012)
Mahoney et al, Sci Rep & Phys. Rev. A (2016)

- q-Machine signal states:

$$|\eta_j(L)\rangle \equiv \sum_{w^L \in |\mathcal{A}|^L} \sum_{\sigma_k \in \boldsymbol{\mathcal{S}}} \sqrt{\Pr(w^L, \sigma_k | \sigma_j)} \; |w^L\rangle \, |\sigma_k\rangle$$

- Density matrix:

$$\rho = \sum_i \pi_i \, |\eta_i\rangle \, \langle \eta_i|$$

- Quantum memory: von Neumann entropy

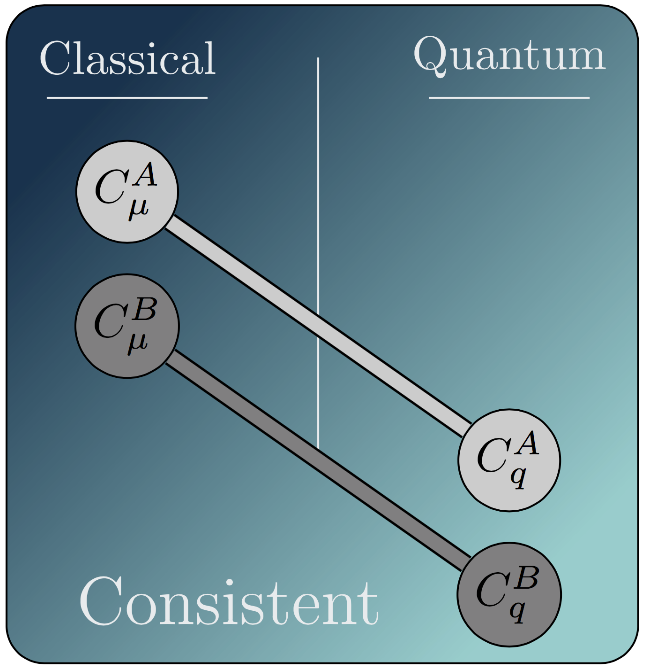
$$C_q = -\mathrm{Tr} \rho \log \rho$$

- Quantum smaller than classical model:

$$C_q \leq C_\mu$$

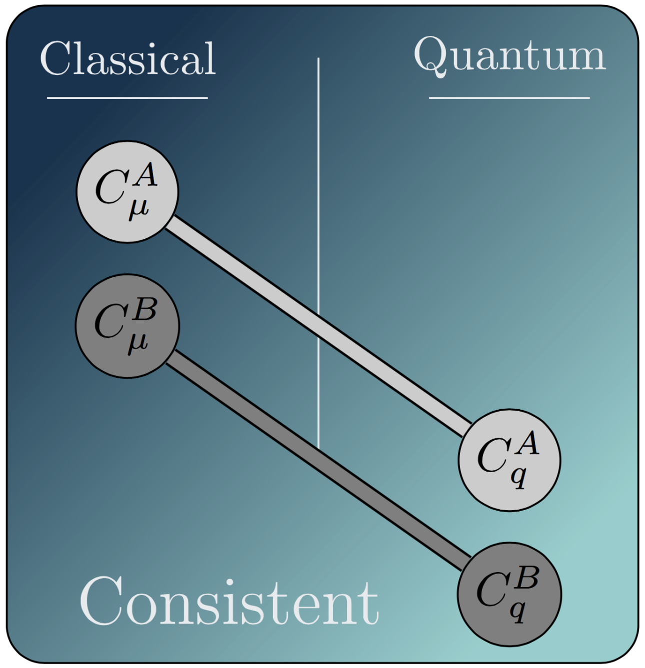
Bibliography: At end; also see Mile Gu, John Mahoney, and others' talks.

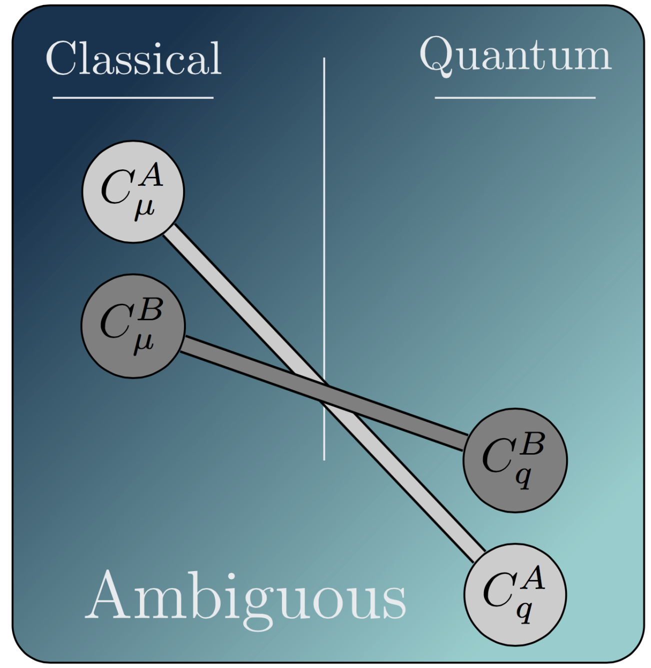
Ambiguity of Simplicity



Aghamohammadi et al
arXiv:1609.03650

Ambiguity of Simplicity

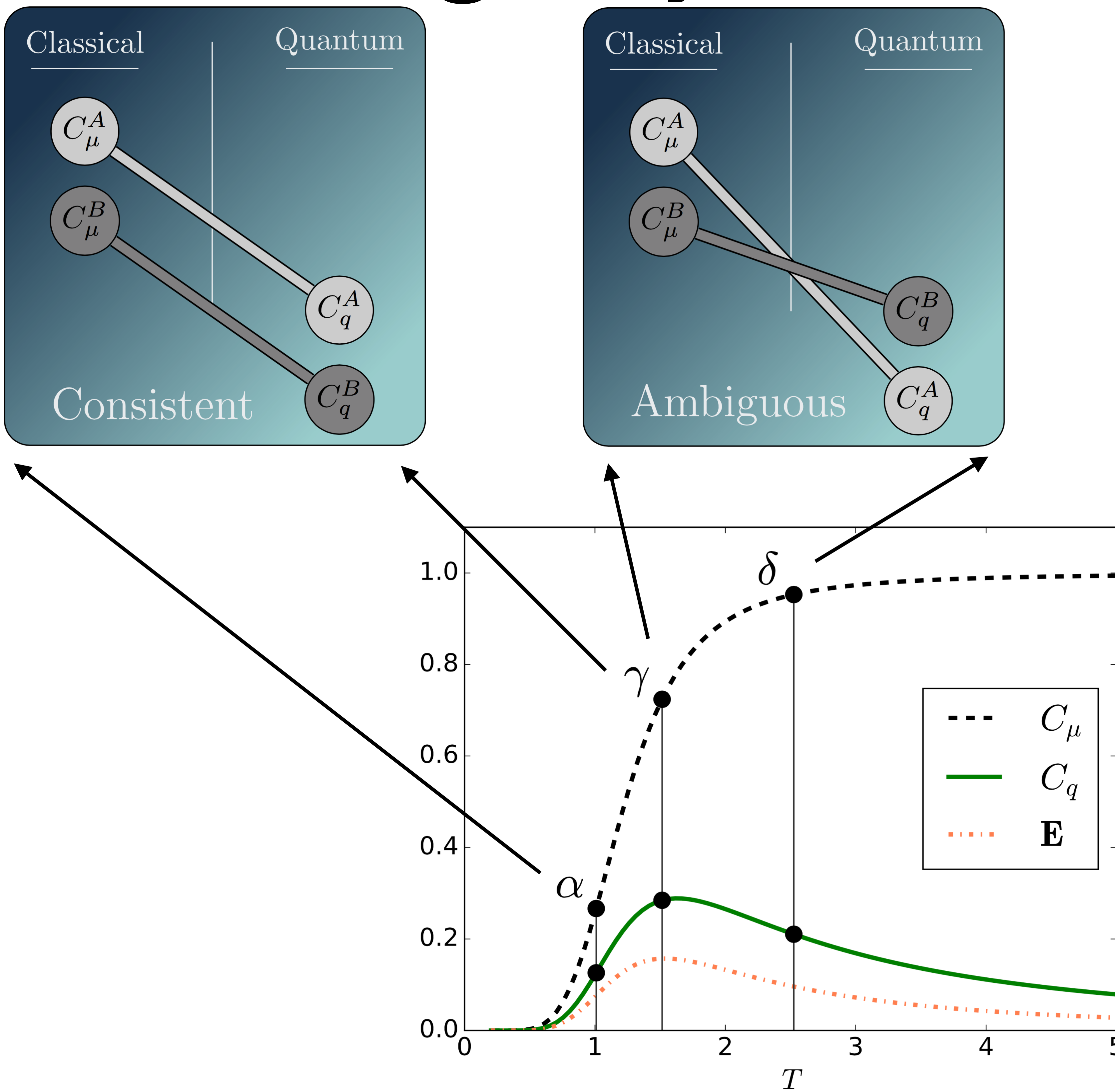


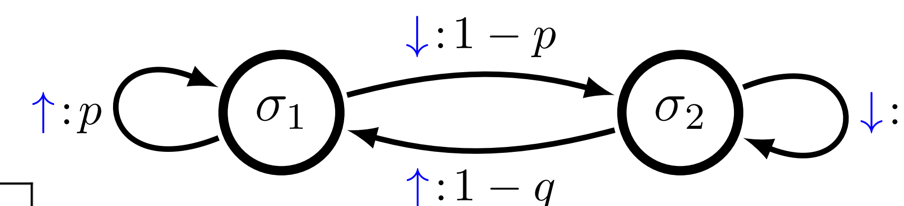


Aghamohammadi et al
arXiv:1609.03650

Ambiguity of Simplicity

Aghamohammadi et al
arXiv:1609.03650





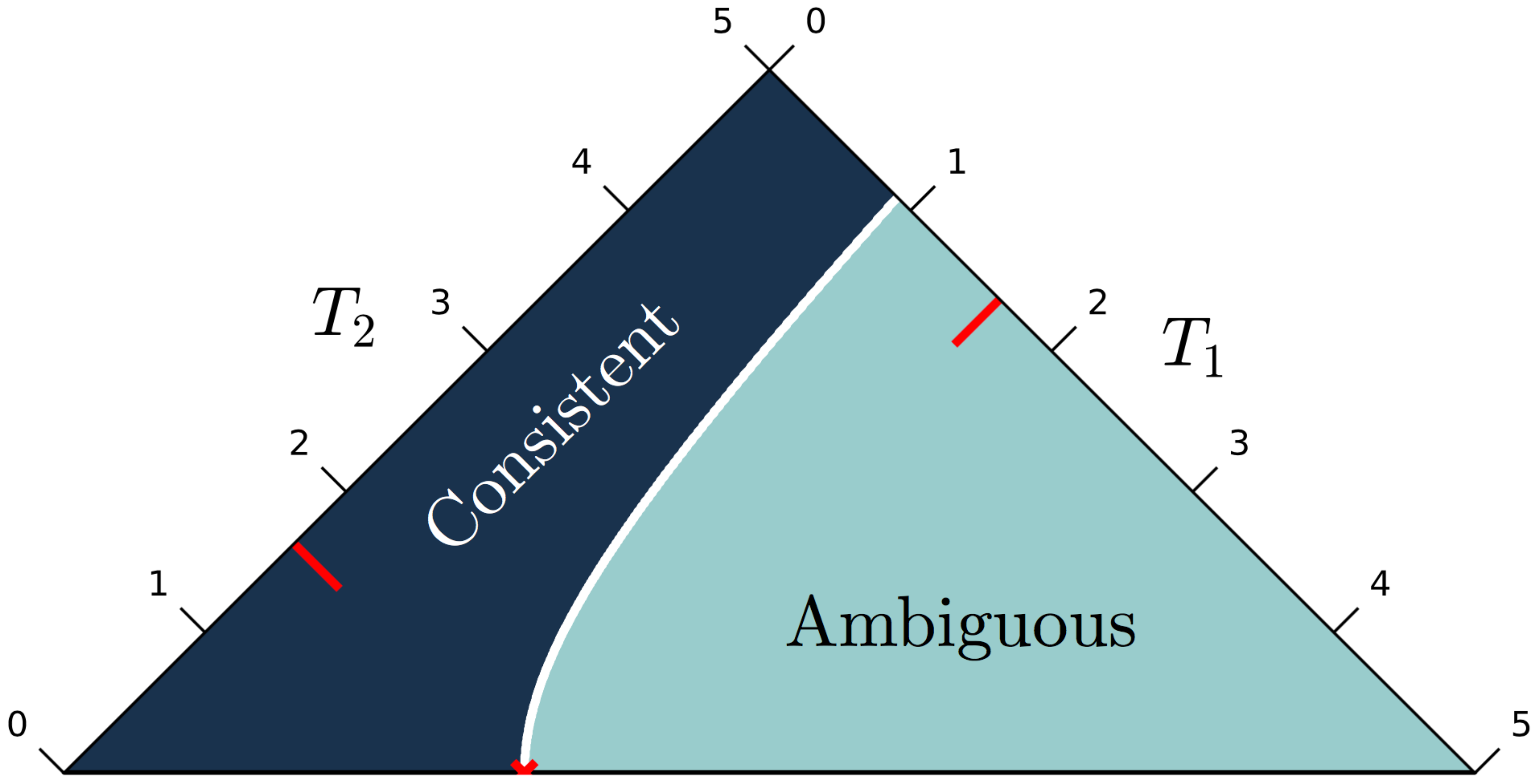
Ising Spin
Chain

Feldman et al (1997)
Suen et al arXiv:1511.05738

Ambiguity of Simplicity

Typical? Yes.

Compare two Ising spin chains at T_1 and T_2:



Ambiguity of Simplicity

Classical and quantum theory

do not agree on

measure of structure

or, even, what structure is.

Consequences

- **Big problem** using information measures to discover *organization in complex systems:*

  Misleading!

- **Bigger problem** in the *physics of complex systems*:

  Classical and quantum theories do not agree.

  Occam's Razor: Out the window!

Where are we?

Information theory is very strong on the negative side, i.e., in demonstrating what cannot be done; on the positive side its application has not produced many results so far; it has not yet led to the discovery of new facts, nor has its application to known facts been tested in critical experiments. To date, a definitive and valid judgment of the value of information theory is not possible.

The view from 1956 …

Information theory is very strong on the negative side, i.e., in demonstrating what cannot be done; on the positive side its application to the study of living things has not produced many results so far; it has not yet led to the discovery of new facts, nor has its application to known facts been tested in critical experiments. To date, a definitive and valid judgment of the value of information theory in biology is not possible.

Joint statement of final panel of the *Symposium on Information Theory in Biology*, Oak Ridge National Laboratory, Gatlinberg, Tennessee. 28-31 October 1956. Henry Quastler, chair.

H. Quastler (ed.), *The Status of Information Theory in Biology*, p. 399 in *Symposium on Information Theory in Biology*, H. P. Yockey, editor, Pergamon Press, New York (1958).

End?

Deal breaker for information theory?

Deal breaker for our notions of pattern?
Pure subjectivity?

No! A fantastic opportunity.

Full Employment Theorem.

End!

References

1. R. G. James, N. Barnett, & J. P. Crutchfield
   Information Flows? A Critique of Transfer Entropies,
   Physical Review Letters **116** (2016) 238701
2. R. G. James & J. P. Crutchfield
   Multivariate Dependence Beyond Shannon Information
   arxiv.org:1609.01233 [math.IT]
3. C. Aghamohammadi, J. R. Mahoney, & J. P. Crutchfield
   The Ambiguity of Simplicity
   arxiv.org:1602.08646 [quant-ph]
4. R. G. James, C. J. Ellison, & J. P. Crutchfield
   Python Discrete Information Theory Package **dit**
   http://github.com/dit/dit

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