Anatomy of a Bit

Reading for this lecture:

CMR articles Yeung and Anatomy.

Anatomy of a Bit ...

Information in a single measurement:

- Alternative rationale for entropy and entropy rate.
- k measurement outcomes
- Stored in $\log_{10} k$ decimal digits or $\log_2 k$ bits.

- Series of n measurements stored in $n \log_2 k$ bits.

- Number M of possible sequences of n measurements with $n_1, n_2, \ldots, n_k$ counts:

$$
\begin{aligned}
M &= \binom{n}{n_1 n_2 \cdots n_k} \\
&= \frac{n!}{n_1! n_2! \cdots n_k!}
\end{aligned}
$$

Anatomy of a Bit ...

- To specify/code which particular sequence occurred store:

$$k \log_2 n + \log_2 M + log_2 n + \cdots \quad \text{bits}$$

Maximum number of bits to store the count n_i of each of the k output values

Number of bits to specify particular observed sequence within the class of sequences that have counts $n_1, n_2, \ldots, n_k$

Bits to specify the number of bits in n itself

- Apply Stirlings approximation: $n! = \sqrt{2\pi n}(n/e)^n$

- To M: $\log_2 M \approx -n \sum_{i=1}^{k} \frac{n_i}{n} \log_2 \frac{n_i}{n}$

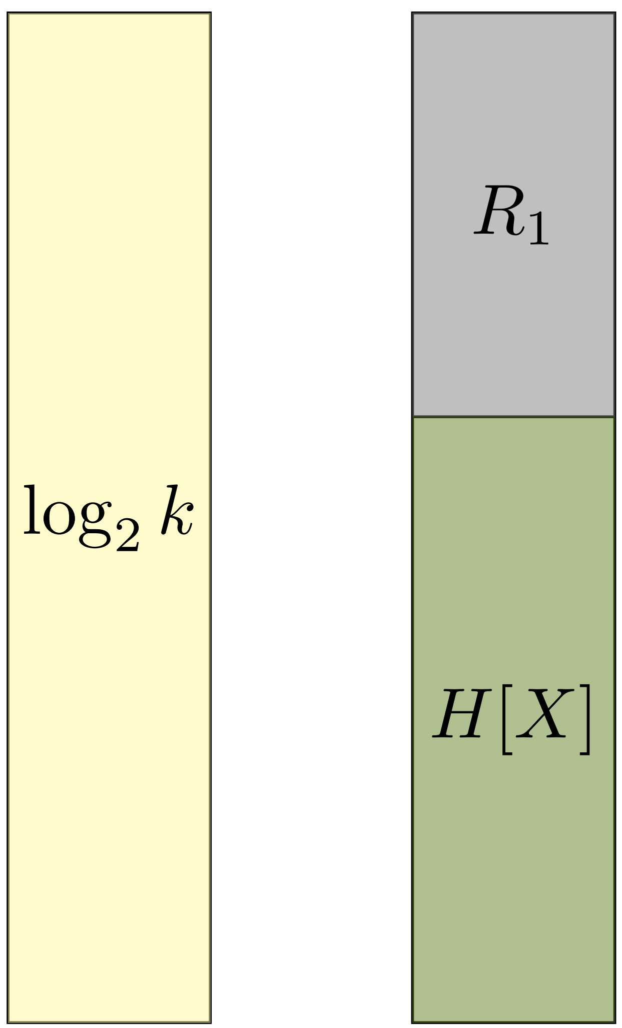
$$= nH\left[\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_k}{n}\right]$$

$$= nH[X]$$

Anatomy of a Bit ...

- Can compress if

$$H[X] < \log_2 k$$



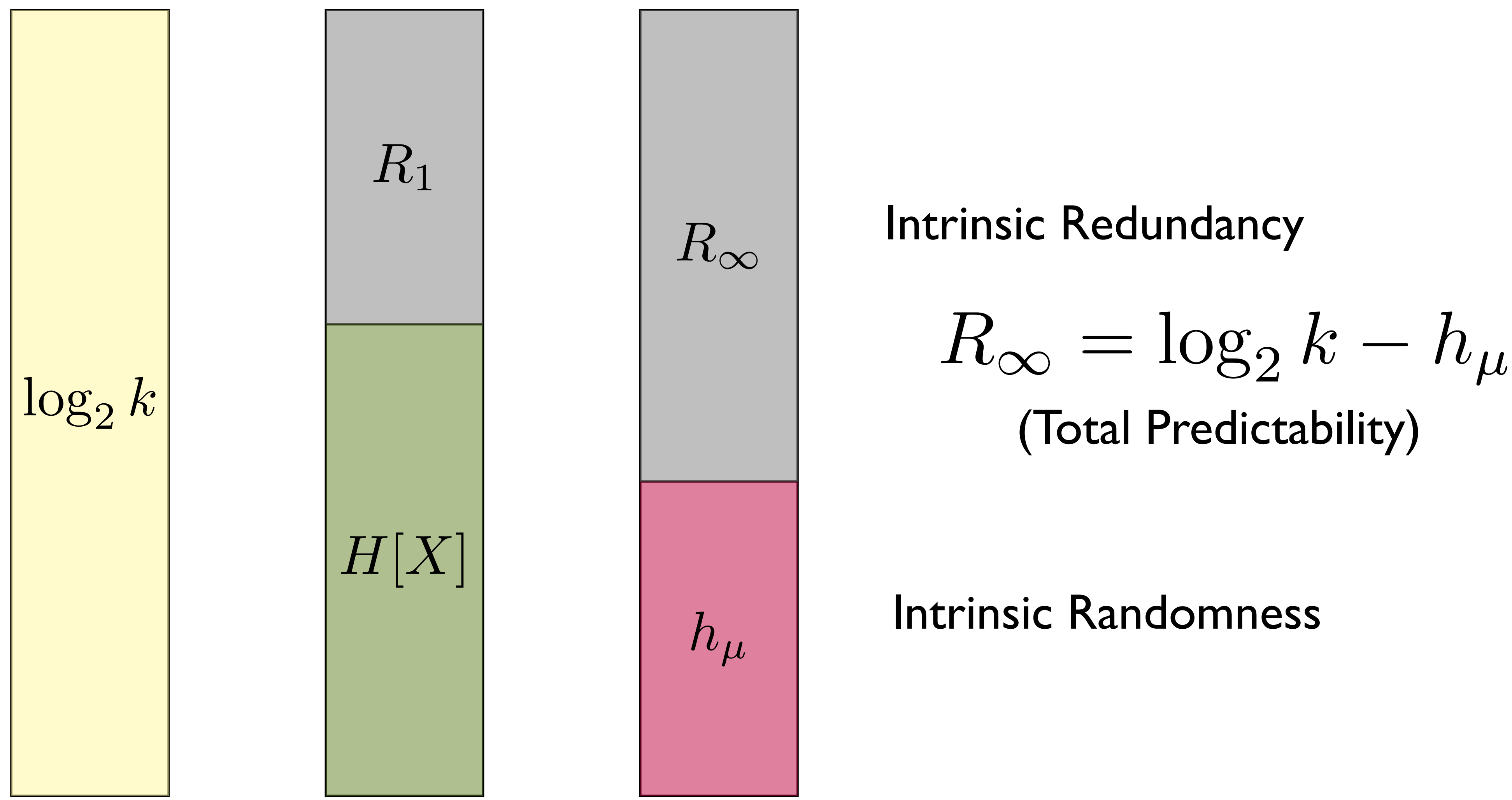
1-Symbol Redundancy

$$R_1 = \log_2 k - H[X]$$

Anatomy of a Bit ...

- Typical sources are not IID, therefore use:

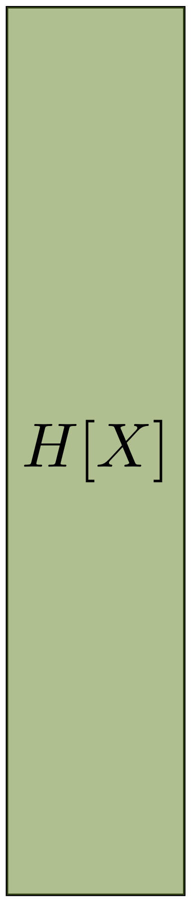
$$h_\mu = \lim_{L\to\infty} \frac{H(L)}{L}$$



Anatomy of a Bit ...

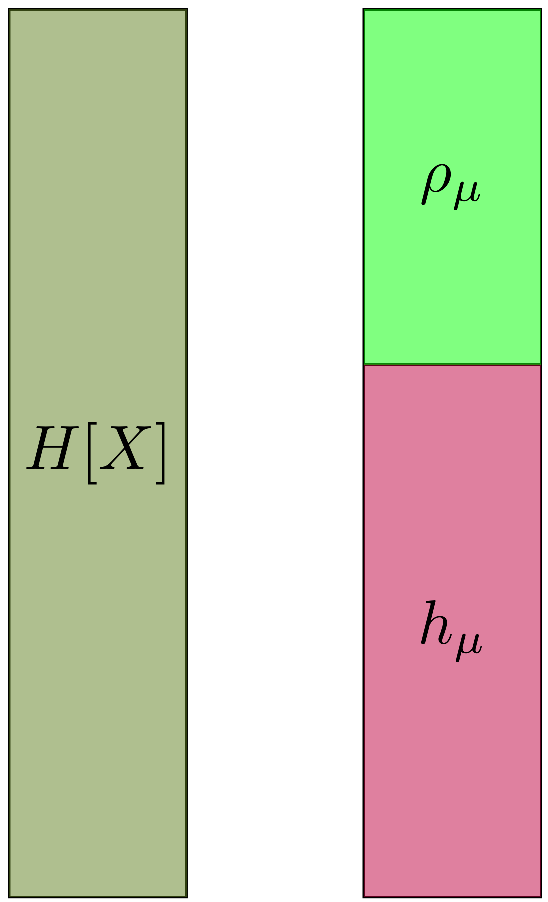
In addition, measurements occur in the context of past and future.

A semantically rich decomposition of a single measurement:



Anatomy of a Bit ...

A semantically rich decomposition of a single measurement ...

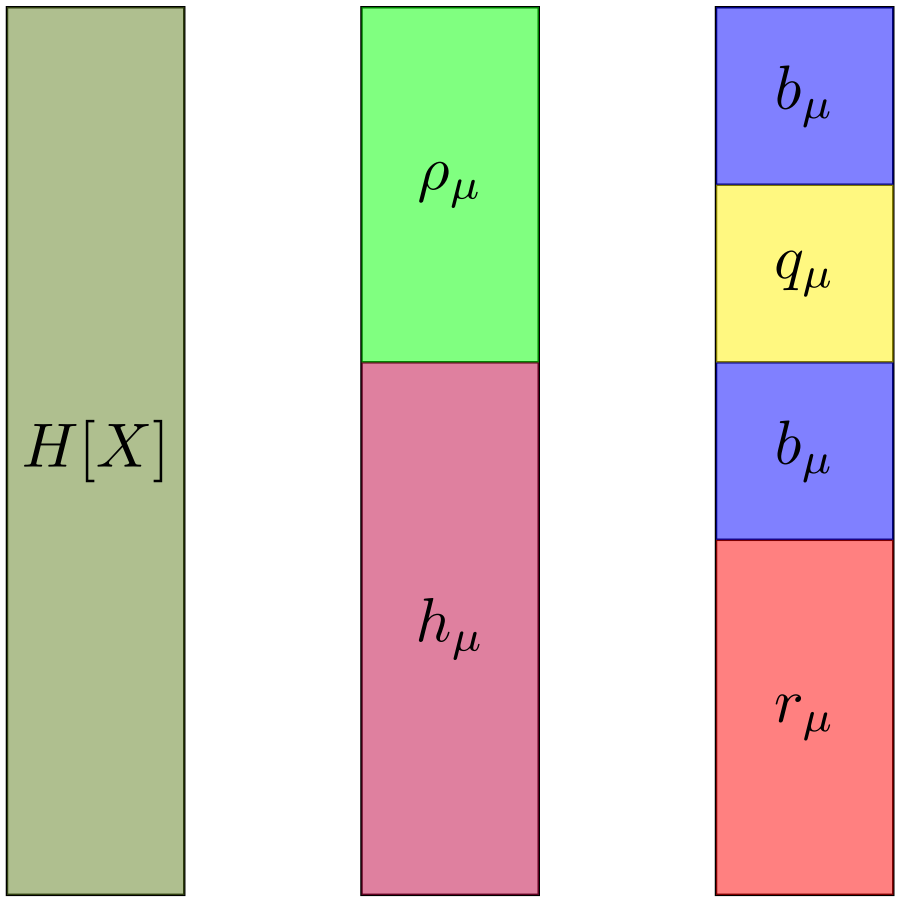


Anticipated from prior observations

Random component

Anatomy of a Bit ...

A semantically rich decomposition of a single measurement ...



Shared between the past and the current observation, but relevance stops there (stationary process only)

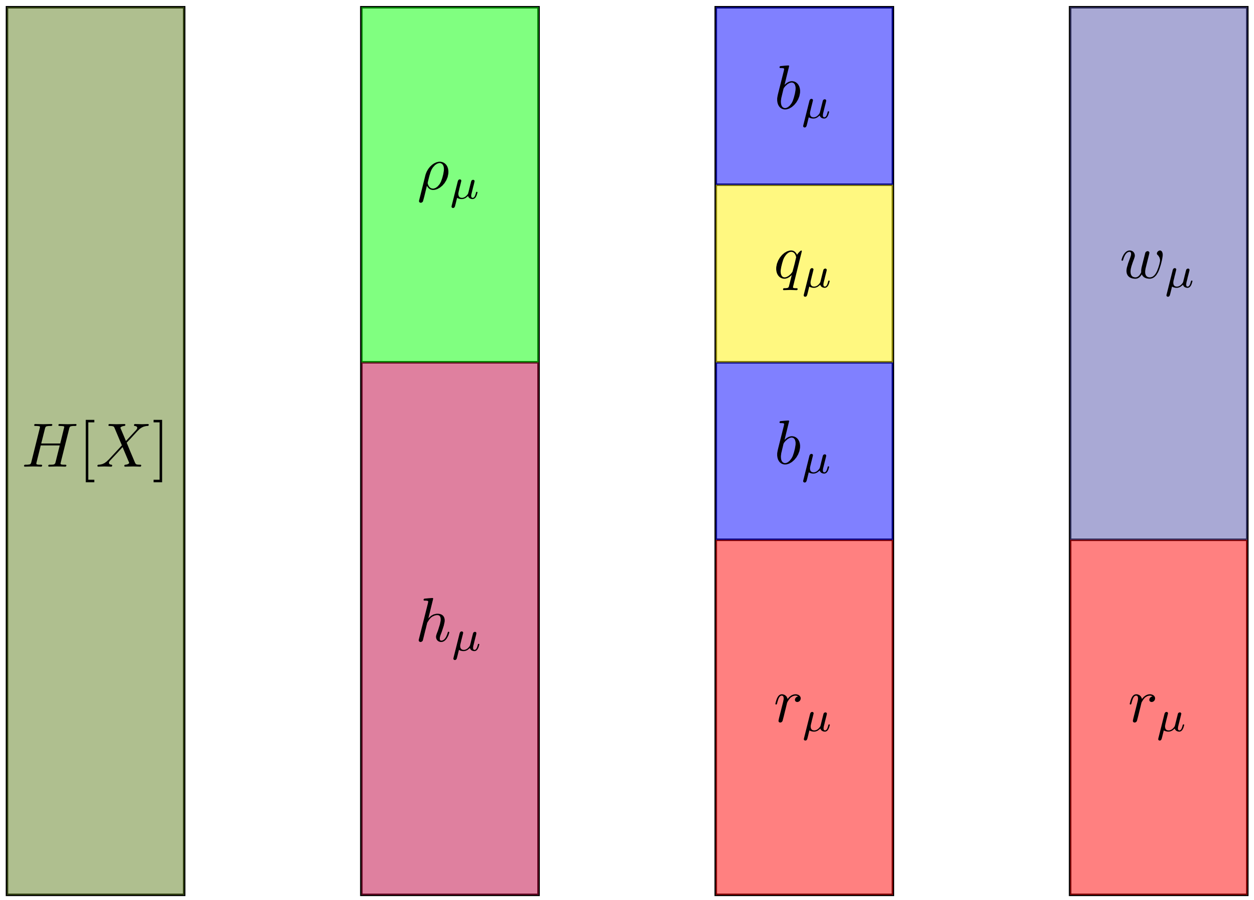
Anticipated by the past, present currently, and plays a role in future behavior. Can be negative!

Relevant for predicting the future

Ephemeral: Exists only for the present

Anatomy of a Bit ...

A semantically rich decomposition of a single measurement ...



Structural information

Ephemeral: exists only for the present

Anatomy of a Bit ...

Settings:

Before: Predict the present from the past, $h_\mu = H[X_0|\overleftarrow{X}]$

$X_{:0}$ $\quad$ X_0

$$\cdots \quad X_{-3} \quad X_{-2} \quad X_{-1} \quad X_0$$

Today, prediction from the omniscient view:
The present in the context of the past and the future.

$X_{:0}$ $\quad$ X_0 $\quad$ $X_{1:}$

$$\cdots \quad X_{-3} \quad X_{-2} \quad X_{-1} \quad X_0 \quad X_1 \quad X_2 \quad X_3 \quad \cdots$$

Anatomy of a Bit ...

Notations:

l-blocks: $X_{t:t+\ell} = X_t X_{t+1} \ldots X_{t+\ell-1}$

Past: $X_{:0} = \ldots X_{-2} X_{-1}$

Present: X_0

Future: $X_{1:} = X_1 X_2 \ldots$

Any information measure $\mathcal{F}$:

$$\mathcal{F}(\ell) = \mathcal{F}[X_{0:\ell}]$$

Anatomy of a Bit ...

Recall block entropy:

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

where:

$$\mathbf{E} = I[X_{:0}; X_{0:}]$$

Recall total correlation:

$$T[X_{0:N}] = \sum_{\substack{A \in P(N) \\ |A|=1}} H[X_A] - H[X_{0:N}]$$

For stationary time series, get block total correlation:

$$T(\ell) = \ell H[X_0] - H(\ell) \quad \text{with} \quad T(0) = 0 \quad T(1) = 0$$

Compares process's block entropy to the case of IID RVs.

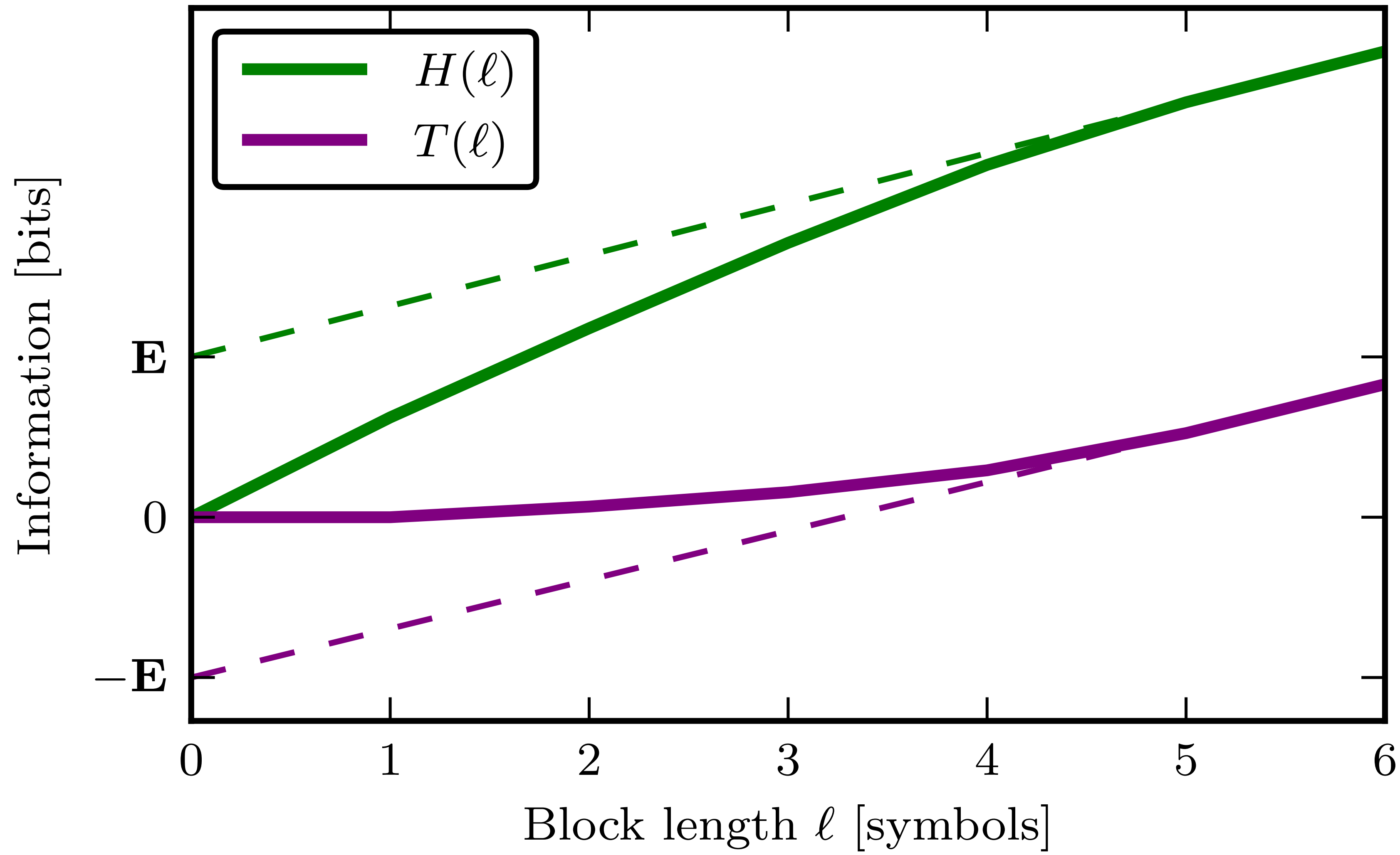
Anatomy of a Bit ...

Block entropy versus total correlation ...

Monotone increasing, with linear asymptote:

$$T(\ell) \approx -\mathbf{E} + \rho_\mu \ell$$

Convex up.



Anatomy of a Bit ...

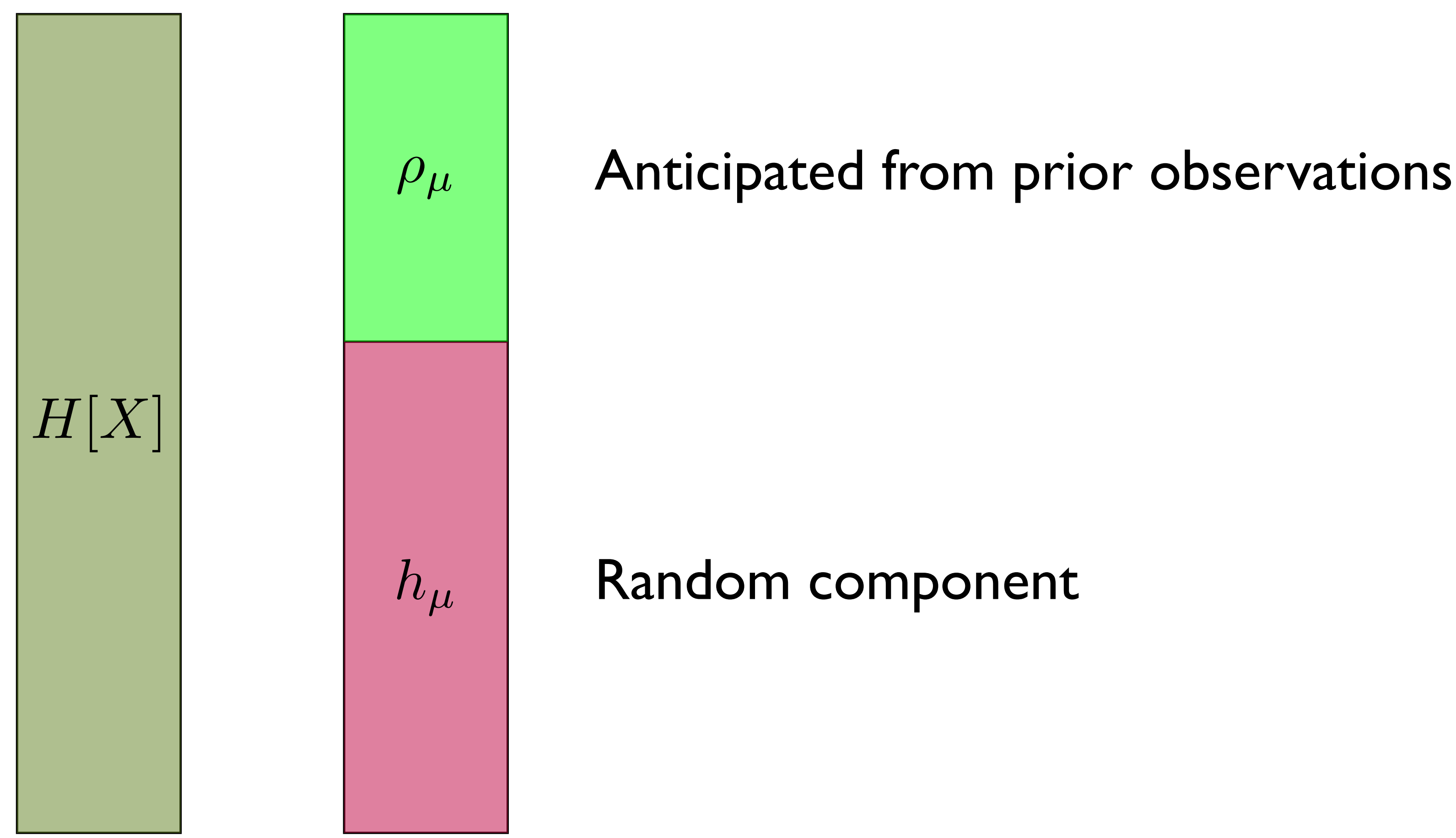
Block entropy versus total correlation ...

Note:

$$H(\ell) + T(\ell) = \ell H(1)$$

Plug in asymptotes gives first decomposition:

$$H(1) = h_\mu + \rho_\mu$$



Anatomy of a Bit ...

Scaling of other block informations:

Bound information $B(\ell) = H(\ell) - R(\ell)$

Enigmatic $Q(\ell) = T(\ell) - B(\ell)$

Local Exogenous $W(\ell) = B(\ell) + T(\ell)$

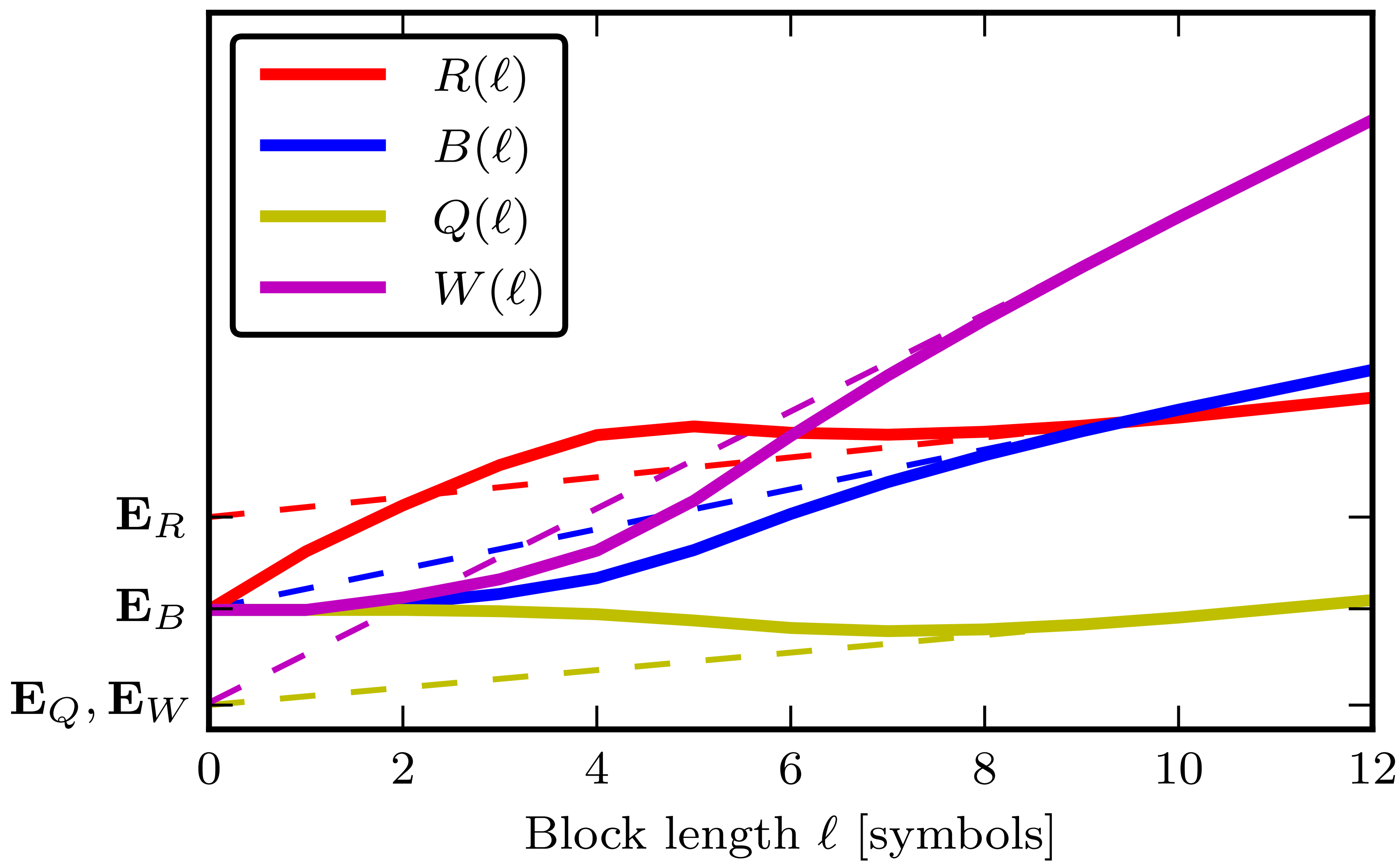
Asymptotes:

$$R(\ell) \approx \mathbf{E}_R + \ell r_\mu$$

$$B(\ell) \approx \mathbf{E}_B + \ell b_\mu$$

$$Q(\ell) \approx \mathbf{E}_Q + \ell q_\mu$$

$$W(\ell) \approx \mathbf{E}_W + \ell w_\mu$$



Anatomy of a Bit ...

Scaling of other block informations:

Block entropy:

$$H(\ell) = B(\ell) + R(\ell)$$
$$= (\mathbf{E}_B + \mathbf{E}_R) + \ell(b_\mu + r_\mu)$$

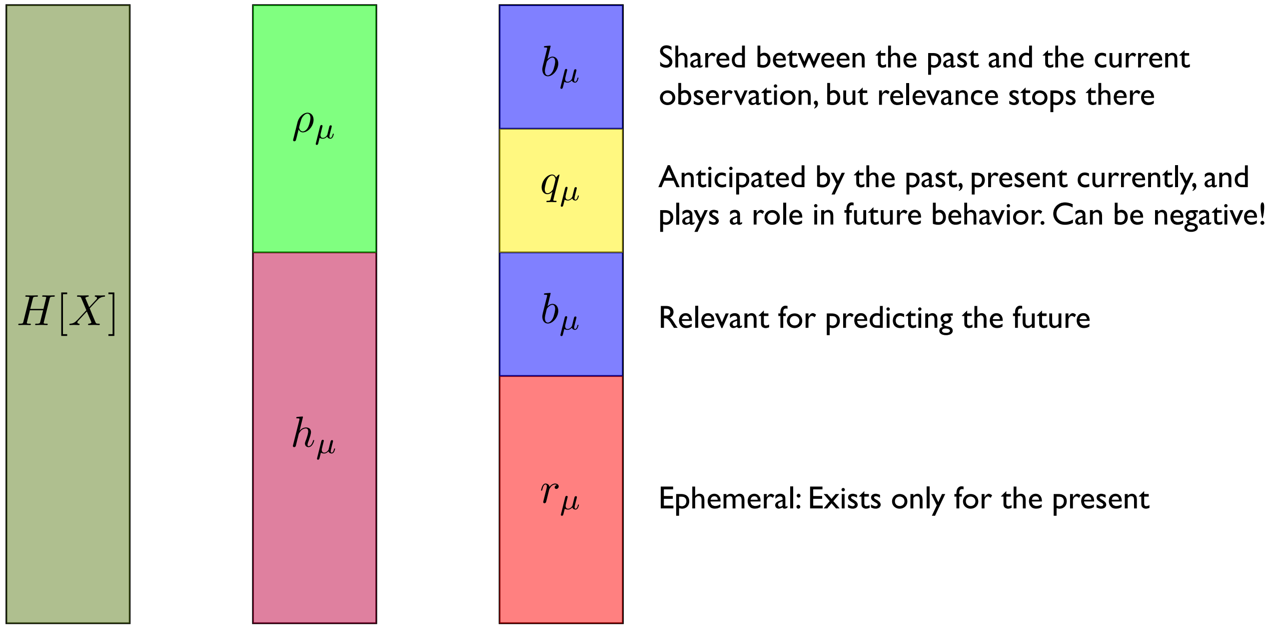
Thus, $h_\mu = b_\mu + r_\mu$ and $\mathbf{E} = \mathbf{E}_B + \mathbf{E}_R$.

Block total correlation:

$$T(\ell) = B(\ell) + Q(\ell)$$
$$= (\mathbf{E}_B + \mathbf{E}_Q) + \ell(b_\mu + q_\mu)$$

Thus, $\rho_\mu = b_\mu + q_\mu$ and $\mathbf{E} = -\mathbf{E}_B - \mathbf{E}_Q$.

Gives second decomposition of $H[X]$.



Anatomy of a Bit ...

Scaling of other block informations:

Block local exogenous:

$$W(\ell) = B(\ell) + T(\ell)$$
$$= (\mathbf{E}_B - \mathbf{E}) + \ell(b_\mu + \rho_\mu)$$

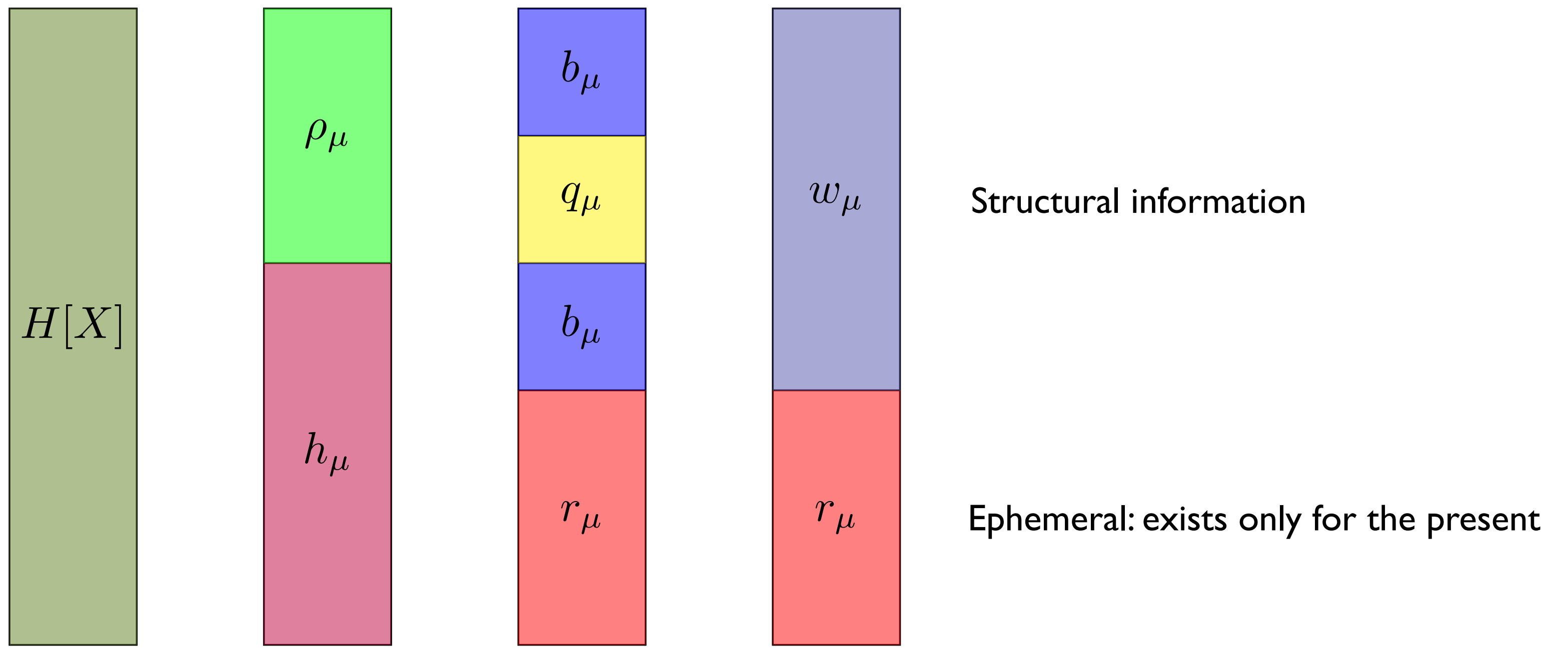
Thus, $w_\mu = b_\mu + \rho_\mu$.

From definitions:

$$R(\ell) + W(\ell) = \ell H(1)$$

Plug in asymptotes gives third decomposition of $H[X]$:

$$H(1) = w_\mu + r_\mu$$



Anatomy of a Bit ...

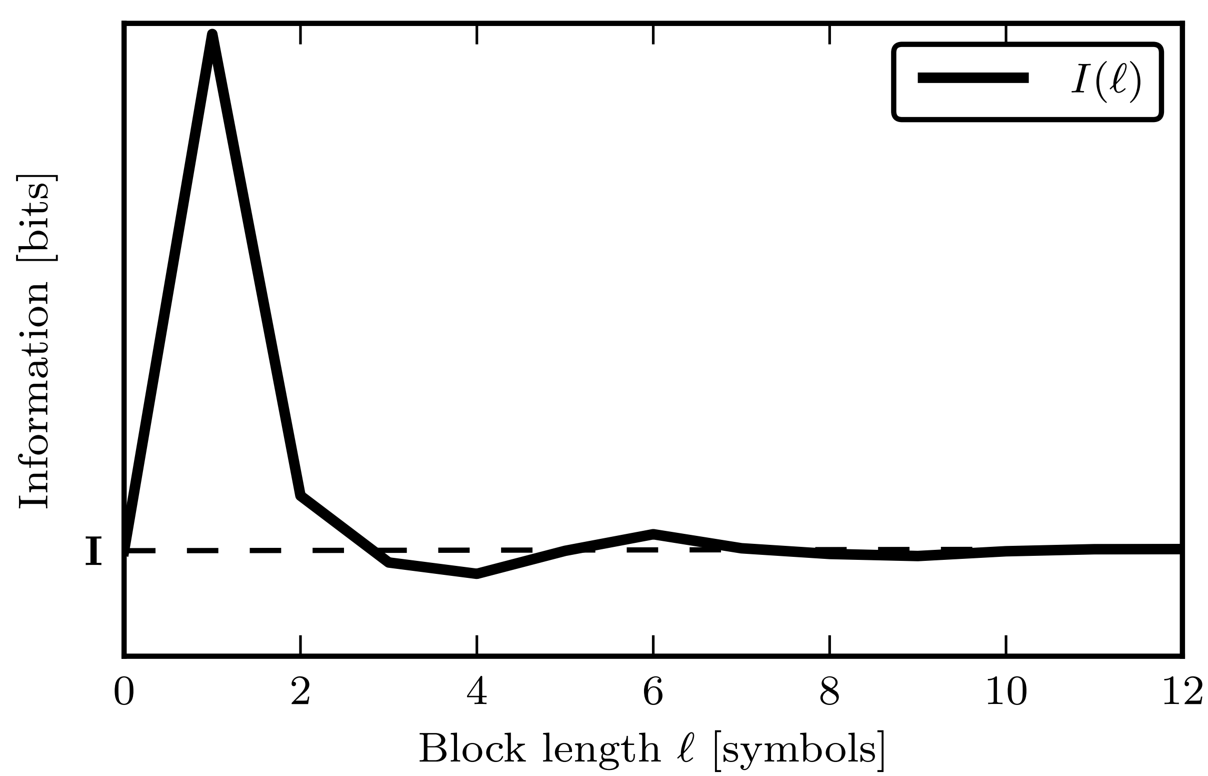
Block multivariate mutual information:

$$I(\ell) = H(\ell) - \sum_{\substack{A \in P(\ell) \\ 0 < |A| < \ell}} I[X_A | X_{\bar{A}}]$$

Asymptote: $I(\ell) \approx \mathbf{I} + \ell i_\mu$?

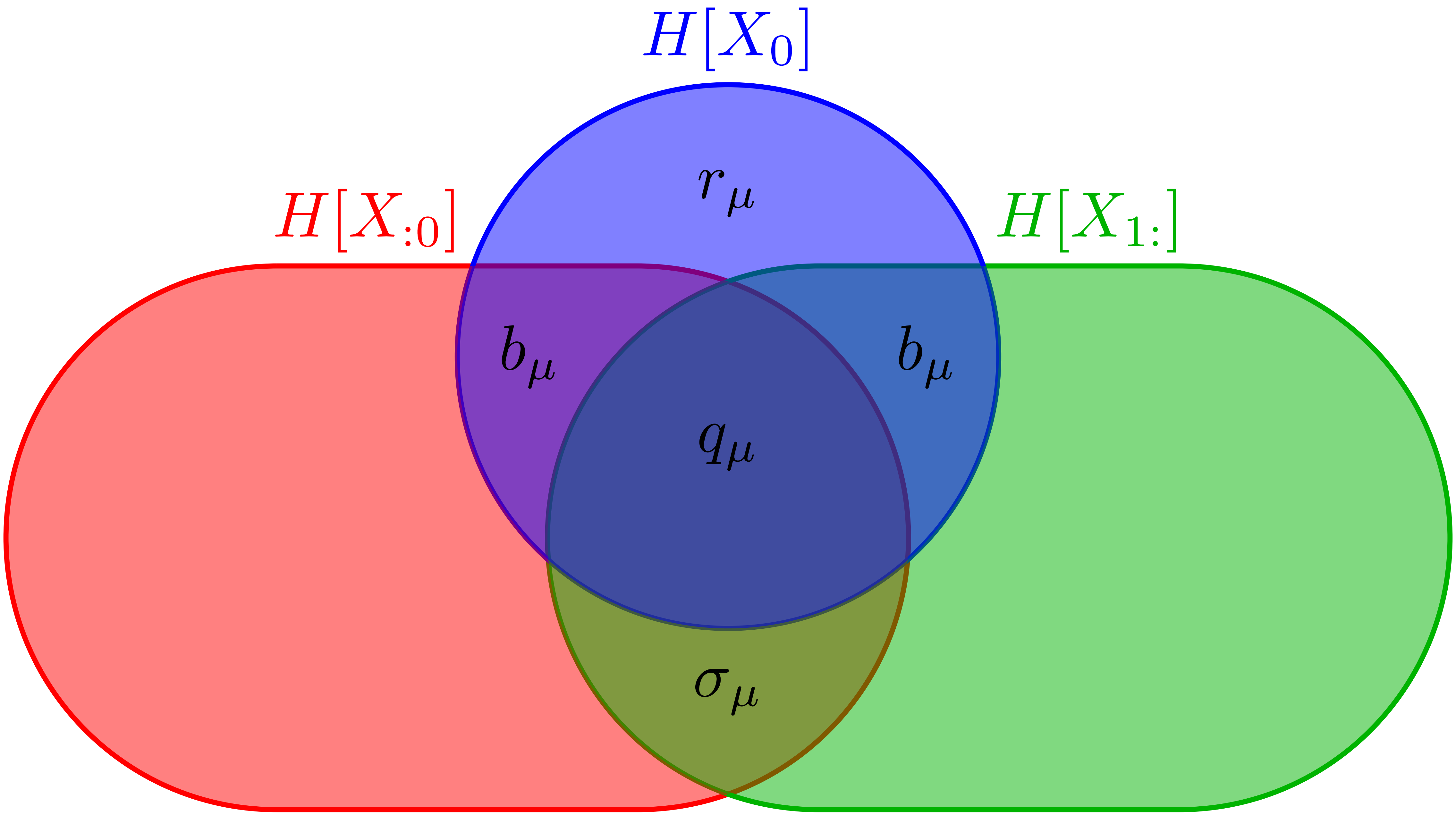
Observations of finite-state process tend to independence:

$$h_\mu > 0 : \qquad \lim_{\ell \to \infty} I(\ell) = 0 \qquad i_\mu = 0 \qquad \mathbf{I} = 0$$



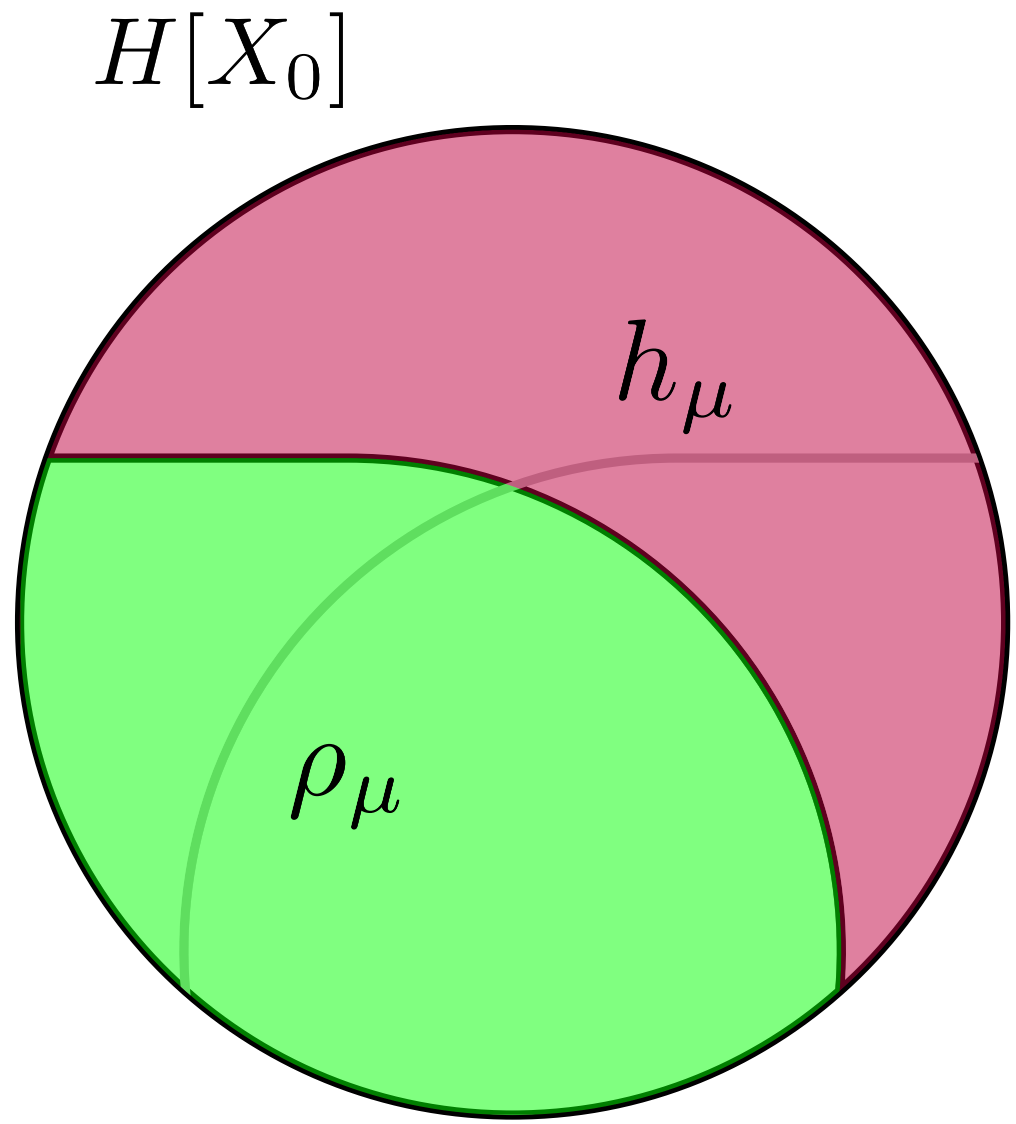
Anatomy of a Bit ...

Measurement Decomposition in the Process I-diagram:



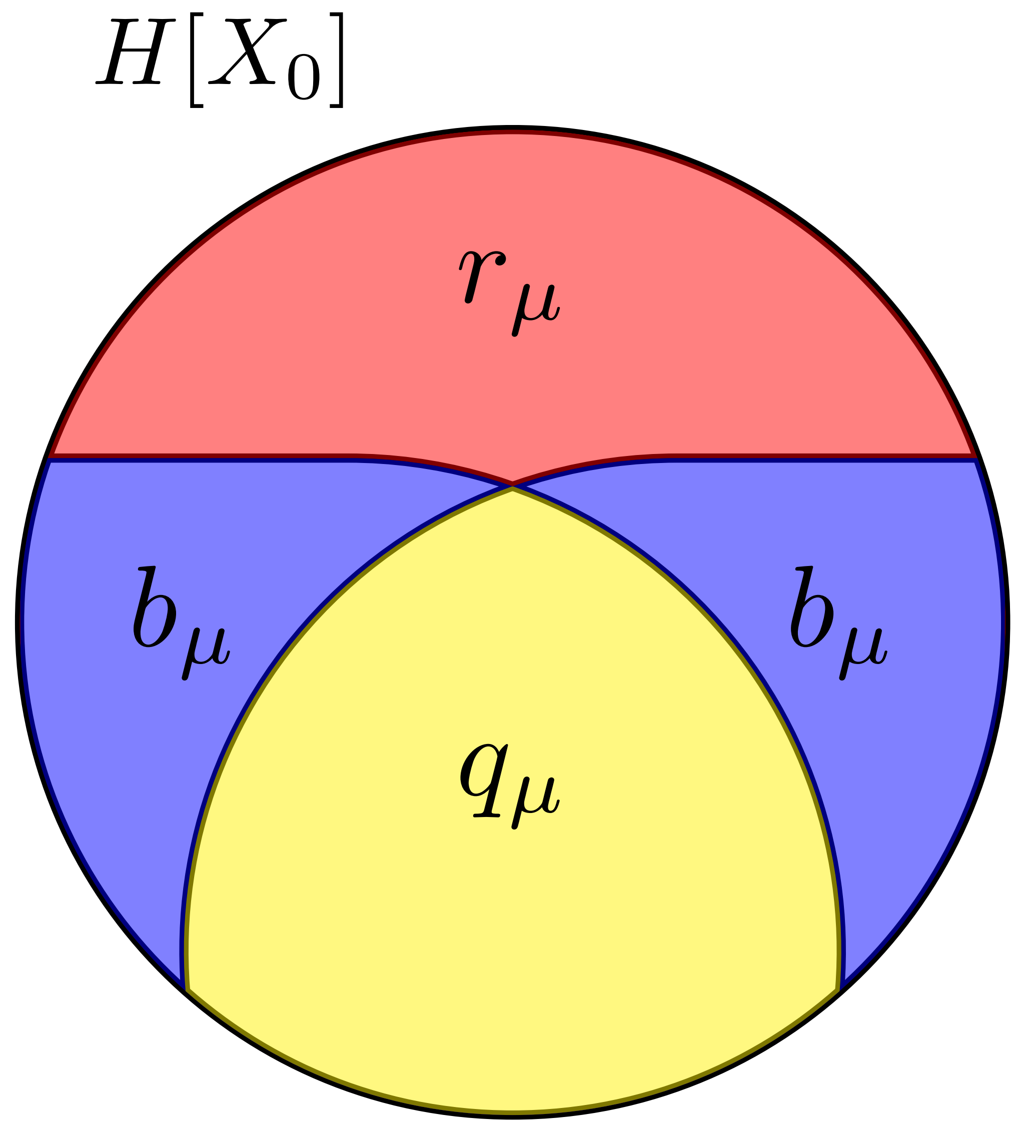
Anatomy of a Bit ...

Predictive decomposition:



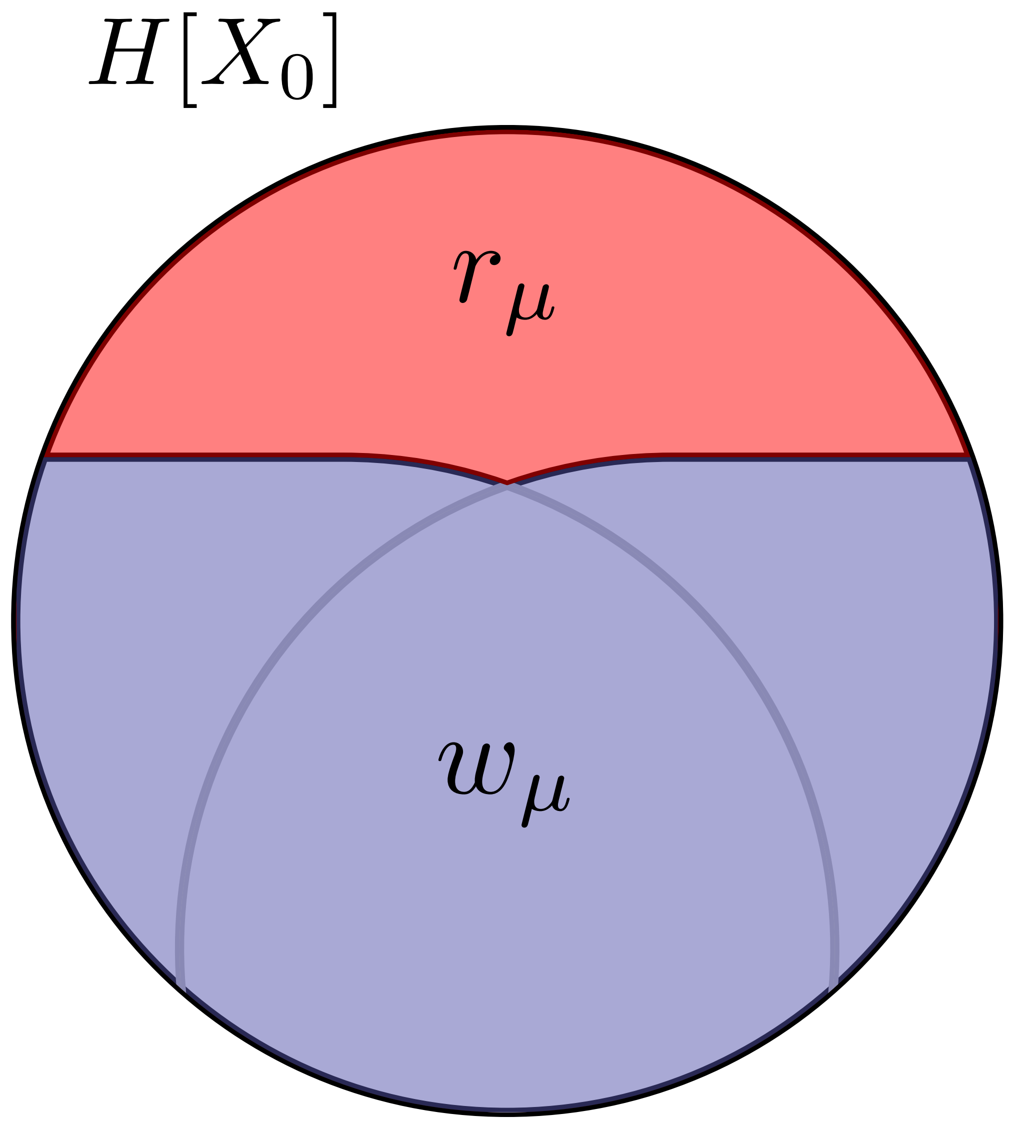
Anatomy of a Bit ...

Atomic decomposition:



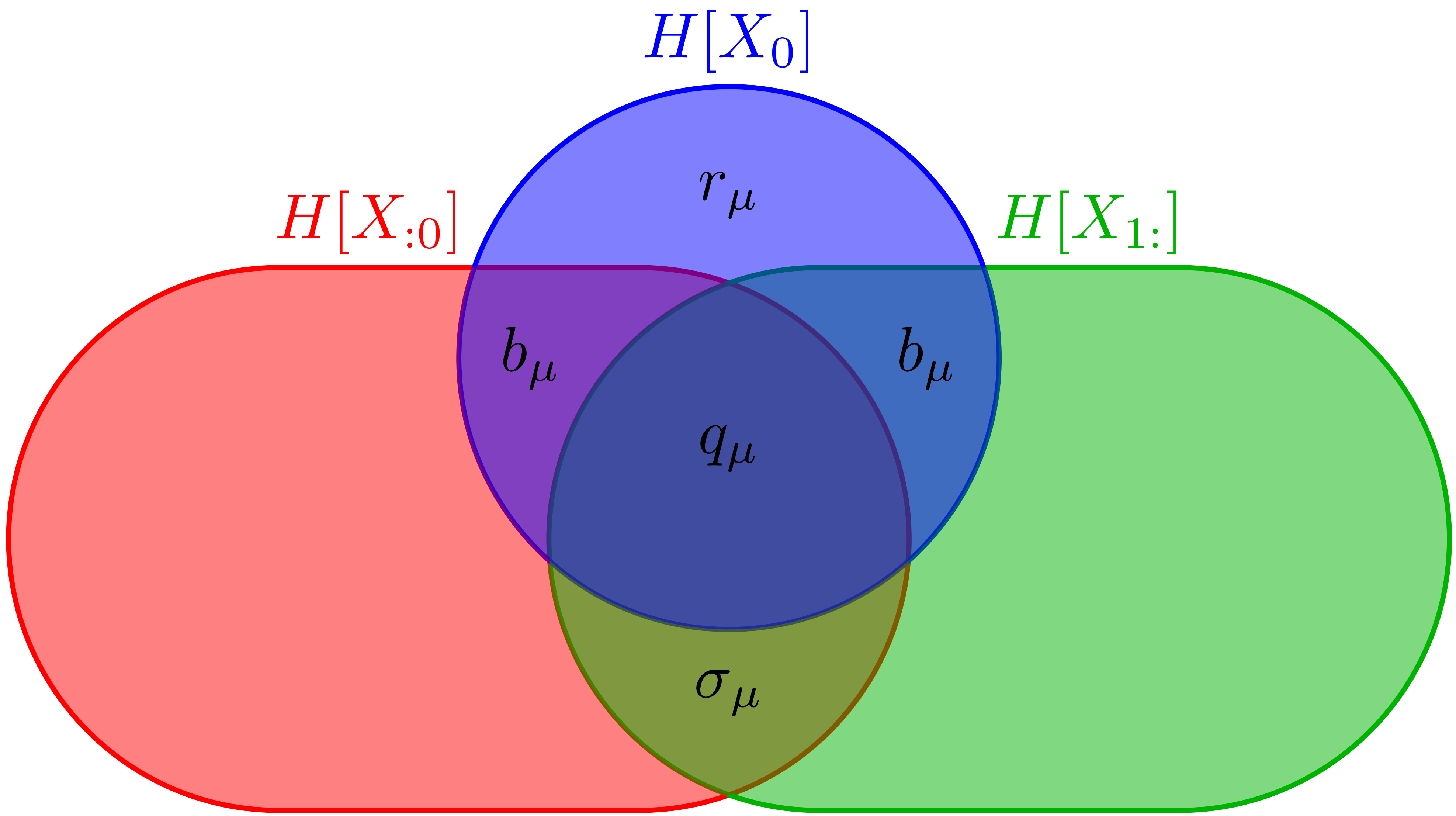
Anatomy of a Bit ...

Structural decomposition:



Anatomy of a Bit ...

Measurement Decomposition in the Process I-diagram ...

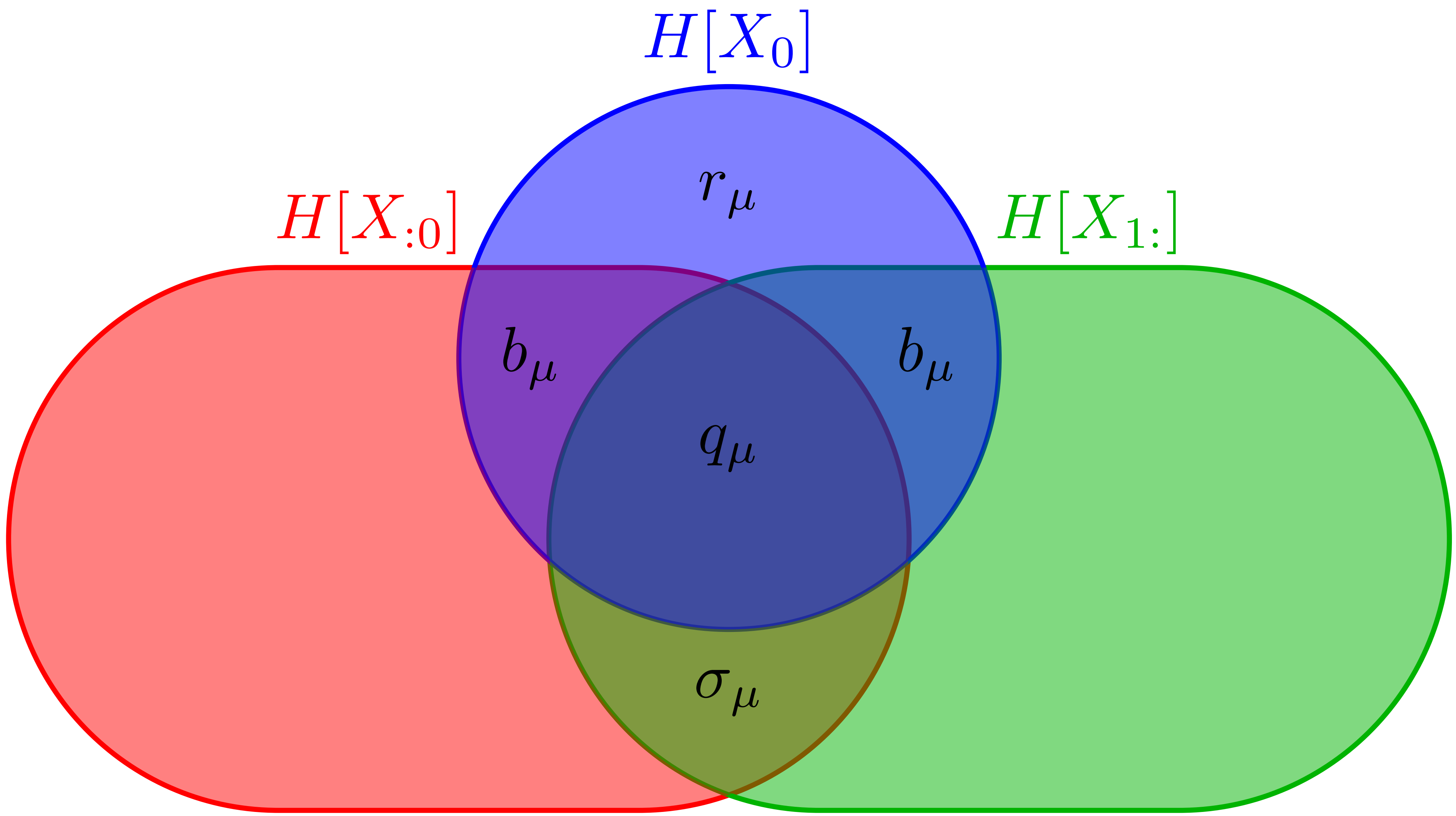


What is σ_μ?

- Not a component of the information in a single observation.
- Information shared between past and future, not in present!
- Hidden state information!

Anatomy of a Bit ...

Measurement Decomposition in the Process I-diagram:

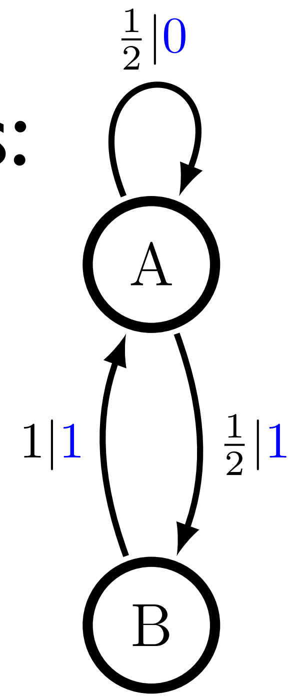


Decomposition of Excess Entropy:

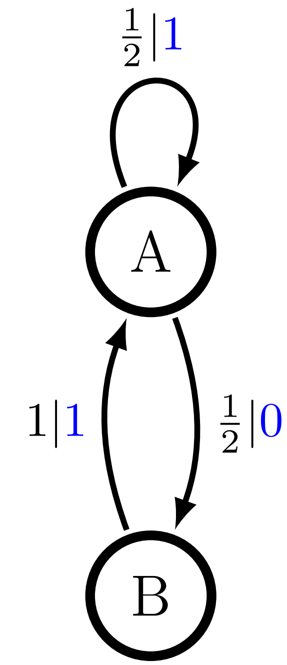
$$\mathbf{E} = b_\mu + q_\mu + \sigma_\mu$$

Anatomy of a Bit ...

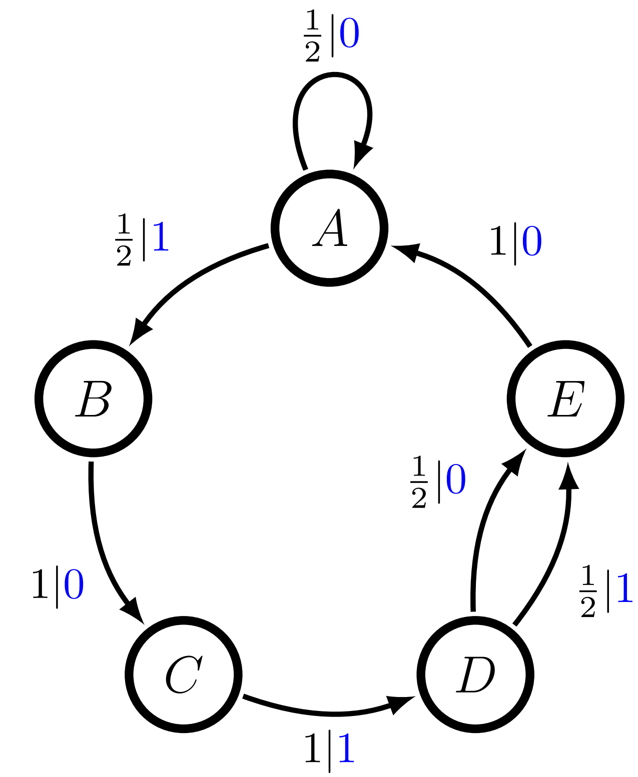
Example Processes:



Even



Golden Mean

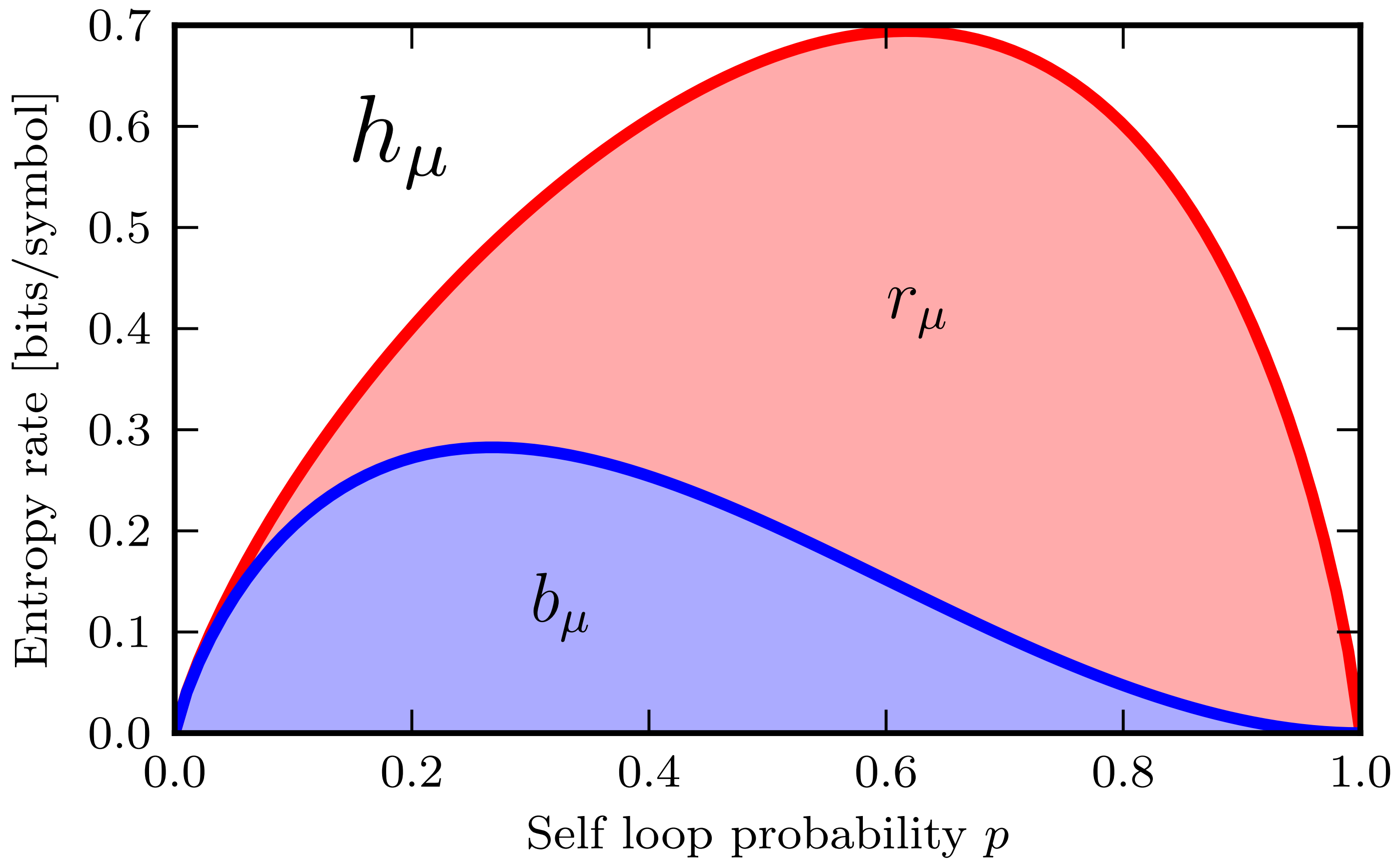


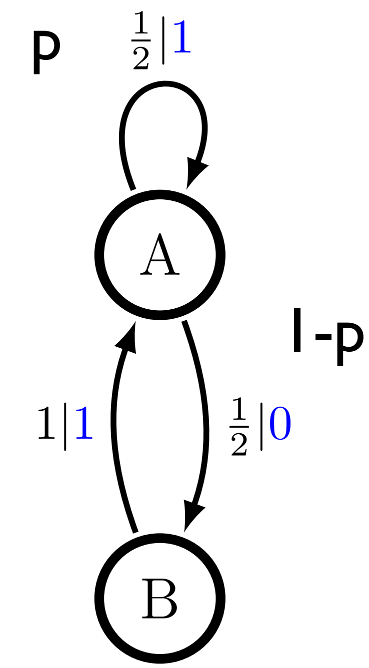
Noisy Random Phase Slip

	Even	Golden Mean	NRPS
$H[1]$	0.91830	0.91830	0.97987
h_μ	0.66667	0.66667	0.50000
ρ_μ	0.25163	0.25163	0.47987
r_μ	0.00000	0.45915	0.16667
b_μ	0.66667	0.20752	0.33333
q_μ	-0.41504	0.04411	0.14654
w_μ	0.91830	0.45915	0.81320
σ_μ	0.66667	0.00000	1.09407

Anatomy of a Bit ...

Example: Parametrized Golden Mean Process:





~ Period 2: Randomness affects phase and so is remembered.

$$r_\mu \approx 0$$

$$h_\mu \approx b_\mu$$

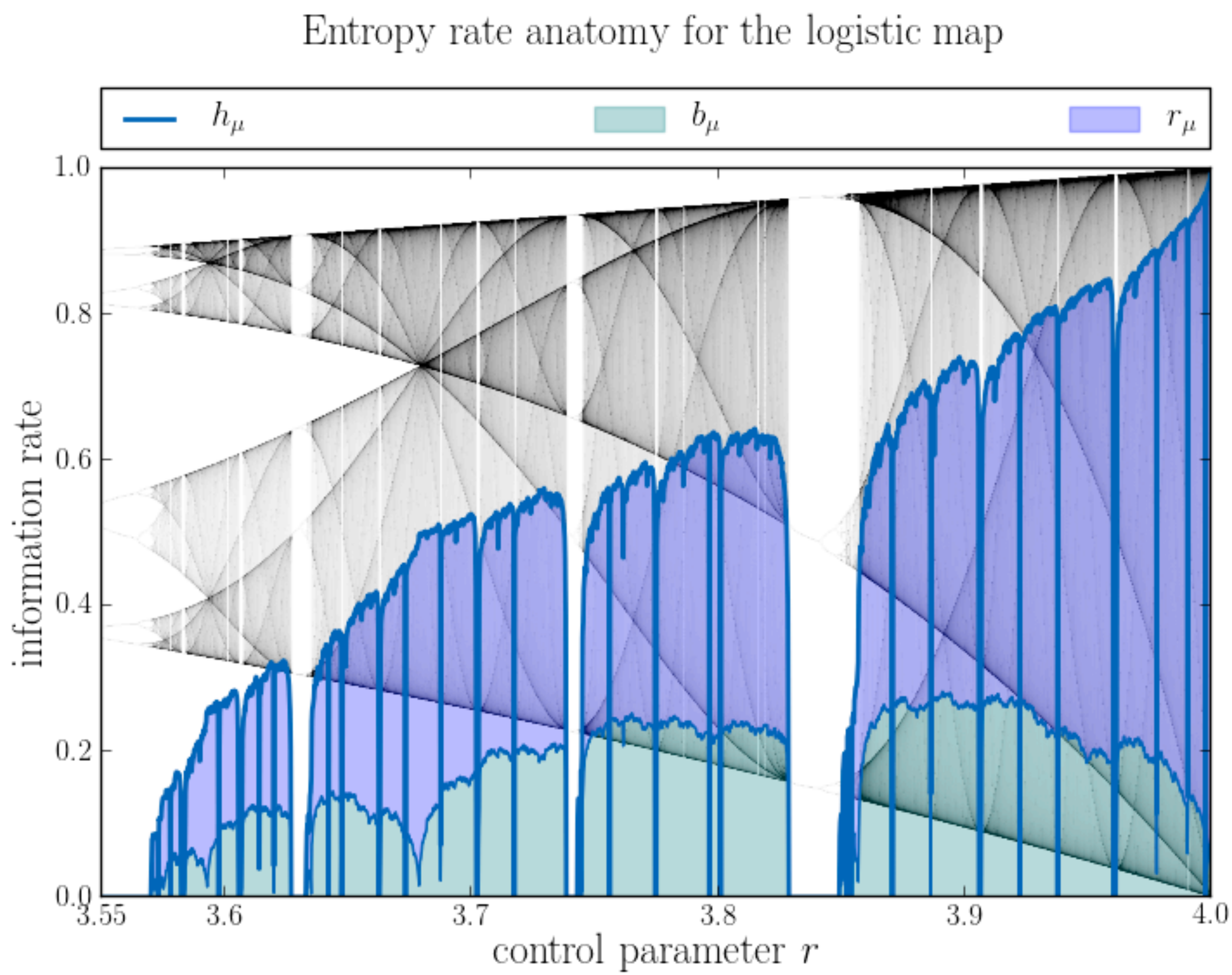
~ Period1: Randomness disrupts all 1s, but is not remembered.

$$b_\mu \approx 0$$

$$h_\mu \approx r_\mu$$

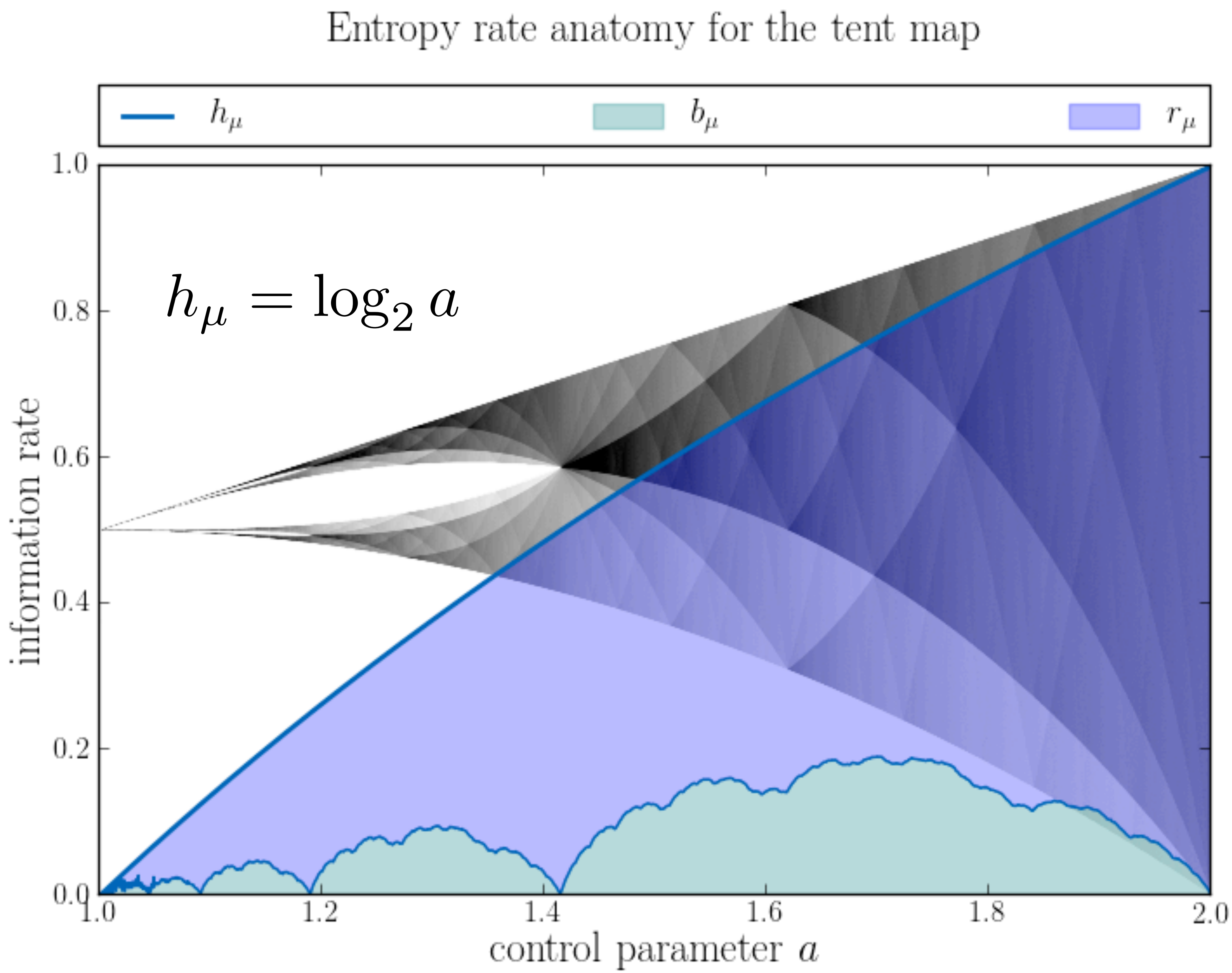
Anatomy of a Bit ...

Entropy-rate Decomposition: Logistic Map



Anatomy of a Bit ...

Entropy-rate decomposition: Tent Map



Anatomy of a Bit ...

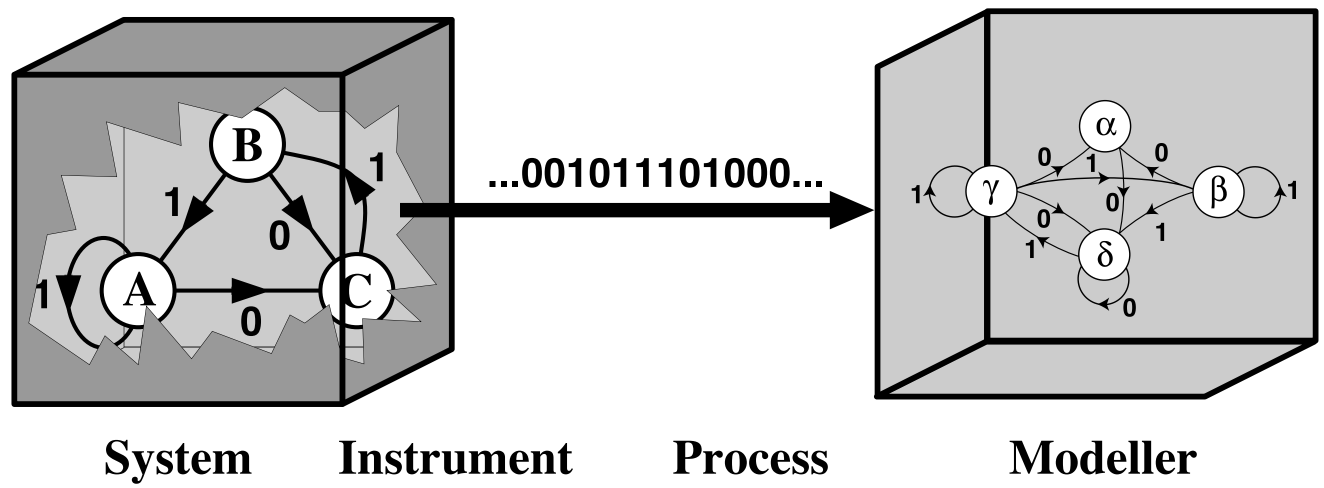
What is information?

Depends on the question!

Uncertainty, surprise, randomness, ...	$H(X) \quad h_\mu$
Compressibility.	$\mathcal{R} = \log_2 \lvert\mathcal{A}\rvert - h_\mu$
Transmission rate.	$I[X;Y]$
"Memory", apparent stored information, ...	$\mathbf{E}$
Synchronization.	$\mathbf{T}$
Ephemeral.	r_μ
Bound.	b_μ
...	

Anatomy of a Bit ...

Analysis Pipeline:



1. An information source: Calculate or estimate $\Pr(s^L)$
   a. Dynamical System:
      deterministic or stochastic
      low-dimensional or high-dimensional (spatial?)
   b. Design instrument (partition)
2. Information-theoretic analysis: $H(L)$
   a. How much information produced? h_μ
   b. How much stored information? $\mathbf{E}$
   c. How does observer synchronize? $\mathbf{T}$

Anatomy of a Bit ...

Course summary:

Dynamical systems as sources of complexity:
1. Chaotic attractors (State)
2. Basins of attraction (Initial conditions)
3. Bifurcation sequences (Parameters)

Dynamical systems as information processors:
1. Randomness:
   Entropy hierarchy: block entropy, entropy convergence, rate, ...
2. Information storage:
   1. Total predictability
   2. Excess entropy
   3. Transient information

Preview of 256B: Physics of Computation

Answers a number of questions 256A posed:

What is a model?
What are the hidden states and equations of motion?
Are these always subjective, depending on the observer?
Or is there a principled way to model processes?
To discover states?

What are cause and effect?

To what can we apply these ideas?
How much can we calculate analytically?
How much numerically?
How much from data?

Anatomy of a Bit ...

Preview of 256B: Physics of Computation

Punch line:

How nature is organized

is

How nature computes

Anatomy of a Bit ...

Reading for next lecture:

CMR articles RURO (Introduction), CAO, and ROIC.

... *next quarter!*