Information Theory on Neuroscience

Short tutorial

Computational Neuroscience & Oscillations Journal Club

05/04

Rodrigo FO Pena

e-mail: pena@njit.edu

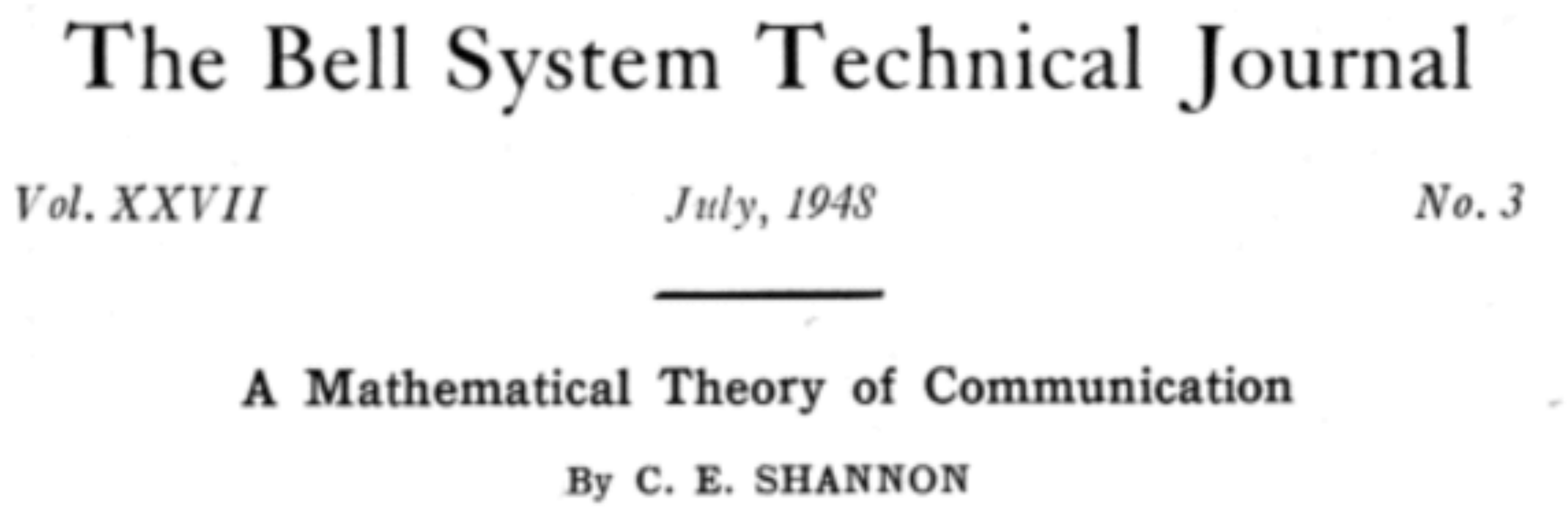
BioDatanamics Lab (H. G. Rotstein), Federated Department of Biological Sciences

New Jersey Institute of Technology & Rutgers University

NJIT New Jersey Institute of Technology

RUTGERS UNIVERSITY | NEWARK

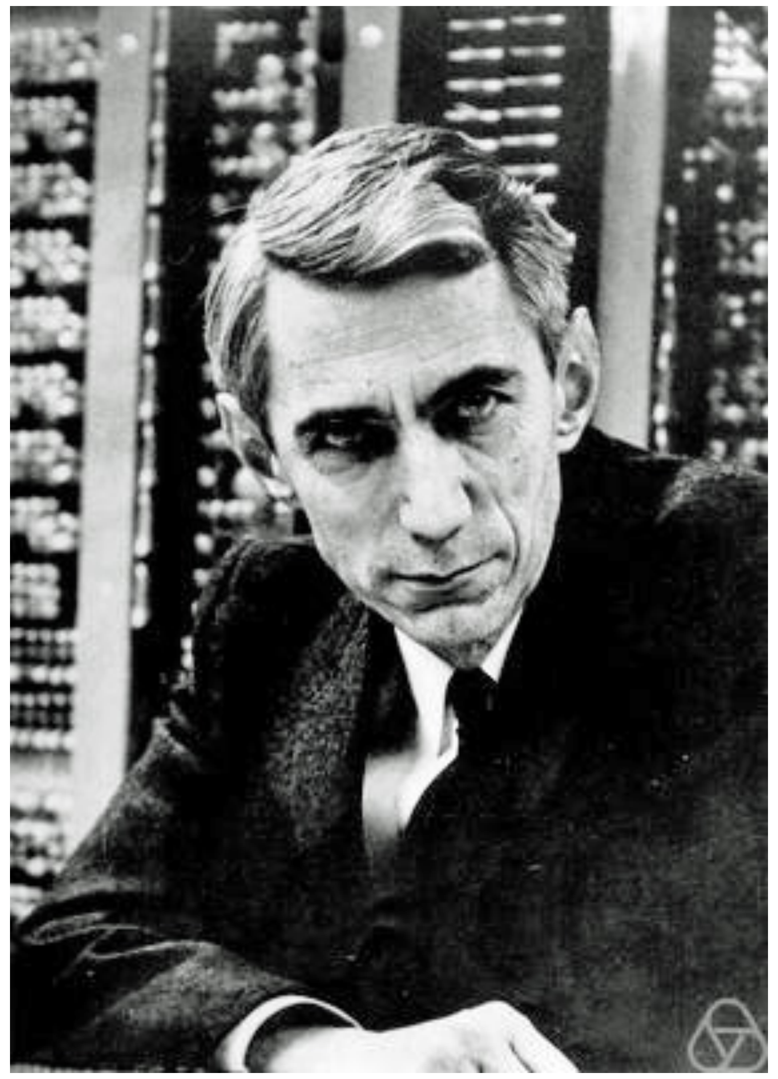
Center for Molecular and Behavioral Neuroscience


The Bell System Technical Journal

Vol. XXVII July, 1948 No. 3

A Mathematical Theory of Communication

By C. E. SHANNON


Claude Shannon

- transmission
- processing
- extraction
- and utilization of information

What is information?

- **Shannon information**

Associates information with uncertainty!

How many yes/no questions to figure out the message?

$\uparrow$ 0

Message 3 characters: spike or silence

Equally likely!

$2^3 = 8$ possible states.

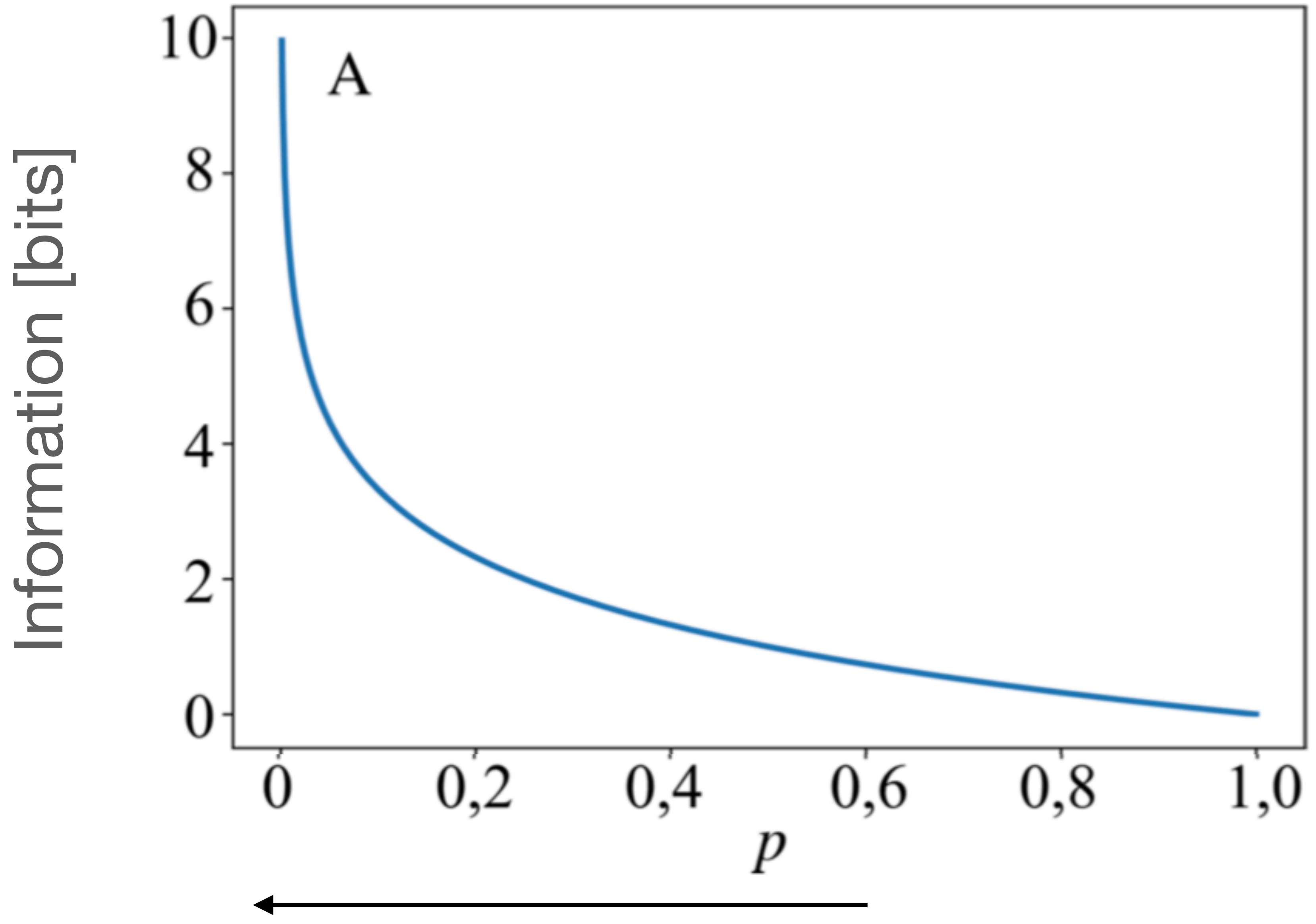
3 yes/no questions: 3 bits of information

$$\boxed{3 = \log_2(8) = -\log_2(1/8) = -\log_2(p_i)}$$

Observing an event is associated with
-> “reduction of uncertainty”
-> “degree of surprise”

How much information is associated with an event *k*?

$$I_k = log_2(1/p_k)$$



You expect more surprise

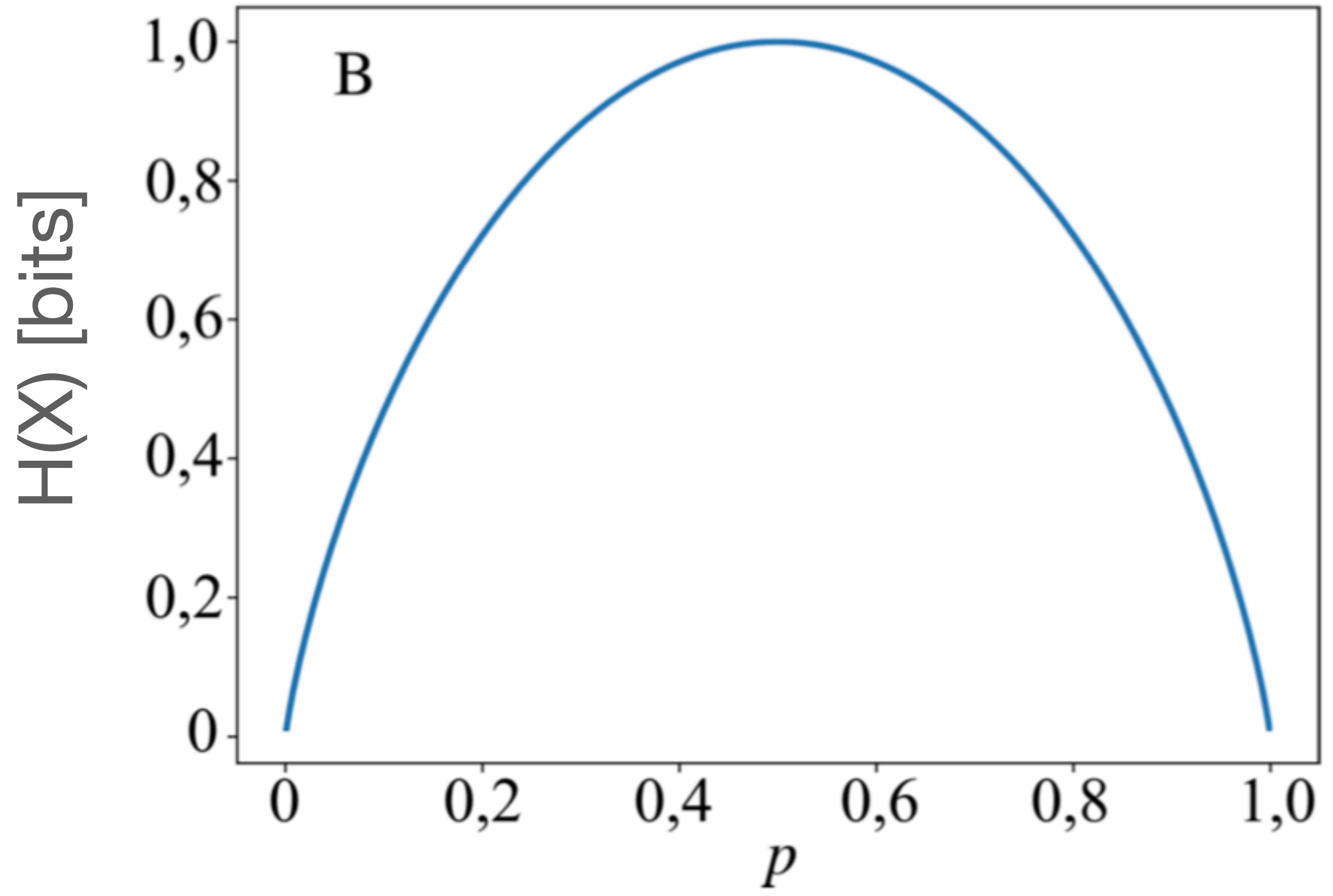
Entropy

Averaging over all states yields the famous Shannon entropy

$$H(X) = -\sum_{k=1}^{m} p_k \log(p_k) = \mathbb{E}[I_k].$$

How many questions we need to figure out the state of a system?

X has two states xk={x1,x2}
with probabilities pk={p,1-p}
-> e.g. toss of a coin



You can figure out the system with
only one question (1 bit)!

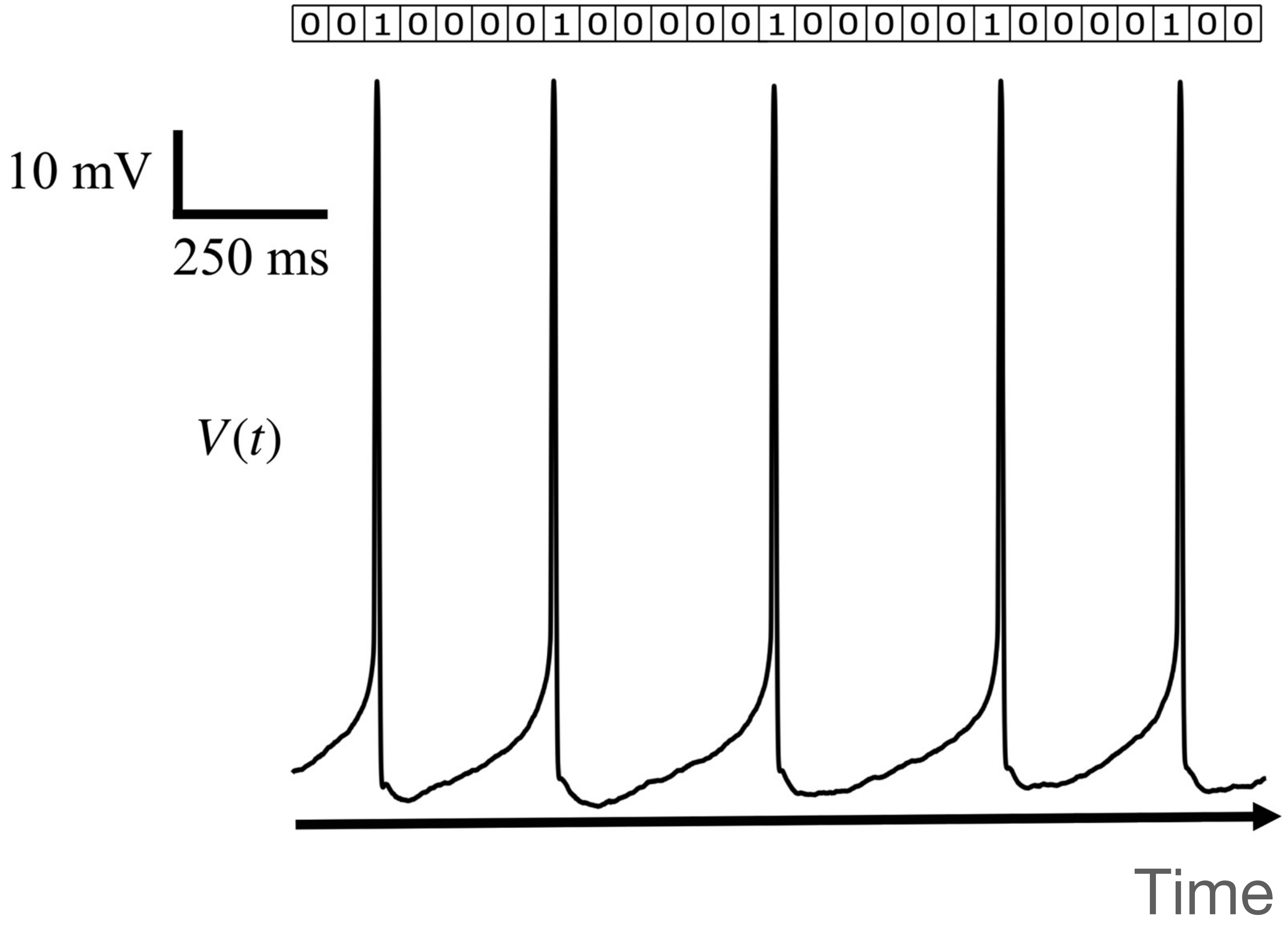
How to measure from spikes:

Spike-train

$$x_i(t) = \sum_{[t_i^f]} \delta(t - t_i^f),$$

$$P_{\rm d} = \frac{\sum_{i=1}^{N} x(t_i)}{N}, \quad \text{(firing)}$$

$$P_{\rm s} = 1 - P_{\rm d}, \quad \text{(silence)}$$



Information rate: divide by the time

Information density rate: divide by the time and by the number of neurons

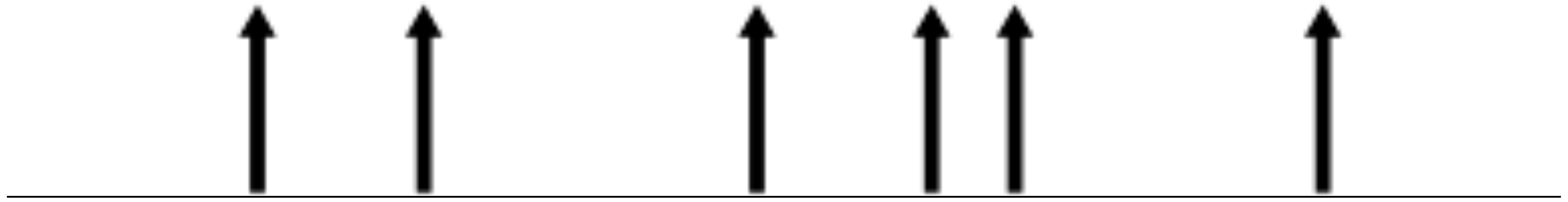
What's the maximum information a spike-train can carry?

(firing)

$$p_f = r_0 \Delta t$$

(silence)

$$p_s = 1 - p_f$$



Information rate of a Poisson process

$$R_{\text{info}} = \frac{r_0}{\ln(2)} \left(1 - \ln(r_0 \Delta t)\right)$$

Conditional entropy

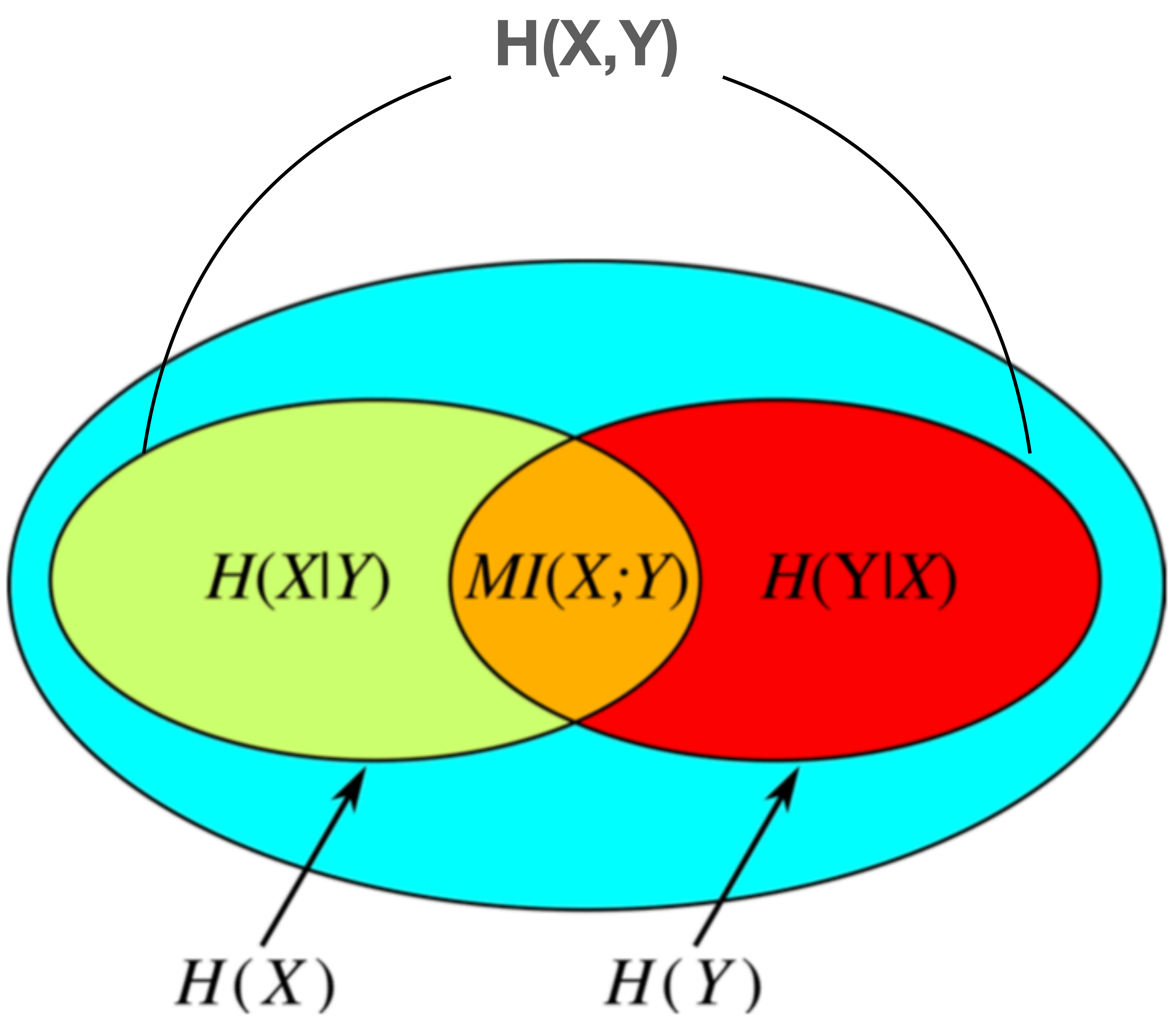
Uncertainty that **remains** in X after you observe Y

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

Mutual information

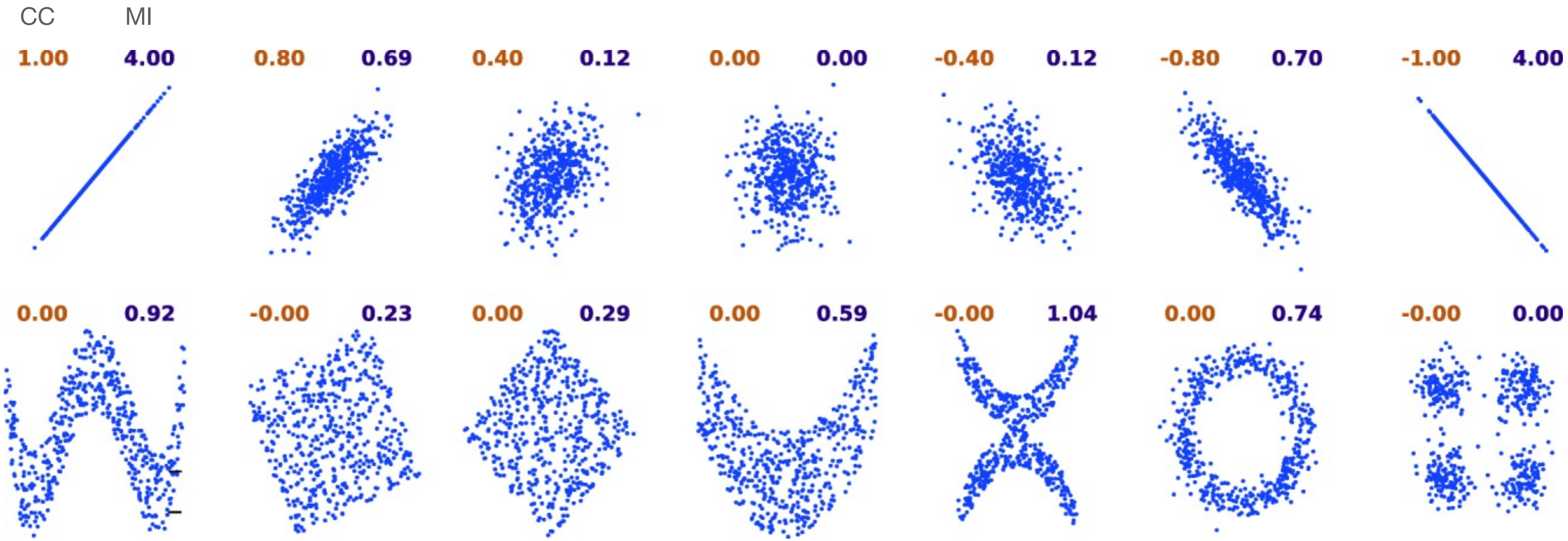
Uncertainty **reduced** in X after you observe Y

$$MI(X;Y) = H(X) - H(X|Y).$$



In other words: is the amount of information you obtain from X after you observe Y

Comparing MI with correlation coefficient (CC)

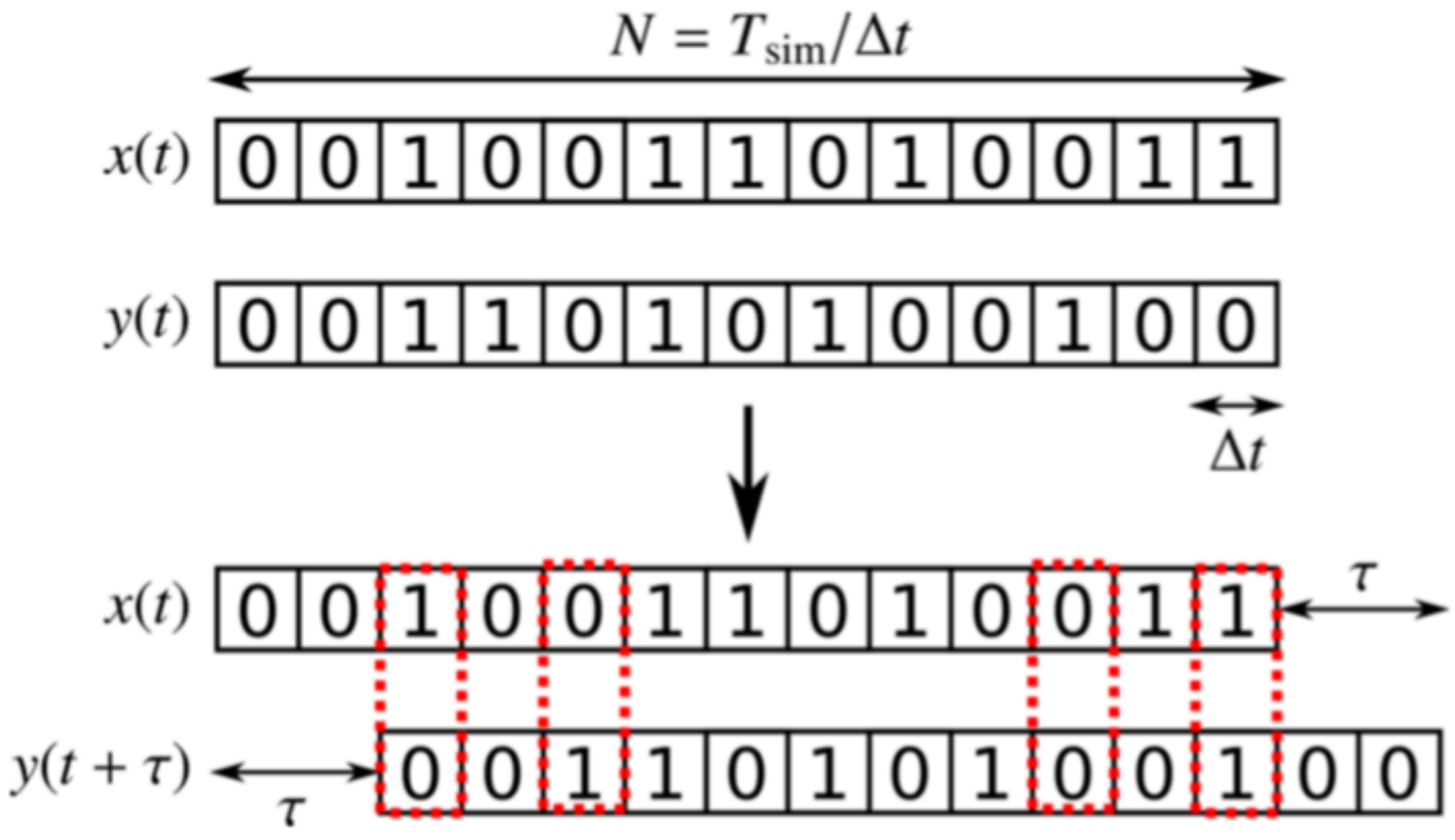


CC

- assumes there is a **linear** relationship;
- Can identify direction of the relationship;

MI

- Doesn't make any assumptions;
- MI is always positive;



H(X): information associated with X

H(Y): information associated with X

Joint entropy

H(X,Y): information associated with the pair X and Y

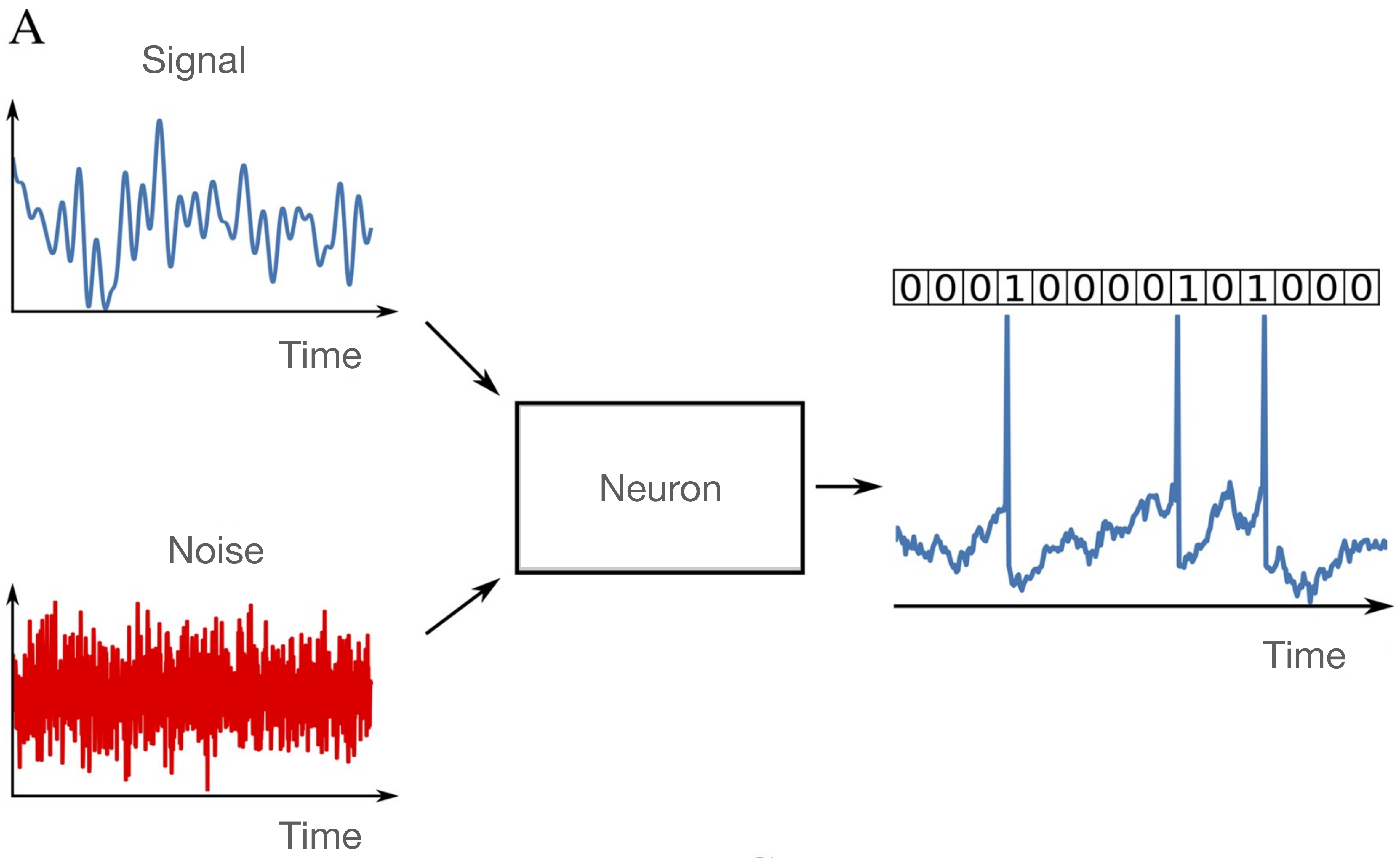
You can write with joint entropy:

MI(X;Y) = H(X) - H(X|Y)

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

MI(X;Y) = H(X) + H(Y) - H(X,Y)

Case of noisy neuron receiving stimulus

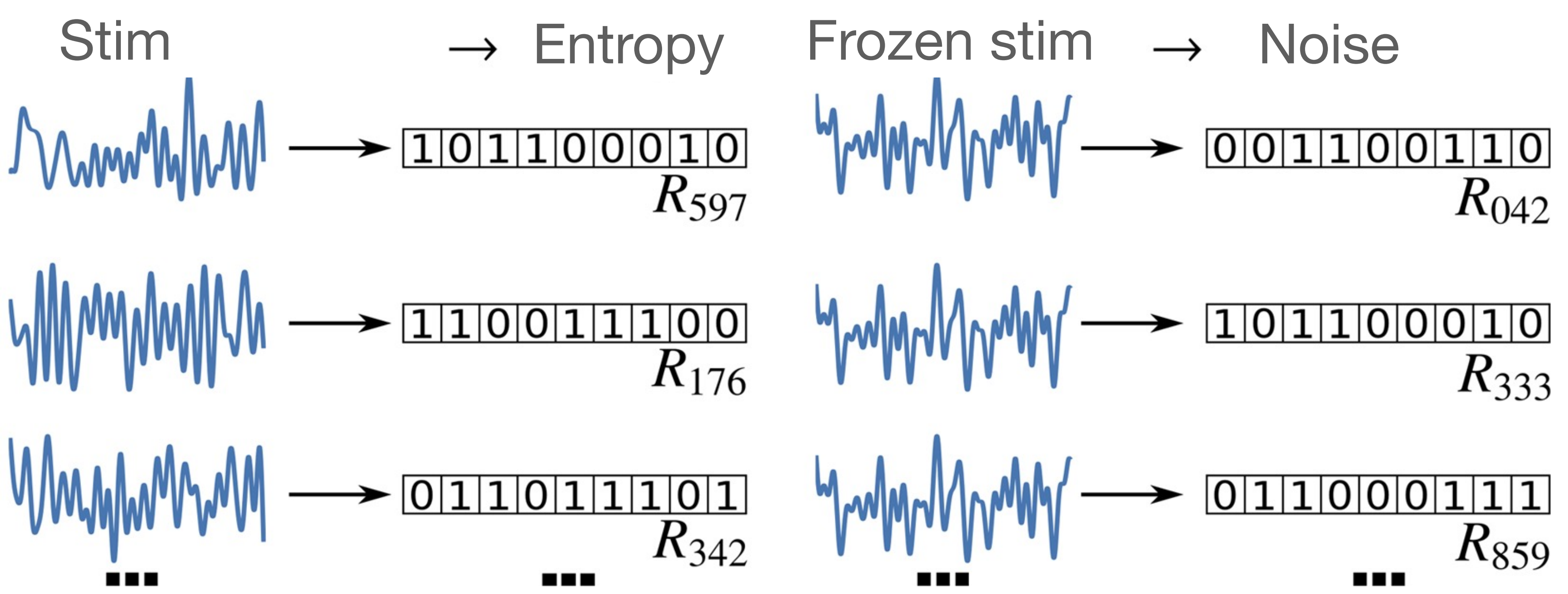


How to obtain MI(X;Y)?

1) direct method

Estimate p(X) and H(X)

Estimate p(X|signal) noise entropy H(noise)



MI(X;Y) = H(X) - H(noise)

Issue: computationally expensive!!
Demands a huge amount of data

2) lower bound mutual information (estimate of MI)

- Coherence:

Cross-spectrum (stimulus-spike train)

$$C(f) = \frac{|S_{x,s}(f)|^2}{S_{x,x}(f)S_{s,s}(f)}$$

Spike train power spectrum

Stimulus power spectrum

Coherence is the squared correlation coefficient in the frequency domain

- Lower Bound:

$$R_{\text{info,LB}} = -\int_0^{f_c} df \log_2[1 - C(f)]$$

Just **-log2(1-C(f))**
Notion of the quantity of information associated to each frequency

Real cells, comparison direct method and lower bound

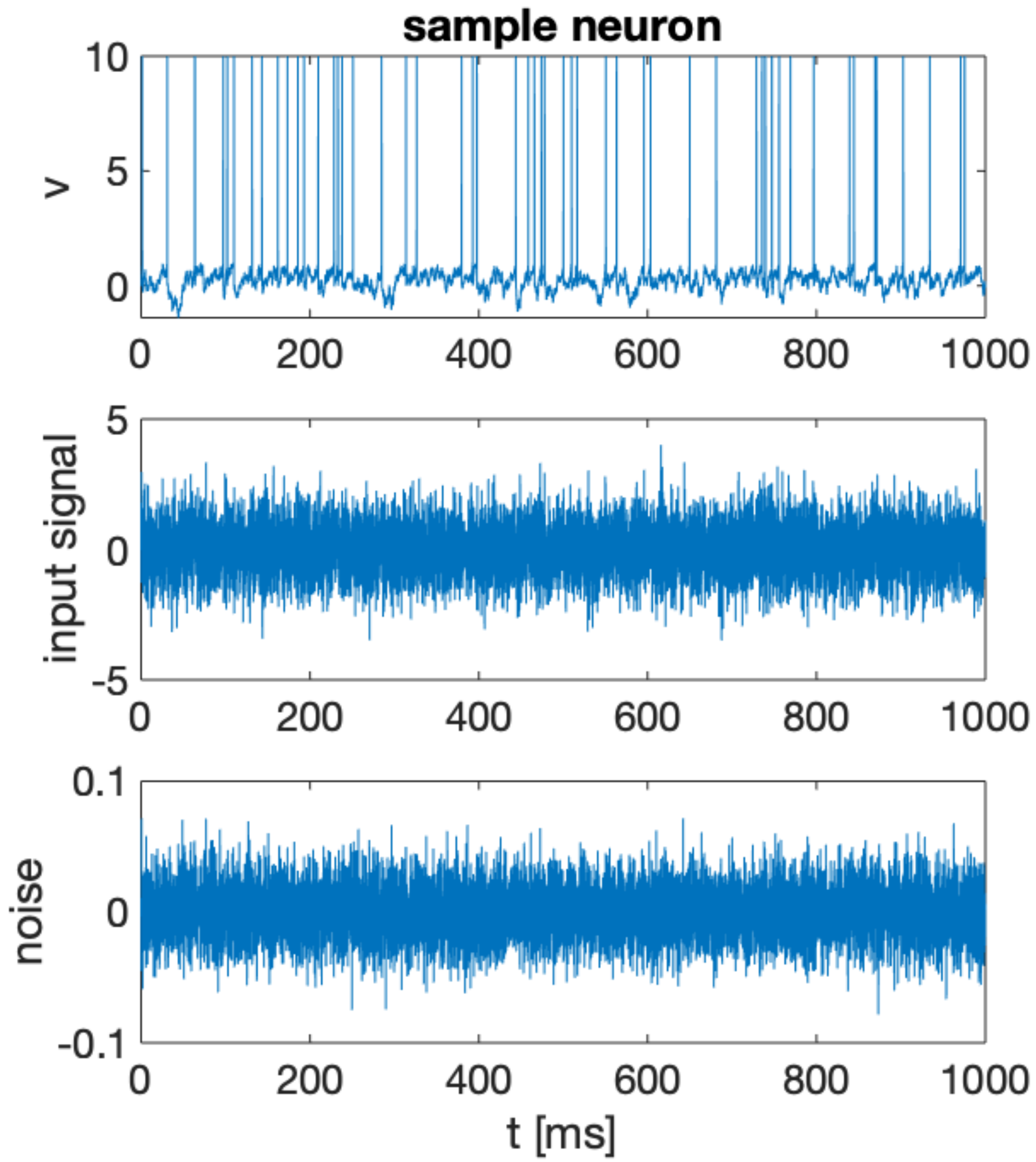
Preparation	Inf Method	Inf Rate	Inf Ratio	Reference
Fly H1	Rev Recon	64	2.5*	[1]
	Direct	81		[5]
Salamander Retina	Rev Recon	3.2	~3.0	[63]
	Direct	~9.6		[62]
Guinea Pig Retina	Rev Recon	3.3	4.6	[7]
	Direct	15.2		
Cat Retina	Rev Recon	61.1/62.2	1.4/1.8†	[61]
	Direct	82.5/109.2		
Cat Thalamus	Rev Recon	~1	~3.6	[16]
	Direct	3.6		[52]
Macaque MT	Rev Recon	5	2.5	[53]
	Direct	12.5		
Fish ELL	Rev Recon	14.7/25.2	1.6/2.1‡	[60]
	Direct	23.1/52.9		
Cricket Cercal INs	Rev Recon	41.1±7.8	2.3	Present Study
	Direct	96.7±19.8		

Aldworth et al.
PLoS Comp. Biol. (2011)

Practical example

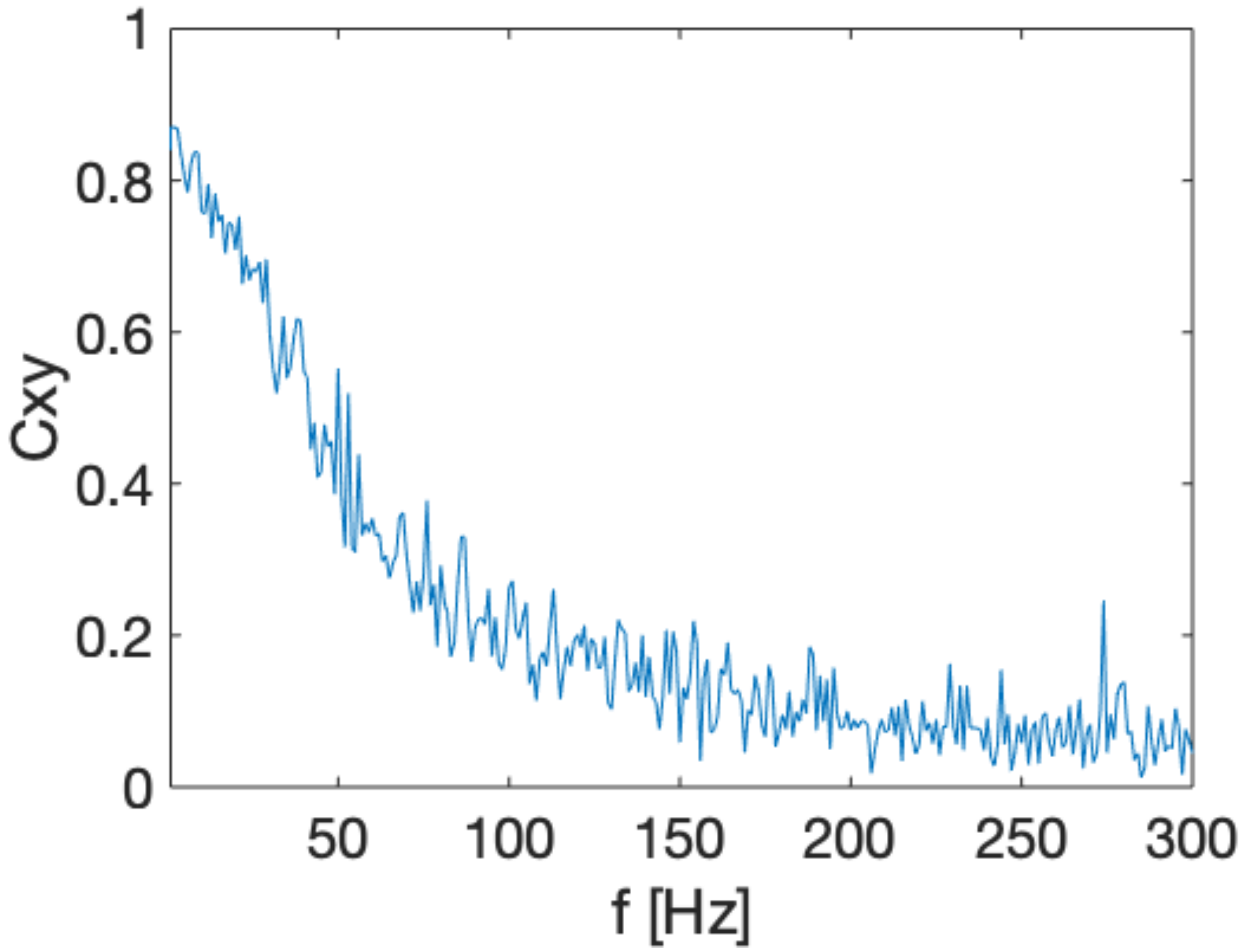
LIF neuron

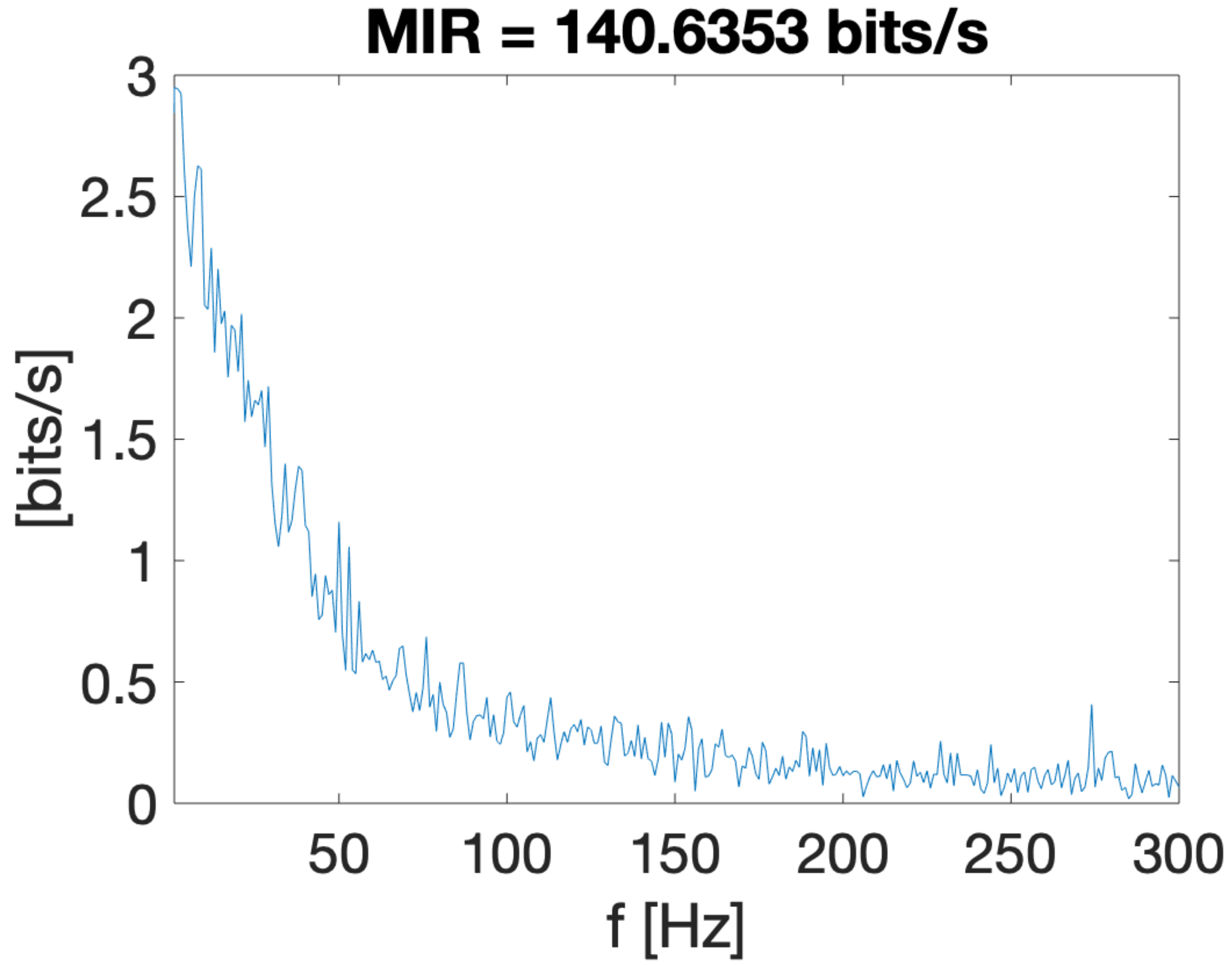
$$\tau_m \dot{v} = -v + \mu + \sqrt{2Dc}\xi_s(t) + \sqrt{2D(1-c)}\xi_n(t),$$



$$C_{xs}(f) = \lim_{T\to\infty} \frac{|\tilde{x}(f)\tilde{s}^*(f)|^2}{\langle|\tilde{x}(f)|^2\rangle\langle|\tilde{s}(f)|^2\rangle}$$

$$I_{lb} = -\int_0^{f_c} \log_2(1 - C_{xs}(f))df.$$





Continuous signals

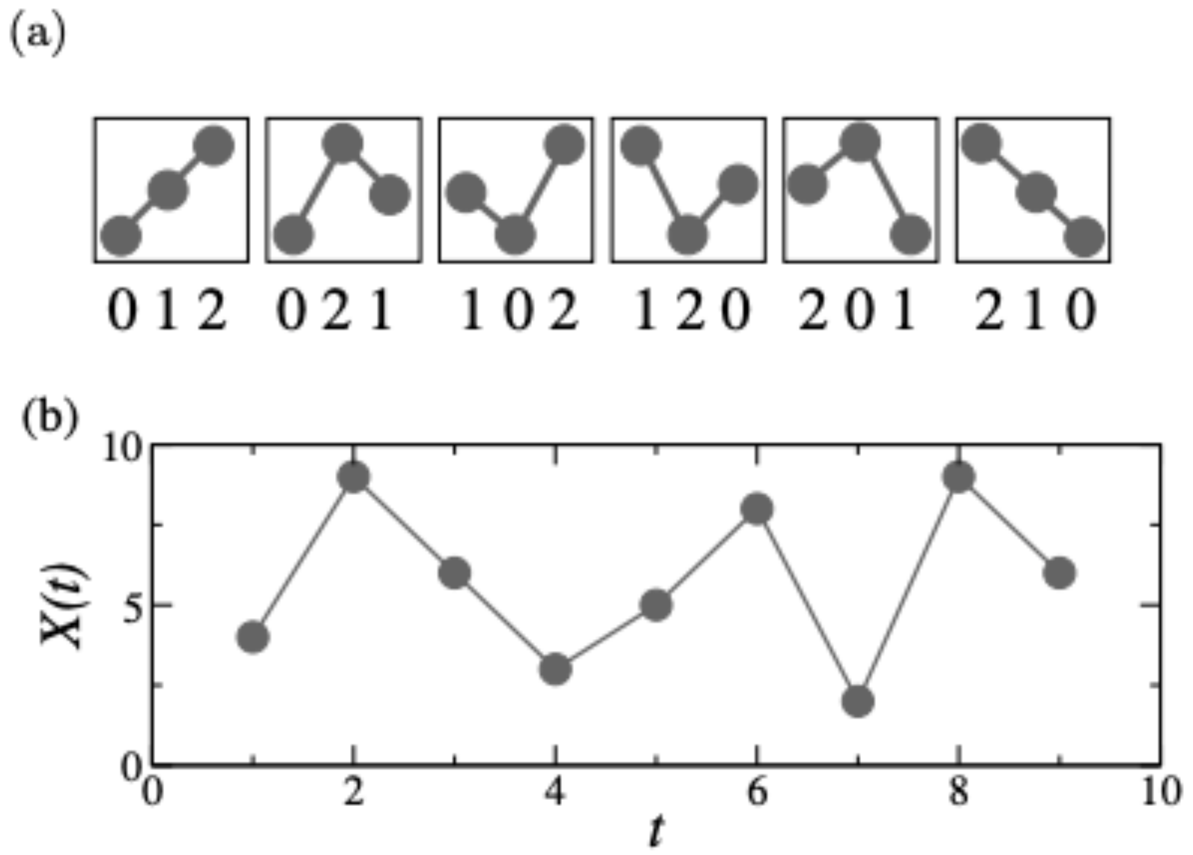
- Method 1

Measure with finite resolution

$$H = -\sum p(x_i)\Delta x \log_2(p(x_i)\Delta x)$$

- Method 2

Symbolic representation



Lofti et al. (2021). *Physical Review E, 103, 012415*

- Method 3

Use kernel estimators to estimate probabilities:

- Schreiber, T. (2000). *Physical review letters*, *85*(2), 461.

- Method 4

Use kNNs to estimate probabilities

KSG estimator: https://github.com/jlizier/jidt

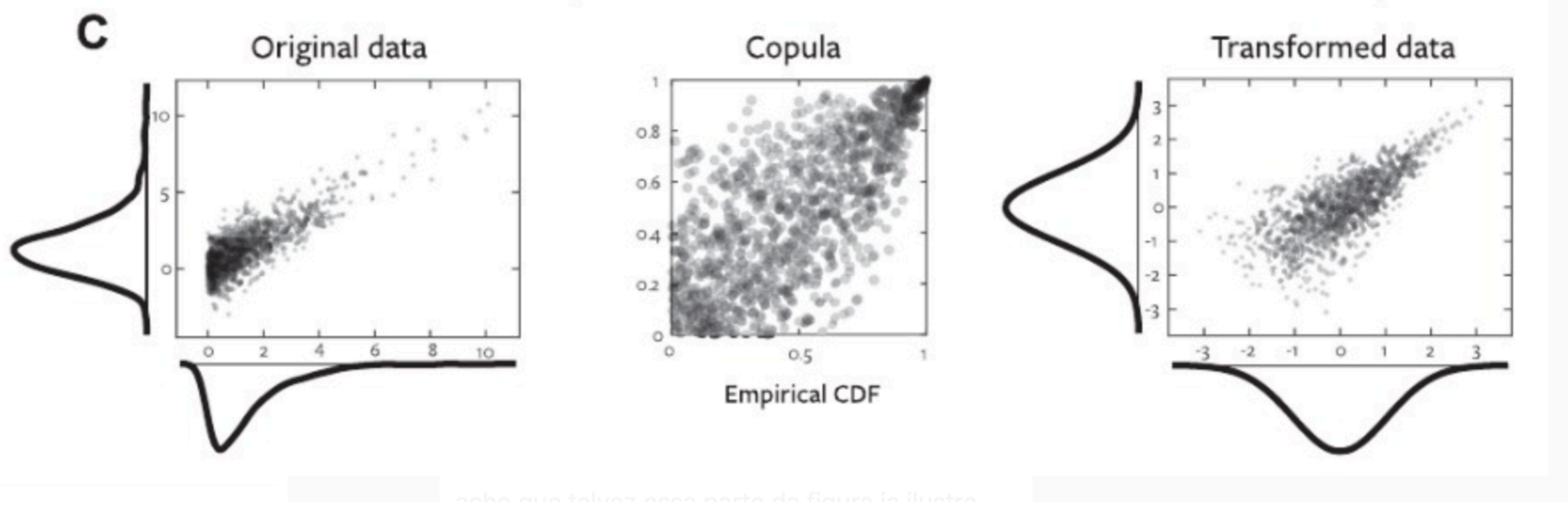
- Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). *Physical review E*, *69*(6), 066138.

- Method 5

Project variables into Gaussian space

- Ince, R. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). *Human brain mapping*, *38*(3), 1541-1573.

$$I(X;Y) = \frac{1}{2\ln 2}\ln\left[\frac{|\Sigma_X||\Sigma_Y|}{|\Sigma_{XY}|}\right]$$



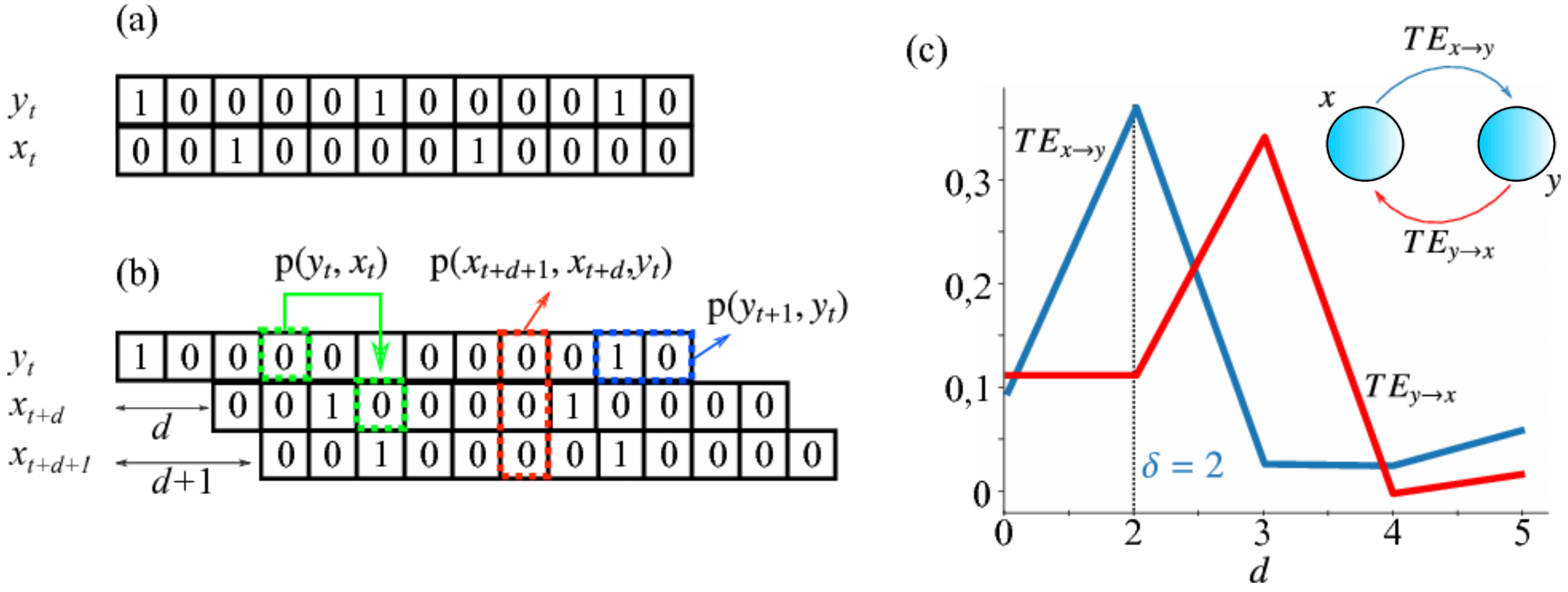
Transfer entropy

TE from y to x:

- knowing the **past** of y
- **uncertainty reduced** in the **future** of x

Flux of information!!

$$TE_{y\to x}(d) = \sum p(x_{t+1+d}, x_{t+d}, y_t) \log_2 \left(\frac{p(x_{t+1+d}, x_{t+d}, y_t)\, p(y_t)}{p(y_{t+1}, y_t)\, p(x_t, y_t)} \right)$$



Kullback-Leibler divergence

or relative entropy

- How much one distribution is different from a second
- Not a “distance” between probability distributions

$$D_{\mathrm{KL}}(P\|Q) = -\sum_{x\in\chi} p(x)\log q(x) + \sum_{x\in\chi} p(x)\log p(x)$$

$$= H_Q(P) - H(P)$$

Cross entropy

Measures the **information lost** when **Q is used to approximate P**

Non-symmetric measure

Widely used in data mining and data science

Interaction information

- Multivariate measure

$$II(X;Y;Z) = MI(X;Y;Z) - (MI(X;Z) + MI(Y;Z))$$

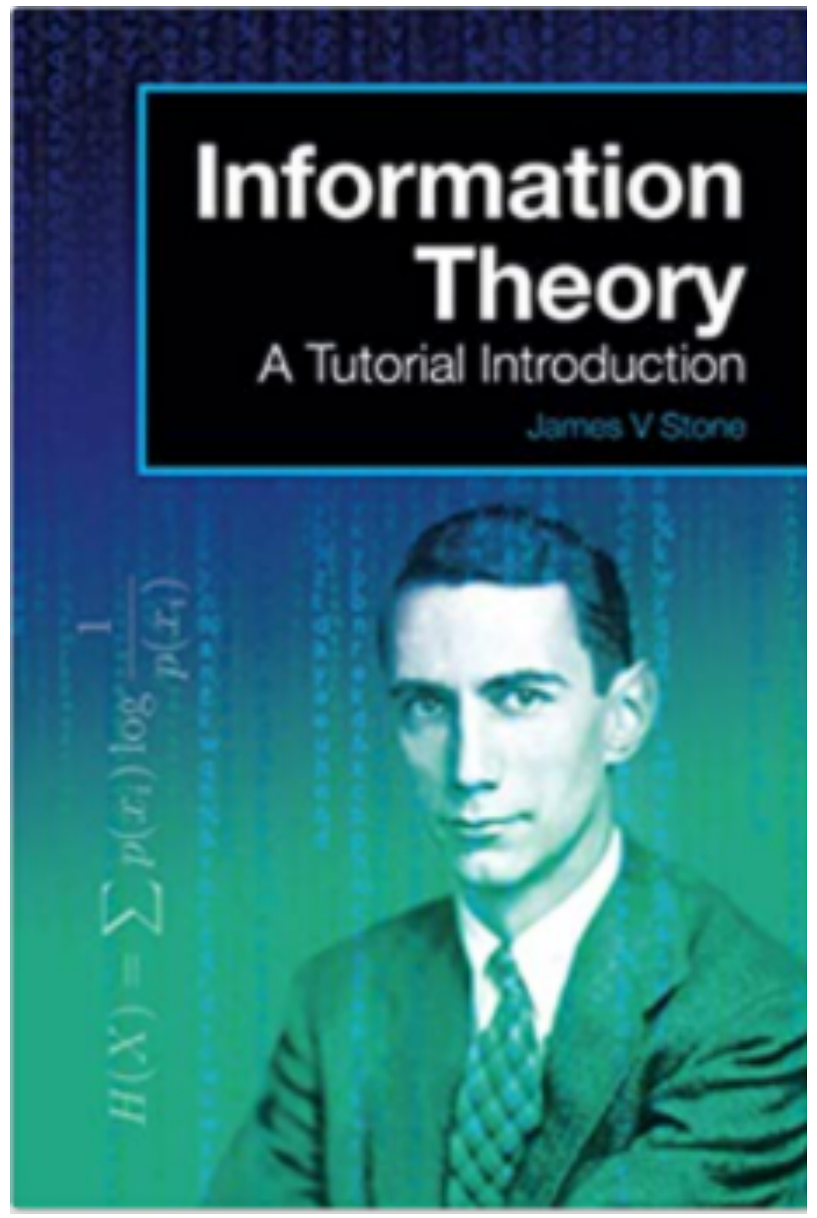
$II > 0$: Synergy

Observing X, Y, and Z has information not found when observing the pairs

$II < 0$: Redundancy

Observing the pairs has repeated information

References



- Stone, James V. "Information theory: a tutorial introduction." (2015).
- Schreiber, T., *Phys Rev Lett* **85**, 461 (2000).
- Kraskov, A., Stögbauer, H., and Grassberger, P., *Phys Rev E*, 69, 066138 (2004).
- Timme, N., Alford, W., Flecker, B. *et al., J Comput Neurosci* **36,** 119–140 (2014).
- D. Bernardi e B. Lindner, *J. Neurophysiol.* **113**, 1342 (2015).