Information Theory for Neuroscience

Measuring the entropy of visual cortices

Carmel Pe’er, Orion Yang, Shivang Rawat, Mathias Casiulis, Stefano Martiniani

Outline

1. Why entropy?
2. Algorithm
3. Discrete implementations
4. Continuous implementations
5. Neural data
6. Future work

Why Entropy?

What is entropy?

- Shannon entropy is a measure of “surprisal”
- Self entropy: or some random variable X following a distribution p

$$H(p,p) = -\sum_x p(x) \log p(x) \, .$$

- Cross entropy: for some random distribution p and an estimator distribution q

$$H(p,q) = -\sum_x p(x) \log q(x) \, .$$

Why Entropy?
Quantifying time reversal symmetry

- Kullback-Leibler divergence (KLD) - measure of distance from one probability distribution to a reference probability distribution.

$$D_{KL}(p \parallel q) = H(p, q) - H(p)$$

- Entropy production - the measure of the distance to time-reversal symmetry

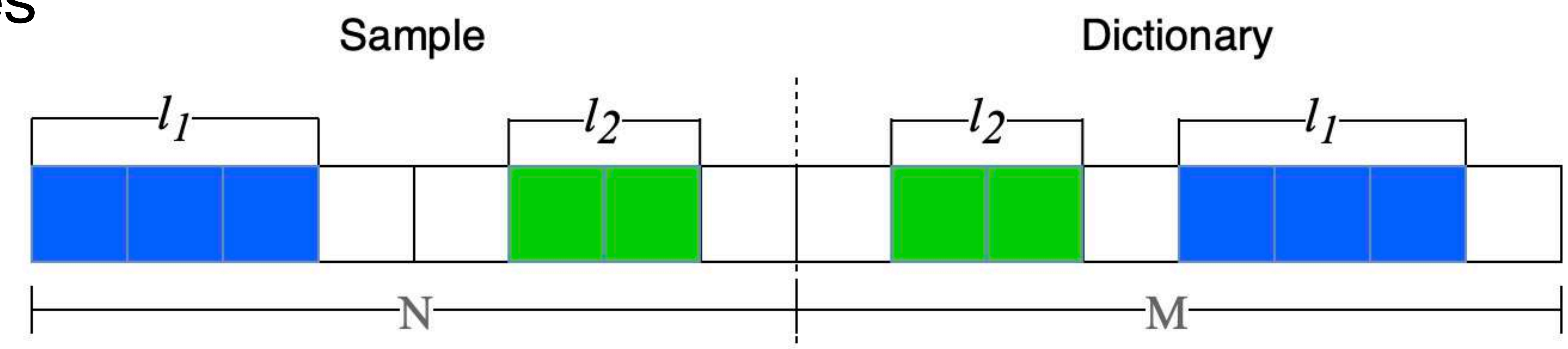
- For variable X' which is the flipped version of X and follows the distribution q,

$D_{KL}(p \parallel q) =$ Entropy production

Algorithm

Estimate the entropy when the distribution is unknown.

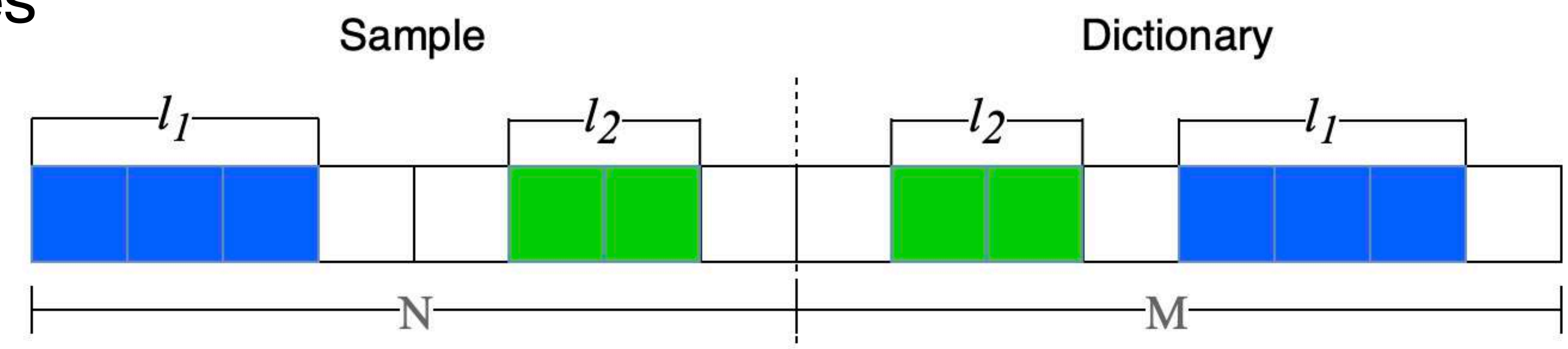
- Input series



Algorithm

Estimate the entropy when the distribution is unknown.

- Input series



- Entropy estimator

$$H = \frac{log_2 M}{< l >}$$

Algorithm

Pattern matching estimator

- Data compression algorithm:

Lempel-Ziv 77 Factorization (LZ77)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X =	a	b	c	a	b	a	b	a	b	a	b	a	b	a	b

$LZ^{77}(X)$ = [a,0], [b,0], [c,0], [0,2], [3,10]

Number of factors $C = 5$

Examples and derivations come from Stefano’s talk at Santa Fe, “The Other Side Of Entropy”

Algorithm

Pattern matching algorithm

$H \leq$ Information required to specify the factor of a finite sequence

$C =$ Number of factors

$$H \leq \frac{C}{N} log_2 N + \frac{1}{N} \sum_{i=1}^{C} log_2 l_i$$

Length of compressed data

Examples and derivations come from Stefano's talk at Santa Fe, "The Other Side Of Entropy"

Algorithm

Pattern matching algorithm

$H \leq$ Information required to specify the factor of a finite sequence

$C =$ Number of factors

$$H \leq \frac{C}{N} log_2 N + \frac{1}{N} \sum_{i=1}^{C} log_2 l_i$$

Length of compressed data

...

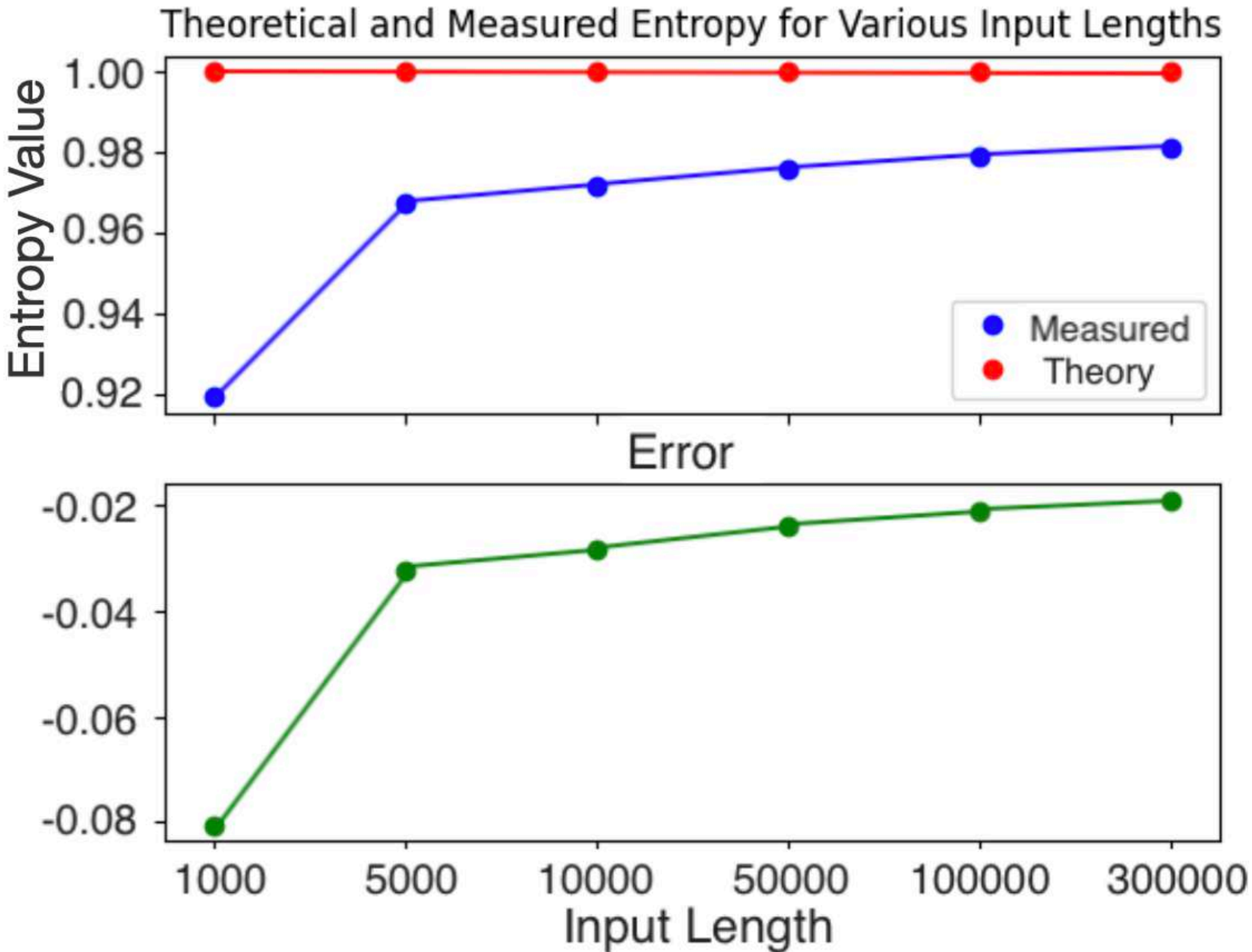
$$H \leq \frac{log_2 M}{< l >}$$

Pattern matching estimator

Examples and derivations come from Stefano's talk at Santa Fe, "The Other Side Of Entropy"

Discrete Implementations

Verification using Bernoulli series



As expected, the error of the estimator decreases as the input length grows.

Discrete Implementations

Introduce the Fast Fourier transform(FFT)

$$H = \frac{log_2 M}{< l >}$$

- Why Fast Fourier Transform?

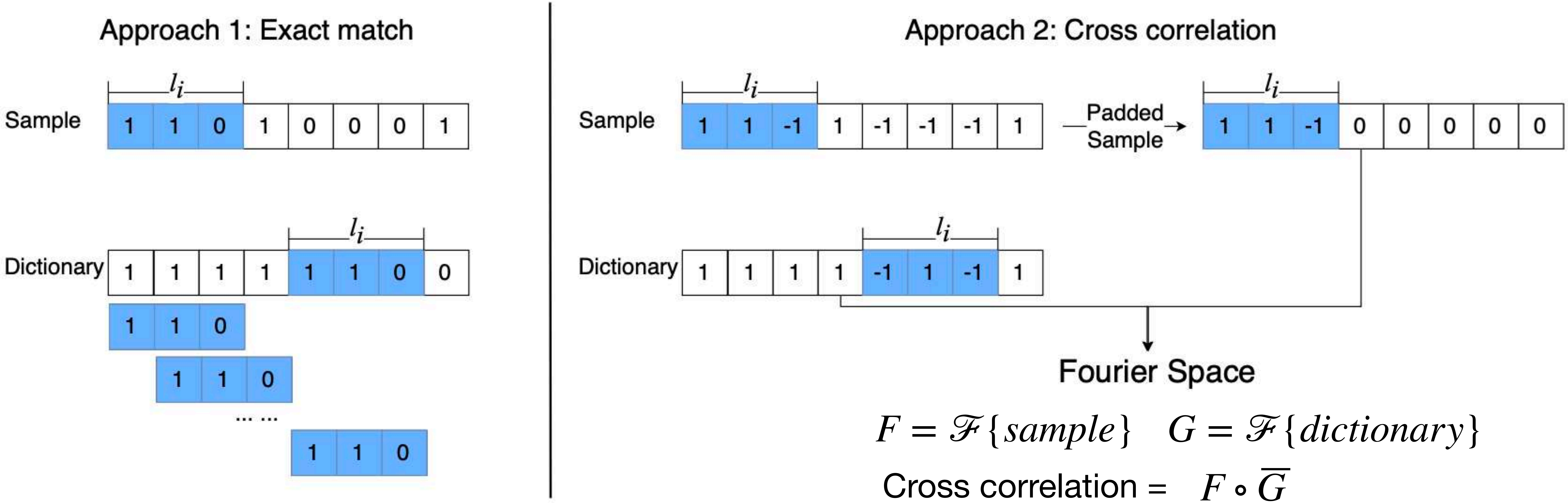
  It’s faster in higher dimensions.

  A good practice before moving to the continuous case

Discrete Implementations

Find the longest length of best matches l_i

$$H = \frac{log_2 M}{< l >}$$



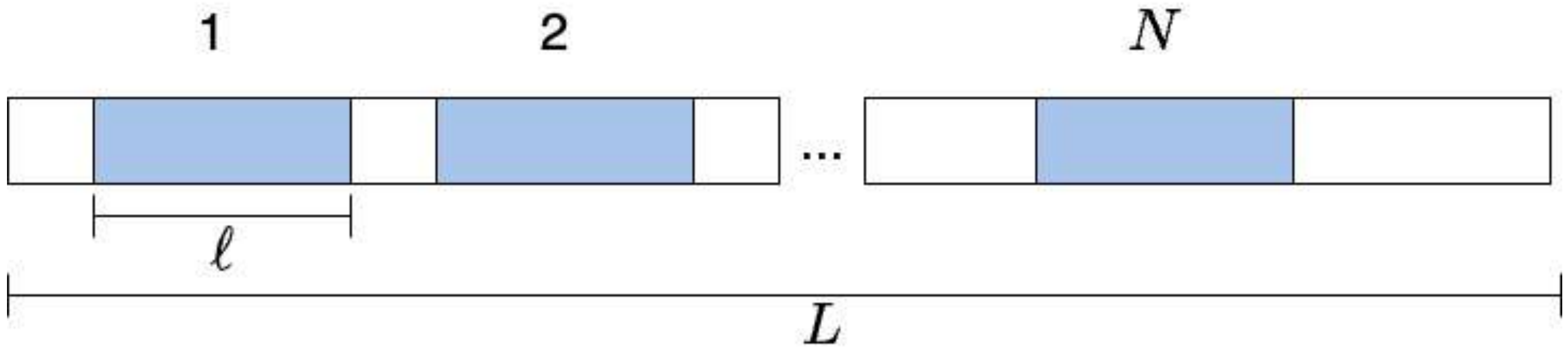
$$F = \mathscr{F}\{sample\} \quad G = \mathscr{F}\{dictionary\}$$

Cross correlation = $F \circ \overline{G}$

Continuous Implementations

Generating hard rods

- Non-overlapping constant-length rods

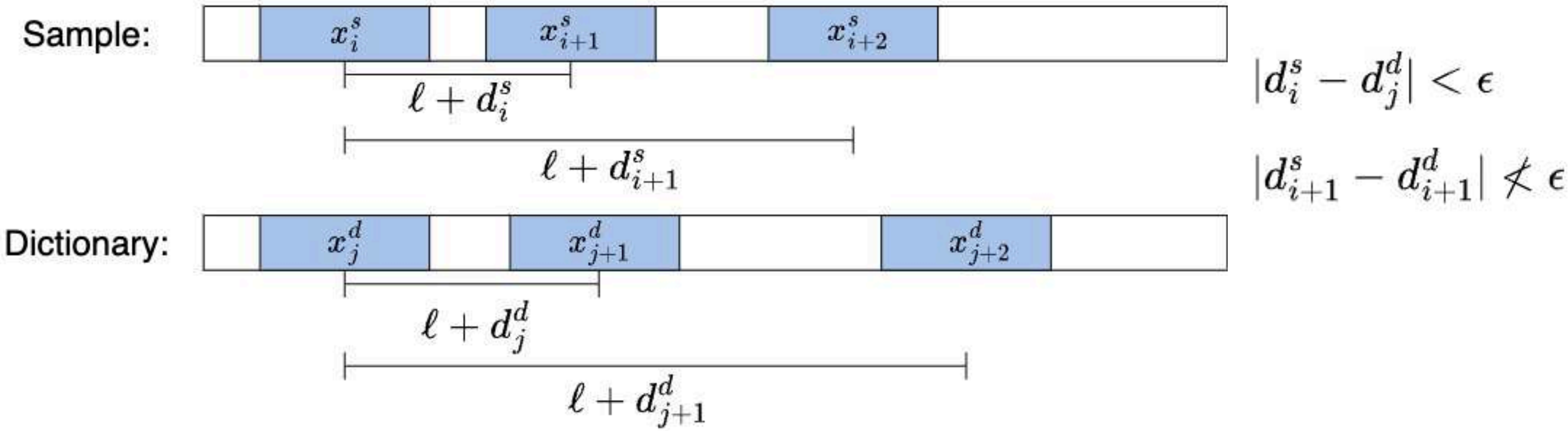


- Analytic entropy is known
- Discretization does not produce a good estimate

Continuous Implementations

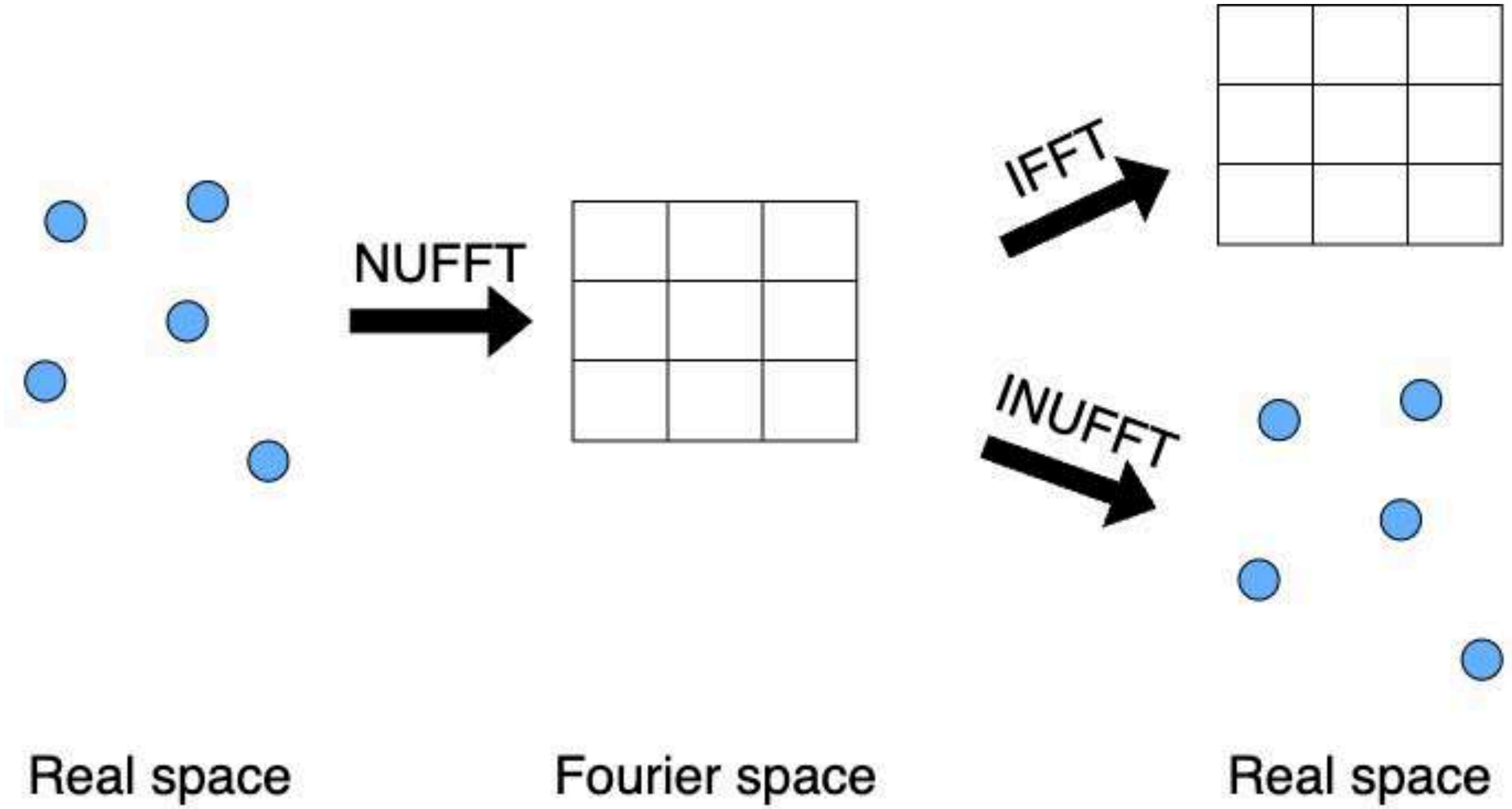
Real-space

- Used hard rods encoded by center location
- Match distances between rod centers with some error ϵ



Continuous Implementations
Fourier-space

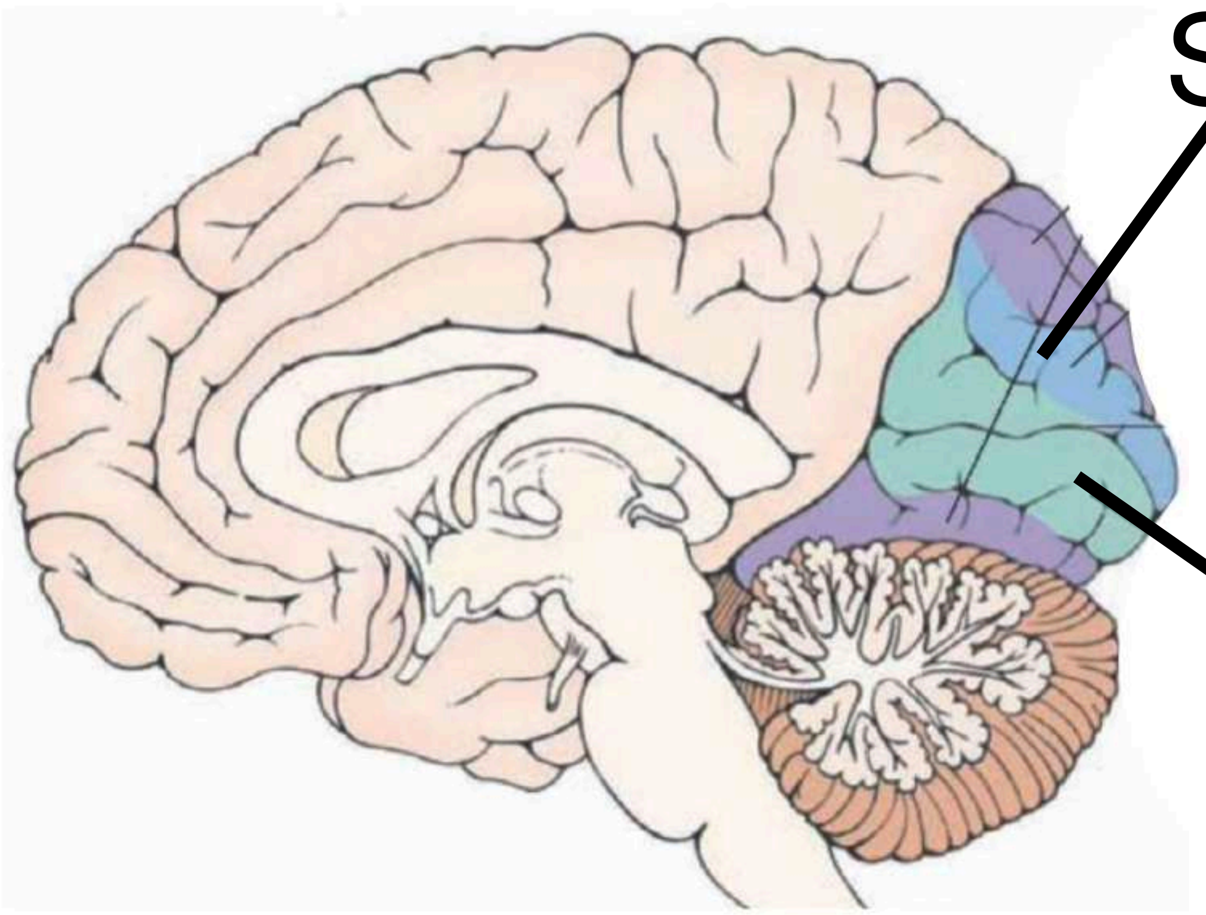
- Use non-uniform FFT (NUFFT)



Neural Data

Neuron activity in the visual cortex

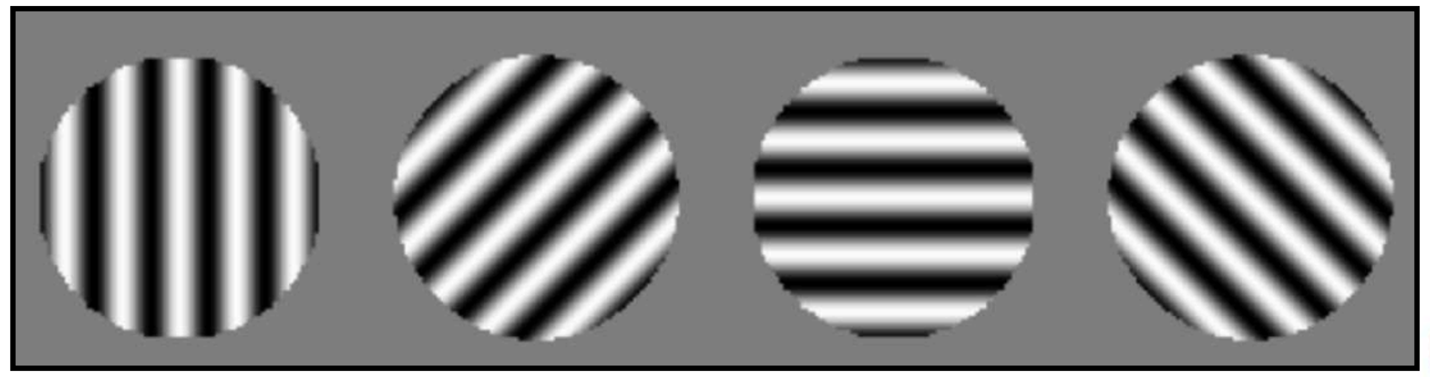
- Spike train of neurons: recording of neuron activations.
- Virtual cortex



Neural Data

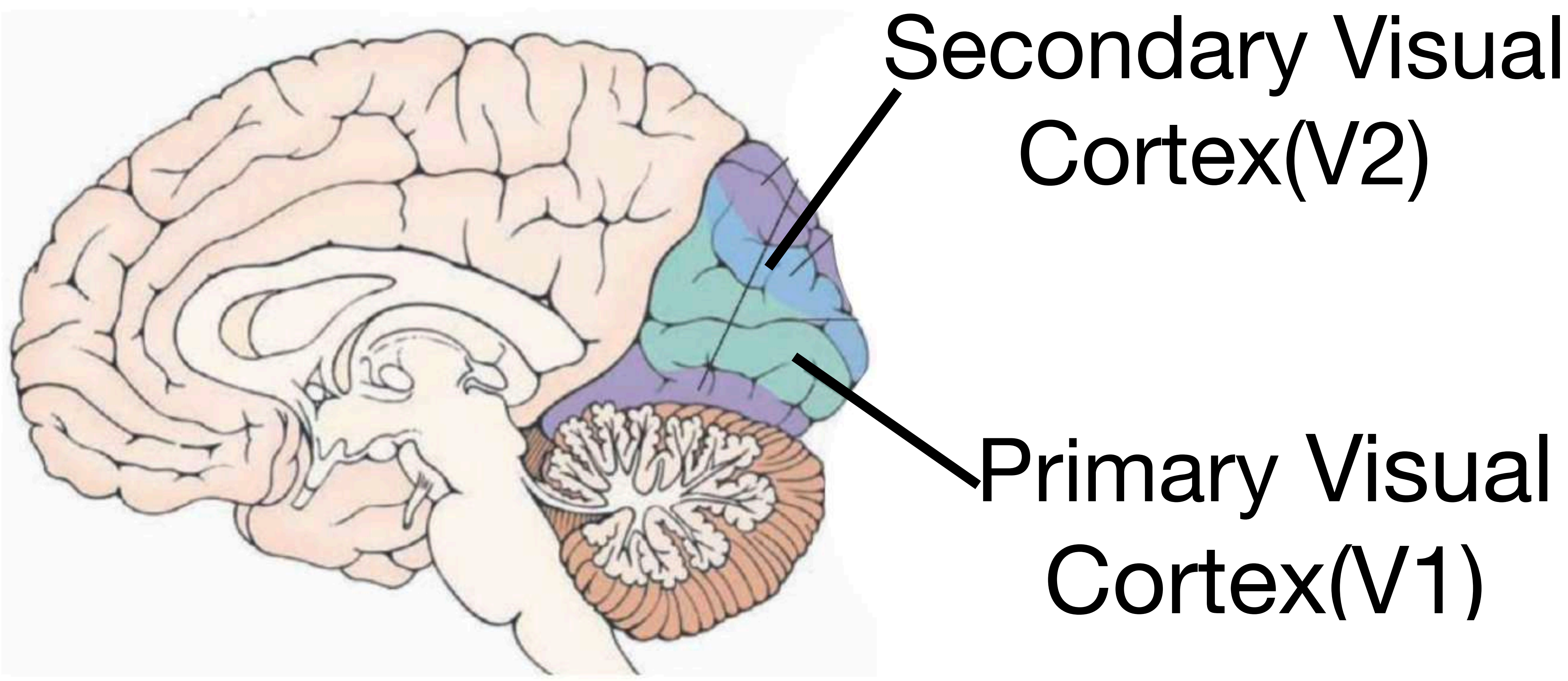
Neuron activity in the visual cortex

- Spike train of neurons: recording of neuron activations.
- Virtual cortex
- Experiment Setup



Oriented gratings

Macaques



Neural Data

Quantities measured

- Hypothesis

  Neural code is more time reversible in different brain areas, revealing the computing happening in the visual cortex.

Neural Data

Quantities measured

- Hypothesis

  Neural code is more time reversible in different brain areas, revealing the computing happening in the visual cortex.

- Entropy: $H(p) = -\sum_x p(x)\log p(x)\,.$
- Cross Entropy: $H(p, q) = -\sum_x p(x)\log q(x)\,.$
- Kullback-Leibler divergence (KLD):

$$D_{KL}(p \parallel q) = H(p, q) - H(p)\,.$$

  - p - distribution of spike train data
  - q - distribution of reversed spike train data

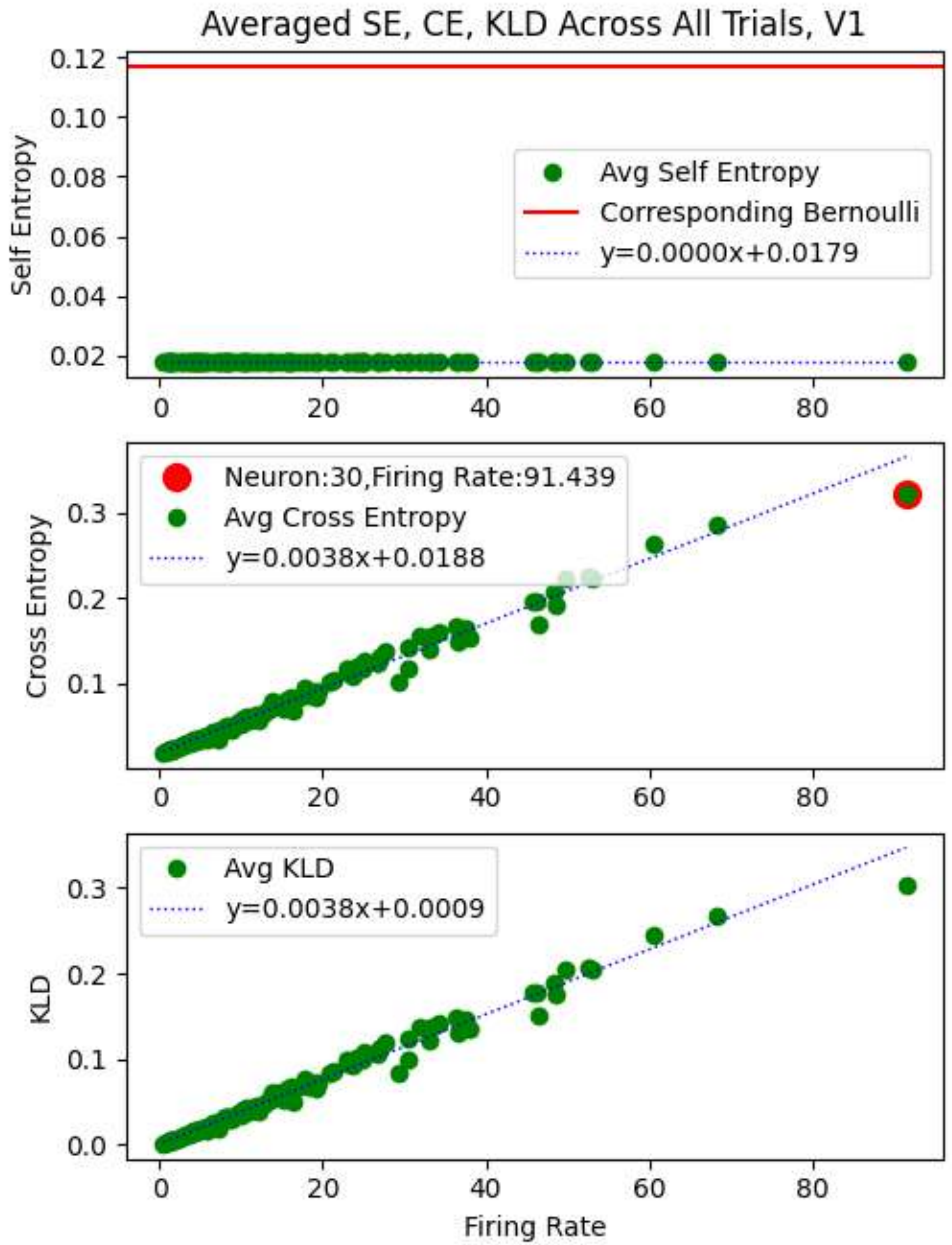
A measurement of time-reversal symmetry

Neural Data

Avg KLD vs. Firing Rate in V1

- Firing rate (Hz): $v = \frac{n_{sp}}{T}$
- In V1

  The KLD has a positive linear relationship with the firing rate of neurons.

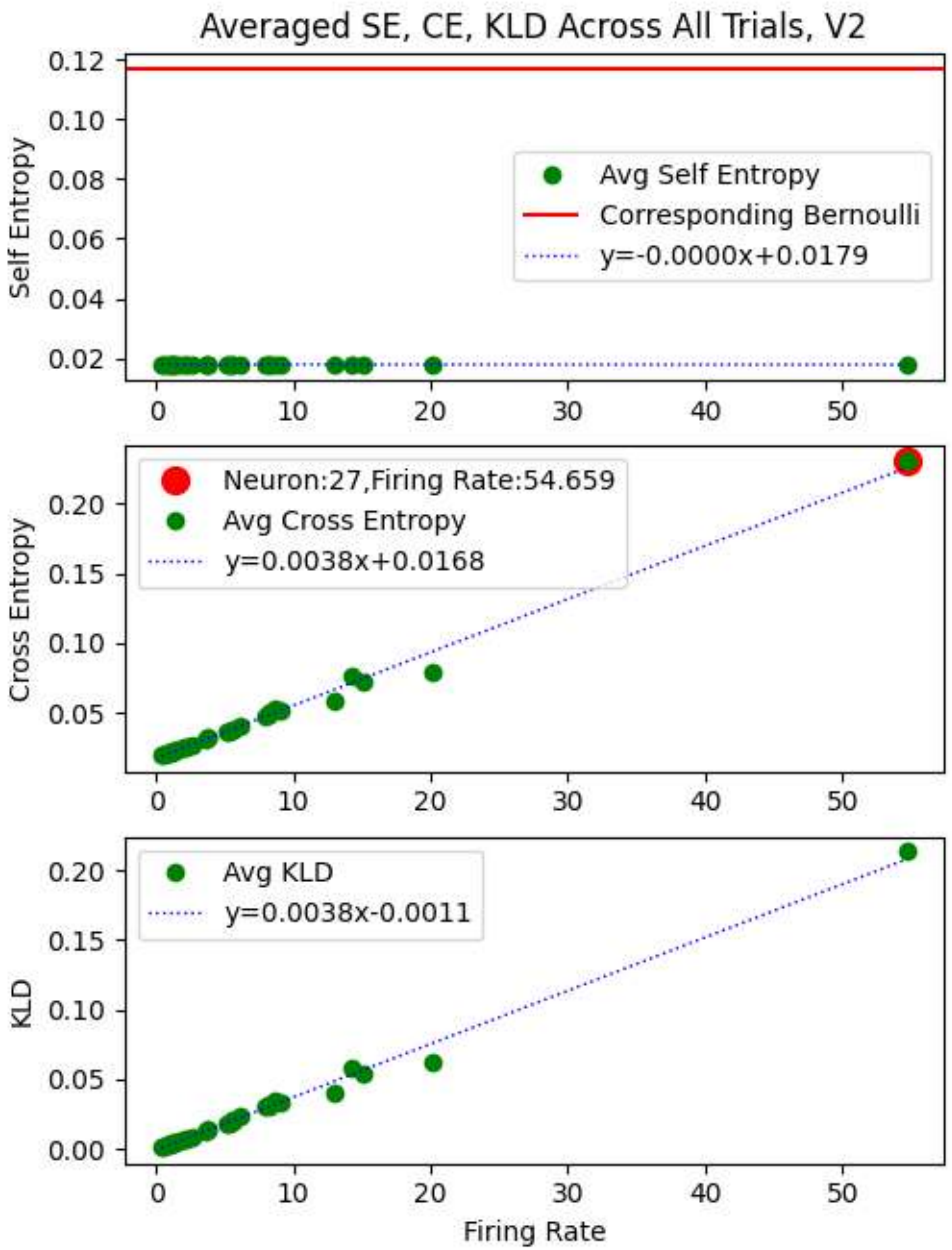


Neural Data

Avg KLD vs. Firing Rate in V2

- Firing rate (Hz): $v = \frac{n_{sp}}{T}$
- The same observation holds in V2:

  The KLD has a positive linear relationship with the firing rate of neurons.



Future work

- Neural data:
  - Apply algorithm to visual cortices of other animals
- Refine continuous implementation in real and Fourier space
- Extend to higher dimensions

Acknowledgments

People

Shivang Rawat

Mathias Casiulis

Stefano Martiniani

Institutions

NYU

NYU COURANT

SIMONS FOUNDATION

Why Entropy (Backup)

Cross-entropy

$$H(p, q) = -\sum_x p(x) \log q(x) .$$

- Measure of bits needed when the distribution is assumed to be q but is actually p
- Useful in machine learning context

Algorithm (Backup)

Pattern matching algorithm

- LZ77 is proved to asymptotically converge to the entropy

$H \leq$ Information required to specify the factor of a finite sequence

$$H \leq \frac{C}{N} logN + \frac{1}{N} \sum_{i=1}^{C} logl_i$$

$$= \frac{C}{N} logN + \frac{C}{N} < logl >$$

$$\leq \frac{C}{N} logN + \frac{C}{N} log < l >$$

$$H \leq \frac{logN}{< l >} + \frac{log < l >}{< l >}$$

$$H \leq \frac{logN}{< l >}$$

Pattern matching estimator

Examples and derivations come from Stefano's talk at Santa Fe, "The Other Side Of Entropy"

Phase Correlation (Backup)

Non-binary input

- Cross correlation is valid only for binary input
- Phase correlation:

$$F = \mathscr{F}\{sample\} \quad G = \mathscr{F}\{dictionary\}$$

$$\text{Phase correlation} = \frac{F \circ \overline{G}}{|F \circ \overline{G}|}$$

Algorithm (Backup)

Pattern matching estimator

- Data compression algorithm:
  - Lempel-Ziv 77 (LZ77) Factorization

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X =	a	b	c	a	b	a	b	a	b	a	b	a	b	a	b

LZ^{77}(X) = [a,0], [b,0], [c,0], [0,2], [3,10]

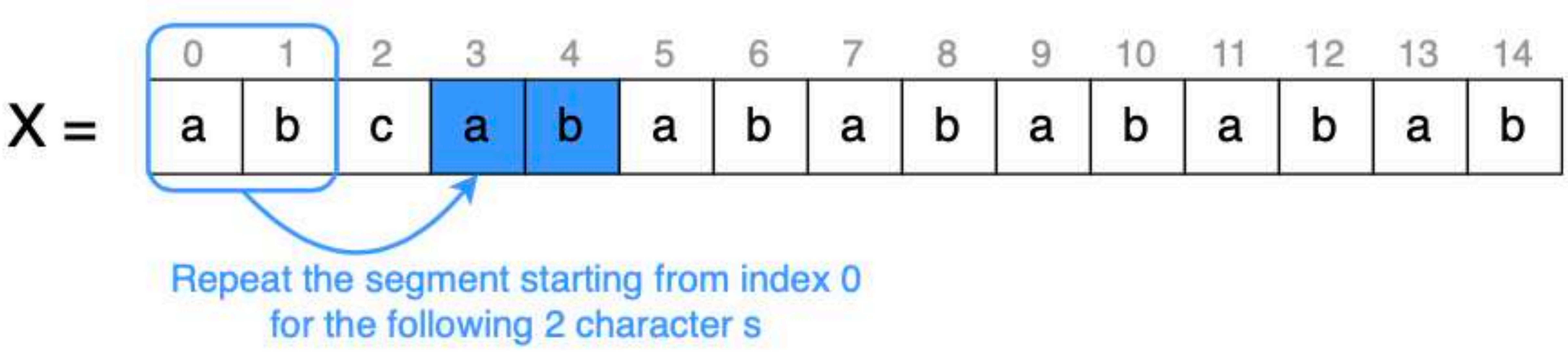
Number of factors $C = 5$

Examples and derivations come from Stefano's talk at Santa Fe, "The Other Side Of Entropy"

Algorithm (Backup)

Pattern matching estimator

- Data compression algorithm:
  - Lempel-Ziv 77 (LZ77) Factorization



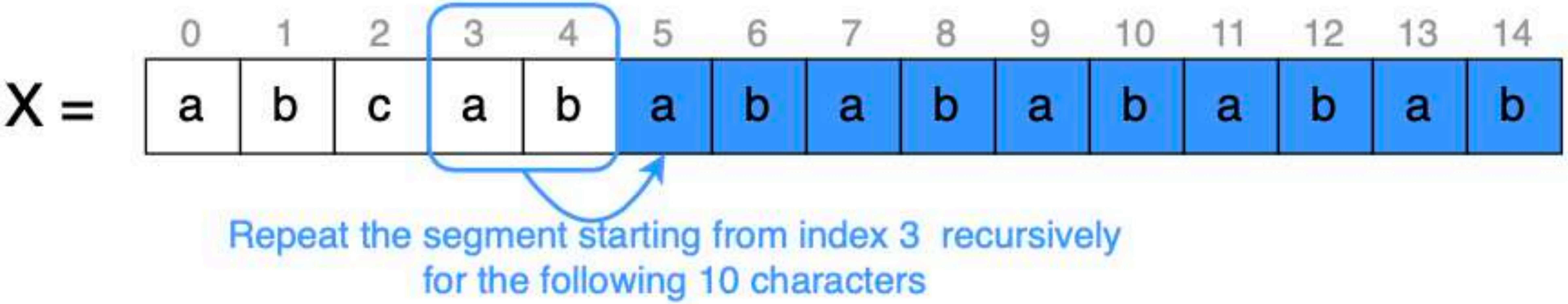
$LZ^{77}(X) = [a,0], [b,0], [c,0], [0,2], [3,10]$

Examples and derivations come from Stefano's talk at Santa Fe, "The Other Side Of Entropy"

Algorithm (Backup)

Pattern matching estimator

- Data compression algorithm:
  - Lempel-Ziv 77 (LZ77) Factorization

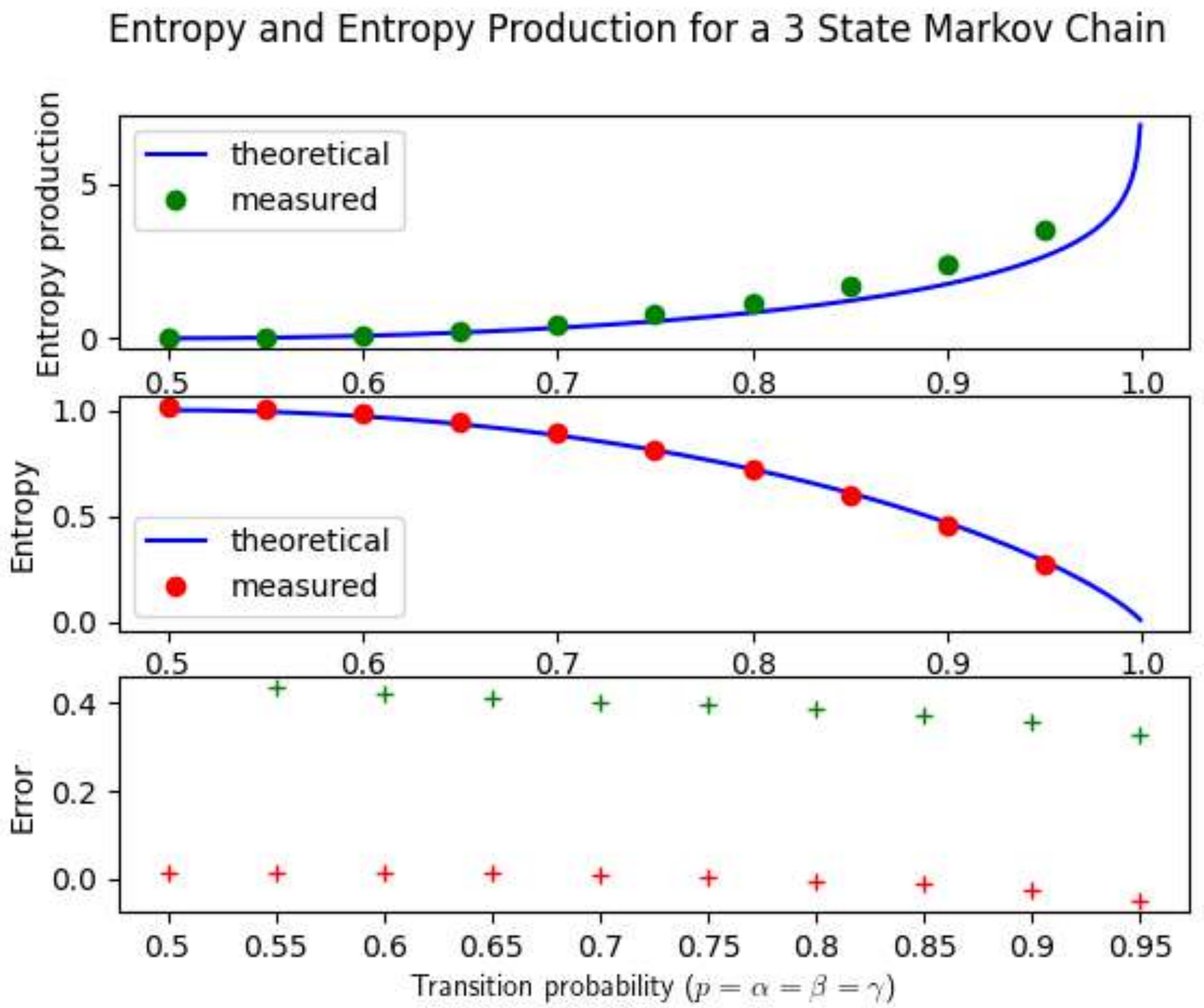


$LZ^{77}(X)$ = [a,0], [b,0], [c,0], [0,2], [3,10]

Examples and derivations come from Stefano's talk at Santa Fe, "The Other Side Of Entropy"

Discrete Implementations (Backup)

Verification using 3-state Markov chain



As expected, the error of the estimator increases as the bias grows.