Correlation in Neuronal Calcium Spiking: Quantification based on Empirical Mutual Information Rate

Sathish Ande\textsuperscript{1}, Srinivas Avasarala\textsuperscript{1}, Ajith Karunarathne\textsuperscript{2}, Lopamudra Giri\textsuperscript{1}, Soumya Jana\textsuperscript{1}

Abstract—Quantification of neuronal correlations in neuron population helps us to understand neural coding rules. Such quantification could also reveal how neurons encode information in normal and disease conditions like Alzheimer’s and Parkinson’s. While neurons communicate with each other by transmitting spikes, there would be change in calcium concentration within the neurons inherently. Accordingly, there would be correlations in calcium spike trains and they could have heterogeneous memory structures. In this context, estimation of mutual information rate in calcium spike trains assumes primary significance. However, such estimation is difficult with available methods which would consider longer blocks for convergence without noticing that neuronal information changes in short time windows. Against this backdrop, we propose a faster method which exploits the memory structures in pair of calcium spike trains to quantify mutual information shared between them. Our method has shown superior performance with example Markov processes as well as experimental spike trains. Such mutual information rate analysis could be used to identify signatures of neuronal behaviour in large populations in normal and abnormal conditions.

Index Terms—Calcium imaging; Spike train; Empirical probability; Mutual information rate; Correlation.

I. INTRODUCTION

Neurons are known to encode and process stimulus information in series of action potentials called spike trains. Additionally, such spike trains are shared between neurons for performing high level functions like memory and learning [1]. While doing so, there would be change in calcium ion concentration within each neuron which in turn cause such change in neighbouring neurons in population [2], [3]. Subsequently, quantification of such correlation based calcium spiking between neurons in terms of mutual information assumes primary significance. Early methods in this direction were based on correlation [4]. Such methods quantify the linear relation between synchronous spike trains but do not reveal how much information shared between them. Furthermore, spike trains could be non-synchronous. In this context, the methods based upon Lempel-Ziv (LZ) algorithm were suggested to estimate the mutual information between spike trains, which are either synchronous or asynchronous [5], [6]. However, these methods have slow convergence i.e., they need large length spike trains for estimation, with out identifying the fact that the neuronal information changes in short time windows [7]. In this backdrop, we propose a fast yet accurate empirical method to estimate the mutual information rate by exploiting memory structures in calcium spike trains [8], [9]. Our method has shown superiority over existing LZ algorithms with example Markov processes and with heterogeneous calcium responses, recorded from hippocampal region of the brain, where memory and learning tasks performed.

The rest of this paper is organized as follows. Section II discusses imaging of neuronal calcium spiking in hippocampal region and deconvolution algorithm for inferring spike trains and introduces the concept of mutual information rate of stochastic processes, and states LZ algorithms and proposed empirical mutual information rate estimator. Further, Section III describes the results illustrating superiority of the proposed method. Finally, Section IV concludes the paper.

II. MATERIALS AND METHODS

The workflow of this paper is shown in Fig. 2 and elaborated in the following.

A. Data Collection and Spike Train Inference

The dataset consists of calcium spiking trains from dissociated hippocampal neurons culture from 1 day postnatal Sprague-Dawley rats. Calcium imaging was performed on 7-th day in vitro (7 DIV) using Leica DMI6000B confocal microscope with Yokogawa CSU-X1 spinning-disk units.
Cells were incubated with 1 μM Fluo-4 (Molecular Probes) dye in 30 min in Hank’s Balanced Salt Solution (HBSS) followed by three washes, each for a duration of 10 min. An excitation and emission wavelength of 488 nm and 510 nm were used for Fluo-4 imaging [10]. Data acquisition was set every 1s, but the gap between two images were found to be from 0.8 s to 1 s due to inherent variabilities arising from the instrument. In order to maintain the temperature at 37°C, it was kept at 37°C for time course of spatially resolved Fluo-4 fluorescence intensity in neuron populations was obtained using Andor software. For the present study, we consider 8 neurons with indices 1–8 from population of 21 neurons for analysis (refer Fig.3 for time course of calcium responses for aforementioned neurons). We inferred spike train for each neuron from its fluorescent time signal using a deconvolution algorithm and those were binarized by adaptive threshold [8], [11]. Such binary spike trains were used for quantification analysis.

B. Mathematical Preliminaries

At this point, we introduce the concept of mutual information rate, and state existing mutual information rate estimators and proposed estimator.

1) Entropy rate: Entropy, a measure of uncertainty in random variable \( X \) with probability mass function \( p_X \) defined on alphabet \( \mathcal{X} \) is defined by \( H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) \). The entropy per symbol, entropy rate, of sequence of random variables \( X_1, X_2, \ldots, X_n, \ldots \) is defined as

\[
\bar{H}(X) = \lim_{n \to \infty} \frac{1}{n} H(X^n),
\]

\( X^n = (X_1, X_2, \ldots, X_n) \). Thus \( \bar{H}(X) \) measures the average asymptotic uncertainty per symbol [12].

2) Mutual information rate: Mutual information \( I(X;Y) \), a measure of mutual dependency between two random variables \( X, Y \) with probability mass functions \( p_X, p_Y \) respectively is defined as

\[
I(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}
\]

(2)

where \( p_{XY}(x,y) \) is joint probability mass function of \( X, Y \). It can be written in-terms of marginal and joint entropy’s of \( X, Y \) as \( I(X;Y) = H(X) + H(Y) - H(X,Y) \).

Further, the mutual information rate \( \bar{I}(X;Y) \) of a sequence of random variables \( X_1, X_2, \ldots, X_n, \ldots \) and \( Y_1, Y_2, \ldots, Y_n, \ldots \) is defined as

\[
\bar{I}(X;Y) = \lim_{n \to \infty} \frac{1}{n} I(X^n;Y^n),
\]

(3)

where \( X^n = X^n_1 = (X_1, X_2, \ldots, X_n), Y^n = Y^n_1 = (Y_1, Y_2, \ldots, Y_n) \). Thus \( \bar{I}(X;Y) \) measures the asymptotic reduction in uncertainty per symbol [12]. It can be written in-terms of marginal and joint entropy rates of \( X^n, Y^n \) as follows.

\[
\bar{I}(X;Y) = \lim_{n \to \infty} \frac{1}{n} I(X^n;Y^n) \]

\[
= \lim_{n \to \infty} \left( \frac{1}{n} H(X^n) + \frac{1}{n} H(Y^n) - \frac{1}{n} H(X^n, Y^n) \right)
\]

(4)

3) Lempel-Ziv estimators of mutual information rate: Versions of Lempel-Ziv (LZ) algorithms have been proposed for data compression, which find the complexity \( K_{LZ} \) of the sequence \( X^n \), a measure of entropy rate.

LZ-78 algorithm: LZ-78 is dictionary based scheme that converts a given sequence into unforeseen sub-blocks to the dictionary.

Denoting by \( b(X^n) \) the number of sub-blocks of a sequence \( X^n \) present in the data, the LZ-78 complexity, the average number of bits per symbol, is defined as [12]

\[
K_{LZ78}(X^n) = \frac{b(X^n) \log (b(X^n))}{n}.
\]

(5)

So, the corresponding entropy rate estimate is \( \hat{H}_{LZ78}(X^n) = \lim_{n \to \infty} K_{LZ78}(X^n) \), where one practically uses large \( n \) in place of the limit. The mutual LZ-78 complexity \( K_{LZ78}(X^n, Y^n) \) between sequences \( X^n, Y^n \) is given as [5]

\[
K_{LZ78}(X^n, Y^n) = \frac{b(X^n, Y^n) \log (b(X^n, Y^n))}{n}.
\]

(6)

where \( b(X^n, Y^n) \) is number of sub-blocks in joint dictionary of \( X^n, Y^n \). The mutual information rate is given as

\[
\bar{I}_{LZ78}(X;Y) = \lim_{n \to \infty} (K_{LZ78}(X^n) + K_{LZ78}(Y^n) - K_{LZ78}(X^n, Y^n))
\]

(7)

Sliding Window Lempel Ziv (SWLZ) algorithm: SWLZ converts the given sequence into sub-blocks such that new block is seen at each position \( i \) is not appear in previous \( i - 1 \) symbols by sliding a window of length, \( L_i \). The SWLZ
Dictionary for $X^n$ | Dictionary for $Y^n$ | Dictionary for $(X^n, Y^n)$
---|---|---
LZ-78 | 0—1—00—11—001—01—110—1 | 01—10—00—0111—1000—0110—011010—1100—110
SWLZ | 0—1—00—01—11—10010—0010—0111—111—1101—101—011—0010100—00100—0100—100 | 01—10—00—0111—11000110—00011—0100101—11011—110010—1011—110010—10
LZ-76 | 0—1—00—11—0010—10110 | 01—10—00—0111—1000110—0010—0100—011010—1100—110

**TABLE I**
LZ-78, SWLZ AND LZ-76 ALGORITHMS: PARSING OF EXAMPLE SEQUENCES $X^n =’010011001011101’$, $Y^n =’100110010100100’$.

| Symbol | Count | Relative frequency | Context | Symbol | Count | Relative frequency |
|---|---|---|---|---|---|---|
| 0 | 7 | 7/15 | 0 | 0 | 2 | 2/7 |
| 1 | 8 | 8/15 | 0 | 1 | 5 | 5/7 |
| 1 | 0 | 4 | 4/7 |
| 1 | 1 | 3 | 3/7 |

**TABLE II**
EMPIRICAL PROBABILITY MODEL OF $X^n$

| Symbol | Count | Relative frequency | Context | Symbol | Count | Relative frequency |
|---|---|---|---|---|---|---|
| 0 | 9 | 9/15 | 0 | 0 | 4 | 4/8 |
| 1 | 6 | 6/15 | 0 | 1 | 4 | 4/8 |
| 1 | 0 | 5 | 5/6 |
| 1 | 1 | 1 | 1/6 |

**TABLE III**
EMPIRICAL PROBABILITY MODEL OF $Y^n$

| Symbol | Count | Relative frequency | Context | Symbol | Count | Relative frequency |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 |

**TABLE IV**
EMPIRICAL PROBABILITY MODEL OF $(X^n, Y^n)$

| Symbol | Count | Relative frequency | Context | Symbol | Count | Relative frequency |
|---|---|---|---|---|---|---|
| 00 | 3 | 3/14 | 00 | 0 | 0 | 0 |
| 01 | 4 | 4/14 | 01 | 1 | 2 | 2/3 |
| 01 | 0 | 0 | 0 |
| 01 | 1 | 1 | 1/5 |
| 01 | 1 | 1 | 1/5 |
| 01 | 0 | 0 | 0 |
| 01 | 1 | 1 | 1/5 |
| 01 | 0 | 0 | 0 |
| 01 | 1 | 1 | 1/5 |
| 01 | 0 | 0 | 0 |
| 01 | 1 | 1 | 1/5 |
| 01 | 0 | 0 | 0 |

Complexity is given by [8]

$$K_{SWLZ}(X^n) = \sum_{i=1}^{n} \frac{n \log n}{L_i(X^n)},$$

(8)

So, the corresponding entropy rate estimate is

$$\hat{H}_{SWLZ}(X) = \lim_{n \to \infty} K_{SWLZ}(X^n).$$

The mutual SWLZ complexity $K_{SWLZ}(X^n, Y^n)$ is given by [8]

$$K_{SWLZ}(X^n, Y^n) = \sum_{i=1}^{n} \frac{2n \log n}{L_i(X^n, Y^n)},$$

(9)

where $L_i(X^n, Y^n)$ is length of sub-block at instant $i$ in joint dictionary of $X^n, Y^n$. The mutual information rate is given as

$$I_{SWLZ}(X; Y) = \lim_{n \to \infty} (K_{SWLZ}(X^n) + K_{SWLZ}(Y^n) - K_{SWLZ}(X^n, Y^n))$$

(10)

**LZ-76 Algorithm:** LZ-76 provides another version of the dictionary-based scheme LZ-78. The LZ-76 complexity is given by [13]

$$K_{LZ76}(X^n) = \frac{b(X^n)}{n} \log n,$$

(11)

where $b(X^n)$ represents total number of sub-blocks. So, the corresponding entropy rate estimate is

$$\hat{H}_{LZ76}(X) = \lim_{n \to \infty} K_{LZ76}(X^n).$$

The mutual LZ-76 complexity $K_{LZ78}(X^n, Y^n)$ is given as

$$K_{LZ76}(X^n, Y^n) = \frac{b(X^n, Y^n)}{n} \log n,$$

(12)

here $b(X^n, Y^n)$ is number of sub-blocks in joint dictionary of $X^n, Y^n$. The mutual information rate is given as

$$I_{LZ76}(X; Y) = \lim_{n \to \infty} (K_{LZ76}(X^n) + K_{LZ76}(Y^n) - K_{LZ76}(X^n, Y^n))$$

(13)

**Illustration:** Parsing by the aforementioned algorithms are illustrated on an example strings $X^n =’010011001011101’$, $Y^n =’100110010100100’$ and the joint string $(X^n, Y^n) =’0110001110100011010101010010’$ in Table III.
C. Proposed Estimator based on Empirical Probabilities

We propose an estimator $\bar{I}_{EP}(X;Y)$ which models the neuronal spiking process as Markov process up to a prescribed order $k$. Further, the conditional probabilities of different contextual symbols up to a prefixed order $k$ are calculated from sequences empirically. For instance, the marginal and joint empirical probability models for the sequences $X^n = '010011001110111011010'$ and the joint string $(X^n,Y^n) = '0110001111000110011011010010'$ are shown in Tables [III] [III] [V] respectively. Here, $k=1$.

The mutual information estimator $\bar{I}(X;Y)$ for sequences $X^n = X_k^n = (X_1,X_2,\ldots,X_n)$ and $Y^n = Y_k^n = (Y_1,Y_2,\ldots,Y_n)$ is as follows.

$$\bar{I}_{EP}(X;Y) = \bar{R}_{EP}(X) - \bar{R}_{EP}(X)$$

where

$\frac{1}{n}H(X_i^k|Y_{i-1}^k) = \frac{1}{n}(H(X_1,Y_1) + \cdots + H(X_{n-1},Y_{n-1}) + H(X_n,Y_n))$ by chain rule

$$\approx H(X_{k+1},Y_{k+1}|X_k^k,Y_k^k)$$

where $(X_k^k,Y_k^k)$ provides the context of order $k$ (i.e., $k$ previous symbols) for $X_{k+1},Y_{k+1}$. In the preceding, $H(U|V)$ denotes the conditional entropy of $U$ given $V$, similarly, the entropy rate estimators for $X^n$ and $Y^n$ are as follows [8].

$$\frac{1}{n}H(X_i^k) \approx H(X_{k+1}|X_k^k)$$

$n >> k$,

$$\frac{1}{n}H(Y_i^k) \approx H(Y_{k+1}|Y_k^k)$$

$n >> k$,

Hence, the mutual information estimator $\bar{I}(X;Y)$ is

$$\bar{I}_{EP}(X;Y) \approx H(X_{k+1}|X_k^k) + H(Y_{k+1}|Y_k^k) - H(X_{k+1},Y_{k+1}|X_k^k,Y_k^k)$$

III. RESULTS

We begin by comparing our method with existing LZ algorithms with an example Markov process. Subsequently, we extend such comparison to neuronal calcium spike trains.

### Table V

| Block length | $I_{SWLZ}(X;Y)$ | $I_{EP}(X;Y)$ |
|--------------|-----------------|---------------|
| k=0          |                 |               |
| 50           | 0.151           | 0.151         |
| 100          | 0.089           | 0.089         |
| 150          | 0.038           | 0.038         |
| 200          | 0.029           | 0.029         |

A. Results for Example Markov Process

We compared estimates by generating four-state (00, 01, 10 and 11) Markov processes $(X(m),Y(m))$ (the number in parenthesis indicating the order) of order $m$ with mutual information rate $\bar{I}(X(m);Y(m))$. $X(m), Y(m)$ were generated by using following equations

$$X_i = X_{i-1} \oplus X_{i-2} \oplus \cdots \oplus X_{i-m+1} \oplus X_{i-m} \oplus u_i$$

(18)

$$Y_i = Y_{i-1} \oplus Y_{i-2} \oplus \cdots \oplus Y_{i-m+1} \oplus Y_{i-m} \oplus v_i$$

(19)

where $(u_i,v_i)$ is an joint source with distribution $Pr(u,v)$ in (20) and the ‘⊕’ indicates modulo-2 addition.

$$Pr(u,v) = \begin{pmatrix} 0 & 1 \\ 0.25-p & 0.25+p \\ 1 & 0.25+p & 0.25-p \end{pmatrix}$$

(20)

Hence,

$$\bar{I}(X;Y) = I(u;v)$$

$$= 2(1 + (a_1)\log_2 a_1 + (a_2)\log_2 a_2)$$

(21)

where $a_1 = 0.25 - p$ and $a_2 = 0.25 + p$. The mutual information rate between $X(m), Y(m)$ is equal to mutual information rate between binary sources $u_i$ and $v_i$. Refer appendix for proof of (21).

Mutual information rate of these processes $\bar{I}(X(m),Y(m)) =0.531$ bits $\forall m$ for chosen $p=0.2$ using (22). The raster plot for a realization of each of the above processes for $m=0,1,2,3$ generated by using (18). (19) is shown in Fig. 4.

We next demonstrate the efficacy of the proposed estimator. Specifically, we calculated $\bar{I}_{EP}(X(m);Y(m))$ using (17), for each of the context lengths $k=0,1,2,3$ and each of the generated processes $(X(0),Y(0)),(X(1),Y(1)),(X(2),Y(2)),(X(3),Y(3))$ (corresponding to $m=0,1,2,3$, respectively), and depict those against the window (sequence) length in Figures 5(a)-(d), respectively. In the respective figures, we compared the proposed method against various LZ estimates, $\bar{I}_{SWLZ}(X;Y), \bar{I}_{SWLZ}(X;Y)$, and $\bar{I}_{LZ}(X;Y)$ (see 5, 8 and 11 respectively). Our method tends to converge.
For experimental analysis, we chose eight neurons, indexed 1, 2, 3, 4, 5, 6, 7, 8 (see Fig. 3), whose calcium responses are shown in Fig. 3. As explained earlier, we first applied the deconvolution method to infer spike trains, and computed the spiking rates and then spiking threshold $I_{th}=0.36$ to remove spurious spikes. In Fig. 3, we showed the inferred spike train for neuron 1, and analogous binary spike sequences were used for further analysis.

B. Experimental Results

For experimental analysis, we chose eight neurons, indexed 1, 2, 3, 4, 5, 6, 7, 8 (see Fig. 3), whose calcium responses are shown in Fig. 3. As explained earlier, we first applied the deconvolution method to infer spike trains, and computed the spiking rates and then spiking threshold $I_{th}=0.36$ to remove spurious spikes. In Fig. 3, we showed the inferred spike train for neuron 1, and analogous binary spike sequences were used for further analysis.

1) Optimization of block length: We studied the behavior of the proposed estimator $I_{EP}(X;Y)$ for neurons 1, 2 for different lengths of time windows, and compared it with $I_{SWLZ}(X;Y)$, as stated earlier. For that purpose, the midpoint of all time windows of length 50, 100, 150, 200 was fixed at 115 and shifted by 5 every time up to 160. Here, we varied the context length within $k = 0, 1, 2$. Subsequently, we estimated mutual information rates in the aforementioned time windows for different $k$, whose standard deviations are furnished in Table V. We observed that mutual information rate estimates had smallest standard deviation for windows of length 200 for each $k$. Such behavior is clearly corroborated in Figure 7. Note that $I_{SWLZ}(X;Y)$ exhibited relatively less deviation in windows of length 200 but it also always overestimated the mutual information rate, and proved unsuitable for present comparison. Thus, the window
length of 200 is most well suited for mutual information rate estimation. However, such analysis should be carried out with more neurons for more comprehensive conclusions.

2) Temporal variation in mutual information rate estimates: We next turn to studying temporal variation in mutual information content. To this end, we again considered with aforementioned neurons 1, 2, 3, 4, 5, 6, 7, 8, and the spike sequences are shown in raster plot in Fig. 9. Specifically, we plotted temporal variation of mean estimated mutual information rates of sequence pairs 1-2, 3-4, 5-6, 7-8 using $\hat{I}_{EP}(X;Y)$ and $\hat{I}_{SWLZ}(X;Y)$ in Fig. 9. Here, the context length $k$ was varied from 0 to 6. The mutual information rate estimates, $I_{EP}(X;Y)$ converges to a value between sequence lengths, 150 and 200 unlike $\hat{I}_{SWLZ}(X;Y)$. For neuron 5-6, the mutual information rate estimates, $\hat{I}_{EP}(X;Y)$ were almost same across time for the orders $k = 0, 1, \ldots, 5$ unlike for $k = 6$, for which slight overestimation is observed (Fig. 9(c)). Further, there was some gap in estimates, $\hat{I}_{EP}(X;Y)$, $k = 0$, and $\hat{I}_{EP}(X)$, $k = 1, \ldots, 6$ for neurons 1-2 and 3-4 as shown in Figures 9(a) and 9(b). This suggested that there could be process memory of order $k = 1$ for neurons 1-2 and 3-4 unlike neurons 5-6, for which order $k = 0$ appears enough. However, the temporal behaviour of neurons 7-8 was completely different from the aforementioned neurons as shown in Figure 9(d). The proposed mutual information rate estimates $\hat{I}_{EP}(X;Y)$ were different and underestimated for each $k = 0, 1, 2$. Further, such estimates were equal to $\hat{I}_{SWLZ}(X;Y)$ for $k = 4, 5, 6$. This suggested that there could be complex process memory structure for this type of neurons. In other words, we clearly observed heterogeneity in mutual information rate estimates, considering only eight neurons. We believe that the proposed estimator can provide a useful tool for studying heterogeneity in process memory in large neuron populations.

IV. CONCLUSION

In this paper, we proposed a fast yet accurate empirical method for estimating mutual information rate between calcium spike trains. Our method outperformed the existing LZ algorithms for lower order memory structures in synthetic as well as experimental spike trains. Furthermore, the heterogeneity in mutual information estimates was observed clearly in neurons. Even though, our method identified the time window required for efficient estimation and tracked temporal variation in information estimates, there is always scope for improvement. Especially, we need to do thorough analysis of complex process memory structures in large neuron populations. The current mutual information rate analysis could be helpful for understanding neuronal information encoding under normal and diseased conditions, and identifying disease signatures.

V. ACKNOWLEDGMENT

We thank Drs. Mennerick and Gautam for providing materials and equipment. Sathish Ande thanks the Ministry of Electronics and Information Technology (MeitY), the Government of India, for fellowship grant under Visvesvaraya PhD Scheme.

REFERENCES

[1] M. B. Kennedy, “Synaptic signaling in learning and memory,” Cold Spring Harbor perspectives in biology, vol. 8, no. 2, p. a016824, 2016.

[2] M. J. Berridge, “Neuronal calcium signaling,” Neuron, vol. 21, no. 1, pp. 13–26, 1998.

[3] S. Peron, T.-W. Chen, and K. Svoboda, “Comprehensive imaging of cortical networks,” Current opinion in neurobiology, vol. 32, pp. 115–123, 2015.

[4] M. R. Cohen and A. Kohn, “Measuring and interpreting neuronal correlations,” Nature neuroscience, vol. 14, no. 7, p. 811, 2011.

[5] J.-L. Blanc, N. Schmidt, L. Bonnier, L. Pezard, and A. Lesne, “Quantifying neural correlations using lempel-ziv complexity,” in Deuxième conférence française de Neurosciences Computationnelles, Neurocomp08”, 2008.

[6] J. Szczepanski, M. Arnold, E. Wajnryb, J. M. Amigó, and M. V. Sanchez-Vives, “Mutual information and redundancy in spontaneous communication between cortical neurons,” Biological cybernetics, vol. 104, no. 3, pp. 161–174, 2011.

[7] E. T. Rolls and A. Treves, “The neuronal encoding of information in the brain,” Progress in neurobiology, vol. 95, no. 3, pp. 448–490, 2011.

[8] S. Ande, J. Regatti, N. Pandey, S. Avasarala, S. Swain, A. Karunarathne, L. Giri, and S. Jana, “Information content in neuronal calcium spike trains: Entropy rate estimation based on empirical probabilities,” in 6th International IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES 2020). IEEE, 2020 (Accepted). [Online]. Available: https://arxiv.org/abs/2102.00723

[9] S. Ande, J. Regatti, N. Pandey, A. Karunarathne, L. Giri, and S. Jana, “Heterogeneity in neuronal calcium spike trains based on empirical distance,” in 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), 2021 (Accepted).
APPENDIX

For the processes \((X(m), Y(m))\), we now prove (21) i.e., the mutual information rate \(\bar{I}(X; Y) = I(u; v)\) irrespective of the choice of \(m, n\). From (41), we have

\[
\bar{I}(X; Y) = \frac{1}{n} H(X^n) + \frac{1}{n} H(Y^n) - \frac{1}{n} H(X^n, Y^n) \quad (22)
\]

Let us elaborate each term as follows.

\[
\frac{1}{n} H(X^n) = \frac{1}{n} (H(X_1) + \sum_{i=2}^{n} H(X_i | X_{i-1}^{i-1})) \quad \text{(by chain rule)} \quad (23)
\]

From (18), \(H(X_1) = H(X_i | X_{i-1}^{i-1}) = H(u) \quad \forall i, m\).

Hence (23) becomes

\[
\frac{1}{n} H(X^n) = \frac{1}{n} (H(u) + \sum_{i=2}^{n} H(u)) = H(u) \quad (24)
\]

Similarly,

\[
\frac{1}{n} H(Y^n) = \frac{1}{n} (H(v) + \sum_{i=2}^{n} H(v)) = H(v) \quad (25)
\]

Hence, the marginal entropy rates of \(X(m)\), \(Y(m)\) are equal to entropy \(H(u)\), \(H(v)\) of binary sources \(u\), \(v\) respectively.

\[
H(u) = H(v) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1 \text{ bit.} \quad (26)
\]

where

\[
\begin{align*}
H(X_i, Y_i | X_{i-m}^{i-1}, Y_{i-m}^{i-1}) &= H(X_i | X_{i-m}^{i-1}, Y_{i-m}^{i-1}) + H(Y_i | X_{i-m}^{i-1}, Y_{i-m}^{i-1}, X_i) \\
&= H(X_i | X_{i-m}^{i-1}, Y_{i-m}^{i-1}) + H(Y_i | X_{i-m}^{i-1}, Y_{i-m}^{i-1}) \\
&= H(X_i | X_{i-m}^{i-1}, Y_{i-m}^{i-1}) + H(Y_i | X_{i-m}^{i-1}, u_i, Y_{i-m}^{i-1}) \\
&= H(X_i | X_{i-m}^{i-1}) + H(v_i | u_i) \\
&= H(u_i) + H(v_i | u_i) = H(u_i, v_i) = H(u, v) \quad \forall i, m
\end{align*}
\]

Similarly,

\[
H(X_1, Y_1) = H(u, v) \quad (28)
\]

Substitute (27), (28) in (26) yields

\[
\frac{1}{n} H(X^n, Y^n) = H(u, v) \quad (29)
\]

Hence, the joint entropy rate between sequences \(X^n\), \(Y^n\) generated from processes \(X(m)\), \(Y(m)\) respectively is equal to joint entropy \(H(u, v)\) between the binary sources \(u\), \(v\).

Substitute (24), (25), (29) in (22) yields

\[
\bar{I}(X; Y) = H(u) + H(v) - H(u, v) = I(u; v) \quad \forall m \quad (30)
\]

Hence, the mutual information rate \(\bar{I}(X; Y)\) between \(X(m)\), \(Y(m)\) is equal to mutual information \(I(u; v)\) between binary sources \(u\) and \(v\).

\[
H(u, v) = \sum_{u, v} p(u, v) \log_2 p(u, v)
\]

\[
= -2(a_1) \log_2 (a_1) - 2(a_2) \log_2 (a_2) \quad \text{(from 20)} \quad (31)
\]

where \(a_1 = 0.25 - p\) and \(a_2 = 0.25 + p\).

Substitute \(H(u) = H(v) = 1 \text{ bit.} \quad (32)\) in (30) yields

\[
\bar{I}(X; Y) = 2(1 + (a_1) \log_2 (a_1) + (a_2) \log_2 (a_2)) \quad \forall m \quad (32)
\]

This completes the proof.