Information Content in Neuronal Calcium Spike Trains: Entropy Rate Estimation based on Empirical Probabilities

Sathish Ande
Department of Electrical Engineering
IIT Hyderabad, Hyderabad, India

Srinivas Avasarala
Department of Electrical Engineering
IIT Hyderabad, Hyderabad, India

Jayanth R Regatti
Department of Electrical Engineering
IIT Hyderabad, Hyderabad, India

Neha Pandey
Department of Electrical Engineering
IIT Hyderabad, Hyderabad, India

Sarpras Swain
Department of Chemical Engineering
IIT Hyderabad, Hyderabad, India

Ajith Karunarathne
Department of Chemistry and Biochemistry
The University of Toledo, Ohio, USA

Lopamudra Giri
Department of Chemical Engineering
IIT Hyderabad, Hyderabad, India

Soumya Jana
Department of Electrical Engineering
IIT Hyderabad, Hyderabad, India

Abstract—Quantification of information content and its temporal variation in intracellular calcium spike trains in neurons helps one understand functions such as memory, learning, and cognition. Such quantification could also reveal pathological signaling perturbation that potentially leads to devastating neurodegenerative conditions including Parkinson’s, Alzheimer’s, and Huntington’s diseases. Accordingly, estimation of entropy rate, an information-theoretic measure of information content, assumes primary significance. However, such estimation in the present context is challenging because, while entropy rate is traditionally defined asymptotically for long blocks under the assumption of stationarity, neurons are known to encode information in short intervals and the associated spike trains often exhibit nonstationarity. Against this backdrop, we propose an entropy rate estimator based on empirical probabilities that operate within windows, short enough to ensure approximate stationarity. Specifically, our estimator, parameterized by the length of encoding contexts, attempts to model the underlying memory structures in neuronal spike trains. In an example Markov process, we compared the performance of the proposed method with that of versions of the Lempel-Ziv algorithm as well as with that of a certain stationary distribution method and found the former to exhibit higher accuracy levels and faster convergence. Also, in experimentally recorded calcium responses of four hippocampal neurons, the proposed method showed faster convergence. Significantly, our technique detected structural heterogeneity in the underlying process memory in the responses of the aforementioned neurons. We believe that the proposed method facilitates large-scale studies of such heterogeneity, which could in turn identify signatures of various diseases in terms of entropy rate estimates.

Index Terms—Calcium imaging; Spike train; Empirical probability; Entropy rate; Heterogeneity.

I. INTRODUCTION

Neurons are known to encode stimulus information in temporal spiking patterns of action potentials, which convey information about high-level functions including learning and memory [1]. In this context, measuring information content of neural spike trains and tracking its variation with time assume significance. However, estimation of entropy rate, the traditional measure of information content [3], faces certain practical difficulty. While entropy rate estimation for stationary processes is not difficult, neuronal spike trains often exhibit nonstationarity, and the challenge lies in accurately capturing temporal variation in the entropy rate [4]. To meet this challenge, we take inspiration from speech analysis [5], assume window-wise stationarity, and estimate the local entropy rate. However, while local stationarity may be assumed within a small window, one requires a large window for the entropy rate estimate to converge to a reliable value [6]. So, the task boils down to choosing an entropy rate estimator that is accurate yet converges quickly, and pairing such an estimator with suitable window size, neither too long, nor too short. In this paper, we propose an entropy rate estimator based on empirical probabilities, and experimentally demonstrate its superiority over competing methods on a theoretical example, as well as on calcium spike trains observed in neurons extracted from the hippocampus, a region of the brain associated with learning and memory [2].
Early research in this direction assumed independent and identically distributed (iid) spiking process, and unrealistically ignored timing information and possible memory \cite{7, 8}. Subsequently, as entropy rate estimator, versions of Lempel-Ziv (LZ) algorithm were suggested \cite{9}, which are guaranteed to converge for large block length \( n \) irrespective of memory structure \cite{10}. Block-based methods were also proposed for large \( n \) \cite{6}. Advanced methods have modeled the spiking process as stationary and Markov \cite{11}. One such method used finite context trees up to certain depth \( k \) and assigned weights to those trees \cite{12}; however, such weights, based on *ad hoc* assumptions, could be unreliable. Another method, assuming hierarchical Dirichlet priors, was demonstrated only for large block length \( n \) \cite{13}. Interestingly, short blocks were recently considered under first and second order Markov models (albeit not in a neuronal study) \cite{14}; however, estimated higher order transition probabilities based on stationary distributions could be error prone for higher orders. In contrast, we propose an entropy rate estimation method that estimates empirical conditional distributions and generalizes to higher order Markov models.

Historically, similar issue of statistical modeling also arose in text compression, where LZ algorithms were found ineffective in compressing short texts \cite{3}. In that context, prediction by partial matching (PPM), based on empirical Markov modeling, proved effective \cite{15}. However, PPM, while closely related to entropy rate estimation, uses certain escape symbols to allow decompression, which is irrelevant for present purpose. Accordingly, we propose an algorithm that makes use the main principle behind PPM, but excludes escape symbols. The rest of this paper is organized as follows. Section II describes collection of calcium imaging data and the deconvolution technique for spike train inference, introduces the notion of entropy rate of stochastic process, and elaborates LZ algorithms and proposed empirical entropy rate estimator. Further, Section III presents the results demonstrating 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

Hippocampal neurons were cultured from 1 day postnatal Sprague-Dawley rats. We performed time-lapse imaging for monitoring intracellular calcium in hippocampal neurons at \( 7^{th} \) day after plating. The neurons were loaded with 1 \( \mu M \) Fluo-4 (Molecular Probes) for 30 min in Hank’s Balanced Salt Solution (HBSS). The cells were then washed with HBSS for 3 times and Fluo-4 intensity was monitored using excitation at 488 nm and emission at 510 nm. We performed imaging using a spinning-disk confocal imaging system comprising a Leica DMi6000B inverted microscope, a Yokogawa CSU-X1 spinning-disk units \cite{16}. The interval between successive images, while set at 1 s, was observed to be between 0.8 s to 1 s due to inherent variabilities. During imaging, neurons were kept in the incubation chamber with the microscope maintained at 37°C and 5% CO\(_2\).

B. Spike Train Inference

The time course of spatially resolved Fluo-4 fluorescence intensity in neuron populations were obtained using Andor software from the time-lapse image data (see Fig. 1 for representative time-lapse image). Regions of interest were marked at subcellular levels (soma regions). Further, intensity was calculated by subtracting the background intensity level, which was then normalized \cite{17}. See Fig. 3 for representative time course of normalized intensity of four neurons indexed in Fig. 1. To infer spike train from time course of normalized fluorescence intensity, we used a fast nonnegative deconvolution algorithm \cite{18}. In this algorithm, the estimated spike rate should be interpreted as the expected number of spikes in the 1s imaging interval around each time point. Unfortunately, such estimated spiking rates are generally noisy, and may lead to erroneous inference. Such errors were mitigated by setting an adaptive threshold

\[
P_{th} = 3 \sigma_p, \tag{1}
\]

\( \sigma_p \) indicating the standard deviation of inferred spiking rate vector \cite{19}. Specifically, the instants were labeled 1, when the spiking rate exceeded \( P_{th} \) (high spiking), and 0, otherwise (low spiking). Such binary spike sequences were used for further analysis.
TABLE I
EMPirical PROBABILITY ESTIMATION

| $k = 0$ (no context) | $k = 1$ | $k = 2$ |
|----------------------|---------|---------|
| Symbol | Count | Relative frequency | Context | Symbol | Count | Relative frequency | Context | Symbol | Count | Relative frequency |
| 0 | 11 | 1/20 | 0 | 0 | 4 | 2/5 | 00 | 0 | 1 | 1/3 |
| 1 | 9 | 9/20 | 0 | 1 | 6 | 3/5 | 00 | 1 | 2 | 2/3 |
| 1 | 0 | 6 | 2/3 | 01 | 0 | 3 | 1/2 |
| 1 | 1 | 3 | 1/3 | 01 | 1 | 3 | 1/2 |
| 10 | 0 | 3 | 1/2 |
| 11 | 0 | 3 | 1 |
| 11 | 1 | 0 | 0 |

C. Mathematical Preliminaries

At this point, we introduce the notion of entropy rate, and describe existing entropy rate estimators.

1) Entropy rate: Entropy $H(Y)$, a measure of uncertainty, of random variable $Y$ with probability mass function $p_Y$ defined on alphabet $\mathcal{Y}$ is defined by $H(Y) = -\sum_{y \in \mathcal{Y}} p_Y(y) \log p_Y(y)$. Further, the entropy rate $\bar{H}(X)$ of a sequence of random variables $X_1, X_2, \ldots, X_n$, is defined by

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

where $X^n = (X_1, X_2, \ldots, X_n)$. Thus $\bar{H}(X)$ measures the average asymptotic uncertainty per symbol [3].

2) Lempel-Ziv estimators of entropy rate: Variants of Lempel-Ziv (LZ) algorithms have been proposed for (text) compression, which also estimate the desired entropy rate. In each case, given a stationary and ergodic sequence $X^n = (X_1, X_2, \ldots, X_n)$, LZ complexity approaches the entropy rate $\bar{H}(X)$ with probability one.

LZ-78 algorithm: LZ-78 is dictionary based scheme that adds unforeseen subsequences to the dictionary. Denoting by $c(X^n)$ the number of sub-sequences of a sequence $X^n$ present in the data, the LZ-78 complexity, the average number of bits per symbol, is defined as [3]

$$K_{LZ78}(X^n) = \frac{c(X^n) \log(c(X^n))}{n}.$$  

(3)

So, the corresponding entropy rate estimate is $\hat{H}_{LZ78}(X) = \lim_{n \to \infty} K_{LZ78}(X^n)$, where one practically uses large $n$ in place of the limit.

Sliding Window Lempel Ziv (SWLZ) algorithm: SWLZ is similar to LZ-78, except that new subsequence is seen at each position $i$ is not appear in previous $i-1$ symbols by sliding a window of length, $L_i$. The SWLZ complexity is given by [14]

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

(4)

So, the corresponding entropy rate estimate is $\hat{H}_{SWLZ}(X) = \lim_{n \to \infty} K_{SWLZ}(X^n)$.

LZ-76 algorithm: LZ-76 provides another variation of the basic dictionary-based scheme. The LZ-76 complexity is given by [9]

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

(5)

where $c(X^n)$ represents total number of sub-sequences. So, the corresponding entropy rate estimate is $\hat{H}_{LZ76}(X) = \lim_{n \to \infty} K_{LZ76}(X^n)$.

Illustration: Parsing by the aforementioned algorithms are illustrated on an example string $X^n = '0111001010100111'$. LZ-78: 0—01—1—00—10—101—001—11... SWLZ: 0—01—1—10—0010—010—101... LZ-76: 0—01—10—010—10011—011... Refer to columns 8–11 in Table II for the empirical transition probability distribution for $X^n = '010011011011001100'$.

D. Proposed Estimator based on Empirical Probabilities

We propose an estimator $\hat{H}_{EP}(X)$ which also models the neuronal spiking process as Markov process up to a prescribed order $k$ as in Section II-C3 (Table I). However, we do not use stationary state distribution, and instead use the estimator

$$\hat{H}_{EP}(X) = \frac{1}{n} \left( H(X_1) + \ldots + H(X_k | X_{k-1}, X_{k-2}, \ldots, X_1) \right) \ldots + H(X_n | X_{n-1}, X_{n-2}, \ldots, X_1)$$

by chain rule

$$= \frac{1}{n} H(X_1) + H(X_2 | X_1) + \ldots + (n-k) H(X_{k-1} | X_k, X_{k-1}, \ldots, X_1)$$

for order $k$

$$\approx H(X_{k+1} | X_k, X_{k-1}, \ldots, X_1)$$

if $n >>> k$,

where $(X_k, X_{k-1}, \ldots, X_2, X_1)$ provides the context of order $k$ (i.e., $k$ previous symbols) for $X_{k+1}$. In the preceding, $H(U | V)$
Empirical probability based

Fig. 5. Comparison of the proposed entropy rate estimate \( \hat{H}(X) \) using the aforesaid deconvolution method [18] (see top pane l in Figure 3).

\[ H(X_m) = \sum H(X_m | X_{m-1}, X_{m-2}, \ldots, X_1) \]

\[ H(X_m) = \sum H(X_m | X_{m-2}) \]

\[ H(X_m) = \sum H(X_m | X_{m-3}) \]

Fig. 6. Spike train inferred for neuron 1 from fluorescence in tensity data

\[ H(X) = \sum H(X_1 | X_0) \]

\[ H(X) = \sum H(X_1 | X_0) \]

\[ H(X) = \sum H(X_1 | X_0) \]

Table II

| Window length | \( H_{5WLZ}(X) \) | Empirical probability based estimate, \( \hat{H}_{EP}(X) \) |
|---------------|-----------------|----------------------------------|
| 40            | 0.0500          | 0.0741   0.0834   0.0712       |
| 80            | 0.0190          | 0.0710   0.0745   0.0621       |
| 120           | 0.0126          | 0.0649   0.0640   0.0530       |
| 160           | 0.0124          | 0.0450   0.0495   0.0475       |
| 200           | 0.0120          | 0.0394   0.0343   0.0422       |

denotes the conditional entropy of \( U \) given \( V \).

III. RESULTS

We begin by considering an example Markov process, and turn to neuronal calcium spike trains subsequently.

\[ P(X_{n-1}, \ldots, X_{n-(m-1)}) = \sum_{X_{n-(m-i-1)}} P(X_{n-1}, \ldots, X_{n-(m-i+1)}) \]

\[ P(X_{n}, \ldots, X_{n-(m-i)}) = P(X_{n}, X_{n-1}, \ldots, X_{n-(m-i)}) \]

\[ P(X_{n}, \ldots, X_{n-(m-i)}) = P(X_{n}, X_{n-1}, \ldots, X_{n-(m-i)}) \]

A. Results for Example Markov Process

One can specify a Markov process of order \( m \) by the \( m \)-th order probability transition matrix \( T_m \) and the \( m \)-th order marginal distribution \( \pi_m \), which in turn specify \( m \)-th order joint distribution [3]. Various lower order joint distributions can be obtained via marginalization

\[ P(X_{n-1}, \ldots, X_{n-(m-i)}) = \sum_{x_{n-(m-i-1)}} P(X_{n-1}, \ldots, X_{n-(m-i+1)}) \]
We considered an example process \( X(3) \) (the number in parenthesis indicating the order) of order \( m = 3 \) having probability transition matrix

\[
T_0 = \begin{pmatrix}
0.04 & 0.95 & 0.88 & 0.99 & 0.93 & 0.09 & 0.92 & 0.09 & 0.99 \\
0.96 & 0.05 & 0.12 & 0.08 & 0.01 & 0.07 & 0.91 & 0.10 & 0.09
\end{pmatrix}
\]

and marginal joint distribution

\[
\pi_3 = [0.02, 0.07, 0.07, 0.06, 0.01, 0.07, 0.70].
\]

Entropy rate of this process \( \hat{H}(X(3)) = 0.4158 \) bits. One may compute the lower order transition matrices \( T_2, T_1 \) and corresponding marginal joint distributions \( \pi_2 \) and \( \pi_1 \) using (7) and (9). In the same vein, \( \pi_0 \) may also be computed. Thus, we obtain processes \( X(2), X(1) \) and \( X(0) \) which are consistent with the original process \( X(3) \). We compute \( \hat{H}(X(2)) = 0.4775 \) bits, \( \hat{H}(X(1)) = 0.5339 \) bits and \( \hat{H}(X(0)) = 0.6386 \) bits. The raster plot for a realization of each of the above processes is shown in Fig. 8.

We next demonstrate the efficacy of the proposed estimator. Specifically, we calculated \( \hat{H}_{EP}(X(m)) \) using (7), for each of the context lengths \( k = 0, 1, 2, 3 \) and each of the generated processes \( X(0), X(1), X(2), X(3) \) (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, \( \hat{H}_{LZS}(X) \), \( \hat{H}_{SWLZ}(X) \), and \( \hat{H}_{LZ6}(X) \), as well as the approximate stationary distribution estimate \( \hat{H}_{ESD}(X) \) (see (3), (4), (5) and (6), respectively). Our method tends to converge to \( \hat{H}(X(\min(m,k))) \), for each pair \( (k,m) \), as designed. In particular, we overestimated entropy rate when the order was underestimated, i.e., \( k < m \). For instance, in Figure 5(c), while \( \hat{H}(X(2)) = 0.4775 \) bits for \( m = 2 \), we overestimate \( \hat{H}_{EP}(X(2)) \approx 0.65 \) bits for \( k = 0 \) and \( \approx 0.59 \) bits for \( k = 1 \). However, when \( k \geq m \), the proposed estimator was accurate, and the convergence to \( \hat{H}(X(m)) \) was faster than the each of the reference algorithms. For instance, in

\[
H(0.2, 0.4, 0.8, 0.8, 0.9, 0.9, 0.9, 0.9) = 0.6386 \text{ bits}
\]

and

\[
H(0.2, 0.4, 0.8, 0.8, 0.9, 0.9, 0.9, 0.9) = 0.6386 \text{ bits}
\]

for neuron 1, and analogous binary spike sequences were used for further analysis.

**B. Experimental Results**

For experimental analysis, we chose four neurons, indexed 1, 2, 3, 4 (see Fig. 1) 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 \( P_i = 1.99 \) (using (11)) to remove spurious spikes. In Fig. 5, we showed the inferred spike train for neuron 1, and analogous binary spike sequences were used for further analysis.

1) Optimization of block/window length: We studied the behavior of the proposed estimator \( \hat{H}_{EP}(X) \) for neuron 1 for different lengths of time windows, and compared it with \( \hat{H}_{SWLZ}(X) \), as stated earlier. For that purpose, the midpoint of all time windows of length 40, 80, 120, 160, 200 was fixed at 100 and shifted by 5 every time up to 140. Here, we varied the context length within \( k = 0, 1, 2 \). Subsequently, we estimated entropy rates in the aforementioned time windows for different \( k \), whose standard deviations are furnished in Table 1. We observed that entropy rate estimates had smallest standard deviation for windows of length 200 for each \( k \). Such behaviour was clearly corroborated in Figure 4. Note that \( \hat{H}_{SWLZ}(X) \) exhibited less deviation in windows of length 200; however, it also always overestimated the entropy rate, and proved unsuitable for present comparison. Thus, the window length of 200 is most well suited for entropy rate estimation. However, such analysis should be carried out with more neurons for more comprehensive conclusions.
2) Temporal variation in entropy rate estimates: We next turn to studying temporal variation in information content. To this end, we again considered with aforementioned neurons 1, 2, 3, and 4, and the spike sequences are shown in raster plot in Fig. 8. Specifically, we plotted temporal variation of mean estimated entropy rates of aforementioned sequences using $\tilde{H}_{EP}(X)$ and $\tilde{H}_{SW LZ}(X)$ in Fig. 9. Here, the context length $k$ was varied from 0 to 6. The entropy rate estimates, $\tilde{H}_{EP}(X)$ converges to a value between sequence lengths, 150 and 200 unlike $\tilde{H}_{SW LZ}(X)$. For neuron 1, the entropy rate estimates, $\tilde{H}_{EP}(X)$ were almost same across time for the orders $k = 0, 1, \ldots, 5$ unlike for $k = 6$, for which slight underestimation is observed (Fig. 9[a]). Further, there was large gap in estimates, $\tilde{H}_{EP}(X), k = 0$, and $\tilde{H}_{EP}(X), k = 1, \ldots, 6$ for neurons 2 and 3 as shown in Figures 9(b) and 9(c). This suggested that there could be process memory of order $k = 1$ for neurons 2 and 3 unlike neuron 1, for which order $k = 0$ appears enough. However, the temporal behaviour of neuron 4 was completely different from the aforementioned neurons as shown in Figure 9(d). The proposed entropy rate estimates $\tilde{H}_{EP}(X)$ were different for each $k$ and those were overestimated for $k = 0, 1, 2$ as compared to $\tilde{H}_{SW LZ}(X)$. This suggested that there could be complex process memory structure for this type of neurons. In other words, we clearly observed heterogeneity in entropy rate estimates, considering only four 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 an empirical probability method for estimating entropy rate to quantify the information content and study its temporal variation in calcium spike trains of hippocampal neurons. We demonstrated our method with synthetic Markov sequences as well as experimental spike sequences and found to have fast convergence in short time windows compared to other methods. Although, this method is able to identify the window for slow temporal statistics and capture the heterogeneity in information estimates, there remains much room for improvement. Importantly, our analysis has to be thoroughly investigated for large number of neurons and neuron populations. Further, complex process memory structure of neurons requires deeper analysis. The present entropy rate analysis could be useful for understanding neuronal information encoding under normal and diseased conditions, and identifying disease signatures. Our tool also enables potential investigations into correlation between neuronal information content and other physiological aspects, such as neuronal activity [20], as well as synchronicity [21].

V. ACKNOWLEDGMENT

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