

# QUANTIFYING NEURAL CORRELATIONS USING LEMPEL-ZIV COMPLEXITY

Jean-Luc Blanc, Nicolas Schmidt, Loic Bonnier and Laurent Pezard

Laboratoire de Neurosciences Intégratives et Adaptatives

UMR 6149 - CNRS - Aix-Marseille Université

Annick Lesne

Institut des Hautes Études Scientifiques

Le Bois-Marie, 35 route de Chartres, 91440, Bures-sur-Yvette, France.

*Permanent address* : Laboratoire de Physique Théorique de la Matière Condensée

UMR 7600 - CNRS - Université Paris 6, France.

## ABSTRACT

Spike train analysis generally focuses on two aims: (1) the estimate of the neuronal information quantity, and (2) the quantification of spikes or bursts synchronization. We introduce here a new multivariate index based on Lempel-Ziv complexity for spike train analysis. This index, called *mutual Lempel-Ziv complexity* (MLZC), can both measure spikes correlations and estimate the information carried in spike trains (i.e. characterize the dynamic process). Using simulated spike trains from a Poisson process, we show that the MLZC is able to quantify spike correlations. In addition, using bursting activity generated by electrically coupled Hindmarsh-Rose neurons, the MLZC is able to quantify and characterize bursts synchronization, when classical measures fail.

## KEY WORDS

Lempel-Ziv complexity, spike trains, synchronization, Hindmarsh-Rose model, mutual information, linear correlation.

## 1 Introduction

Information theory defines the entropy rate of a stochastic process as the amount of novel information produced per unit time. In sensory neuroscience, the entropy rate of a neural spike train quantifies the maximum rate of new information produced about an animal's environment that can be transmitted to the brain. This quantity makes no particular assumption about what information is important for the animal nor on the nature (i.e. deterministic vs stochastic) of the underlying dynamics generating spike trains.

In experimental situations, the computation of entropy rates requires a statistical estimator that is unbiased and converging enough fast to be accurate on a finite data sample. Unfortunately, since the classical definition of entropy rate is moreover based on an asymptotic limit, it does not easily lead to an accurate estimator in the case of finite-size time series.

The concept of complexity, in the sense of Kolmogorov, can be used to obtain accurate estimates of the entropy

rate. In particular, using the implementation of Lempel-Ziv complexity (LZC) [1], one gains the advantage on two finite size issues : an accurate control of the statistical fluctuations (sampling issue) [2] and a better estimation of an asymptotic quantity (convergence issue) [3]. Several studies of neural spike trains have already used Lempel-Ziv complexity [4, 5, 6]. Nevertheless, most of these studies only used one-dimensional neural signals. In this paper we focus on a natural extension of LZC to multidimensional signals [7] to study the estimate of higher order correlations between pairs of neural spike trains.

We first recall the definitions of entropy rate and LZC for one-dimensional binary signals. Then, using the relation between Shannon entropy and mutual information, we give the extension of LZC for bivariate signals and deduce a new criterion that we call the Mutual Lempel-Ziv Complexity (MLZC). In the last part of this study, we use the MLZC to quantify the correlations in neural responses using simulated spike trains.

## 2 Definitions

### 2.1 Entropy rate and Lempel-Ziv complexity

For a bitstring  $X_N = [x_1, \dots, x_N]$  of length  $N$  with  $x_i \in \{0, 1\}$ , a procedure that partitions  $X_N$  into non-overlapping substrings is called a *parsing*. A substring starting at position  $i$  and ending at position  $j$  of  $X_N$  which is the result of a parsing procedure is called a *phrase*  $X_N(i, j)$ . The set of phrases generated by a parsing of  $X_N$  is denoted with  $PX_N$  and the number of phrases  $|PX_N|$  is denoted by  $c(X_N)$ .

Assume that a bitstring  $X_N$  has been parsed up to position  $i$ , so that  $PX_N(1, i)$  is the set of phrases generated so far. According to the original parsing procedure (LZ76) [1] the next phrase  $X_N(i + 1, j)$  will be the first substring which is not yet an element of  $PX_N(1, i)$ . As an illustration, the string 0011001010100111 will be parsed as  $0 \cdot 01 \cdot 10 \cdot 010 \cdot 10100 \cdot 111$  using LZ76 procedure.

For a bitstring  $X_N$ , Lempel-Ziv complexity  $C_{LZ}(X_N)$  is

defined as:

$$C_{LZ}(X_N) = \frac{c(X_N)[\log_k c(X_N) + 1]}{N} \quad (1)$$

where  $k$  is the alphabet size ( $k = 2$  in the binary case).

If the bitstring is generated by a source, assumed to be stationary, the  $n$ -block entropy of the source is  $H_n$  the Shannon entropy of the  $n$ -words  $w_n$ :  $H_n := -\sum_{w_n} p_n(w_n) \log p_n(w_n)$  (with natural logarithm according to the dynamical system convention but at odds with Shannon definition). Then the entropy rate of a stationary and ergodic source is defined as

$$h = \lim_{n \rightarrow \infty} \frac{H_n}{n} = \lim_{n \rightarrow \infty} H_{n+1} - H_n \quad (2)$$

For a random sequence  $X_N$  from an ergodic source, it can be shown that  $h = \lim_{N \rightarrow \infty} C_{LZ}(X_N)$  [8, 9]. This result shows that LZC also quantifies average information quantity in Shannon's sense. Since the LZC is based on the study of recurrence of patterns in a symbolic sequence, this approach provides a tool for the analysis of complex sequences *e.g.* chaotic sequences [10].

## 2.2 Lempel-Ziv complexity for multidimensional sequences

A natural extension of LZC for multidimensional data has been proposed [7]. In the case of a set of  $l$  symbolic sequences  $\{X_N^i\}$  ( $i = 1, \dots, l$ ), Lempel and Ziv's definitions remain valid if one extends the alphabet from scalar values  $x_k$  to  $l$ -tuples elements  $(x_k^1, \dots, x_k^l)$ .

In the case  $l = 2$ , one can thus define the joint LZC  $C_{LZ}(X_N, Y_N)$  for two sequences  $X_N$  and  $Y_N$  is thus defined as

$$C_{LZ}(X_N, Y_N) = \frac{c(X_N, Y_N)[\log_{k^2} c(X_N, Y_N) + 1]}{N} \quad (3)$$

The joint LZC has similar properties as Shannon joint entropy  $H(X_N, Y_N)$ .

Pushing forward the analogy with the Shannon information theory [11] we define the mutual Lempel-Ziv complexity  $MC_{LZ}(X_N; Y_N)$  between sequences  $X_N$  and  $Y_N$  as:

$$MC_{LZ}(X_N; Y_N) = C_{LZ}(X_N) + C_{LZ}(Y_N) - C_{LZ}(X_N, Y_N) \quad (4)$$

The mutual Lempel-Ziv complexity (MLZC) can be understood as a divergence measure between two sequences, by contrast to the mutual information the MLZC can be negative transiently for finite  $N$ , for  $N \rightarrow \infty$  the "true" asymptotic quantity  $MC_{LZ}(X_N; Y_N)$  is positive. In fact the MLZC converges asymptotically to a dynamic extension of the mutual information : *the mutual information rate* [12, 13].

$$\mathcal{I}(X, Y) = \lim_{n \rightarrow \infty} \frac{I_n(X, Y)}{n} = \lim_{n \rightarrow \infty} I_{n+1}(X, Y) - I_n(X, Y) \quad (5)$$

As the rate of mutual information, quantifies all the correlations between the temporal organization of the observed sequences  $X$  and  $Y$  and reflects up to what point they give independent (or related) information on the underlying dynamics of the system. Exactly as the entropy rate  $h(X)$  takes a better account of the whole temporal structure of the sequence  $X$  (compared to linear statistical indices like the correlation function), the production of mutual information per unit time provides a more complete quantification of the interrelations between the two sequences  $X$  and  $Y$ , more thorough than the covariance (like mutual information, MLZC accounts for all correlations, not only the linear ones) and better taking into account the temporal structure of the sequences; indeed, these sequences are more than joint random realizations of two random variables, but rather the joint realization of a random process and MLZC is an integrated index far more meaningful in this respect than a collection of pointwise quantities computed at a given time.

## 3 Numerical simulations

For each simulation we compare the behavior of the linear correlation coefficient (COR), the mutual information (MI) and the mutual Lempel-Ziv complexity (MLZC).

### 3.1 Poisson spike trains

We simulated two correlated binary sequences  $X_N$  and  $Y_N$  representing the discharge of two correlated Poisson neurons according to the following procedure:

1. Two independent Poisson processes are used to generate time series of spike occurrences  $t = t_1, \dots, t_n$  which were translated into bitstrings according to a partition of the time interval  $[0, T]$  into  $N$  bins of equal width  $\Delta t$  ( $N = T/\Delta t$ ). The natural binary encoding (*i.e.*  $b_i = 1$  when one spike occurs in time interval  $i$ ) was then used leading to two independent binary sequences  $B_N^1$  and  $B_N^2$ .
2. We then obtained  $X_N$  and  $Y_N$  from  $B_N^1$  and  $B_N^2$  with:

$$\begin{aligned} X_N &= B_N^1 \oplus \alpha B_N^2 \\ Y_N &= B_N^2 \oplus \alpha B_N^1 \end{aligned} \quad (6)$$

where  $B_N^j \oplus \alpha B_N^i$  means that  $b_k^j$  take the values of  $b_k^i$  with probability  $\alpha$ .

Figure 1 shows the behaviour of the linear correlation coefficient, the mutual information and the MLZC as functions of the mixing parameter  $\alpha$ . We generated here 100 pairs of spike trains with similar firing rate ( $r = 20$  spikes/s, time window  $T = 10s$  and the time resolution  $\Delta t = 1ms$ ). For comparison purpose, we consider here a simple normalization of the mutual Lempel-Ziv complexity (MLZC) which takes value in the unit interval  $[0, 1]$  ( $MC' = (MC - MC_{min}) / (MC_{max} - MC_{min})$ ).

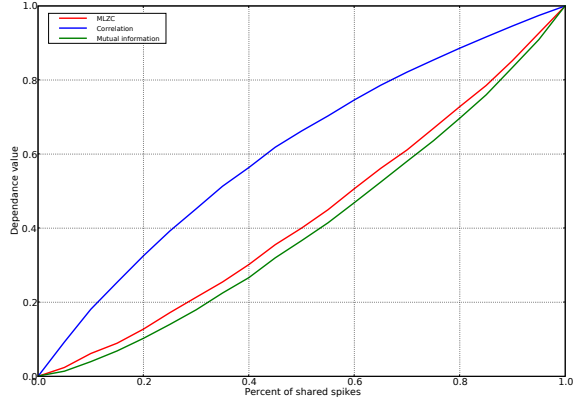


Figure 1. Dependence between two Poisson spike trains : measure of the dependence as a function of the mixing parameter. Blue curve: COR; green curve: MI; red curve: MLZC

### 3.2 Bursting activity

Second we simulated on a more realistic case two coupled neurons, using two Hindmarsh-Rose (HR) models with an electrical coupling [14]. The HR model can reproduce most of the different activity regimes of biological neurons. For different external input currents, the HR neuron may spike and burst regular or chaotically. We consider here the electrical coupling between two HR neurons. Electrical synapses are binary couplings between neurons where the current exchange is simply due to difference between their membrane potential. The time evolution of two electrically coupled HR neurons is described by a set of three differential equations [15].

$$\begin{aligned}
 \dot{x}_i &= y_i + x_i(3x_i - x_i^2) - z_i + I + \epsilon(x_j - x_i) \\
 \dot{y}_i &= 1 - 5x_i^2 - y_i \\
 \dot{z}_i &= -rz_i + rS(x_i + 1.6)
 \end{aligned} \tag{7}$$

where  $x_i$  denotes the membrane potential of neuron  $i$ ,  $y_i$  and  $z_i$  are "fast" and "slow" ionic currents. We set the value of the conductances parameters to  $r = 0.0021$  and  $S = 4$  and set the value of the external current to  $I = 3.38$ . The electric coupling between neurons  $\epsilon$  is a variable parameter who plays the role of a conductivity.

Figure 2 shows the behavior of the linear correlation coefficient, the mutual information and the MLZC as functions of the electrical coupling  $\epsilon$ . HR coupled neurons present a rich and complex dynamical behaviour, so we detail here only the dynamical states of interest in term of dependence between neurons activity.

In the interval  $0 \leq \epsilon \leq 0.03$  the system is uncoupled with neurons following a chaotic uncorrelated evolution, every dependence measures (COR, MI and MLZC) are near 0. Reaching the interval  $0.04 \leq \epsilon \leq 0.22$  chaotic evolution

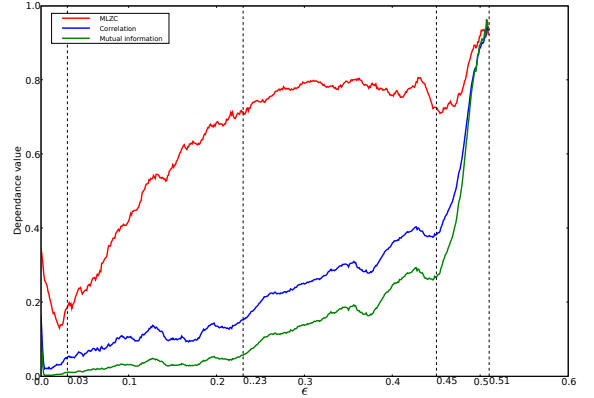


Figure 2. Relation between MLZC, correlation and mutual information of coupled HR neurons

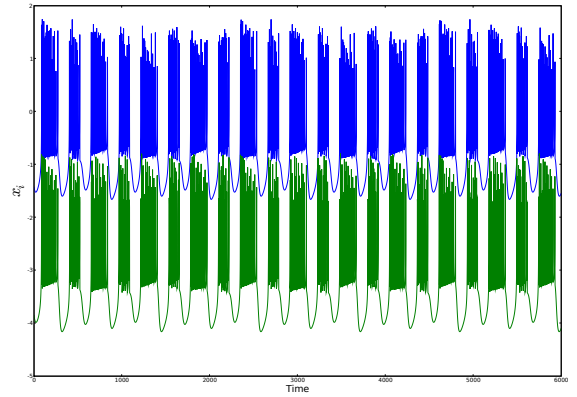


Figure 3. Chaotic time evolution of two HR neurons for  $\epsilon = 0.21$ . Coexisting in phase and anti-phase burst synchronization.

is again dominating, there is no spikes synchronization but some bursts are alternatively synchronized as shown in figure 3. In this case the MLZC value increases faster than the COR and MI values.

In the interval  $0.23 \leq \epsilon \leq 0.44$  several periodic windows of perfect bursts synchronization appear with few spikes synchronization (see figure 4). The MLZC value is high detecting efficiently bursts synchronization, but reaches a plateau due to the periodic regime. The COR and MI values are still weak, but increase slowly.

For  $\epsilon \geq 0.45$  a periodic evolution dominates and the spikes becomes perfectly synchronized for  $\epsilon \geq 0.51$ , behaving as single neuron. Each dependence measure reaches a maximum.

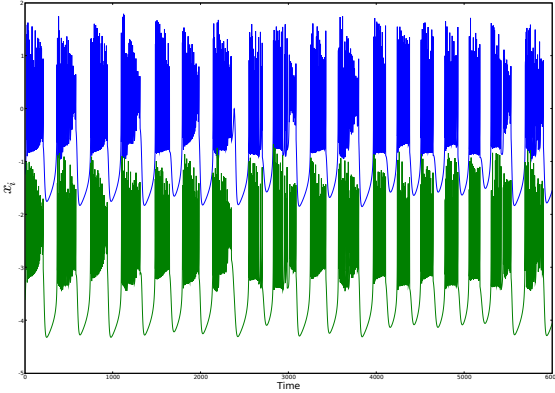


Figure 4. Time evolution of two HR neurons for  $\epsilon = 0.38$ . In-phase burst synchronization with periodic evolution.

## 4 Conclusion

We have introduced a novel spike train analysis method, the MLZC based on a popular data compression algorithm, for the characterization of neuronal correlations.

Considering Poisson spike trains, we find that this measure is able to detect spike synchronization in a better way than using the linear correlation coefficient or the Shannon mutual information. In a more realistic case, considering coupled Hindmarsh-Rose neurons, the MLZC is able to detect bursts synchronization under chaotic and periodic dynamical regimes when the other measures fail. From a theoretical point of view, the MLZC is able to quantify all the correlations between the temporal organization of the observed symbolic sequences and reflecting up to what point they give independent (or related) information on the underlying dynamics of the system.

Further theoretical work should concentrate to establish and exploit mathematical properties of the MLZC measure, in addition further applied work should concentrate on the performance of the MLZC measure to detect spike and burst synchronization, considering in vivo neuronal data.

## References

- [1] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Trans Inform Theory*, IT-22:75–81, 1976.
- [2] J.M Amigo and M.B. Kennel. Variance estimators for the Lempel-Ziv entropy rate estimator. *Chaos*, 16:043102, 2006.
- [3] A. Lesne, J.-L. Blanc, and L. Pezard. Entropy estimation of symbolic sequences : how short is a short sequence ? *Submitted*, 2008.

- [4] J.M. Amigo, J. Szecepanski, E. Wajnryb, and M.V. Sanchez-Vives. Estimating the entropy rate of spike trains via Lempel-Ziv complexity. *Neural. Comput.*, 16:717–736, 2004.
- [5] M. Aboy, R. Hornero, D. Abásolo, and Álvarez D. Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans. Biomed. Eng.*, 53:2282–2288, 2006.
- [6] M. Christen, A. Kohn, T. Ott, and Stoop R. Measuring spike pattern reliability with Lempel-Ziv distance. *Journal of Neuroscience Methods*, 156:342–350, 2006.
- [7] S. Zozor, P. Ravier, and O. Buttelli. On Lempel-Ziv complexity for multidimensional data analysis. *Physica A*, 345:285–302, 2005.
- [8] T.M. Cover and J.A. Thomas. *Elements of information theory*. John Wiley and Son, Inc., New York, 1969.
- [9] R. Badii and A. Politi. *Complexity. Hierarchical structures and scaling in physics*. Cambridge University Press, Cambridge, 1997.
- [10] F. Kaspar and H.G. Schuster. Easily calculable measure for the complexity of spatiotemporal patterns. *Physical Review A*, 36:842–848, 1987.
- [11] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, (27):379–423, 1948.
- [12] S. Ihara. *Information theory for continuous systems*. World scientific, Singapore, 1993.
- [13] M.S. Pinsker. *Information and information stability random process*. Holden Day, San Francisco, 1964.
- [14] J.L. Hindmarsh and R.M. Rose. A model of neuronal bursting using three coupled first order differential equations. *Proceedings of the Royal Society London, B*, 221:87–102, 19984.
- [15] R. Erichsen Jr., M.S. Mainieri, and L.G. Brunet. Periodicity and chaos in electrically coupled Hindmarsh-Rose neurons. *Physical Review E*, 74:061906, 2006.