Estimating the Entropy Rate of Spike Trains via Lempel-Ziv Complexity

José M. Amigó
jm.amigo@umh.es
Centro de Investigación Operativa, Universidad Miguel Hernández, 03202 Elche, Spain

Janusz Szczepański
jszczepa@ippt.gov.pl
Institute of Fundamental Technological Research, Swietokrzyska 21, 00-049 Warsaw, Poland, and Centre of Trust and Certification "Centrast" Co., Poland

Elek Wajnryb
ewajnryb@ippt.gov.pl
Institute of Fundamental Technological Research, Swietokrzyska 21, 00-049 Warsaw, Poland

Maria V. Sanchez-Vives
mavi.sanchez@umh.es
Instituto de Neurociencias, Universidad Miguel Hernández-CSIC, 03550 San Juan de Alicante, Spain

Normalized Lempel-Ziv complexity, which measures the generation rate of new patterns along a digital sequence, is closely related to such important source properties as entropy and compression ratio, but, in contrast to these, it is a property of individual sequences. In this article, we propose to exploit this concept to estimate (or, at least, to bound from below) the entropy of neural discharges (spike trains). The main advantages of this method include fast convergence of the estimator (as supported by numerical simulation) and the fact that there is no need to know the probability law of the process generating the signal. Furthermore, we present numerical and experimental comparisons of the new method against the standard method based on word frequencies, providing evidence that this new approach is an alternative entropy estimator for binned spike trains.

1 Introduction

Neurons respond to external or internal stimuli by firing sequences of discrete, identical action potentials called spike trains (see Figure 1). It is still an open question whether the information about the stimuli is carried by the spike timing, spike patterns, number of spikes, or a different mecha-
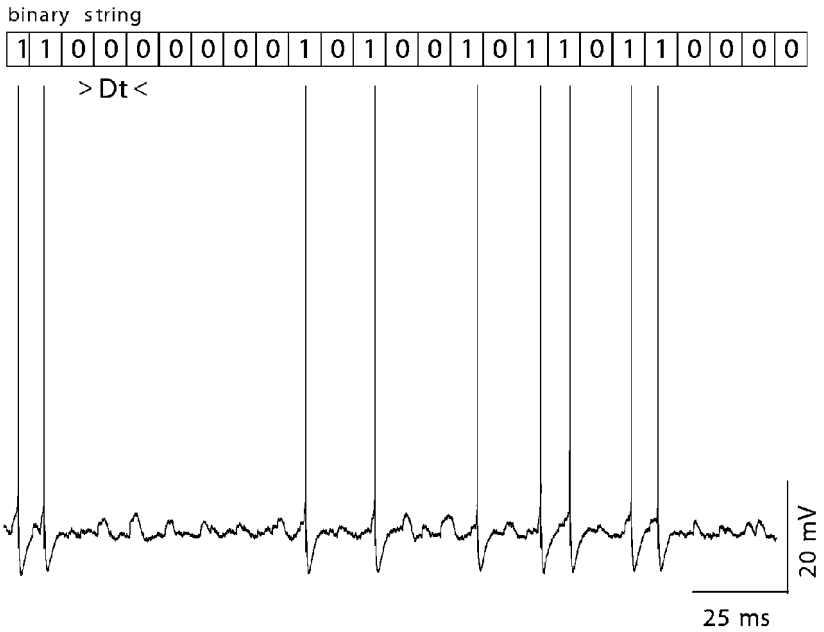
binary string

| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |

> Dt <

20 mV

25 ms

Figure 1: Binary representation of a typical spike train with a given time preci-
sion $\Delta\tau$. By dividing the time axis into discrete bins, a spike train (represented
here as a sequence of impulses) can be represented as a sequence of binary digits,
where 1 denotes at least one spike in the given bin and 0 denotes no spike.

nism (perhaps including some of the previous ones), but independent of
whatever the answer might be, the average amount of information a neu-
ron is transmitting about the stimuli is quantitatively measured in bits per
symbol or bits per second by the mutual information, which is obtained by
subtracting the noise entropy rate (due to the variability of the responses
to a fixed stimulus) from the (Shannon) entropy rate of the entire response
ensemble. (See Shannon, 1948; Cover & Thomas, 1991, for the general the-
ory of information and Rieke, Warland, de Ruyter van Steveninck, & Bialek,
1998; Borst and Theunissen, 1999; Paninski, 2003 for more elaborated and
practical discussions in relation to computational neuroscience.)

In this article, we explore the application of Lempel-Ziv complexity (Lem-
pel & Ziv, 1976) to the measurement of the entropy rate of neurological
signals (not to be mistaken for the complexity measure of the same name
used for data compression, which was proposed later) and compare both
the numerical and experimental results to those obtained with the by now
classic technique of Strong, Koberle, de Ruyter van Steveninck, and Bialek
(1998). The comparison turns out to be very satisfactory: the complexity-

based approach performs better for short sequences, for which undersampling becomes critical. We also address the question of the convergence of the normalized Lempel-Ziv complexity to the entropy rate. The application of Lempel-Ziv complexity to the estimation of mutual information will be studied in a forthcoming article.

2 Signal Codification and Entropy

Before applying the concepts and methods of discrete information theory to a spike train, this has to be transformed into a digital signal or "word," in the parlance of information theory—a sequence of finitely many symbols or "letters." The transformation of a spike train into a word is called the *codification* of the signal and the procedure, the (*en*)*coding*. Codification can be carried out in a variety of ways. Hereafter we will consider only the temporal or binary time bin coding (see Figure 1), which is the most popular encoding since MacKay and McCulloch (1952) used it in their seminal article. Let the first spike of a train under observation occur at time t_0 and the last one at $t_0 + T$. The time interval $[t_0, t_0 + T]$ is then split into n bins $[t_{i-1}, t_i]$ $(1 \le i \le n$, with $t_n = t_0 + T)$ of the same length $\Delta\tau = T/n$. If each bin is now coded by 0 or 1 according to whether it contains no spikes (0) or at least one spike (1), the result is a binary message or word of length n. The bin length $\Delta\tau$ can be interpreted as the time resolution with which the spike train is being observed.

Once a spike train has been codified into a message, this can be viewed as emitted by an information source S that we call, for obvious reasons, a *neuronal source* (Amigó, Szczepánski, Wajnryb, & Sanchez-Vives, 2003). In order to calculate the entropy (or, rather, the entropy rate) of S, we follow Strong et al. (1998) and Reinagel and Reid (2000). First, a representative ensemble of neuronal responses to a given group of stimuli is recorded. Second, one examines segments of the resulting binned spike trains in (overlapping) windows of duration $L \le T$, each segment becoming after codification a binary word of length $l = L/\Delta\tau$. Let $\tilde{p}_i$ be the normalized count of the ith word in the ensemble of words of length l in a set of observations. Then, the corresponding estimate of the neuronal source entropy rate (in bits per second) is

$$H(l, \Delta\tau) = -\frac{1}{l\Delta\tau} \sum \tilde{p}_i \log_2 \tilde{p}_i, \tag{2.1}$$

where the sum is over all the words of length l. As made clear by the notation, such an estimate depends on both l and the time resolution $\Delta\tau$. The true entropy rate $H(S)$ is reached in the limit of vanishing time bins and infinitely long words; mathematically

$$H(S) = \lim_{\Delta\tau \to 0, \, l \to \infty} H(l, \Delta\tau).$$

We want to stress again that the entropy is a property of sources and therefore difficult to evaluate (Strong et al., 1998). In fact, the knowledge of the probability distribution involved in its calculation requires, in principle, an extensive sampling that usually cannot be performed, not to mention the reproducibility of the test conditions. In contrast, the complexity as originally formulated by Lempel and Ziv (1976) is a property of individual sequences that can be used to estimate the entropy or, more generally, to bound it from below. We will sometimes call this concept LZ-76 complexity to distinguish it from LZ-78 complexity, a different definition of complexity also due to Lempel and Ziv (Ziv & Lempel, 1978), which is the basis of typical lossless compression algorithms (e.g., WinZip) in common use in modern computing and other information technologies. Unless otherwise stated, we henceforth refer always to the LZ-76 complexity.

3 Lempel-Ziv Complexity

We define Lempel-Ziv complexity recursively. Given the word $x_1^n := x_1 x_2,$ $\ldots, x_n$ of length n ($x_i \in \{0, 1\}$, $1 \leq i \leq n$), a block of length l ($1 \leq l \leq n$) is just a segment of x_1^n of length l, that is, a subsequence of l consecutive letters, say $x_{i+1}^{i+l} := x_{i+1} x_{i+2}, \ldots, x_{i+l}$ ($0 \leq i \leq n - l$). In particular, letters are blocks of length 1, and blocks of higher length are obtained by juxtaposition of blocks of lower length. Set $B_1 = x_1^1 = x_1$, and suppose that

$$B_1 B_2, \ldots, B_k = x_1^{n_k},$$

where $B_1 B_2, \ldots, B_k$ denotes the juxtaposition of the blocks B_1, $B_2 = x_2^{n_2}$, $,\ldots,, B_k = x_{n_{k-1}+1}^{n_k}$ and $n_{k-1} + 1 \leq n_k < n$ (with $n_0 = 0$ and $n_1 = 1$). Define

$$B_{k+1} := x_{n_k+1}^{n_{k+1}} \quad (n_k + 1 \leq n_{k+1} \leq n)$$

to be the block of minimal length such that it does not occur in the sequence $x_1^{n_{k+1}-1}$. Proceeding in this way, we obtain a decomposition of x_1^n in "minimal" blocks, say,

$$x_1^n = B_1 B_2, \ldots, B_p \tag{3.1}$$

in which only the last block B_p can occasionally appear twice. The complexity $C(x_1^n)$ of x_1^n is then defined as the number of blocks in the (clearly unique) decomposition 3.1:

$$C(x_1^n) := p.$$

The procedure is illustrated by the following example. The decomposition of the binary word $x_1^{19} = 0101101000110111\,0010$ into minimal blocks of new patterns is

0| 1| 011| 0100| 011011| 1001| 0,

where the vertical lines separate the blocks. Therefore, the complexity of x_1^{19} is 7.

Let us mention in passing that to define LZ-78 complexity, one compares the current block $x_{n_k+1}, \ldots, x_{n_k+l_k}$, $l_k = 1, 2, \ldots$, with the previous minimal blocks $B_1, \ldots, B_k$ (instead of looking for the same pattern in the whole segment $x_1, \ldots, x_{n_k+l_k-1}$, as in LZ-76) and sets $B_{k+1} = x_{n_k+1}^{n_k+l_k}$ as soon as it differs from all of them. For the word x_1^{19} of the example above, one now gets the block decomposition 0| 1| 01| 10| 100| 011| 0111| 00| 10, and hence its LZ-78 complexity is 9.

The generation rate of new patterns along x_1^n, a binary word of length n, is measured by the normalized complexity $c(x_1^n)$, which is defined by

$$c(x_1^n) = \frac{C(x_1^n)}{n/\log_2 n} = \frac{p}{n} \log_2 n.$$

Sequences that are not complex (e.g., periodic or quasi-periodic) have a very small normalized complexity. At the opposite end are the random sequences. Although the normalized complexity can take values higher than 1, the normalized complexity of sequences generated by random sources is about 1 with very high probability.

4 Relation Between Entropy and LZ-76 Complexity

To state the relation between the normalized complexity $c(x_1^n, \Delta\tau)$ of a binned spike train x_1^n and the entropy rate per second,

$$H(\Delta\tau) := \lim_{l \to \infty} H(l, \Delta\tau) = -\lim_{l \to \infty} \frac{1}{l\Delta\tau} \sum \tilde{p}_i \log_2 \tilde{p}_i \qquad (4.1)$$

of the neuronal source S, which has produced the binned spike train x_1^n, we still need two definitions. S is said to be stationary if the probability of any block x_i^{i+l-1} of length $l \geq 1$ does not depend on its position $i \geq 1$, which means that the statistical properties of the (in principle, arbitrarily long) words generated by S do not change with time. In general, the entropy fails to exist for nonstationary sources. A stationary source is called ergodic if ensemble averages and time averages coincide almost surely, that is, one can calculate expected values over the word ensemble using the relative frequencies of sufficiently long substrings from a typical word. Hence, ergodicity can be tested in practical cases by sampling typical words; every such word should produce (about) the same averages. As a rule, ergodicity is often encountered in nature for stationary processes. Maybe for this reason, ergodicity is, in general, tacitly assumed—always when entropy is estimated using word relative frequencies from a single sequence without further justification.

It can be proved (Ziv, 1978) that if S is stationary, then

$$\limsup_{n \to \infty} \frac{c(x_1^n, \Delta\tau)}{\Delta\tau} \leq H(\Delta\tau) \quad \text{on average,} \tag{4.2}$$

and, moreover, if S is ergodic, then

$$\limsup_{n \to \infty} \frac{c(x_1^n, \Delta\tau)}{\Delta\tau} = H(\Delta\tau) \quad \text{almost surely.} \tag{4.3}$$

Therefore, equations 4.2 and 4.3 provide ways to bound from below and esti-
mate, respectively, the entropy of a neuronal source (with the corresponding
properties) via the Lempel-Ziv complexity of a sample of spike trains and
a typical (i.e., randomly chosen) spike train produced by it.

Both equations 4.1 and 4.3 involve limits to infinity and hence are difficult
to implement in practice in order to estimate $H(\Delta\tau)$, "$l, n \to \infty$" meaning
that l, n are to be taken as large as possible. But the nature of the difficulty
is different. Whereas large l in $H(l, \Delta\tau)$ leads inevitably to undersampling
of the words of length l requiring some extrapolation technique to be put in
place (see section 5), large n or even small n in $c(x_1^n, \Delta\tau)/\Delta\tau$ can lead to good
approximations of $H(\Delta\tau)$ if the convergence in equation 4.3 is sufficiently
fast. This basic difference hints at the possibility, substantiated by numer-
ical evidence with short sequences, that the entropy rate can be estimated
by means of the normalized complexity in situations where the standard
estimator performs poorly.

5 Undersampling and Stationarity

When estimating the entropy rate of experimental time series, two main
difficulties arise (independent of the method used), which have to do with
the finiteness of the real signals and their stationarity.

First, the definition of $H(\Delta\tau)$, (see equation 4.1) requires words of in-
creasing length l, whereas real spike trains are necessarily finite. Now, in-
creasing l when counting different words out of spike trains of finite length
depletes the word statistics and therefore renders the estimations of the
relative frequencies $\tilde{p}_i$ of the words of length l less and less reliable. Also
as a result of this statistical depletion (or undersampling), $H(l, \Delta\tau)$ gets
artificially smaller than $H(\Delta\tau)$ for sufficiently large l, while $H(1, \Delta\tau) \geq
H(2, \Delta\tau) \geq \cdots \geq H(\Delta\tau)$ should hold.

Furthermore, stationarity is an assumption that cannot be taken for
granted in biological systems because of phenomena such as adaptation
and synaptic plasticity, and should be checked on a case-by-case basis. Usu-
ally one assumes that "short" registers are sufficiently stationary for practi-
cal purposes since the time variability of their statistical properties cannot
become important in the short term. In case of doubt, one should cut the

experimental registers even shorter to ensure stationarity. But, again, this assumption is at odds with the limit $l \to \infty$ in equation 4.1.

Therefore, for one reason or the other, one usually faces in practice short sequences when it comes to evaluating their entropy rate. The by now standard technique for so doing was proposed in Strong et al. (1998) and consists of extrapolating the linear trend (if any) of the graph of $H(l, \Delta \tau)$ versus $1/l$ up to the vertical axis obtaining the intercept as $H(\Delta \tau)$. As a real and practical alternative, we propose next a complexity-based approach to estimate the entropy rate of binned spike trains.

Indeed, equation 4.3 provides a simple way to estimate the entropy of an ergodic neuronal source via the normalized complexity rate of a typical binned spike train x_1^L of total length L produced by it, namely,

$$\frac{c(x_1^l, \Delta \tau)}{\Delta \tau} \simeq H(\Delta \tau),$$

for all $l \leq L$ such that $c(x_1^l, \Delta \tau)/\Delta \tau$ has already converged to $H(\Delta \tau)$. But as already noted, while the performance of the classical entropy rate estimator $H(l, \Delta \tau)$ depends critically on the size of l as compared to L because of undersampling, the applicability of the normalized complexity rate relies rather on the convergence speed of $c(x_1^l, \Delta \tau)$ as $l \to L$, that is, as the pattern count continues. If the convergence is fast, one can also assess the entropy rate of short data series—too short for other methods to deliver. In particular, one expects this to be the case when the source entropy is small (i.e., when the source is far from random), because then the pattern count is low. This is important because stationarity often demands experimental spike trains to be short. Numerical simulations with two-state Markov processes give support to our claim that the normalized complexity rate converges fast to $H(\Delta \tau)$. The next section is devoted to these simulations.

6 Numerical Simulations

The entropy estimation of signals of finite and, especially, short length is a challenging problem. In order to tackle this problem for the estimators in which we are interested, the normalized Lempel-Ziv complexity and the usual estimate based on the word relative frequencies (see equation 2.1), we consider two-state (numbered 0 and 1) Markov processes in discrete time. The conditional probabilities for such processes are completely defined in terms of the transition probabilities from state 0 to state 1, p_{10}, and from state 1 to state 0, p_{01}. The whole Markov transition probability matrix $P_{n+1,n}(x \mid y)$, where $x, y = 0, 1$, is given by

$$P_{n+1,n}(x \mid y) = \begin{pmatrix} 1 - p_{10} & p_{01} \\ p_{10} & 1 - p_{01} \end{pmatrix}. \tag{6.1}$$

The equation for the probability evolution (master equation) then reads

$$\begin{pmatrix} P_{n+1}(0) \\ P_{n+1}(1) \end{pmatrix} = \begin{pmatrix} 1 - p_{10} & p_{01} \\ p_{10} & 1 - p_{01} \end{pmatrix} \begin{pmatrix} P_n(0) \\ P_n(1) \end{pmatrix}, \tag{6.2}$$

and it has the stationary solution

$$\begin{pmatrix} P_{eq}(0) \\ P_{eq}(1) \end{pmatrix} = \begin{pmatrix} p_{01}/(p_{01} + p_{10}) \\ p_{10}/(p_{01} + p_{10}) \end{pmatrix}. \tag{6.3}$$

The entropy rate H of such a process (or, equivalently, Markov source) generating an infinite long binary sequence is given by Cover and Thomas (1991):

$$\begin{aligned} H = P_{eq}(0)(&-p_{10} \log_2 p_{10} - (1 - p_{10}) \log_2(1 - p_{10})) \\ + P_{eq}(1)(&-p_{01} \log_2 p_{01} - (1 - p_{01}) \log_2(1 - p_{01})). \end{aligned} \tag{6.4}$$

On the other hand, the entropy rate estimate obtained as in equation 2.1 from the statistics of l-bit long words in an infinite long binary sequence generated by a Markov source can be written as the convex sum

$$H(l) = \frac{l-1}{l} H + \frac{1}{l} H_{eq}, \tag{6.5}$$

$(l = 1, 2, \ldots)$, where H is the source entropy rate, equation 6.4, and

$$H_{eq} = -P_{eq}(0) \log_2 P_{eq}(0) - P_{eq}(1) \log_2 P_{eq}(1). \tag{6.6}$$

Notice from equation 6.5 that $H(l = 1) = H_{eq}$,

$$\lim_{l \to \infty} H(l) = H \tag{6.7}$$

(as it should by definition) and, in general, $H \leq H(l) \leq H_{eq}$.

In the particular case of a Markov source with

$$p_{01} + p_{10} = 1, \tag{6.8}$$

both extremal values H_{eq} and H coincide ($= -p_{01} \log_2 p_{01} - p_{10} \log_2 p_{10}$) and, hence,

$$H = H(l) = H_{eq} \tag{6.9}$$

for words of any length l. In particular, $H = H(l = 1)$, which shows that if $p_{01} + p_{10} = 1$, one can estimate the entropy rate with arbitrary precision just by sampling a sufficiently large number of single bits. But this is rather an exceptional situation.
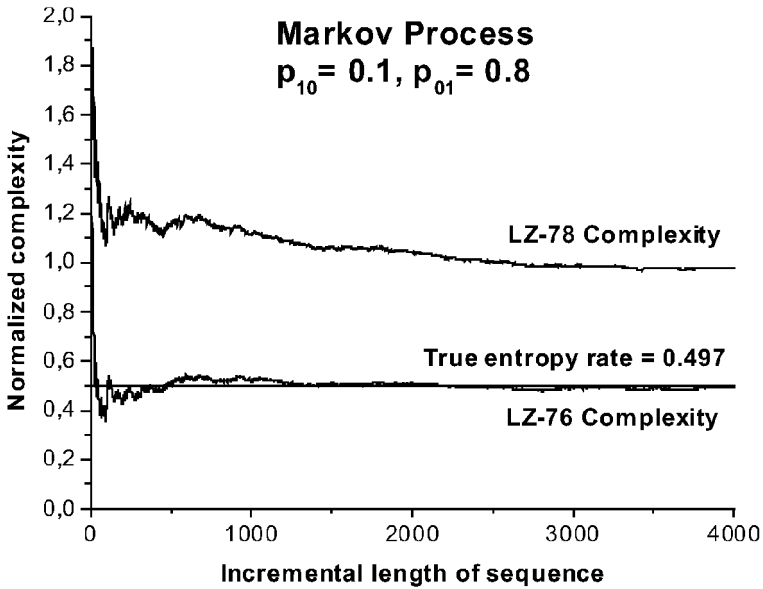
Figure 2: Convergence rate with increasing word length of the normalized Lempel-Ziv complexity to the true entropy rate of the Markov source, which has generated the sequence. Here, $p_{10} = 0.1$ and $p_{01} = 0.8$. The corresponding curve for the normalized LZ-78 complexity is shown for comparison.

In general, $p_{01} + p_{10} \neq 1$, and since $H := \lim_{l \to \infty} H(l)$, the reliable estimation of H through $H(l)$ requires a large sample of long words, while experimental data series are usually too short to satisfy this need. Moreover, given a finite binary sequence of length n, it can be shown that $H(l)$ decays as

$$H_{\text{asym}}(l) = \frac{\log_2(n - l + 1)}{l} \tag{6.10}$$

for sufficiently long words, so that $\lim_{l \to n} H(l) = \lim_{l \to n} H_{\text{asym}}(l) = 0$. This is the undersampling scenario one faces in practice because of the finite length of real data series.

We present in Figures 2 through 6 the results of our numerical simulations pertaining to the estimation of the entropy H via the normalized Lempel-Ziv complexity and by direct sampling of word probabilities. We have performed many such simulations, but only two representative cases will be discussed here.

Specifically, Figures 2 and 3 exhibit the convergence rate of the normalized Lempel-Ziv complexity to H for Markov sources with transition probabilities $p_{10} = 0.1$, $p_{01} = 0.8$ in Figure 2 and $p_{10} = p_{01} = 0.05$ in Figure 3.
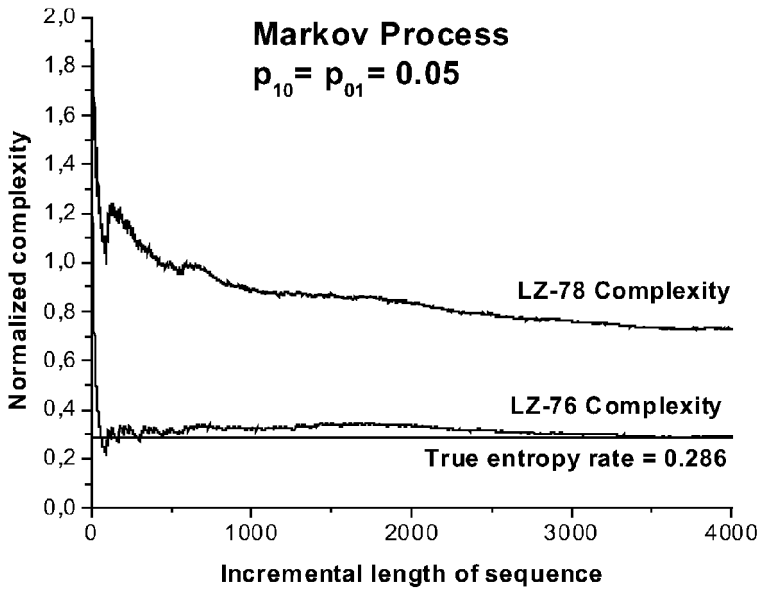
Figure 3: Convergence rate with increasing word length of the normalized Lempel-Ziv complexity to the true entropy rate of the Markov source that has generated the sequence. Here, $p_{10} = p_{10} = 0.05$. The corresponding curve for the normalized LZ-78 complexity is also shown for comparison.

The exact values of the entropy rate for these transition probabilities are, according to equation 6.4, $H = 0.497$ and $H = 0.286$ bits per symbol, respectively. We present also the corresponding curves of the normalized LZ-78 complexity to show that the former (LZ-76) converges much faster than the latter.

Figures 4 and 5 compare both approaches for the extreme case of a binary sequence of only 200 bits generated by two Markov processes with the same transition probabilities as before. Whereas the standard method errs from the exact values of the entropy rate by as much as 22% (the first case) and 24% (the second case), the worst relative error of the complexity-based estimation is only 8%. These and similar simulations support the utility of the complexity-based estimations for short sequences.

Figure 6 shows that for sequences of 4000 bits, the standard estimator performs flawlessly, as expected. The relative error is about 1% for both Markov processes. The corresponding estimations from the normalized complexity can be read in Figures 2 and 3, and the relative error is also 1%.

We conclude after the preceding benchmarking with two-state Markov processes that the normalized complexity can be applied successfully in a variety of situations, and its performance, at least in the controlled envi-
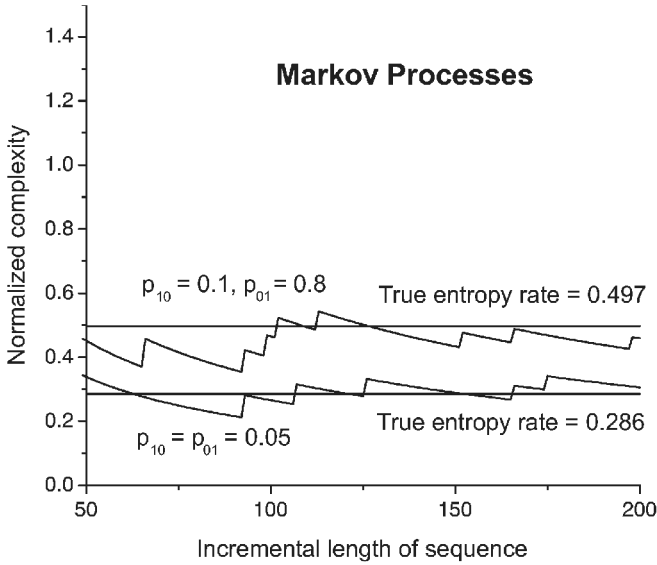
Figure 4: A close-up of the LZ-76 complexity curves of Figures 2 and 3 in the interval $50 \leq n \leq 200$. It can be seen that the normalized complexity rate is already close to the true entropy rate, the relative error being less than 14% around $n = 200$ and 8% at $n = 200$ in either case.
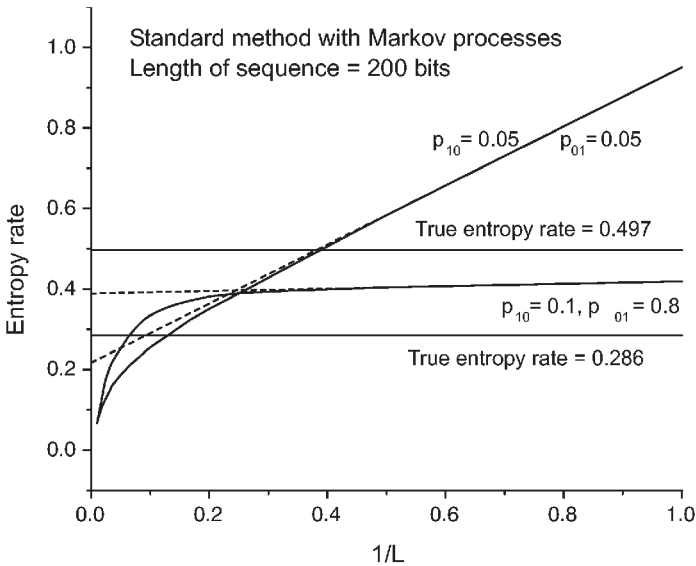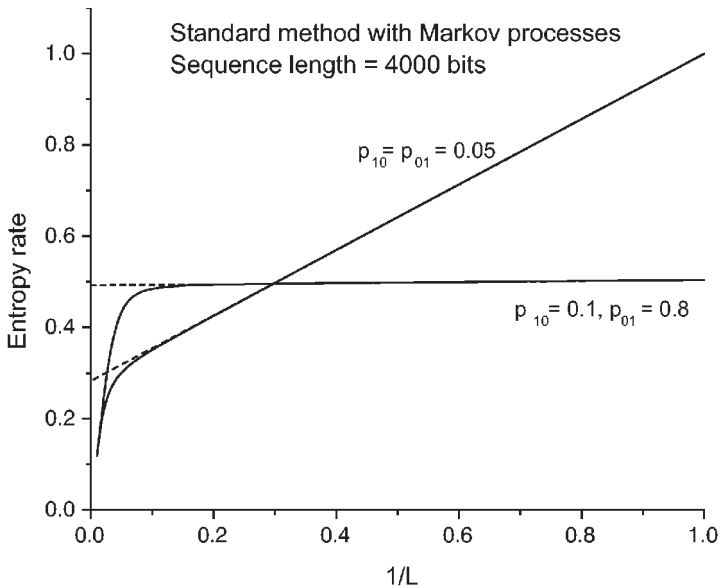

Figure 5: The standard extrapolation technique applied to the same cases as in Figure 4 (sequence length = 200 bits) misses the true entropy rate value by a relative error of 22% (first Markov process) and 24% (second Markov process).

Figure 6: The standard extrapolation technique applied to the same cases as in Figures 2 and 3 (sequence length = 4000 bits). Now the relative error is about 1%, the same as if the entropy rate is read from the corresponding normalized complexity rate curve in Figures 2 or 3.

ronment of numerical simulation, is at least as good as that of the standard method. We turn next to the neural experimental data.

7 Experimental Work and Results

In this section, we continue the comparison of the two methods, this time with experimental data. We have included for this purpose recordings of different discharge patterns obtained both in vivo and in vitro and induced by a variety of stimuli.

The experimental data were obtained from primary cortex recordings both in vivo and in brain slice preparations (in vitro). Intracellular recordings in vivo were obtained from anesthetized adult cats (see Sanchez-Vives, Nowak, & McCormick, 2000a, for details). For the preparation of slices, ferrets 2 to 4 months old of either sex were used (see Sanchez-Vives, Nowak, & McCormick, 2000b, for details). Action potentials were detected with a window discriminator, and the time of their occurrence was collected with a 10 μsec resolution. The resulting time series were used to analyze the neuronal spiking. Concerning the stimuli, they were of three kinds:

1. Intracellular periodic current injection. Intracellular sinusoidal currents were injected in vivo. The frequency of the waveform was 2 Hz, and the intensity ranged between 0.2 and 1.5 nA. The cell recorded ensemble had 8 samples (spike train lengths between 15.56 sec and 47.64 sec).

2. Visual stimulation with sinusoidal drifting gratings. The visual stimulus consisted of a sinusoidal drifting grating presented in a circular patch of 3 to 5 degrees diameter, centered on the receptive field center (in vivo). Only simple cells (classified as in Skottun et al., 1991) were included in this study. In this case, 8 samples were analyzed (spike train lengths between 15.87 sec and 23.62 sec).

3. Intracellular random current injection. Random currents ranging between $-/+ 1.5$ nA with different degrees of correlations were injected during the intracellular recordings from cortical brain slices (in vitro). The ensemble consisted of 20 samples (spike train lengths between 16.32 sec and 35.47 sec).

Figures 7 and 8 present the complexity rates $c(x_1^n, \Delta\tau)/\Delta\tau$ versus the coding frequency $1/\Delta\tau$ of a randomly chosen spike train x_1^n belonging, re-
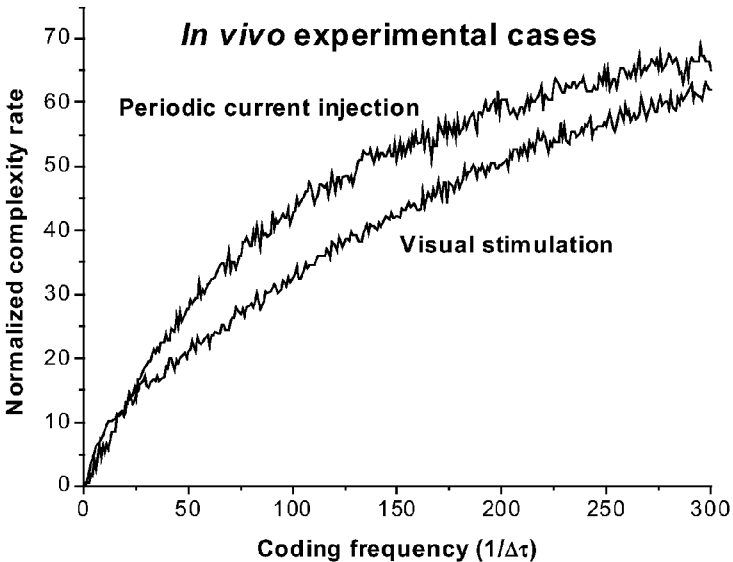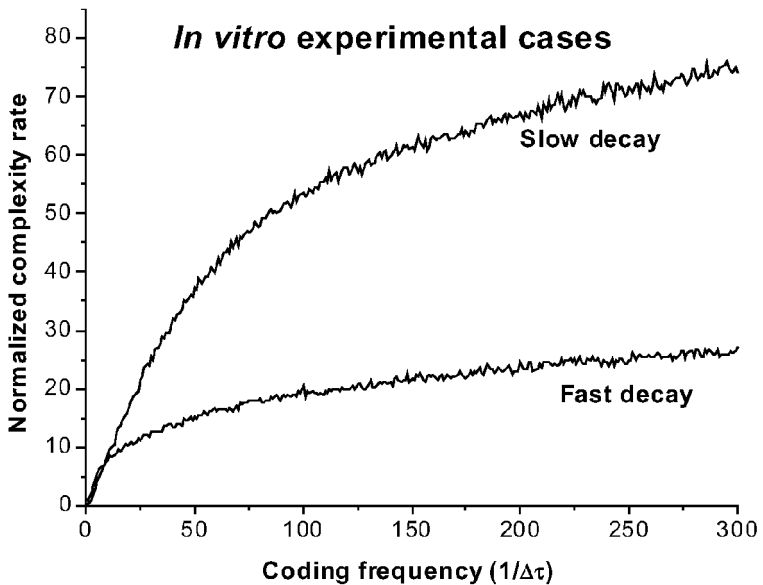

Figure 7: Normalized complexity rate $c(\Delta\tau)/\Delta\tau$ versus coding frequency $1/\Delta\tau$ for the two in vivo experimental cases considered in the text: intracellular periodic current injection and visual stimulation with sinusoidal drifting gratings. The values for $1/\Delta\tau = 100$, 200, and 300 Hz are presented in Tables 1 and 2. The spike train is "typical" in the sense that it was chosen randomly.

Figure 8: Normalized complexity rate $c(\Delta\tau)/\Delta\tau$ versus coding frequency $1/\Delta\tau$ for the two in vitro experimental cases considered in the text: intracellular random current injection with stimuli whose autocorrelation functions decay fast or slow. The spike trains are typical. The values for $1/\Delta\tau = 100, 200,$ and 300 Hz are presented in Tables 3 and 4.

spectively, to the in vivo (periodic current injection and visual stimulation) and in vitro (random current injection) experimental cases just explained. Notice that we have further split the responses to random stimuli into two subsets according to whether the autocorrelation function of the stimuli decays slowly or fast. Only spike trains for which this distinction were clear were considered for analysis. Figures 9 to 12 show the graphs $1/l \mapsto H(l, \Delta\tau)$ with $\Delta\tau = 0.01, 0.005, 0.0033$ sec used to estimate the corresponding entropy rate $H(\Delta\tau)$ by the standard technique of Strong et al. (1998) for the same spike trains as before.

Tables 1 to 4 summarize the entropy rate estimations (in bits per second) obtained by the standard and complexity-based methods applied to the same spike train in each of the four previous experimental cases. The complexity-based estimations are the readings for $1/\Delta\tau = 100, 200, 300$ Hz of the computer program that calculates the complexity curves in Figures 7 and 8. The values listed as standard estimations were numerically extrapolated from the linear trends of the corresponding graphs, Figures 9 to 12.
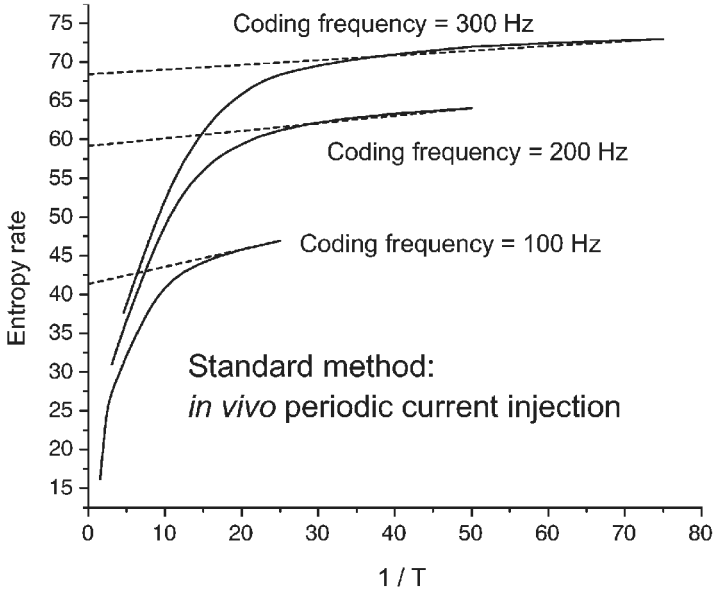
Figure 9: The standard estimate of the entropy rate in the in vivo periodic current injection case for the corresponding spike train of Figure 7, obtained by extrapolating the linear trend of $H(l, \Delta\tau)$ versus $1/T$ ($T = l\Delta\tau$, the window length in time) to infinitely long windows. The time bins were $\Delta\tau = 0.01, 0.005$, and 0.0033 sec (coding frequency = 100, 200, and 300 Hz, respectively). The corresponding estimations are presented in Table 1.

Table 1: $H(\Delta\tau)$ for Periodic Current Injection (in vivo).

Coding Frequency	Standard	Complexity
100 Hz	41.38	42.93
200 Hz	59.20	60.40
300 Hz	68.42	67.00

Table 2: $H(\Delta\tau)$ for Visual Stimulation.

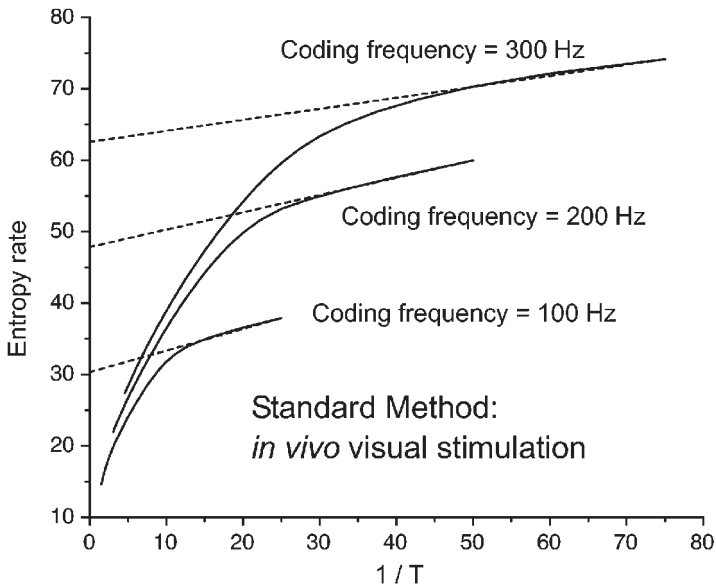| Coding Frequency	Standard	Complexity
100 Hz	30.30	32.78
200 Hz	47.85	50.14
300 Hz	62.55	62.11

Figure 10: The standard estimate of the entropy rate in the in vivo visual stimulation with sinusoidal drifting gratings case for the corresponding spike train of Figure 7. The time bins were $\Delta\tau = 0.01, 0.005$, and 0.0033 sec (coding frequency = 100, 200, and 300 Hz, respectively). The corresponding estimations are presented in Table 2.

Table 1 compares the entropy rate estimations obtained by the two methods for a typical neuron response to periodic current injection (in vivo). The agreement is remarkable in this case. The relative deviation of the estimations decreases from 8% for $1/\Delta\tau = 100$ Hz to below 1% for $1/\Delta\tau = 300$ Hz.

Analogously, Table 2 shows the estimations obtained by both methods from a typical neuron response to visual stimulation with sinusoidal drifting gratings. Here the relative deviation falls slightly from 3% for $1/\Delta\tau = 100$ Hz to 2% for $1/\Delta\tau = 200$ and 300 Hz.

Table 3 shows the results obtained for a spike train elicited by random current injection (in vitro). Furthermore, the spike train belongs to the samples with a slowly decaying autocorrelation function. The relative deviation of the estimations jumps from 2% for $1/\Delta\tau = 100$ and 200 Hz to roughly 5% for $1/\Delta\tau = 300$.

We finish the comparison between the standard and the complexity-based entropy rate estimators with Table 4, which summarizes the results for a spike train with a fast-decaying autocorrelation function elicited by random current injection. This is the case where the agreement is worse, the relative deviation of estimations fluctuating around the 15% level.
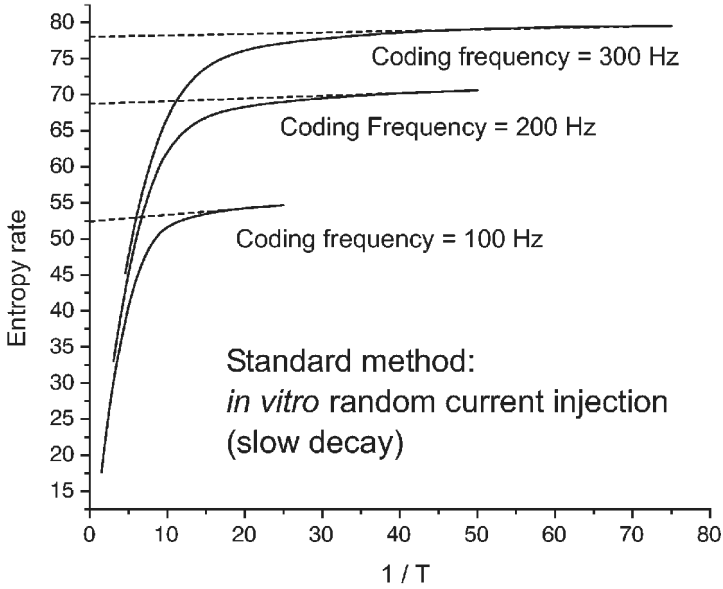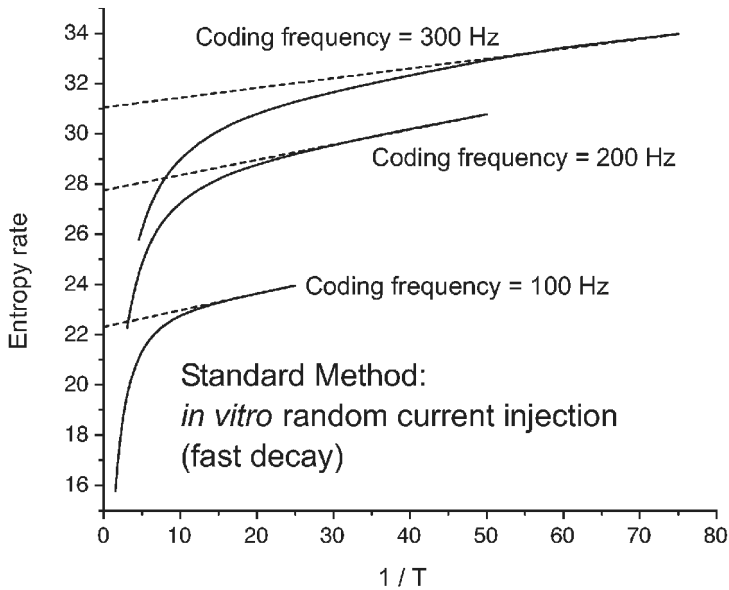
Figure 11: The standard estimate of the entropy rate in the in vitro random current injection with slowly decaying autocorrelation function case for the corresponding spike train of Figure 8. The time bins were $\Delta \tau = 0.01, 0.005,$ and 0.0033 sec (coding frequency = 100, 200, and 300 Hz, respectively). The corresponding estimations are presented in Table 3.

Table 3: $H(\Delta \tau)$ for Random Current Injection (Slow Decay).

Coding Frequency	Standard	Complexity
100 Hz	52.38	53.38
200 Hz	68.69	67.23
300 Hz	78.00	74.70

Table 4: $H(\Delta \tau)$ for Random Current Injection (Fast Decay).

Coding Frequency	Standard	Complexity
100 Hz	22.31	19.00
200 Hz	27.75	24.39
300 Hz	31.05	26.03

Figure 12: The standard estimate of the entropy rate in the in vitro random current injection with fast-decaying autocorrelation function case for the corresponding spike train of Figure 8. The time resolutions were $\Delta\tau = 0.01, 0.005,$ and 0.0033 sec (coding frequency = 100, 200, and 300 Hz, respectively). The corresponding estimations are presented in Table 4.

Summing up, the results are very satisfactory in both in vivo cases and in vitro random injection with a slowly decaying autocorrelation function. For in vitro random injection with a fast-decaying autocorrelation function, we find a 15% discrepancy, the normalized complexity rate lying always below the standard entropy rate estimations. We conjecture that this discrepancy is due to nonstationarity.

8 Conclusions

We propose the normalized Lempel-Ziv complexity (LZ-76 to be more precise) as a reliable estimator of the entropy of neuronal sources. The exact relation between both concepts is given in equations 4.2 and 4.3. Our numerical simulations with two-dimensional Markov processes with different transition probabilities show that the normalized complexity converges fast to the entropy; in section 6, we reported on two such simulations. A further advantage of the complexity approach is that there is no need to know the probability law of the process generating the signal since the normalized

complexity is a property of individual sequences. More important, the numerical evidence gathered over time with Markov processes and partially reported in this article strengthens the validity of the complexity approach. Indeed, the entropy rate estimations of artificial registers by means of the complexity rate compare very favorably with those made with the extrapolation technique of Strong et al. (1998) and, moreover, outperforms this latter method for some sequences too short for it to deliver an accurate estimation but long enough for the normalized complexity rate to be close to the entropy rate. In sum, we conclude that Lempel-Ziv complexity certainly belongs in the toolbox of computational neuroscience as a relevant measure of entropy.

Acknowledgments

References

Amigó, J. M., Szczepański, J., Wajnryb, E., & Sanchez-Vives, M. V. (2003). On the number of states of the neuronal sources. *BioSystems, 68*, 57–66.

Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nature Neurosci., 2*, 947–957.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Lempel, A., & Ziv, J. (1976). On the complexity of an individual sequence. *IEEE Trans. Inform. Theory, IT-22*, 75–88.

MacKay, D., & McCulloch, W. S. (1952). The limiting information capacity of a neuronal link. *Bull. Math. Biophys., 14*, 127–135.

Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Comp., 15*, 1191–1253.

Reinagel, P., & Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *J. Neurosci., 20*, 5392–5400.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1998). *Spikes: Exploring the natural code*. Cambridge, MA: MIT Press.

Sanchez-Vives, M. V., Nowak, L. G., & McCormick, D. A. (2000a). Membrane mechanisms underlying contrast adaptation in cat area 17 in vivo. *J. Neurosci., 20*, 4267–4285.

Sanchez-Vives, M. V., Nowak, L. G., & McCormick, D. A. (2000b). Cellular mechanisms of long-lasting adaptation in visual cortical neurons in vitro. *J. Neurosci., 20*, 4286–4299.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. Journal, 27*, 379–423, 623–656.

Skottun, B. C., DeValois, R. L., Grosof, D. H., Movshon, J. A., Albrecht, D. G., & Bonds, A. B. (1991). Classifying simple and complex cells in the basis of response modulation. *Vision Res., 31*, 1079–1086.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, & Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett., 80*, 197–200.

Ziv, J. (1978). Coding theorems for individual sequences. *IEEE Trans. Inform. Theory, IT-24*, 405–412

Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory, IT-24*, 530–536.