LETTER Communicated by Jack Cowan

Attractor Reliability Reveals Deterministic Structure in Neuronal Spike Trains

P. H. E. Tiesinga
tiesinga@salk.edu
Sloan-Swartz Center for Theoretical Neurobiology and Computational Neurobiology Lab, Salk Institute, La Jolla, CA 92037, U.S.A.

J.-M. Fellous
fellous@salk.edu
Howard Hughes Medical Institute and Computational Neurobiology Lab, Salk Institute, La Jolla, CA 92037, U.S.A.

Terrence J. Sejnowski
terry@salk.edu
Sloan-Swartz Center for Theoretical Neurobiology and Computational Neurobiology Lab, Salk Institute; Howard Hughes Medical Institute, Salk Institute and Department of Biology, University of California–San Diego, La Jolla, CA 92037, U.S.A.

When periodic current is injected into an integrate-and-fire model neuron, the voltage as a function of time converges from different initial conditions to an attractor that produces reproducible sequences of spikes. The attractor reliability is a measure of the stability of spike trains against intrinsic noise and is quantified here as the inverse of the number of distinct spike trains obtained in response to repeated presentations of the same stimulus. High reliability characterizes neurons that can support a spike-time code, unlike neurons with discharges forming a renewal process (such as a Poisson process). These two classes of responses cannot be distinguished using measures based on the spike-time histogram, but they can be identified by the attractor dynamics of spike trains, as shown here using a new method for calculating the attractor reliability.

We applied these methods to spike trains obtained from current injection into cortical neurons recorded in vitro. These spike trains did not form a renewal process and had a higher reliability compared to renewal-like processes with the same spike-time histogram.

1 Introduction

Features that are present in the spike patterns elicited in response to a stimulus repeated across multiple trials can form the basis of a neuronal code. Here, we introduce a novel reliability measure in order to study the

 © 2002 Massachusetts Institute of Technology

reproducibility of sequences of precise spike times produced in vitro by cortical neurons (Mainen & Sejnowski, 1995). We show that spike trains obtained from in vitro cortical neurons and integrate-and-fire model neurons have more deterministic structure than spike trains obtained from renewal processes with the same interspike interval and spike-time probability distribution. This structure cannot be detected from the spike-time histogram.

The response of a neuron to a stimulus can be characterized as an attractor, defined as the voltage trajectory (and associated spike train) to which the neuron's membrane potential converges from different initial conditions (Strogatz, 1994; Jensen, 1998). An attractor is stable against noise: weak noise introduces spike-time jitter, but the sequence of spike times and the input-induced correlations between interspike intervals is conserved. Here, the sequence of spike times is considered the output signal. When the neuron stays on the attractor, it transmits a unique signal.

The dynamics of a model neuron can undergo a bifurcation when its parameter values are varied. During a bifurcation, a small change in the value of a parameter introduces a large change in the spike times: a different attractor emerges. When a neuron is close to a bifurcation point, noise can induce a transition to another attractor. If the new "attractor" is unstable, the neuron will return to the original attractor after a finite time, but if it is stable, the neuron stays on the new attractor. The attractor reliability is defined as the inverse of the number of distinct spike trains that are obtained after a large number of trials.

This article consists of two parts. First, a test is given to determine whether a spike train forms a temporally modulated renewal process. Second, we quantify the extra—non-Poisson—structure present in spike trains using the attractor reliability. These techniques are illustrated using spike trains obtained from experimental data and numerical model simulations. A systematic study of the attractor and bifurcation structure of in vitro and model neurons driven by periodic, quasiperiodic, and random currents will be presented elsewhere.

2 Methods

2.1 Experimental Methods. The voltage response of cortical neurons as measured in a rat slice preparation was described previously (Fellous et al., 2001). Protocols for these experiments were approved by the Salk Institute Animal Care and Use Committee, and they conform to U.S. Department of Agriculture regulations and National Institutes of Health guidelines for humane care and use of laboratory animals. Briefly, coronal slices of rat prelimbic and infralimbic areas of prefrontal cortex were obtained from 2- to 4-week-old Sprague-Dawley rats. Rats were anesthetized with isoflurane and decapitated. Their brain was removed and cut into 350 μm thick slices on a Vibratome 1000. Slices were then transfered to a submerged chamber containing artificial cerebrospinal fluid (ACSF, mM: NaCl, 125; NaH_2CO_3, 25;

D-glucose, 10; KCl, 2.5; $CaCl_2$, 2; $MgCl_2$, 1.3; NaH_2PO_4, 1.25) saturated with 95% O_2/5% CO_2 at room temperature. Whole cell patch clamp recordings were achieved using glass electrodes containing (4-10 MΩ: mM; $KmeSO_4$, 140; Hepes, 10; NaCl, 4; EGTA, 0.1; Mg-ATP, 4; Mg-GTP, 0.3; Phosphocreatine 14). Patch clamp was performed under visual control at 30–32°C. In most experiments, Lucifer yellow (RBI, 0.4%) or Biocytin (Sigma, 0.5%) was added to the internal solution for morphological identification. In all experiments, synaptic transmission was blocked by D-2-amino-5-phosphonovaleric acid (D-APV; 50 μM), 6,7-dinitroquinoxaline-2,3, dione (DNQX; 10 μM), and biccuculine methiodide (Bicc; 20 μM). All drugs were obtained from RBI or Sigma, freshly prepared in ACSF and bath applied. Data were acquired with Labview 5.0 and a PCI-16-E1 data acquisition board (National Instrument) and analyzed with MATLAB (The Mathworks). We used regularly spiking layer 5 pyramidal cells that were identified morphologically.

2.2 Simulation Algorithm. The membrane potential V of an integrate-and-fire neuron driven by a periodic current satisfied,

$$\frac{dV}{dt} = -V + I + A \sin \frac{2\pi}{T} t + \xi(t), \tag{2.1}$$

where I was a time-independent driving current, A was the amplitude of the drive, $T = 2$ was the period, and ξ was a white noise current, with zero mean and variance D, that represented the effects of intrinsic noise. When the voltage V reached threshold, $V(t) = 1$, a spike was emitted, and the voltage was instantaneously reset to zero, $V(t) = 0$.

Dimensionless units were used in model simulations. One voltage unit corresponded to the distance between resting membrane potential and action potential threshold, approximately 20 mV; one time unit corresponded to the membrane time constant, approximately 40 ms.

Equation 2.1 was integrated directly using the fourth-order Runge-Kutta algorithm (Press, Teukolsky, Vetterling, & Flannery, 1992), with step size $dt = 0.01$. The calculated voltage differed by less than 10^{-11} from the voltage obtained by analytically integrating equation 2.1 for a sinusoidal current and with $D = 0$. The spike time t_s, given by the expression $V(t_s) = 1$, was determined by linear interpolation (Hansel, Mato, Meunier, & Neltner, 1998).

3 Results

3.1 Example of Deterministic Structure in Spike Trains. In most experiments, the same stimulus was presented on different trials, and the resulting spike trains were analyzed. The spike trains of a hypothetical experiment are presented in a rasterplot in Figure 1Aa. The computation of reliability and precision based on the spike-time histogram (STH) followed the procedure in Mainen and Sejnowski (1995). The length of a trial was

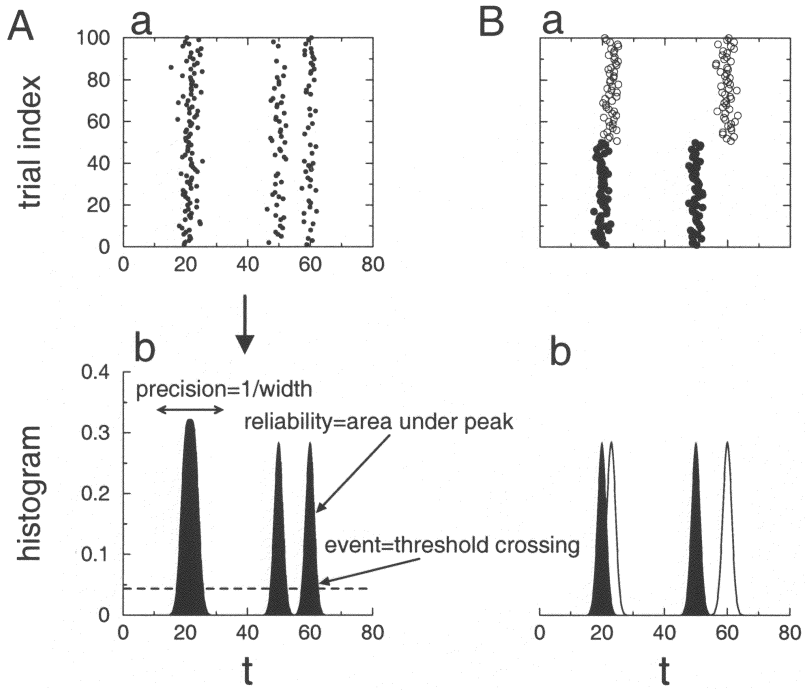

Figure 1: Reliability based on the spike-time histogram is insensitive to deterministic structure in spike trains. (A, B) The same spike times, with the trials ordered differently. On each row: (a) Rasterplot, with each measured spike represented as a circle, its x-ordinate is the spike time and its y-ordinate the trial index; (b) the spike-time histogram (STH)—the number of spikes in a particular time bin normalized by the number of trials. Events were defined as threshold crossings of the spike-time histogram. The event reliability was the fraction of trials during which a spike occurred during the event; the precision was the inverse of the standard deviation of the spike times in the event. The reliability of the response, R_{STH}, was the event reliability averaged over all events. (A) In each of the 100 trials, a spike was present during the first event; hence, its reliability was 1. (B) The trials of A were composed of two distinct spike trains, indicated by the filled and open circles in a. In b, the resulting STH for each group is plotted separately as filled and open peaks, respectively. The first peak in A was in fact made up of two separate events.

divided into discrete bins, and the number of spikes that fell in each bin was counted. The spike count in a bin was normalized by the number of trials. The set of bins was convolved with a smoothing function. The STH so obtained usually contained a number of peaks, each consisting of a set of consecutive bins that contain more spike counts than average. Peaks were

detected by setting a threshold and determining when the smoothed STH crossed it (see Figure 1Ab). The peaks constitute events; the reliability of an event was the fraction of trials during which a neuron spiked during the event. The reliability of the spike train, R_{STH}, was the reliability of each event averaged over all events. (An equivalent procedure for R_{STH} is given in equation 3.5.) The precision of an event was the inverse of the standard deviation of the spike times that were part of the peak. The reliability based on the STH depended on the size of the bins, the smoothing function, and the value of the peak-detection threshold. The STH reliability so defined was insensitive to deterministic structure in the spike trains.

An example of how STH reliability can miss important deterministic structure in spike trains is shown in Figure 1. The data shown in Figure 1Aa were obtained by randomly mixing two types of spike trains: the open and filled circles in Figure 1Ba. Hence, Figures 1Aa and 1Ba are the same, except that the trials are reordered. On each trial, there was a spike present in the first event in Figure 1Ab; hence, its "reliability" was 1. However, that single event in fact consisted of two events (see Figure 1Bb). In other words, there were two distinct spike trains (attractors) present across different trials.

3.2 Testing of Renewal Property. A temporally modulated renewal spike train is fully characterized by its spike-time probability density and its interspike-interval distribution. The defining property of a renewal spike train is that the intervals between spikes are independent. There is no deterministic structure in renewal spike trains apart from the structure induced by a time-varying firing rate (see below). In contrast, for an attractor, the spikes occur at particular times and in a particular sequence (see section 1): there are correlations between the spike times, and there is deterministic structure in the spike train. Hence, it is different from a renewal process. Spike trains that form a renewal process are not reliable. It is therefore important to distinguish renewal processes from nonrenewal processes (Reich, Victor, & Knight, 1998).

The Poisson process is an example of a renewal process with rate λ. The probability of finding a spike between times t and $t + dt$ is λdt and does not depend on previous spike times. Hence, the spike times themselves are also independent. The probability λ is time independent and is estimated from the spike-time histogram. The distribution of interspike intervals is exponential; the probability of obtaining an interspike interval between τ and $\tau + d\tau$ is $\lambda e^{-\tau\lambda} d\tau$. There is no correlation between the length of consecutive intervals. The coefficient of variation (CV), the standard deviation divided by the mean of τ, is equal to 1. Another example of a renewal process is a gamma process of order r. The probability of obtaining a spike between times t and $t + dt$ again is λdt. However, the distribution of interspike intervals is given by a gamma probability distribution and is less disperse, with $CV = \frac{1}{\sqrt{r}}$ and $r > 1$. The gamma process of integer order is closely related

to the Poisson process. For instance, a gamma process of order 2 is obtained from a Poisson process of twice the rate, 2λ, by removing every other spike.

Shuffling procedures can form the basis for a test of the renewal character of spike trains. First, generate a number of different spike trains (trials) using the same process. Then shuffle the spike times randomly across different trials. The probability of obtaining a spike at a given time is conserved, since exactly the same spike times were used. If the spikes were generated by a Poisson process, then they are independent of each other. Hence, the new set of spike trains cannot be distinguished statistically from the original set of spike trains. The shuffling procedure leads to an exponential distribution of interspike intervals, and $CV = 1$. For spike trains generated by a gamma process of order r, $CV = \frac{1}{\sqrt{r}}$. However, after shuffling, the CV changed to 1, and the new set of spike trains can be distinguished from the original set of spike trains by comparing the CVs. Hence, shuffling of spike times is not appropriate for a gamma process. However, when the interspike intervals are shuffled randomly across trials and then used to calculate the new spike times, the interspike interval distribution is conserved. Since the intervals in a renewal process are independent, the new set of spike trains should be indistinguishable from the original set. Note that the actual spike times are not conserved. As a result, the spike-time probability (the spike-time histogram) is different at the boundaries, close to the beginning and end of the trial. However, at some distance from the boundaries, it is approximately the same. An example is discussed below.

The spike-time histogram obtained from neural spike trains usually is time dependent: $\lambda(t)$ is a function of time. The spike trains may still form a temporally modulated renewal process; the probability of obtaining a spike between t and $t+dt$ is $\lambda(t)dt$. The time dependence of $\lambda(t)$ makes testing of the renewal character difficult, since the distribution of interspike intervals depends on time via $\lambda(t)$. However, temporally modulated renewal processes (referred to as "simply modulated renewal process" in Reich et al. (1998)) can be mapped into a time-independent (homogeneous) renewal process by "transforming" time. In what follows, we describe the test of the renewal property in detail. It consisted of three steps. First, time t was mapped into a new time s. Then the interspike intervals were shuffled randomly across trials. Finally, the test statistic was evaluated, and its significance was determined. The procedure is applied to spike trains obtained from repeated current injection into cortical neurons in vitro.

3.2.1 Experimental Data Set. The stimulus wave forms shown in Figure 2Ac were injected in a cortical neuron in current-clamp mode. The stimulus consisted of a 200 ms constant depolarizing current followed by a 1600 ms sinusoidal current with period 100 ms. The height of the initial depolarizing pulse was varied to bring the neuron to different initial voltages at the start of the sinusoidal wave form. For a large enough amplitude, a spike was obtained during the current pulse (see Figure 2Ab). After a transient

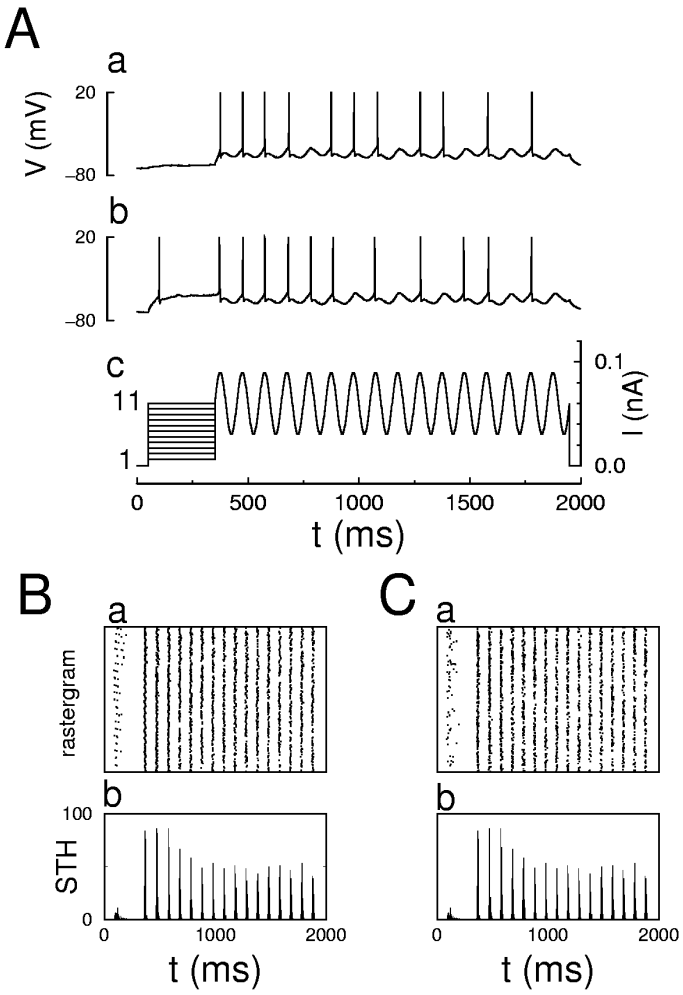

Figure 2: Effect of initial condition on spike timing in cortical neurons. The stimulus wave form consisted of a 200 ms long constant depolarizing current followed by a 1600 ms long sinusoidal current with period 100 ms. The amplitude of the initial depolarizing current was varied over 11 different values. Each of the resulting wave forms was presented 20 times, yielding 220 trials. (A) (a, b) Recorded voltage response during the presentation of the first and eleventh stimulus wave form, respectively; (c) the 11 stimulus wave forms. (B, C) In each row, (a) the rastergram and (b) the spike-time histogram is plotted using (B) the experimental data; (C) Poisson spike trains that were obtained by randomly shuffling spike times across trials. The spike-time histograms were identical, and the rastergrams looked similar. However, the experimental spike trains did not form a renewal process (see Figure 4).

of approximately three cycles, the neuron generated on average about one action potential on every two cycles. The rastergram and spike-time histogram of the spike trains obtained during 20 presentations of 11 different wave forms (220 trials) are shown in Figure 2B. A Poisson process with the same spike-time histogram was generated by randomly shuffling the spike times across different trials (see Figure 2C).

3.2.2 Mapping the Spike Times. The spike time probability $\lambda(t)$ was estimated from the spike-time histogram $P(t)$. Note that $P(t)$ was evaluated in discrete bins; however, for simplicity, the notation used here does not reflect this. Time t was transformed into a new time variable s such that the probability of obtaining a spike between s and $s + ds$ was independent of time s (Reich et al., 1998). This procedure transformed an inhomogeneous renewal process into a homogeneous renewal process. The transformation was $t \to s = g(t)$ with

$$g(t) = T_{STH} \frac{\int_0^t du\, P(u)}{\int_0^{T_{STH}} du\, P(u)}, \tag{3.1}$$

where u was an integration variable and T_{STH} was the length of each trial. The transformation $t_n^i \to s_n^i$ was performed without explicitly calculating $P(t)$ (see Figure 3). Let $\{t_1^i, \cdot, t_{N_i}^i\}$ be the set of N_i spike times during the ith trial, $i = 1, \cdot, N_{tr}$, where N_{tr} was the number of trials (see Figure 3A). The set of all spikes in all trials was collected into one set labeled by a dummy index j, $\{t_1, \cdot, t_M\}$, and sorted in increasing value, $t_{j(1)} \leq t_{j(2)} \leq \cdot \leq t_{j(k)} \leq \cdot \leq t_{j(M)}$, here, $M = \sum_{i=1}^{N_{tr}} N_i$ was the total number of spikes across all trials (see Figure 3B). The transformed time was

$$s_n^i = \frac{k(j)}{M} T_{STH}. \tag{3.2}$$

Here, $k(j)$ was the index of the jth spike time $t_j = t_n^i$ in the sorted set, i was the trial index, and n was the spike index (see Figure 3C). For the experimental data set, the rasterplot of transformed spike times $s_n^i = g(t_n^i)$ consisted of dots that were uniformly distributed in the plane, and the STH was constant (see Figures 4Aa and 4Ab).

3.2.3 Random Shufflings of the Interspike Intervals. The interspike intervals in transformed time were defined as $\nu_n^i = s_n^i - s_{n-1}^i$. The 0th spike time was defined as $s_0^i = 0$, so that there are as many intervals as there are spikes; however, s_0^i was not included in the analysis (and was not shown in the graphs). The intervals ν_n^i were randomly shuffled across trials, and the new spike times were then obtained from the shuffled intervals $\hat{\nu}_n^i$ as, $\hat{s}_n^i = \sum_{j=1}^n \hat{\nu}_j^i$ (j is a dummy index). This procedure was repeated to obtain

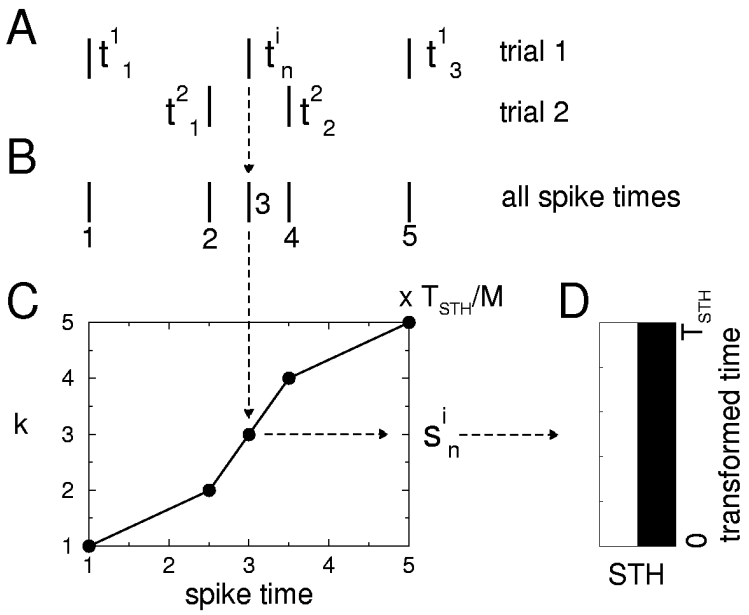

Figure 3: Procedure to obtain a homogeneous renewal process. (A) Spike trains obtained on different trials. t_n^i was the nth spike time in the ith trial. (B) Spike times across all trials were combined into one set and sorted from low to high values (the index is indicated below the ticks). (C) The index k in the ordered set as a function of the spike time. The new spike time was $s_n^i = kT_{STH}/M$ and took values between 0 and T_{STH}. (D) The resulting spike-time histogram was time independent.

N_s different independent realizations of the corresponding renewal process. These realizations will be referred to as surrogate spike trains. One realization generated from the experimental data set is shown in Figure 4B. As mentioned before, the STH (see Figure 4Bb) is reduced compared with the STH of the original spike trains (see Figure 4Ab) near the beginning and end of the stimulus presentation. The structure of the transformed interspike intervals, ν_n^i (see Figure 4Ac) is different from the one obtained from the surrogates (see Figure 4Bc). That difference will be quantified next.

3.2.4 Test Statistics for Renewal Processes. For a temporally modulated renewal process, all the time dependence can be removed by making the transformation $t \to s$. However, if there was nonrenewal structure present in the original spike trains, there could still be a time dependence. Here, we focus on the time dependence of interspike intervals. The time axis was divided in N_ν discrete bins of width Δs. The mean $\nu(m)$ was calculated of

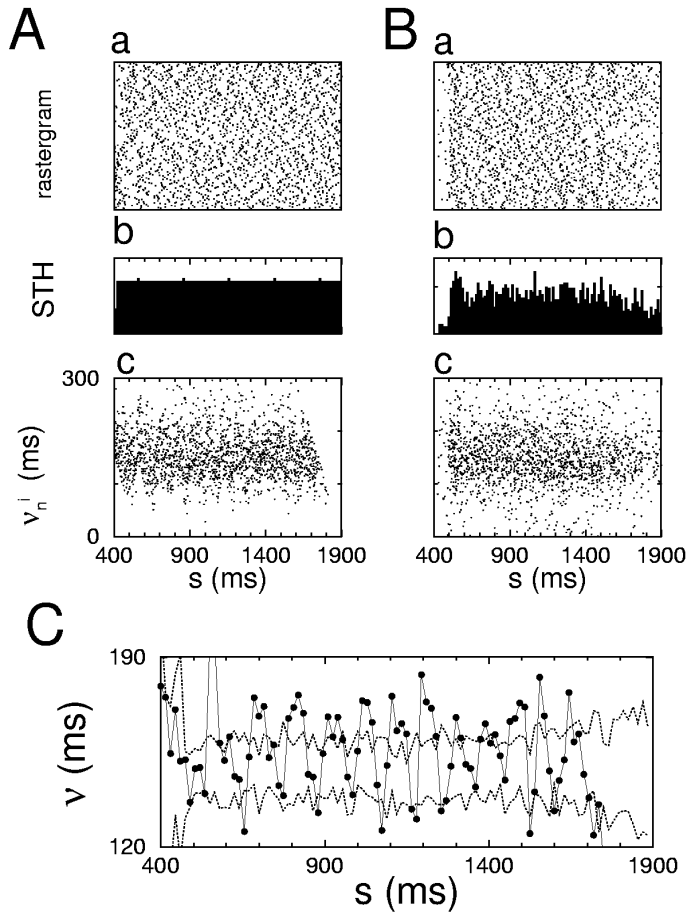

Figure 4: Spike trains of cortical neurons did not form a renewal process. Time was rescaled according to $t \to s = g(t)$, such that the spike-time probability distribution—the spike-time histogram—is independent of time s (see text). In A, rescaled spike trains from Figure 2 and in B renewal spike trains obtained by randomly shuffling the interspike intervals from A across different trials were shown. In each row were plotted (a) the rastergram of rescaled spike times $s_n^i = g(t_n^i)$ (t_n^i is the nth spike time in the ith trial), (b) the spike-time histogram of rescaled spike times, and (c) the rescaled interspike intervals, $\nu_n^i = s_n^i - s_{n-1}^i$, versus s_{n-1}^i. (C) The starting time of each interval was binned (bin width was 1.5 ms), and the average interval $\nu(s)$ was determined for each bin (solid line with filled circles). Confidence intervals, $\hat{\nu} \pm \hat{\sigma}$ (dashed lines), were determined for a renewal process with the same interval distribution using 20 random shufflings (one example was shown in B). The experimental spike trains were significantly different from renewal spike trains, $\chi^2 = 2.8$ for the bins between 700 and 1600 ms at significance $p < 10^{-11}$.

all ν_n^i with starting points s_{n-1}^i that fell in the mth bin, $(m-1)\Delta s \leq s_{n-1}^i < m\Delta s$. The same procedure was performed on all of the N_s surrogate spike trains, $l = 1, \ldots, N_s$, yielding $\hat{\nu}^l(m)$. The mean and standard deviation were subsequently determined:

$$\hat{\nu}(m) = \frac{1}{N_s} \sum_{l=1}^{N_s} \hat{\nu}^l(m),$$

$$\hat{\sigma}(m) = \sqrt{\frac{1}{N_s - 1} \sum_{l=1}^{N_s} (\hat{\nu}^l(m) - \hat{\nu}(m))^2}.$$

The original spike trains are nonrenewal when $\nu(m)$ lies outside the confidence interval given by $\hat{\nu}(m) - \hat{\sigma}(m)$ and $\hat{\nu}(m) + \hat{\sigma}(m)$ of the equivalent renewal proces. The test statistic was:

$$\chi^2 = \frac{1}{N_\nu} \sum_{m=1}^{N_\nu} \left(\frac{\nu(m) - \hat{\nu}(m)}{\hat{\sigma}(m)} \right)^2. \tag{3.3}$$

The hypothesis that the spike train was generated by a renewal process was rejected when $p = 1 - C(N_\nu \chi^2, N_\nu)$ was smaller than a prescribed critical value. Here we assume that $C(N_\nu \chi^2, N_\nu)$ was the cumulative χ^2 probability distribution with N_ν degrees of freedom (Abramowitz & Stegun, 1974; Larsen & Marx, 1986; Press et al., 1992). In the following, the continuous notation $\nu(s)$ will be used instead of $\nu(m)$.

The confidence intervals for the experimental recordings are shown together with $\nu(s)$ in Figure 4C. For the original process, the interval distribution depended on time, whereas for the renewal process, it did not depend on time. The χ^2 statistic based on $N_s = 20$ surrogates was $\chi^2 = 2.8$ for 60 degrees of freedom. Hence, the difference was highly significant, $p < 10^{-11}$, and the discharge produced by the neuron was not a renewal process.

3.3 Attractor Reliability. The attractor reliability was calculated in three steps. First, events were found (see Figure 5). Second, the binary representation for each trial was determined (see Figure 7). Third, the entropy of the distribution of spike trains was calculated (see Figure 8). The procedure was applied to the experimental data from Figure 2 and surrogate spike trains.

3.3.1 Determination of Events. An event was defined as a spike time that occurred across multiple trials. The following algorithm can be used to determine the spike times that are part of a given event. All spike times were combined in one set and ordered from low to high values, yielding $\{t_1, \ldots, t_k, \ldots, t_M\}$. Here, k denotes the index of the spike time in the ordered sequence. The spike time versus k curve had a steplike structure

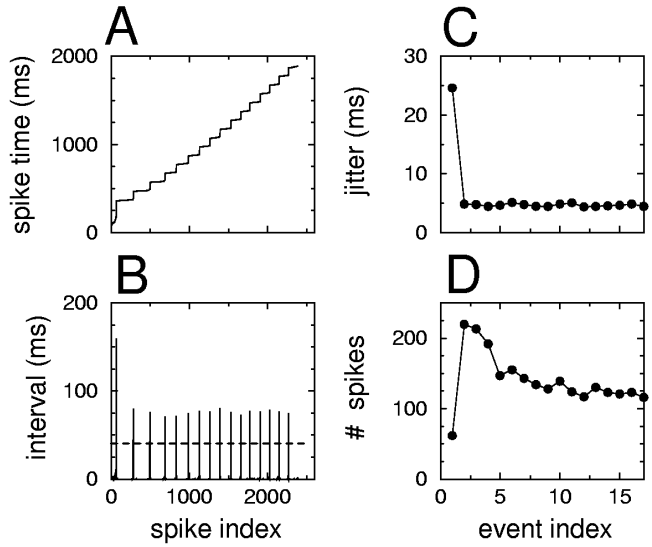

Figure 5: Procedure for determining events in the rastergram. Data from cortical neuron in Figure 2. (A) All the spike times were combined into one set and were ordered increasing from left to right. (B) The first difference of the sorted spike times was thresholded (threshold was 40 ms, dashed line). Spike times between two consecutive upward threshold crossings constituted an event. (C) The standard deviation of the spike times in an event and (D) the number of spikes in an event plotted as a function of event index.

(see Figure 5A). A step was formed by spike times with similar values and corresponded to an event. Between events and steps, spike times changed quickly. The first difference of the ordered spike times, $\tau_k = t_{k+1} - t_k$, was small within an event (roughly, the jitter divided by the number of trials) and large between different events, when k was part of one event and $k+1$ was part of the other event. Hence, the time series τ_k consisted of large values separated by many small values (see Figure 5B).

The set of k values, $k_1, \ldots, k_E$, where τ_{k-1} crossed a threshold from below was determined. E is the number of events that were detected. For $e = 1, \ldots, E$, event e consisted of $\{t_{k_{e-1}}, \ldots, t_{k_e-1}\}$ with the definition $k_{e=0} = 1$. The number of spikes in an event was $N_e = k_e - k_{e-1}$. The spike-time jitter (called "precision" in Mainen & Sejnowski (1995)) was the standard deviation of all the spike times in a given event e,

$$\sigma_e = \sqrt{\left(\frac{1}{N_e}\sum_{j=k_{e-1}}^{k_e-1} t_j^2\right) - t_e^2}, \tag{3.4}$$

$$t_e = \frac{1}{N_e}\sum_{j=k_{e-1}}^{k_e-1} t_j.$$

The threshold should be chosen such that the number of spikes per event is large (but smaller than the number of trials) and the spike-time jitter is small. For the experimental data, we used a threshold of 40 ms, and 17 events were obtained. The spike-time jitter was approximately 5 ms for all events except the first (see Figure 5C). The number of spikes in an event decreased from a maximum of 220 (all trials) at the second event to approximately 110 (half the number of trials) in the last event (see Figure 5D). The STH-based reliability was

$$R_{STH} = \frac{1}{EN_{tr}} \sum_{e=1}^{E} N_e. \tag{3.5}$$

Here, N_{tr} was the number of trials and E was the number of events, as before. For the experimental data, $R_{STH} \approx 0.64$.

3.3.2 Determination of the Binary Representation. Each trial i was then characterized by a binary value,

$$X^i = \sum_{e=1}^{E} n_e^i 2^{E-e}. \tag{3.6}$$

Here, $n_e^i = 1$ when there was a spike during event e on the ith trial and $n_e^i = 0$ otherwise. Binary numbers were also associated with subsets of the spike trains, X_{bL}^i for the events $e = b, \ldots, b + L - 1$ of trial i:

$$X_{bL}^i = \sum_{j=1}^{L} n_{b+j-1}^i 2^{L-j}. \tag{3.7}$$

Here, j was a dummy index, and L was the word length. It follows that $X^i = X_{1E}^i$.

3.3.3 Surrogate Spike Trains. Surrogate spike trains were constructed to compare the experimental spike trains with the equivalent Poisson process. Spike times could be randomly shuffled across trials, as before. However, in that case, there could be more than one spike time during an event on a given trial. This resulted in an ambiguous binary representation that could be resolved by, for instance, using ternary numbers. However, a different Poisson-like surrogate was constructed instead by randomly permuting spike times across trials for each event separately. Let $\{t_{n_i}^i\}$ be all the spike times during event e, and denote the absence of a spike during a given trial i by - (n_i is the index of the spike in the ith trial that is part of event e). Then the spike times during an event could be, for instance, $\{t_{n_1}^1, \text{-}, \text{-}, t_{n_4}^4\}$ (see Figure 6A). A surrogate obtained by a random permutation would be, for instance, $\{t_{n_4}^4, t_{n_1}^1, \text{-}, \text{-}\}$ (see Figure 6B). Consider the case that the original

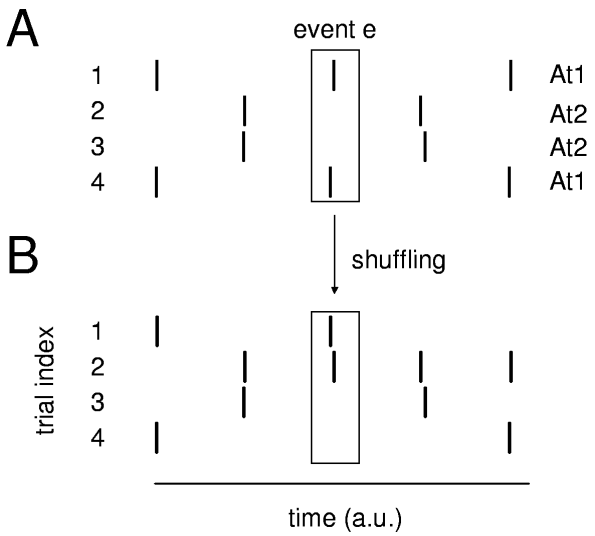

Figure 6: Procedure to generate Poisson-like surrogate spike trains. (A) Original set of spike trains: on trials 1 and 4, attractor 1 (At1) and on trials 2 and 3, attractor 2 (At2) was reached. (B) The spike times in each event e were randomly shuffled across trials. The resulting spike trains no longer resembled the attractors.

spike trains had deterministic structure and were either one of two attractor spike trains (see Figure 6A). The above procedure then breaks up the attractor spike trains and removes the non-Poisson structure (see Figure 6B). Hence, by comparing, for instance, the binary representation of the original and surrogate spike trains, the amount of non-Poisson structure can be assessed.

3.3.4 Binary Representation of Experimental Spike Trains. A transient was discarded, and binary representations were calculated based on 10 events starting from the eighth spike, yielding $X^i_{8,10}$ (see Figure 7Aa). The trials were sorted according to the binary representation starting from the lowest value. There were two plateaus in the $X_{8,10}$ versus index graph (indicated by * in Figure 7Aa). Each of the $X_{8,10}$ values were obtained on 6% of the trials (* in Figure 7Ab). These spike sequences corresponded to the two 1:2 mode-locking attractors: the neuron produced spikes on either the odd cycles or only on the even cycles. On a larger number of trials, neurons were on these attractors for a shorter duration, which led to a triangle-like structure in the rastergram (see Figure 7Ac).

The same analysis was performed on surrogate spike trains, and the results are shown in Figure 7B. There were no plateaus in the $X_{8,10}$ versus index graph (see Figure 7Ba). Each $X_{8,10}$ occurred with approximately equal

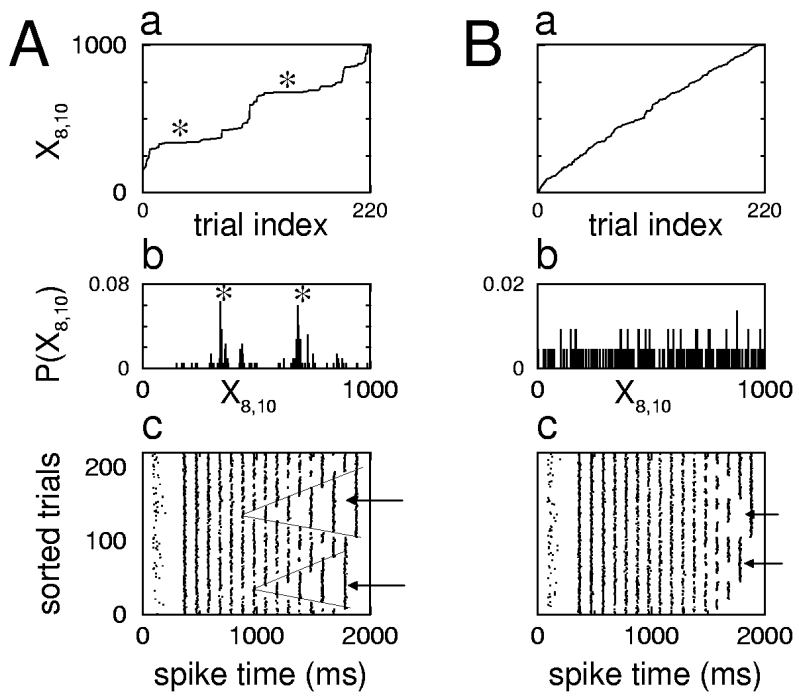

Figure 7: Cortical neuron spike trains had a deterministic structure not present in Poisson-like spike trains. Each trial was assigned a binary representation as described in the text. In *A*, the original data from Figure 2 and in *B* spike trains obtained by randomly shuffling spike times of each event across trials were used. In each row, (a) binary representation $X_{8,10}$, (b) distribution of $X_{8,10}$ values across 220 trials, and (c) rastergram with trials ordered according to value of $X_{8,10}$, increasing from bottom to top. The stars (*) and the arrows (←) are described in the text.

probability (see Figure 7Bb). The length of time that a neuron spent on an attractor on a given trial was reduced compared to the original spike trains. In particular, the triangle-like structure in the ordered rastergram was not present (see Figure 7Bc). The difference between the two sets of spike trains is quantified next using the entropy.

3.3.5 Entropy of Spike Trains. Let $P_{bL}(X_{bL})$ be the probability of obtaining a trial with binary representation X_{bL}. It was estimated using a finite number of trials by determining all the distinct words X_{bL} and counting how often each of them occurred across trials. The count was then normalized by the number of trials to obtain a probability. An example was shown in

Figure 7Ab. The entropy of this distribution was

$$S_{bL} = -\sum_{X_{bL}} P(X_{bL}) \log_2 P(X_{bL}). \tag{3.8}$$

The entropy was then averaged over all allowed b values,

$$S_L = \frac{1}{E-L+1} \sum_{b=1}^{E-L+1} S_{bL}. \tag{3.9}$$

The entropy of the distribution of the binary representation for the whole trial length was $S = S_{1E}$,

$$S = -\sum_{X} P(X) \log_2 P(X). \tag{3.10}$$

The attractor reliability was defined as

$$R_a = 2^{-S}. \tag{3.11}$$

The attractor probability can be interpreted as the inverse of the number of different X^i (this would be exact if each X value occurred with equal probability).

3.3.6 Entropy of Surrogate Spike Trains. The entropy of the surrogates was estimated as the mean of S and S_L over independent surrogates. An analytical expression for the entropy of the surrogate spike train was calculated. Let the probability of obtaining a spike during event e on a given trial be p_e. The probability for a trial X with event occupation numbers $\{n_e\} = \{n_1, \ldots, n_E\}$ was

$$P(X) = \prod_{e=1}^{E} p_e^{n_e} (1-p_e)^{(1-n_e)}, \tag{3.12}$$

and the entropy of this distribution was

$$\begin{aligned} S &= -\sum_{\{n_e\}} P(X) \log_2 P(X) \\ &= -\sum_{e=1}^{E} [p_e \log_2 p_e + (1-p_e) \log_2 (1-p_e)]. \end{aligned} \tag{3.13}$$

Here $\sum_{\{n_e\}} = \sum_{n_1=0}^{1}, \ldots, \sum_{n_E=0}^{1}$ is the sum over all possible combinations of event occupation numbers. The final result followed since the entropy of a

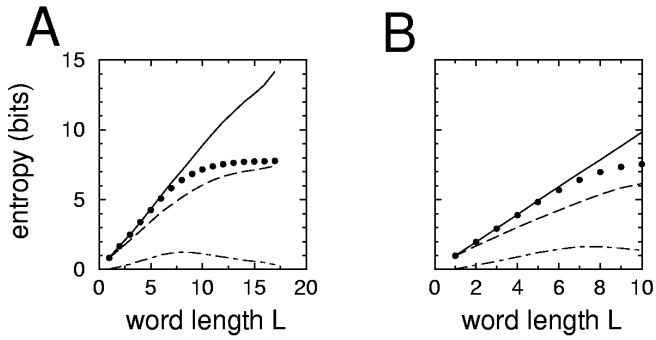

Figure 8: The binary representations of the experimental spike trains were less variable than those of Poisson-like processes. Binary representations with word length L were determined for each trial, and the entropy was calculated as described in the text, starting from (A) the first spike time and (B) the eighth spike time. The entropy of experimental spike trains (dashed line), entropy of surrogate spike trains averaged over 10 random shufflings (filled circles), their difference (dot-dashed line), and the analytical result for the entropy of surrogate spike trains (solid line) are shown. The standard deviation of the surrogate spike-train entropy was of the order of the circle size.

product of probability density functions is the sum over the entropy of each probability distribution. The analytical entropy of the surrogate spike trains increased linearly with E. However, for N_{tr} trials, the maximum entropy was $\log_2 N_{tr}$; hence, the analytical limit may not be reached if too few trials are available for analysis.

The entropy of the experimental spike trains was determined as a function of the word length L (see Figure 8). The entropy of the surrogate spike trains was calculated in two ways: first by determining the mean entropy of 10 surrogate spike trains and then analytically by using equation 3.13. The probability p_e of a spike during event e was estimated as the number of spikes during that event in the original data divided by the number of trials. The entropy of the original spike train was always lower than that of the surrogate spike train. This implies that there was additional structure present in the original spike train that was not present in surrogate spike trains.

The entropy of surrogate spike trains started to differ from the analytical result at $L = 8$ (see Figure 8A). This indicated that the number of trials was not large enough to sample the probability distribution of the binary representations adequately. Initially, the difference between the entropy of the original and surrogate spike trains increased with L, but when the sampling effects occurred, it decreased. As a result, the difference had a maximum that was a sampling artifact (dot-dashed line in Figure 8A). Similar results were obtained for spike trains starting from the eighth event (see Figure 8B).

3.4 Simulated Spike Trains. The statistical test for renewal behavior and the procedure to determine the attractor reliability were also applied to simulated spike trains. As an example, the spike trains of a noisy integrate-and-fire neuron that was 1:2 entrained to a sinusoidal drive were examined (see Figure 9A). The entropy of the original spike trains was lower than the surrogate spike trains (see Figure 9B). The spike trains were nonrenewal; the probability of obtaining the same $\nu(s)$ (see Figure 9C) from a renewal spike train was zero. The attractor reliability was $R_a = 2^{-3.74} \approx 0.075$.

4 Discussion

Other statistical tests to determine whether neural spike trains form a renewal process have been proposed previously, such as the power ratio (Reich et al., 1998). The critical value of the power ratio depended on the interval distribution of the (rescaled) spike train. The test introduced here could be applied to all interval distributions, and its significance value was determined using the χ^2 distribution.

Oram, Wiener, Lestienne, and Richmond (1999) proposed a procedure to determine whether certain patterns of spikes in multiple unit recordings were present above chance levels (see also Abeles & Gat, 2001). The attractor reliability introduced here is related to this procedure since it assesses whether certain patterns—attractors—occur more often than expected for Poisson processes with the same spike-time histogram.

The number of distinct spike trains was determined using a simple procedure. This procedure succeeds if (1) the spike times within an event are sufficiently precise and (2) the distance between different events is larger than the spike-time jitter within an event. The latter condition is not always satisfied; for instance, in Poisson spike trains, spikes can occur with arbitrarily small interspike intervals. In real spike trains, there is always a minimum interspike interval equal to the refractory period. When either of the two conditions is not satisfied, a different method is required to separate spike trains into groups. Two alternatives are the k means method for clustering (Gershenfeld, 1999) or an algorithm based on spike metrics such as in Victor and Purpura (1996). To calculate the entropy of spike trains and compare to equivalent renewal processes, a sufficient number of trials was needed. We found 40 to 1000 trials to be sufficient for most analyses.

Recent experimental studies show that the amplitude of a postsynaptic conductance in response to a presynaptic action potential depends on the previous presynaptic spike times (Markram & Tsodyks, 1996; Abbott, Varela, Sen, & Nelson, 1997; Markram, Wang, & Tsodyks, 1998). As a result, synapses are sensitive to temporal correlations in input spike trains (Brenner, Strong, Koberle, Bialek, & de Ruyter van Steveninck, 2000; Eguia, Rabinovich, & Abarbanel, 2000; Tiesinga, 2001). For instance, a Poisson spike train and a periodic spike train with the same average rate will yield different postsynaptic amplitudes. When there are more distinct spike trains

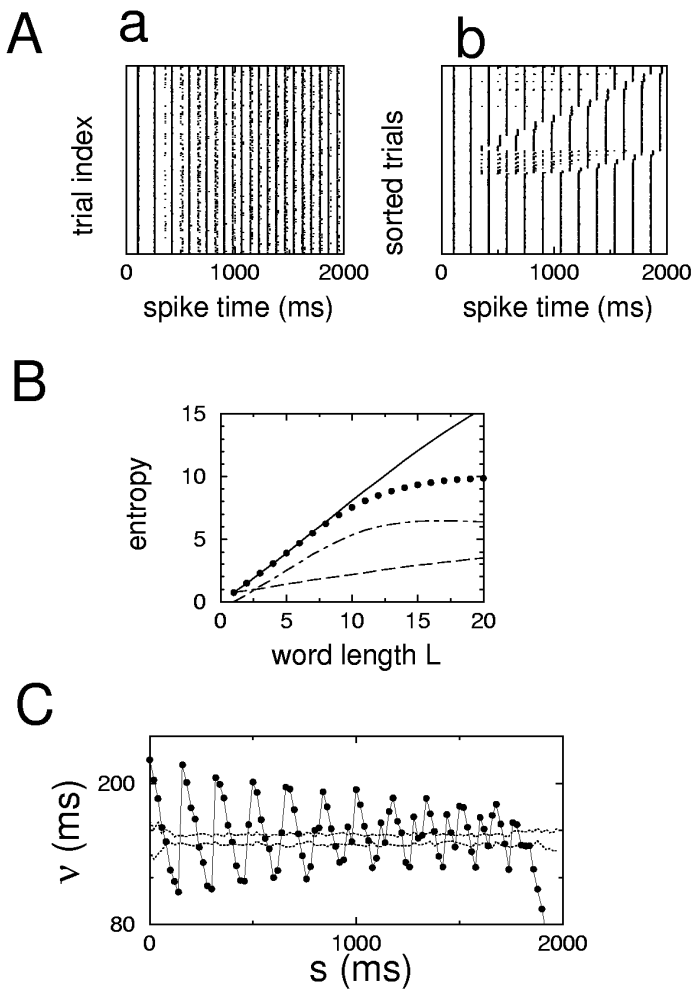

Figure 9: Spike trains obtained from a noisy integrate-and-fire neuron that was 1:2 entrained to a sinusoidal drive. The spike trains had deterministic structure and did not form a renewal process. (A) The rastergrams (a) in original order and (b) ordered on binary representation. (B) Spike-train entropy; curves were annotated as in Figure 8A. (C) Average rescaled interspike interval as a function of time with confidence intervals (notation as in Figure 4C). Model parameters were $I = 1.0, A = 0.17, D = 10^{-4}, T = 2$. To facilitate comparison to experimental data, time was rescaled by a factor of 40 ms.

across trials, the synaptic drive gets more variable. The attractor reliability is a measure for this type of synaptic variability. If the task of a neuron is to transmit information about its input into its output spike train, then this variability would be considered noise since it is not related to the input. The impact on a postsynaptic neuron of the unreliability-induced synaptic variability is not clear. Cortical neurons receive a large number of synaptic inputs from different cells (reviewed in Shadlen & Newsome, 1998) and synapses themselves are unreliable (Bekkers, Richerson, & Stevens, 1990; Allen & Stevens, 1994; Zador, 1998). This issue remains for further study.

A more immediate issue is how reliability depends on the characteristics of the driving input and the intrinsic neuronal dynamics. In previous experimental and theoretical studies (Mainen & Sejnowski, 1995; Nowak, Sanchez-Vives, & McCormick, 1997; Tang, Bartels, & Sejnowski, 1997; Hunter, Milton, Thomas, & Cowan, 1998; Warzecha, Kretzberg, & Egelhaaf, 1998, 2000; Cecchi et al., 2000; Kretzberg, Egelhaaf, & Warzecha, 2001; Fellous et al., 2001), it was shown, using a different reliability measure, that neurons fire unreliably in response to constant depolarizing current, but fire reliably when driven by inputs containing high-frequency components. The reliability measure introduced here forms part of a theoretical framework that allows for the systematic study of neuronal reliability. Elsewhere, we will investigate how the attractor reliability depends on the type and number of attractors and their bifurcation structure.

Acknowledgments

We thank Jack Cowan, Jorge José, Bruce Knight, Susanne Schreiber, and Peter Thomas for discussions and suggestions and Greg Horwitz, Arnaud Delorme, and the anonymous referees for comments that helped improve the presentation of the article. Some of the numerical calculations were performed at the High Performance Computer Center at Northeastern University. This work was partially funded by the Sloan-Swartz Center for Theoretical Neurobiology (P.T.) and the Howard Hughes Medical Institute (J.M.F., T.J.S.).

References

Abbott, L., Varela, J., Sen, K., & Nelson, S. (1997). Synaptic depression and cortical gain control. *Science, 275*, 220–224.

Abeles, M., & Gat, I. (2001). Detecting precise firing sequences in experimental data. *J. Neurosci. Methods, 107*, 141–154.

Abramowitz, M., & Stegun, I. (1974). *Handbook of mathematical functions*. New York: Dover.

Allen, C., & Stevens, C. (1994). An evaluation of causes for unreliability of synaptic transmission. *Proc. Natl. Acad. Sci., 91*, 10380–10383.

Bekkers, J., Richerson, G., & Stevens, C. (1990). Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proc. Natl. Acad. Sci., 87*, 5359–5362.

Brenner, N., Strong, S., Koberle, R., Bialek, W., & de Ruyter van Steveninck, R. (2000). Synergy in a neural code. *Neural Comput., 12*, 1531–1552.

Cecchi, G., Sigman, M., Alonso, J., Martinez, L., Chialvo, D., & Magnasco, M. (2000). Noise in neurons is message dependent. *Proc. Natl. Acad. Sci., 97*, 5557–5561.

Eguia, M., Rabinovich, M., & Abarbanel, H. (2000). Information transmission and recovery in neural communications channels. *Phys. Rev. E, 62*, 7111–7122.

Fellous, J.-M., Houweling, A., Modi, R., Rao, R., Tiesinga, P., & Sejnowski, T. (2001). The frequency dependence of spike timing reliability in cortical pyramidal cells and interneurons. *J. Neurophys., 85*, 1782–1787.

Gershenfeld, N. (1999). *The nature of mathematical modeling*. Cambridge: Cambridge University Press.

Hansel, D., Mato, G., Meunier, C., & Neltner, L. (1998). On numerical simulations of integrate-and-fire neural networks. *Neural Comput., 10*, 467–483.

Hunter, J., Milton, J., Thomas, P., & Cowan, J. (1998). Resonance effect for neural spike time reliability. *J. Neurophysiol., 80*, 1427–1438.

Jensen, R. (1998). Synchronization of randomly driven nonlinear oscillators. *Phys. Rev. E, 58*, 6907–6910.

Kretzberg, J., Egelhaaf, M., & Warzecha, A. (2001). Membrane potential fluctuations determine the precision of spike timing and synchronous activity: A model study. *J. Comput. Neurosci., 10*, 79–97.

Larsen, R., & Marx, M. (1986). *An introduction to mathematical statistics and its applications*. Englewood Cliffs, NJ: Prentice Hall.

Mainen, Z., & Sejnowski, T. (1995). Reliability of spike timing in neocortical neurons. *Science, 268*, 1503–1506.

Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature, 382*, 807–810.

Markram, H., Wang, Y., & Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci., 95*, 5323–5328.

Nowak, L., Sanchez-Vives, M., & McCormick, D. (1997). Influence of low and high frequency inputs on spike timing in visual cortical neurons. *Cereb. Cortex, 7*, 487–501.

Oram, M., Wiener, M., Lestienne, R., & Richmond, B. (1999). Stochastic nature of precisely timed spike patterns in visual system neuronal responses. *J. Neurophysiol., 81*, 3021–3033.

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical recipes*. Cambridge: Cambridge University Press.

Reich, D., Victor, J., & Knight, B. (1998). The power ratio and the interval maps: Spiking models and extracellular recordings. *J. Neurosci., 18*, 10090–10104.

Shadlen, M., & Newsome, W. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci., 18*, 3870–3896.

Strogatz, S. (1994). *Nonlinear dynamics and chaos*. Reading, MA: Addison-Wesley.

Tang, A., Bartels, A., & Sejnowski, T. (1997). Effects of cholinergic modulation on responses of neocortical neurons to fluctuating input. *Cereb. Cortex, 7*, 502–509.

Tiesinga, P. (2001). Information transmission and recovery in neural communications channels revisited. *Phys. Rev. E, 64*, 012901: 1–4.

Victor, J., & Purpura, K. (1996). Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J. Neurophysiol., 76*, 1310–1326.

Warzecha, A., Kretzberg, J., & Egelhaaf, M. (1998). Temporal precision of the encoding of motion information by visual interneuron. *Curr. Biol., 8*, 359–368.

Warzecha, A., Kretzberg, J., & Egelhaaf, M. (2000). Reliability of a fly motion-sensitive neuron depends on stimulus parameters. *J. Neurosci., 20*, 8886–8896.

Zador, A. (1998). Impact of synaptic unreliability on the information transmitted by spiking neurons. *J. Neurophys., 79*, 1219–1229.

Received July 25, 2001; accepted January 4, 2002.